ACHIEVING NUTRITIONAL SECURITY AND FOOD SAFETY THROUGH GENOMICS-BASED BREEDING OF CROPS

EDITED BY: Reyazul Rouf Mir, Sachiko Narita Isobe, Manish Kumar Pandey and Ajay Kumar

PUBLISHED IN: Frontiers in Nutrition and Frontiers in Genetics

frontiers Research Topics

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88966-611-9
DOI 10.3389/978-2-88966-611-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ACHIEVING NUTRITIONAL SECURITY AND FOOD SAFETY THROUGH GENOMICS-BASED BREEDING OF CROPS

Topic Editors:
Reyazul Rouf Mir, Sher-e-Kashmir University of Agricultural Sciences and Technology, India
Sachiko Narita Isobe, Kazusa DNA Research Institute, Japan
Manish Kumar Pandey, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India
Ajay Kumar, North Dakota State University, United States

Citation: Mir, R. R., Isobe, S. N., Pandey, M. K., Kumar, A., eds. (2021). Achieving Nutritional Security and Food Safety Through Genomics-Based Breeding of Crops. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-611-9

Table of Contents

05 ***Editorial: Achieving Nutritional Security and Food Safety Through Genomics-Based Breeding of Crops***
Reyazul Rouf Mir, Ajay Kumar, Manish K. Pandey and Sachiko Narita Isobe

09 ***Nutritional Security in Drylands: Fast-Track Intra-Population Genetic Improvement for Grain Iron and Zinc Densities in Pearl Millet***
Mahalingam Govindaraj, Kedar Nath Rai, Anand Kanatti, Aluri Sambasiva Rao and Harshad Shivade

17 ***An Improved Enzyme-Linked Immunosorbent Assay (ELISA) Based Protocol Using Seeds for Detection of Five Major Peanut Allergens Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8***
Arun K. Pandey, Rajeev K. Varshney, Hari K. Sudini and Manish K. Pandey

28 ***Zinc and Iron Nutrition Status in the Philippines Population and Local Soils***
Alvin D. Palanog, Mark Ian C. Calayugan, Gwen Iris Descalsota-Empleo, Amery Amparado, Mary Ann Inabangan-Asilo, Emily C. Arocena, Pompe C. Sta. Cruz, Teresita H. Borromeo, Antonio Lalusin, Jose E. Hernandez, Cecilia Acuin, Russell Reinke and B. P. Mallikarjuna Swamy

45 ***Association Mapping for 24 Traits Related to Protein Content, Gluten Strength, Color, Cooking, and Milling Quality Using Balanced and Unbalanced Data in Durum Wheat [*Triticum turgidum *L. var.* durum *(Desf).]***
Marina Johnson, Ajay Kumar, Atena Oladzad-Abbasabadi, Evan Salsman, Meriem Aoun, Frank A. Manthey and Elias M. Elias

63 ***Down-Regulation of* CYP79A1 *Gene Through Antisense Approach Reduced the Cyanogenic Glycoside Dhurrin in [*Sorghum bicolor *(L.) Moench] to Improve Fodder Quality***
Arun K. Pandey, Pusuluri Madhu and Basrur Venkatesh Bhat

75 ***Combined QTL-Seq and Traditional Linkage Analysis to Identify Candidate Genes for Purple Skin of Radish Fleshy Taproots***
Tongjin Liu, Jinglei Wang, Chunhui Wu, Youjun Zhang, Xiaohui Zhang, Xiaoman Li, Haiping Wang, Jiangping Song and Xixiang Li

86 ***Single Nucleotide Polymorphisms in Starch Biosynthetic Genes Associated With Increased Resistant Starch Concentration in Rice Mutant***
Selvakumar Gurunathan, Bharathi Raja Ramadoss, Venkataramana Mudili, Chandranayaka Siddaiah, Naveen Kumar Kalagatur, Jutti Rajendran Kannan Bapu, Chakrabhavi Dhananjaya Mohan, Abdulaziz A. Alqarawi, Abeer Hashem and Elsayed Fathi Abd_Allah

96 ***Hypoallergen Peanut Lines Identified Through Large-Scale Phenotyping of Global Diversity Panel: Providing Hope Toward Addressing One of the Major Global Food Safety Concerns***
Arun K. Pandey, Hari K. Sudini, Hari D. Upadhyaya, Rajeev K. Varshney and Manish K. Pandey

111 ***The Genetic Basis and Nutritional Benefits of Pigmented Rice Grain***
Edwige Gaby Nkouaya Mbanjo, Tobias Kretzschmar, Huw Jones, Nelzo Ereful, Christopher Blanchard, Lesley Ann Boyd and Nese Sreenivasulu

129 ***Zinc Biofortified Rice Varieties: Challenges, Possibilities, and Progress in India***
D. Sanjeeva Rao, C. N. Neeraja, P. Madhu Babu, B. Nirmala, K. Suman, L. V. Subba Rao, K. Surekha, P. Raghu, T. Longvah, P. Surendra, Rajesh Kumar, V. Ravindra Babu and S. R. Voleti

142 ***Profiling of Nutraceuticals and Proximates in Peanut Genotypes Differing for Seed Coat Color and Seed Size***
Spurthi N. Nayak, Viresh Hebbal, Pushpa Bharati, Hajisab L. Nadaf, Gopalkrishna K. Naidu and Ramesh S. Bhat

157 ***Combining High Oleic Acid Trait and Resistance to Late Leaf Spot and Rust Diseases in Groundnut (*Arachis hypogaea *L.)***
Dnyaneshwar B. Deshmukh, Balram Marathi, Hari Kishan Sudini, Murali T. Variath, Sunil Chaudhari, Surendra S. Manohar, Ch V. Durga Rani, Manish K. Pandey and Janila Pasupuleti

172 ***Genome-Wide SNP Identification and Association Mapping for Seed Mineral Concentration in Mung Bean (*Vigna radiata *L.)***
Xingbo Wu, A. S. M. Faridul Islam, Naransa Limpot, Lucas Mackasmiel, Jerzy Mierzwa, Andrés J. Cortés and Matthew W. Blair

189 ***Association Mapping of Seed Quality Traits Under Varying Conditions of Nitrogen Application in* Brassica juncea *L. Czern & Coss***
Javed Akhatar, Mohini Prabha Singh, Anju Sharma, Harjeevan Kaur, Navneet Kaur, Sanjula Sharma, Baudh Bharti, V. K. Sardana and Surinder S. Banga

203 ***Enhancing β-Carotene Concentration in Parental Lines of CO6 Maize Hybrid Through Marker-Assisted Backcross Breeding (MABB)***
Senthil Natesan, Thirusenduraselvi Duraisamy, Bharathi Pukalenthy, Sarankumar Chandran, Jagadeeshselvam Nallathambi, Karthikeyan Adhimoolam, Dhasarathan Manickam, Vellaikumar Sampathrajan, Samuel Jeberson Muniyandi, Laishram Joykumar Meitei, Nepolean Thirunavukkarasu, Ganesan Kalipatty Nalliappan and Ravikesavan Rajasekaran

215 ***Genome-Wide SNP Discovery and Mapping QTLs for Seed Iron and Zinc Concentrations in Chickpea (*Cicer arietinum *L.)***
Syed Sab, Ramappa Lokesha, D. M. Mannur, Somasekhar, Kisan Jadhav, Bingi Pujari Mallikarjuna, Laxuman C, Sharanbasappa Yeri, Vinod Valluri, Prasad Bajaj, Annapurna Chitikineni, AnilKumar Vemula, Abhishek Rathore, Rajeev Kumar Varshney, I. Shankergoud and Mahendar Thudi

224 ***Enhancing the Nutritional Quality of Major Food Crops Through Conventional and Genomics-Assisted Breeding***
Kiran B. Gaikwad, Sushma Rani, Manjeet Kumar, Vikas Gupta, Prashanth H. Babu, Naresh Kumar Bainsla and Rajbir Yadav

frontiers in Nutrition

EDITORIAL
published: 04 February 2021
doi: 10.3389/fnut.2021.638845

Editorial: Achieving Nutritional Security and Food Safety Through Genomics-Based Breeding of Crops

Reyazul Rouf Mir[1]*, *Ajay Kumar*[2], *Manish K. Pandey*[3] and *Sachiko Narita Isobe*[4]

[1] Division of Genetics and Plant Breeding, Faculty of Agriculture, Sher-e-Kashmir University of Agricultural Sciences and Technology, Srinagar, India, [2] Depart of Plant Science, North Dakota State University, Fargo, ND, United States, [3] Genomics & Trait Discovery Genetic Gains Program, International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India, [4] Department of Frontier Research and Development, Kazusa DNA Research Institute, Kisarazu, Japan

Keywords: genomics, nutrition, food safety, QTLs, genes, molecular breeding, varieties

OPEN ACCESS

Edited and reviewed by:
Ahmed El-Sohemy,
University of Toronto, Canada

Correspondence:
Reyazul Rouf Mir
imrouf2006@gmail.com;
rrmir@skuastkashmir.ac.in

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Nutrition

Received: *07 December 2020*
Accepted: *18 January 2021*
Published: *04 February 2021*

Citation:
Mir RR, Kumar A, Pandey MK and Isobe SN (2021) Editorial: Achieving Nutritional Security and Food Safety Through Genomics-Based Breeding of Crops. Front. Nutr. 8:638845. doi: 10.3389/fnut.2021.638845

Editorial on the Research Topic

Achieving Nutritional Security and Food Safety Through Genomics-Based Breeding of Crops

The population of the world is continuously increasing and may exceed nine billion by 2050 (34% higher than today). To feed this increasing global population, the crop yields/food production must also increase by 70% by 2050 (1). Thus, the problem of hunger/malnutrition and demand for food security/healthier diets will be also on the rise (2). Although, we have achieved self-sufficiency in producing sufficient food (food quantity) in most countries in the world but the food produced is often deficient in essential nutrients (food quality) resulting in hidden hunger and malnutrition. Therefore, producing sufficient food is not a problem but producing quality food is a problem. The health of future generations will depend on our ability to produce and deliver food with enhanced nutrition for the growing population world-wide. The problem of micro-nutrient deficient food is a global problem now and this problem is seriously affecting people even in developed countries (1). Among nutrients that are most important and often deficient in diets include seed Zn, seed Fe and seed protein content (3, 4).

Conventional plant breeding methods have a long history of improving crop production, productivity, nutritional security, and food safety (5). Several nutrient dense crop varieties in different food crops have been released for cultivation world-wide. HarvestPlus in collaboration with different collaborating partners world-wide has developed several bio-fortified crops including Iron rich beans, iron rich pearl millet, vitamin A rich maize, zinc rich rice and zinc rich wheat and these crops have been released for cultivation in more than 30 countries (https://www.harvestplus.org/what-we-do/crops). In addition to improving nutrition security, the need of the hour is to also to address potential food safety issues related to allergenicity and toxicity of crop products prior to consumer consumption. The screening of crop germplasm for identification of genotypes with significantly reduced or null allergen content is cumbersome and therefore, conventional breeding methods and genomics assisted breeding (GAB) efforts toward hypoallergenic varieties have been undertaken in different crops including wheat, soybean, peanut, etc. (5).

The recent advances in genomic tools and techniques have revolutionized agriculture and changed the way food crop varieties are developed. A variety of genomics tools and techniques have been used in these crop improvement programs Such as high-throughput markers and marker-genotyping platforms, new genome mapping approaches, next-generation genome sequencing. Among genome mapping approaches, quantitative trait locus (QTL) mapping and genome-wide association mapping are two most important approaches being used for discovery of genes/QTLs for simple and complex quantitative traits (6). The advances in genome sequencing technologies have resulted in lowering the cost of sequencing by 100,000-fold since 2000 (7).

The availability of next generation sequencing, high-throughput genotyping and high-throughput phenotyping technologies have completely changed the scenario and have increased the speed of germplasm characterization, allele mining, gene/QTL discovery and their use in marker-assisted breeding(MAB) programs (8). The application of genomics tools and techniques have resulted in the development of next-generation of crop varieties with enhanced yield potential and nutrition that is very critical for food and nutrition security and safety. Several modern breeding approaches including marker-assisted selection (MAS), marker-assisted backcrossing (MABC), marker-assisted recurrent selection (MARS) and genomic selection (GS) have been used for development of improved varieties in almost all important crop plants (6).

The current research topic entitled "**Achieving Nutritional Security and Food Safety Through Genomics-Based Breeding of Crops**" has a total of 17 articles including 13 research articles and four review articles. The articles have been written by authorities in their respective research fields. The articles provide great knowledge about the progress made in genomics for nutrition security in crops like rice, wheat, pearl millet, sorghum, chickpea, groundnut, mung bean, brassica, and radish. The articles deal with research areas that included evaluation of crop germplasm for important traits and development of genetic resources, mapping of genes through quantitative trait locus (QTL) interval mapping, QTL sequencing, association mapping including genome-wide association study (GWAS), QTL sequencing and genomics-assisted breeding (GAB) that included marker-assisted selection (MAS)/marker-assisted back crossing (MABC). Keeping in view the vast range of research areas covered in different manuscripts, we will first discuss the manuscripts that involved germplasm evaluation and development of genetic resources followed by manuscripts involving gene discovery through QTL mapping and association mapping and finally use of these QTLs/genes in development of improved cultivars through GAB approaches including MAS and MABC.

Peanut is one of the most important grain legume crops in the world and peanut allergy is considered serious health concern that affects more than 1% world's population. The allergy caused by peanut consumption is sometimes life-threatening that also affects life quality of allergic patients/their whole families. To alleviate this serious life threatening problem, development and consumption of hypoallergenic peanuts is considered the best solution. Therefore, efforts have been made and two important publications have been published in this research topic dealing with peanut allergy (Pandey, Varshney et al.; Pandey, Sudini et al.). The comprehensive manuscript by Pandey, Varshney et al. reported antibody based ELISA protocol for precisely quantifying and standardizing five major allergen proteins (Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8) in peanut seeds. The deployment of this protocol will allow large scale screening of peanut germplasm for selection of genotypes with minimum allergenicity to ensure food safety. In another study in peanut, Pandey, Varshney et al. used monoclonal antibody-based ELISA protocol to quantify five major allergens by phenotyping of highly diverse peanut germplasm panel. The analysis of trait data led to the identification of peanut lines with less allergenicity. This type of very useful peanut genetic resource will prove useful in future peanut improvement programs aimed at ensuring food safety while consuming peanuts. Similarly, in rice, one of the most important studies was conducted by Gurunathan et al. to develop rice mutant with enhanced resistant starch (RS) through the use of gamma rays as a mutagen. The development of resistant starch mutant will prove useful in human health and nutrition by controlling glucose metabolism. Further characterization of this mutant led to the identification of 31 sequence variants in six (GBSSI, SSI, SSIIa, SSIIIa, SBEIa, and SBEIIb) starch biosynthetic genes (Gurunathan et al.). In addition, three deleterious mutations/variants were also discovered in three starch biosynthetic genes including GBSSI, SSIIa, and SSIIIa with the potential to increase RS content in rice. In sorghum, the manuscript by Pandey, Madhu et al. showed the effectiveness of using antisense CYP79A1 strategy to produce transgenic sorghum plants with low cyanogenic potential, thus making sorghum safer for cattle feed. The desired sorghum plants having less cyanogenic potential were obtained by down-regulation of a key enzyme of dhurrin biosynthesis pathway "**cytochrome P450 CYP79A1**."

Mapping of genes/QTLs for quantitative traits through QTL mapping and association mapping including genome-wide association studies (GWAS) is one of the most important subject areas of crop research. One of the important research articles published in this research topic on QTL mapping involving use of bi-parental mapping population segregating for seed micronutrients like seed Zn and seed Fe concentration is by Sab et al. The authors identified genes/QTLs for seed Zn and seed Fe concentration. For seed Fe concentration, eleven (11) QTLs were identified, while as for seed Zn concentration, eight (8) QTLs were identified. Among these identified QTLs, three QTLs were co-located in the "QTL-hotspot" region on linkage group-04 that also harbors several drought tolerance-related QTLs identified and reported in earlier studies. The knowledge gained in this study could be a helpful in the development of bio-fortified varieties in chickpea through genomics assisted breeding (GAB) approaches.

The other recently emerged mapping approach "**association mapping**" was used by Johnson et al. by involving use of high density SNP markers on a set of 243 durum wheat cultivars/advanced breeding lines developed over a period of 20 years. The results of this genome-wide association study (GWAS) revealed significant marker-trait associations for 24 nutritional/quality traits in durum wheat. The study identified several candidate markers that will help in durum wheat quality improvement through molecular breeding programs in future. Similarly, another study by Akhatar et al. reported use of genome-wide association study (GWAS) for discovery of genes/QTLs and dissect the genetics of oil, protein, and glucosinolates in Indian mustard (*Brassica juncea*). The study involved use of genotyping by sequencing (GBS) approach to identify genes for oil, protein, and glucosinolates. The authors further used meta-analysis of GWAS datasets to improve the

power to recognize associations and identified common SNPs for oil and protein contents. In addition, 21 orthologs of the functional candidate genes related to the biosynthesis of oil, protein, and glucosinolates were also identified. The results of this study will prove useful in improving seed quality traits in *B. juncea* through marker-assisted selection (MAS) programs. In one of the important legume crops "**Mung bean**," genome-wide association study (GWAS) were conducted by Wu et al. using a set of 6,486 SNPs markers identified through genotyping by sequencing (GBS) approach. The analysis of trait data and GBS data led to the identification of 43 significant marker-trait associations (MTAs) for seed calcium, iron, potassium, manganese, phosphorous, sulfur or zinc concentrations. The identified MTAs were distributed across 35 genomic regions explaining on average 22% of the variation for each seed nutrient. In another gene/QTL mapping study, Nayak et al. used single marker analysis (SMA) to identify significant marker-trait associations (MTAs). The study involved use of a set of 60 genotypes of peanut with different kernel color and size and data was recorded on nutritional traits including moisture, fat, ash, crude protein, crude fiber, carbohydrate content, and nutraceuticals including total polyphenol content and total antioxidant activity. The association analysis of trait data with the genotypic data led the identification of 75 major MTAs for most of the nutritional traits. The important markers found associated with nutritional traits and other important yield related traits will prove useful in genomics-assisted breeding (GAB) programs for developing nutrient-rich peanuts.

One of the important manuscript published in this research topic by Liu et al. reports combined use of most recently emerged gene mapping method "**QTL-Seq**" and **traditional linkage analysis** to identify candidate genes for purple skin of radish fleshy taproots. The study reported identification of one dominant gene "*Rsps*" followed by its fine mapping to a 238.51-kb genomic region containing 18 genes. These findings provide insight into the complex anthocyanin biosynthesis regulation in radish and information for molecular breeding to improve the anthocyanin content and appearance of radish taproots.

One of the ultimate and most important aims of germplasm evaluations/development of genetic resources and gene discovery programs is to identify best/suitable genes for their use in crop breeding programs to develop next-generation of crop varieties for cultivation by farmers. In this direction, some very good publications have been published in this research topic. For instance, in groundnut, Deshmukh et al. reported the use of marker-assisted selection (MAS) for introgression of high oleic content, resistance to late leaf spot and rust into popular groundnut cultivar "Kadiri 6." The MAS was exercised for the alleles of target traits through the use of linked allele-specific SSRs and SNP markers. The study reported sixteen homozygous plants with high oleic, late leaf spot and rust resistance alleles. Similarly, Natesan et al. also used marker-assisted backcross breeding (MABB) approach in maize for increasing β-carotene concentration in parental genotypes of maize hybrid "**CO6**." The MABC was exercised by involving the use of β-carotene gene "***crtRB1***" that resulted in the development of six improved maize lines with high recurrent parent genome recovery (90.24–92.42%), good agronomic performance and high β-carotene concentration (7.056–9.232 μg/g). The improved maize genotypes developed in this study will prove useful for ensuring food and nutritional security in the world. Govindaraj et al. reported fast-track intra-population genetic improvement for grain iron and zinc densities in pearl millet by involving three open-pollinated varieties. Progeny selection was found effective for Fe density and selection for Fe was found highly associated with the improvement of Zn density as well. Therefore, selection for one nutrient results in improvement of both the nutrients without yield penalty. This fast-track approach involves direct selection for additive genes and is therefore, important for bio-fortified breeding pipelines to address food-cum-nutritional security.

Four review articles were also published in the research topic. The review article by Gaikwad et al. provided comprehensive overview of literature on QTLs/Genes identified for seed Zn content, seed Fe content, seed proteins, vitamins, oil content, fatty acid content, and other nutritional traits in major food crops. Efforts have been also made to compile literature on recently emerged genomic approaches/genome editing technologies and their use in enhancing nutritional quality in major food crops. The review article by Rao et al. presented detailed information regarding bio-fortification for iron and zinc polished rice cultivars. The article discussed in detail the observations made from the characterization of rice germplasm including bi-parental mapping populations for seed zinc content/concentrations and development of national evaluation system for the release of bio-fortified rice varieties. Another important and comprehensive review by Mbanjo et al. on pigmented rice provided an update on the nutritional value and health benefits of pigmented rice. The review also describes different approaches including QTL mapping and association mapping being used for genetic dissection of pigmented rice. Efforts have been also made to provide insight into the mechanistic basis of grain pigmentation. The Philippines is one of the largest producers and consumer of rice in Asia and most of nutrition of people in Philippines is fulfilled by rice crop. The review article by Palanog et al. highlights the Zn deficiency in Philippines soil and its influence on the Zn nutritional status of the population. The article also highlights the efforts on rice bio-fortification that resulted in the development of two bio-fortified rice varieties approved for commercial release in the Philippines.

AUTHOR CONTRIBUTIONS

The manuscript was written by RM. AK, MP, and SI checked and approved the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Fróna D, Szenderák J, Harangi-Rákos M. The challenge of feeding the world. *Sustainability*. (2019) 11:5816. doi: 10.3390/su11205816
2. FAO. *The Future of Food and Agriculture–Trends and Challenges*. Rome: FAO (2017).
3. Welch RM, Graham RD. Breeding for micronutrients in staple food crops from a human nutrition perspective. *J Exp Bot*. (2004) 55:353–5. doi: 10.1093/jxb/erh064
4. White PJ, Broadley MR. Bio-fortification of crops with seven mineral elements often lacking in human diets–iron, zinc, copper, calcium, magnesium, selenium and iodine. *New Phytol*. (2009) 182:49–84. doi: 10.1111/j.1469-8137.2008.02738.x
5. Kaiser N, Douches D, Dhingra A, Glenn KC, Herzig PR, Stowe EC, et al. The role of conventional plant breeding in ensuring safe levels of naturally occurring toxins in food crops. *Trends Food Sci Tech*. (2020) 100:51–66. doi: 10.1016/j.tifs.2020.03.042
6. Mir RR, Zaman-Allah M, Sreenivasulu N, Trethowan R, Varshney RK. Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. *Theor Appl Genet*. (2012) 125:625–45. doi: 10.1007/s00122-012-1904-9
7. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. (2008) 26:1135–45. doi: 10.1038/nbt1486
8. Mir RR, Reynolds M, Pinto F, Khan MA, Bhat MA. High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci*. (2019) 282:60–72. doi: 10.1016/j.plantsci.2019.01.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Mir, Kumar, Pandey and Isobe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Nutrition

ORIGINAL RESEARCH
published: 24 May 2019
doi: 10.3389/fnut.2019.00074

Nutritional Security in Drylands: Fast-Track Intra-Population Genetic Improvement for Grain Iron and Zinc Densities in Pearl Millet

Mahalingam Govindaraj, Kedar Nath Rai, Anand Kanatti, Aluri Sambasiva Rao and Harshad Shivade*

Crop Improvement, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

Considering the pervasive malnutrition caused by micronutrients, particularly those arising from the deficiencies of iron (Fe) and zinc (Zn), the primary focus of research in pearl millet is on biofortifying the crop with these two minerals. Pearl millet is a highly cross-pollinated crop where open-pollinated varieties (OPVs) and hybrids are the two distinct cultivar types. In view of the severe deficiency of Fe and Zn in Asia and Africa where this crop is widely consumed, crop biofortification holds a key role in attenuating this crisis. The present study included three OPVs previously identified for high-Fe and Zn density to assess the magnitude of variability and test the effectiveness of intra-population improvement as a fast-track selection approach. Large variability among the S_1 progenies was observed in all three OPVs, with the Fe varying from 31 to 143 mg kg^{-1} and Zn varying from 35 to 82 mg kg^{-1}. Progeny selection was effective for Fe density in all three OPVs, with up to 21% selection response for Fe density, and up to 10% selection response in two OPVs for Zn density, for which selection was made as an associated trait. Selection for Fe density had no adverse effect on grain yield and other agronomic traits. These results suggest that effective selection for Fe density in OPVs and composites can be made for these micronutrients and selection for Fe density is highly associated with the improvement of Zn density as well. These genetic changes can be achieved without compromising on grain yield and agronomic traits. Such improved versions could serve as essentially-derived varieties for immediate cultivation and also serve as potential sources for the development of parental lines of hybrids with elevated levels of Fe and Zn density. Therefore, fast-track breeding is essential to produce biofortified breeding pipelines to address food-cum-nutritional security.

Keywords: biofortification, iron, open-pollinated variety, pearl millet, recurrent selection, X-ray fluorescence spectrometer, zinc

OPEN ACCESS

Edited by:
Sachiko Narita Isobe, Kazusa DNA Research Institute, Japan

Reviewed by:
Zhaojun Wei, Hefei University of Technology, China
Sergiy Smetana, German Institute of Food Technologies, Germany

****Correspondence:***
Mahalingam Govindaraj
m.govindaraj@cgiar.org

Specialty section:
This article was submitted to Nutrition and Food Science Technology, a section of the journal Frontiers in Nutrition

Received: *22 December 2018*
Accepted: *01 May 2019*
Published: *24 May 2019*

Citation:
Govindaraj M, Rai KN, Kanatti A, Rao AS and Shivade H (2019) Nutritional Security in Drylands: Fast-Track Intra-Population Genetic Improvement for Grain Iron and Zinc Densities in Pearl Millet. Front. Nutr. 6:74. doi: 10.3389/fnut.2019.00074

INTRODUCTION

Pearl millet (*Pennisetum glaucum* (L.) R. Br.) is a highly cross-pollinated crop with more than 85% outcrossing (1). This floral system provides for open-pollinated varieties (OPVs) and hybrids as the two broad cultivar options. Pearl millet is cultivated in around 26 million ha, primarily for grain production, but it is also valued for its stover as a source of livestock fodder in the arid and semi-arid tropical regions of Africa and Asia, with India having the largest area of about 10 million ha (2). While the development of OPVs continues to be the thrust area of research in Africa, the development of hybrids is the primary focus in India, with

OPVs as the second priority. OPVs are genetically heterogeneous and phenotypically variable populations, which are constituted generally by random mating of 6–10 inbred lines or progenies of suitable populations. Hybrids are highly heterozygous but genetically homogeneous and phenotypically uniform cultivars which are constituted by crossing two distantly related lines. Improved OPVs of pearl millet are cultivated in about 800,000 ha, mostly in the marginal areas of India (3). OPVs have broad genetic base and are thus, less vulnerable to major diseases by horizontal resistance. The development of OPVs is also very important for deriving inbreds and breeding lines through series of selfing and selection that can be used for hybrid development.

Dietary deficiency of iron (Fe) and zinc (Zn) has been reported to be a grave public health problem, affecting more than two billion people worldwide, especially those in the developing countries (4, 5). Therefore, ending all forms of hunger (including hidden hunger), is one of the agenda of the United Nation's Sustainable Development Goals by 2030 (6). Pearl millet is a nutritious cereal with higher contents of protein, more balanced amino acid profile, and higher levels of micronutrients than those found in other cereals (7–9). Pearl millet is found to be more nutritious than other cereals, however, most of the cultivars grown by the farmers contain only very low levels of micronutrients. The lack of Fe/Zn-rich cultivars/hybrids for cultivation is a major lacuna among all pearl millet growers. Minnis-Ndimba et al. (10) studied the spatial distribution of micronutrients in the grains of pearl millet and showed that both Fe and Zn are predominantly concentrated in the germ, consisting of scutellum, embryo and in the seed coat, which also includes the pericarp and aleurone. Recent studies in pearl millet indicated that there is no significant difference in Fe bioavailability (7.0–7.5%) between the biofortified and the non-biofortified cultivars. However, biofortified varieties contain greater amounts of total iron and zinc densities than the non-biofortified varieties resulting in higher intake of these micronutrients in humans (11, 12). This marks the significant contribution of biofortification to tackle these micronutrient deficiencies in the population dependent on this nutritious cereal. However, the per capita consumption of pearl millet has declined significantly from 11.5 kg (during 1972) to 3.06 kg (by 73.4% in 2010) in rural areas and from 4 kg (during 1972) to 1.13 kg (by 71.8% in 2010) in urban India (13). Considering the fact that pearl millet continues to be an important food crop for India and Africa, creating awareness among the public about the nutritional values of pearl millet is necessary, and biofortification adds the value for its endorsement. The contribution of pearl millet to the total nutrient intake (especially Fe and Zn) from all foods/recipies varies widely across rural India. However, in some parts of rural India (Rajasthan, Maharashtra and Gujarat), the contribution of pearl millet in the intake of micronutrients (Fe and Zn) is very high by 30–50% (14). In a recent initiative by the HarvestPlus Challenge Program, the development of crop cultivars, including pearl millet, with high levels of iron and zinc holds much importance in addressing this issue. Results of a study under this project showed larger variability for Fe and Zn density in pearl millet populations, including some of the officially released OPVs (15, 16). ICTP 8203, an OPV released in India in 1988 (17) and under cultivation since then had been found to have the highest level of Fe and Zn density (18). During the course of inbreeding and selection to develop high-iron potential restorer lines from this OPV, 11 S_3 progenies were recombined to develop its better version with higher-Fe version. This improved version, designated as ICTP 8203 Fe 10-2, had 9% higher Fe and 11% more grain yield than ICTP 8203 in multi-location trials. It was officially released and adopted as *Dhanashakti* for cultivation in the country (19). Thus, as a fast-track approach, three of high-Fe OPVs were selected to test the effectiveness of intra-population selection for further increasing the Fe and Zn densities. Recurrent selection programs generally follow S_1 and S_2 progeny testing and very rarely, if ever, go for S_3 progeny testing for yield traits. The S_1 progeny selection has been more frequently used than S_2 progeny selection because of the necessity of one additional generation per cycle in the latter approach. Earlier studies compared S_1 and S_2 selection methods with full-sib (FS) and half-sib (HS) selection methods for grain yield in pearl millet, based on only one cycle of selection. Results showed that S_1 method is more effective than FS or HS methods (20). In the present study, primary emphasis in S_1 progeny testing for improved version of the OPV was given to Fe density *per se*. Further improved versions were developed by primarily considering the agronomic scores and also considering their levels of iron. Thus, the objective of this research was to assess the magnitude of variability and test the effectiveness of S_1 progeny-based genetic improvement in three OPVs of pearl millet for Fe/Zn density.

MATERIALS AND METHODS

Experimental Materials

The basic material for this study consisted of three populations, namely, ICMV 221, AIMP 92901 and ICMR 312. ICMV 221 is an open-pollinated variety (OPV) that was developed at ICRISAT by recombining 124 S_1 progenies derived from a Bold-seeded Early Composite (BSEC). It was released and notified in 1993 for cultivation at national level in India (21). AIMP 92901 is an OPV that was constituted by recombining 272 S_1 progenies derived from BSEC and was jointly released by Marathwada Agricultural University and ICRISAT (22). ICMR 312 is an OPV that was developed by recombining 200 S_1 progenies derived from (BSEC TCP2 C3) and used as a pollinator of a top cross hybrid ICMH 312 (23). All these three OPVs are based on *iniadi* germplasm, which is typically characterized by mostly globular large grains of gray to deep gray color (24). A preliminary trial conducted at ICRISAT showed a few OPVs and breeding lines with higher iron were either based on iniadi germplasm or had a greater percentage of their genes derived from this germplasm (15, 16). Since, AIMP 92901 and ICMR 312 are also based entirely on *iniadi* germplasm; these were included in present study. ICMV 221 was planted in the rainy season of 2008,

Abbreviations: AIMP, Aurangabad/ICRISAT millet population; ICMR, ICRISAT millet restorer; ICMV, ICRISAT millet variety; OPVs, Open-pollinated varieties; XRF, X-ray fluorescence spectrometer.

while AIMP 92901 and ICMR 312 were planted in the summer season of 2009. Each population was planted in 12 rows of 4 m length. The main panicles of about 120 plants in each population were selfed at the boot leaf stage to produce S_1 progenies for field trials.

S_1 Progeny Evaluation and Grain Micronutrient Analysis

The S_1 progenies of ICMV 221 (97), AIMP 92901 (96), and ICMR 312 (106) where seed was adequate to conduct replicated progeny trial and also for subsequent use in random matting to produce C_1 bulks, were planted as three separate experiments in adjacent strips in single row plots of 2-m length and replicated twice in randomized complete block design during the rainy season of 2010 in Alfisols at ICRISAT, Patancheru. The planting was done on ridges 75 cm apart. The over-planted rows were weeded and thinned on the 15th day from planting, leaving single plants that are spaced 10 cm apart. Pre-planting application of DAP (Diammonium phosphate, contains 18%N: 46%P) @ 100 kg ha^{-1} was followed with side dressing of urea (46%N) @ 100 kg ha^{-1}after thinning.

The progenies were visually assessed for agronomic performance on 1–5 scale (1 poor and 5 best) at or after physiological maturity. Seed set is known to affect Fe and Zn density (25) and inbreeding leads, depending on genotypes and environments, to variable seed set in selfing-derived progenies in pearl millet. Therefore, the progenies were also visually assessed for seed set percentage. As none of the progenies had <50% seed set, these were scored based on open-pollinated panicles for four broad categories of seed set (50–60% average, 60–75% above average, 75–90% good, and >90% excellent). No xenia and significant dust contamination effect on open-pollinated grains have been reported in pearl millet, and such grains have been found to provide reliable samples for the analysis of Fe and Zn density (25). Therefore, open-pollinated panicles from 6 to 8 random plants were harvested at or after physiological maturity, sundried for 15–20 days, and hand threshed to produce 30–40 g grain samples for Fe and Zn analysis.

The whole grain samples were analyzed using Energy-Dispersive X-ray Fluorescence (XRF) Spectrometer, calibrated and validated for pearl millet at Flinders University, Adelaide, Australia (26). In the above non-destructive XRF analysis, the calibration of Oxford Instruments X-Supreme 8000 fitted with a 10 place auto-sampler was done. About twenty reference whole grain samples of pearl millet that had wide range of ICP-determined Fe (29–163 mg kg^{-1}) and Zn (35–100 mg kg^{-1}) density were used to calibrate XRF method. Thus, ICP concentrations which were used as reference values were entered into the machine before each sample was scanned. Clean Poly-4 imported film was used for each sample. According to the manufacturer, the X-Supreme 8000 scans a circle of 21 mm diameter with the sample spinner on. All X-Ray scans in this study were performed in this mode, so the scanned area was 346 mm^2 (26). So, background scans fixed uniform emission toward sampling compartment with 60 s acquisition times for each sample cup to determine micronutrients and expressed in mg kg^{-1}. The correlation between XRF and Inductively Coupled Plasma Optical Emission Spectroscopy (ICP) values for Fe and Zn density was highly positive and highly significant (mostly $r = 0.90$; $P < 0.001$). The differences between these two methods were 4–6 mg kg^{-1} for both micronutrients. Based primarily on high Fe density, but also taking seed set and agronomic score into account, 11 progenies in ICMV 221, 10 in AIMP 92901 and 9 in ICMR 312 were selected for recombination to constitute the first improved version of C_1 bulks (**Figure 1**; **Tables S1, S3, S5**). Further, a set of 7 progenies from each OPV, selected principally on the basis of high agronomic score. By also considering high-Fe density, these progenies were used for recombination and the development of the second version of C_1 bulks (**Tables S2, S4, S6**). Six progenies in ICMV 221, 5 in AIMP 92901 and 4 in ICMR 312 were common in both the sets.

Population Bulk Trial and Grain Micronutrient Analysis

The selected progenies for constituting each version of C_1 bulk populations were planted during the summer season of 2011 in 2 row plots of 4 m length on ridges 60 cm apart, maintaining 10 cm plant-to-plant spacing within the rows. In each plot about 45–50 bagged plants were bagged at the boot leaf stage to avoid contamination from foreign pollen. Pollen was collected from bagged plants in each plot, and the composite pollen bulks were prepared from progenies selected for each version of C_1 bulk. These were crossed onto 8–10 plants in each plot corresponding to the appropriate version. This process continued for 3–4 days, and about 25–30 plants were crossed in each plot. At or after physiological maturity, all crossed panicles were harvested in each plot, sundried for 15–20 days, and bulk threshed. Equal amounts of seeds form each plot was pooled to produce the improved version of C_1 bulks of each OPV. The C_0 bulk (original) and improved versions of C_1 bulks were planted in a three-replicate trial in randomized complete block design during the summer and rainy seasons of 2012, and summer season of 2013 in Alfisols at Patancheru.

All these plots were planted in 4 rows of 4 m length, with a spacing of 75 cm between rows in rainy season and 60 cm between the rows in the summer. Days to 50% flower was recorded when the main panicles of 50% plants in a plot had fully emerged stigmas. Plant height and panicle length was recorded on 10 random plants in each plot in each replication. About 15 open-pollinated main panicles from each plot were harvested at or after maturity, sun dried for 15–20 days and bulk threshed to produce grain samples. These grain samples were analyzed for Fe and Zn density at the Waite Analytical Services Laboratory, University of Adelaide, Australia, using Inductively Coupled Plasma Optical Emission Spectroscopy (Spectro Analytical Instruments, Kleve, Germany) as described by Wheal et al. (27). In this destructive method, grain samples were oven-dried overnight at 85°C prior to digestion, grounded enough to pass through 1 mm stainless steel sieve using Christie and Norris hammer mill and stored in screw-top polycarbonate vials. The samples were digested with di-acid (Nitric/Perchloric acid) mixture. After digestion, the volume of the digest was made

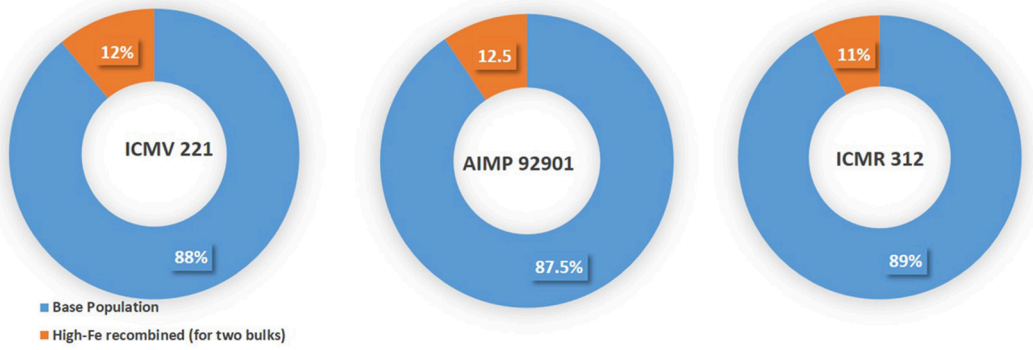

FIGURE 1 | Proportion of S_1 progenies used in each OPV to derive improved version for Fe and Zn density as biofortified varieties.

to 25 mL using distilled water; and the content was agitated for 1 min by vortex mixer. The digests were filtered and the Fe concentration was read at 259.94 nm and Zn concentration at 213.86 nm using ICP-OES and these micronutrients were expressed as mg kg^{-1}. Care was taken at each step to avoid any contamination of the grains with dust particles and any other extraneous matter (28). A random sample of 200 grains from each plot was weighed and multiplied by factor 5 to determine the thousand grain weight. The weight of grains from these 15 panicles were added to the grain weight of remaining panicles of each plot to calculate the grain yield per ha.

Statistical Analysis

The S_1 as well as the population bulk trials were analyzed assuming fixed model following Gomez and Gomez (29) and using Generalized Linear Model procedures (GLM) in Statistical Analysis Systems (SAS) version 9.3 (30). The selection differential (S) was estimated as a deviation of the mean value of the progenies selected as parents for recombination from the mean value of all the progenies included in the S_1 progeny trials (31). Analysis of variance of the population bulk trial was done with population bulks nested within the OPVs following Gomez and Gomez (29) and further partitioned for each OPV.

RESULTS AND DISCUSSION

Pearl millet is a highly out-pollinated crop. Such a reproductive biology facilitate the production of two distinct cultivars, namely, open-pollinated varieties (OPVs) and hybrids. Although hybrids occupy 60% of cultivated area in India, they have not still reached the semi-arid zones of India and the entire Sub-Saharan Africa. Therefore, the development of OPVs as well as hybrids is equally important for the Global Pearl Millet Improvement Program in ICRISAT. Biofortification is one of the new approaches in plant breeding which symbolizes a marriage of profitable cultivation and improved nutrition in primary cereal crops. Pearl millet is a nutritious crop compared to other major cereals but not all the farmers are able to grow varieties or hybrids with improved nutritional or protein profiles. The number of popular cultivars which are rich in Fe/Zn is almost negligible. Unlike the breeding methods for improving yield traits, the methods for utilizing the variability for micronutrients present in the germplasm has not yet been exploited thoroughly. For instance, the average densities of Fe (42 mg kg^{-1}) and Zn (32 mg kg^{-1}) in commercially grown hybrids in India are very low. This is lower than the baseline for Fe density (47 mg kg^{-1}) set by the HarvestPlus Program (16). The total genetic variability for the targeted essential micronutrients (Fe and Zn) in the germplasm pool is revealed by means of nutritional breeding. The fast-track breeding procedure involves the existing commercial OPVs. OPVs have large amounts of trait diversity within acceptable maturity range. This opens up opportunity to exercise progeny selection (for any traits) and to exploit the available variability whereas hybrids, bred for uniformity will not have variability for micronutrients, as these traits are not even monitored in due course of breeding their parental lines. Considering the above facts, the present study aimed to demonstrate the effectiveness of a fast-track breeding approach that uses intra-population variability, in commercial OPVs to enhance the Fe and Zn densities and develop their essentially-derived versions. For the interest of research community and relevance of this study, the results and discussion are combined hereunder.

The OPV ICMV 221 showed significant genetic variability with four-fold difference in Fe density among the S_1 progenies, which range from 31 to 117 mg kg^{-1} (**Table 1**). One progeny had exceptionally high Fe density of 143 mg kg^{-1}, which could have resulted from the seed sample and/or the analytical tool used for the analysis. This specific progeny with this high Fe density had only above-average seed set, and reduction in seed set has been shown to be associated with the overestimation of Fe density in pearl millet (25). Based on the results of several trials, it has also been observed that XRF may occasionally overestimate Fe density (32). Amongst the three OPVs, the progenies of ICMV 221 C_0 bulk had the highest mean Fe density of 75 mg kg^{-1}, and the progenies selected for constituting the two C_1 bulks had 32–33 mg kg^{-1} of Fe, as selection differentials (SD). The Fe density in the progenies of AIMP 92901 ranged from 40 to 104 mg kg^{-1}, with a mean of 68 mg kg^{-1}. Whereas, the selection differential was 23 mg kg^{-1} in the full set of progenies (11) to constitute

TABLE 1 | Mean and range among S_1 progenies tested, and selection differential (SD) for Fe and Zn densities in selected progenies of pearl millet, rainy season (2010), Patancheru.

Population bulk	No. of progenies	Fe density (mg kg^{-1})			Zn density (mg kg^{-1})		
		Range	Mean	*SD*	Range	Mean	*SD*
ICMV 221-C0	97	31–143	75	–	36–82	59	–
ICMV 221-Fe 11-1	11	94–143	108	33	63–80	72	13
ICMV 221-Fe 11-2	7	94–143	107	32	63–78	71	12
AIMP 92901-Fe-C0	96	40–104	68	–	35–79	58	–
AIMP 92901-Fe 11-1	10	84–102	91	23	57–79	68	10
AIMP 92901-Fe 11-2	7	84–102	94	26	58–79	66	8
ICMR 312-C0	106	34–96	65	–	35–69	52	–
ICMR 312-Fe 11-1	9	81–96	88	23	50–69	59	7
ICMR 312-Fe 11-2	7	81–96	87	22	52–65	59	7
SE±			2.1			1.4	

SE, Standard error of mean.

its first C_1 bulk and 26 mg kg^{-1}in the second set (7 progenies) selected to constitute the second C_1 bulk. The S_1 progenies of ICMR 312 had about three-fold difference for Fe density, varying from 34 to 96 mg kg^{-1}with a mean value of 65 mg kg^{-1} and selection differentials of 22–23 for the progenies selected to constitute the two C_1 bulks. About two-fold differences were observed among the S_1 progenies for Zn density in all three populations. The selection differentials of the progenies selected to constitute the two versions of C_1 bulks were 12–13 mg kg^{-1} for ICMV 221, 8–10 mg kg^{-1}for AIMP 92901, and 7 mg kg^{-1} for ICMR 312. Clearly, the selection differentials of the progenies selected to constitute the two versions of C_1 bulks were not different either for Fe or Zn density in all three OPVs because the second set of progenies (in each OPV) included about 50% of those included in the first set.

The mean grain yield in population bulk trial varied from 3.3 t ha^{-1} in the summer season of 2013 to 5.1 t ha^{-1} in the summer season of 2012, indicating large productivity level differences in the environments (**Table S7**). The Fe density across the environments varied from 64 mg kg^{-1} in the summer season of 2012 to 80 mg kg^{-1} in the rainy season of 2012; and the Zn density varied from 53 to 61 mg kg^{-1} under the same seasonal conditions. The difference among the population bulks of ICMV 221 was highly significant ($P < 0.01$), both for Fe and Zn density (**Table 2**). The differences were also highly significant ($P < 0.01$) for Fe density in ICMR 312, and significant for Zn density ($P < 0.05$) in AIMP 92901. Population × environment interaction was not significant for micronutrients and other traits except days to flowering in each individual and across OPVs. Thus, genotype × environment interaction for these micronutrients did not result in any change in the relative ranking of different population across the environments. Similar results have been reported in pearl millet studies for Fe and Zn densities (33, 34) and Cassava for carotenoid contents (35). The first version C_1 bulk of ICMV 221 (ICMV 221 Fe 11-1) had the mean Fe density of 76 mg kg^{-1} (21% higher than the C_0 bulk), and the second version C_1 bulk had 70 mg kg^{-1} Fe density (11% higher than the C_0 bulk) (**Table 3**). The Zn density in these two C_1 version bulks, respectively, was 61 mg kg^{-1} (15% higher than the C_0 bulk), and 58 mg kg^{-1} (9% higher than the C_0 bulk). There were highly significant differences among the population bulks of ICMV 221 for days to 50% flowering and plant height, with the C_1 bulks being 3–4 days later to flowering and 8–13 cm taller in height.

The Fe density of the first C_1 bulk of ICMR 312 was 70 mg kg^{-1} (11% higher than the Co bulk) and that of the second C_1 bulk was 72 mg kg^{-1}mg kg^{-1} (14% higher than the C_0 bulk). Although the difference among the population bulks of AIMP 92901 for Fe density was not significant, the first C_1 version had 66 mg kg^{-1} Fe density (12% higher than the C_0 bulk). It also had 57 mg kg^{-1} Zn density (10% higher than the C_0 bulk). There were no significant differences among the population bulks either in ICMR 312 or in AIMP 92901 for grain yield, time of flowering, plant height, panicle length and seed weight except that the panicles of the C_1 bulks of AIMP 92901 were significantly shorter by 3 cm. It was also found in ICMV 221, that the population bulks did not differ significantly for grain yield, panicle length and 1,000-seed weight.

The above results indicate that the selection for Fe density was effective in all three OPVs, and the magnitude of selection response largely depended on the magnitude of genetic variability and selection differentials. Thus, the largest selection response was observed in ICMV 221, which, among the three OPVs, had the largest variability and the highest selection differential for Fe density. This population also registered significant and the largest selection response for Zn density, which is not unexpected, considering the highly significant and high positive correlation ($r = 0.70$, $P < 0.01$) between Fe and Zn densities observed in the S_1 progeny trial of this OPV. Earlier studies in pearl millet have all shown highly significant and positive correlation between Fe and Zn densities, with the magnitude of correlation coefficient being dependent on the genetic materials used and the environment (15, 18, 25, 36–40). As compared to ICMV 221, the correlation coefficients between Fe and Zn density, although highly significant, were relatively low, both in ICMR 312 ($r = 0.52$; $p < 0.01$) and AIMP 92901 ($r = 0.47$; $p < 0.01$). The selection differentials for Fe and Zn densities in these two OPVs were also less than those observed in ICMV 221. Thus, selection response for Fe density was lower in these two populations, as compared to that observed in ICMV 221, the selection response for Zn density was also lower.

In the present study, a second version of C_1 bulks was constituted primarily based on the visual assessment of agronomic score (that included yielding ability) but also taking the Fe density into account. Nonetheless, its yield did not differ significantly from the yield of C_0 bulk in all three OPVs. This ineffectiveness of selection for grain yield could have resulted from the smaller size of the trial plots (of 2 m length) and the minimal number of replications (only twice) used in this study. Another reason for this result could predominantly be the non-additive mode of genetic variability for grain yield as reported in most of the pearl millet studies (39, 41). In contrast, Fe and Zn densities are the measurable traits and have been reported

TABLE 2 | Mean square for grain iron (Fe) and zinc (Zn) densities, grain yield ,and agronomic traits across three seasons in population bulks of three OPVs in pearl millet, Patancheru.

Source	DF	Mean square						
		Fe (mg kg^{-1})	Zn (mg kg^{-1})	Grain yield (kg ha^{-1})	Days to 50% flowering (d)	Plant height (cm)	Panicle length (cm)	1000-grain weight (g 1000^{-1})
Environment (E)	2	1745.0**	255.7**	23,857,604**	413.9**	16,032**	31.0**	17.1**
Replication/E	6	72.1	25.4	2,30,195	1.1	63.1	6.1**	2.2
Bulk	8	268.3**	88.7**	1,64,068	41.8**	171.2**	5.3**	4.1**
Population	2	336.7**	97.7**	2,14,620	139.3**	188.5*	8.8**	9.8**
Bulk/population	6	245.4**	85.7**	1,47,217	9.2**	165.5**	4.1*	2.2
ICMV 221 bulk	2	406.2**	151.5**	2,68,140	25.4**	452.5**	1.3	2.8
AIMP 92901 bulk	2	125.5	73.2*	1,61,548	2.1	4.9	9.1**	2.2
ICMR 312 bulk	2	204.6*	32.4	11,965	0.1	39.0	2.0	1.7
Bulk × E	16	65.6	22.2	96,542	7.1**	59.0	1.5	0.9
Population × E	4	73.6	34.7	2,23,513	17.6**	110.1*	2.0	0.9
Bulk/population × E	12	62.9	18.1	54,218	3.6**	42.0	1.4	0.9
ICMV 221 bulk × E	4	45.3	6.8	1,01,239	5.9**	81.8	0.3	1.6
AIMP 92901 bulk × E	4	30.7	16.7	31,598	4.7**	7.4	2.6	0.4
ICMR 312 bulk × E	4	112.8	30.8	29,818	0.3	36.8	1.2	0.8
Error	48	39.6	16.6	1,91,134	0.9	38.8	1.4	1.3

*Significant at the 0.05 probability level.
**Significant at the 0.01 probability level.

TABLE 3 | Grain iron (Fe) and zinc (Zn) densities, grain yield and agronomic traits in population bulks of three OPVs in pearl millet, mean of three seasons at Patancheru.

Population bulk	Fe (mg kg^{-1})	Zn (mg kg^{-1})	Grain yield (kg ha^{-1})	Days to 50% flowering (d)	Plant height (cm)	Panicle length (cm)	Grain weight (g 1000^{-1})
ICMV 221-C0	63	53	3,978	39	165	22	14.2
ICMV 221-Fe 11-1	76	61	3,914	42	173	23	15.3
ICMV 221-Fe 11-2	70	58	4,240	43	178	22	14.7
ICMR 312-C0	63	52	4,161	46	176	24	13.1
ICMR 312-Fe 11-1	70	53	4,200	45	178	24	13.5
ICMR 312-Fe 11-2	72	56	4,234	46	177	23	14
AIMP 92901-Fe-C0	59	52	4,335	43	176	25	14.3
AIMP 92901-Fe 11-1	66	57	4,067	42	172	22	13.7
AIMP 92901-Fe 11-2	64	56	4,190	43	174	22	14.6
LSD (5%)[€]	6	4	414	1	6	1	1.1

[€]*Least significant difference.*

to be predominantly under additive genetic control (33, 38, 39), which could enhance the effectiveness of selection during early generations even in smaller plots. Significant selection response for Fe density and lack of significant difference among the population bulks for grain yield in all three populations indicated that effective selection in populations for Fe density can be made, at least, without compromising on grain yield. For the simultaneous selection for Fe density and grain yield to be effective, one needs to use a trial plot of size of at least 4 m length (normal practice in pearl millet improvement), the genotypes being replicated three times, and exercise moderate selection intensity, of say 30–40%, at the S_1 stage and delay high intensity selection to the later selfing generations. Following a sequential selection procedure from S_1 though S_3 generation for estimation of Fe density and visual assessment of agronomic score on yielding ability, a C_1 version of a commercial variety ICTP 8203 selected for high Fe density was developed. This high-Fe version had 9% higher Fe density and 11% higher grain yield than the C_0 version, based on 42 location × year trials in India (19). This result showed that effective simultaneous selection for Fe density and grain yield is possible in pearl millet. Studies in other crops like cassava (42) and maize (43) for carotenoid contents, and strontium (Sr) content in wheat and barley (44) have demonstrated that the mineral

contents can be improved considerably between and within the populations.

The selection for Fe density had no significant effects on any of the agronomic traits studied except for maturity as 3–4 days of delayed flowering and 8–13 cm increase in plant height of the C_1 bulks of ICMV 221, and 3 cm shorter panicles of the C_1 bulks of AIMP 92901. Inconsistency of these changes across the three populations suggests that these changes occurred either as random events, or were characteristics of specific genetic constitutions of the populations, and may not be a general feature reflecting on the association of Fe and Zn densities with the time of flowering and plant height. For instance, the selected high-Fe version of ICTP 8203 mentioned above had similar flowering time but had plants that were 7 cm taller than the original (C_0) bulk (19). While one pearl millet study observed significant positive association between Fe density and time to 50% flowering (36), another study showed that both Fe and Zn densities were negatively correlated with plant height, days to flowering and panicle length, although not significant in all the environments (40). These results would indicate that association of Fe and Zn densities with grain yield and other agronomic traits remains a subject meriting further investigations.

CONCLUSION

The biofortification of pearl millet is focused on improving the levels of both iron and zinc which are associated to one another. Huge genetic variability for Fe and Zn densities was observed among the S_1 progenies in all the three OPVs used in this study. Progeny selection within population was effectively demonstrated with high selection response for both the micronutrients (10–21%). Results are in conformity with the previous reports indicating additive genetic control of both micronutrients which is reflected in the effectiveness of progeny test-based population improvement. Increased cycle of re-selection and subsequent recombination to assemble more favorable alleles controlling Fe/Zn may yield greater difference from the original bulks. This study necessitates the worthiness of genomic tools and analyses that can play major role in discovering such alleles and their loci in the original and improved versions. There were very high correlations between Fe and Zn densities and no significant change in the grain yield as compared to the original population, indicating good prospects of simultaneous genetic improvement for both micronutrients without compromising on grain yield. Nevertheless, conscious selection may be equally important for Zn density along with Fe density for greater genetic gain than that observed in this study. These improved populations can be evaluated in multi-location trials to confirm their stability and superior performances. They can be identified as essentially-derived varieties for higher micronutrients levels with no changes in grain yield and agronomic traits.

AUTHOR CONTRIBUTIONS

MG and KR designed the study and prepared the final manuscript. AK and AR analyzed the data and prepared initial draft. HS managed the field trials and data collection. All of the authors have read and approved this manuscript and was involved in the field experiments.

FUNDING

This research is funded by the HarvestPlus Challenge Program of the CGIAR and was conducted under CGIAR Research Program on Agriculture for Nutrition and Health (A4NH).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2019.00074/full#supplementary-material

REFERENCES

1. Burton GW. Factor affecting pollen movement and natural crossing in pearl millet. *Crop Sci.* (1974) 14:802–5. doi: 10.2135/cropsci1974.0011183X001400060007x
2. Yadav OP, Rai KN. Genetic improvement of pearl millet in India. *Agri Res.* (2013) 2:275–92. doi: 10.1007/s40003-013-0089-z
3. Bantilan MCS, Subba Rao KV, Rai KN, Singh SD. Research on high yielding pearl millet – background for an impact study in India. In: Bantilan MCS, Joshi PK, editors. *Assessing Joint Research Impacts: Proceedings of an International Workshop on Joint Impact Assessment of NARS/ICRISAT Technologies for the Semi-Arid Tropics*. Patancheru: ICRISAT (1998). p. 52–61.
4. WHO (World Health Organization). *Malnutrition.* (2017). Available online at: www.who.int/mediacentre/factsheets/malnutrition/en/ (accessed August 29, 2018).
5. Bouis HE, Saltzman A. Improving nutrition through biofortification: review of evidence from HarvestPlus, 2003 through 2016. *Glob Food Sec.* (2017) 12:49–58. doi: 10.1016/j.gfs.2017.01.009
6. United Nations. *Sustainable Development Goal 2 – Zero Hunger.* (2016). Available online at: www.undp.org/content/undp/en/home/sustainable-development-goals/goal-2-zero-hunger/targets/ (accessed August 29, 2018).
7. Andrews DJ, Kumar KA. Pearl millet for food, feed, and forage. *Adv Agron.* (1992) 48:89–139. doi: 10.1016/S0065-2113(08)60936-0
8. Dwivedi SL, Sahrawat KL, Rai KN, Blair MW, Andersson MS, Pfeiffer WH. Nutritionally enhanced staple food crops. *Plant Breed Rev.* (2012) 36:161–291. doi: 10.1002/9781118358566.ch3
9. Vinoth A, Ravindhran R. Biofortification in millets: a sustainable approach for nutritional security. *Front Plant Sci.* (2017) 8:29. doi: 10.3389/fpls.2017.00029
10. Minnis-Ndimba R, Kruger J, Taylor JRN, Mtshali C, Pineda-Vargas CA. Micro-PIXE mapping of mineral distribution in mature grain of two pearl millet cultivars. *Nucl Instrum Meth B.* (2015) 363:177–82. doi: 10.1016/j.nimb.2015.09.014
11. Kodkany BS, Bellad RM, Mahantshetti NS, Westcott JE, Krebs NF, Kemp JF, et al. Biofortification of pearl millet with iron and zinc in a randomized controlled trial increases absorption of these minerals above physiologic requirements in young children. *J Nutri.* (2013) 143:1489–93. doi: 10.3945/jn.113.176677
12. Tako E, Reed SM, Budiman J, Hart JJ, Glahn RP. Higher iron pearl millet (*Pennisetum glaucum* L.) provides more absorbable iron that

is limited by increased polyphenolic content. *Nutri J.* (2015) 14:11. doi: 10.1186/1475-2891-14-11
13. NRAA. *Products, Diversification, Marketing and Price Discovery of Pearl Millet in India*. *Policy Paper No. 2*. New Delhi: National Rainfed Area Authority (2012). p. 48.
14. Rao PP, Birthal PS, Reddy BVS, Rai KN, Ramesh S. Diagnostics of sorghum and pearl millet grains-based nutrition in India. *Int Sorghum Millets Newslett.* (2006) 47:93–6. Available online at: http://oar.icrisat.org/id/eprint/1119
15. Velu G, Rai KN, Muralidharan V, Kulkarni VN, Longvah T, Raveendran TS. Prospects of breeding biofortified pearl millet with high grain iron and zinc content. *Plant Breed.* (2007) 126:182–5. doi: 10.1111/j.1439-0523.2007.01322.x
16. Rai KN, Yadav OP, Govindaraj M, Pfeiffer WH, Yadav HP, Rajpurohit BS, et al. Grain iron and zinc densities in released and commercial cultivars of pearl millet. *Indian J Agri Sci.* (2016) 86:291–6. Available online at: http://oar.icrisat.org/id/eprint/9394
17. Rai KN, Hash CT. Fertility restoration in male-sterile × maintainer hybrids of pearl millet. *Crop Sci.* (1990) 30:889–92. doi: 10.2135/cropsci1990.0011183X003000040026x
18. Rai KN, Yadav OP, Rajpurohit BS, Patil HT, Govindaraj M, Khairwal IS, et al. Breeding pearl millet cultivars for high iron density with zinc density as an associated trait. *J SAT Agri Res.* (2013) 11:1–7. Available online at: http://oar.icrisat.org/id/eprint/7291
19. Rai KN, Patil HT, Yadav OP, Govindaraj M, Khairwal IS, Cherian B, et al. Variety dhanashakti. *Indian J Genet.* (2014) 74:405–6. Available online at: http://oar.icrisat.org/id/eprint/8605
20. Rai KN, Virk DS. Breeding methods. In: Khairwal IS, Rai KN, Andrews DJ, Harinarayana G, editors. *Pearl Millet Breeding*. New Delhi: Oxford and IBH Publishing Co. Pvt. Ltd. (1999). p. 185–211.
21. Witcombe JR, Rao MNVR, Raj AGB, Hash CT. Registration of ICMV 88904 pearl millet. *Crop Sci.* (1997) 37:1022–3. doi: 10.2135/cropsci1997.0011183X003700030082x
22. Yadav OP, Rai KN, Rajpurohit BS, Hash CT, Mahala RS, Gupta SK, et al. (2012). *Twenty-Five Years of Pearl Millet Improvement in India*. Jodhpur: All India Coordinated Pearl Millet Improvement Project.
23. Witcombe JR, Rao MNVR, Talukdar BS, Singh SD, Rao AM. Registration of ICMR 312 pearl millet topcross pollinator germplasm. *Crop Sci.* (1996) 36:471. doi: 10.2135/cropsci1996.0011183X003600020052x
24. Andrews DJ, Anand Kumar K. Use of the West African pearl millet landrace *Iniadi* in cultivar development. *Plant Genet Resour Newsl.* (1996) 105:15–22.
25. Rai KN, Govindaraj M, Pfeiffer WH, Rao AS. Seed set and xenia effects on grain iron and zinc density in pearl millet. *Crop Sci.* (2015) 55:821–7. doi: 10.2135/cropsci2014.04.0305
26. Paltridge NG, Palmer LJ, Milham PJ, Guild GE, Stangoulis JCR. Energy-dispersive X-ray fluorescence analysis of zinc and iron concentration in rice and pearl millet grain. *Plant Soil.* (2012) 361:251–60. doi: 10.1007/s11104-011-1104-4
27. Wheal MS, Fowles TO, Palmer LT. A cost-effective acid digestion method using closed polypropylene tubes for inductively coupled plasma optical emission spectrometry (ICP-OES) analysis of plant essential elements. *Anal Methods.* (2011) 3:2854–63. doi: 10.1039/c1ay05430a
28. Stangoulis J, Sison C. *Crop Sampling Protocols for Micronutrient Analysis. HarvestPlus Technical Monograph Series 7*. Washington, DC: HarvestPlus (2008). Available online at: http://www.harvestplus.org/sites/default/files/TM7%20complete.pdf (accessed May 25, 2017).
29. Gomez KA, Gomez AA. *Statistical Procedures for Agricultural Research*. 2nd ed. New York, NY: John Wiley and Sons (1984).
30. Statistical Analysis Systems Institute Inc. *SAS/STAT® 9.2 User's Guide.* 2nd ed. Cary, NC: SAS Institute Inc (2009).
31. Hallauer AR, Carena MJ, Miranda Filho JB. *Quantitative Genetics in Maize Breeding*. 2nd ed. Ames, IA: Iowa State University Press (1988).
32. Govindaraj M, Rai KN, Pfeiffer WH, Kanatti A, Shivade H. Energy-dispersive X-ray fluorescence spectrometry for cost-effective and rapid screening of pearl millet germplasm and breeding lines for grain iron and zinc density. *Comm Soil Sci Plant Anal.* (2016a) 47:2126–34. doi: 10.1080/00103624.2016.1228938
33. Velu G, Rai KN, Muralidharan V, Longvah T, Crossa J. Gene effects and heterosis for grain iron and zinc density in pearl millet (*Pennisetum glaucum* (L.) R. Br). *Euphytica.* (2011) 180:251–9. doi: 10.1007/s10681-011-0387-0
34. Govindaraj M, Rai KN, Shanmugasundaram P. Intra-population genetic variance for grain iron and zinc contents and agronomic traits in pearl millet. *Crop J.* (2016b) 4:48–54. doi: 10.1016/j.cj.2015.11.002
35. Ssemakula G, Dixon A. Genotype × environment interaction, stability and agronomic performance of carotenoid-rich cassava clones. *Sci Res Essays.* (2007) 2:390–9.
36. Gupta SK, Velu G, Rai KN, Sumalini K. Association of grain iron and zinc content with grain yield and other traits in pearl millet (*Pennisetum glaucum* (L.) R. Br.). *Crop Improv.* (2009) 36:4–7. Available online at: http://oar.icrisat.org/id/eprint/363
37. Rai KN, Govindaraj M, Rao AS. Genetic enhancement of grain iron and zinc content in pearl millet. *Qual Assur Saf Crop*. (2012) 4:119–25. doi: 10.1111/j.1757-837X.2012.00135.x
38. Govindaraj M, Rai KN, Shanmugasundaram P, Dwivedi SL, Sahrawat KL, Muthaiah AR, et al. Combining ability and heterosis for grain iron and zinc densities in pearl millet. *Crop Sci.* (2013) 53:507–17. doi: 10.2135/cropsci2012.08.0477
39. Kanatti A, Rai KN, Radhika K, Govindaraj M, Sahrawat KL, Rao AS. Grain iron and zinc density in pearl millet: combining ability, heterosis and association with grain yield and grain size. *SprigerPlus.* (2014) 3:763. doi: 10.1186/2193-1801-3-763
40. Kanatti A, Rai KN, Radhika K, Govindaraj M, Sahrawat KL, Srinivasu K, et al. Relationship of grain iron and zinc content with grain yield in pearl millet hybrids. *Crop Improv.* (2014) 41:91–6. Available online at: http://oar.icrisat.org/id/eprint/8894
41. Khairwal IS, Rai KN, Andrews DJ, Harinarayana G. *Pearl Millet Breeding*. New Delhi: Oxford and IBH Publishing (1999).
42. Ceballos H, Morante N, Sánchez T, Ortiz D, Aragón I, Chávez AL, et al. Rapid cycling recurrent selection for increased carotenoids content in Cassava roots. *Crop Sci.* (2013) 53:1–10. doi: 10.2135/cropsci2013.02.0123
43. Dhliwayo T, Palacios-Rojas N, Crossa J, Pixley KV. Effects of S_1 recurrent selection for provitamin A carotenoid content for three open-pollinated maize cultivars. *Crop Sci.* (2014) 54:2449–60. doi: 10.2135/cropsci2013.11.0764
44. Byrne I, Rasmusson DC. Recurrent selection for mineral content in wheat and barley. *Euphytica.* (1974) 23:241–9. doi: 10.1007/BF00035864

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Govindaraj, Rai, Kanatti, Rao and Shivade. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Nutrition

PROTOCOLS
published: 05 June 2019
doi: 10.3389/fnut.2019.00068

An Improved Enzyme-Linked Immunosorbent Assay (ELISA) Based Protocol Using Seeds for Detection of Five Major Peanut Allergens Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8

Arun K. Pandey, Rajeev K. Varshney, Hari K. Sudini and Manish K. Pandey**

International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

OPEN ACCESS

Edited by:
José M. Alvarez-Suarez, University of the Americas, Ecuador

Reviewed by:
Carmen Cuadrado, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Spain
Isabel Ballesteros, University of the Americas, Ecuador

***Correspondence:**
Hari K. Sudini
h.sudini@cgiar.org
Manish K. Pandey
m.pandey@cgiar.org

Specialty section:
This article was submitted to Nutrition and Food Science Technology, a section of the journal Frontiers in Nutrition

Received: *28 February 2019*
Accepted: *25 April 2019*
Published: *05 June 2019*

Citation:
Pandey AK, Varshney RK, Sudini HK and Pandey MK (2019) An Improved Enzyme-Linked Immunosorbent Assay (ELISA) Based Protocol Using Seeds for Detection of Five Major Peanut Allergens Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8. Front. Nutr. 6:68. doi: 10.3389/fnut.2019.00068

Peanut allergy is an important health concern among many individuals. As there is no effective treatment to peanut allergy, continuous monitoring of peanut-based products, and their sources is essential. Precise detection of peanut allergens is key for identification and development of improved peanut varieties with minimum or no allergens in addition to estimating the levels in peanut-based products available in food chain. The antibody based ELISA protocol along with sample preparation was standardized for Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8 to estimate their quantities in peanut seeds. Three different dilutions were optimized to precisely quantify target allergen proteins in peanut seeds such as Ara h 1 (1/1,000, 1/2,000, and 1/4,000), Ara h 2 and Ara h 3 (1/5,000, 1/10,000, and 1/20,000), Ara h 6 (1/40,000, 1/80,000, and 1/1,60,000), and Ara h 8 (1/10, 1/20, and 1/40). These dilutions were finalized for each allergen based on the accuracy of detection by achieving <20% coefficient of variation in three technical replicates. This protocol captured wide variation of allergen proteins in selected peanut genotypes for Ara h 1 (77–46,106 μg/g), Ara h 2 (265–5,426 μg/g), Ara h 3 (382–12,676 μg/g), Ara h 6 (949–43,375 μg/g), and Ara h 8 (0.385–6 μg/g). The assay is sensitive and reliable in precise detection of five major peanut allergens in seeds. Deployment of such protocol allows screening of large scale germplasm and breeding lines while developing peanut varieties with minimum allergenicity to ensure food safety.

Keywords: protocol, peanut allergens, Ara h 1, Ara h 2, Ara h 3, Ara h 6, Ara h 8, ELISA

INTRODUCTION

Food allergies have become the most important food safety concerns around the world. The prevalence of food allergy is more in North America, where 6% of young children and 4% of adults suffer from food allergies (1). Among all the food allergies, ∼1.4% of Americans and 3% Australian are allergic to peanuts and peanut-based products (2, 3). Peanut allergy is very severe and sometimes causes death due to anaphylaxis (4, 5). It is reported that food allergies cause approximately 150 to 200 fatalities per year, based on data from a five year study of anaphylaxis in Minnesota from the Mayo Clinic. Fatal food anaphylaxis is most often caused by peanuts (50–62%)

and tree nuts (15–30%) (6). Proteins are the major cause of food allergy and these allergenic proteins are usually highly resistant to heat and proteolysis (7). Peanut is the largest source for the IgE-mediated food allergies among eight major sources namely tree nuts, egg, cow's milk, fish, crustacean shellfish, soybean, and wheat. There is no effective treatment to peanut allergy and the allergic person is advised strictly to avoid consuming peanut or peanut-based products (8). Peanut being a common food ingredient in many food preparations, it is very challenging for the allergic person to know the composition of these preparations to avoid consumption (9). The threshold of allergen levels differ among the allergic population and even a minute dose of 100 μg of Ara h 1 can trigger an allergic reaction (10). The diagnosis of peanut allergy can be done using different methods such as double-blind, placebo-controlled food challenge (DBPCFC), the basophil activation test, the specific skin prick test (SPT), and the measurement of specific IgE (11–13).

Of the 32 different types of proteins present in peanut seeds, 16 of these proteins show allergenic property (14). Further, out of these 16 peanut allergen proteins, Ara h 1, 2, 3, and 6 are considered as major allergens because they are recognized by the IgE of a majority of allergic patients. Ara h 5, 7, 8, 9, 10, 11, and 12/13 are considered minor allergens and are not thought to be the causative agents in most of the life-threatening allergic reactions (anaphylaxis) (15). Nevertheless, if a person is already sensitive to Bet v 1 allergey caused due to birch pollen, then one of these minor peanut allergens, Ara h 8, shows cross-reactivity with IgE antibodies causing oral allergy syndrome (OAS) (16–18).

Allergic protein belongs to different protein families namely cupin (vicilin-type, 7S globulin, legumin-type, 11S globulin, glycinin), conglutin (2S albumin), profilin, nonspecific lipid-transfer protein 1, pathogenesis-related protein, PR-10, 14 kDa oleosin, and 16 kDa oleosin (19), particularly, Ara h 1, Ara h 2, Ara h 3, and Ara h 6 are the seed storage proteins. Many studies showed that Ara h 1 and Ara h 3 are abundant peanut proteins but, Ara h 2 and Ara h 6 strongly bind with IgE from peanut allergic patient and release mediators from basophils, which confirmed that both are the most important allergens *in vitro* (20–23) and *in vivo* (21, 24, 25) with regards to food allergy (23). Due to the fact that both Ara h 2 and Ara h 6 are recognized more prominently and more frequently than Ara h 1 and Ara h 3 by peanut allergic children with almost all of them showing IgE reactivity to these two allergens (26).

Immuno-assays are the most commonly preferred analytical techniques being used to detect the food allergens in many industries (27). Food industries commonly use ELISA to detect and quantify hidden allergens in food commodities as it is easy and less time-consuming technique (28–30). Further ELISA is highly sensitive, reliable, cost-effective, and accurate. In the past ELISA has been successfully used to detect allergens in several food products such as in meat and meat products, milk and milk products, fish and fish products, soybean, nuts and nut-based products, and fruit juices and their ingredients (31, 32). In ELISA, allergens can be detected by the colorimetric reaction after binding of the antigen with the specific enzyme-labeled antibody. Earlier studies reported the allergenic proteins quantification in peanut-based products through ELISA and HPLC technique (29, 33–35), however, no information on peanut seeds. The key issues in ELISA are the sample preparation and unacceptable variability in multiple measurements of the same sample. The best way of preparing a sample, right dilution of the sample used in the assay, and low coefficient of variation (CV) holds importance in achieving accuracy. As of now there is no standard protocol which can help to detect allergen proteins in peanut seeds on the basis of dilution factor and low coefficient of variation (CV). Further the genetic diversity available in peanut germplasm has not yet been fully exploited and the accumulation of favorable alleles may provide minimum allergen proteins in the improved peanut genotypes. At this juncture a highly sensitive, easy to use and reliable protocol for peanut seeds is required to screen various germplasm sets and breeding material. Therefore, the present study focused on developing a robust protocol to estimate major allergens in peanut seeds through ELISA based methods.

MATERIALS AND METHODS

Plant Material, Reagents and Chemicals

A total of 38 peanut genotypes (see **Supplementary Table 1**) were used while improving and standardizing the ELISA protocol. The seeds of these genotypes after harvest were stored at 10°C and were taken out only when needed for the experiment. ELISA components such as peanut allergen standards, monoclonal antibodies (MAbs; primary and secondary biotinylated antibody), and enzyme conjugates were purchased from INDOOR Biotechnologies Inc. (Charlottesville, VA, USA) for the estimation of five important peanut allergens namely Ara h 1, Ara h 2 Ara h 3, Ara h 6, and Ara h 8. These include monoclonal antibody 2C12 (clone 2C12 A11 A3) and biotinylated antibody 2F7 (clone 2F7 C12 D10) for Ara h 1, the monoclonal antibody 1E8 and biotinylated monoclonal antibody 4G9 for Ara h 3 while monoclonal antibody 3B8 (clone 3B8 B5) and biotinylated antibody 3E12 (clone 3E12 C4 B3) for Ara h 6. Similarly, the monoclonal antibody 1C4 (clone 1C4 G4 C8) and a polyclonal rabbit anti Ara h 2 antibody (AH2) for Ara h 2; capture antibody 4G6 and a polyclonal antiserum raised in rabbit for Ara h 8 were purchased from INDOOR Biotechnologies Inc. Conjugated Goat anti-Rabbit IgG was purchased from Jacksons Laboratories (Bar Harbor, USA Cat No. 111-036-046). ABTS™ (Cat No. 11204521001) and all other reagents and chemicals were purchased from Sigma-Aldrich Co. (Oakville, ON, Canada).

Allergen Standards and Enzyme Solution Preparation

The allergen standards were purchased from INDOOR Biotechnologies Inc. (Charlottesville, VA, USA). These standards were isolated from lightly roasted peanut flour (Runner cultivar) and purified by affinity chromatography. The purified standards were supplied in phosphate buffer, received on the ice and stored at −20°C until further use. Purified natural Ara h 1 (Lot 39285; Conc. 20,000 ng/ml), Ara h 2 (Lot 39158; Conc. 2,500 ng/ml), Ara h 3 (Lot 39051; Conc. 1,250 ng/ml) Ara h 6 (Lot 39198; Conc. 1,000 ng/ml), and Ara h 8 (Lot 39033;

TABLE 1 | The repeatability and precision of the ELISA-based assay in peanut seeds.

S. No	Peanut genotype	Range in peanut seeds (μg/g)	Mean (μg/g)	Standard deviation (SD)	Range of (CV%)	Mean of (CV%)
Ara h 1 ALLERGEN PROTEIN						
1	ICG 311	66–89	77	12	3–5	4
2	ICG 4798	719–936	842	112	5–15	9
3	ICG 5494	1,638–1,689	1,652	23	2–4	3
4	ICG 7963	3,496–4,105	3,761	312	2–11	7
5	ICG 11457	7,493–8,589	8,073	551	3–7	5
6	ICG 13856	18,253–22,080	20,481	1,989	1–10	5
7	ICG 13858	39,058–46,106	42,609	3,524	2–7	14
8	ICG 15380	39,426–43,859	41,008	2,474	1–15	7
Ara h 2 ALLERGEN PROTEIN						
9	ICG 311	253–361	307	54	22–25	24
10	ICG 532	208–331	265	62	12–19	16
11	ICG 12189	230–429	327	100	6–19	13
12	ICG 13491	5,005–6,062	5,426	561	5–17	10
13	ICG 7969	2,488–2,925	2,666	230	9–18	14
14	ICG 9777	2,390–2,467	2,422	41	1–15	7
15	ICG 9842	2,973–3,564	3,321	309	10–18	14
16	ICG 14705	4,136–4,642	4,432	264	4–17	11
Ara h 3 ALLERGEN PROTEIN						
17	ICG 311	406–646	559	32	8–24	8
18	ICG 1487	369–392	382	12	6–9	8
19	ICG 3240	1,274–1,626	1,452	176	9–22	15
20	ICG 3343	5,044–5,679	5,385	320	0.15–19	11
21	ICG 12682	7,161–7,758	7,524	250	2–29	13
22	ICG 12879	12,029–13,195	12,676	593	8–24	17
23	ICG 13603	4,707–5,125	4,909	209	7–12	9
24	ICG 14482	5,731–6,491	6,009	419	7–23	14
Ara h 6 ALLERGEN PROTEIN						
25	ICGV 01328	829–1,065	949	118	0.05–14	5
26	ICGV 87354	1,578–1,629	1,561	77	7–23	14
27	ICGV 93470	2,066–2,412	2,184	197	9–24	14
28	ICG 111	38,174–44,065	39,587	3,453	0.05–14	5
29	ICG 10036	29,711–32,982	31,731	1,809	4–15	10
30	ICGV 91116	41,288–45,998	43,375	1,947	5–11	8
Ara h 8 ALLERGEN PROTEIN						
31	ICG 311	0.275–0.503	0.385	0.114	3–17	11
32	ICG 532	0.415–1	0.752	0.332	4–20	13
33	ICG 12189	0.956–1.302	1.374	0.458	7–12	10
34	ICG 13491	4.8–5.2	5	0.189	5–22	16
35	ICG 7969	5.2–6.3	5.9	0.585	8–23	17
36	ICG 9449	4.2–5.2	4.7	0.477	17–24	21
37	ICG 9842	4.7–5.3	4.9	0.347	5–21	15
38	ICG 14705	5.6–6	6	0.322	11–27	20

The range and mean of different allergen in selected Peanut genotypes. Data represents the mean of three replication ± SD and CV%.

Conc. 2,500 ng/ml) were used as allergen standards for each assay. The standard concentration of each allergen ranged from 2000 - 4 ng/ml for Ara h 1, 250 - 0.5 ng/ml for Ara h 2, 125 - 0.24 ng/ml for Ara h 3, 100 - 0.2 ng/ml for Ara h 6 and 250 - 0.49 ng/ml for Ara h 8. ABTS Substrate is a water-soluble peroxidase substrate that yields a measurable green end product for use in ELISA methods. The ABTS™ (Sigma Aldrich Cat No. 11204521001) was dissolved in 1 mM ABTS solution. The ABTS solution contains 0.1M anhydrous citric acid and 0.2M Dibasic Sodium phosphate.$7H_20_2$. The 274 mg ABTS™ were dissolved in 500 ml ABTS solution and store in an amber color bottle at 4°C until use.

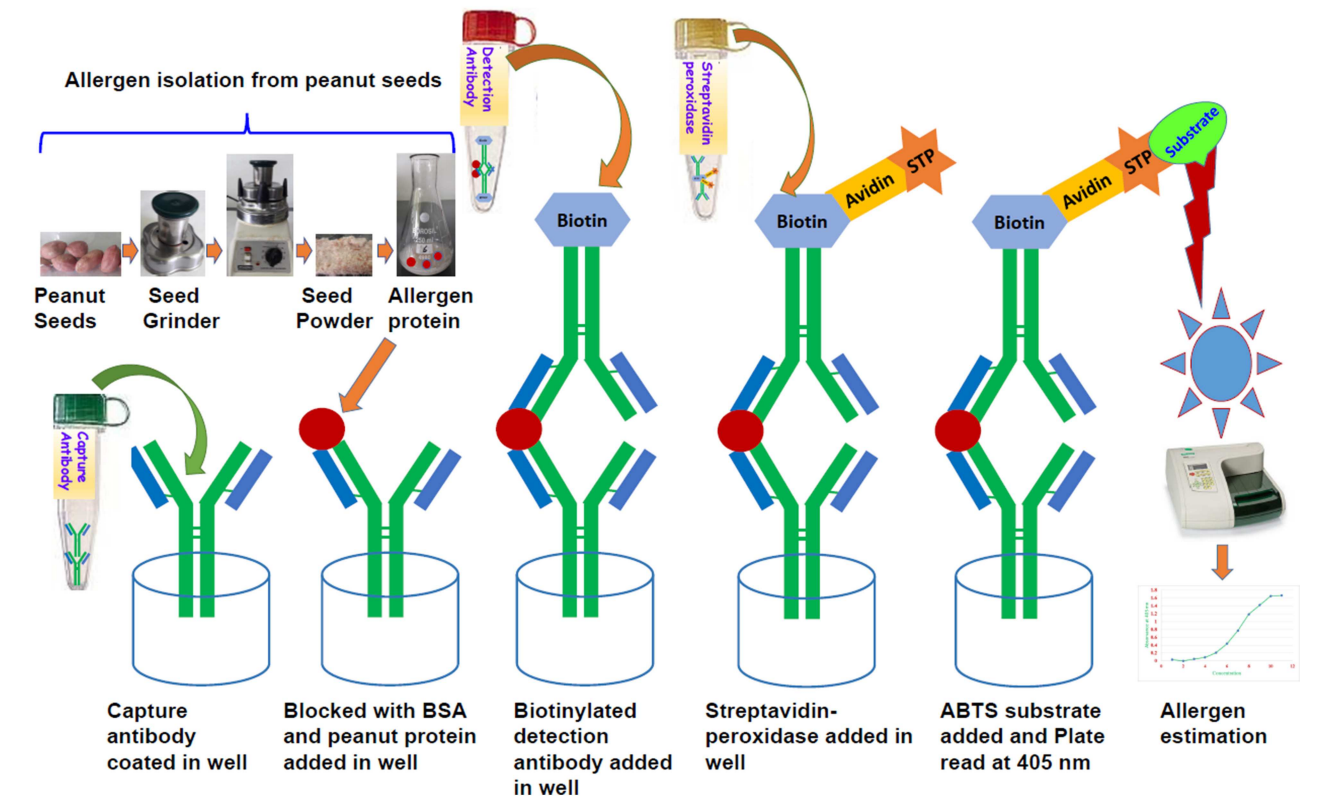

FIGURE 1 | Systematic diagram showing the protocol for allergen estimation in peanut seed through sandwich ELISA.

Development of Protocol for Allergen Estimation

The ELISA is sandwich format having double-antibody based on the specific interaction between antigen and antibody. The peanut allergen proteins are sandwich between two antibodies such as capture antibody and conjugated antibody with streptavidin peroxidase (**Figure 1**). For color development ABTS™ was used. The color intensity depends on the concentration of allergen protein present in the specific sample and measured with an iMark microplate reader (Bio-Rad) at 405 nm. The Microplate Manager (MPM) software was used while analyzing the optical density values of standards and samples. The systemic representation showed the key steps involved to develop the ELISA protocol, to estimate the peanut allergens through peanut seeds (**Figure 1**).

Grinding of Seeds, Homogenization and Purification

Sample extracts were prepared by grinding two grams of peanut seeds in fine powder and then dissolved in 40 ml of PBS-T (0.05% Tween in phosphate buffered saline, pH 7.4) containing 1M NaCl in 50 ml falcon tubes (Sarstedt No:55.476). After 2 h of gentle stirring on a rocking platform at room temperature, the aqueous phase was collected by centrifugation at 2,500 rpm at 4°C for 20 min. The aqueous phase was subsequently centrifuged at 3,500 rpm for 10 min at room temperature to remove residual traces and insoluble particles. Protein extracts were stored at −20°C until use.

Dilution Factor for Different Peanut Allergens

Dilution of a sample extract is critical for an ELISA which in turn determines the values of detection range for antibody and target antigen concentrations. Subsequently, the concentration of that specific allergens sample was estimated by multiplying the concentration found from the graph by the dilution factor (36, 37). By using different dilutions, we estimated the detection range of target antigen and antibody concentration. We standardized the dilution factors to detect each allergen proteins presents in the peanut seeds. Each sample was diluted in three different dilutions to detect the allergic protein present in seeds. The major allergen proteins such as Ara h 1 (1/1,000, 1/2,000, and 1/4,000), Ara h 2 (1/5,000, 1/10,000, and 1/20,000), Ara h 3 (1/5,000, 1/10,000, and 1/20,000), and Ara h 6 (1/40,000, 1/80,000, and 1/160,000) were diluted on a high range as compared to minor allergen protein, Ara h 8, which was diluted in low dilution range (1/10, 1/20, and 1/40). The optimal concentration of HRP-conjugated streptavidin was determined in the same way.

Steps Involved in ELISA

Antibody coating and blocking

Polystyrene microtiter plates (NUNC Maxisorp, Roskilde, Denmark) were coated with 100 μL mAb at 10 μL /10 ml

in 50 mM carbonate buffer (100 μL/well). After overnight incubation at 4°C, the coated wells were washed three times with washing buffer (phosphate buffer containing 0.05% Tween 20) and left to block with 1% BSA for 30 min at room temperature followed by three times washing with washing buffer.

Capture of allergen samples and standard

The standard and samples were diluted in washing buffer containing 1% bovine serum albumin fraction V (Sigma Aldrich Cat No. 10735086001). The standard of each allergen was diluted to make 10 serial doubling dilutions in dilution buffer. Subsequently, the allergen samples (100 μL/well) with three different dilutions were added in respective wells and incubated at room temperature for 1 h.

Adding of detection antibody in plate

After incubation plates were washed three times and specific biotinylated anti Ara h mAb were diluted to 1/1,000 in washing buffer containing BSA (1 mg/mL) was added to the wells (100 μL/well) and incubated for 1 h at room temperature.

Streptavidin-enzyme conjugate

After three washes, HRP-conjugated Streptavidin diluted to 1/1,000 in washing buffer containing BSA (1 mg/mL) was added to the wells and incubated for 1 h at room temperature.

Addition of the substrate

The colorimetric substrate was added to the wells and which formed a colored solution when catalyzed by the enzyme. The wells were washed three times and 100 μL of 1 mM ABTS was added to each well. After 5 min, the color development was observed.

Detection through ELISA reader

The optical density (OD) was measured at 405 nm using BioRad Microplate Reader and the data were processed using Microplate Manager V 6.1 (Bio-Rad Laboratories).

Spike and Recovery Studies

To test the accuracy of peanut allergens estimation, known amount of each allergen was spiked in the peanut extracts. The spike standard concentration for each allergen was ranging from 100 to 1,600 ng/ml for Ara h 1, 25 to 200 ng/ml for Ara h 2, and 6.25 to 100 ng/ml for Ara h 3, 10 to 160 for Ara h 3, and from 10 to 100 ng/ml for Ara h 8. These known amounts of individual standard (Ara h 1, Ara 2, Ara 3 Ara h 6, and Ara h 8) allergens were spiked in peanut extract and later calculating their final content in extract. The recovery was calculated as (B–C)/(A × 100), Where A = Known amount of peanut standard, B = Concentration of spiked standard, C = Concentration of peanut extract.

Statistical Analysis and Sample Analysis Design

All experiments were conducted with three replications based on dilution factor (DF). Mean ELISA plate reading values (Optical Density; OD) for each standard and sample were used to plot a standard curve by placing each allergen standard concentration values on Y axis and respective OD values on X axis on an Excel

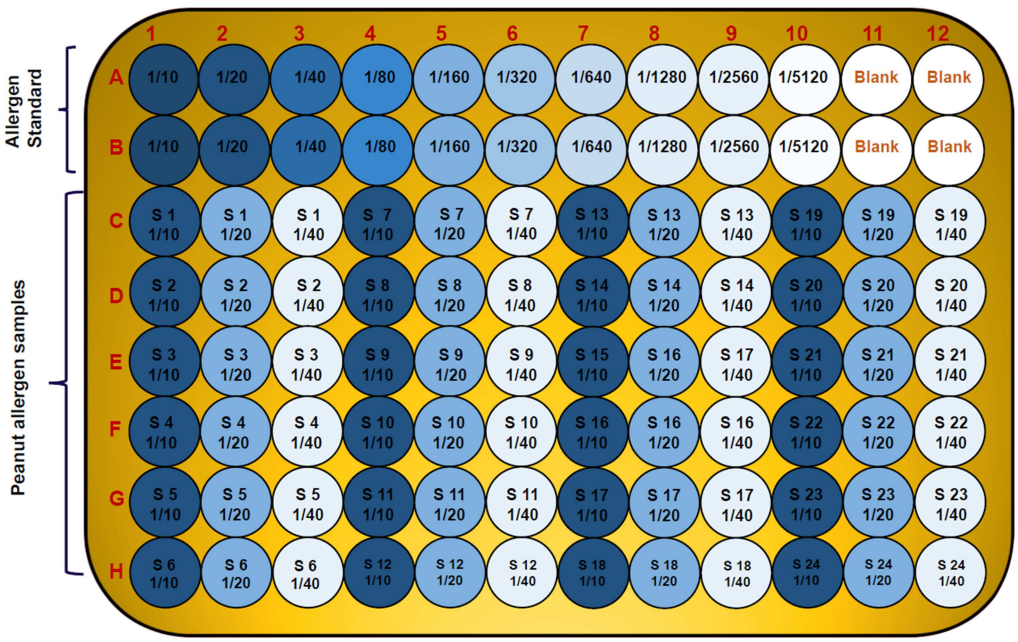

FIGURE 2 | Diagram showing 96-well plate design and the serial dilution of allergen standard and peanut samples in three replicates and three dilutions. The wells A1 to A10 and B1 to B10 contain serial dilution of specific allergen standard; A11 to A12 and B11 to B12 are blank; while the C, D, E, F, G, and H wells are with unknown peanut samples in three different dilutions for specific allergen.

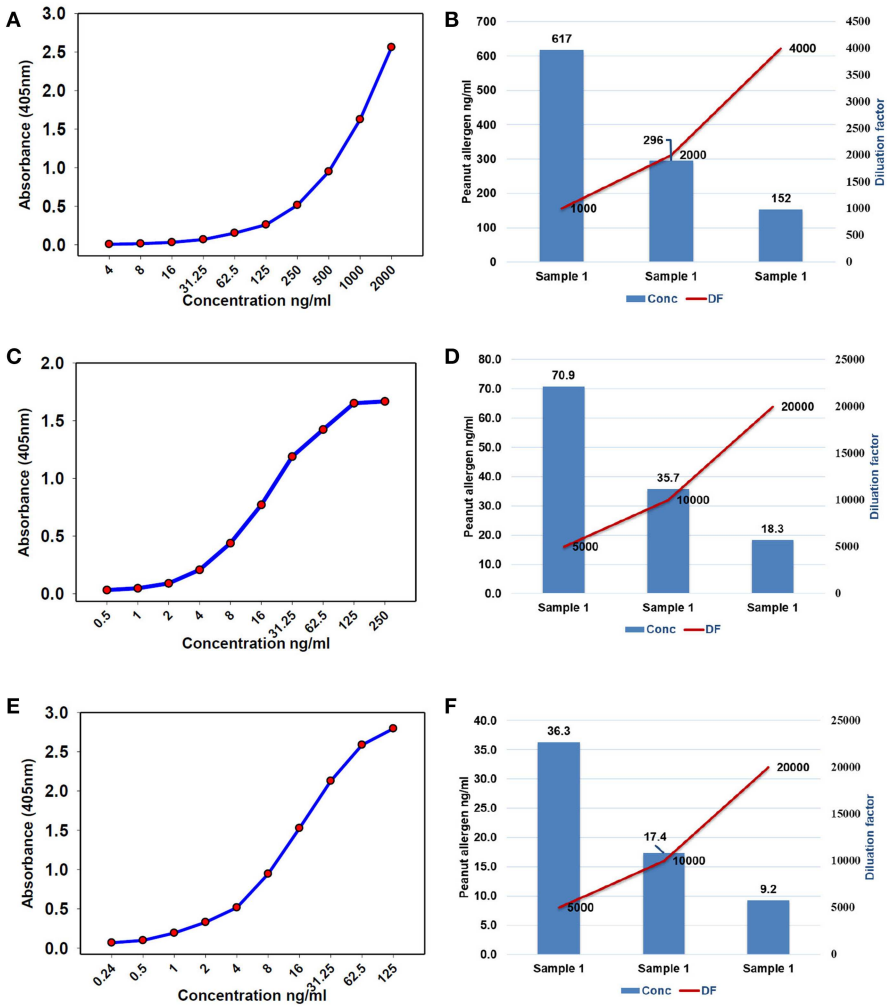

FIGURE 3 | Standard curve and optimized dilutions for Ara h 1, Ara h 2, and Ara h 3 allergens in peanut seeds. **(A)** standard curve for Ara h 1 allergen and **(B)** standardization of dilution factor to estimate Ara h 1. **(C)** standard curve for Ara h 2 allergen and **(D)** standardization of dilution factor to estimate Ara h 2. **(E)** standard curve for Ara h 3 allergen, and **(F)** standardization of dilution factor to estimate Ara h 3.

spreadsheet on the computer. Using regression equation values, we estimated specific allergen concentration in each sample. Data are expressed as means ± standard deviation (SD). Allergen standards and samples were placed in the 96- well format plate as given in **Figure 2**. The statistical analysis of data was performed using SigmaPlot 11.0.

RESULTS

Optimization of Dilution Factors (DFs) for Different Allergens From Peanut Seeds

To find the concentration of the different types of unknown allergens in the peanut seed samples, dilutions were performed on the different scale for specific allergens so that the relative concentrations fall according to their standard graph. Finally, after the absorbance of the specific allergens was determined, the concentrations of these specific allergens were determined by finding the location on the graph that corresponds to the absorbance of the standard. The concentrations of the allergens sample were estimated by multiplying samples concentration with dilution factor. For Ara h 1, the standard ranges from 0.004 to 2 μg/mL (**Figure 3A**), while it ranged from 0.0005 to 0.25 μg/mL and 0.00024 to 0.125 μg/mL for Ara h 2 and Ara h 3 (**Figures 3C,E**), respectively. The standard graph range of Ara h 6 and Ara h 8 was showing range between 0.0002 to 0.10 μg/mL and 0.0005 to 0.250 μg/mL (**Figures 4A,C**).

Based on the standard graph, the dilution factor was optimized on the basis of OD value and dilution of the peanut samples. Each allergen protein was estimated at different dilution factors (DF) in same samples. Here we used a number of dilutions in the peanut samples to detect the specific allergic protein in seeds. The Ara h 1 was detected on three serial doubling dilutions, 1:1,000, 1:2,000, and 1:4,000 (**Figure 3B**) while Ara h 2 and Ara h 3 detected on same dilution 1: 5,000, 1:10,000, and 1: 20,000 (**Figures 3D,F**). In peanut seeds, the Ara h 6 was detected in the high range (1:40,000, 1:80,000, and 1:160,000) dilution factor

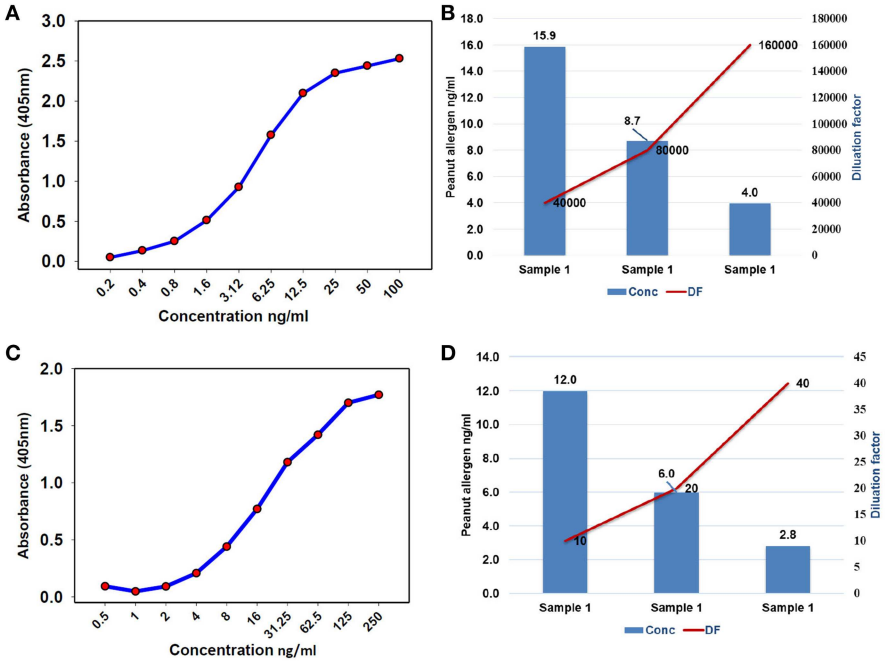

FIGURE 4 | Standard curve and optimized dilutions for Ara h 6 and Ara h 8 allergens in peanut seeds. **(A)** standard curve for Ara h 6 allergen by sandwich ELISA, and **(B)** standardization of dilution factor to estimate Ara h 6. **(C)** standard curve for Ara h 8 allergen by sandwich ELISA, and **(D)** standardization of dilution factor to estimate Ara h 8.

(**Figure 4B**), while Ara h 8 detected in a low range of dilution i.e., 1:10, 1:20, and 1:40 (**Figure 4D**). The result showed that in 1,000 dilution, the amount of Ara h 1 were detected 617 ng/ml while in 4000 dilution it decreased upto 152 ng/ml (**Figures 3B**). Likewise other peanut allergens Ara h 2, Ara h 3, Ara h 6, and Ara h 8 also showing similar pattern in their optimized result. The optimized dilution graphs (**Figures 3D,F**; **4B,D**) showing that when we increasing the dilution, peanut allergen concentration is decreasing accordingly.

Allergens Estimation in Peanut Seeds

The specificity of the sandwich ELISA was investigated by testing sample extracts of 38 peanut genotypes using different dilution factors for different allergens. Here, we used a single sample with three different dilutions and estimated four major (Ara h 1, Ara h 2, Ara h 3, and Ara h 6) and one minor (Ara h 8) allergen proteins in peanut seeds using specific antibodies. **Supplementary Table 1** showed the selected genotypes on the basis of minimum and the maximum amount of allergens present in specific genotypes. The minimum and maximum amount of Ara h 1 was detected in ICG 311 (12.5 μg/g) and ICG 15380 (42733.87 μg/g) genotypes. The low and high level of Ara h 2 was detected in ICG 532 (22.2 μg/g) and ICG 14482 (20600.3 μg/g) while ICG 311 (157.8 μg/g) and ICG 12879 (14556.4 μg/g) showing low and high amount of Ara h 3, respectively. Genotype ICG 3992 (5767.6 μg/g) contained low amount of allergen while ICGV 91116 (42317.6 μg/g) contained high amount of Ara h 6 allergen among the genotypes. Ara h 8 showed low range of allergen ranging from lowest in ICG 311(0.084 μg/g) to highest in ICG 14705 (6.8 μg/g).

Repeatability and Reproducibility

To validate the protocol, we repeated our experiments in three technical replicates with three dilutions optimized for each specific allergen. The assay had a linearity of r^2 ranges from 0.987 to 0.999 for all the specific allergen concentration present in peanut seeds. The coefficient of variation (CV) of each peanut sample in triplicates was consistently <20% (**Table 1**). Assay repeatability and reproducibility was <15% CV as measured by using the three peanut samples (**Table 1**). We found that dilution factor and CV are the key features of this protocol. This method is fairly simple to estimate specific allergen protein present in a single peanut seed. The validation experiments showed linearity of the method and each specific allergen showing different range in different peanut seeds. By using this protocol, we identified wide range of specific allergens, Ara h 1 (77–42,609 μg/g) (**Figure 5A**, **Table 1**), Ara h 2 (265–5,426 μg/g) (**Figure 5B**, **Table 1**), Ara h 3 (382–12,676 μg/g) (**Figure 5C**, **Table 1**), Ara h 6 (949–43,375 μg/g) (**Figure 5D**, **Table 1**), and Ara h 8 (0.385–6 μg/g) (**Figure 5E**, **Table 1**) in peanut genotypes.

The accuracy of the developed method was determined by performing the spike and recovery test. The mean recoveries ($n = 3$) for all the five allergens, individually tested, were in the range of 81–115% (**Table 2**). This indicates that the developed method is accurate for all five peanut allergens through peanut extract.

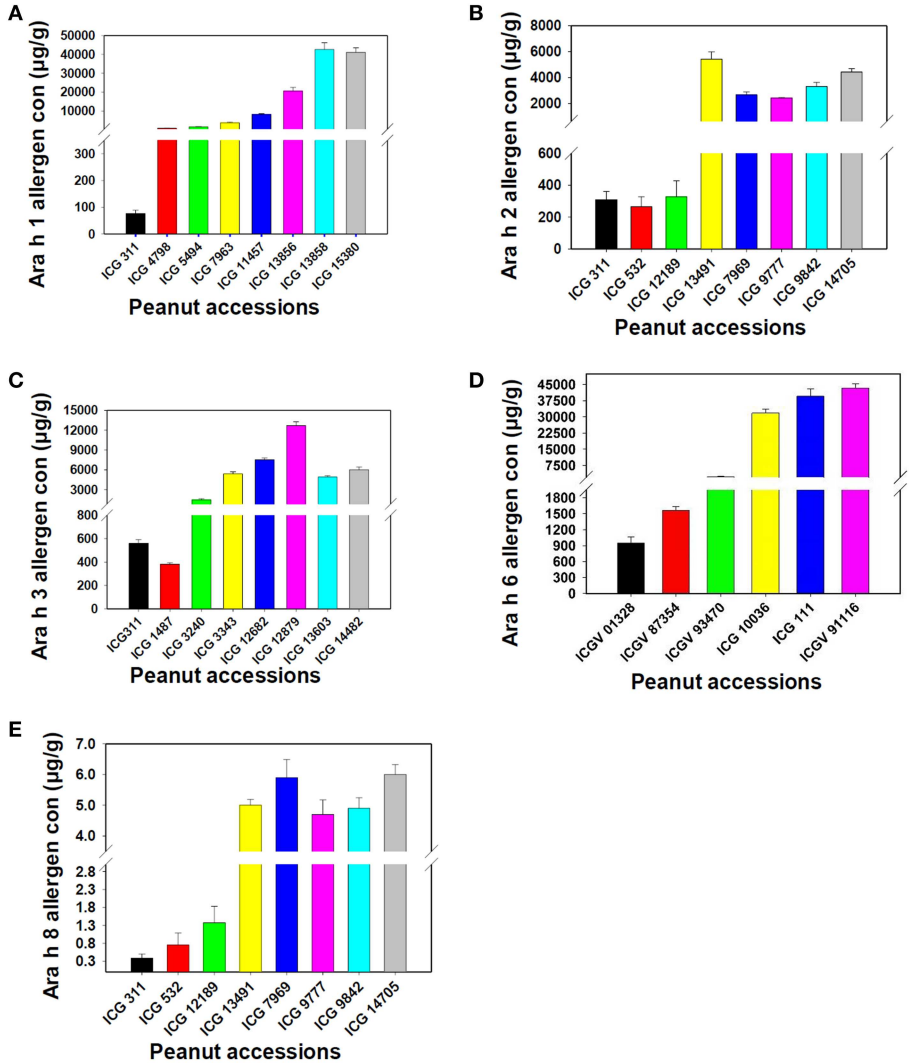

FIGURE 5 | Assessment of the new protocol for its reproducibility for Ara h 1, Ara h 2, Ara h 3, Ara h 6 and Ara h 8 allergen proteins. **(A)** Ara h 1 allergen proteins, **(B)** Ara h 2 allergen proteins, **(C)**: Ara h 3 allergen proteins, **(D)**: Ara h 6 allergen proteins and **(E)** Ara h 8 allergen protein. The figure shows variation in the presence of allergen protein in low to high range in different peanut genotypes.

DISCUSSION

Peanut allergy is one of the major food allergies and leads to death in case of severe allergenic reaction. In order to tackle this problem, efforts toward the development of peanut varieties with minimum or nil allergens needs to be initiated in addition to efforts for developing cure for the allergy caused due to peanut consumption. Peanut genetic improvement program for this particular trait require a reliable, robust and cost-effective assay to screen the large scale germplasm and breeding material for major allergens. In this context, this article reports development of a cost-effective and reliable ELISA-based protocol to estimate five major allergens (Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8) in peanut seeds.

In serological tests, two-fold differences in measurements of replicates of the same sample are considered as acceptable (38). Current studies describe the standardization of dilution factor for peanut seed sample to analyse the specific allergen protein concentration through ELISA method. It has been observed that the concentration and dilution factors are inverse means and the concentration of the allergens decreased with the increasing dilution of the samples. All major allergen proteins (Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8) concentration was gradually decreased while increasing the dilution. For instance, the concentration of Ara h 1 was 617 ng/mL in 1,000 dilution, and in 2,000 dilution the concentration of Ara h 1 was detected as 296 ng/mL while in 4,000 dilution the concentration of Ara h 1 was detected as 152 ng/mL (**Figure 3B**). On the basis of optimized dilution factors, major allergen proteins were estimated in peanut seeds with increased precision. The results from our study show the importance of standard curve and dilution as an important parameter for quantifying the ELISA results (39). Dilution factor

TABLE 2 | Recovery of five major allergens (Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8) from artificial spiked peanut allergens through ELISA in peanut seeds.

S. No.	Concentration used for spiking in samples (ng/ml)	Detected concentration of spiked samples (ng/ml)	% recovery of allergen in spiked samples	CV (%)
RECOVERY OF SPIKED Ara h 1 FROM SAMPLES				
1	100	105 ± 9	105.3	21.6
2	200	206 ± 8	103.0	22.2
3	400	376 ± 7	94.0	6.3
4	800	786 ± 10.9	98.3	17.6
5	1,600	1562 ± 11	97.6	11.4
RECOVERY OF SPIKED Ara h 2 FROM SAMPLES				
1	25	27 ± 3.2	106.8	10.9
2	50	57.9 ± 4.8	115.8	17.8
3	100	104.3 ± 0.3	104.4	12.8
4	150	148.7 ± 10.9	99.2	16.5
5	200	188.6 ± 7.6	94.3	20.7
RECOVERY OF SPIKED Ara h 3 FROM SAMPLES				
1	6.25	5.7 ± 0.77	91.4	9.9
2	12.5	12.17 ± 5.9	97.4	19.0
3	25	20.34 ± 3.24	81.4	7.0
4	50	53.45 ± 5.25	106.9	10.1
5	100	108.96 ± 3.8	109.0	9.4
RECOVERY OF SPIKED Ara h 6 FROM SAMPLES				
1	10	9.70 ± 2.9	97.0	15.5
2	20	17.94 ± 0.25	89.7	10.4
3	40	40.04 ± 4.31	100.1	14.8
4	80	83.97 ± 1.8	105.0	9.9
5	160	157.88 ± 4.37	98.6	8.9
RECOVERY OF SPIKED Ara h 8 FROM SAMPLES				
1	1.25	1.26 ± 0.15	100.78	5.26
2	2.5	2.48 ± 0.05	99.26	20.31
3	5	4.93 ± 0.36	98.53	16.16
4	10	53.45 ± 5.25	105.62	11.58
5	20	108.96 ± 3.8	100.56	12.30

*Each sample was spiked with known amount of peanut allergens in peanut extract. Data represents the mean of three replicates ± SD and CV%. Formula used for detection (% recovery) = (B–C)/A*100. Where A, Known amount of peanut standard; B, Concentration of spiked standard; C, Concentration of peanut extract.*

is an important step in ELISA before analyzing the results. Dilution assay apprises the actual concentration would be taken to experiment. To attained highly sensitive results for protein concentration, optimum values for dilution assay would provide a valid conclusion (39, 40). Earlier few attempts were made to quantify the levels of allergens in peanut seed using immuno assays and chromatography methods. But majority of them were expressed only in terms of percentages out of total protein present in the raw peanut seed. In one of the earlier studies, it was concluded that quantification of a single protein is not possible from a sample mixture of allergenic and non-allergenic proteins (41). Later in another study, using human sera from patients with documented history of peanut allergy, allergens from different peanut introductions were quantified and expressed in terms of ELISA optical density readings (4). In a recent study, Ara h 3 was quantified and found that it is the most abundant protein followed by Ara h 1 and Ara h 2 (34). However, in our study, Ara h 1 found in larger quantities followed by Ara h 6, Ara h 2, Ara h 3, and Ara h 8. Extraction method and proper dilutions of extracts plays crucial role in the precise quantification of these allergen proteins. Here in our study, we have thoroughly studied series of dilutions for each allergen and their impact on the quantification of allergens. These results clearly indicate three different dilutions for each major peanut allergen are required for precise detection in peanut seeds. Further the suggested 96-well plate design can allow the analysis of 24 samples in three replicates.

The newly developed ELISA protocol was validated on assay linearity, accuracy, repeatability and reproducibility. On the basis of primary results, few peanut genotypes were selected with low, medium and higher range of allergen proteins present in seeds. Our results indicated a different level of specific allergen profiles among peanut genotypes. Marked differences in specific peanut allergen profile were observed in peanut flour and peanut-based products such as peanut butter and other confectionary preparations for clinical use (42). In the serological assay, the variability produces continuous value were summarized by the coefficient of variation (CV), often reported in percentage (42, 43). The low CV observed in the current study while preparing dilutions provided more confidence in results produced for multiple genotypes using this protocol. Therefore, this protocol can provide accurate estimation for targeted five allergens and hence, can be very useful in accelerating the peanut research leading to detection and development of hypo-allergic peanuts.

The sample value obtained from an ELISA is dependent upon the interaction between the protein of interest and the ELISA's antibodies, and comparison of this interaction relative to a recombinant protein standard curve. We have validated this newly developed ELISA protocol for estimation of the major peanut allergens in peanut seeds by conducting the recovery experiments. The validation of ELISA protocol depends upon the recovery experiment which allow for a well-ordered presentation of the results (42).

CONCLUSIONS

This study reports successful development and validation of an accurate ELISA method for allergen estimation in peanut seeds. The method developed here is a simple, fast and cost-effective method and can be applied to a large number of samples for specific allergen estimation. Our results clearly indicates three different dilutions for each major peanut allergen are required for precise detection in peanut seeds. Further the suggested 96-well plate design can allow the analysis of 24 samples in three replicates. This protocol can be very useful in accelerating the research in identifying peanut genotypes with minimum allergen proteins. Furthermore, the high level of phenotypic variation in a selected set of peanut germplasm showed very positive indication and hope toward efforts of developing allergen-free peanut genotypes in coming years using genomics-assisted breeding including genome editing.

AUTHOR CONTRIBUTIONS

MP conceived the idea. AP performed the experiment. MP, HS, RV, and AP designed the experiments, analyzed the data, and wrote the manuscript.

FUNDING

This work has been undertaken as part of the CGIAR Research Program on Grain Legumes and Dryland Cereals (GLDC). ICRISAT is a member of CGIAR Consortium.

ACKNOWLEDGMENTS

Authors are thankful to Dr. Hari D. Upadhyaya for providing seeds of genotypes included in the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2019.00068/full#supplementary-material

REFERENCES

1. Waserman S, Watson W. Food allergy. *Allergy Asthma Clin Immunol.* (2011) 7:S7. doi: 10.1186/1710-1492-7-S1-S7
2. Sicherer SH, Sampson HA. Peanut allergy: emerging concepts and approaches for an apparent epidemic. *J Allergy Clin Immunol.* (2007) 120:491–503. doi: 10.1016/j.jaci.2007.07.015
3. Sicherer SH, Sampson HA. Food allergy: epidemiology, pathogenesis, diagnosis, and treatment. *J Allergy Clin Immunol.* (2014) 133:291–307. doi: 10.1016/j.jaci.2013.11.020
4. Dodo H, Marsic D, Callender M, Cebert E, Viquez O. Screening 34 peanut introductions for allergen content using ELISA. *Food Agri Immunol.* (2002) 14:147–54. doi: 10.1080/09540100220145179
5. Sicherer SH, Sampson HA. Food allergy. *J Allergy Clin Immunol.* (2010) S116–25. doi: 10.1016/j.jaci.2009.08.028
6. Lanser BJ, Wright BL, Orgel KA, Vickery BP, Fleischer DM. Current options for the treatment of food allergy. *Pediatric Clin.* (2015) 62:1531–49. doi: 10.1016/j.pcl.2015.07.015
7. Cabanillas B, Maleki SJ, Rodríguez J, Burbano C, Muzquiz M, Jiménez MA, Pedrosa MM, Cuadrado C, Crespo JF (2012). Heat and pressure treatments effects on peanut allergenicity. *Food chemistry*. 132, 360–6. doi: 10.1016/j.foodchem.2011.10.093
8. Wen HW, Borejsza-Wysocki W, DeCory TR, Durst RA. Peanut allergy, peanut allergens, and methods for the detection of peanut contamination in food products. *Compr Rev Food Sci Food Safety.* (2007) 6:47–58. doi: 10.1111/j.1541-4337.2007.00017.x
9. Maleki SJ, Chung SY, Champagne ET, Raufman JP. The effects of roasting on the allergenic properties of peanut proteins. *J Allergy Clin Immunol.* (2000) 106:763–8. doi: 10.1067/mai.2000.109620
10. Lehmann K, Schweimer K, Reese G, Randow S, Suhr M, Becker WM, et al. Structure and stability of 2S albumin-type peanut allergens: implications for the severity of peanut allergic reactions. *Biochem J.* (2006) 395:463–72. doi: 10.1042/BJ20051728
11. Hamilton RG, MacGlashan DW, Saini SS. IgE antibody-specific activity in human allergic disease. *Immunol Res.* (2010) 47:273–84. doi: 10.1007/s12026-009-8160-3
12. Lieberman JA, Sicherer SH. Diagnosis of food allergy: epicutaneous skin tests, in vitro tests, and oral food challenge. *Curr Allergy Asthma Rep.* (2011) 11:58–64. doi: 10.1007/s11882-010-0149-4
13. Nicolaou N, Murray C, Belgrave D, Poorafshar M, Simpson A, Custovic A. Quantification of specific IgE to whole peanut extract and peanut components in prediction of peanut allergy. *J Allergy Clin Immunol.* (2011) 127:684–5. doi: 10.1016/j.jaci.2010.12.012
14. Palladino C, Breiteneder H. Peanut allergens. *Mol Immunol.* (2018) 100:58–70. doi: 10.1016/j.molimm.2018.04.005
15. Hurlburt BK, Offermann LR, McBride JK, Majorek KA, Maleki SJ, Chruszcz M. Structure and function of the peanut panallergen Ara h 8. *J Biol Chem.* (2013) 288:36890–901. doi: 10.1074/jbc.M113.517797
16. Mittag D, Akkerdaas J, Ballmer-Weber BK, Vogel L, Wensing M, Becker WM, et al. Ara h 8, a Bet v 1–homologous allergen from peanut, is a major allergen in patients with combined birch pollen and peanut allergy. *J Allergy Clin Immunol.* (2004) 114:1410–7. doi: 10.1016/j.jaci.2004.09.014
17. Riecken S, Lindner B, Petersen A, Jappe U, Becker WM. Purification and characterization of natural Ara h 8, the Bet v 1 homologous allergen from peanut, provides a novel isoform. *Biol Chem.* (2008) 389:415–23. doi: 10.1515/BC.2008.038
18. Kondo Y, Urisu A. Oral allergy syndrome. *Allergol Int.* (2009) 58:485–91. doi: 10.2332/allergolint.09-RAI-0136
19. Pele M. Peanut allergens. *Roman Biotechnol Lett.* (2010) 15:5204–12. Available online at: https://e-repository.org/rbl/vol.15/iss.2/19.pdf
20. de Jong EC, van Zijverden M, Spanhaak S, Koppelman SJ, Pellegrom H, Penninks AH. Identification and partial characterization of multiple major allergens in peanut proteins. *Clin Exp Allergy.* (1998) 28:743–51. doi: 10.1046/j.1365-2222.1998.00301.x
21. Koppelman SJ, De Jong GAH, Laaper-Ertmann MD, Peeters KABM, Knulst AC, Hefle SL, et al. Purification and immunoglobulin E-binding properties of peanut allergen Ara h 6: evidence for cross-reactivity with Ara h 2. *Clin Exp Allergy.* (2005) 35:490–7. doi: 10.1111/j.1365-2222.2005.02204.x
22. Palmer GW, Dibbern DA Jr, Burks AW, Bannon GA, Bock SA, Porterfield HS, et al. Comparative potency of Ara h 1 and Ara h 2 in immunochemical and functional assays of allergenicity. *Clin Immunol.* (2005) 115:302–12. doi: 10.1016/j.clim.2005.02.011
23. Porterfield HS, Murray KS, Schlichting DG, Chen X, Hansen KC, Duncan MW, et al. Effector activity of peanut allergens: a critical role for Ara h 2, Ara h 6, and their variants. *Clin Exp Allergy.* (2009) 39:1099–108. doi: 10.1111/j.1365-2222.2009.03273.x
24. Koppelman SJ, Knol EF, Vlooswijk RAA, Wensing M, Knulst AC, Hefle SL, et al. Peanut allergen Ara h 3: isolation from peanuts and biochemical characterization. *Allergy.* (2003) 58:1144–51. doi: 10.1034/j.1398-9995.2003.00259.x
25. Peeters KA, Koppelman SJ, Van Hoffen E, Van Der Tas CW, den Hartog Jager CF, Penninks AH, et al. Does skin prick test reactivity to purified allergens correlate with clinical severity of peanut allergy?. *Clin Exp Allergy.* (2007) 37:108–15. doi: 10.1111/j.1365-2222.2006.02628.x
26. Flinterman AE, Van Hoffen E, den Hartog Jager CF, Koppelman S, Pasmans SG, Hoekstra MO, et al. Children with peanut allergy recognize predominantly Ara h2 and Ara h6, which remains stable over time. *Clin Exp Allergy.* (2007) 37:1221–8. doi: 10.1111/j.1365-2222.2007.02764.x
27. Wu X, Li Y, Liu B, Feng Y, He W, Liu Z, et al. Two-site antibody immunoanalytical detection of food allergens by surface plasmon resonance. *Food Anal Methods.* (2016) 9:582–8. doi: 10.1007/s12161-015-0232-5
28. Hefle SL, Lambrecht DM. Validated sandwich enzyme-linked immunosorbent assay for casein and its application to retail and milk-allergic complaint foods. *J Food Prot.* (2004) 67:1933–8. doi: 10.4315/0362-028X-67.9.1933
29. Jayasena S, Smits M, Fiechter D, De Jong A, Nordlee J, Baumert J, et al. Comparison of six commercial ELISA kits for their specificity and sensitivity in detecting different major peanut allergens. *J Agric Food Chem.* (2015) 63:1849–55. doi: 10.1021/jf504741t
30. Stephan O, Vieths S. Development of a real-time PCR and a sandwich ELISA for detection of potentially allergenic trace amounts of peanut (Arachis hypogaea) in processed foods. *J Agric Food Chem.* (2004) 52:3754–60. doi: 10.1021/jf035178u

31. Asensio L, González I, García T, Martín R. Determination of food authenticity by enzyme-linked immunosorbent assay (ELISA). *Food Control.* (2008) 19:1–8. doi: 10.1016/j.foodcont.2007.02.010
32. You J, Li D, Qiao S, Wang Z, He P, Ou D, et al. Development of a monoclonal antibody-based competitive ELISA for detection of β-conglycinin, an allergen from soybean. *Food Chem.* (2008) 106:352–60. doi: 10.1016/j.foodchem.2007.05.040
33. Koppelman SJ, Vlooswijk RAA, Knippels LMJ, Hessing M, Knol EF, Van Reijsen FC, et al. Quantification of major peanut allergens Ara h 1 and Ara h 2 in the peanut varieties Runner, Spanish, Virginia, and Valencia, bred in different parts of the world. *Allergy.* (2001) 56:132–7. doi: 10.1034/j.1398-9995.2001.056002132.x
34. Koppelman SJ, Jayasena S, Luykx D, Schepens E, Apostolovic D, De Jong GA, et al. Allergenicity attributes of different peanut market types. *Food Chem Toxicol.* (2016) 91:82–90. doi: 10.1016/j.fct.2016.02.016
35. Chen H, Zou Z, Tao A. A quantitative method for detecting Ara h 2 by generation and utilization of monoclonal antibodies. *J Immunol Res.* (2018) 2018:4894705. doi: 10.1155/2018/4894705
36. Engvall E. Quantitative enzyme immunoassay (ELISA) in microbiology. *Med Biol.* (1977) 55:193–200.
37. Voller A, Bartlett A, Bidwell DE. Enzyme immunoassays with special reference to ELISA techniques. *J Clin Pathol.* (1978) 31:507–20. doi: 10.1136/jcp.31.6.507
38. Wood RJ, Durham TM. Reproducibility of serological titers. *J Clin Microbiol.* (1980) 11:541–5.
39. Thakur K, Sharma S, Prabhakar S, Gupta P, Anand A. Revisiting the dilution factor as vital parameter for sensitivity of ELISA assay in CSF and Plasma. *Ann Neurosci.* (2015) 22:37. doi: 10.5214/ans.0972.7531.220108
40. Andreasson U, Perret-Liaudet A, van Waalwijk van Doorn LJ, Blennow K, Chiasserini D, Engelborghs S, et al. A practical guide to immunoassay method validation. *Front Neurol.* (2015) 6:179. doi: 10.3389/fneur.2015.00179
41. Kleber-Janke T, Crameri R, Appenzeller U, Schlaak M, Becker WM. Selective cloning of peanut allergens, including profilin and 2S albumins, by phage display technology. *Int Arch Allergy Immunol.* (1999) 265:74. doi: 10.1159/000024203
42. Filep S, Block DS, Smith BR, King EM, Commins S, Kulis M, et al. Specific allergen profiles of peanut foods and diagnostic or therapeutic allergenic products. *J Allergy Clin Immunol.* (2018) 141:626–31. doi: 10.1016/j.jaci.2017.05.049
43. Reed GF, Lynn F, Meade BD. Use of coefficient of variation in assessing variability of quantitative assays. *Clin Diag Lab Immunol.* (2002) 9:1235–9. doi: 10.1128/CDLI.9.6.1235-1239.2002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Pandey, Varshney, Sudini and Pandey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Nutrition

REVIEW
published: 07 June 2019
doi: 10.3389/fnut.2019.00081

Zinc and Iron Nutrition Status in the Philippines Population and Local Soils

*Alvin D. Palanog[1,2,3], Mark Ian C. Calayugan[1,2], Gwen Iris Descalsota-Empleo[1,4], Amery Amparado[1], Mary Ann Inabangan-Asilo[1], Emily C. Arocena[3], Pompe C. Sta. Cruz[2], Teresita H. Borromeo[2], Antonio Lalusin[2], Jose E. Hernandez[2], Cecilia Acuin[1], Russell Reinke[1] and B. P. Mallikarjuna Swamy[1]**

[1] Strategic Innovation Platform, International Rice Research Institute, Los Baños, Philippines, [2] College of Agriculture and Food Science, University of the Philippines, Los Baños, Philippines, [3] PhilRice Negros, Philippine Rice Research Institute, Science City of Muñoz, Philippines, [4] College of Agriculture, University of Southern Mindanao, Kabacan, Philippines

OPEN ACCESS

Edited by:
Reyazul Rouf Mir, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India

Reviewed by:
Mohd Anwar Khan, Sher-e-Kashmir University of Agricultural Sciences and Technology, India
Simonetta Friso, University of Verona, Italy
Rajneesh Paliwal, International Institute of Tropical Agriculture (IITA), Nigeria

***Correspondence:**
B. P. Mallikarjuna Swamy
m.swamy@irri.org

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Nutrition

Received: *03 March 2019*
Accepted: *20 May 2019*
Published: *07 June 2019*

Citation:
Palanog AD, Calayugan MIC, Descalsota-Empleo GI, Amparado A, Inabangan-Asilo MA, Arocena EC, Sta. Cruz PC, Borromeo TH, Lalusin A, Hernandez JE, Acuin C, Reinke R and Swamy BPM (2019) Zinc and Iron Nutrition Status in the Philippines Population and Local Soils. Front. Nutr. 6:81. doi: 10.3389/fnut.2019.00081

The Philippines is one of the major rice-producing and rice-consuming countries of Asia. A large portion of its population depends on rice for their daily caloric intake and nutritional needs. The lack of dietary diversity among poor communities has led to nutritional consequences, particularly micronutrient deficiencies. Iron-deficiency anemia (IDA) and zinc deficiency (ZnD) are two serious nutritional problems that affect the health and economic sector of the country. Since rice dominates the Filipino diet by default, biofortification of rice will help improve the micronutrient status. The Philippine government has proactively initiated various programs and policies to address micronutrient deficiencies, particularly through fortification of basic food commodities. Biofortification, the fortification of rice with micronutrients through breeding, is considered the most sustainable and cost-effective strategy that can benefit large vulnerable populations. However, developing promising genotypes with micronutrient-enriched grains should be coupled with improving micronutrient bioavailability in the soil in order to optimize biofortification. This review documents the prevailing soil Zn-deficiency problems in the major rice production areas in the Philippines that may influence the Zn nutritional status of the population. The article also reports on the biofortification efforts that have resulted in the development of two biofortified varieties approved for commercial release in the Philippines. As nutritional security is increasingly recognized as a priority area, greater efforts are required to develop biofortified rice varieties that suit both farmers' and consumers' preferences, and that can address these critical needs for human health in a sustainable and cost-effective manner.

Keywords: biofortification, iron deficiency, micronutrients, rice, zinc deficiency

INTRODUCTION

Rice is the most important staple food consumed by more than half of the world's population. Unfortified modern rice varieties with small amounts of micronutrients in the grains supply only a fraction of the daily individual requirements. Dependence on rice as the major dietary source of micronutrients contributes to micronutrient deficiency.

The Philippines accounts for 2.8% of global rice production and is the eighth largest rice producer in the world (1). The annual per capita consumption of rice in the country is estimated at 109.5 kg, which provides 56% of the calories, 37% of the proteins, and 3.8% of the fats required per person per day (2). However, most high-yielding and popular varieties currently grown are poor sources of micronutrients such as iron (Fe) and zinc (Zn), especially in their polished form (3, 4), and contain an average of 2 ppm of Fe and 12 ppm of Zn (5). Meanwhile, brown rice contains 6.3–24.4 ppm of Fe and 13.5–28.4 ppm of Zn. The endosperm, bran, and hulls of rice grain contain about 34.6, 27.0, and 38.4% of Fe and 56.6, 23.8, and 19.6% of Zn, respectively (6). The amounts of these micronutrients decrease after polishing (6).

The daily Zn requirement of an adult individual ranges from 8 to 11 mg per day while pregnant and lactating women require the highest intakes at 11–13 mg per day (7). The daily Fe requirement in adults varies from 12 to 28 mg per day while pregnant and lactating women require 30–38 mg per day (8). Current food intakes are unable to meet these requirements. Iron supplements are therefore provided to pregnant and lactating women, and zinc supplements are part of child diarrhea management. Iron fortification has been mandated by law for rice and flour (more on this below). But these efforts are insufficient as at present micronutrient malnutrition affects a large portion of the Philippine population (9) and is an important public health concern domestically.

Iron-deficiency anemia (IDA) is the most prevalent micronutrient deficiency affecting a large proportion of infants (40.5%), pregnant women (26.4%), lactating women (16.7%), and elderly males (23.0%) in the country. Anemia can cause profound impact on human health and productivity and Fe deficiency has pronounced effects on the first 1,000 days of life (10).

Zinc deficiency (ZnD) is also prevalent in 25.6% of the population based on serum Zn level tests. The elderly >60 years old (36.3%), adults 20–59 years old (28.1%), and lactating women (25.2%), followed by adolescents (23.6%), and school children (21.6%), are at high risk for Zn deficiency among the different populations and physiological groups in the country (2). ZnD contributes to impaired immune systems and slow child growth.

In the Philippines, rice is seen as an effective delivery system for food fortification because of its high per capita consumption (109 kg/person/day). An increment in micronutrient content in rice would positively impact human nutrition and health (11). Food fortification is considered a cost-effective approach in increasing the micronutrient content of rice although it poses several disadvantages: fortified foods should be consumed in an adequate amount, they should not affect the sensory properties of foods, and they need a safe delivery system and continued investments (12, 13). A more cost-effective and sustainable approach for rice fortification is biofortification. It is a process of increasing the concentrations of bioavailable micronutrients in the edible parts of the plant by traditional breeding, genetic engineering, or agronomic approaches (14).

Conventional plant breeding and genetic engineering are means of improving the genetic traits of crops to enhance the accumulation of bioavailable micronutrients (15). Both approaches have proven effective in increasing Zn content while genetic engineering is more effective in improving Fe content. Conventional plant breeding is a time-consuming process and its effectiveness is influenced by the availability of micronutrients in the soil (16). Thus, it is important to consider the various soil or edaphic factors and management options that improve Fe and Zn soil availability, which influences the accumulation of micronutrients in biofortified rice.

In this review, we briefly discuss rice production and dietary consumption in the Philippines and the importance of Fe and Zn in the health of Filipinos. This article highlights the micronutrient deficiency status of the Philippines soil and Filipino population and their possible association. It also gives an overview of the progress in rice biofortification efforts in the country and the programs and interventions that are implemented by the government to address micronutrient deficiency problems.

RICE IS AN IMPORTANT COMPONENT OF THE FILIPINO DIET AND AN EFFICIENT VEHICLE FOR FORTIFICATION

The Philippines is a rice-producing and -consuming country. The area allotted for rice production in the country was about 4.81 million hectares in 2017 (17), the majority of which was irrigated (69.7%), followed by rainfed (30.0%), and the rest was upland and saline-prone areas (18). Majority of the rice farmers use high quality inbred rice seeds (45%), farmers good seeds (FGS) {43%}, hybrid seeds (9%) and lastly, traditional varieties (3%). Highest volume produced was from inbred rice with 8.3 million metric tons (mmt) followed by hybrid (2.08 mmt), FGS (0.16 mmt), and traditional varieties (0.06 mmt). Largest area was planted with FGS (2.5 million hectares) followed by inbred seeds (∼2.0 million hectares) and both hybrid and traditional varieties were planted to < 1 million hectares (17). Rice production increased significantly from 10.5 million metric tons (mmt) in 1995 to 17.63 mmt in 2016, with a productivity of 4.38 tons per hectare (19, 20). Per capita rice consumption (PCRC) has increased significantly over the last decade, with an average of 109.5 kg in 2015 (21). PCRC is used in estimating rice import requirements and setting rice self-sufficiency targets (22). About 10% of the annual consumption requirement is augmented from imports. In 2010 and 2011, the Philippines was the biggest rice importer globally, importing 2.38 million tons. Supplying enough rice for every Filipino exerts a lot of pressure on the domestic rice industry, especially with the constantly growing population, increasing at a rate of almost 2% per annum (18). PCRC has risen by 13% nationally and has risen in all regions. Central Visayas has the largest increase of 41%, while Calabarzon has the least increase of 4% (23). Most of the low-income families live in rural areas, spend more on rice, and have higher PCRC than high-income and urban people (8, 22). A dietary survey conducted by the Department of Science and Technology—Food and Nutrition Research Institute (DOST-FNRI) showed that rice intake was lowest among the highest-income households (1,120 grams per

day) and highest among low-income households (1,300 grams per day) (2).

A typical Filipino diet is composed of rice-vegetable-fish combinations, with the highest percentage (37.2%) of rice and rice products in the total consumption and hypothetical approximation of percent Estimated Average Requirement (EAR) met for Zn and Fe for various population groups (**Figure 1**) (2). The diet can be characterized as high in carbohydrates, marginal in protein, and low in fat (8). Rice supplied the highest daily per capita calories, protein, and fats to Filipinos (**Table 1**) compared with other commodities (8). Rice, in its unpolished form, provides good amounts of nutrients such as vitamins B1, B6, and E; protein; minerals; unsaturated fat; and dietary fiber that can be found particularly in the embryo of the grain. The largest part of the grain, the endosperm, is mostly composed of carbohydrates with some incomplete protein and traces of minerals. On the other hand, the bran is mainly composed of carbohydrate cellulose, traces of vitamin B, minerals (including Fe, phosphorus (P), magnesium (Mg), and potassium (K), and incomplete proteins (23). The bran layers and majority (75–90%) of vitamins and minerals are removed during polishing to produce the commonly consumed white rice rich in starch but relatively low in micronutrient content (5, 23).

Rice is a prime source of carbohydrate but it can also be a source of protein although it lacks several of the necessary amino acids required for good health; thus, it should be combined with other protein sources such as nuts, seeds, legumes, fish, or meat (26). Clearly, rice alone cannot provide the essential nutrients needed by the body to be healthy. The over dependence of Filipinos on rice as their main source of energy and nutrients, and the lack of diversity in diets, especially among the low-income class, make them vulnerable to various micronutrient deficiencies. Increasing the micronutrient concentration of rice will therefore have a significant impact on the nutritional status of rice-consuming Filipinos (11). A biofortified rice variety does not require additional input and can be grown by farmers in a sustainable way to produce micronutrient-dense rice grains that may be more accessible to nutrient-deficient populations, especially in agricultural areas (11).

IMPORTANCE OF IRON AND ZINC IN HUMANS

The human body requires almost 50 known nutrients to carry out vital metabolic functions properly. Insufficient amounts of even one of these nutrients will lead to adverse metabolic disturbances resulting in poor health, impaired growth, and development in children, sickness, and eventually high economic losses to society (27, 28). This article will focus only on the importance and health consequences of deficiency in two micronutrients in review: Fe and Zn.

Iron and Iron-Deficiency Anemia (IDA)

Iron acts as the central atom of hemoglobin (29) and a component of myoglobin (30), which functions in the storage of oxygen in muscle tissue, and of the cytochrome system (29), which is essential in the energy-releasing process of cellular respiration. In the body, it is present in the erythrocytes (60%), in readily mobilizable Fe stores (25%), and in myoglobin muscle tissue and in various enzymes involved in oxidative metabolism and other cell functions (31). Since Fe is involved in numerous and diverse cellular functions, a constant balance among iron uptake, transport, storage, and use is required to maintain Fe

TABLE 1 | Daily per capita calories, protein, and fat supplied by rice in the Philippine population from 2011 to 2017 (in grams).

Nutrients	2011	2012	2013	2014	2015	2016	2017
Calories	1139.08	1162.64	1130.94	1118.45	1097.99	1054.76	1156.51
Protein	23.93	24.43	23.76	23.50	23.07	22.16	24.30
Fat	5.74	5.86	5.70	5.64	5.54	5.32	5.83

Source: (21).

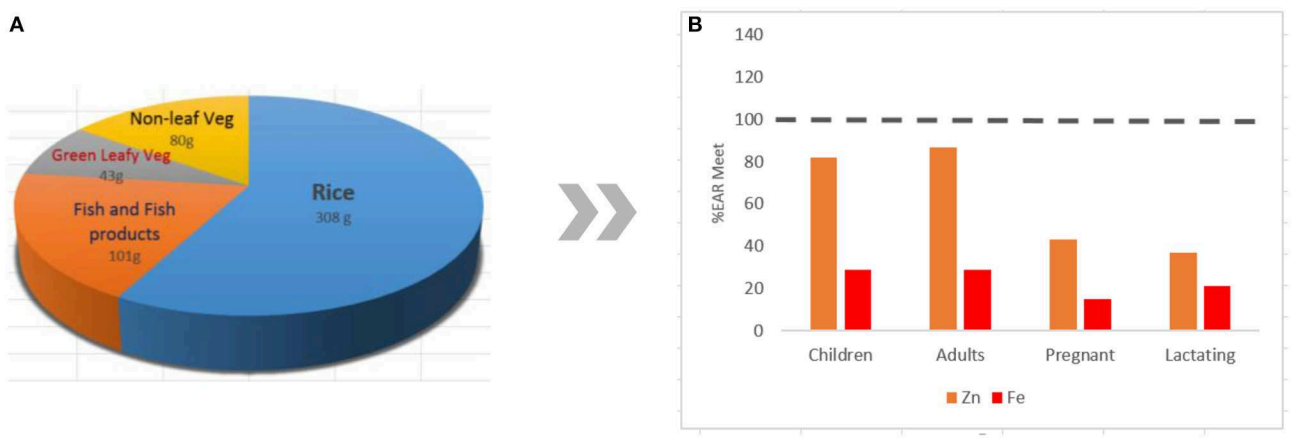

FIGURE 1 | The typical mean daily dietary composition of Filipinos **(A)** and the hypothetical approximation of percent Estimated Average Requirement (%EAR) of Zn and Fe met for various population groups **(B)** (2). Bioavailability in the body is not considered in the approximation.

homeostasis in the body (32). Imbalances in the uptake or absorption and metabolic demand result in the depletion of Fe stores, which leads to deficiency (32).

Iron deficiency results when the body's supply of available Fe is too low. People with this condition cannot produce an adequate amount of hemoglobin to meet the body's oxygen-transport needs. The relationship between Fe deficiency and brain function is an important consideration in developing strategies to combat the former (33–35). The liver and brain have the same magnitude of Fe in their structures. The supply of Fe to the brain cells starts in the early phase of brain development; thus, early deficiency leads to irreversible damage to the brain (36). In humans, about 10% of the normal content of Fe is found in the brain at birth and this content reaches 50% at the age of 10 years old, with the optimum amount reached at the age of 20–30 years old. Fe deficiency can also cause immune system impairment, which renders the individual susceptible to infection (37, 38). When the deficiency becomes severe, the condition is diagnosed as IDA. The cut-off levels for moderate and severe anemia are 100 and 70 g/L (hemoglobin/blood), respectively. Common symptoms of IDA include tiredness and weakness and paleness in the hands and eyelids due to inadequate oxygen supply to the body's cells. Further, IDA can cause major health concerns in an individual at different life stages and at different ages. It can cause underdeveloped cognitive functions and poor physical coordination in children <2 years of age (which can be irreversible depending on the duration and severity), reduced physical endurance, and impaired immune system among other symptoms (39). It also affects memory, intelligence, and sensory perception, which directly impact academic performance and later influence economic productivity and earning potential (39, 40). IDA has also been shown to negatively influence the results of iodine supplementation (12). During pregnancy, IDA can cause irreversible changes to fetal kidney and neural development (41, 42).

Zinc and Zinc Deficiency (ZnD)

Zinc plays essential roles in a wide variety of biochemical processes that affect growth, development, and reproduction and virtually all aspects of cellular metabolism (43). Zn has catalytic, structural, or regulatory roles and more than 300 mammalian enzymes are Zn-dependent (44, 45). Zn participates directly in enzyme catalysis or plays a role in the structural stability of proteins by supporting their folding and oligomerization (46). Further, Zn forms the prosthetic group of numerous enzymes as well as the receptor proteins for steroid and thyroid hormones and vitamins A and D (47). It also affects the function of enzymes involved in growth and development of bones (48). Zn enzymes are ubiquitous to all known classes of enzymes such as oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases that are involved in the synthesis and/or degradation of metabolic products such as carbohydrates, lipids, proteins, and nucleic acids (49).

Zinc deficiency (ZnD) results from an inadequate intake of absorbable Zn, an increase in Zn physiological requirements, inefficient absorption of Zn, increased losses, and impaired use of Zn. Serum Zn, the most available measure of ZnD, is used as an indicator of deficiency at the population level. The acceptable level of serum Zn is ∼65 μg/dL (50). Severe deficiency can cause growth retardation and stunting, changes in neuro-behavioral development, impaired reproduction, and immune disorders (48, 51). It can also contribute to vitamin A deficiency (VAD) since the lack of Zn can reduce the synthesis of retinol binding protein (52). ZnD in a newborn or growing animal is considered fatal (48). Recently, global awareness of the importance of Zn nutrition to human health has increased dramatically. Although any age group can be at risk, infants, children, adolescents, and pregnant women are especially at risk because of their high requirements (53). Infants and young children have an increased requirement for growth (54). Adolescents reach the peak of physiological Zn requirement during the time of their pubertal growth spurt, which occurs in 10–15-year-old girls and 12–15-year-old boys (55). Pregnant and lactating women have increased Zn nutritional demands during these periods, especially during lactation (56). Elder persons often have an inadequate intake of Zn even in rich countries due to reduced consumption of Zn-rich foods and decreased efficiency of Zn absorption with age (31). Zn is an important nutrient for humans and its deficiency poses a significant widespread risk to human health. Age and sex are established variables that affect serum Zn content. Marginal caloric intake consisting primarily of starchy foods and low consumption of meat in the diets of older children contributes to increased risks of ZnD. On the other hand, the vulnerability of males to ZnD is largely attributed to higher growth rate and greater proportion of muscle per kilogram of body weight since muscles contain higher Zn content than fat (57).

IRON AND ZINC STATUS OF THE PHILIPPINE POPULATION

Malnutrition is a large and growing problem worldwide with more than two billion people suffering from micronutrient malnutrition (58). It is estimated that 25% of pre-school children and 37% of non-pregnant women of reproductive age are Fe deficient (59). On the other hand, it is estimated that 17.3% of the world's population has low Zn intake, with the highest prevalence in Africa (23.9%) and Asia (19.4%) (60), contributing to global development challenges—eradicating hunger, reducing child and maternal mortality, and ending all forms of malnutrition (61)—and this can cause inter-generational consequences (27). Deficiencies of these micronutrients are concentrated in the semi-arid tropics, particularly in South and Southeast Asia and sub-Saharan Africa (62). Malnourished women may give birth to infants with low birth weight and greater risk of micronutrient deficiencies (61). Meeting nutritional requirements before and during pregnancy, at childbirth, and immediately post-partum positively impacts the survival, growth, and development of the fetus and newborn.

The 2015 Philippine National Nutrition Survey conducted by the Department of Science and Technology-Food and Nutrition Research Institute (DOST-FNRI) indicated that the mean per capita intake per day of meat and fish products, which are considered as the best sources of micronutrients, accounted

for only 7.3 and 11.9% of the total daily intake, respectively. Meanwhile, a large percentage (70%) of the total daily intake of the Filipino diet came from plant sources, primarily rice and corn (maize) that contain high amounts of phytate, a potent inhibitor of micronutrient absorption (63).

Iron-Deficiency Anemia (IDA) in the Philippines

Iron-deficiency anemia (IDA) is considered the most common form of anemia in the Philippine population (9, 64); hence, anemia is used as its proxy indicator. The highest prevalence of anemia in the country were observed in infants aged 6–11 months (40.1%), pregnant women (24.6%), elderly males (23.0%), elderly females (19.1%), and lactating women (16.7%) (65, 66). Rural areas have higher levels of anemia than urban areas (11.7 vs. 10.7%) (67). Moreover, the prevalence of anemia among children aged 6 months to 1 year from low-income families is considerably higher (16.5%) than among those from high-income families (7.9%) and this trend persists in older children (68). Anemia is more common among teenage pregnant women (30.6%) than among their adult counterparts (25.4%). More cases of pregnancy anemia were observed in urban areas (30.3%) than in rural areas (21.8%) (66).

IDA in pregnant women increases the incidence of infant mortality (the Philippines being among the countries with the highest rates in Southeast Asia), low birth weight, and premature birth (69). Although a substantial drop in the overall prevalence rate of anemia from 28.9% in 1993 to 11.1% in 2013 was observed, the prevalence rate among infants aged 6 months to <1 year is still at 40% and is considered a "severe" public health problem (66, 67).

IDA negatively affects the economy of the country in terms of "cost of illness" (which deals with health-care costs and productivity losses) and in terms of "burden of the disease," which identifies the impact on the affected individual in terms of years lived with disability (disability-adjusted years; DALYs) and years of life lost (51). Weiser et al. (51) estimated the cost of illnesses due to IDA, VAD, and ZnD in two age groups (6–23 and 24–59 months) of Filipino children by socioeconomic strata. The total lifetime costs of IDA, VAD, and ZnD amounted to USD 30 million for direct medical costs, USD 618 million for production losses, and 122,138 DALYs for intangible costs. The cost of IDA is mainly due to its impact on the mental development of children aged 6 to 23 months, which translated into 100% production losses and 84% DALY losses. For 24- to 59-month-old children, additional effects of IDA are limited to DALY losses because of reduced physical activity. Production and intangible losses are also mainly contributed by IDA (51). A decrease in cognitive performance in children due to Fe deficiency also leads to a decrease of 4% in hourly earnings in their later life (69). The estimated decrease in gross domestic product (GDP) is as much as 2% per year due to widespread Fe deficiency in the country that could be translated into income losses of PHP 172 or 0.9% in severely stunted workers, and for workers engaged in moderate (5%) and heavy (17%) physical labor (61). The bulk of the costs incurred came from projected lifetime costs, costs resulting from impaired mental and physical development, and costs of premature death. Results of the study emphasized the importance of addressing micronutrient deficiencies in high-risk infants and young children belonging to the lowest-income households (51).

Zinc Deficiency (ZnD) in the Philippines

ZnD ranks fifth among the top 10 leading illnesses and diseases in developing countries. In the Philippines, the first Zn status assessment study based on serum Zn level conducted by FNRI showed a high magnitude of prevalence of ZnD across population and physiological groups (63, 66). The occurrence rate of ZnD was high in school children (30.8%), male adolescents (32.2%), adults (31%), and lactating women (39.7%). The study covered 79 provinces in 17 regions, including the National Capital Region (NCR). It was observed that Zn-deficient pre-school and school children manifested a significantly higher rate of stunting than their normal counterparts. Stunting at an early age has been associated with poor cognitive ability (70) and increased child morbidity and mortality (71). A stunted child is also highly at risk for chronic diseases, fat-impaired oxidation, reduced glucose tolerance (31), and increased risk of severe infections such as diarrhea and acute respiratory infection (51, 72). Zn-deficient school children were also likely to have higher prevalence of anemia, VAD, and iodine deficiency. Anemia and stunting are the alternative indicators of ZnD (50).

ZnD has resulted in economic and human costs that have lifetime consequences, based on the study conducted by Weiser et al. (51). This study quantified the cost of micronutrient deficiencies (including ZnD) in the country. ZnD dominated the direct medical costs (89% of the total) because of the cost of treating diarrhea, respiratory diseases, and measles. Future production losses due to stunting brought about mainly by deficiency contributed 28% of the total production losses. However, compared with other micronutrient deficiencies (IDA and VAD), ZnD has lower intangible costs or DALYs mainly because of the absence of its impact on mortality (51). A systematic review of Zn supplementation trials showed that Zn had no significant effect on mortality (73).

Given the high prevalence of ZnD in the country that exceeds 20% (63), the International Zinc Nutrition Consultative Group (IZiNCG) recommended further assessment and planning for strategic intervention programs to be initiated (74). Hotz and Brown (74) suggested three intervention strategies: supplementation, fortification, and dietary diversification/modification. Current evidence indicates the beneficial effects of such interventions in reducing or eliminating the risk of ZnD (72). However, minimal efforts were carried out to control ZnD and resolve micronutrient malnutrition in the country (63).

Clearly, the impacts of IDA and ZnD on health and the economic sector of the country are alarming and require immediate attention. Furthermore, this highlights the importance of implementing intervention strategies to reach the vulnerable sectors of the population.

IRON AND ZINC STATUS OF PHILIPPINE SOILS

Micronutrients are essential in plant nutrition and production. They are required in minute quantities generally below 100 parts per million (ppm) in plant tissues (75) but are equally as important as macronutrients in plants. The relationship between micronutrient supply and crop growth has been examined and trace elements such as Zn, Mn, and copper (Cu) were shown to contribute to achieving higher yields (76, 77). Further, several studies suggest that the application of micronutrients to crops might result in more vigorous seedlings, increased disease resistance, and possibly improved drought tolerance (76, 78–80).

Iron in Plants, Soil, and the Rice System

Iron is an important micronutrient in all living organisms since it plays an important role in numerous metabolic processes such as DNA synthesis, respiration, and photosynthesis (81). In particular, it is involved in the synthesis of chlorophyll and maintaining chloroplast structure and function in plants. The photosynthetic cells of plants contain ∼80% of the total Fe in plants, emphasizing the important role of Fe in the electron transport system, biosynthesis of cytochromes and heme-molecules, and construction of Fe-S cluster assemblies (82). Essentially, the cytochromes act as electron carriers in the respiratory chain (83). Fe is a major component of plant redox systems because of its high affinity to active metalloprotein sites where it acts as a co-factor in redox reactions (84). For example, mitochondria contain a large amount of metalloproteins that require Fe to function (85). Several Fe-containing proteins are involved in both the respiratory chain and tricarboxylic acid cycle. Crucial steps in Fe-S cluster assembly for the entire cell take place in mitochondria, suggesting the importance of this compartment in Fe handling by the cell (86).

Fe is the fourth most abundant element in the lithosphere and is present in large quantities in the soil; however, it is also considered as the third most limiting nutrient in plant growth. Fe is predominantly found in the Fe^{3+} form mainly as a constituent of oxyhydroxide polymers whose very low solubility renders them insufficient for plant needs in aerobic environments (81, 87, 88) and alkaline soils (89). Deficiency in the soil is due to several factors: low concentration of soluble Fe^{2+} in upland soils; insufficient soil reduction under submerged conditions with low organic status; high-pH alkaline soils; excessive P fertilizer application; excessive concentrations of Mn, Cu, Zn, Mo, Ni, and Al in soil; and genotypes with low potential for excretion of organic acids to solubilize Fe (89). In contrast to ZnD, Fe deficiency is common in high-pH aerobic soils. Widespread Fe toxicity was observed in rice grown in wetland acid sulfate soils (90). However, Fe deficiency may occur in flooded soils when there is insufficient organic matter decomposition to drive the reduction of Fe^{3+} to Fe^{2+} and redox potential remains high. Fe deficiency in rice plants is manifested by interveinal chlorosis and yellowing of emerging leaves, poor root formation, and growth retardation (91). Since Fe is non-mobile within rice, symptoms initially appear on young leaves. In severe cases, the whole plant becomes chlorotic and dies. Fe deficiency is an important concern in dryland soils but it usually disappears 1 month after planting (89). Crops suffering from this deficiency have retarded growth and are more susceptible to diseases (16, 92). The deficiency of this nutrient leads to a decrease in dry matter production, chlorophyll content, and activity of enzymes involved in sugar metabolism.

Iron deficiency is a costly and difficult nutrient deficiency to correct. Unless applied in large amounts, inorganic Fe fertilizers are often ineffective in overcoming Fe deficiency. Deficiency is usually treated with the application of solid $FeSO_4$ or Fe chelates through broadcasting, banding, or foliar applications (93). Applying organic fertilizers such as rice crop residues or manures as nutrient sources also helps alleviate Fe deficiency in the soil. A combination of foliar application of $FeSO_4$ (3% solution) at different sowing intervals followed by soil application (150 and 305 kg $FeSO_4$) was found to be an efficient and cost-effective method in improving Fe availability in aerobic rice soil conditions (94).

Zinc in Plants, Soil, and the Rice System

Zinc is an essential micronutrient required for crop growth. Zn has chemical properties similar to those of other nutrient elements such as Fe and Mg. As such, competition between these elements can occur during absorption (95). Zn acts as an activator of many enzymes in plants and the metal component of enzymes or a functional structure, and regulatory co-factor of a large number of enzymes. It is also an essential component of RNA polymerase and a constituent of ribosomes, which is essential for their structural integrity (75). It is involved in several biochemical processes in the rice plant such as cytochrome (96) and nucleotide synthesis (97), chlorophyll production (96), and enzyme co-factor and activator, and it maintains membrane integrity (98–100) and auxin metabolism (101, 102). This micronutrient is bioavailable as Zn^{2+} in the soil solution or as exchangeable Zn on cation-exchange sites, organically complex Zn in solution, or organically complex Zn in soil solids. The normal exchangeable Zn extracted from the soil solution ranges from 0.05 to 0.25 ppm using water and ranges from 0.1 to 2 ppm using ammonium acetate. Lower values can be observed in soils where ZnD occurs in plants (103).

ZnD is common in rice soils, especially in anaerobic irrigated lowland fields, calcareous soils, and in rice soils with high pH and high organic content (75, 89, 103). The deficiency of this micronutrient in relation to flooded rice soil conditions has been well-documented (104, 105). Submergence in conventional flooding depletes soil oxygen, decreases redox potential, and increases pH, which eventually reduces the solubility of Zn in the soil (106). Meanwhile, low Zn concentration in calcareous soils is associated with high bicarbonate content and high pH (107) that inhibit translocation of Zn from roots to shoots (108). Commonly, transplanted seedlings (2–4 weeks) are susceptible to ZnD manifested by stunting, which eventually leads to delayed maturity and reduced yield (109). Zn and P ratio was observed to be related to the manifestation of ZnD. The interaction of Zn and P in the soil is highly affected by the infection of roots with vesicular arbuscular mycorrhiza on which infection of the roots facilitates absorption of Zn. However, increasing the supply of P in the soil strongly suppresses infection of the roots. The connection of P toxicity to ZnD is still not clear (110).

Plants grown in Zn-deficient soil tend to have stunted growth manifested by a reduction in stem length and shortening of internodes. Plant leaves exhibit rosetting, mottling, and interveinal chlorosis, and sometimes red spots appear due to the deposition of anthocyanin (75). Rice grown in Zn-deficient soils exhibits small brown spots and poor root development. Zn is one of the important micronutrients that affect growth and yield in rice and it has been an integral factor that determines the rice production of numerous countries. ZnD in soils is effectively corrected by Zn fertilization (89). It can be applied to soil through surface application or soil incorporation (105, 109) through seed treatment (111–114), foliar application (115–119), and by dipping seedlings into Zn solution (120–122) or in combinations. Numerous available Zn sources have been tested and studied individually or in combinations (103). Among the Zn sources, Zn sulfate is commonly used in rice and is usually applied through soil incorporation during flooding or after transplanting (89, 105). Although soil application has been proven to be an effective strategy in improving Zn concentration in tissues as well as improving yield in rice (123), it is not economically effective because of the high cost (103). Thus, other application methods or complementary approaches can be explored to maximize the availability of Zn in the soil at a lower cost.

Several agronomic approaches exist to improve the availability of Zn in rice production systems; these include tillage, crop rotations, and intercropping, manure application, fertilizer management (103), and water management. Franzluebbers and Hons (124) reported higher extractable P, K, and Zn in no-tillage land preparation vis-à-vis the conventional tillage system. However, Bhaduri and Purakayastha (125) recently reported that tillage operation has little effect on Zn availability in rice. On the other hand, crop rotation of rice with other crops such as wheat and maize has been reported to be beneficial in improving the availability of Zn in the soil such as the case of rice-maize rotation, which increases Zn fertilizer-use efficiency compared with continuously flooded rice (126). Manure application has also been reported to improve Zn availability in the soil by improving micronutrient availability through soil chemical, physical, and biological changes (127). This contributes to the accumulation of Zn under calcareous soils by reducing soil pH and improving mobilization of Zn. Furthermore, Azolla application in submerged soil conditions in rice also prevents ZnD by reducing bicarbonates in the soil (128). The availability of the Zn in the soil is highly affected by water regimes and aerobic conditions (129). Alternate wetting and drying (AWD) (130) has been proven to effectively facilitate absorption of Zn by increasing the redox potential of the soil. High concentrations of P in the soil create an antagonistic effect on the availability of Zn; hence, the application of large quantities of P fertilizers in the soil can induce deficiency in Zn and increase Zn requirements in plants (75).

Micronutrient Soil Status of the Major Rice Areas in the Philippines

Limited information is available on the micronutrient status of Philippine soils since most of the soil fertility surveys conducted in the country focus only on the major macronutrients such as nitrogen (N), P, and K. The Bureau of Soils and Water Management (BSWM) of the Department of Agriculture (DA), the agency mandated to conduct soil surveys, mapping, and classification in the Philippines, is currently planning to conduct a soil fertility survey focusing on micronutrient levels of the country's soils (131). Previous accounts of the Zn-deficient soils in the country have been reported by Yoshida et al. (132) and these were estimated to be around 500,000 ha of irrigated rice. Katyal and Vlek (133) also reported ZnD in rice cultivated on Hydrosols (Gleysols) that caused severe yield reduction in plots not treated with Zn fertilizer. ZnD was observed in calcareous soils in the Philippines in which half of the 13 field trials conducted in the country showed deficiency in Zn (134). Alloway (135) also included the Philippines on the list of countries with widespread Zn deficiency in their soils. On the other hand, no previous survey of Fe-deficient areas in the country has been conducted.

A recent soil fertility survey was conducted by Magahud et al. (24) on paddy soils of major rice production areas in the Philippines in which they determined the concentrations of various nutrient elements in the soil and in the plants that included macronutrients such as K, Ca, and S and micronutrients such as Fe and Mn. The study reported toxic levels of the micronutrients Fe and Mn in most of the rice areas in Central Luzon, which was attributed to periodic soil submergence and increased soil acidity due to continuous cropping. The high concentration of Fe in San Leonardo, Nueva Ecija, was possibly due to the occurrence of Fe in clay-sized soil fractions in the area while high Fe and Mn in Sta. Cruz, Zambales, is caused by the input of mine wastes. Toxic concentrations of Fe and Mn (1,000 ppm and 300 ppm, respectively) were eventually observed in rice plants (24). It appears that Fe is highly available in rice; however, the bioavailability of this nutrient in the rice grain is poor (136).

In another study conducted by Magahud et al. (25), which assessed the levels of heavy metals, including Zn, in the country's major rice production areas found that Zn concentrations are generally low in rice production areas. **Figure 2** presents the results of the studies on the concentrations of micronutrients, particularly Fe and Zn. The prevalence of ZnD in the rice-growing provinces was previously reported by Descalsota et al. (137) using soil samples collected from major rice areas in the country and identified the deficient elements in various provinces. The results indicated that all of the sites sampled were deficient in Zn. Meanwhile, sufficient concentrations of the micronutrients Ni and Mo were found at all of the sites sampled whereas few sites exhibited deficiency in Cu, particularly in Muñoz, Llanera, and Zaragosa in Nueva Ecija; La Paz in Tarlac; Sta. Cruz in Zambales; Casiguran in Sorsogon; and Zara in Iloilo.

CO-LOCATION OF ZN DEFICIENCIES IN PHILIPPINE SOILS AND POPULATION

Micronutrient deficiencies in the soil will produce micronutrient-deficient crops that provide insufficient amounts of micronutrients to humans, consequently contributing to micronutrient deficiencies in the population. There are strong indications of a causal relationship between micronutrient content in the soil and human micronutrient status (138). In

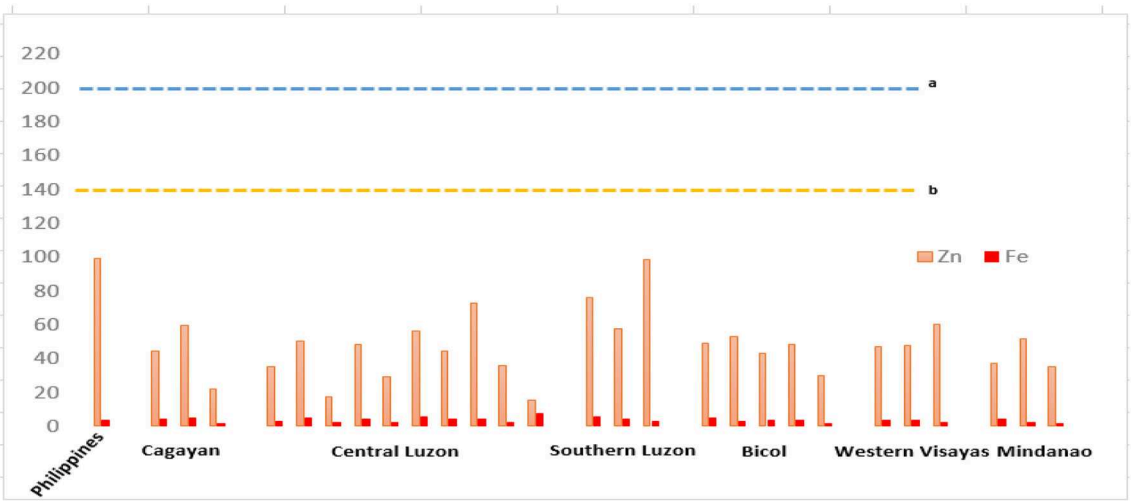

FIGURE 2 | Micronutrient concentrations of major rice paddies in the Philippines (24, 25) in various regions in the country, with each bar representing different sample sites (provinces), showing deficiency in Zn based on (a) Canadian soil guidelines for agricultural soils and (b) Netherlands target values for soils. The concentration of nutrients is measured in ppm. Provinces in each region include Cagayan (Cagayan and Isabela), Central Luzon (Nueva Ecija, Tarlac, Bulacan, Pangasinan, and Zambales), Southern Luzon (Laguna), Bicol Region (Camarines Sur, Albay, and Sorsogon), Western Visayas (Iloilo and Negros Occidental), and Mindanao (Maguindanao, North Cotabato, and Sultan Kudarat).

order for crops to provide the necessary micronutrients for humans, they must obtain sufficient amounts of micronutrient initially from the soil. In rice, the environment, particularly the soil, has a significant impact on Fe concentration in rice grains (11, 134). It has been estimated that, among the agricultural soils in the world, 49% are deficient in Zn, 31% in B, 15% in Mo, 14% in Cu, 10% in Mn, and 3% in Fe. It appears that these micronutrient deficiencies are reflected in the human population. Among the micronutrient deficiencies, Bo receives less priority while Zn has the highest priority in human nutrition (136).

In the Philippines, a majority of the soils have multiple macronutrient and micronutrient deficiencies (24, 137). Zn is predominantly deficient based on the critical concentration (100 ppm for Zn) in the paddy soils where the survey was conducted (24). Zn-deficient paddy soils were identified for rice plants based on the results of the Minus-One Element Technique (MOET) followed by Descalsota et al. (137). A significant positive correlation and linear regression between soil Zn and plant Zn were observed by Magahud et al. (25), indicating that higher soil concentrations of Zn lead to higher Zn concentration in rice plants. The concentration of Zn in rice, particularly in the grains, is highly affected by the amount of bioavailable Zn in the soil (117, 130). The uptake of Zn in rice grains is suggested to be affected by the supply of Zn during the grain Zn loading period (139). The supply of Zn to the grains is via root uptake in Zn-sufficient conditions while grain Zn accumulation is via root uptake and remobilization in Zn-deficient conditions (111, 112, 140). Zn uptake and transport are the major bottlenecks for Zn biofortification in rice (141).

As stated earlier in this review, ZnD was observed across age groups and was widespread in all population groups, with overall prevalence of more than 20% (63). The widespread ZnD in paddy soils somehow corresponds to the incidence of ZnD observed in the populations that grow crops in the same soils. **Figure 3** shows the overlapping locations of Zn-deficient soils and Zn-deficient populations in the country. Most of the areas that exhibited high occurrence of ZnD in the soil have ZnD in the population as well. The Zn-deficient soils where soil sampling was conducted represent the major rice production areas in the country and were also located in rural areas where the populations have a high prevalence of micronutrient deficiencies. Moreover, it can be observed that ZnD in both soils and populations is widespread throughout the Philippines. This is similar to the case in China, where micronutrient-deficient soils are located in areas of intensive agricultural production with high population density (116). Some areas exhibited both Zn and Cu deficiencies in the soils. Of all the micronutrients, Zn is directly linked to the food chain since deficiency in both humans and food crops is extensive (136). This observation needs further investigation to establish the relationship among micronutrient contents of soil, plants, and the human population in the country. A comprehensive evaluation of soil micronutrient contents and the development of a soil micronutrient map would be useful for this undertaking.

GOVERNMENT PROGRAMS THAT ADDRESS MICRONUTRIENT DEFICIENCIES

The Philippines has a Philippine Plan of Action for Nutrition (PPAN) as an integral component of the Philippine Development Plan for 2017–2022. It aims to expand and increase access to economic opportunities, especially for food for poor

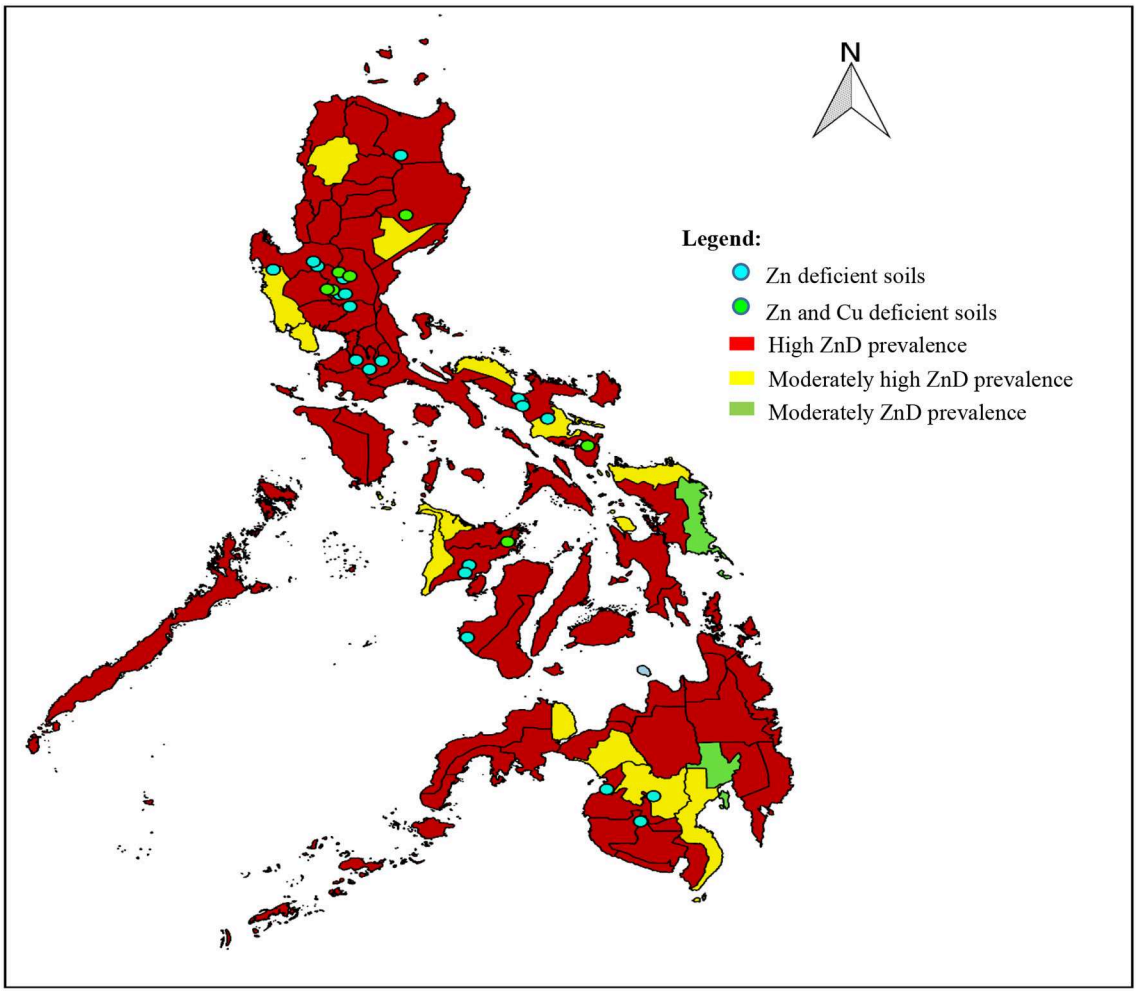

FIGURE 3 | Locations of Zn-deficient soils and ZnD occurrence across age group populations in the Philippines. Sources: (24, 57, 137).

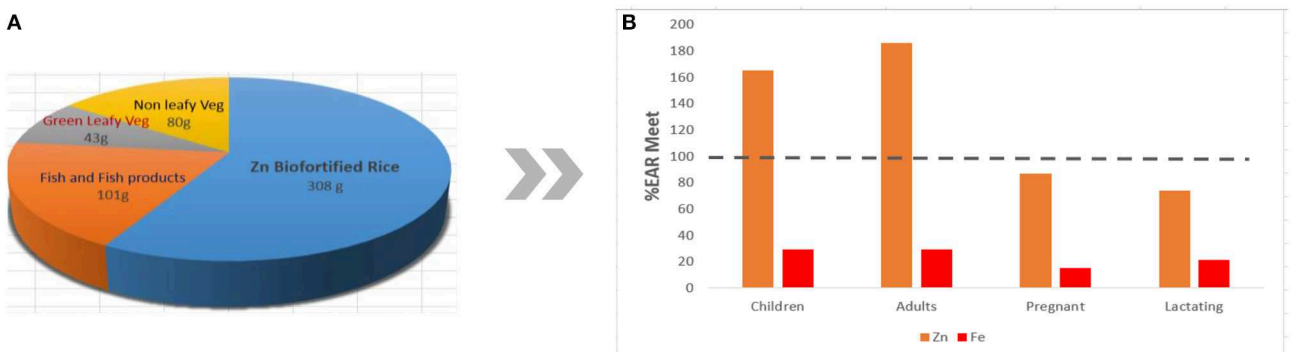

FIGURE 4 | Hypothetical approximate %EAR met for Zn and Fe of various population groups **(B)** if rice is replaced with Zn-biofortified rice in the typical Filipino diet **(A)**, assuming that the target content of 28 ppm is achieved while bioavailability in the body is not considered in the computation.

households, and reduce inequality in human development outcomes, as described in "Ambisyon 2040," the country's 25-year development plan. The target outcomes of the PPAN are to reduce wasting from 7 to <5%, reduce stunting from 33.4 to 28%, reduce micronutrient deficiencies, and have no further increase in overweight. Nutrition-specific programs, nutrition-sensitive programs, nutrition-supportive programs, and nutrition-enabling programs that involve the active participation of

local government units are planned. Specific programs to combat micronutrient malnutrition include micronutrient supplementation and mandatory food fortification, that is, flour fortification with Fe and vitamin A, cooking oil fortification with vitamin A, sugar fortification with vitamin A, salt iodization, and rice fortification with Fe (142).

Although numerous intervention programs are currently implemented to alleviate the problems in micronutrient deficiencies, awareness of these programs poses a major challenge. A recent survey conducted by FNRI on various age groups (13–59) across the country revealed that, out of 80,064 respondents, only 8.2% were aware of micronutrients, their deficiencies, and clinical outcomes. Nutritional campaigns and public health awareness targeting highly at-risk groups need to be strengthened to effectively achieve the objectives of the government programs.

Tables 2, **3** list the government policies and intervention programs aimed at reducing micronutrient deficiencies, general malnutrition, and other health-related problems in the country.

Government Policies

The Philippines is regarded as a "strong" nation in nutrition governance (152) and it has signed on to the Scaling Up Nutrition (SUN) Movement (146) to show its commitment to alleviating the problem of malnutrition. The Philippine government has crafted numerous policies and laws to address malnutrition and micronutrient deficiencies in the country (**Table 3**).These policies aim to reduce micronutrient deficiencies in the population by prioritizing the high-risk groups (infants, children, and pregnant and lactating women) through the introduction of various strategic intervention programs such as fortification, supplementation, and diversification. The government's initiative to eradicate micronutrient deficiencies, particularly Fe deficiency by enhancing Fe content in rice (Fortification Law), has paved the way for researchers to introduce alternative strategies to significantly increase the micronutrient content in rice through biofortification.

International Support

International agencies such as United Nations International Children's Emergency Fund (UNICEF), World Health Organization (WHO) (153), Food and Agriculture Organization (FAO) are staunch supporters in combating malnutrition in the country. In particular, UNICEF has actively advocated and promoted nutritional fortification of food to combat micronutrient malnutrition, iodization of food grade salt, pushed breastfeeding in the first 6 months of life, and paved the way for vaccination, Vitamin A supplementation and deworming of children (154). On the hand, FAO complemented government's efforts through technical and development interventions in the field of agriculture, fisheries, forestry and rural development (155). WHO aims to help the country reach every citizen with health coverage based on international standards. **Supplementary Table 1** presents some of the programs, projects, and collaborative efforts with international agencies to support Philippine government in tackling malnutrition.

HEALTHIER RICE VARIETIES: BIOFORTIFICATION EFFORTS IN THE PHILIPPINES

Developing rice varieties with high micronutrient concentration within the grains is a cost-effective and sustainable strategy to address micronutrient needs among populations unable to access a healthy and diverse diet. Food fortification, supplementation, and diet diversification have been suggested strategies to combat micronutrient malnutrition (156) but they have been found to have varying degrees of success in addressing micronutrient deficiencies in developing countries (12, 13). Biofortification is a complementary solution that offers numerous advantages compared with other fortification strategies such as accessibility for malnourished individuals in rural areas. It is cost-effective, low-cost, and sustainable (157). A biofortified rice variety can be grown indefinitely and the micronutrient-dense grains can be produced in a sustainable way without any additional input to reach malnourished populations. Biofortification has no marginal cost compared with fortification and supplementation that require continuous financial outlays. Further, biofortified rice varieties can be evaluated and adapted to other locations, which multiplies the initial investment (158). The Copenhagen Consensus ranked biofortification and other interventions in reducing micronutrient deficiencies as the highest value-for-money investments for economic development (159).

The large genetic variability in rice germplasm for Zn micronutrient concentration makes it a suitable crop for biofortification through conventional breeding. Similarly, rice germplasm has significant genetic variation for Fe content (11). Developing rice varieties with high Fe and Zn content through biofortification also targets improving agronomic traits and increasing yield. The *HarvestPlus* program objectives have set the guidelines for breeding toward high Fe and Zn content. *HarvestPlus* is a CGIAR initiative that spearheads biofortification efforts through an umbrella of international agricultural research centers in developing new varieties of vitamin- and mineral-rich staple foods (28). Biofortification breeding should maintain or improve crop productivity, increase micronutrient content to positively impact human nutritional status in a measurable way, ensure the stability of improved traits across environments and climatic conditions, test the bioavailability of micronutrients in biofortified lines in humans to ensure benefits to people preparing and consuming them in traditional ways within normal conditions, and test consumer acceptance to ensure maximum impact on nutritional health (160).

HarvestPlus has screened more than 7,500 rice lines and has identified promising donor lines in breeding for high-Zn rice varieties in other Asian countries, such as Bangladesh and India (161). The Bangladesh Rice Research Institute (BRRI) has released five conventionally bred high-Zn rice varieties: BRRI dhan62, BRRI dhan72, BRRI dhan64, BRRI dhan74, and BRRI dhan84 (162). Meanwhile, the ICAR-Indian Institute of Rice Research has released DRR dhan 45, a high-Zn rice variety with 22.6 ppm Zn content in grains in polished form (163). China has released a rice variety with high Fe content, Zhongguanxian 1, with 7 ppm Fe content in polished form, released in 2010 (164).

TABLE 2 | Government programs that address malnutrition in the Philippine population.

Programs	Target population	Nutrients provided	Sources
Pantawid Pamilyang Pilipino Program (4Ps)	Children and pregnant women	General nutrition	(143)
PhilHealth	All	General nutrition	(143)
Expanded Program on Immunization (EPI)	Infants	Vaccine	(143)
Food fortification	Infants, children, and adults	Fe, Zn, iodine, vitamin A, lysine	(9, 41, 60, 144)
Vitamin A supplementation	Children	Vitamin A	(143)
Fe supplementation	Young children and pregnant women	Fe	(61, 66, 143)
Multi-Nutrient Rice Kernel (MNERK)	All	Fe, Zn, vitamin A, B_9, B_{12}, B_1, and iodine	(66)
Micronutrient Powder (MNP)	Children	Fe, Zn, and vitamin A	(143)
Gulayan sa Paaralan	Children	General nutrition	(143)
Rice biofortification	All	Zn and Fe	(16, 79, 145)

TABLE 3 | Government policies crafted to combat malnutrition in the Philippines.

Policies	Objectives	Target population	Program/service provided	Sources
Philippine Plan of Action (PPAN) for 2017–2022	Address nutritional problems such as stunting, wasting, nutrient deficiencies, hunger, food security, and maternal nutrition	General population, especially children and pregnant women	General nutrition	(146)
DOH Memorandum No. 2011-0303: Micronutrient powder supplementation	Reduce under-five mortality rate by 66% and maternal mortality rate by 75%	Children and pregnant women	Vitamin and mineral supplements	(147)
Revised policy on child growth standard, 2010	Reduce undernutrition among children and compare growth of children against international standard	Infants and children	General nutrition	(148)
Philippine code on marketing and breast milk substitute (Revised 2010)	Support provision of safe and adequate nutrition of infants through protection and promotion of breastfeeding	Infants	Proper use of breast milk substitutes and supplements	(149)
Revised policy on micronutrient supplementation	Address micronutrient deficiencies	Children, pregnant women, and lactating women	Micronutrient supplements	(150)
Strategy for maternal, newborn, and child health and nutrition	Provide strategy to reduce maternal and neonatal mortality	Pregnant women and infants	Integrated maternal, newborn, and child health	(148)
Philippine Fortification Law (R.A. 8976)	Compensate inadequacy of micronutrients in the Filipino diet	All	Philippine food fortification program	(151)
Early Childhood Development Act, 2000 (R.A. 8980)	Promotes the rights of children to survival, development, and special protection	Children	National system for childhood care and development	(150)

The International Rice Research Institute (IRRI), situated in Los Baños, Laguna, Philippines, has been evaluating rice germplasm for Fe content variability since 1992 (11) and started working on Zn in 1995 (165). This research effort was influenced by the Philippine government's initiative to eliminate Fe malnutrition in the country by artificially enriching rice with Fe (11). Initially, improved lines with enhanced grain Fe and Zn coupled with high grain yield and desirable agronomic traits were developed. The most popular line was IR68144-3B-2-2-3, a line developed from crossing IR72, a high-yielding line from IRRI, and Zawa Bonday, a tall traditional variety from India, which was identified as a donor parent for high Fe content in grains (11). Two years after the establishment of the regional iron rice biofortification program (2003), NSIC Rc172 (MS13), tagged as the first high-Fe rice variety, was released. In 2016, the first zinc-rich variety, NSIC Rc460, was released by the National Seed Industry Council (NSIC) for production, 14 years following the release of MS13. This variety is a product of collaborative efforts between IRRI and the Philippine Rice Research Institute (PhilRice).

The Department of Agriculture (DA) through PhilRice has been working closely with IRRI in developing micronutrient-dense elite lines. Breeding materials from both research institutes are assembled and evaluated, and field trials are conducted at various PhilRice stations and commenced in the last wet season of 2014. Line development at PhilRice has generated 24 crosses involving various donor parents containing high grain Fe or Zn with high yield ability, resistance to various pests, and good grain quality. Currently, there are 54 entries in Advanced Yield Trials (AYTs), 22 entries in Observational Yield Trials, 22 entries

in Preliminary Yield Trials (PYTs), 43 entries in Pre-National Cooperative Testing (NCT), and 482 F_6 lines being evaluated at different branch stations of PhilRice. Entries are composed of lines coming from both IRRI and PhilRice (166). As of the moment, two promising lines bred by PhilRice and six promising lines from IRRI are being evaluated in NCT and will be released later as varieties upon passing evaluation. The breeding target for biofortified Zn rice follows the target set by *HarvestPlus*, which was designed to meet the dietary needs of the population, and should provide 60–80% of the Estimated Average Requirement (EAR). Currently, the baseline Zn content of rice is at 16 ppm, with a final target content of 28 ppm (158). **Figure 4** shows the approximate EAR met for Zn and Fe if rice is biofortified with Zn.

Breeding for rice varieties with enhanced nutritional content and desirable agronomic traits such as high grain yield are important breeding targets in rice (80). Development of high-yielding varieties with high level of Zn is doable (167). High Zn rice varieties with high-yield potential have been released for commercial cultivation in the Philippines and Bangladesh (162). Incorporating high Zn in hybrid rice can be a promising option. The current micronutrient content of hybrid varieties is relatively comparable with non-biofortified inbred varieties (168), increasing Zn content of hybrids can be an excellent breeding target. Several studies have revealed positive evidence on the possibility of developing hybrid rice varieties with high Zn content (23, 169). At IRRI efforts are being made to mainstream Zn breeding to incorporate grain Zn content as a key component of all the future rice varietal releases and also to develop hybrid varieties with high Zn content.

The limited number of micronutrient-enriched varieties available for farmers and consumers will be the bottleneck in the promotion of the varieties. The current released variety (NSIC Rc460) is still in the stage of seed multiplication and there are no recorded farmer users and areas planted to the variety. Awareness of the availability of this kind of variety and its purpose is still a major concern. The absence of an X-ray fluorescence (XRF) analyzer for evaluating the micronutrient content of promising lines delays the development of genotypes of PhilRice with high Zn and Fe content. The acquisition of an XRF analyzer is crucial for developing promising lines with high micronutrient content and an important factor in accelerating breeding programs for micronutrient-dense varieties.

FUTURE PROSPECTS

Filipinos depend on rice as their main source of nutrition. The high dependence on rice and the lack of diversity in the diet have led to nutritional and health problems, particularly micronutrient deficiencies. Fe and Zn deficiencies are widespread and affect millions of Filipinos across ages and social strata with health and economic consequences. These consequences necessitate the design and evaluation of effective programs and policies for reducing micronutrient deficiencies.

Biofortification is an attractive strategy to enrich rice with micronutrients. It is an effective and sustainable approach that can target large portions of vulnerable populations. However, breeding efforts in the Philippines are slow. Currently, only two biofortified rice varieties, Zn rice, NSIC Rc460, and Fe rice, NSIC Rc172 (MS13), are commercially available for production and multiplication. However, a few more high-Zn rice varieties are in the final stage of testing in national trials and they will be released in the next 1 or 2 years. The availability of biofortified varieties that are high-yielding and suit farmers' preferences and consumers' taste will be the major prerequisite in effectively promoting biofortified rice to target populations. The limited options in the number of local varieties reduce the possibility of successful farmer adoption and delivery to target consumers. The lack of equipment for effective evaluation of grain micronutrient content of promising lines is the bottleneck in the development of varieties in local rice breeding programs.

Recently, the Department of Agriculture has been promoting the use of hybrid rice varieties under the High Yielding Technology Adoption (HYTA) program primarily due to their promising yield potential. Development of hybrid rice varieties with high micronutrient content can be a plausible option to address concerns on malnutrition and low yield. The program can be a perfect avenue to promote both technologies.

Once biofortified varieties are developed, the next step would be their promotion to farmers for production. Varieties are presumed to be tailored to farmers' preferences and, therefore, increasing awareness of their availability and their health benefits would be the next agenda. Nationwide farm demonstrations coupled with technology introduction during field days and gatherings would be effective means to promote varieties to farmers and seed growers as well as to increase use by consumers. Recently, the country has been actively advocating the consumption of brown rice through Presidential Proclamation No. 494 since this rice is more nutritious than polished rice and can potentially help to alleviate micronutrient deficiencies. The nationwide initiative of PhilRice and the DA in promoting the consumption of healthy rice such as brown rice can also be tapped in introducing other nutritious types of rice such as biofortified rice varieties. Furthermore, local governments can be tapped to issue ordinances and resolutions to support the advocacy of introducing and consuming healthy rice to rice stakeholders (170). The consumption of diversified food, especially other staples with higher micronutrients and vegetables, will be the ideal strategy, but an immediate change in diet and its availability and accessibility will be the limitation.

After biofortified rice has been introduced to the market and eventually to consumers, an impact assessment will evaluate the effect of consuming it on the micronutrient status of the target population. This could be similar to the study conducted by Haas et al. (171), which tested the efficacy of the additional Fe from the biofortified rice in the intake of Filipino women who are at risk of Fe deficiency. The study showed that 9 months of consumption of Fe-biofortified rice improved the Fe stores in the blood of non-anemic Filipino women (171). However, examining the bioavailability of micronutrients from plants to humans is difficult because of the governing interactions of various factors, particularly in determining the bioavailability of a particular micronutrient to an individual with a diversified diet (172, 173). The micronutrient status of

test subjects affects the regulation of these nutrients within the body. It is suggested that, to maximize the response of subjects to a test meal, the subjects should be marginally depleted with micronutrients (172, 173). Ideally, human trials should be conducted in a controlled environment to generate reliable and conclusive data.

The locations of micronutrient-deficient populations, particularly with ZnD, coincide with the locations of Zn-deficient soils in the Philippines. It is crucial that breeding for superior genotypes be coupled with sound agronomic practices (such as Zn fertilizer application) that promote micronutrient availability in the soil to maximize micronutrient content in the grains. Increasing grain micronutrient concentration is not solely dependent on the genotype's capacity to accumulate micronutrients in its grains but is also dependent on soil micronutrient concentration. Philippine soils have been found to be Zn-deficient, especially in paddy soils. Deficiency in micronutrients in the soil leads to deficiency in plants, which eventually leads to lower amounts of micronutrients consumed by humans. It is therefore essential to consider the soil-plant-human system in optimizing micronutrient biofortification in rice that combines plant breeding and nutritional management of the soil-plant system with the aim of improving nutritional status in humans. This approach is more sustainable because it involves regulation of the micronutrient flow from soils to plants and eventually to humans. Further, interdisciplinary collaboration of plant breeders, plant nutritionists, soil scientists, and human nutritionists could result in the development of micronutrient-efficient and micronutrient-dense rice genotypes (116).

Breeding for biofortified rice rich in Zn and Fe should be a top priority in the Philippines and other rice-growing countries because it is a win-win approach since it is beneficial to both rice plants and human nutrition. Other approaches for biofortification, such as genetic engineering, should be explored, especially in the case of Fe since natural variability in this micronutrient is limited, particularly in polished rice. Lastly, the impact of biofortified rice can be realized through measurable improvements in the micronutrient status of target populations through community-based human feeding trials with a given duration.

AUTHOR CONTRIBUTIONS

AP, GD-E, and BS designed the review. AP wrote the whole manuscript. AA and MC wrote some parts of the manuscript. MI-A, EA, PC, TB, AL, JH, CA, RR, and BS edited and contributed significant inputs.

FUNDING

AP is funded by the Department of Science and Technology and Philippine Rice Research Institute. We would like to thank HarvestPlus for funding the development of high Zinc rice varieties.

ACKNOWLEDGMENTS

The authors are grateful to the Food and Nutrition Research Institute-Department of Science and Technology (FNRI-DOST) for providing important data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2019.00081/full#supplementary-material

REFERENCES

1. Food and Agriculture Organization. *FAO Statistical Yearbook*. Bangkok: FAO. (2014).
2. Department of Science and Technology-Food and Nutrition Research Institute (DOST-FNRI). *Philippine Nutrition Facts and Figures 2015: Dietary Survey*. FNRI Bldg., DOST Compound, Gen. Santos Ave., Bicutan, Taguig City, Metro Manila, Philippines. (2016).
3. Kennedy G, Burlingame B, Nguyen, VN. Nutritional Contribution of Rice and Impact of Biotechnology and Bio-Diversity in Rice-Consuming Countries. In: *20th International Rice Commission,* Bangkok, Thailand (2002). p. 23–6.
4. Sharma A, Patni B, Shandhdhar SC. Zinc: an indispensable micronutrient. *Physiol Mol Biol Plants*. (2013) 19:11–20. doi: 10.1007/s12298-012-0139-1
5. International Rice Research Institute. *Healthier Rice Varieties. High-Iron and High-Zinc Rice*. Los Baños: IRRI (2017).
6. Lu K, Li L, Zheng X, Zhang Z, Mou Y, Yu Z. Quantitative trait loci controlling Cu, Ca, Zn, Mn, and Fe content in rice grains. *J Genet*. (2008) 87:305–10. doi: 10.1007/s12041-008-0049-8
7. King J, Cousins R. Zinc. In: Shils M. editor. *Modern Nutrition Health and Disease*. 10th ed. Philadelphia, PA: Lippincott Williams and Wilkins (2006). p. 272–85.
8. Department of Science and Technology-Food and Nutrition Research Institute (DOST-FNRI). *Philippine Dietary Reference Intake*. FNRI Bldg., DOST Compound, Gen. Santos Avenue, Bicutan, Taguig: Philippines. (2015).
9. Detzel P, Wieser S. Food fortification for addressing iron deficiency in Filipino children: benefits and cost-effectiveness. *Ann Nutr Metab*. (2015) 66:35–42. doi: 10.1159/000375144
10. Shaw JG, Friedman JF. Iron deficiency anemia: focus on infectious diseases in lesser developed countries. *Anemia*. (2011) 2011:260380. doi: 10.1155/2011/260380
11. Gregorio GB, Senadhira D, Htut H, Graham RD. Breeding for trace mineral density in rice. *Food Nutr Bull*. (2000) 21:382–6. doi: 10.1177/156482650002100407
12. Lyons G, Stangouilis J, Graham R. High-selenium wheat: biofortification for better health. *Nutr Res Rev*. (2008) 16:45–60. doi: 10.1079/NRR200255
13. Misra BK, SharmaRK, Nagarajan S. Plant breeding: a component of public health strategy. *Curr Sci*. (2004) 84:1210–5. Retrieved from http://www.jstor.org/stable/24109928
14. White PJ Broadley MR. Physiological limits to zinc biofortification of edible crops. *Front Plant Sci*. (2011) 2:80. doi: 10.3389/fpls.2011.00080
15. Gomez-Gallera S, Rojas E, Sudhakar D, Zhu CF, Pelacho AM, Capell T, et al. Critical evaluation of strategies for mineral fortification of staple food crops. *Transgenic Res*. (2010) 19:165–80. doi: 10.1007/s11248-009-9311-y
16. Cakmak I. Enrichment of cereal grains with zinc: agronomic or genetic biofortification? *Plant Soil*. (2008) 302:1–17. doi: 10.1007/s11104-007-9466-3
17. Philippine Statistics Authority. *Selected Statistics in Agriculture*. Quezon City (2018).
18. Philippine Statistics Authority. *Selected Statistics in Agriculture*. Quezon City (2016).

19. Global Rice Science Partnership (GRiSP). *Rice Almanac, 4th Edn.* Los Banos: International Rice Research Institute. (2013).
20. Philippine Statistics Authority. *Selected Statistics in Agriculture*. Quezon City (2017).
21. Philippine Statistics Authority. *Food Consumption and Nutrition*. Quezon City: Agricultural Indicator System. (2016).
22. Francisco SR, Mataia AB, Eusebio AM, Sanguyo EB, Constantino MAS, Causon EP, et al. *Increasing Per Capita Rice Consumption in the Philippines: Myths and Realities, a Paper Presented in the Monthly Seminar Series of the Philippine Rice Research Institute Held at PhilRice*. Munoz, Nueva Ecija, Philippines.
23. Babu VR. Importance and advantages of rice biofortification with iron and zinc. *J SAT Agric Res.* (2013) 11:1–6.
24. Magahud JC, Badayos RB, Sanchez PB, Sta. Cruz PC. Level and sources of potassium, calcium, sulfur, iron, and manganese in major paddy soils of the Philippines. *Int J Phil Sci Technol.* (2015) 8:1–8. doi: 10.18191/2015-08-2-014
25. Magahud JC, Badayos RB, Sanchez PB, Sta. Cruz PC. Levels and potential sources of heavy metals in major irrigated rice areas in the Philippines. *IAMURE Mult Res.* (2015) 15:28–59. doi: 10.7718/ijec.v15i1.1000
26. Wu CY, Lu LL, Yang XE, Geng Y, Wei YY, Hao HL, et al. Uptake translocation and remobilization of zinc absorbed at different stages by rice genotypes of different Zn densities. *J Agric Food Chem.* (2010) 58:6767–73. doi: 10.1021/jf100017e
27. Branca F, Ferrari M. Impact of micronutrient deficiencies on growth: the stunting syndrome. *Ann Nutr Metab.* (2002) 46:8–17. doi: 10.1159/000066397
28. Ramakrishnan U, Manjrekar R, Rivera J, Gonzales-Cossio T, Martorell R. Micronutrients and pregnancy outcome: a review of literature. *Nutr Res.* (1999) 19:103–59. doi: 10.1016/S0271-5317(98)00178-X
29. Tuman RW, Doisy RJ. Chapter 10: The role of trace elements in human nutrition and metabolism. In: Ioannis S, Scarpa H, Chilton K, editors. *Sourcebook on Food and Nutrition*. Chicago, IL: Marquis Academic Media. (1978). p. 156–77.
30. Zhang MW, Gou BJ, Peng ZM. Genetic effects on Fe, Zn, Mn, and P, contents in Indica black pericarp rice and their genetic correlations with grain characteristics. *Euphytica.* (2004) 135:315–23. doi: 10.1023/B:EUPH.0000013340.98344.60
31. Adrollo-Sanchez M, Hininger-Favier I, Meunier N, Toti E, Zaccaria M, Brandolini-Bunlon M, et al. Zinc intake and status in middle-aged and older European subjects. The Zenith study. *Eur J Clin Nutr.* (2005) 59:S37–41. doi: 10.1038/sj.ejcn.1602296
32. IOM. Institute of Medicine. Iron. In: *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*. Washington, DC: National Academy Press. (2001). p. 290–393.
33. Lozoff B, Jimenez E, Wolf A. Long-term developmental outcome of infants with iron deficiency. *N Engl J Med.* (1991) 325:687–94. doi: 10.1056/NEJM199109053251004
34. Youdim MBH. *Brain Iron: Neurochemical and Behavioral Aspect*. New York, NY: Taylor and Francis (1988).
35. Beard JL, Connor JR, Jones BC. Iron in the brain. *Nutr Res.* (1993) 1:157–70. doi: 10.1111/j.1753-4887.1993.tb03096.x
36. Food and Agriculture Organization/World Health Organization. *Human Vitamin and Mineral Requirements*. Rome: Food and Nutrition Division, FAO (2001).
37. Dunne JR, Malone D, Tracy JK, Gannon C, Napolitano LM. Perioperative anemia: an independent risk factor for infection, mortality, and resource utilization surgery. *J Surg Res.* (2002) 102:237–44. doi: 10.1006/jsre.2001.6330
38. Jonker FA, Van Boele HM. Anaemia, iron-deficiency and susceptibility to infections. *J Infect.* (2014) 69(Suppl 1):S23–7. doi: 10.1016/j.jinf.2014.08.007
39. Lynch SR. Fe/Physiology. In: Benjamin Caballero, editor. *Encyclopedia of Food Sciences and Nutrition*. 2nd edn. Oxford: Elsevier Science Ltd. (2003) p. 1–6000.
40. Benton D. The influence of children's diet on their cognition and behavior. *Eur J Nutr.* (2008) 47:(Suppl 3):25–37. doi: 10.1007/s00394-008-3003-x
41. Amin SB, OrlandoM, Wang H. Latent iron deficiency in utero is associated with abnormal auditory neural myelination in ≥35 weeks gestational age infants. *J Pediatr.* (2013) 163:1267–71. doi: 10.1016/j.jpeds.2013.06.020
42. McArdle HJ, Gambling L, Kennedy C. Iron deficiency during pregnancy: the consequences for placental function and fetal outcome. *Proc Nutr Soc.* (2014) 73:9–15. doi: 10.1017/S0029665113003637
43. Ruz M. ZINC/Properties and determination. In: Benjamin C. editor. *Encyclopedia of Food Sciences and Nutrition.* 2nd Edn. Oxford: Elsevier Science Ltd. (2003).
44. Valle BL, David SA. Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry.* (1990) 29:5647–59. doi: 10.1021/bi00476a001
45. Coleman JE. Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins. *Annu Rev Biochem.* (1992) 61:896–946. doi: 10.1146/annurev.bi.61.070192.004341
46. Babusiak M. *The Biochemistry of Zinc and Iron. Proteomic Studies*. Thesis 1st Faculty of Medicine, Charles University, Prague (2007).
47. Bender A, Bender AE. *Benders' Dictionary of Nutrition and Food Technology*. 7th ed. Boca Raton, FL: CRC Press (1999).
48. Solomons NW Cousins RJ. Absorption and malabsorption of mineral nutrients. In: Solomons NW, Rosenberg IH, editors. *Zinc.* New York, NY: Alan R. Liss. (1984). p. 125–97.
49. Solomons NW. ZINC/Deficiency. In: Benjamin C. editor. *Encyclopedia of Food Sciences and Nutrition*. 2nd ed. Oxford: Elsevier Science Ltd. (2003). p. 1–6000.
50. Food and Nutrition Research Institute (FNRI). *Department of Science and Technology: Philippine Nutrition Facts and Figures*. Taguig: FNRI. (2010).
51. Weiser S, Plessow R, Eichler K, Olivia M, Capanzana MV, Agdeppa I, et al. Burden of micronutrient deficiencies by socio-economic strata in children aged 6 months to 5 years in the Philippines. *BMC Public Health.* (2013) 13:1167. doi: 10.1186/1471-2458-13-1167
52. Bender DA, Bender AE. *Benders' Dictionary of Nutrition and Food Technology*. 7th ed. Boca Raton, FL: CRC Press (1999).
53. Simmer K, Khanum S, Carlsson L, Thompson RPH. Nutritional rehabilitation in Bangladesh: the importance of zinc. *Am J Clin Nutr.* (1998) 47:1036–40. doi: 10.1093/ajcn/47.6.1036
54. Brown KH, Rivera JA, Bhutta Z, Gibson RS, King JC. International zinc nutrition consultative group technical document #1. Assessment of the risk of zinc deficiency in population and options for its control. *Food Nutr Bull.* (2004) 25:99–203.
55. Maret W, Sandstead HH. Zinc requirements and the risks and benefits of zinc supplementation. *J Trace Elem Med Biol.* (2006) 20:3–18. doi: 10.1016/j.jtemb.2006.01.006
56. King JC. Determinants of maternal zinc status during pregnancy. *Am J Clin Nutr.* (2000) 71:1334S−3. doi: 10.1093/ajcn/71.5.1334s
57. Food and Nutrition Research Institute-Department of Science and Technology. *Annual Report.* FNRI-DOST, Taguig City, Metro Manila, Philippines (2015).
58. Development Initiatives. *Global Report: Nourishing the SDGs*. Bristol: Development Initiatives (2017).
59. Petry N, Olofin I, Hurrell RF, Boy E, Wirth JP, Moursi M, et al. The proportion of anemia associated with iron deficiency in low, medium, and high human development index countries: a systematic analysis of national surveys. *Nutrients.* (2016) 8:693. doi: 10.3390/nu8110693
60. BaileyRL, West KP Jr, Black RE. The epidemiology of global micronutrient deficiencies. *Ann Nutr Metab.* (2015) 66:22–3. doi: 10.1159/000371618
61. Department of Health (DOH). *MNCHN Strategy Manual of Operations*. Manila, Philippines (2011).
62. Reddy BVS, Ramesh S, Longvah T. Prospects of breeding for micronutrients and ß-carotene-dense sorghums. *ISMN.* (2005) 46:10–4. Retrieved at http://oar.icrisat.org/1210/1/ISMN-46_10-14_2005.pdf
63. Marcos JM, Perlas LA, Trio PZ, Ulanday JRC, Cheong RL, Desnacido JA, et al. Zinc status of Filipinos by serum zinc level. *Philipp J Sci.* (2015) 144:139–48. Retrieved at https://www.researchgate.net/profile/Juanita_Marcos/publication/304951529_Zinc_status_of_Filipinos_by_serum_zinc_level/links/57d88cc308ae601b39afad57.pdf
64. Food and Agriculture Organization. *Nutrition country profiles: Philippines*. FAO, Singapore (2010).

65. Rohner F, Woodruff BA, Aaron GJ, Yakes EA, Lebanan MA, Raycon-Solon P, et al. Infant and young child feeding practices in urban Philippines and their associations with stunting, anemia, and deficiencies of iron and vitamin A. *Food Nutr Bull.* (2013) 34:S17–34. doi: 10.1177/1564826513034 2S104
66. Food and Nutrition Research Institute. *National Nutrition Survey*. Manila: FNRI. (2013).
67. Kreissl A. Malnutrition in the Philippines: perhaps a double burden? *JEM.* (2009) 11:24. Retrieved at https://www.kup.at/journals/summary/8113.html
68. Herrin AN. *Putting Prevention of Child Stunting in the Forefront of the Nutrition Agenda: A Nutrition Sector Review*. Philippine Institute for Development Studies. Quezon City, Philippines. (2016).
69. De Jon N, Romano AB, Gibson RS. Zinc and iron status during pregnancy of Filipino women. *Asia Pac J Clin Nutr.* (2002) 11:186–93. doi: 10.1046/j.1440-6047.2002.00289.x
70. Mendez MA, AdairLS. Severity and timing of stunting in the first two years of life affect performance on cognitive tests in late childhood. *J Nutr.* (1999) 129:1555–62. doi: 10.1093/jn/129.8.1555
71. Tomkins A. Malnutrition, morbidity, and mortality in children and their mothers. *Proc Nutr Soc.* (2000) 59:135–46. doi: 10.1017/S0029665100000161
72. Black MM. Zinc deficiency and child development. *Am J Clin Nutr.* (1998) 68(suppl.):464S—9S. doi: 10.1093/ajcn/68.2.464S
73. Brown KH, Peerson JM, Baker SK, HessSY. Preventive zinc supplementation among infants, pre-schoolers, and older prepubertal children. *Food Nutr Bull.* (2009) 30:S12–S40. doi: 10.1177/15648265090301S103
74. Hotz C, Brown KH. International Zinc Nutrition Consultative Group (IZiNCG) Technical Document #1. Assessment of risk of zinc deficiency in population and options for its control. *Food Nutr. Bull.* (2004) 25:S94–200. doi: 10.4067/S0717-75182010000200014
75. Lohry R. Micronutrients: functions, sources, and application methods. In: *Indiana CCA Conference Proceedings*. Indianapolis, IN. (2007).
76. Welch RM, GrahamRD. Breeding for micronutrients in staple food crops from a human nutrition perspective. *J Exp Bot.* (2004) 55:353–64. doi: 10.1093/jxb/erh064
77. Gupta UC, Gupta SC. Quality of animal and human life as affected by selenium management of soils and crops. *Commun Soil Sci Plant Anal.* (2002) 33:2537–55. doi: 10.1081/CSS-120014464
78. Frossard E, Bucher M, Machler F, Mozafar A, Hurrell R. Potential for increasing the content and bioavailability of Fe, Mn, Zn, and Ca in plants for human nutrition. *J Sci Food Agric.* (2000) 80:861–79. doi: 10.1002/(SICI)1097-0010(20000515)80:7<861::AID-JSFA601>3.3.CO;2-G
79. Welch RM, Graham RD. Breeding crops for enhanced micronutrient content. *Plant Soils.* (2002) 245:205–14. doi: 10.1023/A:1020668100330
80. Bouis E. Micronutrient fortification of plants through plant breeding: can it improve nutrition in man at low cost? *Proc Nutr Soc.* (2003) 62:403–11. doi: 10.1079/PNS2003262
81. Rout GR, Sahoo S. Role of iron in plant growth and metabolism. *Rev Agric Sci.* (2015) 3:1–24. doi: 10.7831/ras.3.1
82. Briat JF, Gaymard CC. Iron utilization and metabolism in plants. *Curr Opin Plant Biol.* (2007) 10:276–82. doi: 10.1016/j.pbi.2007.04.003
83. Perur NG, Smith RL, Wiebe HH. Effect of iron chlorosis on protein fractions of corn leaf tissue. *Plant Physiol.* (1961) 36:736–9. doi: 10.1104/pp.36.6.736
84. Mende, K, Kirby E. *Principles of Plant Nutrition*. International Potash Institute (1987).
85. Bertini I, Rosato A. From genes to metalloproteins: a bioinformatics approach. *Eur J Inorg Chem.* (2007) 2007:5526–39. doi: 10.1002/ejic.200700186
86. Vigani G, Maffi D, Zocchi G. Iron availability affects the function of mitochondria in cucumber root. *N Phytol.* (2009) 182:127–36. doi: 10.1111/j.1469-8137.2008.02747.x
87. Samaranayake P, Peiris BD, Dssanayake S. Effect of excessive ferrous (Fe^{2++}) on growth and iron content in rice (*Oryza sativa*). *Int J Agric Biol.* (2012) 14:296–8. Retrieved from https://search.proquest.com/docview/994047549?accountid=173015
88. Zo, Y, Zhang F. Soil and crop management strategies to prevent iron deficiency in crops. *Plant Soil.* (2011) 339:83–95. doi: 10.1007/s11104-010-0566-0
89. Dobermann A Fairhurst T. *Rice: Nutrient Disorders and Nutrient Management*. Handbook Series, Potash and Phosphate Institute (PPI), Potash and Phosphate Institute of Canada (PPIC), and International Rice Research Institute, Philippines, (2000). p. 191.
90. Sahrawat KL. Iron toxicity in wetland rice and role of other nutrients. *J Plant Nutr.* (2004) 27:1471–504. doi: 10.1081/PLN-200025869
91. Kobayashi T, Nakayama Y, Itai RN, Nakanishi H, Yoshira T, Mori S. Identification of novel cis-acting elements, IDE1 and IDE2, of barley IDS2 gene promoter conferring iron deficiency-inducible, root specific expression in heterogeneous tobacco plants. *Plant J.* (2003) 36:780–93. doi: 10.1046/j.1365-313X.2003.01920.x
92. Rashid A, Ryan J. Micronutrient constraints to crop production in soils with Mediterranean-type characteristics: a review. *J Plant Nutr.* (2004) 27:959–75. doi: 10.1081/PLN-120037530
93. Martens DC, Westernmann DT. Fertilizer applications for correcting micronutrient deficiencies. In: Mortvedt JJ, Cox FR, Shuman LM, Welch RM, editors. *Micronutrients in Agriculture*. Madison, WI: Soil Science Society of America (1991). p. 549–92.
94. Pal S, Datta PS, Rattan RK, Singh AK. Diagnosis and amelioration of Fe deficiency under aerobic rice. *J Plant Nutr.* (2008) 5:919–40. doi: 10.1080/01904160802043262
95. Neue HU, Quijano C, Senadhira D, Sette T. Strategies of dealing with micronutrient disorders and salinity in lowland rice systems. *Field Crops Res.* (1998) 56:139–55. doi: 10.1016/S0378-4290(97)00125-1
96. Chen W, He ZL, Yang, X, Feng Y. Zinc efficiency is correlated with root morphology, ultrastructure, and anti-oxidative enzymes in rice. *J Plant Nutr.* (2009) 32:287–305. doi: 10.1080/01904160802608627
97. Broadley MR, White PJ, Hammond JP, Zelko L, Lux A. Zinc in plants. *N Phytol.* (2007) 173:677–702. doi: 10.1111/j.1469-8137.2007.01996.x
98. Brown PH, Cakmak I, Zhang Q. Form and function of zinc in plants. In: Bobson AD editor. *Zinc in Soil and Plants*. Dordrecht: Kluwer Academic Publishers (1993). p. 90–106. doi: 10.1007/978-94-011-0878-2_7
99. Marschner H. *Mineral Nutrition of Higher Plants*. London: Academic Press (1995). p. 889.
100. Sadeghzadeh B, Rengel Z. Zinc in soils and crop nutrition. In: Hawkesford, MJ, Barraclough P, editors. *The Molecular Basis of Nutrient Use Efficiency in Crops*. London: Wiley (2011). p. 335–76. doi: 10.1002/9780470960707.ch16
101. Skoog F. Relationship between zinc and auxin in the growth of higher plants. *Am J Bot.* (1940) 27:937–51. doi: 10.1002/j.1537-2197.1940.tb13958.x
102. Alloway BJ. *Zinc in Soils and Crop Nutrition*. International Zinc Association. (2003). Available online at: http://zinc-crops.org~p~114.
103. Rehman S, Aziz T, Farooq M, Wakeel A, Rengel Z. Zinc nutrition in rice production systems: a review. *Plant Soil.* (2012) 361:203. doi: 10.1007/s11104-012-1346-9
104. Fageria NK, Dos Santos AB, Cobucci T. Zinc nutrition in lowland rice. *Soil Sci Plant Anal.* (2011) 42:1719–27. doi: 10.1080/00103624.2011.584591
105. Naik SK, Das DK. Effect of split application of zinc on yield of rice (*Oryza sativa* L.) in an inceptisol. *Arch Agron Soil Sci.* (2007) 53:305–13. doi: 10.1080/03650340701311398
106. Renkou X, Zhao A, Li Q, Kong X, Ji G. Acidity regime of the Red Soils in a subtropical region of southern China under field conditions. *Geoderma.* (2003) 115:75–84. doi: 10.1016/S0016-7061(03)00077-6
107. Qadar A. Selecting rice genotypes tolerant to zinc deficiency and sodicity stress. I. Differences in zinc, iron, manganese, copper, phosphorus concentrations and phosphorus/zinc ratio in their leaves. *J Plant Nutr.* (2002) 25:457–73. doi: 10.1081/PLN-120003376
108. Forno DA, Yoshida S, Asher CJ. Zinc deficiency in rice I. Soil factors associated to deficiency. *Plant Soil.* (1975) 42:537–50. doi: 10.1007/BF00009941
109. Neue HU, LantinRS. Micronutrient toxicities and deficiencies in rice. In: Yeo AR, Flowers TJ. editors. *Soil Mineral Stresses: Approaches to Crop Improvement*. Berlin: Springer. (1994). p. 175–200. doi: 10.1007/978-3-642-84289-4_8
110. Epstein E, Bloom AJ. *Mineral Nutrition of Plants: Principles and Perspectives*. 2nd ed. Sunderland, MA: Sinauer Associates (2005).

111. Jiang W, Struik PC, Zhao M, Van Keulen H, Fan TQ, Stomph TJ. Indices to screen for grain yield and grain zinc mass concentration in aerobic rice at different soil Zn levels. *NJAS Wageningen J Life Sci.* (2008) 55:181–97. doi: 10.1016/S1573-5214(08)80036-X
112. Stomph TJ, Hoebe N, Spaans E, Van der Putten PEL. *The Relative Contribution of Post-Flowering Uptake of Zinc to Rice Grain Zinc Density*. In: 3rd International Zinc Symposium 10-14, Hyderabad, India (2011).
113. Farooq M, Wahid A, Siddique KHM. Micronutrient application through seed treatments: a review. *J Soil Sci Plant Nutr.* (2012) 12:125–42. doi: 10.4067/S0718-95162012000100011
114. Singh MV. *Micronutrient Seed Treatments to Nourish the Crops at the Critical Stage of Growth.* Tech. Bull. IISS, Bhopal (2003). p. 1–93.
115. Welch RM. Breeding strategies for biofortified staple plant foods to reduce micronutrient malnutrition globally. *J Nutr.* (2002) 132:495–9. doi: 10.1093/jn/132.3.495S
116. Yang XE, Chen WR, Feng Y. Improving human micronutrient nutrition through biofortification in the soil plant system: China as case study. *Environ Geochem Health.* (2007) 29:413–28. doi: 10.1007/s10653-007-9086-0
117. Cakmak I. Enrichment of fertilizers with zinc: an excellent investment for humanity and crop production in India. *J Trace Elem Med Biol.* (2009) 23:281–9. doi: 10.1016/j.jtemb.2009.05.002
118. Johnson SE, Lauren JG, Welch RM, Duxbury JM. A comparison of the effects of micronutrient seed priming and soil fertilization on the mineral nutrition of chickpea (*Cicer arietenum*), lentil (*Lens culinaris*), rice (*Oryza sativa*), and wheat (*Triticum aestivum*) in Nepal. *Exp Agric.* (2005) 41:427–48. doi: 10.1017/S0014479705002851
119. Karak T Das D. *Effect of Foliar Application of Different Sources of Zn Content, Uptake, and Yield of Rice (Oryza sativa L.).* In: 18th World Congress of Soil Science, July 9-15, Philadelphia, Pennsylvania, USA (2011).
120. Yoshida S. *Fundamentals of Rice Crop Science*. Los Baños: International Rice Research Institute (1981).
121. Katyal JC, Ponnamperuma FN. Zinc deficiency: a widespread nutritional disorder in Agusan del Norte, Philippines. *Agric J.* (1979) 58:79–89.
122. De Datta SK. *Principles and Practices of Rice Production*. New York, NY: John Wiley and Sons (1981).
123. Khan MU, QasimM, Subhan M, Jamil M, Amad RD. Response of rice to different methods of Zn application in calcareous soils. *Pak J Appl Sci*. (2003) 3:524–9. doi: 10.3923/jas.2003.524.529
124. Franzluebbers AJ, Hons FM. Soil-profile distribution of primary and secondary plant-available nutrients under conventional and no tillage. *Soil Till Res.* (1996) 39:229–39. doi: 10.1016/S0167-1987(96)01056-2
125. Bhaduri D, Purakayastha TJ. *Soil Available Zn: a Potent Soil Quality Indicator in Rice-Wheat System*. In: International Zinc Symposium, 10–14, Hyderabad, India (2011).
126. Mandal B, Hazra GC, Mandal LN. Soil management influence in zinc desorption for rice and maize nutrition. *Soil Sci Soc Am J.* (2000) 64:1699–705. doi: 10.2136/sssaj2000.6451699x
127. Eghball B, Ginting D, Giley JE. Residual effect of manure and compost applications on corn production and soil properties. *Agron J.* (2004) 96:442–7. doi: 10.2134/agronj2004.4420
128. Mandal B, Chatterjee J, Hazra GC, Mandal LN. Effect of preflooding on transformation of applied zinc and its uptake by rice in lateric soils. *Soil Sci.* (1992) 153:250–7. doi: 10.1097/00010694-199203000-00009
129. Gao S, Tanji KK, Scardaci SC, Chow AT. Comparison of redox indicators in paddy soil during rice growing season. *Soil Sci Soc Am J.* (2002) 66:805–17. doi: 10.2136/sssaj2002.8050
130. Tuyogon DSJ, Impa SM, Castillo OB, Lazaro W, Johnson-Beebout SE. Enriching rice grain zinc through zinc fertilization and water management. *Soil Sci Soc Am J.* (2016) 80:121–34. doi: 10.2136/sssaj2015.07.0262
131. Bureau of Soils and Water Management-Department of Agriculture (BSWM-DA). *Personal communication.* Manila: BSWM (2018).
132. Yoshida S, Ahn JS, Forno DA. Occurrence, diagnosis, and correction of zinc deficiency in lowland rice. *Soil Sci Plant Nutr*. (1973) 19:83–93. doi: 10.1080/00380768.1973.10432522
133. Katyal JC, Vlek PLG. Micronutrient problems in tropical Asia. In: Vlek PLG. editor. *Micronutrients in Tropical Food Production*. Dordrecht: Martinus Nijhoff/Dr. W. Junk Publishers (1985). p. 69–94. doi: 10.1007/978-94-009-5055-9_3
134. Sillanpaa M. *Micronutrient and Nutrient Status of Soils, a Global Study*. Rome: FAO Soils Bull (1990). p. 414.
135. Alloway BJ. *Zinc in Soils and Crop Nutrition*. Brussels: IZA and IFA (2004).
136. Graham RD. Micronutrient deficiencies in crops and their global significance. In: Alloway BJ editor. *Micronutrient Deficiencies in Global Crop Production*. Dordrecht: Springer (2008). p. 41–61. doi: 10.1007/978-1-4020-6860-7_2
137. Descalsota JP, Mamaril CP, San Valentin JO. Fertility status of selected rice soils in the Philippines. *Philipp J Crop Sci.* (2005) 28:45–57. Retrieved from https://www.cabi.org/GARA/FullTextPDF/2009/20093019309.pdf
138. Nube M, Voorman RL. *Simultaneously Addressing Micronutrient Deficiencies in Soil, Crops, Animal and Human Nutrition: Opportunities for Higher Yields and Better Life*. Center for World Food Studies. Staff Working Paper (2006).
139. Speretto RA, Ricachenevsky FK, Waldow V, MullerALH, Dressler VL, Fett JP. Rice grain Fe, Mn, and Zn accumulation: How important are flag leaves and seed number? *Plant Soil Environ.* (2013) 59:262–6. doi: 10.17221/841/2012-PSE
140. Impa SM, Morete MJ, Ismail AM, Schulin R, Johnson-Beebut SE. Zn uptake, translocation and grain Zn loading in rice (*Oryza sativa*) genotypes selected for Zn-deficiency tolerance and high grain Zn. *J Exp Bot.* (2013) 64:2739–51. doi: 10.1093/jxb/ert118
141. Nakandalage N, Nicolas M, Norton R, Hirotsu N, Milham P, Seneweera S. Improving rice zinc biofortification success rates through genetic and crop management approaches in a changing environment. *Front Plant Sci.* (2016) 7:764. doi: 10.3389/fpls.2016.00764
142. Department of Health. *National Nutrition Council.* Philippine Plan of Action for Nutrition 2017–2022. Manila, Philippines (2017).
143. Lutsey PL, Dawe D, Villate E, Valencia S, Lopez O. Iron supplementation compliance among pregnant women in Bicol. *Pub Health J.* (2007) 11:76–82. doi: 10.1017/S1368980007000237
144. Mannar V Gallego EB. Fe fortification: country level experience and lessons learned. *J Nutr.* (2002) 132:856S—8. doi: 10.1093/jn/132.4.856S
145. Bouis H. Enrichment of food staples through plant breeding: a new strategy for fighting micronutrient malnutrition. *Nutrition.* (1996) 54:131–7. doi: 10.1111/j.1753-4887.1996.tb03915.x
146. Food and Nutrition Research Institute. *Philippine Nutrition Profile*. Manila: FNRI-DOST. (2014).
147. Department of Health. *Department Memorandum No. 2010-0010*. DOH, Manila, Philippines (2010).
148. Food and Nutrition Research Institute. *FNRI Digest*. Vol 27. (2010). Manila, Philippines.
149. Available online at: http://pcw.gov.ph/law/executive-orders-no-51 (accessed November 10, 2018)
150. Department of Health. *Department Memorandum No. 2011-0303*. DOH, Manila, Philippines (2011).
151. Available online at: https://www.doh.gov.ph/micronutrient-program (accessed November 10, 2018)
152. Engesveen K, Nishida C, Prudhon C, Shrimpton R. Assessing countries' commitment to accelerate nutrition action demonstrated in PRSPs, UNDAFs, and through the nutrition governance. *SCN News.* (2009) 37:10–6. Retrieved from https://www.researchgate.net/publication/285667381_Assessing_countries'_commitment_to_accelerate_nutrition_action_demonstrated_in_PRSP_UNDAF_and_through_nutrition_governance
153. World Health Organization. *Western Pacific Region*. Report of the WHO Representative. The Work of World Health Organization in the Philippines (2011).
154. UNICEF. *UNICEF: Seven Decades of Upholding the Rights of Filipino Children.* Mandaluyong: UNICEF. (2018).
155. FAO Representation in the Philippines. *Philippines and FAO: Achievements and success stories.* Makati City, Philippines (2011).
156. Stein AJ, Meenaskshi JV, Quiam M, Nestel P, Sachdev SPS, Bhutta ZA. Health benefits of biofortification – ex-ante analysis of iron-rice and iron-wheat in India. In: *Association Annual Meeting, Providence, Rhode Island.* Providence, RI. (2005).
157. Boui HE, Hotz C, McClafferty B, Meenakshi JV, Wolfgang PH. Biofortification: a new tool to reduce micronutrient malnutrition. In:

Paper presented at the 19th International Congress of Nutrition, Bangkok, Thailand. (2009).
158. Bouis HE, Saltzman A. Improving nutrition through biofortification: a review of evidence from HarvestPlus, 2003 through 2016. *Global Food Security*. (2017) 12:49–58. doi: 10.1016/j.gfs.2017.01.009
159. Hoddinott J, Rosegrant M, Torero M. *Investment to Reduce Hunger and Undernutrition*. Cph. Consens. Chall. Pap. (2012).
160. IRRI. *Bringing Hope, Improving Lives, Strategic Plan 2007-2015*. International Rice Research Institute. (2007).
161. Virk K. Zinc Rice. In: *The 2nd Global Conference on Biofortification: Getting Nutritious Foods to People*. CIAT-HarvestPlus.
162. Available online at: www.harvestplus.org (accessed November 11, 2018)
163. Yadava DK, Choudhury PR, Hossain F, Kumar D. *Biofortified Varieties: Sustainable Way to Alleviate Malnutrition*. Indian Council of Agricultural Research, New Delhi. (2017).
164. Lei X. Biofortification in China. In: *The 2nd Global Conference on Biofortification: Getting Nutritious Foods to People* Kigali, Rwanda. (2014).
165. Graham RD, Senadhira D, Beebe SE, Iglesias C, Ortiz-Monasterio I. Breeding for micronutrient density in edible portions of staple food crops: conventional approaches. *Field Crops Res*. (1999) 60:57–80. doi: 10.1016/S0378-4290(98) 00133-6
166. Arocena EC, Ticman HT, Geneston KB, Pini M, Palanog AD, Jimenez H. *Expanded GxE Experiments in Different Agro-Ecologies in Support of Bangladesh and Eastern India High-zinc Rice Profiles: Multi-location Evaluation of Recombinant Inbred Lines for Identifying Most Adapted Line for Varietal Promotion (Philippines)*. Annual Report 2016. Philippine Rice Research Institute.
167. Swamy BPM, Rahman MA, Inabangan-Asilo MA, Amparado A, Manito C, Chadha-Mohanty P, et al. Advances in breeding for high Zinc in Rice. *Rice*. (2016) 9:49–65. doi: 10.1186/s12284-016-0122-5
168. Babu VR, Shreya K, Dangi KS, Usharani G, Nagesh P. Evaluation of Popular Rice (*Oryza sativa* L.) Hybrids for quantitative, qualitative, and nutritional aspects. *Int J Sci Res Publicat*. (2013) 3:1–8. Retrieve from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.299.4137
169. Stangoulis J, Huynh BL, Welch RM, Choi EY, Graham RD. Quantitative trait loci for phytate in rice grain and their relationship with grain micronutrient content. *Euphytica*. (2007) 154:289–94. doi: 10.1007/s10681-006-9211-7
170. Available online at:www.philrice.gov.ph/eat-rice-riceponsable/ (accessed November 11, 2018)
171. Haas JD, BeardJL, Murray-Kolb LE, del Mundo AM, Felix A, Gregorio G. Iron-biofortified rice improves the iron stores of non-anemic Filipino women. *J Nutr*. (2005) 135:2823–30. doi: 10.1093/jn/135.12.2823
172. House WA. Trace elements bioavailability as exemplified by Fe and zinc. *Field Crops Res*. (1999) 60:115–41. doi: 10.1016/S0378-4290(98) 00136-1
173. Van Campen DR, Glahn RP. Micronutrient bioavailability techniques: accuracy, problems and limitations. *Field Crops Res*. (1999) 60:93–113. doi: 10.1016/S0378-4290(98)00135-X

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Palanog, Calayugan, Descalsota-Empleo, Amparado, Inabangan-Asilo, Arocena, Sta. Cruz, Borromeo, Lalusin, Hernandez, Acuin, Reinke and Swamy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 August 2019
doi: 10.3389/fgene.2019.00717

Association Mapping for 24 Traits Related to Protein Content, Gluten Strength, Color, Cooking, and Milling Quality Using Balanced and Unbalanced Data in Durum Wheat [*Triticum turgidum* L. var. *durum* (Desf).]

Marina Johnson, Ajay Kumar, Atena Oladzad-Abbasabadi, Evan Salsman, Meriem Aoun, Frank A. Manthey and Elias M. Elias*

Department of Plant Sciences, North Dakota State University, Fargo, ND, United States

OPEN ACCESS

Edited by:
Steven H. Zeisel,
University of North Carolina at Chapel Hill, United States

Reviewed by:
Liu Jin Dong,
Chinese Academy of Agricultural Sciences, China
Mallikarjuna Swamy,
International Rice Research Institute, Philippines
Pawan L. Kulwal,
Mahatma Phule Agricultural University, India

****Correspondence:***
Elias M. Elias
Elias.Elias@ndsu.edu

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Genetics

Received: *05 February 2019*
Accepted: *08 July 2019*
Published: *16 August 2019*

Citation:
Johnson M, Kumar A, Oladzad-Abbasabadi A, Salsman E, Aoun M, Manthey FA and Elias EM (2019) Association Mapping for 24 Traits Related to Protein Content, Gluten Strength, Color, Cooking, and Milling Quality Using Balanced and Unbalanced Data in Durum Wheat [Triticum turgidum L. var. durum (Desf).]. Front. Genet. 10:717. doi: 10.3389/fgene.2019.00717

Durum wheat [*Triticum durum* (Desf).] is mostly used to produce pasta, couscous, and bulgur. The quality of the grain and end-use products determine its market value. However, quality tests are highly resource intensive and almost impossible to conduct in the early generations in the breeding program. Modern genomics-based tools provide an excellent opportunity to genetically dissect complex quality traits to expedite cultivar development using molecular breeding approaches. This study used a panel of 243 cultivars and advanced breeding lines developed during the last 20 years to identify SNPs associated with 24 traits related to nutritional value and quality. Genome-wide association study (GWAS) identified a total of 179 marker–trait associations (MTAs), located in 95 genomic regions belonging to all 14 durum wheat chromosomes. Major and stable QTLs were identified for gluten strength on chromosomes 1A and 1B, and for PPO activity on chromosomes 1A, 2B, 3A, and 3B. As a large amount of unbalance phenotypic data are generated every year on advanced lines in all the breeding programs, the applicability of such a dataset for identification of MTAs remains unclear. We observed that ~84% of the MTAs identified using a historic unbalanced dataset (belonging to a total of 80 environments collected over a period of 16 years) were also identified in a balanced dataset. This suggests the suitability of historic unbalanced phenotypic data to identify beneficial MTAs to facilitate local-knowledge-based breeding. In addition to providing extensive knowledge about the genetics of quality traits, association mapping identified several candidate markers to assist durum wheat quality improvement through molecular breeding. The molecular markers associated with important traits could be extremely useful in the development of improved quality durum wheat cultivars using marker-assisted selection (MAS).

Keywords: durum wheat, genome wide association mapping, gluten strength, grain and pasta color, Infinium iSelect 90k, pasta, protein content, unbalanced data

INTRODUCTION

Durum wheat, the hardest of all wheats (Miller et al. 1982), is one of the most important crops in the world. Annual worldwide durum wheat production is estimated to be around 36 million tons, with approximately, 2.5 million tons produced in the United States (NASS, 2018). The market price of durum wheat is generally higher than other wheat classes, which makes durum wheat attractive for growers. Apart from grain yield and disease resistance, quality is a strong criterion of durum wheat variety selection. Grain quality parameters differ based on the end use but usually include visual appearance, vitreousness (VIT), weather damage (falling number), and protein quantity and quality. During milling, durum wheat endosperm is ground into coarse particles called semolina, which is then used for making pasta and couscous. The quality of semolina is determined by many factors including ash content, color, and particle-size distribution. Grain protein concentration is important for high-quality pasta production due to its effect on the firmness of cooked pasta and tolerance to overcooking (Dexter and Matsuo, 1977). Gluten strength can also affect pasta characteristics such as tolerance to overcooking, reduced stickiness, and minimal loss of solids during cooking. The degree of kernel translucency, and thus the apparent degree of vitreousness, is related to the degree of kernel compactness. Generally, more desirable coarse semolina is produced from highly vitreous wheats (Hoseney 1987). Milling yield is another important criterion for the milling industry, as higher semolina yield means higher profits for durum wheat millers. Pasta color is an important consideration by the consumers. Therefore, processors prefer clear, bright yellow semolina, which generally produces a superior end product. Polyphenol oxidases (PPOs) produce dark/brown polyphenols (Anderson and Morris, 2003). When PPO activity reaches substantial levels, an unappealing brown color can occur in end products (Sissons, 2008). Cooked quality for pasta, which is determined by its cooked firmness, cooking loss, and cooked weight, is important to producers. Durum wheat with high grain and end-product quality can receive premium prices in national and international markets.

The phenotypic evaluation of many important quality traits requires large amounts of grain per line, which is not feasible in early generations of breeding due to the large number of lines and limited resources in most breeding programs. Marker-assisted selection provides an opportunity for breeders to identify superior lines in early generations, thus saving significant resources and speeding up the process of cultivar development. However, most of the quality traits in wheat are quantitatively inherited and influenced by multiple QTL, QTL × QTL, the environment, and the interaction between QTL and environment (McCartney et al., 2005; Li et al., 2012; Kumar et al., 2013; Tadesse et al., 2015; Kumar et al., 2018; Merida-Garcıa et al., 2019). Therefore, a complete knowledge about the genetics of target traits and identification of markers tightly linked to those target traits is essential for the successful integration of MAS in the breeding program. Traits like grain protein content (Blanco et al., 2006; Blanco et al., 2012; Kumar et al., 2018), gluten strength (Kumar et al., 2013; Kumar et al., 2018), and color (Pozniak et al., 2007; Zhang and Dubcovsky, 2008b; Roncallo et al., 2012) have been investigated in some details in durum wheat. However, a limited number of studies have been conducted to genetically dissect other important quality traits related to milling, PPO (Si et al., 2012), and pasta cooking (Zhang et al., 2008a), which determine the end-use value of durum wheat (Fiedler et al., 2017; N'Diaye et al., 2017). Also, most of those genetic dissection studies conducted in durum wheat were based on bi-parental mapping and low-resolution maps. Genome-wide association mapping studies based on advanced breeding lines may have more relevant results compared with those generated from bi-parental populations and diverse panels because of their direct application in the breeding program.

During the process of developing germplasm and cultivars in a breeding program, a large amount of phenotypic data are routinely collected on breeding lines in advanced generations. However, each year, only a small number of advanced breeding lines are evaluated. A few of those are replaced with new lines over the next few years, resulting in data from a large collection of advanced breeding lines composed of genotypes that are evaluated in different years and locations, generating unbalanced data. With genotyping cost decreasing day by day, the AM approach for identifying QTL could be even more cost-effective if a large amount of phenotypic data routinely collected by breeding programs could be used to gain insight of the genetics of quantitative traits and to identify MTAs for molecular breeding. However, information about the application of such unbalanced data for genetic studies is mostly missing. Therefore, in this study, a panel of advanced breeding lines and cultivars of durum wheat were used to identify 1) genomic regions associated with 24 quality traits, 2) associated markers suitable for MAS, and 3) whether historic unbalanced data are suitable for AM analysis and QTL identification. To our knowledge, this is the first such comprehensive study demonstrating the application of historic unbalanced phenotypic data for genome-wide association studies of complex, low, and moderate heritability traits in wheat.

Abbreviations: AFLP, amplified fragment length polymorphism; AM, association mapping; CLOSS, cooking loss; CV, coefficient of variance; CWT, pasta cooked weight; FIRM, pasta firmness; FN, falling number; GI, gluten index; GWAS, genome-wide association study; LD, linkage disequilibrium; MAF, minor allele frequency; MAS, marker-assisted selection; MIXO, mixogram score; MSD, mean squared difference; PCA, principal component analysis; FDR, false discovery rate; RFLP, restriction fragment length polymorphism; SASH, semolina ash; SDS, sodium dodecyl sulfate micro-sedimentation; SEXT, semolina extraction; SNP, single-nucleotide polymorphism; SPROT, semolina protein; SSR, simple sequence repeat; TEXT, total extraction; TYP, total yellow pigment; WG, wet gluten; WPROT, whole-wheat protein; WTS, work to shear.

MATERIALS AND METHODS

Plant Material and Field Evaluation

The AM panel consisted of 243 durum wheat cultivars and inbred lines ($F_{5:9}$) entered into the Uniform Regional Durum Nursery (URDN) from 1997 to 2014 (except in 2010 and 2011 due to severe weather conditions). These cultivars and inbred lines were chosen based on the phenotypic data available for the

agronomic and quality traits routinely collected over the years. The historic unbalanced data were collected on these lines from 1997 until 2014. Each year, a subset of those lines was evaluated at five locations in North Dakota (Williston, Minot, Langdon, Carrington, and Prosper) in a randomized complete block design (RCBD) with four replications. Some of the lines were replaced each year with new inbred lines and evaluated in the same manner. So the unbalanced phenotypic data for the panel were obtained for the 80 environments belonging to 16 growing seasons. In 2015, a panel of 256 lines (13 additional lines were added to make a 16 × 16 lattice) was also evaluated together at two locations (Prosper and Langdon) to collect balanced data for the traits under study. Genotypes were planted in a 16 × 16 simple lattice design with two replicates. For both datasets, individual plots consisted of four 2-m-long rows spaced 0.3 m apart. Plots were harvested with a plot combine (HEGE 140), and the grain was collected in individual sacks. The grain was dried to approximately 12% moisture content and stored at 16°C until further processing. Quality tests were performed in the Durum Wheat and Pasta Processing Laboratory at North Dakota State University (NDSU).

Data Collection

Sample Preparation

Post-harvest cleaning was conducted using the Carter-Day Dockage Tester (Simon-Carter-Day Company, Minneapolis, Minnesota) configured with a number 25 riddle and a number 2 top sieve and number 2 bottom sieve. The samples were cleaned again using the Carter-Day Dockage Tester configured with a number 25 riddle, a number 8 top sieve, and a number 2 bottom sieve. For quality traits analyses, subsamples from the total amount of seed available were taken and ground into whole-wheat flour using a Udy Cyclone Mill (UDY Corporation, Boulder, Colorado) fitted with a 60-mesh sieve and stored in plastic bags at 4°C until tests were performed. Ash content, protein content, sedimentation volume, falling number, and total yellow pigment content tests were performed using the whole-wheat samples.

Subsamples for milling were taken from the available grain. Prior to milling, all subsamples were tempered in two stages based on grain moisture. In the first stage, they were tempered to 12.5% moisture for at least 72 h; and in the second stage, they were tempered to 15% moisture for 24 h before milling. The samples were milled into semolina using a Quadramat Jr. Mill (C.W. Brabender Instruments, Inc., South Hackensack, New Jersey) according to the American Association of Cereal Chemists International (AACCI) method 26-50.01 (AACC International, 2008). Semolina samples were kept at 4°C until further analysis. Different approaches for semolina mixing and extrusion were used for the two datasets due to the amount of semolina available. For the historic unbalanced dataset, 1,000 g of semolina was hydrated and mixed in a Hobart C-100-T mixer equipped with a pastry knife agitator. The mixer was set on low speed for 10 s while distilled water was added and then on high speed for 50 s. The semolina was mixed for an additional 2 min on high speed to ensure it reached a complete premix stage. Processing was done in a semicommercial-scale pasta extruder (DeMaco, Melbourne, Florida) and extruded through an 84-strand 0.043-inch Teflon spaghetti die. A jacketed extrusion tube (23-cm length × 4.4-cm inside diameter) was attached to the pasta extruder to allow a longer time for semolina hydration to minimize white specks in the spaghetti. Actual conditions for dough extrusion were a screw rotation speed at 28–29 rpm, a vacuum at 0.8–1.05 kg/cm^2, and a jacket temperature at 46–48°C. Room temperature and relative humidity were maintained at 25°C and 40–45%, respectively.

For the 2015 balanced dataset, 300 g of semolina was hydrated and mixed in a KitchenAid commercial mixer. The mixer was set on low speed for 10 s while distilled water was added and then on high speed for 50 s. The semolina was mixed for an additional 2 min on high to ensure it reached a complete premix stage. Processing was done using a commercial tabletop electric pasta machine (Arcobaleno, Lancaster, PA, model AEX18) and extruded through a 35-strand 1.09-mm Teflon spaghetti die.

Due to the large number of samples evaluated in 2015 and the amount of time needed for drying, the drying process was omitted for the 2015 balanced dataset. For the historic unbalanced dataset, the extruded spaghetti was dried in a laboratory pilot-scale dryer (Standard Industries, Fargo, North Dakota) on the low-temperature cycle with a total drying time of 18 h. The low-temperature cycle typically has an 18-h total drying time at 40°C (Yue et al., 1999).

Two different spaghetti cooking times were used for the two datasets. In the historic unbalanced dataset, dry spaghetti (10 g) was broken into lengths of approximately 5 cm and placed in 300 ml of boiling water for 12 min. However, based on preliminary results, the cooking time for the fresh spaghetti in the 2015 balanced dataset was reduced to 4 min. Fresh spaghetti (10 g) was cut into lengths of approximately 5 cm and placed in 300 ml of boiling water for 4 min. The optimum cooking time was determined using AACC method 66-50 (AACC International, 2008).

Phenotypic data for following quality traits were recorded:

Protein

Grain protein content (WPROT): Protein content was determined using an Infratec 1226 Whole Grain Analyzer (FOSS Tecator, Höganäs, Sweden). The data on protein content were adjusted to 14% moisture content.

Semolina protein content (SPROT): Semolina protein content was determined using AACC method 39-25.01 (AACC International, 2008) adapted for the FOSS Infratec 1241 Grain Analyzer (Foss North America, Eden Prairie, Minnesota).

Milling-Related Traits

To produce durum wheat semolina and flour, the samples were milled using a Quadramat Jr. Mill (C.W. Brabender Instruments, Inc., South Hackensack, New Jersey) according to AACC method 26-50.01 (AACC International, 2008). Total extraction refers to the portion

of the durum wheat kernel that can be milled into flour and semolina. Semolina extraction is only that portion of the wheat kernel that is milled into semolina. For uniformity, all extractions were adjusted to 14% moisture and expressed on constant moisture basis.

Total extraction (TEXT): Total extraction was expressed as a percentage weight per weight (w/w) of semolina and flour from tempered durum wheat samples.

Semolina extraction (SEXT): Semolina extraction was expressed as a percentage weight per weight (w/w) of semolina from tempered durum wheat samples.

Semolina ash content (SASH), moisture content: Approved methods (AACC International, 2008) were used to determine ash (method 08-01.01) and moisture content (method 44-15.02).

Vitreousness (VIT): The percentage of vitreous kernels was determined by cutting 100 kernels taken at random transversally with a farinator (grain splitter) and identifying the grains that were not fully vitreous according to the appearance of the sectional areas of the endosperm. Vitreous grains are translucent and transparent when cut, while starchy grains are white and opaque due to the existence of air pockets in the endosperm (Hoseney, 1986).

Gluten-Related Traits

Sedimentation volume (SDS): The sedimentation volume was measured using sodium dodecyl sulfate (SDS) micro-sedimentation test as described by Dick and Quick (1983).

Mixogram score (MIXO): Ten grams of semolina, based on 14% moisture, was weighed. Water was added based on the grain protein content using a formula ($Y = 1.5X + 43.6$) described in AACC method 54-40A (2008), where Y = amount of water (ml) added to the sample and X = protein content at 14% mb. The 10-g bowl mixograph (National Manufacturing, TMCO Division, Lincoln, Nebraska) was used to measure the dough mixing strength of semolina. Mixing tolerance was scored using a scale of 1 (weak) to 8 (strong).

Wet gluten (WG) and gluten index (GI): Wet gluten and gluten index were determined with the glutomatic instrument (Perten Instruments, Springfield, Illinois) using AACC method 38-12.02 (AACC International, 2008).

Glutork (GLUT): Water binding capacity (water bound in wet gluten) was determined with Glutork 2020 (Perten Instruments, Springfield, Illinois) using AACC method 38-12.02 (AACC International, 2008) and expressed as difference between wet and dry gluten content (%).

Color-Related Traits

Semolina color (color a, color b, and color L): Semolina color was determined using the Minolta colorimeter CIEL CR410 (Hunter lab *L*, a, b). Value "*L*" indicates lightness or brightness, value "b" indicates yellowness, and value "a" indicates "greenness."

Difference in color a (dif_a), difference in color b (dif_b), and difference in color L (dif_L):

A semolina dough sheet was made using a modified method described by Fu et al. (2011). A total of 30 g of semolina was hydrated to 38% moisture at 45°C and mixed for 1 min in a KitchenAid mixer (4.3 L KitchenAid CLASSIC Stand Mixer 5K45SS) at speed 4. After being mixed, the dough was sheeted twice in a pair of sheeting rolls with a gap of 1 mm. The resulting dough sheet was folded twice and sheeted twice in a pair of sheeting rolls with a gap of 3 mm without folding. The smooth dough sheet was transferred to a plastic bag and stored in a closed drawer at room temperature. Color was measured on the dough sheet at intervals of 0.5 and 24 h using a Minolta colorimeter CIEL CR410 (Hunter lab *L*, a, b). Differences in color a, color b, and color L were measured between the time intervals.

Dry pasta color (color): Visual color was determined under the constant light source and assigned a numerical visual color score from 1 to 12, with 12 as the best score. The scores were generated according to the color map designed by Debbouz (1994).

Total yellow pigment (TYP): Total yellow pigment (TYP) content was determined using the water-saturated *n*-butanol AACC method 14-50.01 (AACC International, 2008) as modified by using 2 g of ground whole meal. Water-saturated *n*-butanol (10 ml) was added to 2 g of whole meal and shaken for 2 min. After a 30-min rest, the extracts were centrifuged at 12,000 rpm for 10 min, and the supernatant was carefully transferred to cuvettes. Absorbance was measured using a spectrophotometer (Beckman Coulter DU 720 General Purpose UV/Vis Spectrophotometer) at a wavelength of 435.8 nm. Measurements per extracted sample were recorded, and values averaged and converted to yellow pigment concentration (μg/g) using the extinction coefficient (1.6632) for β-carotene (Sims and Lepage, 1968).

Polyphenol oxidase (PPO): Polyphenol oxidase activity was determined using intact kernels as described by Anderson and Morris (2001) using AACC method 22-85.01 (AACC International, 2008). A 1.5 ml aliquot of 10 mM of L-DOPA (L-3,4-dihydroxyphenylalanine) containing 0.02% v/v Tween-20 as a substrate in a 50 mM of MOPS [3-(*N*-morpholino)propane sulfonic acid] buffer with a pH of 6.5 was added to five undamaged seeds in a 2-ml microcentrifuge tube. The tubes were placed on an orbital shaker (Glas-Col, Terre Haute, Indiana) and rotated for 1 h at room temperature to allow the reaction to occur. Polyphenol oxidase activity was measured as the change in absorbance at 475 nm using a Beckman Coulter spectrophotometer (Beckman Coulter DU 720 General Purpose UV/Vis Spectrophotometer, Fullerton, California). The L-DOPA solution was made fresh daily. Each sample was measured in duplicate.

Cooking-Related Traits

Cooked weight (CWT): After being cooked, samples were rinsed off with distilled water in a Buchner funnel and drained. Spaghetti strips were weighed and reported in grams.

Cooking loss (CLOSS): Cooking loss (% weight of solids) was measured by evaporating the cooking water to dryness in a forced-air oven at 110°C overnight. The residue was weighed and reported as percentage of the dry spaghetti.

Cooked spaghetti firmness (FIRM): Cooked firmness was measured using a plexiglass blade probe attached to a Texture Analyzer (Model TA-XT, Texture Technology Corporation, Scarsdale, New York) as described by Walsh (1971). Five strands of cooked spaghetti were placed on a plexiglass plate and sheared at a 90° angle with a plexiglass tooth probe. Firmness was measured as the maximum shear strength of curve (g).

Work to shear (WS): Five strands of cooked spaghetti were placed on a plexiglass plate and sheared at a 90° angle with a plexiglass tooth probe. A TA-XT2 texture analyzer was used to calculate the area under the curve (g·cm), indicating the amount of work required to shear the cooked spaghetti (the CF score). The average of three CF scores was used to report CF.

Statistical Analysis

The analysis of variance was conducted using the Statistical Analysis System (SAS) computer package version 9.3 (SAS, 2004). The unbalanced historic dataset was analyzed using a mixed linear model (MLM) with Proc Mixed method III, where genotypes were the fixed effects, and environments and replicates within environments were the random effects. The balanced dataset was analyzed using Proc GLM method III. Least square (LS) means were used for the analyses (Steel and Torrie, 1980).

The entry means plot-based heritability for all the traits were estimated using the method described by Holland et al. (2003). The variance and covariance parameters were calculated using the COVTEST and ASYCOV options of the MIXED procedure (SAS, 2004), with environments and genotypes deemed random. Trait correlations were calculated and plotted in R 3.0 (Venables et al., 2017) using cor.matrix and corrplot from the corrplot package. Correlation values were considered significantly different from zero at $P \leq 0.05$.

DNA Isolation and SNP-Marker Genotyping and Analysis

Four seeds from each genotype were planted into potting mix in the greenhouse in the fall of 2014. Three young leaf tissues from each genotype were harvested and sent to the USDA-ARS Cereal Crops Genotyping Laboratory in Fargo, ND, for DNA isolation. The extracted DNA samples were genotyped using the Illumina 90k iSelect BeadChip platform, and the markers were called using the diploid version of GenomeStudio software (Wang et al., 2014). FastPHASE 1.3 software with the default settings (Scheet and Stephens, 2006) was used to impute missing loci data using a "likelihood"-based imputation. The heterozygotes were considered missing. Only markers having a minor allele frequency (MAF) > 0.05 were considered for further analysis.

Linkage disequilibrium for all pairwise comparisons between intra-chromosomal SNP was computed, and the genome-wide LD decay was estimated using JMP Genomics 8.1 software (SAS, 2004). The LD was computed as the squared correlation coefficient (R^2) for each of the marker pairs. Genome-wide LD decay was estimated by plotting LD estimates (R^2) from all 14 chromosomes against the corresponding pairwise genetic distances (cM). Smoothing spline Fit (lambda = 338064.8) was applied to the estimate of LD decay.

Association Mapping (AM) Analysis

Association mapping analysis was done using JMP Genomics 8.1 software (SAS, 2004; Zhao et al., 2007). Population structure (Q matrix), which can be defined as the differential relatedness among genotypes, was controlled with principal component analysis (PCA). The identity-by-state (IBS) matrix (K matrix) representing the proportion of shared alleles for all pairwise comparisons in each population was applied. Five regression models were generated to analyze marker–trait association: 1) naive, 2) kinship, 3) kinship plus population structure (the first two principle components (PCs) collectively explained 11.3% of genotype variation), 4) kinship plus population structure (the first three PCs collectively explained 15.46% of genotype variation), and 5) the kinship plus population structure (the first four PCs collectively explained 19.2% of genotype variation). The best model was determined according to the Bayesian information criterion (BIC), where the lowest BIC value is preferred (Ghosh et al., 2006; Zhang et al., 2010). The *P* values of the selected models were later adjusted by calculating the corresponding false discovery rate (FDR) (Benjamini and Yekutieli, 2001). Marker–trait associations were considered significant at an FDR ≤ 0.1.

RESULTS

Phenotypic Data Analysis

There were significant differences among genotypes for most of the traits in both balanced and unbalanced datasets (**Table 1**). Also, the environment had a significant effect on most of the traits as indicated by the significant genotype by environment interactions (**Table 1**). The correlation analysis showed significant correlation between related traits. Grain protein (WPROT) was consistently positively correlated with VIT, WG, FIRM, and GLUT and negatively correlated with CWT and CLOSS (**Figure 1**). Semolina ash (SASH) was positively correlated with WPROT. Total yellow pigment (TYP) was positively correlated with semolina color b and negatively correlated with semolina color L (**Figure 1**). Gluten strength, measured by GI, SDS, and MIXO, was significantly positively correlated with spaghetti firmness (FIRM), but SDS, MIXO, and GI had no significant correlation with WPROT, CWT, or CLOSS (**Figure 1**). These

TABLE 1 | Analysis of variance and heritability estimates for different quality traits in durum wheat association mapping panel.

Traits	Mean squares					H¶	SE#
	Genotype (G)	Location (L)	Rep(L)	G*L	Error		
Unbalanced data							
Grain protein (WPROT)	1.92**	87.19*	88.74**	0.41**	0.35	0.209	0.020
Semolina protein (SPROT)	1.45**	82.62*	79.82**	0.36**	0.27	0.194	0.020
Total extraction (TEXT)	6.96**	546.61**	89.63**	1.24*	1.07	0.288	0.024
Semolina extraction (SEXT)	8.84**	405.46**	79.43**	1.28	1.20	0.338	0.024
Semolina ash (SASH)	0.008**	0.313**	0.233**	0.002**	0.002	0.224	0.022
Kernel vitreousness (VIT)	47,310**	32,988**	300,035**	16,500	152,094	0.223	0.021
Sedimentation volume (SDS)	643.98**	2,636.55**	1,953.89**	28.72**	21.77	0.574	0.023
Gluten index (GI)	3,531.53**	6,918.65**	3,226.44**	116.41**	91.77	0.655	0.021
Wet gluten (WG)	44.04**	766.45*	773.30**	7.69**	5.43	0.291	0.025
Mixogram score (MIXO)	12.09**	8.95	11.54**	0.54**	0.42**	0.551	0.023
Pasta color (Color)	0.62**	14.42**	9.59**	0.12**	0.10	0.272	0.025
Firmness (FIRM)	1.85**	136.52*	135.37**	0.28**	0.22	0.262	0.022
Cooking loss (CLOSS)	0.54**	13.48	18.44**	0.21**	0.16	0.116	0.017
Cooked weight (CWT)	1.87**	37.84**	24.67**	0.72**	0.60	0.088	0.011
Balanced data							
Grain protein (WPROT)	0.69**	382.30**	7.05**	0.26**	0.10	0.385	0.043
Semolina protein (SPROT)	0.86**	570.61**	5.66**	0.27**	0.10	0.447	0.042
Kernel vitreousness (VIT)	205.03**	43,076.09**	14,301.21**	72.65**	45.93	0.358	0.040
Sedimentation volume (SDS)	210.49**	9,123.01**	5,242.92**	28.01**	16.62	0.670	0.027
Gluten index (GI)	1,000.44**	405,220.08**	6,920.06**	176.54**	78.81	0.619	0.032
Wet gluten (WG)	18.30**	14,402.79**	235.27**	6.49**	2.16	0.411	0.043
Glutork (GLUT)	0.07**	47.78**	0.61**	0.03**	0.01	0.297	0.056
Color a (Color_a)	0.66**	77.01**	6.09**	0.22**	0.11	0.403	0.041
Color b (Color_b)	11.63**	645.64**	18.65**	1.19**	0.75	0.730	0.023
Color L (Color_L)	2.29**	1,030.62**	13.72**	1.30**	0.75	0.195	0.042
Difference in color a (Dif_a)	0.13**	97.65**	0.38**	0.08*	0.06	0.169	0.038
Difference in color b (Dif_b)	1.90**	219.48**	0.38	0.95**	0.64	0.232	0.041
Difference in color L (Dif_L)	7.45	585.54**	47.02**	7.29	6.60	0.005	0.033
Total yellow pigment (TYP)	3.33**	128.28**	7.16**	0.24**	0.14	0.803	0.018
Polyphenol oxidase activity (PPO)	0.0568**	0.0486**	0.0026	0.0018**	0.0011	0.905	0.009
Firmness (FIRM)	1.98**	135.67*	136.76**	0.28**	0.22	0.456	0.023
Cooking loss (CLOSS)	0.16	52.68**	38.04**	0.18**	0.14	0.103	0.045
Cooked weight (CWT)	0.85	2.79	8.81**	0.85	0.80	0.156	0.044

*,** *Significance at P < 0.05 and 0.01, respectively; ns not significant at P < 0.05; G*L, genotype by location interaction; Rep(L), replicates within location; Error, plots residuals;*
¶*broad sense heritability on plot basis calculated for the RILs.*
#*standard error for heritability.*

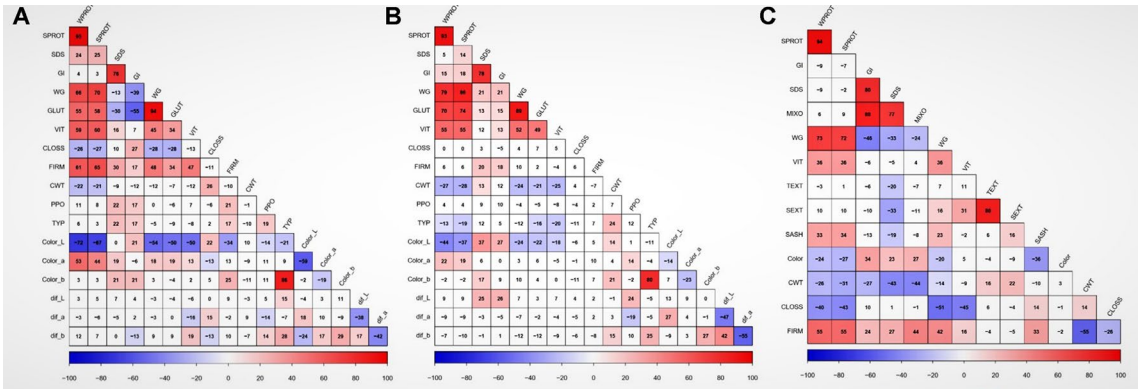

FIGURE 1 | Correlation between traits and locations based on adjusted means. Data showing the relationship between quality traits in **(A)** Prosper, **(B)** Langdon, and **(C)** unbalanced combined data. WPROT, whole-wheat protein; SPROT, semolina protein; SDS, sedimentation test; GI, gluten index; WG, wet gluten; GLUT, glutork; VIT, vitreousness; CLOSS, cooking loss; FIRM, firmness; CWT, cooked weight; PPO, polyphenol oxidase; TYP, total yellow pigment; Color_L, semolina color (Hunter lab value); Color_a, semolina color (Hunter lab value a); Color_b, semolina color (Hunter lab value b); Dif_L, difference in semolina dough sheet color L after 24 h; dif_a, difference in semolina dough sheet color a after 24 h; dif_b, difference in in semolina dough sheet color b after 24 h; Color, spaghetti color; TEXT, total extraction; SEXT, semolina extraction; SASH, semolina ash; MIXO, mixogram score. Cells with correlation values not significant at *P* value < 0.01 have a white background. ## GLUT, WTS, Color_L, Color_a, color_b, Dif_L, Dif_a, Dif_b, PPO, and TYP were only measured in balanced dataset, while TEXT, SEXT, SASH, MIXO, and Color were only measured in historic unbalanced dataset.

findings suggest that both protein quantity and quality/composition play an independent role in the end-use product as has been observed in the past (Ciaffi et al., 1991; Samaan et al., 2006). Overall, the correlation analysis showed that the protein quantity and quality characteristics were associated with the cooking properties.

Marker Properties and Linkage Disequilibrium (LD) Analysis

A total of 6,492 SNP markers showed polymorphism among durum lines (**Supplementary Table 1**). Out of those, 4,196 SNP markers were selected for LD/association mapping (AM) after excluding the markers with MAF <5%, missing data points >10%, and markers with no genetic position on the consensus durum wheat map (Maccaferri et al., 2015). Markers were ordered according to the scaled map positions of the tetraploid wheat SNP consensus map (Maccaferri et al., 2015). The LD decayed to 0.2 within 5 cM, on average (**Figure 2**). Significantly associated SNPs that were ≤5 cM apart and/or located between the pairwise LD (R^2) ≥ 0.7 were considered a single QTL.

Population Structure, Kinship Analysis, and Regression Model Selection for AM

Population structure was inferred using principal component analysis (PCA). Principal component analysis showed that two, three, and 10 PCs explain a cumulative 11.3%, 15.4%, and 26.8% of the genotype variation, respectively. The first three PCs clustered the collection into three sub-populations (**Figure 3**). The familial relatedness was estimated using an identity-by-state matrix (K matrix), and kinship between accessions was calculated (**Figure 4**). Some hotspots with related lines were observed on the heat map, suggesting intermediate familial relationships among genotypes. Accounting for the population structure and familial relationship between individuals in the AM analysis reduces the number of false-positive associations. Based on the BIC values of the five regression models (as explained in the Materials and Methods) tested, no single model fits best for all traits in different environments (**Supplementary Table 2**). For most of the traits, mixed models (KQ) incorporating information about familial relatedness (K matrix) and population structure (Q matrix) were found more suitable (meaning that they have lower BIC values).

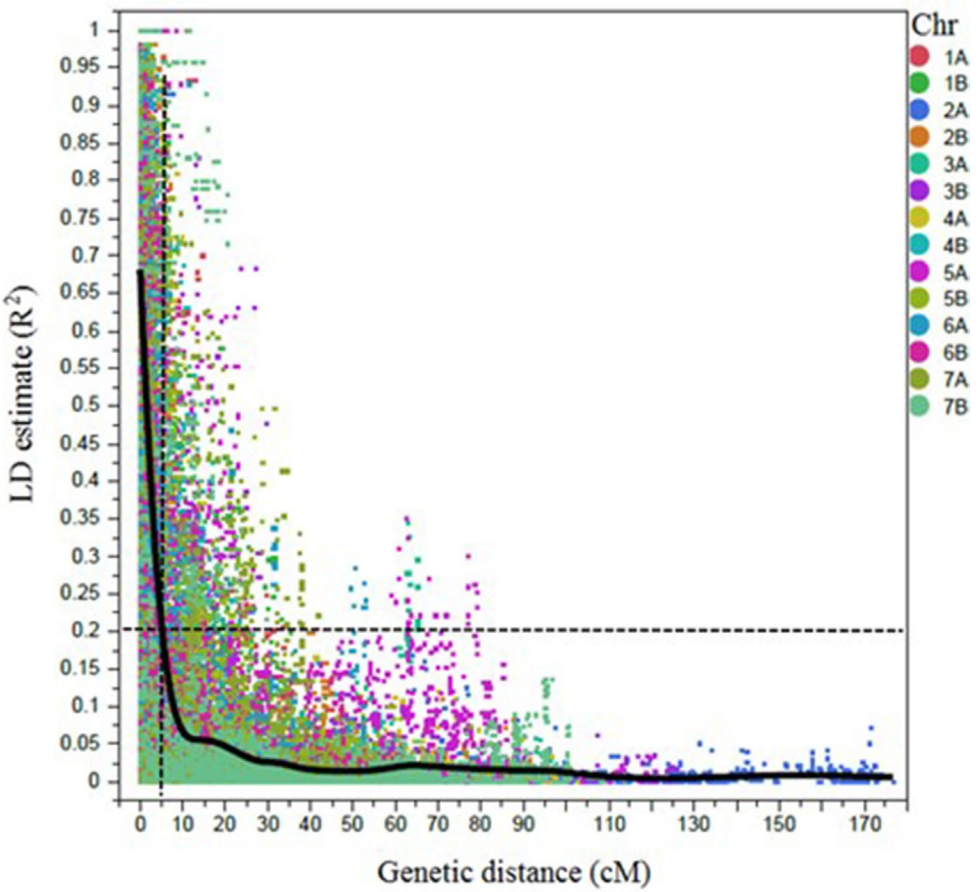

FIGURE 2 | Scatter plot showing the linkage disequilibrium (LD) decay across the chromosomes (Chr) for 243 durum wheat genotypes. The genetic distance in centimorgan (cM) is plotted against the LD estimate (R^2) for pairs of SNPs. Smoothing spline fit, lambda = 338,064.8, R^2 = 0.56902, and sum of squares error = 6,804.7.

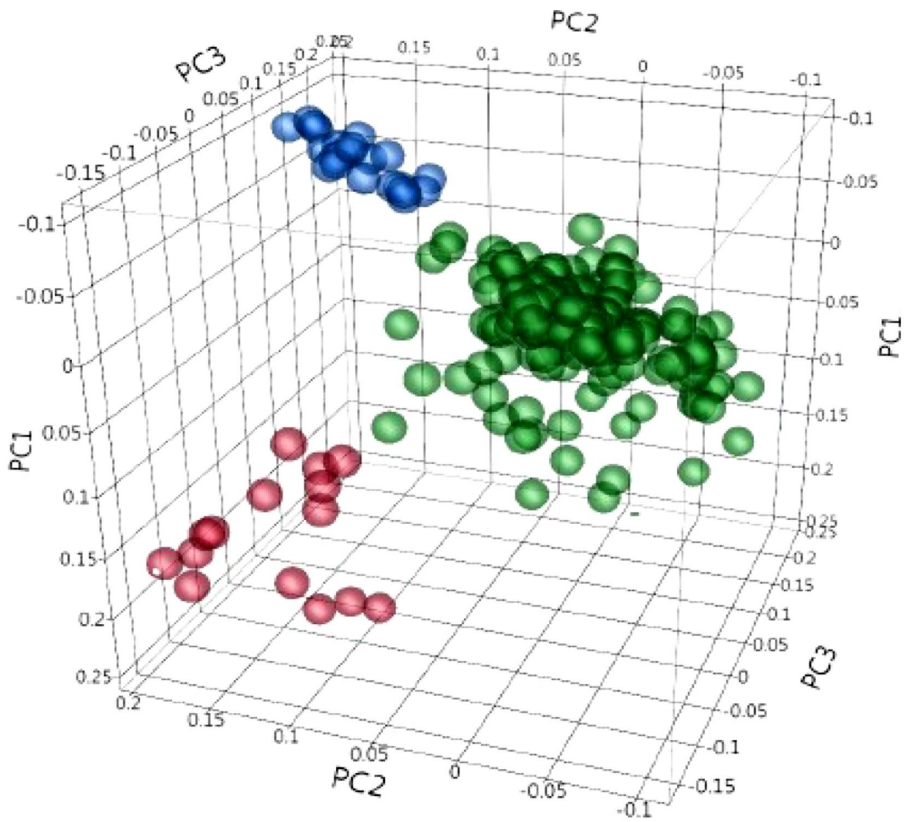

FIGURE 3 | Principal component (PC) analysis obtained from 4,196 polymorphic SNPs, indicating the population structure in 243 durum wheat accessions. PC1, PC2, and PC3 explain 6.8%, 4.5%, and 4.1% of the variation, respectively. The colors represent three different clusters.

In some cases of QK model fitting, three PCAs and in some cases four PCAs were found more appropriate. If BIC values for a particular traits were the same using three and four PCAs, we preferred to use three PCAs in the final GWAS model. For few other traits, adding information about population structure in the model did not help. So to avoid any false negatives because of model overfitting, we analyzed those traits with regression models incorporating only familial relatedness (or K matrix).

Identification of Marker–Trait Associations (MTAs) in Breeding Program Germplasm

Most quality traits were interrelated (**Figure 1**), which was reflected in their significant correlations and co-localization of the associated loci (**Table 2**; **Supplementary Table 3**). The detailed results of marker–trait association are presented in **Supplementary Table 3**, but significant QTLs identified in two or three datasets (except for traits which were analyzed in single environment) are presented in **Table 2**].

Protein content: Both whole grain protein and semolina protein were genetically dissected in this study. Association mapping identified three loci for grain protein and five loci for semolina protein, including two common loci, located one each on 5B and 7A (**Supplementary Table 3**). The QTLs for protein were located in six genomic regions belonging to five different chromosomes (1B, 5B, 6A, 7A, and 7B). All these loci were identified in both balanced and unbalanced datasets. All these loci were associated with minor effects (R^2 = 3.2–5.2%) on protein content.

Milling-related traits: Milling-related traits are represented by total extraction (TEXT), semolina extraction (SEXT), semolina ash (SASH), and kernel vitreousness (VIT). All these traits except VIT were evaluated in the unbalanced dataset only. A total of four, six, eight, and 11 loci were associated with TEXT, SEXT, SASH, and VIT, respectively. These 29 QTL were located in 24 genomic regions belonging to all durum wheat chromosomes except 7B. All the four loci associated with TEXT (located on chromosomes 2A, 2B, 4A, and 5A) were also associated with SEXT (**Supplementary Table 3**) as well, suggesting that those two traits are closely associated. On the other hand, SASH and VIT showed some independent genetic control, as only a single QTL for each of those traits was also found associated with another milling-related trait. However, five out of 11 QTL for VIT were also found associated with one or more color-related traits. Two major QTL (R^2 > 10%) for kernel vitreousness were identified, one each on chromosomes 4B (17.7–22.2 cM) and 5B (146.14–149 cM). This genomic region on 4B was also found associated with Dif_b, CLOSS, Color, Color_a, and WG. Another QTL on 7A was also associated with WPROT, SPROT, and WG. The phenotypic variation contributed by loci associated with milling-related traits varied from 3% to 16.3%.

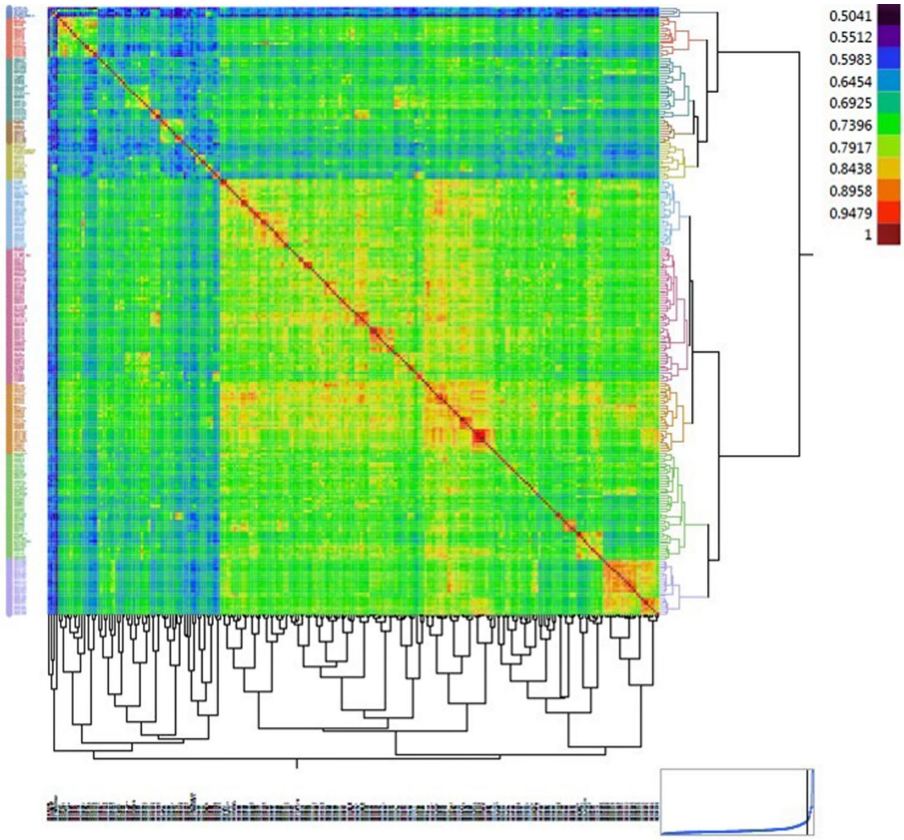

FIGURE 4 | Heat map displaying the relationship matrix among durum wheat genotypes. The red diagonal represents a perfect relationship of each accession with itself. The symmetric off-diagonal elements represent the relationship measures (identity-by-state) for pairs of genotypes. The blocks of warmer colors on the diagonal show clusters of closely related genotypes.

Gluten-related traits: Five gluten-related traits were genetically dissected in this study. These were sedimentation volume (SDS), gluten index (GI), wet gluten (WG), mixogram score (MIXO), and glutork (GLUT). The number of loci identified for these traits ranged from five (MIXO) to 11 (SDS). The 39 QTLs identified for these five gluten-related traits were located in 25 genomic regions, belonging to 12 different chromosomes (all except 5A and 5B) (**Supplementary Table 3**). Two consistently detected loci (explaining 7% to 11% of the phenotypic variation) associated with multiple gluten-related traits were identified in close proximity on the short arm of chromosome 1B. These loci were associated with SDS, GI, and MIXO and were identified in both balanced and unbalanced datasets. These loci were also associated with other quality traits like cooking weight and pasta firmness. The SNP marker IWB70674 associated with the locus on 1B at 15.2 cM was significant and stable across all locations and datasets. The FDR value was low in both balanced and unbalanced datasets. Another important QTL for gluten strength was identified on chromosome 1A and explained about 7.7% of the phenotypic variation. This locus associated with GI, FIRM, and CWT and was found in both balanced and unbalanced datasets. The SNP marker IWB6234 associated with this locus on chromosome 1A could also be useful for MAS. Based on the information for these two markers, two haplotypes were identified for gluten strength measured by SDS, GI, and MIXO (**Table 3**). Few other consistent QTLs associated independently with one of the gluten traits were also identified on other chromosomes as well (**Supplementary Table 3**).

Color-related traits: Nine color-related traits were measured in the present study, including dough color a, color b, color L before and after 24 h, indicated pigment loss (dif_a, dif_b, and dif_L), dried pasta color ("color"), total yellow pigment (TYP) in whole-wheat flour, and polyphenol oxidase (PPO) in whole wheat flour. "Color" represents spaghetti color and was only measured in the unbalanced dataset. A total of 58 QTLs, located on 41 different genomic regions, belonging to all 14 chromosomes were identified for nine color-related traits measured in both datasets. Ten genomic regions harbored QTL for two to four different color-related traits. For individual color-related traits, the number of loci ranged from four (color_L) to 13 (PPO). Five regions, located one each on chromosomes 3B, 4A, 4B, 6A, and 7A, could be considered important as they were associated with three to four colored-related traits. The QTL on 3B (86.4–89.4 cM) was associated with color_a, dif_b, and TYP. The QTL on 4AL was associated with pasta color, color_a, color_b, dif_b, and TYP. The locus on 4B was associated with pasta color, color_a,

TABLE 2 | Genetic location of some important (stable) marker–trait associations (MTAs)* for quality traits in durum wheat. Detailed results of marker–trait association for 24 traits are presented in **Supplementary Table 3**.

Trait/Chr. ##	Position (cM)[a]	Genomic region	Trials[b]	$-\log_{10}$ (*P* value)	R^2¶
a. Protein-related traits					
Grain protein (WPROT)					
5B	204.7–206.1	61	II††, III††	2.86	4.4
7A	59.5–62.5	79	II†, III†	3.02	4.7
7B	62.2–67.3	90	I††, III††	2.31	3.4
Semolina protein (SPROT)					
5B	204.7–206.1	61	I††, III†	3.14	4.9
6A	111.9–113.5	68	I††, II†	3.33	5.1
7A	59.5–59.8	79	I††, III	3.31	5.2
b. Milling-related traits					
Total extraction (TEXT)					
2A	145.9	11	III††	2.23	3.3
2B	158.3–161.5	24	III	5.14	8.5
4A	25.8	41	III†	3.18	5
Semolina extraction (SEXT)					
2A	145.8–145.9	11	III††	2.99	3.6
2B	161.5	24	III††	2.66	4.1
4A	25.2–25.8	41	III†	3.1	4.8
Kernel vitreousness (VIT)					
2B	181.6–183.1	26	I††, III††	2.45	3.6
4B	17.7–22.5	46	I, II	5.63	11.6
7A	59.5–62.5	79	I††, II††	2.84	4.2
c. Gluten-related traits					
Sedimentation volume (SDS)					
1A	1.3–4.6	1	I†, II†, III†	4.73	7.7
1B	0.3–6.1	4	I†, II†, III	5.2	8.7
1B	15.2–15.7	5	I†, II, III	6.11	10.2
3A	79.5	31	I††, II††, III††	2.86	4.3
3B	75.5–79.1	36	II††, III††	2.91	4.4
4A	0	40	I††, II†	3.31	5.1
7B	65.5	90	I†, II††	3.18	4.9
Gluten index (GI)					
1B	0.3–6.1	4	I†, II††, III†	4.3	7.3
1B	15.2–15.7	5	I†, II†, III	6.59	11
3A	170.1–176.9	32	I†, II††	3.49	5.1
3B	75.5–86.9	36	I††,II†, III††	3.15	4.8
6A	67.9–72.4	64	I†, II††, III††	3.47	5.4
7B	169.8–175.9	94	I††, II††, III††	2.67	3.3
Wet gluten (WG)					
2A	186.2–189.8	14	I††, III††	2.86	4.3
2B	146.8	22	II†, III†	3.24	5.3
4B	22.5–28.8	46	II†, III††	3.48	5.4
7A	59.5	79	I††, III††	2.69	4.3
Glutork (GLUT)					
6A	67.9–72.4	64	I††, II††	2.65	3.9
d. Color-related traits					
Color_a					
4A	139.2–143.7	44	I†, II†	4.73	7.6
7A	180.3–181.8	88	I†, II††	3.31	5.1
7B	195.9–196.5	95	I†, II††	4.37	6.9
Color_L					
2A	189.8	14	I††, II††	2.91	4.4
6A	0.1–3.1	62	I††, II††	2.28	3.3
Difference in color b (dif_b)					
4A	159.5	45	I†, II††	3.53	5.5
4B	22.5–26.4	46	I†, II††	3.81	5.1
7B	120.4–123.2	91	I†, II†	3.16	4.8
7B	138.3–140.4	93	I†, II††	3.54	5.5
Difference in color L (dif_L)					
2B	17.7–19	17	I†, II†	4.22	6.7
2B	181.6–183.1	26	I††, II††	2.48	3.7

(*Continued*)

TABLE 2 | Continued

Trait/Chr. ##	Position (cM)[a]	Genomic region	Trials[b]	$-\log_{10}$ (*P* value)	R^2¶
Total yellow pigment (TYP)					
4A	139.2–143.7	44	I††, II††	2.72	4.1
7A	180.3–181.8	88	I†, II††	3.22	5
Polyphenol oxidase activity (PPO)					
1A	6.6	1	I, II†	5.14	8.3
2B	120.2–124.9	21	I†, II	6.67	11
3A	170.1–176.9	32	I, II	9.03	15
3A	183.8–184	33	I, II	7.49	12.3
3B	190.4	38	I, II	9.03	15
3B	198.5–205.1	39	I, II	9.03	15
5A	167.1–167.4	57	I††, II†	3.39	5.2
5B	63.4	59	I†, II	5.95	9.7
6A	105.7	67	I†, II†	3.84	6
6B	27.1	72	I†, II†	5.15	8.3
e. Cooking-related traits					
Firmness (FIRM)					
1B	3–8.5	4	II††, III†	4.93	8.1
3A	7.3–9	27	I††, II††	2.71	4.1
3B	4.2–7.4	34	I†, II†	3.75	5.9
7A	180.3–184.1	88	I††, II††	2.82	4.2
7B	62.2–67.3	90	I††, III††	2.9	4.5
Cooking loss (CLOSS)					
3B	4.2–7.4	34	I††, II†	3.75	5.9
4A	129.3	43	I††, III††	2.45	3.6
6A	127.1–130.0	71	II††, III††	2.56	3.8
Work to shear (WTS)					
2B	153.4	23	I††, II††	2.32	3.4
Cooked weight (CWT)					
1B	3–8.5	4	I†, III†	3.8	6.2
2B	80.6–84	20	I†, III†	4.77	7.8
3A	7.3–9	27	I†, III††	3.85	6
3B	4.2–7.4	34	II, III††	4.3	7
7B	62.2-67.3	90	I††, III††	2.47	3.6

[a]*cM, marker position on the consensus durum map of Maccaferri et al. (2015).*

[b]*I = Prosper balanced trial; II = Langdon balanced trial; III = unbalanced combined dataset where an SNP marker was detected above the pFDR value.*

The specific model to used identify significant MTAs for individual traits are reported in **Supplementary Table 2.*

†*SNP marker that was detected above $-\log_{10}$ (P value) of 3 but below the pFDR value in that trial (environment).*

††*SNP marker that was detected above $-\log_{10}$ (P value) of 2 but below the pFDR value in that trial (environment).*

¶*R^2, proportion of phenotypic variation explained by the individual marker.*

##*GLUT, WTS, Color_L, Color_a, color_b, Dif_L, Dif_a, Dif_b, PPO, and TYP were only measured in the balanced dataset, while TEXT, SEXT, SASH, MIXO, and Color were only measured in the historic dataset.*

pigment loss defined by dif_b, and VIT (**Supplementary Table 3**). Chromosome 6A harbored a genomic region (124–129.4 cM), which was associated with four color-related traits (color_a, color_b, color_L, and dif_b). Another consistent QTL associated with three color-related traits (TYP, col_b, and dif_b) was detected on chromosome 7A at 180–181.8 cM. Co-localized QTLs for TYP and dough color b were identified on chromosomes 2B and 7A. Dough color a, dough color b, pigment loss (measured as difference in dough color b after 24 h), and overall spaghetti color had overlapping QTL on chromosomes 4A, 4B, and 7A.

Markers strongly associated with PPO were identified on chromosomes 1A (IWA5150), 2B (IWA1488), 3A (IWB69399), and 3B (IWB23604) (**Supplementary Table 3**). All SNP markers had low FDR values. The phenotypic variation explained by these color related QTL ranged from 3.1% to 15%.

Pasta cooking-related traits: Pasta cooking parameters, such as cooked firmness (FIRM), cooking loss (CLOSS), work to shear (WTS), and cooked weight (CWT), were scored on fresh pasta in the balanced dataset from both Langdon and Prosper locations. In the unbalanced dataset, the same cooking parameters were taken on dry pasta. A total of 45 QTLs located in 32 genomic regions belonging to 13 different chromosomes (except 5B) were identified for four cooking-related traits (**Supplementary Table 3**). The PVE explained by these QTL ranged from 3% to 14%. Four QTLs, located one each on 1B, 2B, 3A, and 3B, explained >10% PV. An important genomic region for cooking-related traits was identified on the short arm of 3B (4.2–7.4 cM). This region was associated with all four cooking-related traits (CWT, CLOSS, WTS, and FIRM) and was identified in both balanced and unbalanced datasets. The PVE for this 3BS locus ranged from 5.9% to 7.0%. Other important regions that were associated with multiple cooking-related traits and detected in more than one dataset were located on 1B (CWT, FIRM, GI, MIXO, and SDS), 2B (CWT and WTS), 3A (CWT and FIRM), 7A (CWT, WTS, and FIRM), and 7B (CWT, FIRM, SDS, and WPROT). The telomeric region on 1BS, which was also found associated with cooking traits FIRM and CWT, also harbors loci for gluten-related traits (GI, MIXO, and SDS). Ten out of 45 QTL were identified in both

TABLE 3 | Phenotypic means and *t*-test *P* values for lines in the association mapping panel with various combinations of tightly associated markers for gluten strength and PPO activity.

Marker for MAS	Phenotype	Number of genotypes	Phenotypic mean	Range of the phenotypic trait	Actual nucleotide
IWB70626					
Unbalanced	High GI haplotype	179	60.2	22.8–92.5	G
	Low GI haplotype	42	43.9	9.2–69.5	T
t-test			**6.66236E–06**		
Mean PL*					
	High GI haplotype	191	47.6	12.9–82.8	G
	Low GI haplotype	49	36.7	1.04–64.7	T
t-test			**5.44463E–05**		
IWB6234 and IWB70626					
Unbalanced	High GI haplotype	14	69.0	50.9–87.6	GC
	Low GI haplotype	35	44.8	9.2–69.5	TT
			7.45444E–07		
IWB70626					
Unbalanced	High SDS haplotype	187	52.3	39.0–67.4	C
	Low SDS haplotype	47	45.6	25.2–54.6	T
t-test			**1.15538E–12**		
Mean PL	High SDS haplotype	190	55.3	38.8–74.8	C
	Low SDS haplotype	49	50.1	23.5–65.5	T
t-test			**2.50915E–05**		
IWB6234 and IWB70626					
Unbalanced	High SDS haplotype	15	58.4	49.4–65.3	GC
	Low SDS haplotype	39	45.0	25.2–50.9	TT
t-test			**7.08108E–11**		
Mean PL	High SDS haplotype	15	66.3	57.3–74.8	GC
	Low SDS haplotype	39	49.0	23.5–61.8	TT
			5.93077E–10		
IWB70626					
Unbalanced	High Mixo score	187	6.5	4.9–8.3	C
	Low Mixo score	47	5.5	2.5–6.6	T
t-test			**3.29702E–12**		
IWB6234 and IWB70626					
	High Mixo score	15	6.7	5.8–7.5	GC
	Low Mixo score	39	5.4	2.5–6.6	TT
t-test			**2.758E–07**		
IWA5150					
Mean PL	Low PPO	237	0.11	0.04–0.54	T
	All other lines	2	0.41	0.34–0.48	G
t-test			**0.140956287**		
IWA1488					
PL					
	Low PPO	235	0.10	0.04–0.54	A
	All other lines	4	0.43	0.038–0.048	G
t-test			**0.000149301**		
IWB69399					
Mean PL	Low PPO	229	0.10	0.038–0.048	A
	All other lines	10	0.38	0.06–0.54	G
			2.20191E–05		
IWB23604	Low PPO	231	0.10	0.038–0.490	T
Mean PL	All other lines	8	0.422	0.10–0.54	G
			0.000360083		
IWA1488 + IWB69399 + IWB23604					
Mean PL	Low PPO	228	0.95	0.038–0.480	AAT
	All other lines	11	0.43	0.10–0.54	GGG
			7.32631E–07		
IWA5150 + IWA1488 + IWB69399 + IWB23604					
Mean PL	Low PPO	225	0.09	0.04–0.50	TAAT
	All other lines	14	0.40	0.06–0.54	GGGG
t-test			**2.56311E–06**		

**PL, mean of data from Prosper and Langdon locations.*

the balanced and unbalanced datasets. Differences in spaghetti type (dry vs. fresh) during cooking might account for the low number of shared QTLs between the two datasets.

Distribution and co-localization of marker–trait associations (MTAs or QTLs): Co-localized or closely linked QTLs may help in improving several traits simultaneously when desirable alleles for each trait are contributed by the same parent. In this study, a total of 179 MTAs were identified for 24 different quality traits. The number of MTAs per chromosome ranged from 3 to 19. The highest number of MTAs were observed for chromosomes 2B and 6A (19 each), followed by 7A (18) and 1B (17). The lowest number of MTAs was observed for chromosomes 1A and 4B (three each). MTAs for different traits were considered in the same genomic region if they were identified at the same position or very close to each other (<1 cM apart) on the durum consensus map from Maccaferri et al. (2015). Based on this classification, these 179 marker–trait associations were located on 95 genomic regions belonging to all 14 durum wheat chromosomes. Durum A-genome chromosomes harbored 90 MTAs located in 50 genomic regions, whereas B-genome chromosomes showed 89 MTAs that represented 45 genomic regions. Among those 95 genomic regions, 53 regions were associated with only a single MTA or QTL each, while the remaining 42 genomic regions harbored multiple co-localized MTAs (**Supplementary Table 3**). The total MTAs associated with those 42 genomic regions were 126, with an average of three MTAs mapped in a particular genomic region. Individually, those regions harbor two to eight MTAs (genomic regions 36 and 88). This means that a major portion of the MTAs or QTLs for different quality traits was co-localized due to either tight linkage or pleiotropy.

Comparison of GWAS results from balanced and unbalanced datasets: To the best of our knowledge, there is no study on durum wheat that compares results from GWAS based on unbalanced and balanced phenotypic data. In this study, a total of nine traits, including grain protein (WPROT), semolina protein (SPROT), sedimentation volume (SDS), wet gluten (WG), gluten index (GI), kernel vitreousness (VIT), pasta firmness (FIRM), cooking loss (CLOSS), and cooked weight (CWT), were evaluated in both balanced and unbalanced trials. However, because FIRM, CLOSS, and CWT were measured on dry pasta in the historic unbalanced data and on fresh pasta in the balanced dataset, we decided not to include their comparison here. For the remaining six traits that were measured in both balanced and unbalanced datasets, we observed 40 and 31 marker–trait associations (MTAs), respectively (**Table 4**). A total of 26 (84%) of the MTAs identified in unbalanced dataset were also identified using balanced datasets. Interestingly, for the five traits (except VIT), all the MTAs identified using the unbalanced dataset were also identified in the balanced dataset. For kernel vitreousness, among the six and eight MTAs identified in balanced and unbalanced dataset, only three were common (**Table 4**).

Important primary candidates for MAS in durum wheat: Stable and highly significant QTLs are listed in **Table 2**. However, there were some major QTLs identified in this study, which could be the primary target for MAS.

TABLE 4 | Number of marker–trait associations (MTAs) identified for different traits using balanced and unbalanced datasets.

Trait	Balanced data	Unbalanced data	Common
Whole grain protein (WPROT)	3	3	3
Semolina protein (SPROT)	4	3	3
Kernel vitreousness (VIT)	6	8	3
Sedimentation volume (SDS)	11	6	6
Gluten index (GI)	8	6	6
Wet gluten (WG)	8	5	5
Total	**40**	**31**	**26**

For example, markers associated with major QTL for gluten strength on chromosomes 1A and 1B and for PPO activity on chromosomes 1A, 2B, 3A, and 3B could also be excellent primary candidates for MAS in durum wheat breeding programs (**Table 3**). We conducted a *t*-test for markers and alleles significantly associated with increased phenotypic values to determine their possible usefulness for MAS. The lower the *P* value, the more useful the markers (or alleles) and their combination are for MAS. Two haplotypes for gluten strength and five for polyphenol oxidase activity (PPO) were identified (**Table 3**). In the unbalanced and balanced datasets, the two alleles of the marker IWB70626 explained 16.3% and 10.9%, gluten index mean difference. The combination of two markers (IWB70626 and IWB6234) was able to explain the mean difference of 24.8% for the unbalanced dataset. Similarly, for gluten strength measured with SDS, the two haplotypes (IWB6234 and IWB70626) explained 13.4- and 17.3-ml mean difference in SDS. The two alleles for PPO explained 0.35-ppm mean difference for the phenotype.

Discussion

The majority of the earlier studies that aimed to genetically dissect protein content, gluten strength, and other quality traits in durum wheat were based on bi-parental mapping populations and low-density linkage maps (Jernigan et al., 2017; Kumar et al., 2018). Genetic dissection using association mapping populations provides a more detailed understanding of QTL responsible for the particular phenotype (Gupta et al., 2014; Gupta et al., 2019). Further, genome-wide association mapping studies based on advanced breeding lines offer the additional advantage that gained information about important alleles and associated markers can be directly applied into the breeding program. Also, the deployment of high-density marker systems like Illumina's iSelect 90k SNP array in genome-wide studies has the tremendous potential to identify tightly linked markers associated with target traits and map QTL/genes more precisely. This study based on advanced breeding lines and 90k SNP Infinium array was used to dissect a large number of quality traits in durum wheat. Therefore, this study, probably provides the most comprehensive knowledge about the genetic architecture of important durum wheat quality traits. The information gained from this study has direct implications for durum wheat breeding using genomics-based tools.

Genetics of Traits Related to Nutritional Value Enhancement and Grain Quality in Durum Wheat and Identification of some Major Loci

Protein- and dough-related traits: The protein- and dough-related characteristics like grain protein (WPROT), semolina protein (SPROT), gluten index (GI), wet gluten (WG), sedimentation volume (SDS), mixogram score (MIXO), and glutograph (GLUT) are considered the most important parameters of pasta quality and thus represent major breeding targets in the durum cultivar development program. QTLs for grain protein were previously identified on all wheat chromosomes (Kulwal et al., 2005; Bogard et al., 2013; Echeverry-Solarte et al., 2015; Kumar et al., 2018). The limited number of QTL for protein content in this study may be due to the low genetic diversity within the genotypes because of the fixation of high-protein-content QTL alleles. The QTL identified on chromosome 7A was located in the same region, which has been reported in earlier studies (Groos et al., 2003; Prasad et al., 2003; Sun et al., 2009). However, QTLs on chromosomes 5B and 7B were not previously reported in durum wheat and may be novel for high grain protein. Our durum germplasm also lacks the major protein content-associated gene *Gpc-B1* (Uauy et al., 2006), thus offering an opportunity to introgress the functional allele of this major gene to enhance grain protein content. Several examples now exist for the successful introgression of this high GPC gene (*Gpc-B1*) into the adapted germplasm, through MAS (Kumar et al., 2011; for review see, Kumar et al., 2018).

Kernel vitreousness (VIT) is also an important characteristic associated with many grain, flour, semolina, and pasta quality traits, including milling and pasta firmness. A high percentage of vitreous kernels maximizes semolina yield (Dexter et al., 1994). Vitreous areas of the endosperm are known to be higher in protein than mealy ones (Matsuo and Dexter, 1980; Dexter et al., 1994). Protein content showed a positive correlation with kernel vitreousness in this study, as has been reported earlier (Sissons 2004; Bilgin et al., 2010; Sieber et al., 2015). The QTL for grain protein that was identified on chromosome 7A (59.5–62.5 cM) was also associated with vitreousness. This locus could help simultaneous improvement in both the traits.

Previous studies have also shown that the high-molecular-weight glutenin subunits (HMW-GS) are particularly important for determining dough elasticity and correlated positively with dough baking quality (Anjum et al., 2007). Group 1 chromosomes harbor genes for glutenin subunits: HMW-GS loci (Glu-A1, Glu-B1, and Glu-D1) on their long arms (Payne and Lawrence, 1983) and LMW-GS loci (Glu-A3, Glu-B3, and Glu-D3) on their short arms (D'Ovidio and Masci, 2004). Also, the QTLs for gluten strength have been identified on most durum and hexaploid wheat chromosomes; the major and most consistent effect across environments is associated with group 1 chromosomes, particularly 1B (Elouafi et al., 2000; Patil et al., 2009; Conti et al., 2011; Kumar et al., 2013). In this study also, two closely located loci on 1BS (0.3–6.1 and 15.2–15.7 cM) have the most significant effects associated with three gluten-related traits (SDS, GI, and MIXO) across environments. Similarly, the homeologous region on 1AS (1.3–4.6 cM) was also associated with multiple traits across different datasets. The association of these two regions with multiple gluten-related traits was expected because of the high correlations between GI, SDS, and MIXO. The regions on 1AS and 1BS were also associated with cooked weight (CWT) and pasta firmness (FIRM), suggesting the importance of these regions not only for gluten strength but for pasta cooking related quality traits as well. Although QTLs for gluten strength parameters on chromosomes 3B and 7B explained minor phenotypic variation (3–6%), they were consistent across environments and datasets. Conti et al. (2011) also reported QTLs for gluten strength on chromosome 3B. The QTLs on chromosomes 6A and 7B were also earlier reported (Patil et al., 2009; Kumar et al., 2013). In the present study, QTLs for wet gluten and glutograph were identified on chromosomes 1B, 2A, 2B, 6A, and 7A. They shared more common QTLs with semolina protein content than does gluten strength (GI, SDS, and MIXO), suggesting the importance of protein quantity and quality for dough strength and pasta production (D'Egidio et al., 1990; Sissons, 2008).

Milling quality: The aim of the durum grain milling process is to maximize semolina and minimize flour production through successive steps of grinding and sieving (Posner, 2009). The process is complex as it depends on different factors, such as the moisture content of the grain, impurities and broken durum wheat kernels, the size and texture of the grain, and grain protein content (González, 1995). The identification of a large number of QTLs for milling quality parameters in this study confirms their polygenic nature (Sun et al., 2009; Wu et al., 2015). All QTLs associated with TEXT were also associated with SEXT, suggesting that these two traits measure the same process, which is most likely under the same genetic control. It is also possible that these loci have a pleiotropic effect on both traits. Similar results were reported in other studies (Hessler et al., 2002; Russo et al., 2014) and were also expected based on phenotypic correlation analysis (**Figure 1**). The QTL on chromosome 1B could be the same as the QTL reported by Zhang et al. (2008a). The QTLs on chromosomes 2B and 6B were not previously reported and could be novel.

Semolina ash (SASH), another important milling trait, has complex genetics. A decrease in grain weight always results in higher ash content or lower extraction rates (Breseghello and Sorrells, 2007; Brevis et al., 2010). The QTLs for SASH on chromosomes 1B and 6A could be the same QTL reported in a previous study (Zhang et al., 2008a).

Color-related traits: Essentially, all the QTLs for dough color b, a, and total yellow pigment (TYP) had been reported earlier (Parker et al., 1998; Mares and Campbell, 2001; Hessler et al., 2002; Carrera et al., 2007; Pozniak et al., 2007; Zhang and Dubcovsky, 2008b; Garbus et al., 2009; Verlotta et al., 2010; N'Diaye et al., 2017). Major QTLs for yellowness on chromosomes 7A and 7B have been identified earlier in durum wheat (Zhang et al., 2008b). In durum wheat, chromosome 4B is known to harbor *Lpx-B1.1* and *Lpx-B1.2* genes. Previous studies show that deletion of *Lpx-B1.1* is associated with carotenoid pigment degradation during pasta processing (Hessler et al., 2002; Carrera et al., 2007; Garbus et al., 2009; Verlotta et al., 2010). Earlier studies in durum wheat have also reported a linkage between semi-dwarfing gene

Rht-B1b and *Lpx-B1.1* (Peng et al., 1999; N'Diaye et al., 2017). Similar results were observed in our study where we found a QTL associated with pigment loss on the short arm of chromosome 4B (18.4 to 28.8 cM), a region that harbors a reduced height gene. Markers for pigment loss on chromosome 4B did not show an association with dough color b† and total yellow pigment, confirming that *Lpx-B1.1* deletion has an effect on LOX activity during processing, but not on initial semolina or pasta color (Borrelli et al., 1999; Carrera et al., 2007). Like our study, the distal region on chromosomes 7A and 7B has been found to be associated with total yellow pigment in other tetraploid wheat studies as well (Pozniak et al., 2007; Zhang and Dubcovsky, 2008b). QTLs for color b† and TYP on chromosome 7B were in the same vicinity as the previously reported *Phytoene synthase 1* locus (PSY-B1) (Pozniak et al., 2007).

For pasta color a† (green-red chromaticity), detection of four loci suggests complex genetic control of this trait. Considering the association of chromosome 3B, 4A, and 4B regions with color, col_b, and dif_b, as well as the negative correlation between dough colors a† and b† ($r = -0.26$) and pigment loss as measured by colors a† and b† ($r = -0.63$), a genetic linkage between these two traits could be suggested. This study's findings and those of N'Diaye et al. (2017) support the undesirable association between pasta redness and pasta yellowness. Therefore, much effort should focus on breaking the LD to facilitate selecting against redness in dough color. The positive correlation between dif_b and dif_L ($r = 0.31$), as well as the negative correlations between dif_L and dif_a ($r = -0.41$) and dif_b and Dif_a ($r = -0.63$), and a single QTL on chromosome 6A associated with col_b and dif_b, may suggest an indirect masking effect of col_a on col_L by directly influencing col_b, especially in semolina dough over time.

Previously reported QTLs for polyphenol oxidase (PPO) activity were independent from other color-related traits, suggesting that their effect on pasta quality is mostly due to a browning reaction rather than the influence of semolina color components (Zhai et al., 2016). The phenotypic and genetic analyses suggested similar findings in our study. Polyphenol oxidase activity did not show any significant correlations with any other color-related trait, and only two (one each on 1A and 7B) out of 13 genomic regions for PPO showed association with other color-related traits. The major QTL for PPO on chromosome 2B was located in a proximate region compared with that in earlier-reported studies (Beecher et al., 2012; Si et al., 2012). Previously reported major QTLs for PPO on chromosome 2A in tetraploid (Watanabe et al., 2006) and hexaploid wheats (Zhang et al., 2005; He et al., 2007; Wang et al., 2009) could not be identified in the present study. The major QTLs for PPO on 3A and 3B seem to be novel and could be attributed to the different sources of germplasm used in this study.

Cooking-related traits: Many QTLs for four cooking-related traits (firmness, cooking loss, work to shear, and cooked weight) shared common regions, suggesting a close association between those traits. Two major regions associated with cooking-related traits were also associated with gluten strength on chromosome 1B and whole grain protein on chromosome 7B. Zhang and Dubcovsky (2008b) also reported QTL for mixograph peak height and width near the QTL for firmness and cooking loss on chromosome 1B. Independent QTLs for firmness on chromosomes 6A and 7A suggest that parameters other than protein quantity and quality also affect pasta firmness and cooking loss. For instance, genes responsible for amylose synthesis are reported on chromosome 7A (Miura et al., 1999).

Application of Unbalanced Historic Phenotypic Data for Genetic Studies and Molecular Breeding

The collection of phenotypic data requires extensive efforts and resources. Breeding programs collect large amounts of phenotypic data from advanced breeding lines every year for selection purposes. In the yield trial stages, the number of such advanced breeding lines tested each year is relatively small, and some of these lines are replaced by other breeding lines, resulting in an unbalanced dataset of advanced breeding lines developed over time. If these historic unbalanced data could be effectively used for genetic studies in crops, they could save significant amounts of resources and provide useful information for molecular breeding of crops. However, not many studies have been conducted to show the utility of such unbalanced historical data for genetic studies in plants. This could be addressed by comparing the results of genetic analysis using structured balanced data and unbalanced datasets. In barley, Wang et al. (2012) evaluated a set of 384 breeding lines to identify QTL for heading date, a highly heritable trait. The study showed that the unbalanced data could be used to identify the three QTLs that were discovered using balanced dataset. However, a careful consideration of population size and experimental design is needed to reduce false-discovery rate, which was higher in case of unbalanced data. While Wang et al., (2012) studied only a single trait with high heritability, most of the target traits in a breeding program show low-to-moderate heritability and complex genetics. In this study, we included more complex traits, having low (grain protein, semolina protein, and kernel vitreousness) to moderate heritability (sedimentation volume, gluten index, and wet gluten) (**Table 1**). Association mapping identified more MTAs using the balanced data (40 for six traits) compared with the unbalanced data (31 for the same six traits). We observed that about 65% of the MTAs identified by balanced data were also detected by historic unbalanced data, suggesting that either balanced data have slightly higher power in QTL detection or that unbalanced historic data might be associated with false-negative MTAs (Type II error). On the other hand, the majority of the MTAs (84%) identified using historic unbalanced data were also detected using balanced dataset, which indicates that historic unbalanced data did not detect false positives as observed by Wang et al. (2012). Our results clearly demonstrate that historic unbalanced data are suitable for genetic studies of both high and low heritable traits. The fact that no false-positive QTLs were detected using unbalanced data for complex and low heritability traits like protein content offers greater prospects of using historic unbalanced data from the breeding program to generate information for molecular breeding of both simple and complex traits. A couple of recent studies have also shown that historical data could be useful for genomic selection as well (Dawson et al., 2013; Rutkoski et al., 2015).

In the last few years, genotyping has become inexpensive, and a large amount of genotypic data can now be generated quickly. Also, an annotated reference sequence of whole wheat genome is now available. In this scenario, our ability to genetically dissect complex traits using routinely collected phenotypic data by the breeders will be extremely promising. This could help us save resources spent on genetic studies and may enable us to speed up genetic gain through molecular breeding tools.

HIGHLIGHTS

Multi-environment phenotypic data and high-density SNP platform were used to identify markers associated with 24 nutritional value enhancement and quality traits for genomics-assisted durum wheat breeding. The study also showed the application of historic unbalanced phenotypic data for genetic studies.

CONCLUSION

The study aimed to dissect the genetics of durum wheat quality and identify useful marker–trait association for 24 different traits using high-density 90k Infinium SNP marker data. Genome-wide association studies revealed that MTAs for durum quality traits are distributed across the whole genome. Markers associated with some major QTL for gluten strength on chromosomes 1A and 1B and PPO activity on chromosomes 1A, 2B, 3A, and 3B could also be excellent easy candidates for MAS in durum wheat breeding programs. The information gained on extensive genetic dissection of durum wheat quality traits and the resources developed in this study may prove extremely useful to assess quality in early generations by incorporating molecular breeding tools in the breeding program. Another main objective of this study was to explore the possibilities of using a large amount of multi-year, multi-location unbalanced historical data generated by the breeding programs for genetic studies. A large number of common SNPs detected in both the unbalanced historic and balanced datasets suggest that the unbalanced data collected by the plant breeding programs over space and time could be used to gain knowledge about the genetics of important traits and identify MTAs for molecular breeding. This would save huge amounts of resources invested on conducting genetic studies using specifically designed populations. Also, as this study was based on advanced breeding lines, the MTAs identified in this study are easily accessible and should provide more directly useful information for local-knowledge-based breeding.

AUTHOR CONTRIBUTIONS

MJ, FM, and EE conceived and designed the experiments. MJ, AO-A, and ES performed the experiment. MJ, AK, and MA analyzed the data. MJ and AK wrote the manuscript with inputs from all the co-authors.

ACKNOWLEDGMENTS

The authors would like to thank North Dakota State University's durum wheat breeding project members Stan Stancyk, Sally Mann, and Megan Shawgo for their assistance in field research process; Mary Osenga of USDA-ARS, Small Grains Genotyping Laboratory, Fargo, ND for genotyping; and Yu (Gaby) Liu and Hiroshi Ando of NDSU durum wheat quality lab Fargo, ND, for helping to analyze "countless" numbers of grain and semolina samples. Thanks are also due to North Dakota Wheat Commission for funding this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00717/full#supplementary-material

REFERENCES

AACC International (2008). *Approved methods of the AACCI*. 11th edn. St. Paul: The association.

Anderson, J. V., and Morris, C. F. (2001). An improved whole-seed assay for screening wheat germplasm for polyphenol oxidase activity. *Crop Sci.* 41, 1697–1705. doi: 10.2135/cropsci2001.1697

Anderson, J. V., and Morris, C. F. (2003). Purification and analysis of wheat grain polyphenol oxidase (PPO) protein. *Cereal Chem.* 80, 135–143. doi: 10.1094/CCHEM.2003.80.2.135

Anjum, F. M., Khan, M. R., Din, A., Saeed, M., Pasha, I., and Arshad, M. U. (2007). Wheat gluten: high molecular weight glutenin subunits-structure, genetics, and relation to dough elasticity. *J. Food Sci.* 72, 56–63. doi: 10.1111/j.1750-3841.2007.00292.x

Beecher, B. S., Carter, A. H., and See, D. R. (2012). Genetic mapping of new seed-expressed polyphenol oxidase genes in wheat (*Triticum aestivum* L). *Theor. Appl. Genet.* 124 (8), 1463–1473. doi: 10.1007/s00122-012-1801-2

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188. doi: 10.1214/aos/1013699998

Bilgin, O., Korkut, K. Z., Baser, I., Daglioglu, O., Öztürk, I., Balkan, A., et al. (2010). Variation and heritability for some semolina characteristics and grain yield and their relations in durum wheat (*Triticum durum* Desf). *World J. Agric. Sci.* 6, 301–308.

Blanco, A., Mangini, G., Giancaspro, A., Giove, S., Colasuonno, P., Simeone, R., et al. (2012). Relationships between grain protein content and grain yield components through quantitative trait locus analyses in a recombinant inbred line population derived from two elite durum wheat cultivars. *Mol. Breed.* 30, 79–92. doi: 10.1007/s11032-011-9600-z

Blanco, A., Simeone, R., and Gadaleta, A. (2006). Detection of QTLs for grain protein content in durum wheat. *Theor. Appl. Genet.* 112, 1195–1204. doi: 10.1007/s00122-006-0221-6

Bogard, M., Allard, V., Martre, P., Heumez, E., Snape, J. W., Orford, S., et al. (2013). Identifying wheat genomic regions for improving grain protein concentration independently of grain yield using multiple inter-related populations. *Mol. Breed.* 31, 587–599. doi: 10.1007/s11032-012-9817-5

Borrelli, G. M., Troccoli, A., Di Fonzo, N., and Fares, C. (1999). Durum wheat lipoxigenase activity and other quality parameters that affect pasta colour. *Cereal Chem.* 76, 335–340. doi: 10.1094/CCHEM.1999.76.3.335

Breseghello, F., and Sorrells, M. E. (2007). QTL analysis of kernel size and shape in two hexaploid wheat mapping populations. *Field Crops Res.* 101, 172–179. doi: 10.1016/j.fcr.2006.11.008

Brevis, H. C., Morris, C., Manthey, F., and Dubkovsky, J. (2010). Effect of the grain protein content locus Gpc-B1 on bread and pasta quality. *J. Cereal Sci.* 51, 357–365. doi: 10.1016/j.jcs.2010.02.004

Carrera, A., Echenique, V., Zhang, W., Helguera, M., Manthey, F., Schrager, A., et al. (2007). A deletion at the Lpx-B1 locus is associated with low lipoxygenase activity and improved pasta color in durum wheat (*Triticum turgidum* ssp durum). *J. Cereal Sci.* 45, 67–77. doi: 10.1016/j.jcs.2006.07.001

Ciaffi, M., Benedettelli, S., Giorgi, B., Porceddu, E., and Lafiandra, D. (1991). Seed storage proteins of *Triticum turgidum* ssp *dicoccoides* and their effect on the technological quality in durum. *Wheat Plant Breed.* 107, 309–319. doi: 10.1111/j.1439-0523.1991.tb00554.x

Conti, V., Roncallo, P. F., Beaufort, V., Cervigni, G. L., Miranda, R., Jensen, C. A., et al. (2011). Mapping of main and epistatic effect QTLs associated to grain protein and gluten strength using a RIL population of durum wheat. *J. Appl. Genet.* 52, 287–298. doi: 10.1007/s13353-011-0045-1

D'Egidio, M. G., Mariani, B. M., Nardi, S., Novaro, P., and Cubadda, R. (1990). Chemical and technological variables and their relationships: a predictive equation for pasta cooking quality. *Cereal Chem.* 67, 275–281. doi: 10.12691/wjar-2-1-3

D'Ovidio, R., and Masci, S. (2004). The low-molecular-weight glutenin subunits of wheat gluten. *Cereal Sci.* 39, 321–339. doi: 10.1016/j.jcs.2003.12.002

Dawson, J. C., Endelman, J. B., Heslot, N., Crossa, J., Poland, J., and Dreisigacker, S. (2013). The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Res.* 154, 12–22. doi: 10.1016/j.fcr.2013.07.020

Debbouz, A. (1994). What's new in pasta color. *Pasta J.* 76, 31–32.

Dexter, J. E., and Matsuo, R. R. (1977). Influence of protein content on some durum wheat quality parameters. *Can. J. Plant Sci.* 57, 717–727. doi: 10.4141/cjps77-105

Dexter, J. E., Preston, K. R., Martin, D. G., and Gander, E. J. (1994). The effects of protein content and starch damage on the physical dough properties and bread-making quality of Canadian durum wheat. *J. Cereal Sci.* 18, 139–151. doi: 10.1006/jcrs.1994.1054

Dick, J. W., and Quick, J. S. (1983). A modified screening test for rapid estimation of gluten strength in early-generation durum wheat breeding lines. *Cereal Chem.* 60, 315–318.

Echeverry-Solarte, M., Kumar, A., Kianian, S., Simsek, S., Alamri, M. S., Mantovani, E. E., et al. (2015). New QTL alleles for quality-related traits in spring wheat revealed by RIL population derived from supernumerary × non-supernumerary spikelet genotypes. *Theor. Appl. Genet.* 128, 893–912. doi: 10.1007/s00122-015-2478-0

Elouafi, I., Nachit, M. M., Elsaleh, A., Asbati, A., and Mather, D. E. (2000). QTL-mapping of genomics regions controlling gluten strength in durum (*Triticum turgidum* L var *durum*). *Options Mediterr.* 40, 505–509.

Fiedler, J. D., Salsman, E., Liu, Y., Michalak de Jiménez, M., Hegstad, J. B., Chen, B., et al. (2017). Genome-wide association and prediction of grain and semolina quality traits in durum wheat breeding populations. *Plant Genome* 10 (3), 1–12. doi: 10.3835/plantgenome2017.05.0038

Fu, B., Schlichting, L., Pozniak, C. J., and Singh, A. K. (2011). A fast, simple, and reliable method to predict pasta yellowness. *Cereal Chem.* 88, 264–270. doi: 10.1094/CCHEM-12-10-0173

Garbus, I., Carrera, A. D., Dubcovsky, J., and Echenique, V. (2009). Physical mapping of durum wheat lipoxygenase genes. *Cereal Sci.* 50, 67–73. doi: 10.1016/j.jcs.2009.02.008

Ghosh, J. K., Delampady, M., and Samanta, T., (2006). *An introduction to Bayesian analysis: theory and methods*. New York: Springer-Verlag.

González, A. T. J. (1995). "Milling process of durum wheat," in *Durum wheat quality in the Mediterranean region*. Eds. N. Di Fonzo, F. Kaan, and M. Nachit (Zaragoza, Spain: CIHEAM), 43–51.

Groos, C., Robert, N., Bervas, E., and Charmet, G. (2003). Genetic analysis of grain protein content, grain yield and thousand-kernel weight in bread wheat. *Theor. Appl. Genet.* 106, 1032–1040. doi: 10.1007/s00122-002-1111-1

Gupta, P. K., Kulwal, P. L., and Jaiswal, V. (2014). Association mapping in crop plants: opportunities and challenges. *Adv. Genet.* 38, 109–147. doi: 10.1016/B978-0-12-800271-1.00002-0

Gupta, P. K., Kulwal, P. L., and Jaiswal, V. (2019). Association mapping in plants in the post-GWAS genomics era. *Adv. Genet*. 104, 75–154. doi: 10.1016/bs.adgen.2018.12.001

He, X. Y., He, Z. H., Zhang, L. P., Sun, D. J., Morris, C. F., Fuerst, E. P., et al. (2007). Allelic variation of polyphenol oxidase (PPO) genes located on chromosomes 2A and 2D and development of functional markers for the PPO genes in common wheat. *Theor. Appl. Genet*. 115, 47–58. doi: 10.1007/s00122-007-0539-8

Hessler, T. G., Thomson, M. J., Benscher, D., Nachit, M. M., and Sorrells, M. E. (2002). Association of a lipoxygenase locus, Lpx-B1, with variation in lipoxygenase activity in durum wheat seeds. *Crop Sci.* 42, 1695–1700. doi: 10.2135/cropsci2002.1695

Holland, J. B. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant Breed. Rev.* 22, 4–54.

Hoseney, R. C. (1986). "Structure of cereals," in *Principles of cereal sciences and technology*. Ed. R. C. Hoseney (St Paul, MN: American Association of Cereal Chemistry), 1–33.

Hoseney, R. C. (1987). Wheat hardness. *Cereal Foods World* 32, 320–322.

Jernigan, K. L., Morris, C. F., Zemetra, R., Chen, J., Garland-Campbell, K., and Carter, A. (2017). Genetic analysis of soft white wheat end-use quality traits in a club by common wheat cross. *J. Cereal Sci.* 76, 148–156. doi: 10.1016/j.jcs.2017.06.005

Kulwal, P., Kumar, N., Kumar, A., Gupta, R., Balyan, H., and Gupta, P. (2005). Gene networks in hexaploid wheat: interacting quantitative trait loci for grain protein content. *Integr. Genomics* 5, 254–259. doi: 10.1007/s10142-005-0136-3

Kumar, A., Elias, E. M., Gavami, F., Xu, X., Jain, S., Manthey, F. A., et al. (2013). A major QTL for gluten strength in durum wheat (*Triticum turgidum* L var *durum*). *J. Cereal Sci.* 57, 21–29. doi: 10.1016/j.jcs.2012.09.006

Kumar, A., Jain, S., Elias, E. M., Ibrahim, M., and Sharma, L. K., (2018). "An overview of QTL identification and marker-assisted selection for grain protein content in wheat," in *Eco-friendly agro-biological techniques for enhancing crop productivity*. Eds. R. Sengar and A. Singh (Singapore: Springer), 245–274. doi: 10.1007/978-981-10-6934-5_11

Kumar, J., Jaiswal, V., Kumar, A., Kumar, N., et al. (2011). Introgression of a major gene for high grain protein content in some Indian bread wheat cultivars. *Field Crops Res.* 123, 226–233. doi: 10.1016/j.fcr.2011.05.013

Li, F., Cui, J., Ding, A. M., Zhao, C. H., Wang, X.-Q., Wang, L., et al. (2012). QTL detection of seven quality traits in wheat using two related recombinant inbred line population. *Euphytica* 183, 207–226. doi: 10.1007/s10681-011-0448-4

Maccaferri, M., Ricci, A., Salvi, S., Milner, S. G., Noli, E., Martelli, P. L., et al. (2015). A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. *Plant Biotechnol. J.* 13, 648–663. doi: 10.1111/pbi.12288

Mares, D. J., and Campbell, A. W. (2001). Mapping components of flour and noodle colour in Australian wheat. *Aust. J. Agric. Res.* 52, 1297–1309. doi: 10.1071/AR01048

Matsuo, R. R., and Dexter, J. E. (1980). Comparison of experimentally milled durum wheat semolina to semolina produced by some Canadian mills. *Cereal Chem.* 57, 117–122.

McCartney, C. A., Somers, D. J., Humphreys, D. G., Lukow, O., Ames, N., Noll, J., et al. (2005). Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452 × 'AC Domain'. *Genome* 48, 870–883. doi: 10.1139/g05-055

Merida-Garcıa, R., Liu, G., He, S., Gonzalez-Dugo, V., Dorado, G., Galvez, S., et al. (2019). Genetic dissection of agronomic and quality traits based on association mapping and genomic selection approaches in durum wheat grown in Southern Spain. *PLoS ONE* 14 (2), e0211718. doi: 10.1371/journal.pone.0211718

Miller, B. S., Afeork, S., Pomeranz, Y., Bruinsma, B. L., and Booth, G. D. (1982). Measuring hardness of wheat. *Cereal Foods World* 27, 61–64.

Miura, H., Araki, E., and Tarui, S. (1999). Amylose synthesis capacity of the three Wx genes of wheat cv Chinese Spring. *Euphytica* 108, 91–95. doi: 10.1023/A:1003604328227

N'Diaye, A., Haile, J. K., Cory, A. T., Clarke, F. R., Clarke, J. M., Knox, R. E., et al. (2017). Single marker and haplotype-based association analysis of semolina and pasta colour in elite durum wheat breeding lines using a high-density consensus map. *PLoS One* 12, e0170941. doi: 10.1371/journal.pone.0170941

NASS. (2018). National Agricultural Statistics Service. https://www.nass.usda.gov. Access 5 August 2018.

Parker, G. D., Chalmers, K. J., Rathjen, A. J., and Langridge, P. (1998). Mapping loci associated with flour colour in wheat (*Triticum aestivum* L). *Theor. Appl. Genet.* 97, 238–245. doi: 10.1007/s001220050891

Patil, R. M., Oak, M. D., Tamhankar, S. A., and Rao, V. S. (2009). Molecular mapping of QTLs for gluten strength as measured by sedimentation volume and mixograph in durum wheat (*Triticum turgidum* L ssp durum). *Cereal Sci.* 49, 378–386. doi: 10.1016/j.jcs.2009.01.001

Payne, P. I., and Lawrence, G. J. (1983). Catalogue of alleles for the complex loci, Glu-A1, Glu-B1 and Glu-D1 which code for high-molecular-weight subunits of glutenin in hexaploid wheat. *Cereal Res. Comm.* 11, 29–35.

Peng, Z. S., Su, Z. X., and Cheng, K. C. (1999). Characterization of dwarf trait in the tetraptoid wheat landrace. *Aiganfanma Wheat Inf. Serv.* 89, 7–12.

Posner, E. S. (2009). "Wheat flour milling," in *Wheat: chemistry and technology*. Eds. K. Khan and P. R. Shewry (St Paul, MN: American Association of Cereal Chemistry), 119–152. doi: 10.1094/9781891127557.005

Pozniak, C. J., Knox, R. E., Clarke, F. R., and Clarke, J. M. (2007). Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. *Theor. Appl. Genet.* 114, 525–537. doi: 10.1007/s00122-006-0453-5

Prasad, M., Kumar, N., Kulwal, P. L., Roder, M. S., Balyan, H. S., Dhaliwal, H. S., et al. (2003). QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. *Theor. Appl. Genet.* 106, 659–667. doi: 10.1007/s00122-002-1114-y

Roncallo, P. F., Cervigni, G. L., Jensen, C., Miranda, R., Carrera, A. D., Helguera, M., et al. (2012). QTL analysis of main and epistatic effects for flour colour traits in durum wheat. *Euphytica* 185, 77–92. doi: 10.1007/s10681-012-0628-x

Russo, M. A., Ficco, D. B. M., Laidò, G., Marone, D., Papa, R., and Blanco, A. (2014). A dense durum wheat × *T. dicoccum* linkage map based on SNP markers for the study of seed morphology. *Mol. Breed.* 34, 1579–1597. doi: 10.1007/s11032-014-0181-5

Rutkoski, J., Singh, R. P., Huerta Espino, J., Bhavani, S., Poland, J., Jannink, J. L., et al. (2015). Efficient use of historical data for genomic selection: a case study of stem rust resistance in wheat. *Plant Genome J.* 8, 1–10. doi: 10.3835/plantgenome2014.09.0046

Samaan, J., El-Khayat, G. H., Manthey, F. A., Fuller, M. P., and Brennan, C. S. (2006). Durum wheat quality: II. The relationship of kernel physicochemical composition to semolina quality and end product utilization. *Int. J. Food Sci. Technol.* 41, 47–55. doi: 10.1111/j.1365-2621.2006.01313.x

SAS Institute Inc. (2004). *SAS/ETS 9.1 User's Guide*. Cary, NC: SAS Institute Inc.

Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi: 10.1086/502802

Si, H. Q., Zhou, Z. L., Wang, X. B., and Ma, C. X. (2012). A novel molecular marker for the polyphenol oxidase gene located on chromosome 2B in common wheat. *Mol. Breed.* 30, 1371–1378. doi: 10.1007/s11032-012-9723-x

Sieber, A.-N., Wurschum, T., and Longin, C. F. H. (2015). Vitreosity, its stability and relationship to protein content in durum wheat. *Cereal Sci.* 61, 71–77. doi: 10.1016/j.jcs.2014.10.008

Sims, R. P. A., and Lepage, M. A. (1968). A basis for measuring the intensity of wheat flour pigments. *Cereal Chem.* 45, 605–611.

Sissons, M. (2004). "Pasta," in *Encyclopedia of grain science*. Eds. C. Wrigley, H. Corke, and C. Walker (London: Elsevier Academic Press), 409–418. doi: 10.1016/B0-12-765490-9/00123-3

Sissons, M. (2008). Role of durum wheat composition on the quality of pasta and bread. *Food* 2, 75–90.

Steel, R. G. D., and Torrie, J. H. (1980). *Principles and procedures of statistics*. 2nd ed. New York: McGraw-Hill.

Sun, X.-Y., Wu, K., Zhao, Y., Kong F.-M., Han, G.-Z., Huang X.-J., et al. (2009). QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica* 165, 615–624. doi: 10.1007/s10681-008-9794-2

Tadesse, W., Ogbonnaya, F. C., Jighly, A., Sanchez-Garcia, M., Sohail, Q., Rajaram, S., et al. (2015). Genome-wide association mapping of yield and grain quality traits in winter wheat genotypes. *PLoS One* 10, e0141339. doi: 10.1371/journal.pone.0141339

Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., and Dubcovsky, J. (2006). A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* 314, 1298–1301. doi: 10.1126/science.1133649

Venabes, W. N., Smith, D. M., R Development Core Team. (2017). An introduction to R, a programming environment for data analysis and graphics, version 3.6.1. The R Foundation for Statistical Computing. Available at: http://cran.r-project.org/doc/manuals/R-intro.pdf. Accessed Sept. 21, 2017.

Verlotta, A., Simone, V. D., Mastrangelo, A. M., Cattivelli, L., Papa, R., and Trono, D. (2010). Insight into durum wheat Lpx-B1: a small gene family coding for the lipoxygenase responsible for carotenoid bleaching in mature grains. *BMC Plant Biol.* 10, 263–269. doi: 10.1186/1471-2229-10-263

Walsh, D. E. (1971). Measuring spaghetti firmness. *Cereal Sci. Today* 16, 202.

Wang, H., Smith, K. P., Combs, E., Blake, T., Horsley, R. D., and Muelbauer, G. J. (2012). Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. *Theor. Appl. Genet.* 124, 111–124. doi: 10.1007/s00122-011-1691-8

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 00 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183

Wang, X. B., Ma, C. X., Si, H. Q., Qiao, Y., Chang, C., He, X. F., et al. (2009). Gene markers for grain polyphenol oxidase activity in common wheat. *Mol. Breed.* 23, 163–170. doi: 10.1007/s11032-008-9223-1

Watanabe, N., Akond, A. S., and Nachit, M. M. (2006). Genetic mapping of the gene affecting polyphenol oxidase activity in tetraploid durum wheat. *J. Appl. Genet.* 47, 201–205. doi: 10.1007/BF03194624

Wu, Q. H., Chen, Y. X., Zhou, S. H., Fu, L., and Chen, J. J. (2015). High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817 × Beinong6. *PLoS One* 10, e0118144. doi: 10.1371/journal.pone.0118144

Yue, P., Rayas-Duarte, P., and Elias, E. (1999). Effect of drying temperature on physicochemical properties of starch isolated from pasta. *Cereal Chem.* 76, 541–547. doi: 10.1094/CCHEM.1999.76.4.541

Zhai, S., He, Z., Wen, W., Jin, H., Liu, J., Zhang, Y., et al. (2016). Genome-wide linkage mapping of flour color-related traits and polyphenol oxidase activity in common wheat. *Theor. Appl. Genet.* 129, 377–394. doi: 10.1007/s00122-015-2634-6

Zhang, L. P., Ge, X. X., He, Z. H., Wang, D. S., Yan, J., Xia, X. C., et al. (2005). Mapping QTLs for polyphenol oxidase activity in a DH population from common wheat. *Acta Agron. Sin.* 31, 7–10. doi: 10.1007/s1032–005–6618–0

Zhang, W., Chao, S., Manthey, F., Chicaiza, O., Brevis, J. C., Echenique, W., et al. (2008a). QTL analysis of pasta quality using a composite microsatellite and SNP map of durum wheat. *Theor. Appl. Genet.* 117, 1361–1377. doi: 10.1007/s00122-008-0869-1

Zhang, W., and Dubcovsky, J. (2008b). Association between allelic variation at the Phytoene synthase 1 gene and yellow pigment content in the wheat grain. *Theor. Appl. Genet.* 116, 635–645. doi: 10.1007/s00122-007-0697-8

Zhang, Z., Ersoz, E., Lai, C., Todhunter, R., Tiwari, H., Gore, M., et al., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., et al. (2007). An Arabidopsis example of association mapping in structured samples. *PLoS Genet.* 3, e4. doi: 10.1371/journal.pgen.0030004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Johnson, Kumar, Oladzad-Abbasabadi, Salsman, Aoun, Manthey and Elias. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Nutrition

ORIGINAL RESEARCH
published: 30 August 2019
doi: 10.3389/fnut.2019.00122

Down-Regulation of *CYP79A1* Gene Through Antisense Approach Reduced the Cyanogenic Glycoside Dhurrin in [*Sorghum bicolor* (L.) Moench] to Improve Fodder Quality

Arun K. Pandey[1,2]*, *Pusuluri Madhu*[1,2] *and Basrur Venkatesh Bhat*[1]*

[1] ICAR-Indian Institute of Millets Research (IIMR), Hyderabad, India, [2] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University,
United States

Reviewed by:
Cecilia Karstin Blomstedt,
Monash University, Australia
Shyam Solanki,
North Dakota State University,
United States

***Correspondence:**
Arun K. Pandey
pandeyarun80@gmail.com
Basrur Venkatesh Bhat
bhatv@millets.res.in

Specialty section:
This article was submitted to Nutrition and Food Science Technology, a section of the journal Frontiers in Nutrition

Received: 11 April 2019
Accepted: 24 July 2019
Published: 30 August 2019

Citation:
Pandey AK, Madhu P and Bhat BV (2019) Down-Regulation of CYP79A1 Gene Through Antisense Approach Reduced the Cyanogenic Glycoside Dhurrin in [Sorghum bicolor (L.) Moench] to Improve Fodder Quality. Front. Nutr. 6:122. doi: 10.3389/fnut.2019.00122

A major limitation for the utilization of sorghum forage is the production of the cyanogenic glycoside dhurrin in its leaves and stem that may cause the death of cattle feeding on it at the pre-flowering stage. Therefore, we attempted to develop transgenic sorghum plants with reduced levels of hydrogen cyanide (HCN) by antisense mediated down-regulation of the expression of cytochrome P450 CYP79A1, the key enzyme of the dhurrin biosynthesis pathway. *CYP79A1* cDNA was isolated and cloned in antisense orientation, driven by rice *Act1* promoter. Shoot meristem explants of sorghum cultivar CSV 15 were transformed by the particle bombardment method and 27 transgenics showing the integration of transgene were developed. The biochemical assay for HCN in the transgenic sorghum plants confirmed significantly reduced HCN levels in transgenic plants and their progenies. The HCN content in the transgenics varied from 5.1 to 149.8 μg/g compared to 192.08 μg/g in the non-transformed control on dry weight basis. Progenies with reduced HCN content were advanced after each generation till T_3. In T_3 generation, progenies of two promising events were tested which produced highly reduced levels of HCN (mean of 62.9 and 76.2 μg/g, against the control mean of 221.4 μg/g). The reduction in the HCN levels of transgenics confirmed the usefulness of this approach for reducing HCN levels in forage sorghum plants. The study effectively demonstrated that the antisense *CYP79A1* gene deployment was effective in producing sorghum plants with lower HCN content which are safer for cattle to feed on.

Keywords: dhurrin, HCN, CYP79A1, antisense, *Sorghum bicolor*, transgenics, forage quality

INTRODUCTION

The sorghum crop [*Sorghum bicolor* (L.) Moench] is grown worldwide in 45.38 million hectares and accounts for the production of 6.37 million tons of grains (1). Nearly 80% of the cultivated area lies in Asia and Africa. It is the fifth major cereal crop in the world, mainly cultivated in semi-arid regions. Sorghum is an important grain and forage crop of semi-arid regions due to its high adaptability and suitability to rain-fed low input agriculture. One of the major factors

limiting the utilization of sorghum forage, especially till the flowering stage, is the production cyanogenic glycoside, i.e., dhurrin, which is toxic to the feeding livestock (2). Hydrocyanic acid (HCN) toxicity due to hydrolysis of dhurrin from sorghum forage in the rumen of the cattle causes their death. Leaves and stems of all sorghum species contain dhurrin. Other plants also produce HCN, but in lesser amounts, whereas in sorghum it is produced in larger amount which is hazardous to the animal species (3).

Sorghum seedling synthesizes the cyanogenic glycoside dhurrin (β-D-glucopyrinosyloxy-(S)-p-hydroxy-mandelontrile) using L-tyrosine as the precursor amino acid (4). Tyrosine is converted to p-hydroxyphenyl acetaldoxime by two multifunctional cytochrome P450s (CYP79A1, CYP71E1) each encoded by a single gene (5). CYP79A1 catalyes two consecutive N-hydroxylation reactions, followed by a dehydration and decarboxylation reaction (6) where conversion of tyrosine to p-hydroxyphenyl-acetaldoxime occurs (5, 6). The oxime is then converted by CYP71E1 to the aglycon p-hydroxymandelonitrile (5, 7) which is subsequently acted upon by a soluble UDPG-glycosyl transferase to produce dhurrin (5, 8). Sorghum mutants were identified through chemical mutagenesis which are deficient in the production of dhurrin and cyanide release (9). One of these mutations t*cd1* had a lowered activity of CYP79A1, the first enzyme in dhurrin production. This behaved like a recessive allele and followed a Mendelian pattern of inheritance. Though cyanogenesis often appears to follows classical Mendelian inheritance, its quantitative expression at a particular point in space and time is highly plastic (10). The Caudatum group of sorghums exhibited the highest and the Guinea group showed the least dhurrin content in the sorghum leaf (11). They suggested that genes for both biosynthesis and catabolism are important in determining natural variation for leaf dhurrin in sorghum in different environments.

This investigation was aimed at obtaining sorghum plants with reduced expression levels of the gene *CYP79A1,* which is responsible for the production of cyanogenic glycosides, using the antisense approach. This approach was thought to be feasible and effective, and has been shown to reduce linamarin in cassava (12) and obtain cassava plants with 94% reduction in cyanogen production in transgenics with anti-sense of *CYP79D1* and *CYP79D2* genes that regulate the cyanogenesis pathway in cassava. Moreover, the effectiveness of these sorghum genes inducing cyanogen production has been demonstrated in arabidopsis and tobacco (13). Other robust tools such as RNA interference-mediated down-regulation technologies (14) were not deployed since it was not intended to completely block the dhurrin biosynthesis pathway as a small quantity of the dhurrin may be desirable as a defense against insects (15). Antisense technology is known to substantially down-regulate but not block the target gene completely (12, 16, 17). For sorghum, the *CYP79A1* gene is the candidate of choice as it is the first enzyme involved in the dhurrin biosynthesis pathway and the down-regulation of *CYP79A1* would lead to no accumulation of secondary products.

MATERIALS AND METHODS

Gene Isolation and Development of the Antisense–Vector

Based on the sequence of the *CYP79A1* cDNA sequence [Genbank ID U32624; Koch et al. (5)], primers were designed (forward primer CYLS1 5′- ATG GCG ACA ATG GAG GTA GAG GCC GCG GCC G-3′ and reverse primer CYLS2 5′- TCA GAT GGA GAT GGA CGG GTA GAG GTG CG-3′) to amplify the largest ORF (1.67 kb) of the *CYP79A1* gene. From the RNA of the leaf tissue of sorghum genotype CSV 15, *CYP79A1* cDNA was synthesized and amplified by RT-PCR (OneStep RT-PCR kit, Qiagen). The amplicon was cloned in the pTZ57R/T vector (Fermentas, USA) and confirmed by sequencing. The clone with antisense orientation with regards to T7 promoter (*Sma*I-*Xba*I in antisense direction) was selected for sub-cloning in pJS108.

The ORF of *CYP79A1* was excised out from the pTZ57R/T vector at the restriction enzyme sites of *Xba*I and *Sma*I and subcloned into the pJS108 vector. The *UidA* gene was excised out of pJS108 using *Hind*III and *Xba*I. The *Hind*III restricted cohesive terminus (at 5′ end) was filled using T_4 DNA polymerase to create a blunt end. Thus, the vector had an *Xba*I terminus and a blunt end and the *CYP79A1* fragment (to be inserted) had a blunt end (3′ end) and an *Xba*I terminus (5′ end). Semi-directional cloning was hence possible leading to the antisense orientation of the insert. The ligated vector was transformed into *E. coli* and designated pJS108-*CYP79A1*-AS. Restriction analysis of the plasmid from the *E. coli* confirmed the presence and the antisense orientation of the *CYP79A1* fragment in the plasmid (**Figure 1**).

Genetic Transformation in Sorghum to Generate the Putative Transgenic

Shoot apices were obtained from aseptically germinated seedlings of dual-purpose (grain and fodder purpose) sorghum variety *CSV 15*. A cut was made at the base of the seedling apex, just below the attachment of the largest expanded leaf. The top portion with the parts of the unexpanded primordial leaves intact was used as explant (18). The pH of the Murashige and Skoog basal medium was adjusted to pH 5.8. Before bombardment, explants were placed on osmotic medium (MS salts + 17.7 μM BAP + 2.3 μM kinetin + 0.9 μM 2, 4-D + 0.4 M mannitol +0.4 M sorbitol) for 4 h. Twenty-five to thirty explants were placed on each culture plate (94 mm diameter) for bombardment. A home-made particle inflow gun (PIG) was used for gene transfer (19). Genetic transformation of the explants was carried out using tungsten particles (1.0–1.5 μm) coated with the plasmid (1 μg/μl) at a ratio of 1:1 (5 μl each) under a helium gas pressure of 12 kg/cm^2 and partial vacuum (600 mm Hg). The bombarded explants were then left on the osmotic medium and incubated in culture room at 25°C for 24 h in 16 h light/8 h dark condition.

The bombarded explants were sub-cultured on the somatic embryo induction medium (MS salts + 17.7 μM BAP + 2.3 μM kinetin + 2.2 μM 2, 4-D). After 2 weeks in the medium, the enlarged primordial leaves were trimmed while retaining the bulged shoot apex. These explants were further sub-cultured for 2 weeks on somatic embryo maturation medium (MS salts +

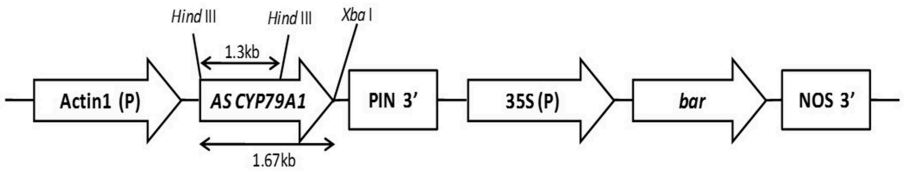

FIGURE 1 | Partial map of *CYP79A1* Antisense RNA construct (pJS108-*CYP79A1-AS*) Vector backbone from plasmid pJS108 (courtesy: Dr. Ray Wu, Cornell University, USA); Act1- Promoter of rice actin (*Act 1*) gene; A.S *CYP79A1*- cDNA of *CYP79A1* in antisense orientation; 35S—CaMV35S promoter; *bar* encodes the enzyme Phosphinothricin acetyl transferase that confers Basta resistance; PIN and NOS—termination sequences.

17.7 μM BAP + 2.2 μM 2, 4-D). After 2 weeks, the meristematic masses containing multiple buds were dissected into 2 to 3 pieces and sub-cultured on somatic embryo germination medium (MS basal + 17.7 μM BAP + 1.3 μM 2,4-D + 7.57 μM Basta), where the multiple bud initials develop into plantlets in 2 weeks. A two-step selection strategy was imposed using Basta @ 7.57 μM for the first 20 days followed by Basta @ 15.14 μM with sub-culturing after every 10 days in each medium, thus imposing selection for 40 days in total. Surviving shoot buds were transferred to shoot elongation medium (MS basal + 4.4 μM BAP + 2.45 μM IBA) and sub-cultured every 2 weeks. All growth stages since bombardment were maintained at 25°C in 16 h light/8 h dark conditions until hardening and transfer to the glasshouse. The healthy shoots were transferred to hormone-free MS medium until they set roots. Later they were transferred to pots in glasshouse and raised to maturity.

Molecular Characterization of Generated Transgenic Plants

Polymerase Chain Reaction (PCR)

PCR analysis was done to verify the presence of the *bar* gene in the regenerated plants. Genomic DNA was isolated from the leaves of one-month old transformants by CTAB method (20). The presence of *bar* transgene was determined by PCR using 250 ng of genomic DNA, deploying 1.0 μM of each of forward primer PAT-F (5′-ACC ATC GTC AAC CAC TAC ATC G-3′) and reverse primer PAT-R (5′-TCT TGA AGC CCT GTG CCT C-3′), in 20 μl reaction containing 1X PCR buffer (Fermentas), 1.2 μM $MgCl_2$, 0.5 μM dNTPs (Fermentas) and 2.5 units of Taq DNA polymerase (Fermentas). The first 10 cycles consisted of denaturation for 1 min at 95°C, followed by annealing for 1 min at 60°C in the first cycle, reducing by 1°C each cycle (Touchdown PCR), and extension for 1.5 min at 72°C. This was followed by 25 cycles of 1 min at 95°C, 1 min at 50°C, and 1.5 min at 72°C. Final extension was done at 72°C for 10 min.

Southern Hybridization

Southern blot hybridization was carried out using 20 μg of genomic DNA from leaves of 1 month old transformed plants according to the standard protocol (21). To check for transgene integration, the *Hind*III enzyme that releases the major fragment (1.3 kb) of *CYP79A1* antisense gene was deployed. To estimate the number of sites of integration of transgenes, the enzyme *Xba*I that has a single restriction site in the plasmid was used. The digested DNA was separated by electrophoresis and transferred to a nylon membrane (Hybond-N^+, Amersham) by vacuum blotting. The probe for verifying transgene integration was prepared by digesting pJS108-*CYP79A1*-AS with *Hind*III. The 1.3 Kb fragment was purified with Qiagen Gel extraction kit and used as probe. Probe for copy number was prepared through PCR. The *bar* gene probe was amplified by PCR using plasmid (pJS108-*CYP79A1*-AS), deploying forward primer PAT-F and reverse primer PAT-R as done in PCR analysis. The length of the probe was 294 bp. The membrane was probed with respective P^{32}dCTP radioactive labeled probes as per manufacturer's instructions (NEBlot Kit, New England Biolabs). The hybridization of the probe was detected on X-ray film by autoradiography.

Reverse Transcription Polymerase Chain Reaction (RT-PCR)

The mRNA was isolated from putative transgenic using the PolyATract® mRNA isolation system III (Promega). RT-PCR amplification was performed in a total volume of 20 μl containing 1X OneStep RT-PCR buffer (Qiagen), dNTPs 10 mM, 5X Q Solution (Qiagen), 1.0 μM of each primer (PAT-F and PAT-R), 1.0-unit Reverse Transcriptase enzyme mix (Qiagen) and 300 ng of mRNA. For reverse transcription reaction, the mix was incubated at 55°C at 30 min followed by initial PCR activation at 95°C for 15 min. The touch-down PCR was performed as described in the PCR section above. RT-PCR was also done to amplify the *CYP79A1* antisense cDNA product whose sequence was found to be identical to the corresponding reverse sequence of a native *CYP79A1* fragment.

Quantitative Real Time PCR (qRT-PCR)

Total RNA was isolated from leaf tissues of T_4 generation plants originating from two insertion events using RNeasy Plant Mini kit (Qiagen) and treated with DNase according to the manufacturer's protocol. 4 μg of RNA was reverse transcribed using Superscript III (Invitrogen) and RNase H treated according to the protocol. A 1:3 dilution of the cDNA was used for quantitative PCR performed on a LightCycler 480 II machine (Roche) using LightCycler 480 SYBR Green I Master Kit (Roche) according to the manufacturer's protocol. For assaying CYP79A1, primers 5′- ATGGCGACAATGGAGGTAGA-3′ and 5′-CACCTCCGGGTTGTTCACCA-3′ were used. Reference gene 18s rRNA (22) was used as internal control with primers 5′- GGCTCGAAGACGATCAGATACC-3′ and 5′- TCGGCATCGTTTATGGTT-3′. The relative expression

analysis module of LightCycler 480 software release 1.5 was used adopting the cDNA from non-transformed control as reference. Melt curve analyses were done to ensure the absence of multiple peaks and primer dimers. The experiment was performed in three technical replicates.

Phenotyping of Generated Transgenic to Estimate the Level of Hydrogen Cyanide (HCN)

Samples were collected and HCN content was analyzed 45 days after sowing (23). Whole plants excluding root portion were collected (three biological replicates) and were finely chopped using a scalpel blade. 1.0 g of finely chopped plant material was poured immediately into 15 ml capacity air-tight glass test tube and 200 μl of chloroform was added. A strip of Whatman No. 1 filter paper (1 cm × 10 cm size) that was pre-dipped in saturated alkaline picric acid was stuck to the cork stopper and immediately suspended in the air-tight tube by closing the cork stopper. This test tube with chopped plant sample and filter paper strip was kept at room temperature for 24 h. Next day the filter paper strip was placed in another tube with 10 ml of distilled water and was washed in water carefully by vortex until all color of the strip comes into solution. Optical density of the solution was measured at 515 nm using a spectrophotometer (Multiscan spectrum, Thermo). Another set of filter paper strips were introduced in the air-tight tube with plant samples and assayed as before after 48 h. Standard curve was prepared with the help of KCN. HCN content was calculated using the standard curve. The percent dry matter in a fresh sample was separately estimated after drying in oven, to express HCN content on dry weight basis.

Visual Phenotyping of Generated Transgenic Plants Through Leaf Painting Assay

Transgenic plants were characterized in terms of sensitivity to the DL-Phosphinothricin (PPT) swab tests performed 25 days after sowing. For leaf painting, a line was drawn on the upper, fully emerged leaves using a black permanent marker. PPT solution (0.5 mg/ml) containing 0.01% Tween-20 (wetting agent) was painted below the line and toward the leaf tip with the help of cotton swab. The leaves were scored for symptoms of leaf scorching after 48 h of painting (24). Symptoms of PPT glutamine synthetase damage manifest as chlorotic discoloration giving a scorched appearance (**Figure S1**). Those plants with <50% leaf area scorched were classified as resistant (due to *bar* transgene expression). All the control plants recorded scorching in 90–100% leaf area. The non-transformed plants of tissue culture origin were used as control.

RESULTS

Reduced levels of HCN is required to improve fodder quality. Through genetic modification we are able to achieve this goal.

Particle Inflow Gum Mediated Genetic Transformation and Generation of Putative Transgenics With Reduced HCN

Particle bombardment is an efficient method of genetic transformation of cereals, wherein biological molecules are driven at high velocity into the target tissue. It offers advantages like the introduction of multiple genes, the simplicity of introducing genes, and transformation in those plants, where agro-infection is difficult. Genetic transformation of sorghum using shoot meristem explants was carried out, using tungsten particles (1–1.5 μm diameter) coated with the *CYP79A1* antisense RNA construct (pJS108-*CYP79A1-AS*) (**Figure 1**) under a helium gas pressure of 12 kg/cm^2 and partial vacuum (600 mm Hg). The bombarded explants were then left on the induction medium followed by selection (**Figure 2A**).

In vitro selection against the selectable marker (*bar*) using Basta was initiated 4 weeks after bombardment. The 4 weeks incubation period allowed for the recovery and proliferation of the cells transformed by bombardment, which might have been damaged during the process of delivering DNA. The meristems enlarged while keeping in the induction and maturation media (**Figure 2B**). After 2 weeks, the meristematic masses containing multiple buds were allowed to develop into auxiliary plantlets on somatic embryo germination medium with selection agent Basta for 40 days. Only one-sixth of the transformed explants (2,400 in total) survived Basta selection (**Figures 2C,D**). Transformed shoot forming sectors could be seen as green to yellow colored growing tissue against a background of brown to black-colored necrotic tissue and the regenerated shoots. The surviving shoots were transferred to hormone-free MS medium to set roots (**Figures 2E,F**).

Confirmation of Transgenics Using Different Molecular Biology Techniques

Stable integration and effective expression of foreign genes are of critical importance for the successful application of genetic engineering in agricultural crops. The most common and powerful ways of detecting the foreign DNA in a transgenic are PCR and Southern analysis. The expression of transgene may be verified by RT-PCR and assaying for expected biochemical changes in traits such as cyanogenic potential.

Polymerase Chain Reaction (PCR) and Southern Blotting-Based Confirmation of Putative Transgenics

PCR was used to confirm the presence of the transgene in the regenerated plants and their progenies till T_3 generation. Since both the *CYP79A1* native sense and antisense sequences are likely to give the same amplification, PCR analysis was done to verify the presence of *bar* gene located on the antisense construct. Out of 45 putative transgenics tested, 30 were found to contain the transgene, as confirmed by the amplification of 294 bp fragment of the *bar* gene (**Figure 3**). The T_1 generation progenies (15–20 plants) of the confirmed 27 plants were raised in glasshouse. On PCR analysis, some of them did not show the presence of the *bar* gene indicating segregation of the transgene.

Southern blot hybridization was carried out using 20 μg of genomic DNA. To confirm the presence of transgene, the

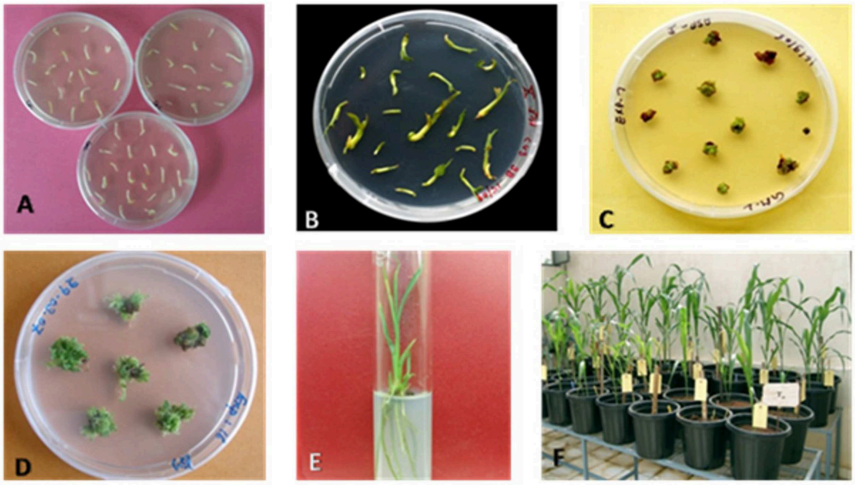

FIGURE 2 | Different steps of Sorghum transformation to generate putative transgenices. **(A)** Shoot meristem explants on induction medium; **(B)** Growth and bulging of explants on induction medium after 12–14 days of incubation; **(C)** Survival of transformed explants in selection; **(D)** Explants with shoot formation in elongation medium; **(E)** Well-developed plants in the rooting medium; **(F)** Putative transgenic plants acclimatized in the glasshouse.

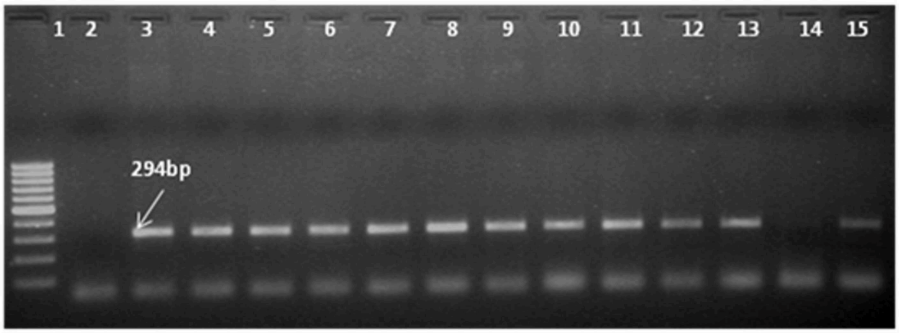

FIGURE 3 | Identification of transgene presence in T_0 plants by Polymerase Chain Reaction (PCR) with primers specific to *bar* gene shows amplification of 294 bp product. Ingel **Lane 1** signifies 100 bp DNA ladder (Fermentas). **Lane 2** and **3** designate negative control and positive control. Plasmid DNA of construct employed in transformation was used as positive control and genomic DNA of untransformed plant was used as negative control. **Lanes 4–15** indicate the line number of respective T_0 plants (S18, S28, S13, S6, S14, S32, S17, S31, S27, S25, S29, S9, S11).

*Hind*III enzyme that releases the major fragment (1.3 kb) of the *CYP79A1* antisense gene was deployed. The integration of the transgene could be confirmed in 27 of the 30 transgenics as evident from the 1.3 kb fragment lighting up in the *Hind*III restricted genomic DNA of the transgenics compared to the DNA from non-transformed control plant (**Figure 4**). Southern analysis was also used to probe the number of sites of integration of the transgene in the transgenic plant DNA. This was done in plants that were confirmed for the presence of transgene through *Hind*III restriction analysis. A single restriction enzyme *Xba*I, that cuts the plasmid only once at a specific site in the transgene, was employed (**Figure 5**). Once the transgene is integrated into the genome at a particular site, the restriction analysis performed using such enzyme would lead to generation of fragments of varying length since the other end of the fragment would be at a random site in the plant genome. It was observed that the plants exhibited 2–5 sites of integration as visualized by the fragments that showed up by hybridizing with the transgene *bar* probe. Though large proportion of plants exhibited 2 to 3 sites of transgene integration, some of them had up to five sites of integration. The Plants S13 and S34 showed one fragment followed by S15 and S35 with two fragments. While S6, S8, S18, and S25 exhibited three fragments, at least four fragments were seen in the plant S4, and five fragments could be observed in S16 and S32. Each fragment indicated one site of transgene integration and thus multiple sites of integration were observed in these transgenics. Each site of integration may carry one or more copy of the transgene.

Reverse Transcription (RT-PCR) Based Confirmation

RT-PCR of transgenics was done in 30 PCR confirmed T_0 plants. RT-PCR using the *bar* primers revealed that 27 transgenics

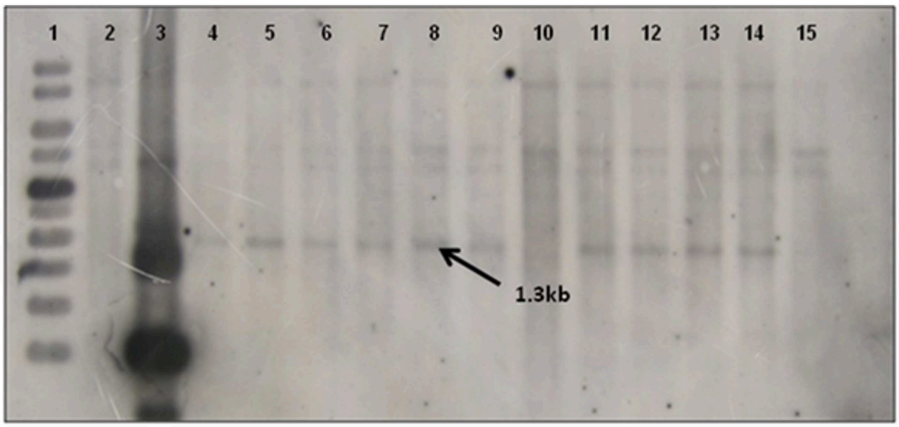

FIGURE 4 | Southern blot of T_0 Transgenics lines for gene integration using pJS108-*CYP79A1-AS* gene as probe. The genomic DNA of T_0 plants was digested with *Hind*III enzyme and blot was probed with radiolabelled pJS108-*CYP79A1-AS* gene. **Lane 1**- GeneRuler 1Kb (Fermentas), **Lane 2**- Wild type genomic DNA (negative control), **Lane 3**- Plasmid DNA of pJS108-CYP79A1-AS construct used for transformation, **Lanes 4–15** indicate the line number of respective T_0 plants (S28, S13, S6, S35, S4, S32, S15, S22, S25, S2, S9, S7). The presence of 1.3 Kb band indicates the presence of transgene (pJS108-CYP79A1-AS gene).

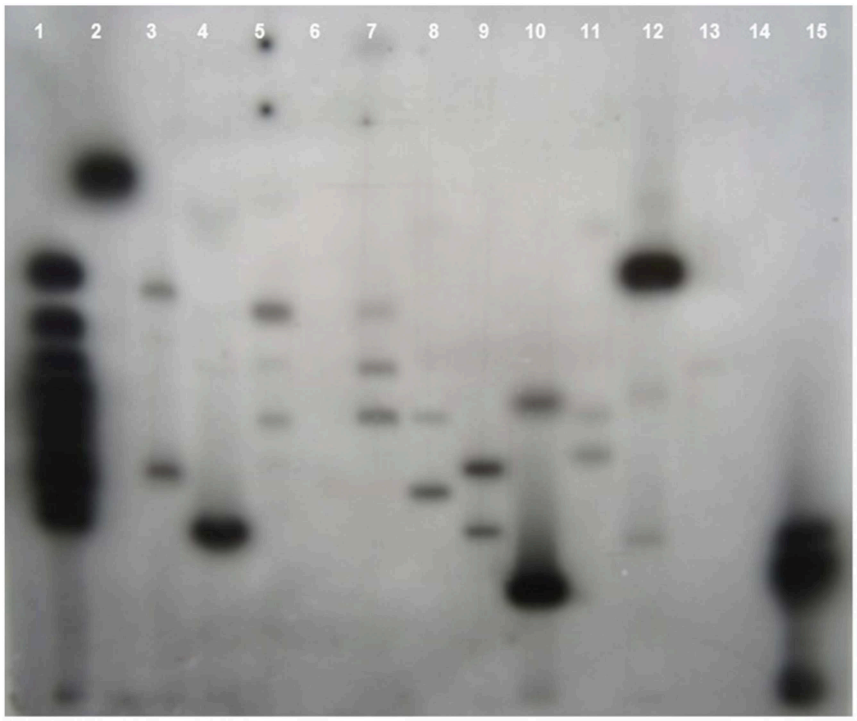

FIGURE 5 | Southern blot of T_0 Transgenics lines for sites of transgene integration. The genomic DNA of T_0 plants was digested with *Xba1* enzyme and blot was probed with radiolabelled *bar* gene. **Lane 1**- GeneRuler 1Kb DNA ladder (Fermentas), **Lanes 2–13** indicate the line number of respective T_0 plants (S13, S18, S4, S32, S17, S6, S35, S15, S25, S8, S16, S34), **Lane 14**- Wild type genomic DNA (Negative control), **Lane 15** Plasmid DNA of pJS108-CYP79A1-AS construct used for transformation as positive control.

expressed the transgene in the events tested (**Figure 6**). *Bar* gene-specific RT-PCR product sequence matched the known sequence of the marker gene. RT-PCR was also done to amplify the *CYP79A1* antisense cDNA product. The reverse sequence of the native gene portion was recovered by sequencing the product of antisense *CYP79A1*-specific RT-PCR (**Figure 7**).

Visualized Selection Marker Phenotype Through Leaf Painting Assay

Identification of the functionality of the transgene in the transgenic plants and tracking the inheritance of the transgenes in their progeny was done through leaf painting assay. Seeds from 36 T_0 transgenic plants along with non-transformed controls

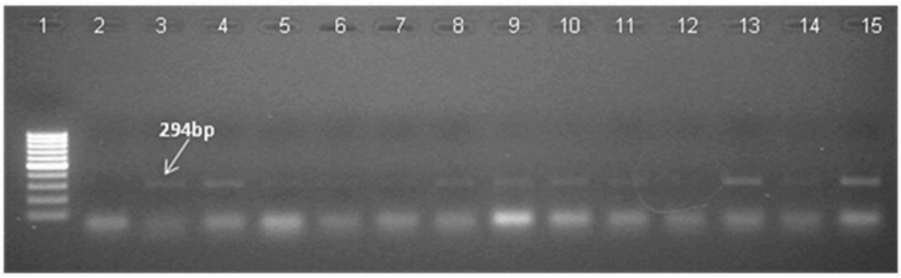

FIGURE 6 | Reverse transcriptase PCR (RT-PCR) of T_0 Transgenics lines for *bar* gene RT-PCR with primers specific to *bar* gene shows amplification of 294 bp product. Lane **1** signifies 100 bp DNA ladder (Fermentas). **Lane 2** designate negative control. Wild type RNA of untransformed plant was used as negative control, **Lanes 3–15** indicate the line number of respective T_0 plants (S18, S28, S13, S6, S35, S4, S32, S17, S22, S11, S25, S9, S16). The presence of 294 bp band indicates the expression of *bar* gene in transgenics plants.

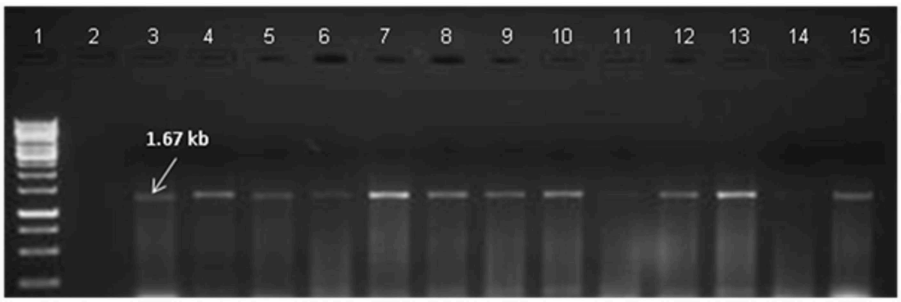

FIGURE 7 | Reverse transcriptase PCR (RT-PCR) of T_0 Transgenics lines for antisense gene. RT-PCR analysis of To Transgenics lines using antisense *CYP79A1* gene-specific primers shows amplification of 1.67 Kb product. **Lane 1** signifies GeneRuler 1 kb DNA ladder (Fermentas), **Lane 2**- designate negative control. Wild type RNA of untransformed plant was used as negative control; **Lanes 3–15** indicate the line number of respective T_0 plants. (S18, S28, S13, S6, S35, S4, S32, S17, S22, S11, S25, S9, S16).

were germinated. Leaf painting assay were performed in 25 days old plants with 0.05% PPT concentration. Progenies from 36 T_0 plants were screened by PPT leaf painting of which 20 progenies (S10, S11, S16, S22, S23, S29, S32, S33, S36, S5, S8, S12, S18, S20, S4, S24, S25, S27, and S26) were without chlorosis in at least 50% plants.

Developed Transgenic Events Showed Reduced HCN Level in the Advanced Generations (T_0, T_1, T_2, T_3)

The HCN content was determined in T_0 transgenics 45 days after transferring them to pots. The HCN content was reduced in the confirmed T_0 transgenic plants which ranged from 17.25 to 178.66 μg/g with a mean of 104.06 μg/g as compared to 192.08 μg/g in the non-transformed control. There were 14 transgenic plants with <100 μg/g HCN and another 13 possessed more than 100 μg/g (**Table 1**). Glasshouse-raised T_1 progenies of 24 events were assayed for HCN content in leaves 45 days after sowing. Each event was assayed along with non-transformed control. Events with a greater proportion of plants exhibiting substantially reduced HCN content were observed. The HCN content in T_1 progenies varied from 5.1 to 149.8 μg/g as against mean HCN content of non-transformed control of 211.5 μg/g. Antisense mediated down-regulation of *OMT* gene in maize resulted in significantly lower *O*-methyltransferase activity in T_1 transgenics with some plants showing a 60% reduction (16).

T_1 generation plants confirmed by PCR were characterized for HCN production and those with lowered HCN were advanced to next generation. The progenies (T_2) of the promising 43 T_1 plants were characterized by PCR and assayed for HCN as earlier for advancement to T_3. The plants with reduced HCN consistently across the generations were identified (**Table 2**). T_1 progeny of S4 had minimum HCN of 50.3 μg/g and maximum of 156 μg/g with a mean HCN of 88.3 μg/g. Its progeny S4-2 showed 88.2 μg/g of HCN which was further advanced to T_2. Its progenies exhibited HCN levels from 11.5 to 131.0 μg/g. Similarly, in the S35 event HCN content was 132.6 μg/g while in T_1 progeny it ranged between 64.3 and 326.2 μg/g. From this event, 2 progenies (S35–10 and S35–12) were advanced to T_2 which had HCN ranges of 2.8- 133.7 and 16.0- 154.8 μg/g. T_3 generation is an advanced stage where lines are expected to be relatively stable for inheritance and transgene expression. Hence, 7T_2 plant progenies from two T_0 plants (S4 and S32) that were consistently positive in molecular analyses and possessed lowered HCN content across generations were advanced to T_3. The progenies of individual T_0 plants viz., S4, S6, S10, S17, S3, and S36 showed Mendelian segregation (3:1 ratio) for HCN content (lower than control: higher than control). For the hemizygous transgene, T_1 progeny segregated in 3:1 ratio for HCN content.

TABLE 1 | Frequency distribution of *CYP79A1* antisense transgenics (T_0) for HCN content.

Class range (HCN in μg/g)	0–20	21–40	41–60	61–80	81–100	101–120	121–140	141–160	161–180	181–200
No. of transgenics	1	1	4	3	5	1	5	3	4	0

HCN (μg/g) in (i) non-transformed control = 192.08 ± 2.15 (n = 20); (ii) Transformed (but no gene integration was detected) control = 207.0 (n = 20).

TABLE 2 | HCN content (in μg/g) in promising events in T_1 and T_2 generations.

S No.	Event (T_O)	HCN content in T_O	HCN content in T_1 progeny			Selected T_1 plant ID	HCN in selected T_1 plant	HCN content in T_2 progeny		
			Minimum	Maximum	Mean			Minimum	Maximum	Mean
1	S4	154.7 ± 7.2	50.3	156.0	88.3	S4-2	88.2	11.5	131.0	55.5 ± 7.2
2	S16	68.12 ± 17.5	5.1	112.4	57.4	S16-2	70.9	36.6	171.5	113.9 ± 28.5
3	S17	60.8 ± 19.8	33.5	169.1	104.9	S17-6	68.6	4.0	191.9	92.2 ± 26.7
4	S20	115.0 ± 23.3	53.1	174.4	97.4	S20-3	71.3	0.5	160.1	50.5 ± 24.1
5	S25	46.8 ± 37.4	30.7	295.1	144.5	S25-15	50.3	4.7	99.9	42.4 ± 23.5
6	S26	87.4 ± 38.8	68.7	370.0	151.6	S26-7	68.7	9.5	106.9	38.4 ± 26.4
7	S32	86.6 ± 39.6	51.2	209.3	104.2	S32-1	51.2	8.6	134.8	66.5± 26.4
8	S33	139.1 ± 45.7	56.7	495.2	292.5	S33-4	100.0	2.3	150.3	55.7 ± 22.3
9	S35	132.6 ± 28.3	64.3	326.2	198.2	S35-10	94.7	2.8	133.7	43.3± 28.0
10	S35	–	–	–	–	S35-12	67.9	16.0	154.8	88.3 ± 27.0
11	Control*	192.08	89.3	399.0	213.1	Control	211.5	100.1	295.0	189.1 ± 18.5

**Each of the T_1 and T_2 progenies were compared against separate non-transformed controls and the range and mean across all those controls is given above.*

Phenotyping of T_3 Generation of Transgenics Plants Having Reduced Level of Hydrogen Cyanide

A total of seven T_2 plant progenies from S4 and S32 transgenic events that consistently possessed lower HCN content across generations were evaluated in T_3 generation. More than 90 percent of the T3 progenies were positive for presence of transgene as indicated by PCR (data not shown) suggesting that advancing of generations had effectively eliminated the non-transgenic sergeants.

The HCN analysis done in 45 days old plants showed consistently low HCN levels in these plant progenies (**Table 3**). The level of HCN ranged from 6.1 μg/g (in S4-26) to 107.8 μg/g (in S4-25) with a mean of 62.9 μg/g compared to a range of 184.2–398.1 μg/g in control with a mean of 221.4 μg/g. In the progenies of S32 segregants, the range was 25.1 μg/g (in S32–21) to 150.2 μg/g (in S32–21) with a mean of 76.2 μg/g. Thus, the reduced cyanogen potential of T_1 plants was inherited stably in these progenies. It is proposed that a total of 3–5 progenies with lowest HCN content from each event may be advanced to next generation as they are expected to be the most promising for lower cyanogenic potential.

TABLE 3 | HCN content in T_3 progenies of transgenics (in μg/g).

S. no.	Plant ID	Range (reduced HCN)	Mean	Control (mean)	No. of plants
1	S32-12	32.4–96.9	63.2 ± 9.1	217.2 ± 3.6	18
2	S32-14	13.7–127.8	82.3 ± 5.7	200.7 ± 2.4	19
3	S32-21	25.1–150.2	75.4 ± 2.4	196.9 ± 13.4	8
4	S4-25	13.7–107.8	53.8 ± 9.3	240.8 ± 3.6	11
5	S4-26	6.1–104.3	60.9 ± 6.0	187.0 ± 19.7	16
6	S4-21	49.3–82.9	67.3 ± 17.0	216.8 ± 17.6	13

Reduction of *CYP79A1* Gene Expression in Developed and Stable Transgenic Plant Progenies

The reduction of *CYP79A1* gene expression in the transgenic plant progenies was assayed through quantitative PCR (Real-time PCR) in selected progenies of two promising events S4 and S32 (**Table 4**; **Figure 8**). The plants from the event S4 showed from 7 to 12,180 times reduced expression of the *CYP79A1* (**Table 4**). Similarly, the reduction was from 11 times to 42,017 times in the plants of event S32. Both events had recorded lower HCN levels in the range of 6–84 μg/g. The correlation of the HCN content with normalized Cp ratio was statistically significant (Pearson product-moment correlation $r = 0.532$, $p < 0.0375$). Therefore, it can be concluded that the reduced HCN levels in the transgenics are due to the lowered expression of *CYP79A1* gene.

DISCUSSION

Sorghum (*Sorghum bicolor* L. Moench) is grown extensively throughout the world owing to its wider adaptability for varied soil and moisture conditions and utility as food, fodder, fuel, and fiber. Mixed crop and livestock farming systems are mostly prevalent in India. Therefore, both grain and fodder are of equal importance, underscoring the interdependence of crops and livestock. Sorghum is an ideal forage crop due to its quick

growth, high yielding ability (350–400 q/ha and 700–750 g/ha of green fodder from single cut and multi-cut varieties/hybrids, respectively), high dry matter content (25–30%), better quality (crude protein 5–9%) and its suitability for various forms of utilization like green chop, silage and hay.

The particle inflow gun was deployed in the present investigation for sorghum transformation using the plasmid carrying the *CYP79A1* antisense strand under the control of *Actin1* promoter (**Figure 1**). The protocols for producing transgenic sorghum are still being standardized using various methods of DNA transfer (25, 26). Genetic transformation of sorghum using shoot meristem explants was carried in this experiment (27, 28). The advantage of adopting shoot meristem explants (**Figure 2A**) for the transformation lies in its year-round availability unlike explants such as immature embryos and immature inflorescences (27, 29–31). This system could be conveniently adapted for sorghum transformation as large number of multiple buds and somatic embryos capable of regeneration are obtained in a relatively short period. Heterogeneity of the explant is also minimized as the shoot apices have very limited mother plant tissue, which also resulted in enhancing the rate of multiplication.

TABLE 4 | Reduction in *CYP79A1* gene levels as estimated by quantitative PCR.

Sample no.	Progeny ID	Target/reference Cp ratio*(normalized)	Reduction in *CYP79A1* levels (number of folds)	HCN (μg/g)
1	S4-25-2	0.00058	1,737	17
2	S4-25-6	0.00030	3,342	14
3	S4-25-12	0.04092	24	25
4	S4-26-6	0.00494	203	23
5	S4-26-10	0.01033	97	6
6	S4-26-17	0.03951	25	56
7	S4-26-18	0.15234	7	84
8	S4-21-1	0.00008	12,180	57
9	S32-12-6	0.09238	11	32
10	S32-12-7	0.00008	12,903	51
11	S32-21-4	0.00002	42,017	25
12	S32-14-4	0.00026	3,824	27

**Control (non-transformed) ratio was considered as unity and relative expressions were worked out.*

Thirty putative transgenics were confirmed by showing the amplification of 294 bp fragment of the *bar* gene in the generated transgenics. Transformed tissues with the selective marker gene may be capable of detoxifying the selective agent so efficiently that non-transformed tissue in close proximity can also survive (32) and such a phenomenon can lead to false positives during *in vitro* selection. Out of 30 PCR confirmed transgenics, 27 showed the transgene integration. In the transgene integration southern blot, upper fragments that appear in all plants including control indicated the native *CYP79A1* DNA. Similar results were obtained in cassava (12). The Plants S13 and S34 showed single copy insertation and comparatively low dhurrin expression, but due to the unavailability of seeds, we could not advance these lines. Interestingly, as the data of HCN content in T_1 reveal, plants with more sites of integration such as S4, S16, and S32 (4–5 sites of integration) had lower HCN content.

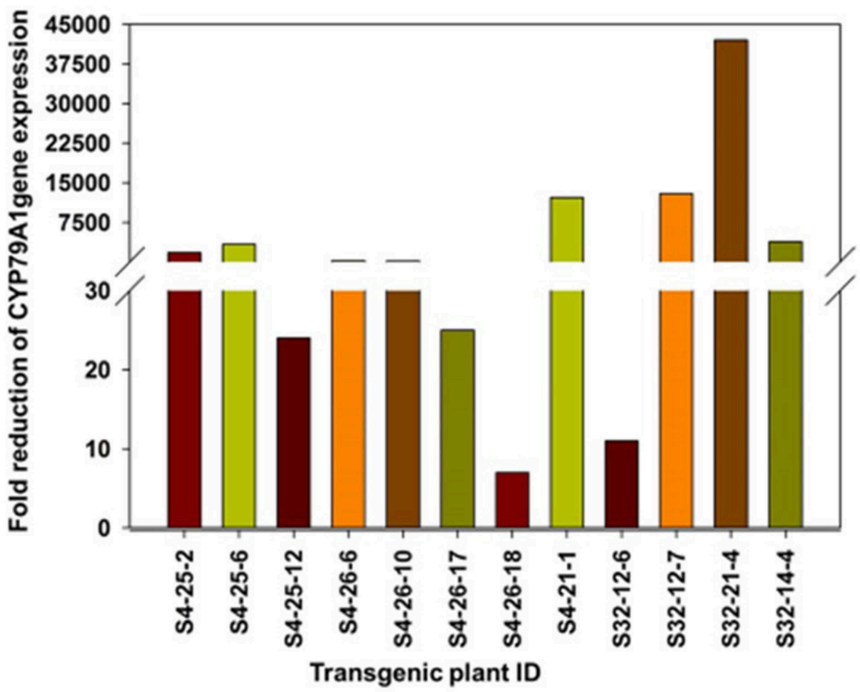

FIGURE 8 | Fold reduction (on log scale) of *CYP79A1* gene expression as determined by quantitative Real time PCR in selected progenies of two promising events S4 and S32. The plants from the event S4 and S32 showed reduced expression of the CYP79A1.

However, in progenies of plant S35 (2 sites of integration), some plants did exhibit higher HCN content. The integration of multiple copies of transgene have been reported by several workers using the biolistic transformation method (33) observed that in rice, the higher copy number of transgenes led to its higher expression levels, where more β-carotene was produced. Similarly, transgenic rice with more than one copy of *Bt* gene performed better in field conditions suggesting effective levels of transgene expression (34, 35). Also, observations of high transgene expression levels were reported in wheat transformed with multiple copies of reporter genes (36). Transformants selected for producing higher levels of pharmaceuticals were found to contain three or more transgene copies (37). Besides, events with a single copy losing the expression of transgene in T_1 and later generations have also been observed (38). Transgene "switching off" is largely due to the presence of one or more rearranged copies rather than multiple copies of the transgene (39). However, on the contrary, multiple copies of transgenes were thought to result in gene silencing due to co-suppression (40). It was also believed that the expression of a multi-copy transgene may vary over generations compared to a low copy number transgene event (41). The important problem for the potential use of transgenic plants for crop improvement is the instability of transgene expression. The expression of the integrated transgene was confirmed by reverse transcription PCR (RT-PCR). Few of the T_0 regenerants in the present study failed to show up in RT-PCR, but were found positive in PCR and/or Southern for transgene(s). They exhibited HCN levels similar to that of control (186–219 μg/g). Gene silencing and interactions between multiple copies of the same transgene or different transgenes are known to result in unexpected expression patterns of foreign genes (42). This transgene inactivation in T_0 plants maybe because of the hemizygous state of the transgene (43). Methylation can be induced by interactions between homologous transgene copies or may reflect genomic position effects (32). The gene silencing may occur at transcriptional (44), post-transcriptional levels (45), and reported a functional role for methylation in gene silencing (42).

The leaf painting assay also confirmed the presence of the *bar gene* by showing various ranges of chlorosis in 36 T_0 progenies. Out of 36 putative transgenics, 20 progenies were without chlorosis in plants. The progenies of T_0 plants that showed low HCN content also showed resistance to PPT indicating the expression of both *bar* and *CYP79A1* antisense transgenes in these plants. The assay was also successfully used to confirmed transgenic integration in Indicia rice (46).

Antisense approach had been successful in down-regulation of several pathways in plant species. Transgenic tomatoes were developed with reduced sucrose synthase (SuSy) activity in fruit by expressing an antisense RNA fragment of the *TOMSSF* gene under the control of the cauliflower mosaic virus 35S promoter (47). The transgenic citrus lines that produce higher level (over expression) of antisense ACS RNA were found to repress the increase of ACC content following the chilling treatment (48). Cassava transgenic were also developed with reduced levels of *CYP79D1* and *CYP79D2* enzymes resulting in the inhibition of cyanogen production by antisense technology (12). Other robust tools such as RNA interference-mediated down-regulation technologies (14) were not deployed since it was not intended to completely block the dhurrin biosynthesis pathway. The synthesis of a small quantity of the dhurrin may be desirable as a defense against insects (49, 50). Antisense technology is known to substantially down-regulate but not to block the target gene completely (12, 16, 17). Generated transgenics were also showed the Mendelian segregation for HCN content (lower than control: higher than control). Earlier researchers have indicated that hemizygous nature of transgenics and the transgene segregation in 3:1 ratio of dominant: recessive in selfed progeny are common (51, 52). Transgene expression heterogeneity may also be due to the influence of factors like position effects, gene rearrangement, gene silencing and co-suppression (43, 51, 53).

CONCLUSIONS

In the present study, the antisense RNA approach was used to down-regulate dhurrin synthesis. The inheritance of the reduced HCN content was studied up to T_3 generation by advancing the progenies based on screening for transgene by PCR and assaying the HCN content. The HCN content in the transgenics varied from 5.1 to 149.8 μg/g compared to 192.08 μg/g in the non-transformed control on dry weight basis. Progenies with reduced HCN content were advanced after each generation till T3. Other robust tools such as RNA interference-mediated down-regulation technologies were not deployed since it was not intended to completely block the dhurrin biosynthesis pathway. The synthesis of a small quantity of the dhurrin may be desirable as a defense against insects. This investigation demonstrated that the antisense *CYP79A1* strategy is effective in producing sorghum plants with lower cyanogenic potential. The frequency and inheritance of the transgene improved with the advancement of generations perhaps due to elimination of non-carrier plants of transgene by molecular and HCN analyses. The rice *actin1* promoter is adequately driving the antisense *CYP79A1* expression in the leaf tissues of sorghum. Absence of other unintended agronomic variations in the transgenics (in terms of fodder and grain yield components; data not shown) indicated that the transgenics thus obtained would be useful as cultivars. The present study effectively demonstrated that the antisense strategy was effective in producing sorghum plants with low cyanogenic potential and the biolistic method of transformation remains a useful tool for obtaining transgenic sorghum plants.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

AP and BB conceived and designed the experiments, analyzed the data, and wrote the paper. AP and PM performed the experiments.

FUNDING

Initial research funding received from the Department of Biotechnology, Government of India (Grant No. BT/PR6031/AGR/02/307/2005) is duly acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2019.00122/full#supplementary-material

REFERENCES

1. F.A.O. *FAOSTAT*. Food and Agricultural Organization (2018). Available online at: http://fao.org/faostat/en/#home (accessed August 10, 2018).
2. Haskins FA, Gorz HJ, Johnson BE. Seasonal variation in leaf hydrocyanic acid potential of low and high Dhurrin sorghum. *Crop Sci.* (1987) 27:903–6. doi: 10.2135/cropsci1987.0011183X002700050014x
3. Kojima M, Poulton JE, Thayer SS, Conn EE. Tissue distributions of dhurrin and enzymes involved in its metabolism in leaves of *Sorghum bicolor*. *Plant Physiol.* (1979) 63:1022–8. doi: 10.1104/pp.63.6.1022
4. Kahn RA, Bak S, Svendsen I, Halkier BA, Moller BL. Isolation and reconstitution of cytochrome P450ox and *in vitro* reconstitution of the entire biosynthetic pathway of the cyanogenic glucoside dhurrin from sorghum. *Plant Physiol.* (1997) 115:1661–70. doi: 10.1104/pp.115.4.1661
5. Koch BM, Sibbesen O, Halkier BA, Svendsen I, Moller BL. The primary sequence of cytochrome P450tyr, the multifunctional N-Hydroxylase catalyzing the conversion of L-Tyrosine to p-Hydroxyphenyl acetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L) Moench. *Arch. Biochem. Biophys.* (1995) 323:177–86. doi: 10.1006/abbi.1995.0024
6. Sibbesen O, Koch B, Halkier BA, Moller BL. Cytochrome P450 is a multifunctional heme-thiolate enzyme, which catalyzes the conversion of l-tyrosine into p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogene glucoside dhurrin in *Sorghum bicolor* (L.) *Moench. J. Biol. Chem.* (1995) 270:3506–11. doi: 10.1074/jbc.270.8.3506
7. Bak S, Kahn RA, Nielsen HL, Møller BL, Halkier BA. Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. *Plant Mol Biol.* (1998) 36: 393–405.
8. Jones PR, Moller BL, Hoj PB. The UDP glucose-p-hydroxymandelo-nitrile-o-glucosyl transferase that catalyzes the last step in synthesis of the cyanogenic glucoside dhurrin from *Sorghum bicolor*. *J Biol Chem.* (1999) 274:35483–91. doi: 10.1074/jbc.274.50.35483
9. Blomstedt CK, Gleadow RM, O'Donnell N, Naur P, Jensen K, Laursen T, et al. A combined biochemical screen and TILLING approach identifies mutations in *Sorghum bicolor* L. Moench resulting in acyanogenic forage production. *Plant Biotechnol J.* (2012) 10:54–66. doi: 10.1111/j.1467-7652.2011.00646.x
10. Gleadow RM, Møller BL. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annu Rev Plant Biol.* (2014) 29:155–85. doi: 10.1146/annurev-arplant-050213-040027
11. Hayes CM, Burow GB, Brown PJ, Thurber C, Xin Z, Burke JJ. Natural variation in synthesis and catabolism genes influences dhurrin content in sorghum. *Plant Genome.* (2015) 8:1–9. doi: 10.3835/plantgenome2014.09.0048
12. Siritunga D, Sayre RT. Generation of cyanogen-free transgenic cassava. *Planta.* (2003) 217:367–73. doi: 10.1007/s00425-003-1005-8
13. Bak S, Olsen CE, Halkier BA, Møller BL. Transgenic tobacco and Arabidopsis plants expressing the two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in dhurrin biosynthesis. *Plant Physiol.* (2000)123:1437–48. doi: 10.1104/pp.123.4.1437
14. Kusaba M. RNA interference in crop plants. *Curr Opin Biotechnol.* (2004) 15:139–43. doi: 10.1016/j.copbio.2004.02.004
15. Hansen KS, Kristensen C, Tattersall DB, Jones PR, Olsen CE, Bak S, et al. The *in vitro* substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor*. *Phytochemistry.* (2003) 64:143–51. doi: 10.1016/S0031-9422(03)00261-9
16. He X, Hall MB, Gallo-Meagher M, Smith RL. Improvement of forage quality by downregulation of maize O-methyltransferase. *Crop Sci.* (2003) 43:2240–51. doi: 10.2135/cropsci2003.2240
17. Lin C, Tsai ZT, Wang D. Role of antisense RNAs in evolution of yeast regulatory complexity. *Genomics.* (2013) 102:484–90. doi: 10.1016/j.ygeno.2013.10.008
18. Harshavardhan D, Rani TS, Ulaganathan K, Seetharama N. An improved protocol for regeneration of Sorghum bicolor from isolated shoot apices. *Plant Biotechnol.* (2002) 19:163–71. doi: 10.5511/plantbiotechnology.19.163
19. Finer JJ, Vain P, Jones MW, McMullen MD. Development of the particle inflow gun for DNA delivery to plant cells. *Plant Cell Rep.* (1992) 11:323–8. doi: 10.1007/BF00233358
20. Saghai MM, Biyashev RM, Yang GP, Zhang Q, Allard RW. Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. *Proc Natl Acad Sci USA.* (1994) 91:5466–70. doi: 10.1073/pnas.91.12.5466
21. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*. New York, NY: Cold Spring Harbor Laboratory Press (1989).
22. Baerson SR, Dayan FE, Rimando AM, Nanayakkara ND, Liu CJ, Schröder J, et al. A functional genomics investigation of allelochemical biosynthesis in Sorghum bicolor root hairs. *J Biol Chem.* (2008) 283:3231–47. doi: 10.1074/jbc.M706587200
23. Haque MR and Bradbury JH. Total cyanide determination of plants and foods using the picrate and acid hydrolysis methods. *Food Chem.* (2002) 77:107–14. doi: 10.1016/S0308-8146(01)00313-2
24. Dennehey BK, Retersen WL, Ford-Santino C, Pajeau M, Armstrong CL. Comparison of selective agents for use with the selectable marker gene bar in maize transformation. *Plant Cell Tissue Organ Cult.* (1994) 36:1–7. doi: 10.1007/BF00048308
25. Gurel S, Gurel E, Kaur R, Wong J, Meng L, Tan HQ, et al. Efficient, reproducible Agrobacterium-mediated transformation of sorghum using heat treatment of immature embryos. *Plant Cell Rep.* (2009) 28:429–44. doi: 10.1007/s00299-008-0655-1
26. Visarada KB, Padmaja PG, Saikishore N, Pashupatinath E, Royer M, Seetharama N, et al. Production and evaluation of transgenic sorghum for resistance to stem borer. *In Vitro Cell Dev Biol Plant.* (2014) 50:176–89. doi: 10.1007/s11627-013-9561-5
27. Girijashankar V, Sharma HC, Sharma KK, Swathisree V, Prasad LS, Bhat BV, et al. Development of transgenic sorghum for insect resistance against the spotted stem borer (*Chilo partellus*). *Plant Cell Rep.* (2005) 24:513–22. doi: 10.1007/s00299-005-0947-7
28. Yellisetty V, Reddy LA, Mandapaka M. In planta transformation of sorghum (*Sorghum bicolor (*L.) Moench) using *TPS1* gene for enhancing tolerance to abiotic stresses. *J Genet.* (2015) 94:425–34. doi: 10.1007/s12041-015-0540-y
29. Zhong H, Wang W, Sticklen M. *In vitro* morphogenesis of *Sorghum bicolor* (L.) Moench: efficient plant regeneration from shoot apices. *J Plant Physiol.* (1998) 153:719–26. doi: 10.1016/S0176-1617(98)80226-5
30. Tadesse Y, Sagi L, Swennen R, Jacobs M. Optimisation of transformation conditions and production of transgenic sorghum (*Sorghum bicolor*) via microparticle bombardment. *Plant Cell Tissue Organ Cult.* (2003) 75:1–8. doi: 10.1023/A:1024664817800
31. Devi P, Zhong H, Sticklen M. Production of Transgenic sorghum plants with related HVA1 gene. In: Seetharama N and Godwin ID, editors. *Sorghum Tissue Culture and Transformation*. New Delhi: Oxford Publishers (2004). p. 75–79.
32. Christou P. Strategies for variety-independent genetic transformation of important cereals, legumes and woody species utilizing particle bombardment. *Euphytica.* (1995) 85:13–27. doi: 10.1007/978-94-011-0357-2_2

33. Datta K, Baisakh N, Oliva N, Torrizo L, Abrigo E, Tan J, et al. Bioengineered 'golden'indica rice cultivars with β-carotene metabolism in the endosperm with hygromycin and mannose selection systems. *Plant Biotechnol J.* (2003) 1:81–90. doi: 10.1046/j.1467-7652.2003.00015.x
34. Tu J, Zhang G, Datta K, Xu C, He Y, Zhang Q, et al. Field performance of transgenic elite commercial hybrid rice expressing *Bacillus thuringiensis* δ-endotoxin. *Nat Biotechnol.* (2000) 18:1101–4. doi: 10.1038/80310
35. Ye GY, Tu J, Hu C, Datta K, Datta SK. Transgenic IR72 with fused Bt gene *cry1Ab/cry1Ac* from *Bacillus thuringiensis* is resistant against four lepidopteran species under field conditions. *Plant Biotechnol.* (2001)18:125–33. doi: 10.5511/plantbiotechnology.18.125
36. Stoger E, Williams S, Keen D, Christou P. Molecular characteristics of transgenic wheat and the effect on transgene expression. *Transgenic Res.* (1998) 7:463–71. doi: 10.1023/A:1008833324193
37. Nandi S, Suzuki YA, Huang J, Yalda D, Pham P, Wu L (2002). Expression of human lactoferrin in transgenic rice grains for the application in infant formula. *Plant Sci*. 163, 713–22. doi: 10.1016/S0168-9452(02)00165-6
38. Fu X, Kohli A, Twyman RM, Christou P. Alternative silencing effects involve distinct types of non-spreading cytosine methylation at a three-gene, single-copy transgenic locus in rice. *Mol Gen Genet.* (2000) 263:106–18. doi: 10.1007/PL00008669
39. Kohli A, Gahakwa D, Vain P, Laurie DA, Christou P. Transgene expression in rice engineered through particle bombardment: molecular factors controlling stable expression and transgene silencing. *Planta.* (1999) 208:88–97. doi: 10.1007/s004250050538
40. Matzke MA, Matzke AJ. How and why do plants inactivate homologous (trans) genes? *Plant Physiol.* (1995) 107:679–85. doi: 10.1104/pp.107.3.679
41. Jones HD. Wheat transformation: current technology and applications to grain development and composition. *J Cereal Sci.* (2005) 41:137–47. doi: 10.1016/j.jcs.2004.08.009
42. Kumpatla SP, Hall TC. Recurrent onset of epigenetic silencing in rice harboring a multi-copy transgene. *Plant J.* (1998) 14:129–35. doi: 10.1046/j.1365-313X.1998.00097.x
43. Elmayan T, Vaucheret H. Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally. *Plant J.* (1996) 9:787–97. doi: 10.1046/j.1365-313X.1996.9060787.x
44. Meyer P, Heidmann I, Niedenhof I. Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. *Plant J.* (1993) 4: 89–100. doi: 10.1046/j.1365-313X.1993.04010089.x
45. Smith RH, Hood EE. *Agrobacterium tumefaciens* transformation of monocotyledons. *Crop Sci.* (1995) 35:301–9. doi: 10.2135/cropsci1995.0011183X003500020001x
46. Datta SK, Datta K, Soltanifar N, Donn G, Potrykus I. Herbicide-resistant Indica rice plants from IRRI breeding line IR72 after PEG-mediated transformation of protoplasts. *Plant Mol Biol.* (1992) 20:619–29. doi: 10.1007/BF00046447
47. D'Aoust MA, Yelle S, Nguyen-Quoc B. Antisense inhibition of tomato fruit sucrose synthase decreases fruit setting and the sucrose unloading capacity of young fruit. *Plant Cell.* (1999) 11:2407–18. doi: 10.1105/tpc.11.12.2407
48. Wong WS, Li GG, Ning W, Xu ZF, Hsiao WW, Zhang LY, et al. Repression of chilling-induced ACC accumulation in transgenic citrus by over-production of antisense 1-aminocyclopropane-1-carboxylate synthase RNA. *Plant Sci.* (2001) 161:969–77. doi: 10.1016/S0168-9452(01)00505-2
49. Woodhead S, Bernays E. Changes in release rates of cyanide in relation to palatability of Sorghum to insects. *Nature.* (1997) 270:235–36. doi: 10.1038/270235a0
50. Tattersall DB, Bak S, Jones PR, Olsen CE, Nielsen JK, Hansen ML, et al. Resistance to an herbivore through engineered cyanogenic glucoside synthesis. *Science.* (2001) 293:1826–8. doi: 10.1126/science.1062249
51. Maqbool SB, Christou P. Multiple traits of agronomic importance in transgenic indica rice plants: analysis of transgene integration patterns, expression levels and stability. *Mol Breed.* (1999) 5:471–80. doi: 10.1023/A:1009634226797
52. Kim WS, Kim J, Krishnan HB, Nahm BH. Expression of *Escherichia coli* branching enzyme in caryopses of transgenic rice results in amylopectin with an increased degree of branching. *Planta.* (2005) 220:689–95. doi: 10.1007/s00425-004-1386-3
53. Iyer LM, Kumpatla SP, Chandrasekharan MB, Hall TC. Transgene silencing in monocots. *Plant Mol Biol.* (2000) 43:323–46. doi: 10.1023/A:1006412318311

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Pandey, Madhu and Bhat. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 20 September 2019
doi: 10.3389/fgene.2019.00808

Combined QTL-Seq and Traditional Linkage Analysis to Identify Candidate Genes for Purple Skin of Radish Fleshy Taproots

Tongjin Liu[1†], Jinglei Wang[1,2†], Chunhui Wu[1], Youjun Zhang[1], Xiaohui Zhang[1], Xiaoman Li[1], Haiping Wang[1], Jiangping Song[1] and Xixiang Li[1*]*

[1] Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, Beijing, China, [2] Institute of Vegetables Research, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

OPEN ACCESS

Edited by:
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India

Reviewed by:
Bilal A. Padder,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India
Zhisheng Xu,
Nanjing Agricultural University, China
Ru Gang Yu,
Huaibei Normal University, China

***Correspondence:**
Youjun Zhang
zhangyoujun@caas.cn
Xixiang Li
lixixiang@caas.cn

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Genetics

Received: 18 March 2019
Accepted: 02 August 2019
Published: 20 September 2019

Citation:
Liu T, Wang J, Wu C, Zhang Y, Zhang X, Li X, Wang H, Song J and Li X (2019) Combined QTL-Seq and Traditional Linkage Analysis to Identify Candidate Genes for Purple Skin of Radish Fleshy Taproots. Front. Genet. 10:808. doi: 10.3389/fgene.2019.00808

Taproot skin color is a crucial visual and nutritional quality trait of radish, and purple skin is most attractive to consumers. However, the genetic mechanism underlying this character is unknown. Herein, F_2 segregating populations were constructed to investigate radish genomic regions with purple skin genes. Segregation analysis suggested that pigment presence was controlled by one dominant gene, *Rsps*. A bulk segregant approach coupled to whole-genome sequencing (QTL-seq) and classical linkage mapping narrowed the *Rsps* location to a 238.51-kb region containing 18 genes. A gene in this region, designated *RsMYB1.1* (an *Arabidopsis PAP1* homolog), was a likely candidate gene because semiquantitative RT-PCR and quantitative real-time PCR revealed *RsMYB1.1* expression in only purple-skinned genotypes, sequence variation was found between white- and purple-skinned radishes, and an InDel marker in this gene correctly predicted taproot skin color. Furthermore, four *RsMYB1.1* homologs (*RsMYB1.1-1.4*) were found in "XYB36-2" radish. *RsMYB1.1* and the previously mapped and cloned *RsMYB1.4* (contributing to red skin) were located on different chromosomes and in different subclades of a phylogenetic tree; thus, they are different genes. These findings provide insight into the complex anthocyanin biosynthesis regulation in radish and information for molecular breeding to improve the anthocyanin content and appearance of radish taproots.

Keywords: radish, purple taproot skin, inheritance, QTL-seq, candidate gene

INTRODUCTION

Radish (*Raphanus sativus*, 2n = 2x = 18), a member of the Brassicaceae (Cruciferae) family, is an important root vegetable grown and consumed throughout the world. In addition to its root, the leaf, siliques, and seed oil of some cultivars are consumed in some regions. The taproots of radish exhibit significant variation in shape, size, color, and flavor. There are many different taproot skin colors, including purple, red, pink, white, green, and even black and yellow.

The color of agricultural products is an important visual quality trait that directly affects the choice behavior of consumers. For this reason, a number of studies have attempted to identify major genes that control the color of different vegetables by map-based cloning. In cucumber,

two genes responsible for fruit color were identified by map-based cloning. A single-nucleotide insertion in *APRR2* was responsible for a white immature fruit color (Liu H. et al., 2016), while a mutation in MYB36 resulted in the formation of a yellow-green peel mutant (Hao et al., 2018). A 4-bp insertion in a putative R3 MYB repressor, *SlMYBATV*, led to anthocyanin accumulation in tomato fruit (Cao et al., 2017). A previous study mapped a purple leaf gene in ornamental kale to a 44.8-kb interval, and a dihydroflavonol reductase-coding gene was identified as a candidate gene (Liu et al., 2017). Through map-based cloning, *BnAPR2* was identified as a candidate gene responsible for the purple leaf trait in *Brassica napus* (Li et al., 2016). More recently, *RsMYB1*, an R2R3-MYB transcription factor, was identified in the red-skinned radish cultivar "Lian Yan No. 1" by map-based cloning (Yi et al., 2018). However, to date, no map-based cloning studies have been performed for the major genes responsible for the purple skin of radish taproots, which is an important visual and nutritional quality character.

In recent years, with an increase in people's consumption level, red and purple radishes have gained popularity because their enriched glucosinolates and anthocyanins have health benefits (Park et al., 2011). Developing radish cultivars with different skin colors has become an important breeding objective, and illuminating the mechanism of skin color formation will significantly accelerate progress in skin color breeding. Previous studies have indicated that anthocyanin accumulation results in the red/purple skin of radish (Park et al., 2011; Muleke et al., 2017). The anthocyanin biosynthesis pathway is relatively conserved in plants and has been extensively studied (Park et al., 2011; Zhang Y. et al., 2015; Zhang et al., 2015; Chen et al., 2016). The regulation of this process is extremely complex, and some members of the MYB, basic helix–loop–helix (bHLH), WD40, lateral organ boundary domain (LBD), WRKY, and NAC transcription factor families have been reported to be involved in its positive or negative regulation (Johnson et al., 2002; Koes et al., 2005; Matsui et al., 2008; Rubin et al., 2009; Stracke et al., 2010; Zhou et al., 2015). Recently, most structural genes involved in a multi-enzymatic pathway that controls anthocyanin biosynthesis were reported in radish (Park et al., 2011; Chen et al., 2016; Muleke et al., 2017; Sun et al., 2018). However, only two transcription factors (*RsMYB1* and *RsTT8*) that regulate anthocyanin accumulation in radish taproots identified by homology or map-based cloning have been reported (Lim et al., 2016; Lim et al., 2017; Yi et al., 2018).

Some early studies showed that a single locus or multiple loci may control root skin color and that red is dominant over white (Yarnell, 1956; He et al., 1997). He et al. (1997) indicated that red skin was probably controlled by three pairs of independent genes and that some genes controlling skin color likely affected flesh color inheritance (He et al., 1997). In contrast, Yi et al. (2018) reported that red skin was determined by a single dominant gene. These conflicting results most likely occur because the inheritance of radish root skin color is complex, and different inheritance mechanisms may be involved in different materials. The inheritance patterns and chromosomal locations of purple taproot skin genes in radish are still unclear. Therefore, the identification and cloning of major genes responsible for the purple skin of radish deserve further study.

In the present study, the radish inbred line CX16Q-25-2 (purple skin with white flesh) was crossed with the inbred line CX16Q-1-6-2 (white skin and flesh) to construct an F_2 population, which was used for fine genetic mapping of the genes for purple skin. Furthermore, genes located within the mapping region were analyzed, a candidate gene was identified and named *RsMYB1.1*, and four *RsMYB1.1* homologous genes distributed on three chromosomes of "XYB36-2" radish were identified and named *RsMYB1.1-1.4* according to their chromosome position. These findings will facilitate elucidation of the molecular mechanism for purple skin formation in radish taproots.

MATERIALS AND METHODS

Plant Materials

The white-skinned male radish inbred line CX16Q-1-6-2 (**Figure 1A**) was crossed with the purple-skinned female radish inbred line CX16Q-25-2 (**Figure 1B**) to generate the F_1 generation. F_2 populations (**Figure 1C**) were generated from self-pollination of a single F_1 plant and further self-pollinated to generate the $F_{2:3}$ population. All plants were grown at the Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS), Beijing, China.

Phenotyping and DNA Extraction

Skin color was identified visually 40 days after planting. Young leaves of each plant were collected, freeze dried, and ground into a fine powder for DNA extraction. Genomic DNA was extracted from 100 mg of leaf powder of each plant using the CTAB method (Doyle and Doyle, 1987). The quality of the DNA was assessed by 1% agarose gel electrophoresis and a NanoPhotometer (Implen, CA, USA), and the concentrations were measured by a Qubit DNA Assay Kit in a Qubit 2.0 Fluorometer (Life Technologies, CA, USA).

Total Anthocyanin Content Measurements

Total anthocyanins were extracted from 0.5 g of finely ground taproot skin from CX16Q-25-2, CX16Q-1-6-2, and different F_2 individuals (purple, red, pink, and white taproot skin) and evaluated by measuring the absorbance of the extract at 530 and 657 nm, as described by Chu et al. (2013). Anthocyanins were quantitated using the following formula: $Q = (A_{530} - 0.25 \times A_{657}) \times M^{-1}$, where Q is the anthocyanin content, A_{530} and A_{657} are the absorbance at the wavelengths 530 and 657 nm, respectively, and M is the weight of the sample. All samples were measured as triplicates in three independent biological replicates. Error bars indicate the standard error (SE) of the average of the anthocyanin content.

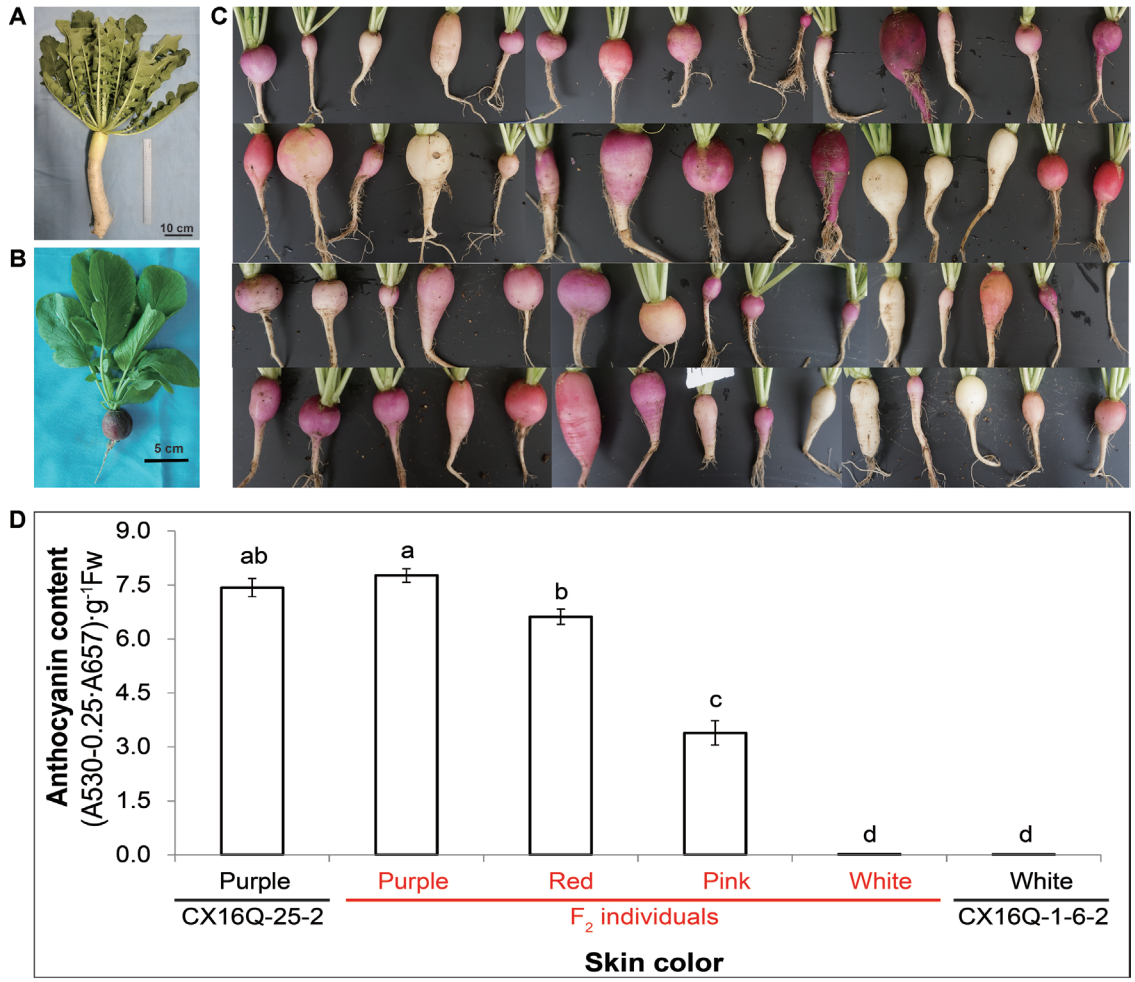

FIGURE 1 | Skin color of the parental lines and their derived F_2 population. **(A)** CX16Q-1-6-2; **(B)** CX16Q-25-2; **(C)** F_2 individuals of CX16Q-25-2 × CX16Q-1-6-2 divided into purple skins with varying shades and white skins; **(D)** relative total anthocyanin content of the skin of taproots with different skin colors.

Bulk Segregant Approach Coupled to Whole-Genome Sequencing (QTL-seq) Analysis

For the QTL-seq analysis, a purple-skinned DNA pool (P-pool) and a white-skinned DNA pool (W-pool) were constructed by mixing equal amounts of DNA from 40 purple-skinned and 40 white-skinned F_2 plants from the 2017 fall experiment, respectively. Four sequencing libraries, namely, P-pool, W-pool, and the two parental inbred lines (CX16Q-25-2 and CX16Q-1-6-2), were prepared for paired-end (150 bp) sequencing with the Illumina HiSeq platform. Clean reads were obtained by trimming adaptors of raw sequence reads (".fastq" format), removing reads with ambiguous "N" nucleotides exceeding 5% of the read length, and eliminating reads containing a number of bases with a Q value below 5 exceeding 50% of the total length. High-quality clean reads were used for QTL-seq analysis and insertion and deletion (InDel) marker development (Gordon, 2009).

A QTL-seq pipeline (http://genome-e.ibrc.or.jp/home/bioinformatics-team/mutmap, developed by the Iwate Biotechnology Research Center, Japan) was used to identify candidate genomic regions. Briefly, the cleaned reads of CX16Q-25-2 were first aligned to the reference genome "XYB36-2" (Zhang et al., 2015b) using an inbuilt BWA aligner (Li and Durbin, 2009). The reference-guided assembly of the CX16Q-25-2 parent was developed by substituting the bases with high-confidence nucleotides called for the CX16Q-25-2 parent. Then, the reads from the P and W bulks were aligned to the developed assembly, and variants were called for both bulks. The single nucleotide polymorphism (SNP) index for both bulks was calculated at each SNP position as suggested by Hiroki et al. (2013). The ΔSNP index was obtained by subtracting the SNP index of the P-pool from that of the W-pool. The average ΔSNP index located in a certain genomic interval was calculated using a sliding window analysis with a 4-Mb window size and 20-kb increment. Statistical confidence intervals of the ΔSNP index were calculated under the null hypothesis of no quantitative trait locus (QTL) following the description by Takagi et al., and 95% confidence intervals of the ΔSNP index were used to identify

genomic regions containing candidate major QTLs (Hiroki et al., 2013).

QTL Analyses With InDel Markers

QTL-seq analysis identified the major QTL for *Rsps* (purple skin). To fine map the *Rsps* gene, InDel markers were designed within the *Rsps* target region according to the polymorphism information of two resequenced parental inbred lines: CX16Q-25-2 and CX16Q-1-6-2. The short reads from parental inbred lines were aligned to the "XYB36-2" reference genome (Zhang et al., 2015b) with BWA software (Li and Durbin, 2009). InDel calling was performed by SAMtools software (Li and Durbin, 2009). Primer pairs were designed using Primer3 (Triinu and Maido, 2007), and specific primers were selected by alignment with the "XYB36-2" reference genome using electronic PCR (e-PCR) software (Schuler, 1997). Polymorphic markers were screened using the DNA of P-pool, W-pool, and the two parental inbred lines (CX16Q-25-2 and CX16Q-1-6-2) and then applied to the plants in the F2 population from the 2017 fall experiment.

PCR amplifications were performed in a 15-μl volume as described by Zhang et al. (2015a). The primers are listed in **Supplementary Table 1**. The amplified products were separated on an 8% polyacrylamide gel with 1× Tris/borate/EDTA (TBE) buffer at a constant voltage of 180 V for 1 h and then visualized with silver staining. The marker bands of individuals with the CX16Q-25-2, CX16Q-1-6-2, and F_1 alleles were recorded as "a", "b", and "h", respectively. JoinMap 4.0 was used to perform linkage analysis (Van Ooijen et al., 2002).

Candidate Gene Identification

To amplify the candidate gene, the DNA sequence was extracted from the radish genome (Zhang et al., 2015b), and a specific PCR primer was designed using Primer3web (Untergasser et al., 2012) (see **Supplementary Table 1**). PCR amplification was performed in a total volume of 50 μl using KOD-Plus-Neo polymerase (Toyobo, Osaka, Japan) according to the manufacturer's specifications with the two-step cycle procedure (Liu T. et al., 2016). The PCR products were subjected to Sanger sequencing using an ABI 3730 instrument (Applied Biosystems, CA, USA).

RNA Extraction, Semiquantitative RT-PCR, and Quantitative Real-Time PCR (qPCR) of the Candidate Gene

Total RNA was extracted from the taproot skin of CX16Q-25-2, CX16Q-1-6-2, and different F_2 individuals (purple, red, pink, and whitetaproot skin), treated with DNase and then reverse transcribed into cDNA as previously described (Liu T. et al., 2016). The specific primer was designed using Primer3web (Untergasser et al., 2012) and is shown in **Supplementary Table 1**. Semiquantitative RT-PCR for the candidate gene was carried out between CX16Q-25-2 and CX16Q-1-6-2 in a total volume of 50 μl using KOD-Plus-Neo polymerase with a T100 thermal cycler (Bio-Rad, Hercules, CA, USA). The following cycling parameters were used: 2 min of predenaturation at 94°C, followed by 35 cycles of denaturation (98°C, 10 s), annealing (63°C, 30 s), and extension (68°C, 30 s). Ten microliters of PCR product was then separated on a 2% agarose gel stained with GoldView. To investigate the pigment-controlling role of the candidate gene, qPCR analysis was performed in the taproot skin of parental lines and different F_2 individuals (purple, red, pink, and white). The cDNA was diluted fivefold as templates for qPCR using SuperReal PreMix Plus (TIANGEN BIOTECH CO., LTD, Beijing, China). The reaction volume was 20 μl, including 0.6 μl of forward and reverse primers at 10 mM each, 10 μl of 2 × SuperReal PreMix Plus, 2.0 μl of the cDNA sample, 2.0 μl of 50 × ROX Reference Dye, and 4.8 μl of ddH_2O. The thermal cycling profile was 95°C for 15 min; 40 cycles of 95°C for 10 s, 60°C for 30 s; then 95°C for 15 s, 60°C for 1 min, and ramping to 95°C for 15 s. Three independent biological and technical replicates were performed. Data were analyzed using StepOne™ Software v.2.0 (Applied Biosystems). Radish glyceraldehyde 3-phosphate dehydrogenase (*RsGADPH*) transcripts were used as a control for semiquantitative RT-PCR and an internal reference for qPCR (Liu et al., 2019).

Search for Homologs of the Candidate Gene

The predicted protein sequence of the candidate gene was used to search for corresponding homologs by BLASTP with previously published radish genome data (Zhang et al., 2015). Related protein sequences of other species were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov/). DNAMAN8 software (Lynnon, Quebec, Canada) was used to align sequences with the default parameters, and MEGA5 software was used to generate a phylogenetic tree according to the predicted protein sequences (Tamura et al., 2011).

RESULTS

Inheritance of the Purple Skin of Radish Taproots

To determine the inheritance of purple skin in radish, purple-skinned (CX16Q-25-2) and white-skinned (CX16Q-1-6-2) radish inbred lines were used to generate F_1 and F_2 populations (**Figure 1**). The fleshy taproot skin color of all F_1 individuals was purple (**Table 1**). The 557 F_2 individuals exhibited complex

TABLE 1 | Segregation of taproot skin color in the F_1 and F_2 populations.

Population	Plants tested	Pigmented: white	Mendelian expectation	χ2 value[a]	*p*
CX16Q-25-2 ×CX16Q-1-6-2					
F_1	all	Pigmented	1:0		
F_2	557	419:138	3:1	0.33	0.57

CX16Q-25-2 and CX16Q-1-6-2 are radish inbred lines with purple and white skin, respectively.
[a]*χ2 (0.05, 1) = 3.84*

skin color segregation including white, pink, red, and purple of different shades. The total anthocyanin content of the parental lines and different F2 radish roots are significantly different (**Figure 1D**). This result indicated that the genetics of purple taproot skin were complex, with multiple factors involved. However, the segregation was consistent with a 3:1 Mendelian segregation ratio for pigmented (pink, red, and purple with different shades) and white roots, which indicated that pigment presence was controlled by a single dominant gene, designated *Rsps* (purple skin). The present study was mainly focused on the identification of this dominant gene.

Preliminary Mapping of *Rsps* by QTL-seq

The two parental lines (CX16Q-25-2 and CX16Q-1-6-2) and two extreme pools (purple-skinned pool (P-pool) and white-skinned pool (W-pool)) from the F_2 population were paired-end (150 bp) sequenced with an Illumina HiSeq platform. In total, 5.6, 5.9, 23.4, and 30.0 Gb of clean reads from CX16Q-25-2 (11.2× depth coverage), CX16Q-1-6-2 (11.8× depth coverage), P-pool (46.8× depth coverage), and W-pool (60× depth coverage) were generated, respectively. These short reads were aligned to the "XYB36-2" reference genome, and 1,330,575, 1,130,541, 2,406,284, and 2,452,285 SNPs were identified between CX16Q-25-2 and "XYB36-2," CX16Q-1-6-2, and "XYB36-2," P-pool and "XYB36-2," W-pool and "XYB36-2," respectively. Subsequent QTL-seq analysis identified only one genomic region from 31.45 to 33.00 Mb on "XYB36-2" radish chromosome 2 that had a Δ(SNP index) value that was significantly different from 0 at the 95% confidence level (**Figure 2A**). These results indicated that there was a major QTL controlling purple skin in the 31.45- to 33.00-Mb region on chromosome 2 in radish (**Figure 2A**), which was designated *Rsps*.

QTL Analyses With InDel Markers

To verify the accuracy of QTL-seq in the detection of *Rsps* and narrow down its location, traditional QTL analysis was carried out with 352 F_2 plants. Three polymorphic InDel primer pairs were identified from markers developed in our previous studies (**Supplementary Table 1**). In addition, a total of 674 pairs of InDel primers were developed in the preliminary mapped regions with potential variation of >4 bp (**Supplementary Table 2**). Of the pairs, 52 were synthesized and used to screen polymorphic markers between CX16Q-25-2 and CX16Q-1-6-2 and between P-pool and W-pool, and five markers were determined to be useful (**Supplementary Table 1**). Then, these eight polymorphic markers were applied to the F_2 populations for QTL mapping. The *Rsps* locus was localized to a 1.3-cM interval between R02-7 and R02-24, flanking the gene at genetic distances of 0.6 cM and 0.7 cM, and this region matched that from QTL-seq (**Figure 2B**). Based on the marker locations of R02-7 and R02-24 in the reference genome of "XYB36-2," the 238.51-kb interval (31584238–31822749 bp) on chromosome 2 harbors the candidate gene for *Rsps*.

Candidate Gene Prediction for Purple Taproot Skin in Radish

There were 18 genes within the 238.51-kb candidate region in the "XYB36-2" reference genome (**Table 2**) (http://brassicadb.org/brad/datasets/pub/Genomes/Raphanus_sativus/) (Zhang et al., 2015b). Based on the functional annotations, *Rsa10008423* was predicted to be a member of the MYB family of transcription factors.

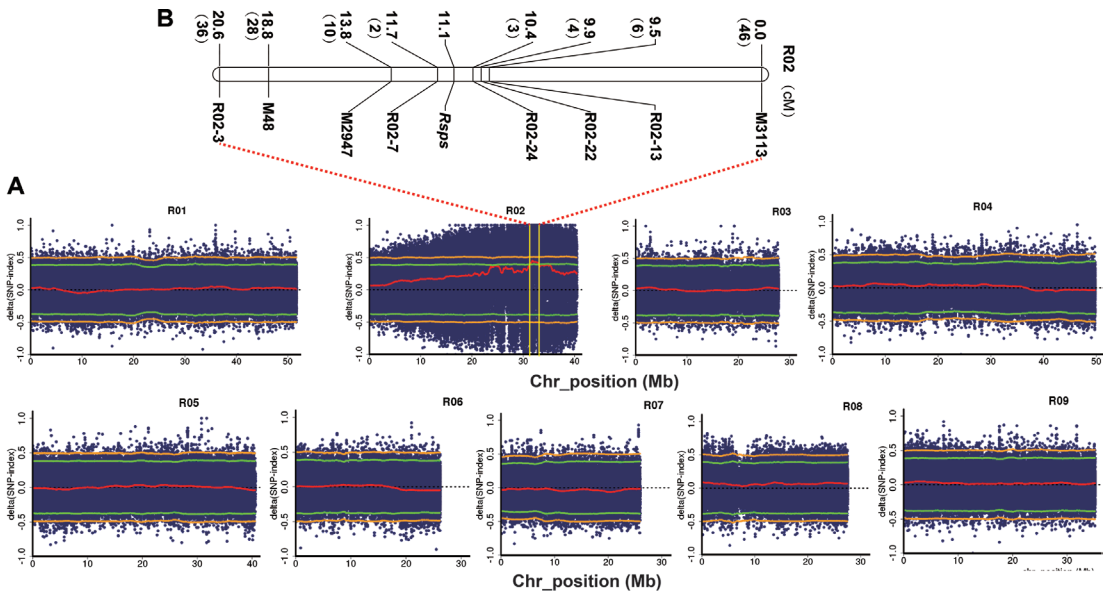

FIGURE 2 | Combined QTL-seq and traditional linkage analysis to map a purple skin locus (*Rsps*) in radish. **(A)** Δ(SNP index) plot of the purple and white bulks with confidence intervals under the null hypothesis of no QTLs (green, $p < 0.05$; orange, $p < 0.01$); **(B)** genetic linkage analysis with InDel markers confirming the position of *Rsps*. The numbers in brackets refer to the number of recombinants of each marker.

TABLE 2 | Annotation of radish genes in the candidate region.

Radish genes[a]	Gene position on R02[b]	AT ID[c]	E-value	AT GO annotation[d]
Rsa10008436	31582579-31584948 bp	AT1G66170	0.0	PHD domain-containing protein required for male meiosis
Rsa10008435	31595569-31596809 bp	AT1G66190	1e-52	Hypothetical protein
Rsa10008434	31601403-31606044 bp	AT1G07440	e-115	NAD(P)-binding Rossmann-fold superfamily protein
Rsa10008433	31618713-31620726 bp	AT1G66200	0.0	Cytosolic glutamate synthetase
Rsa10008432	31627059-31652390 bp	AT5G48720	1e-06	XRI1 (X-ray-induced 1)
Rsa10008431	31653659-31654085 bp	AT2G41520	7e-30	One of the 36 carboxylate clamp-tetratricopeptide repeat proteins
Rsa10008430	31678702-31680338 bp	AT3G46920	0.22	Kinase superfamily with an octicosapeptide/Phox/Bem1p domain-containing protein
Rsa10008429	31686259-31688854 bp	AT1G66250	0.0	O-glycosyl hydrolase family 17 protein
Rsa10008428	31689231-31689558 bp	AT2G01918	2e-24	Encodes a protein homologous to each PQL protein
Rsa10008427	31690647-31693371 bp	AT1G66330	e-137	Senescence-associated family protein
Rsa10008426	31696303-31698148 bp	AT1G66345	0.0	pentatricopeptide repeat protein involved in splicing of the nad4 intron, which affects biogenesis of the respiratory complex I
Rsa10008425	31714251-31715573 bp	AT1G17690	3e-31	U3 small nucleolar RNA-associated protein
Rsa10008424	31716933-31718411 bp	AT1G66350	0.0	Negative regulator of gibberellic acid responses, member of the GRAS family of transcription factors
Rsa10008423	31734491-31736148 bp	AT1G56650	1e-103	MYB domain-containing transcription factor involved in anthocyanin metabolism and radical scavenging
Rsa10008422	31737844-31741218 bp	AT1G65320	2e-29	Cystathionine beta-synthase family protein
Rsa10008421	31777128-31778529 bp	AT1G66470	1e-92	ROOT HAIR DEFECTIVE6
Rsa10008420	31805026-31806054 bp	AT2G29880	2e-65	Myb/SANT-like DNA-binding domain protein
Rsa10008419	31815194-31816694 bp	AT3G44840	e-125	S-adenosyl-L-methionine-dependent methyltransferase superfamily protein

[a]*Radish genes in the candidate region of the reference genome "XYB36-2".*
[b]*The physical position of genes in the candidate region of the reference genome "XYB36-2".*
[c]*The best hits for the genes in the candidate region compared to Arabidopsis thaliana (AT).*
[d]*Gene Ontology(GO) annotation of the radish genes in the candidate region matched to AT best-hit genes obtained from TAIR (https://www.arabidopsis.org/).*

Blast alignment results showed that the coding sequence (CDS) identity of *Rsa10008423* with the *Arabidopsis* PRODUCTION OF ANTHOCYANIN PIGMENT1 (*AtPAP1*, *AT1G56650*) gene was up to 82.40% (**Supplementary Figure1A**), and the predicted encoded proteins had 77.51% sequence identity (**Supplementary Figure 1B**). *AtPAP1* was reported to be involved in regulating the expression of anthocyanin biosynthetic genes (Borevitz et al., 2000). *Rsa10008423* is a homologous gene of *AtPAP1* in radish; therefore, *Rsa10008423* was designated as *RsMYB1.1*.

Analysis of the Candidate Gene

We investigated the expression patterns of *RsMYB1.1* with semiquantitative RT-PCR in two parental lines to analyze whether the expression level of *RsMYB1.1* may contribute to the variance in taproot skin color (**Figure 3**). The *RsMYB1.1* transcript was detected only in the taproot skin of purple-skinned radish (**Figure 3A**).

The DNA sequences of the two parental lines and the CDS of *RsMYB1.1* in CX16Q-25-2 were cloned, sequenced, and analyzed to investigate their sequence differences (**Figure 3B**). The alignment of *RsMYB1.1* DNA and CDSs revealed multiple mutations (**Figure 3B**). The DNA sequence is 1,610-bp long, with a 750-bp predicted open reading frame encoding a polypeptide of 249 amino acids. *RsMYB1.1* consists of three exons and two introns, and the first intron (which consists of 769 nucleotides) is significantly longer (8.5 times) than the second intron and represents 47.8% of the total length of the gene. The first and second exons are highly conserved between the two parental lines, while six SNP mutations were found in the third exon. In addition, two SNP mutations were found in the second intron. More notably, the first intron exhibited more SNP and InDel mutations between the two parental lines. Accordingly, an InDel marker was designed in this region, as shown in **Figure 3B**. Forty-six pigmented-skinned (pink, red, and purple with different shades) and 22 white-skinned radishes were selected from the F_2 population to determine the consistency between these marker genotypes and the taproot skin color phenotypes. Furthermore, the present InDel marker was validated in 17 randomly selected pigmented- and white-skinned radishes from the $F_{2:3}$ population. The results showed that all taproots were predicted correctly (**Figures 3C**, **D**).

The expression levels of *RsMYB1.1* in the taproot skin of both parents and F_2 individuals with purple, red, pink, and white skin were measured by qPCR. *RsMYB1.1* had significantly higher expression levels in pigmented-skinned individuals than in white taproot skin (**Figure 4**). Among pigmented-skinned individuals, purple-skinned plants had the significantly highest expression levels of *RsMYB1.1*, red-skinned plants had the second highest levels, and pink-skinned plants had the significantly lowest expression. The expression trend of *RsMYB1.1* in the taproot skin of parents and F_2 individuals correlated with their anthocyanin accumulation (**Figures 1** and **4**). These results suggest that *RsMYB1.1* may be a candidate gene controlling purple skin in CX16Q-25-2 radish.

MYB1 Genes Were Expanded in the Radish Genome

Four and three genes were identified by homology searches in "XYB36-2" (Zhang et al., 2015b) and "WK10039" (Jeong et al., 2016) protein datasets, respectively, with the deduced protein

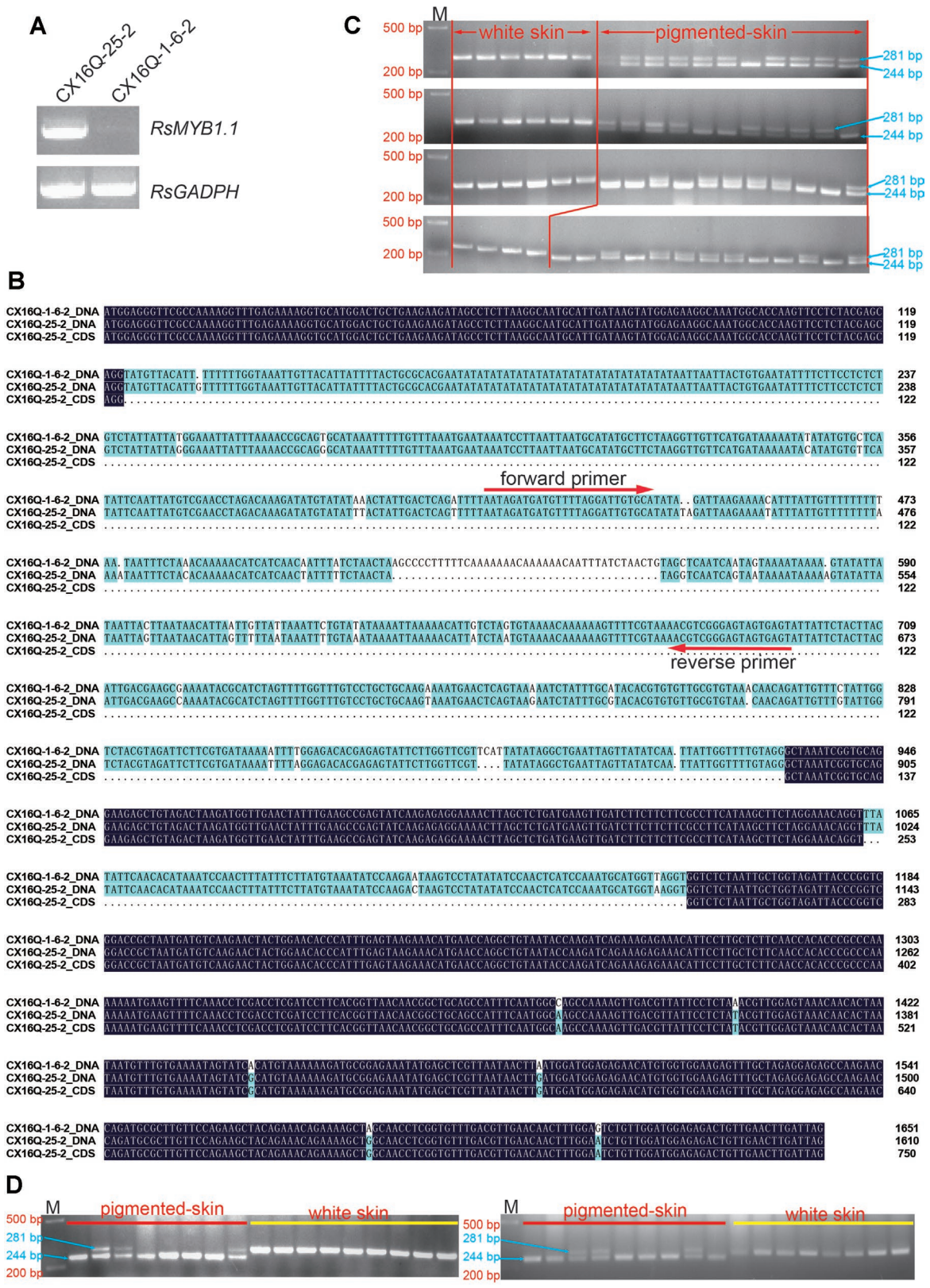

FIGURE 3 | Analysis of candidate genes. **(A)** Expression patterns of *RsMYB1.1* with semiquantitative RT-PCR in the two parental lines; **(B)** alignment of *RsMYB1.1* sequences of white and purple radishes; the positions of the forward and reverse primer pairs designed according to the parental sequence variation used to genotype the F_2 individuals are shown with arrows; **(C)** genotypes of F_2 individuals for PCR markers; **(D)** genotypes of $F_{2:3}$ individuals for PCR markers. M, size marker.

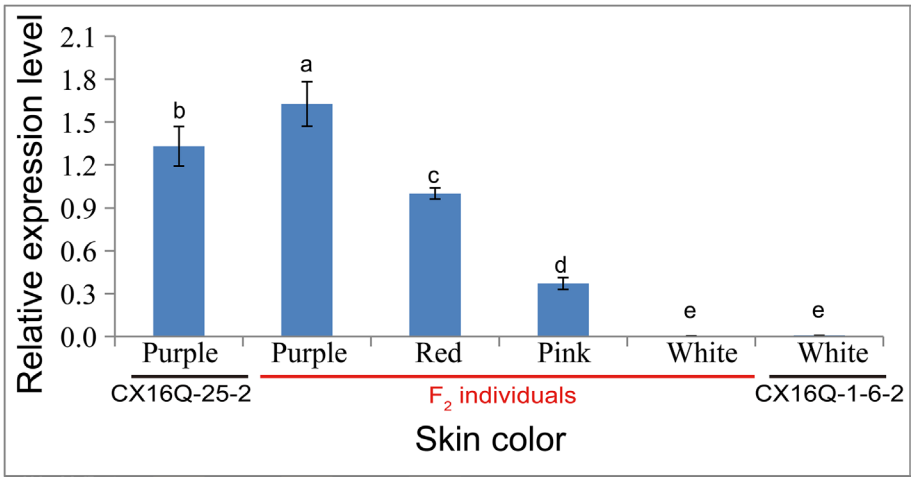

FIGURE 4 | Relative expression of *RsMYB1.1* in the skin of taproots with different skin colors detected by quantitative real-time PCR analysis. (a-e) indicate that the expression values were significantly different.

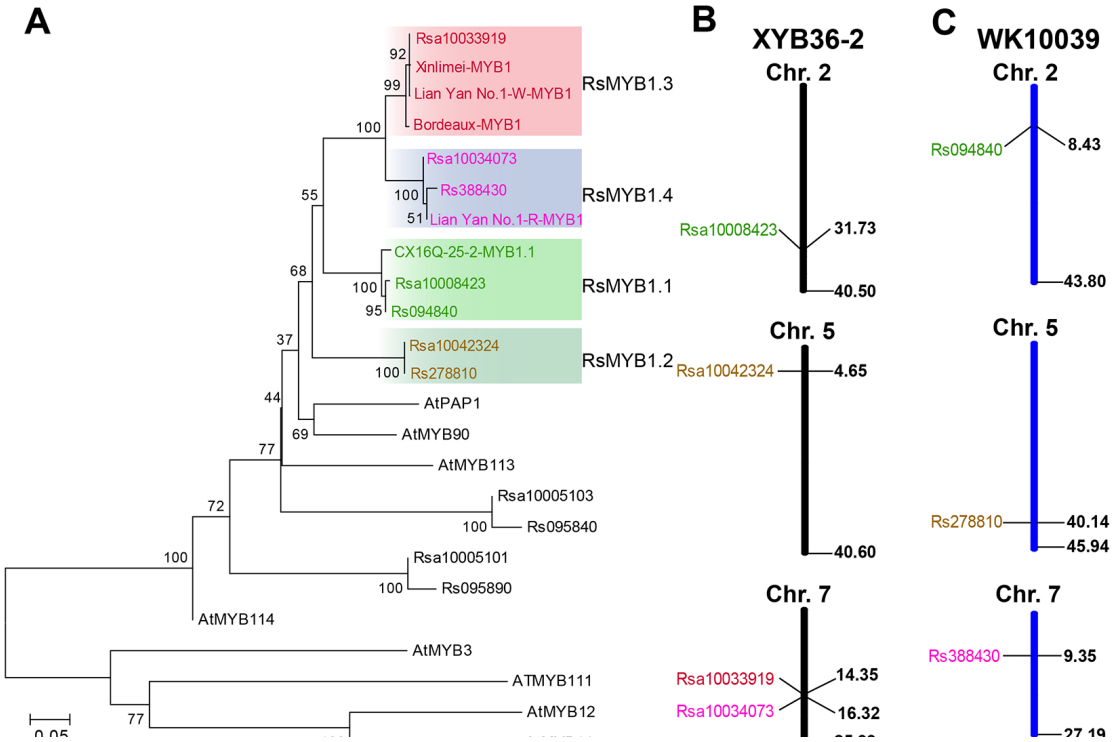

FIGURE 5 | Phylogenetic tree of RsMYB1.1 homologous genes in radish and anthocyanin R2R3-MYB regulators in *Arabidopsis* based on their amino acid sequences. **(A)** The chromosome locations of these genes in the "XYB36-2" **(B)** and "WK10039" **(C)** radish genomes. GenBank accession numbers are as follows: AtMYB11 (NP_191820), AtMYB12 (NP_182268.1), AtMYB75 (NP_176057.1), AtMYB90 (NM_105310), AtMYB111 (NP_199744), AtMYB114 (NP_176812), and Bordeaux-MYB1 (AKM95888.1).

sequence of *RsMYB1.1* from CX16Q-25-2 as a reference. All of these homologous genes were best matched to *Arabidopsis* PRODUCTION OF ANTHOCYANIN PIGMENT1 (*AtPAP1*, *AT1G56650*) in The Arabidopsis Information Resource (TAIR) database (https://www.arabidopsis.org/). In a phylogenetic tree of anthocyanin-regulated R2R3-MYB proteins from *Arabidopsis* and radish, four "XYB36-2" and three "WK10039" homologous genes and the published radish gene MYB1 all fell in the AtPAP1 subgroup (**Figure 5A**). Sequence alignments showed 96.79%, 99.60%, and 100% identity at the amino acid level between Rsa10034073 and Rs388430, Rsa10008423 and Rs094840, and Rsa10042324 and Rs278810, respectively (**Supplementary**

Figure 2), indicating that these three gene pairs were most likely the same gene in different radish cultivars. In addition, Rsa10033919 was identified in the "XYB36-2" radish genome but not in the "WK10039" genome (**Figures 5B, C**). This gene falls within the same subclade as the previously reported MYB1 genes from the "Xinlimei" (green skin with red flesh) and "Bordeaux" (red skin with red flesh) radish cultivars and red-skinned "Lian Yan No. 1" radish progeny, indicating that Rsa10033919 is a homolog of RsMYB1. Therefore, according to their chromosome locations in the "XYB36-2" genome, *Rsa10008423* (*Rs094840*), *Rsa10042324* (*Rs278810*), *Rsa10033919*, and *Rsa10034073* (*Rs388430*) were designated *RsMYB1.1*, *RsMYB1.2*, *RsMYB1.3*, and *RsMYB1.4*, respectively.

DISCUSSION

Fleshy taproot skin color is an important visual and nutritional quality trait of radish that influences radish production and consumption. The identification and cloning of major genes responsible for the skin color of radish will promote effectiveness in the breeding process. To date, possible inheritance patterns of radish taproot skin color have been reported by numerous studies (Yarnell, 1956; He et al., 1997). In different hybrid combinations, green root skin might be controlled by one or two pairs of independent genes or two pairs of linked genes (He et al., 1997). Previous research also indicated that a single locus or multiple loci may control red root skin, and red is dominant over white (Yarnell, 1956; He et al., 1997; Yi et al., 2018). In the present study, a cross between purple (CX16Q-25-2) and white (CX16Q-1-6-2) inbred lines produced F_2 plants with complex skin color segregation, which was consistent with a previous report that the genetics of purple coloration are complex and involve multiple factors (He et al., 1997). However, because the segregation of pigmented and white roots was consistent with a 3:1 Mendelian segregation ratio, pigment presence was controlled by a single dominant gene. This hypothesis was also verified by subsequent QTL-seq analysis results with two extreme pools (purple- and white-skinned pools) in which only one QTL region was detected. The results are consistent with the finding of a recent study that a single genetic factor distinguishes between red- and white-skinned progenies derived from the Chinese commercial F_1 cultivar "Lian Yan No. 1" (red skin with white flesh) (Yi et al., 2018). Therefore, our present study mainly focused on the identification of candidate genes located in this QTL region.

We used QTL-seq combined with classical methods to rapidly identify a major QTL, *Rsps*, which was localized to a 238.51-kb physical interval on chromosome 2 (**Figure 3**). Only 18 genes are located in this interval, based on the "XYB36-2" reference genome (**Table 2**). The present study indicated that combined QTL-seq, traditional linkage analysis, and high-quality whole-genome re-sequencing can significantly promote progress in gene mapping and subsequent cloning.

The taproot skins of some radish cultivars are purple and red because of the accumulation of a large concentration of anthocyanins (Park et al., 2011; Yi et al., 2018). The biosynthesis of anthocyanins is regulated by some transcription factors, especially members of the MYB gene family. Among the 18 genes located in the 238.51-kb candidate region of *Rsps* (**Table 2**), the expression pattern (**Figures 3A** and **4**), gene structure (**Figure 3B**), segregation analysis (**Figures 3C**, **D**), and phylogenetic analysis (**Figure 5A**) of *RsaMYB1.1* suggested that it could be a candidate gene for *Rsps*. RsaMYB1.1 shared 77.51% amino acid sequence identity with *Arabidopsis* AtPAP1 and was expressed in purple-skinned radish plants but not in white-skinned ones. There are multiple mutations in *RsaMYB1.1* between the two parental lines, and the marker designed according to a variable locus separated the purple-skinned and white-skinned radishes correctly (**Figures 3C**, **D**). Therefore, it is reasonable to postulate that *RsaMYB1.1* is the candidate gene for purple taproot skin in radish. However, further functional research is needed to validate the present findings. The MBW complex composed of R2R3-MYB, basic helix–loop–helix, and WD40 is highly conserved in plants involved in the regulation of anthocyanin biosynthesis (Koes et al., 2005). Among MBW complex components, R2R3-MYB plays a more important role because it is thought to directly bind to the promoter sequences of anthocyanin biosynthesis genes and activate their expression (Hichri et al., 2011). Numerous studies have reported that *MYB* homologs are involved in regulating the biosynthesis of anthocyanin in many plant species, including radish (Koes et al., 2005). A subgroup of *Arabidopsis* R2R3-MYBs including *AtPAP1* (*AtMYB75*), *AtPAP2* (*AtMYB90*), *AtMYB114*, and *AtMYB113* demonstrates a high degree of amino acid conservation and is involved in anthocyanin regulation (Allan et al., 2008). *MdMYBA*, *MdMYB1*, and *MdMYB10* are reportedly responsible for apple anthocyanin accumulation (Takos et al., 2006; Ban et al., 2007; Espley et al., 2007). *BoMYB2* is involved in controlling anthocyanin biosynthesis in purple cauliflower (Chiu and Li, 2012). *SmMYB1* is reported to participate in the regulation of anthocyanin biosynthesis in the peel of eggplant (Zhang et al., 2014). In radish, *RsMYB1* is reported to function in the activation of anthocyanin biosynthesis (Lim et al., 2016; Yi et al., 2018). Therefore, it is not surprising that an *AtPAP1* homologous gene was identified as the candidate gene responsible for the purple skin of CX16Q-25-2 radish.

A previous study suggested that the taproot color of radish was determined by the ratio of pelargonidin and cyanidin, which are responsible for the red and purple coloration, respectively, and the *H* locus was proposed to be involved in the biosynthesis of pigments from pelargonidin to cyanidin (Yi et al., 2018). However, our present research did not detect a QTL region responsible for this trait. The most likely reason for this lack of detection was that the purple or red skin trait in the present F_2 population was easily affected by environmental conditions. The F_2 population generated by red- and purple-skinned radish inbred lines may be ideal for mapping the *H* locus.

Four *RsMYB1.1* homologs were identified in the "XYB36-2" radish genome and named *RsMYB1.1*, *RsMYB1.2*, *RsMYB1.3*, and *RsMYB1.4* (**Figure 5B**). *RsMYB1.1* was first reported in the present study and may function in the regulation of anthocyanin biosynthesis in radish taproot skin. The function of *RsMYB1.2* is still unknown and requires further research. In the phylogenetic analysis, the RsMYB1.3 subclade included the previously reported *RsMYB1* gene from the "Bordeaux" homozygous F_3 progeny (Lim et al., 2016) and "Xinlimei" (Liu et al., 2019), both of which have red flesh. These results indicate that RsMYB1.3 plays an important role in the regulation of anthocyanin biosynthesis in radish taproot

flesh. More recently, another *RsMYB1*gene was identified in the red-skinned radish cultivar "Lianyan No. 1" by map-based cloning (Yi et al., 2018). This gene falls in the RsMYB1.4 subclade with Rsa10034073 from "XYB36-2" and Rs388430 from "WK10039". *RsMYB1.4* and *RsMYB1.1* are located on different chromosomes and fall into different subclades, indicating that they are two different genes, but they might have the same function in the regulation of anthocyanin biosynthesis in radish taproot skin. The identification of different genes previously in "Lianyan No. 1" and in CX16Q-25-2 in the present study may be attributable to the fact that different cultivars were used in mapping candidate genes. Further research is needed to confirm the role of these genes, which would enrich the research on anthocyanin biosynthesis regulation in radish.

DATA AVAILABILITY

Sequences data generated in this study are available on the Sequence Read Archive http://www.ncbi.nlm.nih.gov/Traces/sra/, at accession PRJNA563272.

AUTHOR CONTRIBUTIONS

XXL conceived and supervised the study. TL, JW, and CW analyzed the data and drafted the manuscript. TL, XXL, HW, and JS performed the experiments. JW, TL, and XML helped with mapping population construction and sample collection. XXL, YZ, XZ, and JW revised the manuscript. All authors have read and approved the manuscript.

FUNDING

This work was supported by a grant from the National Key Research and Development Program of China (2016YFD0100204-2), the National Natural Science Foundation of China (31801858), the China Postdoctoral Science Foundation Funded Project (2017M620971), and the Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-XTCX2016016-4-4, CAAS-XTCX2016001-5-2, and CAAS-XTCX2016017).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00808/full#supplementary-material

SUPPLEMENTARY TABLE 1 | The primers used in the present study.

SUPPLEMENTARY TABLE 2 | InDel primers developed in the preliminary mapping regions with potential variation of >4 bp.

SUPPLEMENTARY FIGURE 1 | Alignment of radish Rsa10008423 with the Arabidopsis PAP1 (MYB75, AT1G56650) gene coding sequence (a) and their deduced amino acid sequences (b).

SUPPLEMENTARY FIGURE 2 | Alignment of radish Rsa10034073 and Rs388430 (a), Rsa10008423 and Rs094840 (b), and Rsa10042324 and Rs278810 (c) based on their deduced amino acid sequences.

REFERENCES

Allan, A. C., Hellens, R. P., and Laing, W. A. (2008). MYB transcription factors that colour our fruit. *Trends Plant Sci.* 13, 99–102. doi: 10.1016/j.tplants.2007.11.012

Ban, Y., Honda, C., Hatsuyama, Y., Igarashi, M., Bessho, H., and Moriguchi, T. (2007). Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. *Plant Cell Physiol.* 48 (7), 958–970. doi: 10.1093/pcp/pcm066

Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* 12, 2383–2393. doi: 10.1105/tpc.12.12.2383

Cao, X., Qiu, Z., Wang, X., Giang, T., Liu, X., Wang, J., et al. (2017). A putative *R3 MYB* repressor is the candidate gene underlying *atroviolacium*, a locus for anthocyanin pigmentation in tomato fruit. *J. Exp. Bot.* 68, 5745–5758. doi: 10.1093/jxb/erx382

Chen, F. B., Xing, C. Y., Huo, S. P., Cao, C. L., Yao, Q. L., and Fang, P. (2016). Red pigment content and expression of genes related to anthocyanin biosynthesis inradishes (*Raphanus sativus* L.) with different colored flesh. *J. Agr. Sci.* 8, 126–135. doi: 10.5539/jas.v8n8p126

Chiu, L. W., and Li, L. (2012). Characterization of the regulatory network of BoMYB2 in controlling anthocyanin biosynthesis in purple cauliflower. *Planta* 236, 1153–1164. doi: 10.1007/s00425-012-1665-3

Chu, H., Jeong, J. C., Kim, W. J., Chung, D. M., Jeon, H. K., Ahn, Y. O., et al. (2013). Expression of the sweet potato R2R3-type *IbMYB1a* gene induces anthocyanin accumulation in *Arabidopsis. Physiol. Plantarum.* 148 (2), 189–199. doi: 10.1111/j.1399-3054.2012.01706.x

Doyle, J., and Doyle, J. (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bull.* 19, 11–15.

Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., and Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* 49 (3), 414–427. doi: 10.1111/j.1365-313X.2006.02964.x

Gordon, A. (2009). Fastx-toolkit. http://hannonlab.cshl.edu/fastx_toolkit/.

Hao, N., Du, Y., Li, H., Wang, C., Wang, C., Gong, S., et al. (2018). *CsMYB36* is involved in the formation of yellow green peel in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* 131, 1659–1669. doi: 10.1007/s00122-018-3105-7

He, Q., Zhao, S., Shi, H., An, Z., Lang, F., and Wang, S. (1997). Preliminary study on skin color inheritance of Chinese radish. *J. Shandong Agric. Sci.* 2, 4–9. doi: 10.14083/j.issn.1001-4942.1997.02.001

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J. Exp. Bot.* 62, 2465–2483. doi: 10.1093/jxb/erq442

Hiroki, T., Akira, A., Kentaro, Y., Shunichi, K., Satoshi, N., Chikako, M., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

Jeong, Y. M., Kim, N., Ahn, B. O., Oh, M., Chung, W. H., Chung, H., et al. (2016). Elucidating the triplicated ancestral genome structure of radish based on chromosome-level comparison with the Brassica genomes. *Theor. Appl. Genet.* 129, 1357–1372. doi: 10.1007/s00122-016-2708-0

Johnson, C. S., Kolevski, B., and Smyth, D. R. (2002). *Transparent testa GLABRA2*, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. *Plant Cell* 14, 1359–1375. doi: 10.1105/tpc.001404

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* 10, 236–242. doi: 10.1016/j.tplants.2005.03.002

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Zhu, L., Yuan, G., Heng, S., Yi, B., Ma, C., et al. (2016). Fine mapping and candidate gene analysis of an anthocyanin-rich gene, *BnaA.PL1*, conferring

purple leaves in *Brassica napus* L. *Mol. Genet. Genomics* 291, 1523–1534. doi: 10.1007/s00438-016-1199-7

Lim, S. H., Kim, D. H., Kim, J. K., Lee, J. Y., and Ha, S. H. (2017). A radish basic helix–loop–helix transcription factor, *RsTT8* acts a positive regulator for anthocyanin biosynthesis. *Front. Plant Sci.* 8, 1917. doi: 10.3389/fpls.2017.01917

Lim, S. H., Song, J. H., Kim, D. H., Kim, J. K., Lee, J. Y., Kim, Y. M., et al. (2016). Activation of anthocyanin biosynthesis by expression of the radish R2R3-MYB transcription factor gene *RsMYB1*. *Plant Cell Rep.* 35, 641–653. doi: 10.1007/s00299-015-1909-3

Liu, H., Jiao, J., Liang, X., Liu, J., Meng, H., Chen, S., et al. (2016). Map-based cloning, identification and characterization of the *w* gene controlling white immature fruit color in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* 129, 1247–1256. doi: 10.1007/s00122-016-2700-8

Liu, T., Zhang, X., Yang, H., Agerbirk, N., Yang, Q., Wang, H., et al. (2016). Aromatic glucosinolate biosynthesis pathway in *Barbarea vulgaris* and its response to *Plutella xylostella* infestation. *Front. Plant Sci.* 7, 83. doi: 10.3389/fpls.2016.00083

Liu, T., Zhang, Y., Zhang, X., Sun, Y., Wang, H., Song, J., et al. (2019). Transcriptome analyses reveal key genes involved in skin color changes of 'Xinlimei' radish taproot. *Plant Physiol. Biochem.* 139, 528–539. doi: 10.1016/j.plaphy.2019.04.006

Liu, X., Gao, B., Han, F., Fang, Z., Yang, L., Zhuang, M., et al. (2017). Genetics and fine mapping of a purple leaf gene, *BoPr*, in ornamental kale (*Brassica oleracea* L. var. acephala). *BMC Genomics* 18, 230. doi: 10.1186/s12864-017-3613-x

Matsui, K., Umemura, Y., and Ohme-Takagi, M. (2008). AtMYBL2, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in Arabidopsis. *Plant J.* 55, 954–967. doi: 10.1111/j.1365-313X.2008.03565.x

Muleke, E. M., Fan, L., Wang, Y., Xu, L., Zhu, X., Zhang, W., et al. (2017). Coordinated regulation of anthocyanin biosynthesis genes confers varied phenotypic and spatial–temporal anthocyanin accumulation in radish (*Raphanus sativus* L.). *Front. Plant Sci.* 8: 1243 doi: 10.3389/fpls.2017.01243

Park, N. I., Xu, H., Li, X., Jang, I. H., Park, S., Ahn, G. H., et al. (2011). Anthocyanin accumulation and expression of anthocyanin biosynthetic genes in radish (*Raphanus sativus*). *J. Agric. Food Chem.* 59, 6034–6039. doi: 10.1021/jf200824c

Rubin, G., Tohge, T., Matsuda, F., Saito, K., and Scheible, W. R. (2009). Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in Arabidopsis. *Plant Cell* 21, 3567–3584. doi: 10.1105/tpc.109.067041

Schuler, G. D. (1997). Sequence mapping by electronic PCR. *Genome Res.* 7, 541. doi: 10.1101/gr.7.5.541

Stracke, R., Jahns, O., Keck, M., Tohge, T., Niehaus, K., Fernie, A. R., et al. (2010). Analysis of Production of flavonol glycosides-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. *New Phytol.* 188, 985–1000. doi: 10.1111/j.1469-8137.2010.03421.x

Sun, Y., Wang, J., Qiu, Y., Liu, T., Song, J., and Li, X. (2018). Identification of 'Xinlimei' radish candidate genes associated with anthocyanin biosynthesis based on a transcriptome analysis. *Gene* 657, 81–91. doi: 10.1016/j.gene.2018.03.001

Takos, A. M., Jaffé, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., and Walker, A. R. (2006). Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* 142 (3), 1216–1232. doi: 10.1104/pp.106.088104

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Triinu, K., and Maido, R. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291. doi: 10.1093/bioinformatics/btm091

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40, e115. doi: 10.1093/nar/gks596

Van Ooijen, J. W., Boer, M. P., Jansen, R. C., and Maliepaard, C. (2002). MapQTl4.0, software for the calculation of QTL positions on genetic maps. *Plant Res. Int. Wagenigen*.

Yarnell, S. H. (1956). Cytogenetics of the vegetable crops. II. Crucifers. *Bot. Rev.* 22, 81–166. doi: 10.1007/BF02872468

Yi, G., Kim, J. S., Park, J. E., Shin, H., Yu, S. H., Park, S., et al. (2018). MYB1 transcription factor is a candidate responsible for red root skin in radish (*Raphanus sativus* L.). *PLoS One* 13, e0204241. doi: 10.1371/journal.pone.0204241

Zhang, X., Liu, T., Wei, X., Qiu, Y., Song, J., Wang, H., et al. (2015a). Expression patterns, molecular markers and genetic diversity of insect-susceptible and resistant *Barbarea* genotypes by comparative transcriptome analysis. *BMC Genomics* 16, 486. doi: 10.1186/s12864-015-1609-y

Zhang, X., Yue, Z., Mei, S., Qiu, Y., Yang, X., Chen, X., et al. (2015b). A *de novo* genome of a Chinese radish cultivar. *Hortic. Plant J.* 1, 155–164. doi: 10.16420/j.issn.2095-9885.2016-0028

Zhang, Y., Hu, Z., Chu, G., Huang, C., Tian, S., Zhao, Z., et al. (2014). Anthocyanin accumulation and molecular analysis of anthocyanin biosynthesis-associated genes in eggplant (*Solanum melongena* L.). *J. Agric. Food Chem.* 62, 2906–2912. doi: 10.1021/jf404574c

Zhang, Y., Hu, Z., Zhu, M., Zhu, Z., Wang, Z., Tian, S., et al. (2015). Anthocyanin accumulation and molecular analysis of correlated genes in purple kohlrabi (*Brassica oleracea* var. gongylodes L.). *J. Agric. Food Chem.* 63, 4160–4169. doi: 10.1021/acs.jafc.5b00473

Zhou, H., Lin-Wang, K., Wang, H., Gu, C., Dare, A. P., Espley, R. V., et al. (2015). Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *Plant J.* 82, 105–121. doi: 10.1111/tpj.12792

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Liu, Wang, Wu, Zhang, Zhang, Li, Wang, Song and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 November 2019
doi: 10.3389/fgene.2019.00946

Single Nucleotide Polymorphisms in Starch Biosynthetic Genes Associated With Increased Resistant Starch Concentration in Rice Mutant

Selvakumar Gurunathan[1,2], Bharathi Raja Ramadoss[3], Venkataramana Mudili[2], Chandranayaka Siddaiah[4], Naveen Kumar Kalagatur[2], Jutti Rajendran Kannan Bapu[1], Chakrabhavi Dhananjaya Mohan[5], Abdulaziz A. Alqarawi[6], Abeer Hashem[7] and Elsayed Fathi Abd_Allah[6]*

[1] Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, India, [2] DRDO-BU-Centre for Life Sciences, Bharathiar University Campus, Coimbatore, India, [3] Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada, [4] Department of Studies in Biotechnology, University of Mysore, Mysore, India, [5] Department of Studies in Molecular Biology, University of Mysore, Mysore, India, [6] Department of Plant Production, College of Food and Agricultural Sciences, King Saud University, Riyadh, Saudi Arabia, [7] Botany and Microbiology, Department, College of Science, King Saud University, Riyadh, Saudi Arabia

OPEN ACCESS

Edited by:
Manish Kumar Pandey,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

Reviewed by:
Larry Parnell,
Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, United States
Vikas Kumar Singh,
International Rice Research Institute, Philippines

***Correspondence:**
Selvakumar Gurunathan
selvakumaragri@gmail.com

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Genetics

Received: 25 January 2019
Accepted: 05 September 2019
Published: 15 November 2019

Citation:
Gurunathan S, Ramadoss BR, Mudili V, Siddaiah C, Kalagatur NK, Bapu JRK, Mohan CD, Alqarawi AA, Hashem A and Abd_Allah EF (2019) Single Nucleotide Polymorphisms in Starch Biosynthetic Genes Associated With Increased Resistant Starch Concentration in Rice Mutant. Front. Genet. 10:946. doi: 10.3389/fgene.2019.00946

Resistant Starch (RS), plays a crucial role in human health and nutrition by controlling glucose metabolism. RS or dietary fibre content in rice is low because it goes through a variety of process before it is ready for cooking and consumption. Hence, this study was carried out to develop a rice mutant with increased RS. The rice mutant (γ278) with increased RS was developed by utilizing gamma (γ) rays as a mutagen. Mutant γ278 was characterized for mutations in the starch biosynthetic genes *viz*., *GBSSI, SSI, SSIIa, SSIIIa, SBEIa*, and *SBEIIb* to reveal the functional mutations/variations led to high RS content in rice. A total of 31 sequence variants/mutations in six genes were identified. We report the discovery of three deleterious mutation/variants each in *GBSSI, SSIIa*, and *SSIIIa* with the potential to increase RS content in rice. Further, *wild* × *mutant* crosses were made to develop an F_2 population to study the effect of combination of deleterious mutations. The SNP (*GBSSI*:*ssIIa*:*ssIIIa*) combination responsible for high RS content in F_2 population was identified and recorded highest amylose content (AC) (26.18%) and RS (8.68%) content. In conclusion, this marker combination will be highly useful to develop a rice variety with increased RS.

Keywords: amylose, resistant starch, starch biosynthesis, rice, mutation/variants

INTRODUCTION

Diabetes mellitus (DM) is a major global threat and its prevalence is increasing at an alarming rate worldwide, especially in Asian countries (Cummings and Englyst, 1991; Chan et al., 2009). About 415 million adults suffered with type 2 diabetes mellitus (T2DM) globally during 2015, which is predicted to increase to 642 million by 2040 (Herman, 2017). Apart from obesity and physical inactivity, another major cause of high T2DM prevalence is the consumption of rapidly digestible carbohydrate-rich foods capable of increasing the blood sugar concentration. "Rice is life" for human being especially in the Asian subcontinent, where 90% of world's rice is grown and consumed by 60% of the population (Khush and Virk, 2000). Although rice is a major cereal

globally, efforts to address or enhance the nutritional qualities of rice through crop improvement programs are rare. Rice grain lacks several micronutrients including vitamins and rapidly releases energy after consumption, therefore, has a relatively higher glycemic response compared to other starchy foods (Chassy et al., 2008).

Earlier dietary carbohydrates used to be derived from whole coarse grains loaded with sufficient dietary fibre however, they are currently replaced with refined carbohydrates devoid of any dietary fibre, e.g. modern day rice resulting from advanced milling technology (Chattopadhyay, 2005). It is known that refined carbohydrate-rich diet enhances the plasma glucose level, insulin, triglycerides (TAG), and non-esterified fatty acid (NEFA) thus play a significant role to insulin resistance (Wolever and Mehling, 2003). Due to lifestyle changes, cells of the body did not respond properly to the hormone insulin which precedes the development/onset of T2DM. In a recent systemic review based on four prospective cohort studies with a total of 13,284 incident cases of T2DM among 352,384 Asian participants, Hu et al. (2012) concluded that the higher consumption of white rice was associated with a significant increase T2DM risk. In contradiction, the scientific investigation from Miller et al. (1992) indicated that all processed forms of rice (white, brown, or parboiled) fell under the category of high-glycemic index (GI foods). This signifies the necessity to develop rice variety(ies) with reduced starch hydrolysis or glycemic index utilizing molecular breeding approaches.

Starch is a complex glucose polymer stored as distinct water insoluble granules in rice grains (~90%) primarily composed of one-quarter amylose and three-quarters amylopectin (Ramadoss et al., 2019). Starch is divided into three fractions based on the enzymatic hydrolysis, readily digestible (RDS), slowly digestible (SDS), and resistant starch (RS) (Englyst and Hudson, 1996). Amylose, an important determinant of RS formation as it showed a positive correlation with amylose (Ramadoss et al., 2015), is more resistant to enzymatic hydrolysis compared to its counterpart amylopectin (Leeman et al., 2006; Lehmann and Robin, 2007) and therefore, is being widely used to enhance the RS content of processed foods**.** Therefore, amylose concentration in starch can be targeted to improve RS content in the rice grains. Increasing amylose and RS content in rice endosperm is envisaged as a potential target to enhance its starch quality and promote human health and nutrition (Fuentes-Zaragoza et al., 2010). RS has a significantly lower hydrolysis index compared to other fractions of starch and free glucose moieties thus assist to normalize blood sugar levels in diabetic people. In addition, RS is fermented by large intestinal microflora into short chain fatty acids promoting the growth of beneficial bacteria and lowering the pH of the intestinal microenvironment thus alleviating the risk of gastrointestinal (GI) disorders. Extensive studies also have shown that RS possess physiological functions like those of dietary fiber (Eerlingen and Delcour, 1995).

Three genetic mechanisms have been described to increase amylose and RS content in cereals. Interestingly, the specific nature and impact of each mechanism can differ among cereals despite their close relatedness. The first mechanism includes overexpression of granule-bound starch synthase I (*GBSSI*) to increase amylose content (AC) (Flipse et al., 1996; Umemoto et al., 2004). The second mechanism to increase amylose accumulation in the developing grain is to down-regulate *SSIIa* or *SSIIIa* activity and decease amylopectin biosynthesis. Moreover, suppressed expression of *SSIIIa* has been associated with up-regulation of *GBSSI* and therefore, to increased amylose content (Umemoto et al., 2008; Sestili et al., 2012). The most prominent third mechanism, to increase apparent AC in cereals, targets suppression of major starch branching enzyme (*SBEIIb)* activity as structural changes in starch granule may have a great impact on RS content, particle size distribution and crystallinity. Mutation in starch synthases (*SSIa*, *SSIIa*, and *SSIIIa)* produce altered amylopectin chain length distribution in rice endosperm (Fujita et al., 2007). The suppression of starch branching enzyme *SBEI* alone has either no detectable effect (Regina et al., 2004), or only a minor effect on amylopectin chain length. Moreover, suppression of *SBEIIb* leads to elevated AC in maize and rice with altered amylopectin chain length distribution (Butardo et al., 2011) and suppression of both *SBEIIb* and *SBEI* in rice led to high AC (Zhu et al., 2012). Previous studies have been focused on the identification of mutations and characterization to understand the role of individual mutations in starch biosynthesis. This study was aimed to discover mutations responsible for increased RS content in the identified mutant line γ278 and to understand the role/contribution of individual mutations and/or combination of mutation towards trait improvement. Consequently, the present study characterized a high RS rice mutant line (mutant γ278), developed through gamma mutagenesis and biochemical analysis. The developed mutant was characterized for key candidate genes in the starch biosynthesis (*GBSSI, SSI, SSIIa, SSIIIa, SBEIa,* and *SBEIIb*) which is responsible for high RS content and to find out the sequence variants/mutations between the high RS mutant (γ278) and its wild type ADT43. Further, we analyzed the functional sequence variants/mutations for their contribution on the trait of interest through the development of segregating population (F_2) of "*wild type* × *mutant*" cross and revealed the role of gene-specific mutations towards an increase in rice RS content.

MATERIALS AND METHODS

Plant Material and Mutational Dose Determination and Screening

The cultivar ADT 43 was chosen for inducing mutations which is a short duration variety derived from a cross between IR 50 and Improved White Ponni (IWP). To create an ADT43 rice mutant population, gamma radiation (Tamil Nadu Agricultural University-Radiation facility) was used and LD_{50} value (250 Gy) was determined using kill curve analysis. Fixed LD_{50} was used to treat and to develop the 5,000 M_1 plants (Agasimani et al., 2013). Then the 3,000 M_2 families were raised and three plants from each family have been tagged. The seeds were harvested in a single plant basis and a portion of seeds was processed to white rice and used for biochemical analysis (data not shown). The RS content of M_2 to M_6 generation was estimated on dry weight basis following Goni et al. (1996)

using the Megazyme RS assay kit (Cat#K-RSTAR; Megazyme International Ireland Ltd., Ireland). Identified mutant (γ278) line with increased RS was used for further molecular studies to identify the functional mutations.

DNA Isolation

Total genomic DNA was extracted from the leaves using DNeasy 96 Plant kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions.

Candidate Gene Selection and Gene Model Construction

The existing literature was reviewed to select the candidate genes associated with starch composition and quality (Rahman et al., 2000; Nakamura, 2002; Hirose et al., 2006; Waters and Henry, 2007). Based on these literatures, a total of six key candidate starch synthesis pathway genes including *Granule Bound Starch Synthase I* (*GBSSI*), *Starch Synthase I (SSI), Starch Synthase IIa (SSIIa), Starch Synthase IIIa (SSIIIa), Starch Branching Enzyme Ia (SBEIa),* and *Starch Branching Enzyme IIb (SBEIIb)* were undertaken to reveal the influence of target gene mutations in RS enhancement (**Table 1**). Targeted gene sequences were retrieved from the NCBI genome database (http://www.ncbi.nlm.nih.gov/Genbank) gene model was constructed by using Splign (https://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi) bioinformatics pipeline.

Target Gene Amplification and Mutation Identification

By using the target gene sequences and expressed sequence tags (EST), a total of seven primer pairs were designed using Primer3 software (Primer3; https://www.ncbi.nlm.nih.gov/tools/primer-blast/), and used to amplify the target genes. Primer sequences and target amplicon lengths were presented in **Table 2**. PCR reactions (50 µl) were performed using 10 µl of 5× long AMP Taq Reaction buffer, 1.5 µl of dNTPs (10 mM), 2 µl of each forward and reverse primer (10 µM), 2µl of DMSO, 5 µl of DNA (50ng/µl), 2 µl of Long Amp® Taq polymerase (5 unit), and 25.5 µl sterile water. Step down PCR cycling was performed using ABI 2720 (Applied Biosystems, Foster City, CA) thermal cycler.

Site-Specific Gene Sequencing and Mutant Validation

Upon successive amplification of target genes, site-specific target gene sequencing was carried out using Ion Torrent Next Generation Sequencing (NGS) platform with 1Gb data output using Ion 316™ Chip by following the manufacturer's instruction. After the sequencing of libraries, aligning of sequencing information were carried out by using Torrent Suite 1.5 with wild type (ADT 43) reference sequences. After sequence alignment, variant caller bioinformatic pipeline was used with selected parameters such as min–max distance, mismatch cost, length fraction, and similarity were selected to pull out rare SNPs from

TABLE 1 | Details of selected candidate genes with their function used in this study.

S no.	Gene	ID of mRNA	Genomic ID	Size of the gene (kb)	Function
1.	Granule Bound Starch Synthase (*GBSSI*)	AB425323.1	NC_008399.2	3.479	Amylose biosynthesis
2.	Starch Synthase I (*SSI*)	AK109458.1	NC_008399.2	6.815	Amylopectin synthesis
3.	Starch Synthase IIa (*SSIIa*)	AB115918.1	NC_008399.2	4.419	Amylopectin synthesis
4.	Starch Synthase IIIa (*SSIIIa*)	GQ151020.1	AK061604	11.083	Amylopectin synthesis
5.	Starch Branching Enzyme Ia (*SBEIa*)	AK065121.1	NC_008399.2	4.745	Amylopectin synthesis
6.	Starch Branching Enzyme IIb (*SBEIIb*)	D16201.1	NC_008395.2	10.899	Amylopectin synthesis

TABLE 2 | Details of gene specific primers designed to amplify selected candidate genes.

S no.	Gene	Order	Primer Sequence (5′ to 3′)	Amplicon size (Kb)
1	Granule Bound Starch Synthase (*GBSSI*)	F	TTCATCTGATCTGCTCAAAG	5.525
		R	CCAGAAGAGTACAACATCAAAC	
2	Starch Synthase I (*SSI*)	F	ATCACTTCACAAACCCATAAC	7.604
		R	GAAAGACAGGAAGATTGAGG	
3	Starch Synthase IIa (*SSIIa*)	F	AAAGTAACTCGCTTCTGGAG	5.112
		R	AAGAAGTAACATCGCATCAAT	
4	Starch Synthase IIIa (*SSIIIa*)	F	CTTCTATGCCCTCGGAGCAG	11.018
		R	ACATAGCTGATACATACTCCCA	
5	Starch Branching Enzyme Ia (*SBEIa*)	F	GGTGACTGTTGTGGAGGAGG	4.730
		R	CGTCAGAAGACCGAAACACA	
6	Starch Branching Enzyme IIb (*SBEIIb*)	F1	AGCACAGGAGTAGCAAGTAG	6.770
		R1	AACTACTGCATCAGCATCAG	
		F2	GCACCAAGTAGTCGTTTCGG	6.508
		R2	TGGAGCATAGACAACGCAGG	

F, Forward primer; R, Reverse primers.

the aligned sequence contigs in comparison with their wild type reference sequences. The minimum variant frequency of 0.5 and minimum coverage of 20 was used to detect rare SNPs which gives variants on or above 0.5% were considered as SNPs.

Identified SNP sequence variants were analyzed by the PARSESNP program (Taylor and Greene, 2003), which provides information on the location along with the details about amino acid changes and severity of mutations. SIFT (http://sift.jcvi.org/www/SIFT_seq_submit2.html.) a web-based tool which predicts whether amino acid substitution affects protein function and structure based on sequence homology and the physical properties of amino acids (Ng and Henikoff, 2003) was employed to discover functional mutations. The predicted SIFT score ranges from 0 to 1. The amino acid substitution is predicted to be damaging if the score is <0.05 and tolerated if the score is >0.05.

Starch Structural Analysis

Starch structural analysis was carried out using a scanning electron microscope (SEM) (FEI, USA) as described by Fujita et al. (2003).

Trait Association and SNP Analysis Through Molecular Breeding

Wild Type × Mutant Crosses

For trait association, F_1 plants were developed by crossing γ278 with wild type ADT 43. Five F_1 plants were selected and tagged for SNP genotyping with G→A transition in *SSIIa* gene to identify the true hybrid. SNP genotyping was carried out using KASPar technology (LGC genomics, Ipswich, UK). The true F_1 plants (possessing the SNP markers in a heterozygous state) were self-pollinated and produced 538 F_2 plants.

SNP Genotyping of F_2 of "*Wild* × *Mutant*" Cross

Mutant segregation analysis and their contribution towards the trait of interest were analyzed with 192 F_2 plants of the "*wild type* × *mutant*" crosses. The observed and expected distribution of mutation in the F_2 plants were analyzed using simple excel workbook suggested by Montoliu (2012). SNP genotyping of F_2 plants was carried out using KASPar technology (LGC genomics, Ipswich, UK) to determine the SNPs. Seeds of each tagged plants were harvested and stored for further study.

Estimation of AC and RS Concentration

The RS **concentration** of grain samples from the 28 F_2 plants representing various allelic combination was estimated along with their parent traits including ADT43 and γ278 (Goni et al., 1996) using the Megazyme RS analysis kit. AC estimation was carried out using the method explained earlier (Raja et al., 2017b).

RESULTS

Mutational Dose Determination and Screening

LD_{50} value of γ-radiation was determined as 250 Gray for ADT 43 rice variety. Fixed LD_{50} was used to treat and to develop the mutant population (Agasimani et al., 2013). The M_2 mutant population was analyzed for their RS and AC (data not shown). Results of RS estimation further revealed that there is a significant change in the RS concentration ranged from 3.61 ± 0.4% to 6.62 ± 0.4% (data not shown). Further F_2 analysis revealed that the rice mutant named γ278 showed a maximum RS concentration of 8.63 ± 0.53%. Additional molecular characterization studies were undertaken to reveal the genes and its mutations responsible for quantitative up-gradation of RS concentration in the identified rice mutant line γ278.

Target Gene Amplification and Mutation Identification

PCR conditions were optimized for selective amplification of target genes. Various PCR cycling conditions and master mix composition were standardized for the amplification of different candidate gene in the mutant and wild type genotype. After successful amplification (**Figure 1**), PCR products of wild type and mutants were pooled across the target genes and then barcoded for easy interpretation of the sequence analysis. The average read depths were recorded in the range of 432 to 969. Sequencing results revealed that target gene *SBEIa* recorded with a maximum average read depth of 969 followed by *SSIIa* (879), *GBSSI* (824), *SSIIIa* (709), *SSI* (689), and *SBEIIb* (432). Mutation with higher confidence levels in the target genes were identified using variant caller and the mutations were filtered with the quality score of 60 and above. From 47.28 kb of six candidate genes, 31 (28 SNPs and three Indels) variants were identified (**Table S1**). Eight variants were discovered in *SSI* (seven SNPs and one Indel), *SSIIIa* (eight SNPs) followed by five each in *GBBSI* (four SNPs and one Indel), and *SBEIIb* (four SNPs and one Indel). Three SNP variants were discovered in *SSIIa*. *SBEIa* recorded two SNP variants and all of them were found to be in intron.

Further position analysis revealed that five sequence variants were identified in *GBSSI*, two were in exon region and rest was detected in the intron region. In the case of *SSIIa*, out of three sequence variants identified, two were found in exon regions. Interestingly, a total of eight sequence variants were found in *SSIIIa*, out of which four were found in exon regions and the rest of them were in the intron regions.

Functional Validation of SNPs

Functional analysis of the exon residing mutations indicated that 62.5% were silent and 32.5% were deleterious mutations. The discovered exon residing mutations in the γ278 along with their SIFT score were furnished in **Table 3**. Two exon residing mutations observed in the GBSS I gene, SIFT analysis of exon residing SNPs revealed that only one SNP at position 2078 (C→T) of the reference sequence in ninth exon of the gene is a functional mutation leading to Proline to Serine substitution at 415th amino acid residue with a SIFT score of 0.00. This functional mutation in C→T (2078) lead to the creation of restriction site *Hpy188I* and deletion of *ApaI, AsuI, DraII, HaeIII*, and *NlaIV* restriction sites further indicated the possibility of generating CAPS/dCAPS marker. On the other hand, the sequence variation T→C (1804) was found to be a silent variation with the SIFT score of 1.00, as it did not cause any amino acid change.

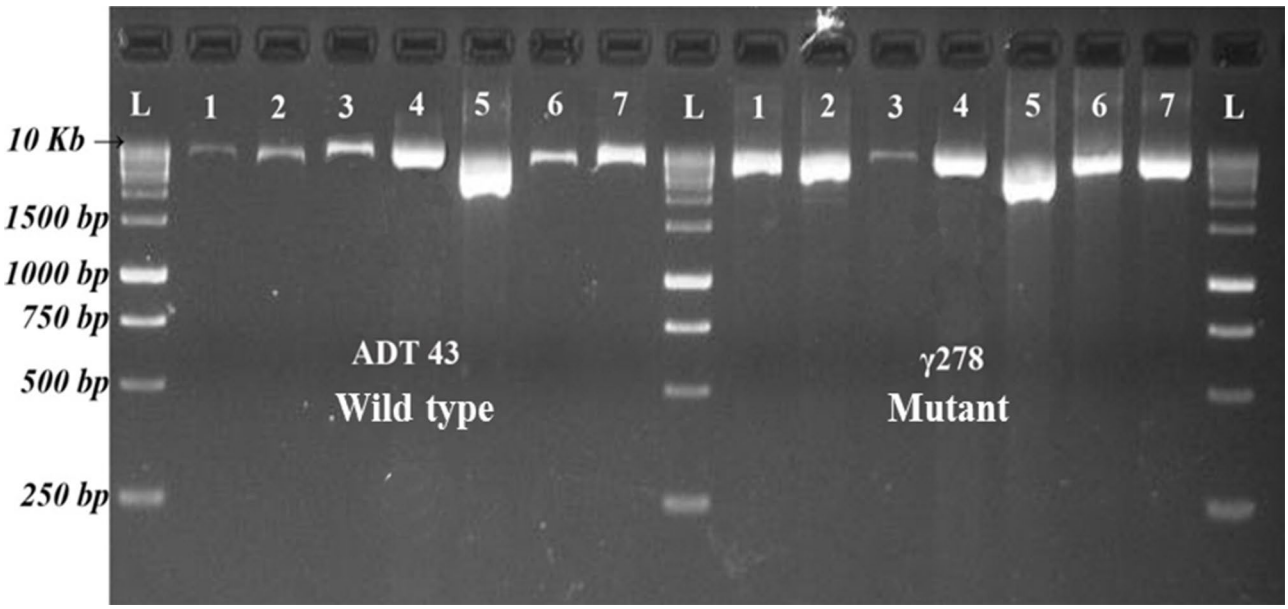

FIGURE 1 | Resolved PCR products of candidate genes in ADT43 and γ278. L = Size Marker, 1 = *GBSS I*, 2 = *SS IIa*, 3 = *SS IIIa*, 4 = *SS I*, 5 = *SBE Ia*, 6 = *SBE IIb* Fragment 1, and 7 = *SBE IIb* Fragment 2.

TABLE 3 | Deleterious mutations identified in the selected candidate genes.

S. no.	Candidate gene	Nucleotide change[a]	Effect on protein sequence[b]	Restriction site		SIFT Score	Effect of mutation
				Gained in variant	Lost from reference		
1	*GBSSI*	T1804C	P362=	*BseYI*	*MaeI*	1.00	Silent
		C2078T*	P415S	*Hpy188I*	*ApaI, AsuI, DraII, HaeIII, NlaIV*	0.00	Deleterious
2	*SSIIa*	G3797A*	G604S	*Eco57MI, GsuI*	*HpaII*	0.00	Deleterious
		G3901T	G638=	–	–	–	Silent
3	*SSIIIa*	C1615T*	A195V	–	*AciI, SfaNI*	0.00	Deleterious
		T2276C	H415=	–	*AvaIII, NlaIII*	1.00	Silent
		C3135A	R702=	–	–	1.00	Silent
		G5515A	L1256=	–	–	1.00	Silent

[a] *The letter on the left side of the numeral specifies the wild type nucleotide; letter on the right side of the numeral specifies the altered nucleotide; numeral specifies the base pair position of the nucleotide change with respect to the gene sequence.*

[b] *Signifies a synonymous change; the numeral specifies the residual number of the amino acid based on the gene model; letter on the left of the numeral specifies wild type amino acid; letter on the right side of the numeral specifies the altered amino acid.*

* *Indicates that the nucleotide change resulted in deleterious mutation.*

The two identified exon variants (G→A at 3797 and G→T at 3901) of *SSIIa*, were found in the region of eighth exon. The mutation G→A triggered an amino acid substitution Glycine to Serine at the position of 604 in protein sequence and caused the intolerant mutation and showed the deleterious effect on the protein function with a sift score of 0.00. On the other hand, the mutational sequence variation G→T at 3901 did not yield any amino acid change in the target protein sequence.

Four exon residing mutations (position of nucleotide change) C→T (1615), T→C (2276), C→A (3135), and G→A (5515) were identified in *SSIIIa*. Among them, three found in the region of third exon and rest was present in the fifth exon. Out of four mutations discovered in the third exon, C→T (1615) change resulted in amino acid substitution of Alanine to Valine at the position of 195 and showed deleterious effect with a SIFT score of 0.00 and rest of the variants did not show any significant change in the amino acid sequence. C→T variation further leads to the deletion of *AciI* and *SfaNI* restriction sites without the addition of new restriction site.

Phenotypic Discrimination and Grain Structural Analysis

To study the effect of mutations in the grain phenotype, observations were recorded on morphology based on visual observation a significant difference in the mutant and wild type

grains (**Figure 2A**). Also, to study the effect of mutation on starch structural properties, SEM analysis of mutant and wild type grains were studied (**Figure 2B**). SEM analysis revealed that there is a significant structural alteration in the starch granule arrangement of mutant grain in comparison with its wild type (**Figure 2B a, b, c**). The structural changes were recorded as granules with loosely packed in nature and presence of numerous airspaces between the granules (**Figure 2B d**) in mutant grain (γ278), whereas, fewer airspaces with compact granule arrangement were observed in the wild type (ADT43). Furthermore, γ278 was recorded with smaller, round-shaped starch granules accumulated in the amyloplast of endosperm cells, in contrast to the sharp-edged polygonal granules in the ADT-43 (**Figure 2B e**). Moreover, the phenotypic observations revealed that a higher proportion of smaller granule size coupled with white core belly which is absent in the wild type genotype ADT 43.

Role of Mutation in RS Enhancement

The role of identified mutations in RS enhancement as individually and as in combination was determined by employing "*wild type* × *mutant*" hybridization to generate the F_1 population. The true F_1 plants were confirmed through KASPar–SNP genotyping. A total of 192 F_2 plants were taken for genotyping to study the effect of individual mutation. Interestingly, all the studied mutations are in line with the Mendelian segregation ratio (1:2:1) in 192 F_2 plants (**Table 4**). Amylose and RS concentration analysis of F_2 plants harboring the 8 pre-determined mutant SNP marker combinations revealed that out of eight homozygous marker combinations (**Table 5**), two of them recorded with higher mean RS content than the mutant γ278, whereas five of them recorded intermediate mean RS concentration between wild type and mutant γ278. Moreover, the highest mean concentration of amylose (26.18%) and RS (8.68%) were recorded in F2 plants with the marker combinations of CC : AA:TT among the target genes of *GBSSI: SSIIa: SSIIIa* respectively and followed by CC : GG:TT combination which is recorded 25.23% of AC and 7.79% of RS concentration (**Table 5**). The marker combination TT : AA:TT recorded with RS concentration of 7.03% even it having lower AC than the marker combinations CC : AA:CC (RS — 6.19%) and TT : GG:TT (RS — 5.74%). The lowest value for AC (20.89%) and RS concentration (2.27%) were recorded in the marker combinations of TT : GG:CC.

DISCUSSION

The existing cereals contributing a major portion to the daily diet lack the appropriate concentration of RS to meet out the per day requirement, therefore, daily intake of RS is much lesser than the

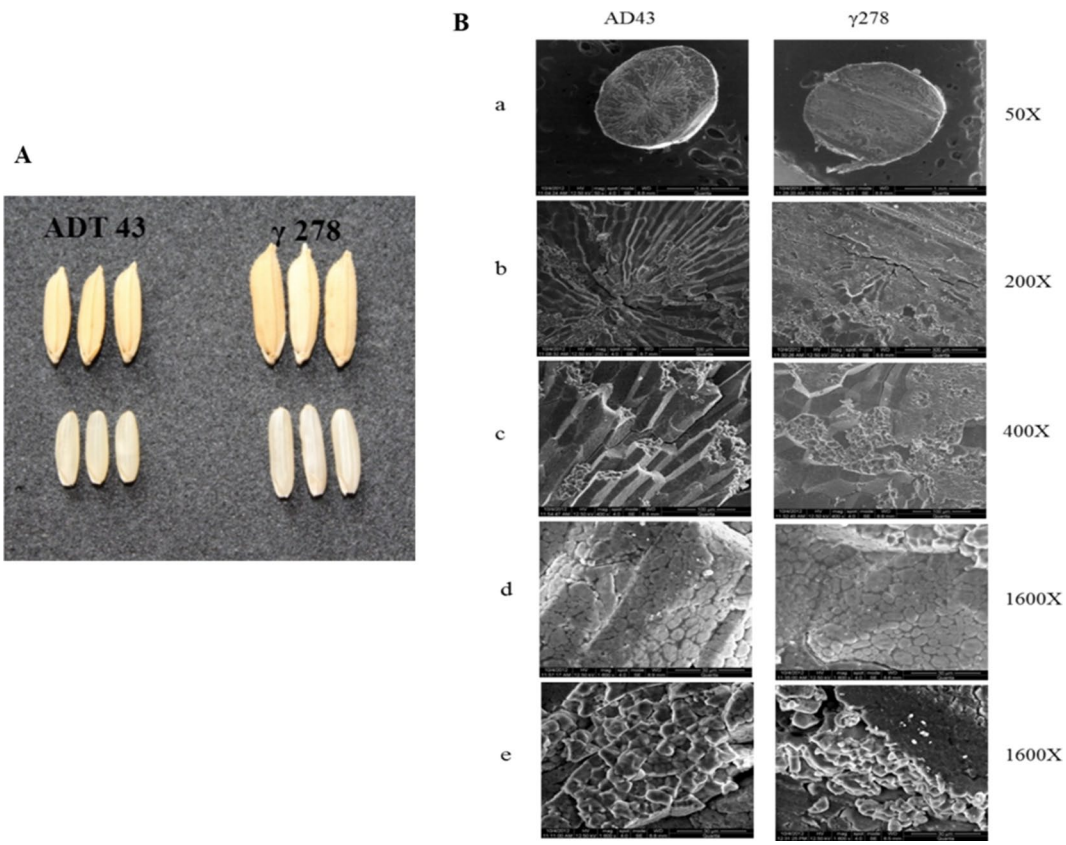

FIGURE 2 | Phenotypic and starch grain variation in the wild type and mutant grains. **(A)** Phenotypic variation of mutant and wild type genotype. **(B)** Scanning electron micrograph of starch in the mutant and wild type grains in different magnification (a-e).

TABLE 4 | Chi Square table representation of SNP segregation in target genes.

GBSSI					*SSIIa*				*SSIIIa*			
	Expected	**Observed**	**O–E**	**Chi square value**	**Expected**	**Observed**	**O–E**	**Chi square value**	**Expected**	**Observed**	**O–E**	**Chi square value**
Wild	48 (1)	58	10	2.08	48 (1)	51	3	0.18	48 (1)	49	1	0.02
Heterozygous	96 (2)	99	3	0.09	96 (2)	102	6	0.37	96 (2)	99	3	0.09
Mutant	48 (1)	35	−13	3.52	48 (1)	39	−9	1.68	48 (1)	44	−4	0.33
Total	192	192	0	5.69^{NS}	192	192	0	2.25^{NS}	192	192	0	0.44^{NS}

^{NS} No significant deviation.

minimum recommended levels (6 g *per* meal) for health benefits (Dian et al., 2005). Currently, researchers and nutritionists have shown increased emphasis on research exploring the starches resistant to human digestive enzymes. Rice is one of the major staple cereals with low RS concentration utilized among the globe. Introducing an important trait such as RS in rice grains will assist to prepare food products rich in RS/dietary fibre, alleviating the burden of T2DM, and improve the socio-economic condition of the society. Earlier researchers postulated that RS concentration is positively correlated with amylose concentration (AC) in cereal crops (Benmoussa et al., 2007; Sang et al., 2008; Ramadoss et al., 2015) however, it would be disadvantageous to the cooking quality and palatability. Alternatively, other starch properties such as granule size, architecture, crystalline pattern, the degree of crystallinity, surface pores, or channels and degree of polymerization also influence starch digestibility (Tester et al., 2006; Noda et al., 2008). However, amylose concentration and amylopectin chain length variation influence the formation of RS concentration in rice endosperm (Fujita et al., 2007; Nakamura et al., 2010). Mutation/nucleotide changes in starch biosynthetic genes (*GBSSI, SSI, SSIIa, SSIIIa, SBEIa*, and *SBEIIb*) lead to variation in the starch quality/quantity in rice endosperm. Therefore, the present study was undertaken to study the effect of allelic variation in the target genes which contribute towards the enhancement of RS concentration in rice endosperm.

Exposing seeds to the radiation is one of the well-known approaches in mutation breeding. To know the effect of radiation on the change of gene structure and enzyme function, we characterized important candidate genes which are responsible for increasing the RS fraction of starch in the rice grains by adopting PCR and re-sequencing mutation/variant discovery method through *in silico* and *in vivo* approaches.

NGS sequencing results together with PARSESNP analysis revealed that the mutation C→T (2078) might cause an amino acid substitution of proline to serine at the 415 position in amino acid sequence of *GBSSI* gene. Mutations in *GBSSI* showed reduced activity by affecting starch-binding capacity (Liu et al., 2014), and its ADP–glucose-binding capability or its protein stability. Moreover, Delrue et al. (1992) reported that the Chlamydomonas mutant lacking *GBSSI* activity generated amylose free starch with the altered structure of amylopectin. Therefore, a deleterious mutation in *GBSSI* in the present study was expected to have lesser AC in γ278 mutant than wild type, but surprisingly the determined AC in γ278 was found to be more than the corresponding wild type. This could be explained by the mutations in the key candidate genes (*SSIIa* and *SSIIIa*) of amylopectin biosynthesis that might be responsible for the altered ratio of amylose and amylopectin which may lead to an incremental effect on AC than amylopectin in the total starch.

Mutation G→A (3797) in *SSIIa* caused an amino acid substitution of Glycine to Serine at the position of 604 in the mutant γ278. The same amino acid substitution (G604S) was reported to result in amylopectin chain length variation between *japonica* (Serine) and *indica* (Glycine) types (Nakamura et al., 2005). Among the four exon residing mutations identified in *SSIIIa* gene, the variant C→T (1615) cause amino acid substitution of Alanine to Valine at the position of 195 with deleterious in

TABLE 5 | Amylose and RS concentrations of selected F2 plants with different combinations of mutations discovered in this study.

S no.	Marker combinations	Number of plants	Well position	Plant	AC (%)	Mean AC (%)	RS (%)	Mean RS (%)
1.	CC : GG:CC	4	H01	F2-8	22.36 ± 0.6	22.35 ± 0.5	3.90 ± 0.2	3.80 ± 0.2
	GBSSI: SSIIa: SSIIIa		F06	F2-46	22.79 ± 0.8		3.81 ± 0.1	
			B02	F2-106	22.11 ± 0.5		3.68 ± 0.3	
			H05	F2-136	22.14 ± 0.4		3.84 ± 0.2	
2.	CC : GG:TT	5	H06	F2-48	25.25 ± 0.3	25.23 ± 0.4	7.82 ± 0.4	7.79 ± 0.3
	GBSSI: SSIIa: ssIIIa		G08	F2-63	25.16 ± 0.7		7.79 ± 0.3	
			F09	F2-70	25.31 ± 0.4		7.64 ± 0.5	
			B12	F2-90	25.21 ± 0.2		7.94 ± 0.2	
			H03	F2-120	25.23 ± 0.4		7.77 ± 0.1	
3.	*CC : AA:TT*	2	E05	F2-133	26.03 ± 0.5	26.18 ± 0.6	8.71 ± 0.4	8.68 ± 0.3
	GBSSI: ssIIa: ssIIIa		D09	F2-164	26.34 ± 0.8		8.65 ± 0.3	
4.	CC : AA:CC	6	F08	F2-62	24.89 ± 0.2	24.61 ± 0.4	6.23 ± 0.2	6.19 ± 0.4
	GBSSI: ssIIa: SSIIIa		D09	F2-68	24.46 ± 0.4		6.31 ± 0.3	
			D03	F2-116	24.42 ± 0.6		6.04 ± 0.4	
			G07	F2-151	24.67 ± 0.4		6.27 ± 0.2	
			C10	F2-171	24.48 ± 0.6		6.13 ± 0.5	
			C12	F2-187	24.74 ± 0.5		6.18 ± 0.6	
5.	TT : AA:CC	2	F10	F2-174	23.34 ± 0.3	23.42 ± 0.4	5.75 ± 0.4	5.58 ± 0.4
	gbssI: ssIIa: SSIIIa		E11	F2-181	23.51 ± 0.5		5.42 ± 0.5	
6.	TT : GG:CC	4	A02	F2-9	20.68 ± 0.2	20.89 ± 0.3	2.44 ± 0.1	2.27 ± 0.3
	gbssI: SSIIa: SSIIIa		C08	F2-59	21.10 ± 0.2		2.01 ± 0.5	
			D11	F2-84	20.87 ± 0.5		2.24 ± 0.3	
			H11	F2-184	20.93 ± 0.3		2.41 ± 0.4	
7.	TT : GG:TT	2	G08	F2-159	24.33 ± 0.5	24.40 ± 0.4	5.67 ± 0.5	5.74 ± 0.4
	gbssI: SSIIa: ssIIIa		H12	F2-192	24.47 ± 0.3		5.81 ± 0.4	
8.	TT : AA:TT	3	F01	F2-6	23.51 ± 0.6	23.48 ± 0.4	6.98 ± .05	7.03 ± 0.5
	gbssI: ssIIa: ssIIIa		D01	F2-100	23.43 ± 0.4		7.10 ± 0.4	
			G01	F2-103	23.52 ± 0.2		7.02 ± 0.5	
9.	ADT 43					22.17 ± 0.2		3.61 ± 0.4
10.	γ 278					23.40 ± 0.3		7.26 ± 0.4

nature. Raja et al. (2017a) also reported the same mutation in rice found to have increased RS concentration. Previous reports also indicated that the loss of *SSIIIa* function leads to accumulation of white-core floury endosperm in rice (Fujita et al., 2007). Further, it was found that the accumulation of white core floury endosperm is due to enrichment of amylose and the loose packing of starch granules in the mutants (Fujita et al., 2007). Similar results were also reported by Yoshii (2000) and Nagato and Ebata (1958).

Upon successive mutant detection, SNP genotyping and trait association analysis was carried out to reveal the role of each mutation with respective to RS enhancement in the wild *type* × *mutant* derived F_2 population. Highest RS concentration of 8.68% was recorded in the marker combination of *GBSSI*: *ssIIa*: *ssIIIa.* Here we discovered wild type *GBSSI* and defective *ssIIa* and *ssIIIa* enhanced the RS level up to 8.68%. Zhou et al. (2016) reported active *GBSSI* and defective *SSIIIa* leads to high RS formation in rice endosperm. Marker combination *GBSSI : SSIIa*:*ssIIIa* was recorded RS value of 7.79%. Mutation free GBSS*I* resulted in enhanced AC, conversely mutations in *SSIIa* and *SSIIIa* leads to the formation of more short chains in amylopectin led to enhanced RS concentration. Similar results were also reported in rice by Sasaki et al. (2009). On the other hand, in spite of its lower AC, the marker combination of *gbssI:ssIIa:ssIIIa* recorded with 7.03% RS, than the marker combinations of *GBSSI:ssIIa : SSIIIa* (RS — 6.19%) and *gbssI: SSIIa: ssIIIa* (RS — 5.74%) with high AC. The marker combination *gbssI*:*ssIIa*:*ssIIIa* having mutant SNP in all the three genes lead to decreased AC and in-turn increasing the significant structural modifications by increasing the α -1-6 linkages which are less prone to α-amylose (Regina et al., 2015). Similar kind of findings were reported by Zihua et al. (2007) who reported that the concentrated α-1,6 linkages in amylopectin lead to the decrease in overall enzyme digestion rate. Because of mutant *gbssI,* lowest recorded AC (20.89%) and RS concentration (2.27%) were analyzed in marker combinations of *gbssI*:*SSIIa : SSIIIa*.

CONCLUSION

Among the analyzed marker combinations in the present study, wild type *GBSSI* with mutant gene alleles of *ssIIa* and *ssIIIa* will be the promising combinations to screen a large number of mutant/germplasm lines for enhanced RS in rice. This marker combination can also be used to develop a cultivar with increased RS concentration through marker-assisted breeding. High RS line developed in this study can be used to manage the global burden of T2DM lines. The outcomes of the present study may lead to propose a novel insight into the development of high yield rice varieties and hybrids with value-added properties in the near future.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

SG performed all the experiments and data analysis. BR helped in execution of all the experiments. VM, CS, and NK helped in execution of bioinformatic and structural analysis. JB and SG together conceptualized this study. SG, BR, CM, AA, AH, and EA drafted the manuscript with the input from JB. All authors read and approved the final manuscript.

FUNDING

Financial support for this study was provided by the pioneer international hi-bred student fellowship of the pioneer international USA to the first author to carry out this work as a part of doctoral program.

ACKNOWLEDGMENTS

This work was supported by pioneer international hi-bred student fellowship of the pioneer international USA to the first author to carry out this work as a part of the doctoral program. We thank Dr. S. Ganesh Ram, Professor and Head, Department of Plant Genetic Resources, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India for offering mutant line, lab facility and also for his technical assistance for this investigation. Also, thankful to Joint Director, DRDO-BU-CLS for his kind support and encouragement to carry out structural analysis. The authors would also like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding to the Research Group number (RG-1435-014).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00946/full#supplementary-material

REFERENCES

Agasimani, S., Selvakumar, G., Joel, A. J., and Ganesh Ram, S. (2013). A simple and rapid single kernel screening method to estimate amylose content in rice grains. *Phytochem. Anal.* 24, 569–573. doi: 10.1002/pca.2433

Benmoussa, M., Moldenhauer, K. A., and Hamaker, B. R. (2007). Rice amylopectin fine structure variability affects starch digestion properties. *J. Agri. Food Chem.* 55, 1475–1479. doi: 10.1021/jf062349x

Butardo, V. M., Fitzgerald, M. A., Bird, A. R., Gidley, M. J., Flanagan, B. M., Larroque, O., et al. (2011). Impact of down-regulation of starch branching enzyme IIb in rice by artificial microRNA-and hairpin RNA-mediated RNA silencing. *J. Exp. Bot.* 62, 4927–4941. doi: 10.1093/jxb/err188

Chan, J. C., Malik, V., Jia, W., Kadowaki, T., Yajnik, C. S., Yoon, K. H., et al. (2009). Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *Jama* 20, 2129–2140. doi: 10.1001/jama.2009.726

Chassy, B., Egnin, M., Gao, Y., Glenn, K., Kleter, G. A., Nestel, P., et al. (2008). Recent developments in the safety and nutritional assessment of nutritionally improved foods and feeds. *Comp. Rev. Food Sci. Food Safety* 7, 50–113. doi: 10.1111/j.1541-4337.2007.00029_4.x

Chattopadhyay, P. K. (2005). "Postharvest technology for rice in India: a changing scenario," in *Rice is life: scientific perspectives for the 21st century (proceedings of the world rice research conference)*, http://large.stanford.edu/courses/2016/ph240/mccall1/docs/toriyama.pdf. International Rice Research Institute and Tsukuba (Japan): Japan International Research Center for Agricultural Sciences. 294–296.

Cummings, J. H., and Englyst, H. N. (1991). What is dietary fibre? *Trends Food Sci. Technol.* 2, 99–103. doi: 10.1016/0924-2244(91)90638-Y

Delrue, B., Fontaine, T., Routier, F., Decq, A., Wieruszeski, J. M., Van, D., et al. (1992). Waxy Chlamydomonas reinhardtii: monocellular algal mutant defective in amylose biosynthesis and granule-bound starch synthase accumulate a structurally modified amylopectin. *J. Bacteriol.* 174, 3612–3620. doi: 10.1128/jb.174.11.3612-3620.1992

Dian, W., Jiang, H., and Wu, P. (2005). Evolution and expression analysis of starch synthase III and IV in rice. *J. Exp. Bot.* 56, 623–632. doi: 10.1093/jxb/eri065

Eerlingen, R. C., and Delcour, J. A. (1995). Formation, analysis, structure and properties of type III enzyme resistant starch. *J. Cereal Sci* 22, 129–138. doi: 10.1016/0733-5210(95)90042-X

Englyst, H. N., and Hudson, G. J. (1996). The classification and measurement of dietary carbohydrates. *Food Chem.* 57, 15–21. doi: 10.1016/0308-8146(96)00056-8

Flipse, E., Keetels, C. J. A. M., Jacobsen, E., and Visser, R. G. F. (1996). The dosage effect of the wildtype GBSS allele is linear for GBSS activity but not for amylose content: absence of amylose has a distinct influence on the physico-chemical properties of starch. *Theor. Appl. Genet.* 92, 121–127. doi: 10.1007/BF00222961

Fuentes-Zaragoza, E., Riquelme-Navarrete, M. J., Sánchez-Zapata, E., and Pérez-Álvarez, J. A. (2010). Resistant starch as functional ingredient: A review. *Food Res. Int.* 43, 931–942. doi: 10.1016/j.foodres.2010.02.004

Fujita, N., Kubo, A., Suh, S. D., Wong, K. S., Jane, J. L., Ozawa, K., et al. (2003). Antisense inhibition of isoamylase alters the structure of amylopectin and the physicochemical properties of starch in rice endosperm. *Plant Cell Physiol.* 44, 607–618. doi: 10.1093/pcp/pcg079

Fujita, N., Yoshida, M., Kondo, T., Saito, K., Utsumi, Y., Tokunaga, T., et al. (2007). Characterization of SSIIIa-deficient mutants of rice: the function of SSIIIa and pleiotropic effects by SSIIIa deficiency in the rice endosperm. *Plant Physiol.* 144, 2009–2023. doi: 10.1104/pp.107.102533

Goni, I., Garcia-Diz, L., Mañas, E., and Saura-Calixto, F. (1996). Analysis of resistant starch: a method for foods and food products. *Food Chem.* 56, 445–449. doi: 10.1016/0308-8146(95)00222-7

Herman, W. H. (2017). "The Global Burden of Diabetes: An Overview," in *Diabetes Mellitus in Developing Countries and Underserved Communities* (Berlin: Springer International Publishing), 1–5. doi: 10.1007/978-3-319-41559-8_1

Hirose, T., Ohdan, T., Nakamura, Y., and Terao, T. (2006). Expression profiling of genes related to starch synthesis in rice leaf sheaths during the heading period. *Physiol. Plant* 128, 425–435. doi: 10.1111/j.1399-3054.2006.00758.x

Hu, E. A., Pan, A., Malik, V., and Sun, Q. (2012). White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. *B.M.J.* 344, 1–9. doi: 10.1136/bmj.e1454

Khush, G. S., and Virk, P. S. (2000). Rice breeding: achievements and future strategies. *Crop Improv.* 2, 115–144.

Leeman, A. M., Karlsson, M. E., Eliasson, A. C., and Björck, I. M. (2006). Resistant starch formation in temperature treated potato starches varying in amylose/amylopectin ratio. *Carbohydr. Polym.* 65 (3), 306–313. doi: 10.1016/j.carbpol.2006.01.019

Lehmann, U., and Robin, F. (2007). Slowly digestible starch — its structure and health implications: a review. *Trends Food Sci. Technol.* 18, 346–355. doi: 10.1016/j.tifs.2007.02.009

Liu, D., Wang, W., and Cai, X. (2014). Modulation of amylose content by structure-based modification of OsGBSS1 activity in rice (Oryza sativa L.). *Plant Biotechnol. J.* 12, 1297–1307. doi: 10.1111/pbi.12228

Miller, J. B., Pang, E., and Bramall, L. (1992). Rice: a high or low glycemic index food? *Am. J. Clin. Nutr* 56, 1034–1036. doi: 10.1093/ajcn/56.6.1034

Montoliu, L. (2012). Mendel: a simple excel workbook to compare the observed and expected distributions of genotypes/phenotypes in transgenic and knockout mouse crosses involving up to three unlinked loci by means of a χ 2 test. *Transgenic Res.* 21, 677–681. doi: 10.1007/s11248-011-9544-4

Nagato, K., and Ebata, M. (1958). Studies on white-core rice kernel. I. On the occurrence of white core. *Proc. Crop Sci. Soc. Jpn.* 27, 49–51. doi: 10.1626/jcs.27.49

Nakamura, Y. (2002). Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. *Plant Cell Physiol.* 43, 718–725. doi: 10.1093/pcp/pcf091

Nakamura, Y., Francisco, P. B., Hosaka, Y., Sato, A., Sawada, T., Kubo, A., et al. (2005). Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between japonica and indica rice varieties. *Plant Mol. Biol.* 58, 213–227. doi: 10.1007/s11103-005-6507-2

Nakamura, Y., Utsumi, Y., Sawada, T., Aihara, S., Utsumi, C., Yoshida, M., et al. (2010). Characterization of the reactions of starch branching enzymes from rice endosperm. *Plant Cell Physiol.* 51, 776–794. doi: 10.1093/pcp/pcq035

Ng, P. C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi: 10.1093/nar/gkg509

Noda, T., Takigawa, S., Matsuura Endo, C., Suzuki, T., Hashimoto, N., Kottearachchi, N. S., et al. (2008). Factors affecting the digestibility of raw and gelatinized potato starches. *Food Chem.* 110, 465–470. doi: 10.1016/j.foodchem.2008.02.027

Rahman, S., Li, Z., Batey, I., Cochrane, M. P., Appels, R., and Morell, M. (2000). Genetic alteration of starch functionality in wheat. *J. Cereal Sci.* 31, 91–110. doi: 10.1006/jcrs.1999.0291

Raja, R. B., Agasimani, S., Jaiswal, S., Thiruvengadam, V., Sabariappan, R., Chibbar, R. N., et al. (2017a). EcoTILLING by sequencing reveals polymorphisms in genes encoding starch synthases that are associated with low glycemic response in rice. *BMC Plant Biol.* 17, 1–13. doi: 10.1186/s12870-016-0968-0

Raja, R. B., Anusheela, V., Agasimani, S., Jaiswal, S., Thiruvengadam, V., Chibbar, R. N., et al. (2017b). Validation and applicability of single kernel-based cut grain dip method for amylose determination in rice. *Food Anal. Methods* 10, 442–448. doi: 10.1007/s12161-016-0607-2

Ramadoss, B. R., Manu, P. G., Somanath, A., Sarita, J., Thiruvengadam, V., Ganesh Ram, S., et al. (2019). Starch granule size and amylopectin chain length influence starch *in vitro* enzymatic digestibility in selected rice mutants with similar amylose concentration. *J. Food Sci. Technol.* 56, 391–400. doi: 10.1007/s13197-018-3500-8

Ramadoss, B. R., Precilla, P., and Ganesh Ram, S. (2015). Characterization of selected rice (*Oryza sativa* L.) land races for high amylose and resistant starch (RS). *Trends Biosci.* 8, 548–552.

Regina, A., Berbezy, P., Kosar-Hashemi, B., Li, S., Cmiel, M., Larroque, O., et al. (2015). A genetic strategy generating wheat with very high amylose content. *Plant Biotechnol. J.* 13, 1276–1286. doi: 10.1111/pbi.12345

Regina, A., Kosar-Hashemi, B., Li, Z., Rampling, L., Cmiel, M., Gianibelli, M. C., et al. (2004). Multiple isoforms of starch branching enzyme-I in wheat: lack of the major SBE-I isoform does not alter starch phenotype. *Funct. Plant Biol.* 31, 591–601. doi: 10.1071/FP03193

Sang, Y., Bean, S., Seib, P. A., Pedersen, J., and Shi, Y. C. (2008). Structure and functional properties of sorghum starches differing in amylose content. *J. Agril. Food Chem.* 56, 6680–6685. doi: 10.1021/jf800577x

Sasaki, T., Kohyama, K., Suzuki, Y., Okamoto, K., Noel, T. R., and Ring, S. G. (2009). Physicochemical characteristics of waxy rice starch influencing the *in vitro* digestibility of a starch gel. *Food Chem.* 116, 137–142. doi: 10.1016/j.foodchem.2009.02.024

Sestili, F., Botticella, E., Proietti, G., Janni, M., D'Ovidio, R., and Lafiandra, D. (2012). Amylose content is not affected by overexpression of the Wx-B1 gene in durum wheat. *Plant Breed.* 131, 700–706. doi: 10.1111/j.1439-0523.2012.02004.x

Taylor, N. E., and Greene, E. A. (2003). PARSESNP: a tool for the analysis of nucleotide polymorphisms. Nucleic acids research, 31 (13), 3808–3811. doi: 10.1093/nar/gkg574

Tester, R. F., Qi, X., and Karkalas, J. (2006). Hydrolysis of native starches with amylases. *Anim. Feed Sci. Technol.* 130, 39–54. doi: 10.1016/j.anifeedsci.2006.01.016

Umemoto, T., Aoki, N., Lin, H., Nakamura, Y., Inouchi, N., Sato, Y., et al. (2004). Natural variation in rice starch synthase IIa affects enzyme and starch properties. *Funct. Plant Biol.* 31, 671–684. doi: 10.1071/FP04009

Umemoto, T., Horibata, T., Aoki, N., Hiratsuka, M., Yano, M., and Inouchi, N. (2008). Effects of variations in starch synthase on starch properties and eating quality of rice. *Plant Prod. Sci.* 11, 472–480. doi: 10.1626/pps.11.472

Waters, D. L. E., and Henry, R. J. (2007). Genetic manipulation of starch properties in plants: patents 2001–2006. *Recent Pat. Biotechnol.* 1, 52–259. doi: 10.2174/187220807782330174

Wolever, T. M., and Mehling, C. (2003). Long-term effect of varying the source or amount of dietary carbohydrate on postprandial plasma glucose, insulin, triacylglycerol, and free fatty acid concentrations in subjects with impaired glucose tolerance. *Am. J. Clin. Nutr* 77, 612–621. doi: 10.1093/ajcn/77.3.612

Yoshii, M. (2000). *Microstructure by SEM during brewing.* The 24th Sakamai Kondankai Koen Yoshishu, 1–7.

Zhou, H., Wang, L., Liu, G., Meng, X., Jing, Y., Shu, X., et al. (2016). Critical roles of soluble starch synthase SSIIIa and granule-bound starch synthase Waxy in synthesizing resistant starch in rice. *Proc. Natl. Acad. Sci.* 113, 12844–12849. doi: 10.1073/pnas.1615104113

Zhu, L., Gu, M., Meng, X., Cheung, S. C., Yu, H., Huang, J., et al. (2012). High-amylose rice improves indices of animal health in normal and diabetic rats. *Plant Biotech. J.* 10, 353–362. doi: 10.1111/j.1467-7652.2011.00667.x

Zihua, A. O., Simsek, S., Zhang, G., Venkatachalam, M., Reuhs, B. L., Hamaker, B. R. et al. (2007). Starch with a Slow Digestion Property Produced by Altering Its Chain Length, Branch Density, and Crystalline Structure. *J. Agric. Food Chem.* 55, 4540–4547. doi: 10.1021/jf063123x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Gurunathan, Ramadoss, Mudili, Siddaiah, Kalagatur, Bapu, Mohan, Alqarawi, Hashem and Abd_Allah. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 27 November 2019
doi: 10.3389/fgene.2019.01177

Hypoallergen Peanut Lines Identified Through Large-Scale Phenotyping of Global Diversity Panel: Providing Hope Toward Addressing One of the Major Global Food Safety Concerns

*Arun K. Pandey, Hari K. Sudini, Hari D. Upadhyaya, Rajeev K. Varshney and Manish K. Pandey**

International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

OPEN ACCESS

Edited by:
Suresh T. Mathews, Samford University, United States

Reviewed by:
Hongbing Chen, Nanchang University, China
Guang Ren, University of Alabama at Birmingham, United States

***Correspondence:**
Manish K. Pandey
m.pandey@cgiar.org

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Genetics

Received: *29 May 2019*
Accepted: *24 October 2019*
Published: *27 November 2019*

Citation:
Pandey AK, Sudini HK, Upadhyaya HD, Varshney RK and Pandey MK (2019) Hypoallergen Peanut Lines Identified Through Large-Scale Phenotyping of Global Diversity Panel: Providing Hope Toward Addressing One of the Major Global Food Safety Concerns. Front. Genet. 10:1177. doi: 10.3389/fgene.2019.01177

Peanut allergy is one of the serious health concern and affects more than 1% of the world's population mainly in Americas, Australia, and Europe. Peanut allergy is sometimes life-threatening and adversely affect the life quality of allergic individuals and their families. Consumption of hypoallergen peanuts is the best solution, however, not much effort has been made in this direction for identifying or developing hypoallergen peanut varieties. A highly diverse peanut germplasm panel was phenotyped using a recently developed monoclonal antibody-based ELISA protocol to quantify five major allergens. Results revealed a wide phenotypic variation for all the five allergens studied *i.e.,* Ara h 1 (4–36,833 μg/g), Ara h 2 (41–77,041 μg/g), Ara h 3 (22–106,765 μg/g), Ara h 6 (829–103,892 μg/g), and Ara h 8 (0.01–70.12 μg/g). The hypoallergen peanut genotypes with low levels of allergen proteins for Ara h 1 (4 μg/g), Ara h 2 (41 μg/g), Ara h 3 (22 μg/g), Ara h 6 (829 μg/g), and Ara h 8 (0.01 μg/g) have paved the way for their use in breeding and genomics studies. In addition, these hypoallergen peanut genotypes are available for use in cultivation and industry, thus opened up new vistas for fighting against peanut allergy problem across the world.

Keywords: reference set, peanut allergens, Ara h 1, Ara h 2, Ara h 3, Ara h 6, Ara h 8, enzyme-linked immunosorbent assay

INTRODUCTION

Food allergy causes severe health issues throughout the globe and the incidences are increasingly recorded across the globe. Even though, approximately 5% of young kids and 4% of adult in western countries are affected by food allergens (Sicherer et al., 2010), the problem has now become more common in developing countries (Liew et al., 2013; Leung et al., 2018). About 40% of the food allergies occur due to the consumption of plants and plant-derived products. Peanut is identified as one of the major sources of food allergy in addition to milk, egg, dry fruits (almonds, cashews, hazelnuts, pistachios, pecans, and walnuts), fish, shellfish, soy, and wheat, with ~90% cumulative contribution among food allergies in human (Hefle et al., 1996). A large number of population

across the world are affected by peanut allergy and several reports are coming more frequently. For instance, the 1% population of Canadian children are allergic to peanuts (Ben-Shoshan et al., 2009) while the prevalence of peanut-based allergy in France and Denmark ranged between 0.3-0.75% and 0.2-0.4%, respectively (Morisset et al., 2005; Osterballe et al., 2005). About 3% of Australians are allergic to peanuts and peanut-based products (Sicherer and Sampson, 2007; Sicherer and Sampson, 2014). It is a big problem in the United Kingdom (UK) as well, and the prevalence of sensitization increased from 1.3 to 3.2% in 3 years old kids (Grundy et al., 2002). Importantly, the prevalence of peanut-based allergy in United States of America (USA) has been increased from 0.4 to 1.4% between 1997 to 2008 (Sicherer et al., 2010). Similarly, peanut allergy has also been reported in the Asian countries such as Singapore and Philippines where 0.47 and 0.43% school children, respectively, were found allergic to peanuts (Shek et al., 2010). Although information from China is not available, however, the situation in China may be similar to Singapore as 76.8% of Singapore residents are Chinese in origin (http://www.singstat.gov.sg/). Although there may be several cases of food allergy in India (Mahesh et al., 2016), however, not much information is available from India on peanut allergy. Keeping in mind the interdependence and trade among countries for producing raw material, processing, and consumption of peanut, such health-concerning features of the crop reduces its importance in international trade and commerce (Pandey et al., 2012; Pandey and Varshney 2018; Varshney et al., 2018; Varshney et al., 2019). Therefore, the countries producing the peanuts and peanut based product with the most safe, nutritious, and healthy features will get a competitive advantage over other producing countries.

All the major food allergies, including peanut, may induce anaphylaxis leading to life-threatening reactions (Dodo et al., 2005; Sicherer and Sampson, 2010) and it is almost impossible to avoid accidental ingestion of peanut-based products (Berger and Smith, 1998; Kagan et al., 2003). Remarkably, food-based allergies cause around 150–200 deaths per year (http://www.startribune.com/peanut-allergy-kills-22-year-old-twin-cities-man/366152021/), largely due to the consumption of peanuts (50–62%) and tree nuts (15–30%) in USA (Lanser et al., 2015). Proteins are the major cause of food allergy, and these proteins are usually highly resistant to heat and proteolysis (Cabanillas et al., 2012). Peanut is the largest source of the immunoglobulin E (IgE)-mediated food allergies and there is no effective treatment due to which the allergic person is forced to avoid consuming peanut or peanut-based products (Wen et al., 2007). However, the peanut being a common food ingredient in many food preparations, it is very challenging for the allergic person to know the composition of these preparations to avoid consumption (Maleki et al., 2000). The threshold of allergen levels differ among the allergic population and even a minute dose of 100 μg of Ara h 1 can trigger an allergic reaction (Warner, 1999). The diagnosis of peanut allergy can be done using different methods such as double-blind, placebo-controlled food challenge (DBPCFC), the basophil activation test, the specific skin prick test (SPT), and the measurement of specific IgE (Hamilton et al., 2010; Lieberman and Sicherer, 2011; Nicolaou et al., 2011).

Of the 32 different types of proteins present in peanut seeds (Pele, 2010), 18 of these proteins show the allergic property (Iqbal et al., 2016). Further, out of 18 peanut allergen proteins mainly Ara h 1, Ara h 2, Ara h 3, and Ara h 6 are considered as major allergens due to their life-threatening reactions recognized by the IgE leading to anaphylaxis (Krause et al., 2010). The remaining allergen proteins are considered as minor allergens as they don't cause life-threatening allergic reactions (anaphylaxis). Nevertheless, if a person is already sensitive to Bet v 1 allergen caused due to birch pollen, then one of these minor peanut allergens, Ara h 8, shows cross-reactivity with IgE antibodies causing oral allergy syndrome (OAS) (Mittag et al., 2004; Riecken et al., 2008; Kondo and Urisu, 2009). Allergic protein belongs to different protein families namely cupin (vicilin-type, 7S globulin, legumin-type, 11S globulin, glycinin), conglutin (2S albumin), profilin, nonspecific lipid-transfer protein 1, pathogenesis-related protein (PR-10) 14 kDa, oleosin (16 kDa), and seed storage proteins particularly Ara h 1, Ara h 2, Ara h 3, and Ara h 6 (Pele, 2010). Many studies have shown that the most abundant peanut-based allergens (Ara h 1 and Ara h 3 but, Ara h 2 and Ara h 6) bind strongly with peanut allergic IgE and release basophils mediators, which were confirmed in vitro (de Jong et al., 1998; Koppelman et al., 2005; Palmer et al., 2005; Porterfield et al., 2009) and in vivo (Koppelman et al., 2003; Koppelman et al., 2005; Peeters et al., 2007) with regards to food allergy (Porterfield et al., 2009). Although all the five peanut allergens (Ara h 1 and Ara h 3 but, Ara h 2 and Ara h 6) show IgE reactivity to these peanut allergens, however, the Ara h 2 and Ara h 6 allergens are more commonly recognized in children (Flinterman et al., 2007).

Possible solutions to peanut allergy include the development of vaccine or development of allergen-free peanut varieties. Much research has been focussed on diagnosis and cure to minimize the impact of allergens in the human population, however, reducing allergen proteins in peanut varieties and their products can be the best solution. Unfortunately, insufficient scientific information on a total number of allergen genes in the peanut genome and level of phenotypic variability in existing peanut germplasm hinders further research in this area. Therefore, the main objective of this research was to identify hypoallergen peanut lines by screening a large number of diverse germplasm in the peanut reference set (Upadhyaya, 2009). The hypoallergen peanut varieties that have been identified will promote their commercialization and use in the peanut-based industry. These lines and information generated out of this work can be of great importance in efforts toward fighting peanut allergy and ensuring food safety across the world.

MATERIALS AND METHODS

Plant Materials

The peanut "reference set" consisting of 300 diverse accessions representing 51 countries (Upadhyaya, 2009) (**Supplementary Table 1** and **Figure 1A**) were selected from the composite collection. The reference set included the 184 accessions of the peanut mini core collection (Upadhyaya et al., 2002) that represented diversity of the core collection (1,704 accessions)

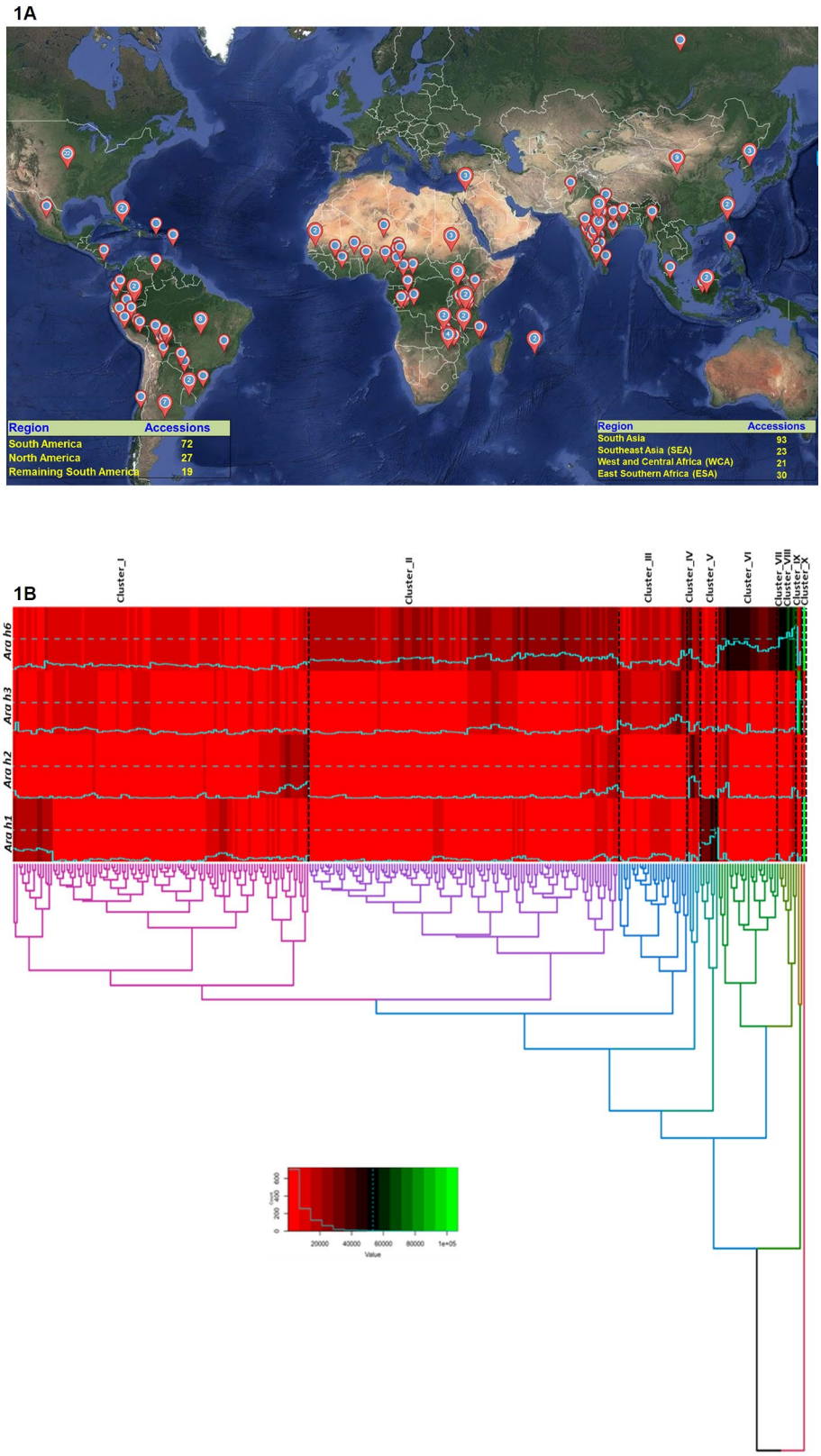

FIGURE 1 | Continued

FIGURE 1 | Graphical representation showing global coverage of peanut "reference set" across 51 countries. **(1A)** The world map was constructed by using Maptive software (http://fortress.maptive.com). Out of 300 accessions of the ICRISAT peanut reference set, 93 belongs to South Asia, 23 from Southeast Asia, 21 from West and Central Africa 30 from East and Southern Africa, 72 from South America, 27 from North America, and 19 from remaining South America. **(1B)** Hierarchical clustering on the basis of phenotyping data grouped 300 accessions in 10 cluster. Cluster 1 having 112 accessions, cluster 2 having 117 accessions while cluster 3, and 6 having 27 and 23 accessions respectively. Cluster 7, 8, 9, and 10 showing a high amount of Ara h 6 in their germplasm lines. In cluster 1 all the accessions having a lower amount of allergen content.

which in turn represents the diversity of entire International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) peanut genebank collection (14,310 accessions) (Upadhyaya et al., 2003). This set comprised of 264 cultivated species (*Arachis hypogaea*) and 36 wild species and has fair representation for its two subspecies namely *fastigiata* (154) and *hypogaea* (95). The subspecies *fastigiata* was further classified into four botanical varieties namely Fastigiata (70 accessions), Peruviana (5 accessions), Vulgaris (78 accessions), and Aequitoriana (1 accession). Similarly, the other subspecies *hypogaea* was classified into two botanical varieties namely Hirsuta (2 accessions) and Hypogaea (93 accessions) (Holbrook and Stalker 2003). These cultivated genotypes can also be classified into four agronomic types based on their growth habit namely Spanish bunch (73), Valencia bunch (70), Virginia bunch (51), and Virginia runner (33). The seeds for cultivated genotypes (264) were collected from two seasons (Rainy 2016 and Post-rainy 2016–17) for estimating allergen content. Since many of the wild accessions have annual growth period, seeds from two different lots were taken for allergen estimation.

Sample Preparation and Protein Isolation

Sample preparation and protein isolation were performed following the protocol mentioned in (Pandey et al., 2019). In brief, 2 g of seeds were grinded to make a fine powder and then dissolved in 40 ml of PBS-T (0.05% Tween in phosphate buffered saline, pH 7.4) containing 1 M NaCl in 50 ml falcon tubes (Sarstedt No: 55.476). After 2 h of gentle stirring at room temperature on the rocking platform, the aqueous phase was collected by centrifugation at 2,500 rpm at 4°C for 20 min. The aqueous phase was subsequently centrifuged to remove residual traces and insoluble particles at 3,500 rpm for 10 min at room temperature. Until use, extracts of proteins were stored at −20°C.

Allergens Estimation Using Enzyme-Linked Immunosorbent Assay

Sandwich format ELISA was used in the study. The peanut allergen proteins were first sandwiched between two antibodies, and then streptavidin-peroxidase was captured. Each peanut sample contains a different quantity of allergen proteins which makes very difficult to estimate the accurate amount of allergen proteins present in seed samples. Dilution is a vital step for ELISA experiment which in turn determines the values of detection range for antibody and target antigen concentrations.

The estimation of peanut allergen through sandwich ELISA were performed according to the recently published protocol (Pandey et al., 2019). Each allergen protein was estimated at different dilution factors (DF). We used a number of dilutions in the peanut samples to detect the specific allergic protein in seeds. The Ara h 1 was detected on three serial doubling dilutions, 1:1,000, 1:2,000, and 1: 4,000 while Ara h 2 and Ara h 3 detected on same dilution 1: 5,000, 1:10,000, and 1: 20,000. In peanut seeds, the Ara h 6 was detected in the high range (1: 40,000, 1: 80,000, and 1:160,000) DF while Ara h 8 detected in a low range of dilution, i.e., 1:10, 1:20, and 1:40.

Cluster and Data Analysis on the Basis of Allergen Content

Statistical analysis was performed to identify the wide variation of peanut allergens among samples using SigmaPlot (http://www.sigmaplot.co.uk/products/sigmaplot/sigmaplot-details.php). Hierarchical clustering was done using average allergens content of five major allergens (Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8) on the basis of similarity matrix generated using HCA (hierarchical cluster analysis). Dendextend statistical package (Galili, 2015) was used for clustering the genotypes on the basis of similarity of average allergen content. This software provides a set of functions for cluster analysis and construction of dendo-gram. The heat map was generated using R package gplots (Warnes et al., 2016) for allergen content of 300 samples.

RESULTS

Phenotypic Variation for Allergens Between Cultivated and Wild Gene Pool

Phenotyping of 300 diverse accessions (264 cultivated and 36 wild accessions) showed wide phenotypic variations for all the five allergens. Among cultivated accessions, the phenotypic variation was very high for all five major allergens namely Ara h 1 (4–36,833 μg/g), Ara h 2 (52–77,042 μg/g), Ara h 3 (22–106,765 μg/g), Ara h 6 (829–103,892 μg/g), and Ara h 8 (0.01–70 μg/g) as compared to the wild relatives i.e., Ara h 1 (28–1,293 μg/g), Ara h 2 (516–33,055 μg/g), Ara h 3 (1,185–20,474 μg/g), Ara h 6 (4,381–56,858 μg/g), and Ara h 8 (0.3–11 μg/g) (**Table 1**). Except Ara h 3 (6,857 μg/g), the average content of Ara h 1 (404 μg/g), Ara h 2 (606 μg/g), Ara h 6 (13,420 μg/g), and Ara h 8 (2 μg/g) was found lower in wild accessions compared to cultivated genotypes (**Table 1** and **Figure 2A**).

Phenotypic Variation for Allergens Between Subspecies *hypogaea* and *fastigiata*

Phenotyping result showed a wide variation for all the five allergens between two subspecies of cultivated peanut i.e., *A. hypogaea* ssp. *hypogaea* (96 accessions) and *A. hypogaea* ssp.

TABLE 1 | Range and mean value for five major peanut allergens in peanut germplasm lines of the reference set in the rainy 2016 and 2016–17 postrainy seasons, International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India.

Name	Number of accessions	Ara h 1 (μg/g) Range and mean	Ara h 2 (μg/g) Range and mean	Ara h 3 (μg/g) Range and mean	Ara h 6 (μg/g) Range and mean	Ara h 8 (μg/g) Range and mean
Variation in cultivated and wild accessions						
Cultivated	264	4–36,833	52–77,042	22–106,765	829–103,892	0.01–70.12
		3,789	7,664	6,008	19,962	6
Wild	36	28–1,293	516–33,055	1,185–20,474	4,381–56,858	0.3–11
		404	606	6,857	13,420	2
Variation between two subspecies						
A. hypogaea ssp. *hypogaea*	96	29–22,000	560–62,350	93–63,720	3,264–57,504	0.5–20
		2,005	8,077	7,015	19,844	6
A. hypogaea ssp. *fastigiata*	148	4.0–36,833	41–57,915	22–106,765	829–103,892	0.1–38
		4,694	6,690	5,276	18,855	5
Variation in agronomic type						
Spanish bunch	72	4.0–8,267	41–57,915	22–20,574	3,022–63,867	0.01–70.12
		1,576	6,139	4,120	17,887	5
Valencia bunch	68	28–36,833	1,017–20,734	68–32,380	829–103,892	0.1–12
		8,295	7,084	5,241	19,581	5
Virginia bunch	53	30–12,470	560–22,695	104–63,720	5,311–57,504	0.5–19
		1,700	7,929	6,201	20,432	6
Virginia runner	33	91–4,594	791–62,350	337–42,508	7,037–46,240	0.7–20
		1,332	8,449	7,468	20,301	6
Variation among accessions with the different biological status						
Landraces	105	28–13,977	41–62,350	22–106,765	3,657–51,024	0.01–70.12
		2,272	7,630	7,970	19,251	6
Breeding material	52	60–36,833	1,874–25,900	131–323,801	829–103,892	0.1–12
		9,917	8,193	5,887	18,300	5
Improved cultivars	67	13–19,235	263–57,915	68–14,645	3,022–57,504	0.1–20
		1,915	7,187	3,038	20,183	5
Wild accessions	33	28–1,293	516–33,055	1,185–20,474	4,381–56,858	0.3–11
		404	606	6,857	13,420	2
Variation among accessions from different geographical regions						
South Asia (SA)	93	28–36,833	561–20,734	61–43,408	829–103,892	0.2–19
		7,121	6,216	6,420	17,993	5
Southeast Asia (SEA)	23	50–11,014	660–25,900	89–20,574	3,910–63,867	0.1–19
		2,283	6,653	5,561	18,285	7
West and Central Africa (WCA)	21	64–6,980	95.7–21,847	337–34,818	4,381–54,830	1.5–24
		1,962	9,130	10,595	17,098	7
East and Southern Africa (ESA)	30	60–13,977	1,017–62,350	105–16,132	6,395–47,041	0.9–70.12
		1,805	11,210	4,137	22,010	7
South America	72	28–6,727	516–33,055	68–63,720	3,657–56,858	0.3–15
		1,184	7,615	5,624	16,272	4
North America	27	12.5–12,470	262–22,696	98–9,431	5,756–57,504	0.1–20
		1,514	6,494	2,637	20,132	5
Remaining South America	19	288–7,867	2,812–15,267	42–106,765	5,586–42,552	0.1–11
		2,243	8,338	8,476	20,958	5

fastigiata (148 accessions). In ssp. *hypogaea,* the allergen content ranged from 29 to 22,000 μg/g for Ara h 1, 560 to 62,350 μg/g for Ara h 2, 93 to 63,720 μg/g for Ara h 3, 3,264 to 57,504 μg/g for Ara h 6, and 0.5 to 20 μg/g for Ara h 8 whereas in ssp. *fastigiata*, the allergen content varied from Ara h 1 (4–36,833 μg/g), Ara h 2 (41–57,915 μg/g), Ara h 3 (22–106,765 μg/g), Ara h 6 (829–103,892 μg/g), and Ara h 8 (0.3–11 μg/g) (**Table 1**). Excluding Ara h 1 (2,005 μg/g), the average amount of Ara h 2 (8,007 μg/g), Ara h 3 (7,015 μg/g), Ara h 6 (19,844 μg/g), and Ara h 8 (6 μg/g) was higher in ssp. *hypogaea* as compared to ssp. *fastigiata* i.e., Ara h 1 (4,694 μg/g), Ara h 2 (6,690 μg/g), Ara h 3 (5,276 μg/g), Ara h 6 (18,855 μg/g), and Ara h 8 (5 μg/g) (**Table 1** and **Figure 2B**). Even the average amount of allergens was detected high in ssp. *hypogaea,* the hypoallergen lines *i.e.,* ICG 1534 (4 μg/g) for Ara h 1, ICG 13491 (41 μg/g) for Ara h 2, ICG 6375 (22 μg/g) for Ara h 3, and ICGV 1328 (829 μg/g) for Ara h 6 were identified from ssp. *fastigiata* (**Table 1**). The above results were also supported by HCA. In HCA analysis, cluster 1 formed on the basis of the

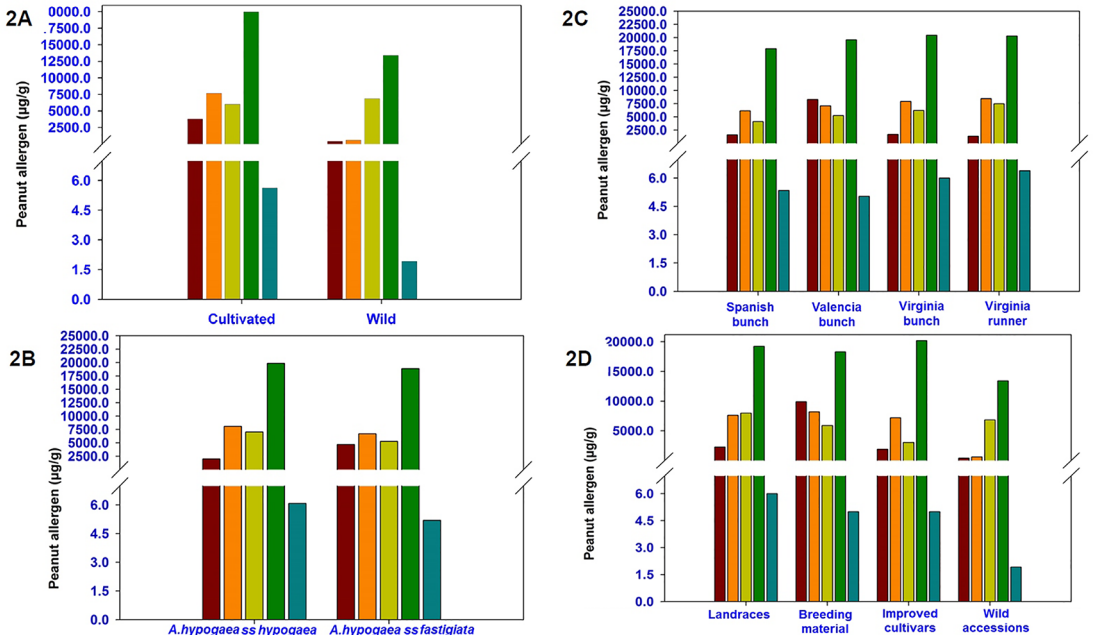

FIGURE 2 | Phenotypic variation for five major peanut allergens **(2A)** between cultivated and wild germplasm accessions **(2B)** between subspecies *hypogaea* and *fastigiata* **(2C)** among different agronomic types **(2D)** and among the biological types of peanut reference set collection.

low quantity of allergens in their accessions. In cluster 1 the allergen content ranged from 12.5 to 27,375 µg/g for Ara h 1, 40.5 to 20,600 µg/g for Ara h 2, 55.43 to 23,306 µg/g for Aar h 3, 829 to 16,483 µg/g for Ara h 6, and 0.067 to 70 µg/g for Ara h 8 (**Supplementary Table 2** and **Figure 1B**).

Phenotypic Variation for Allergens Among Different Peanut Agronomic Types

Phenotyping of accessions from different agronomic types namely Spanish bunch (72 accessions), Valencia bunch (68 accessions), Virginia bunch (53 accessions), and Virginia runner (33 accessions) indicated significant variation (**Table 1** and **Figure 2C**). Although the average amount of Ara h 1 was lowest in Virginia runner (1,332 µg/g) followed by Spanish bunch (1,576 µg/g), Virginia bunch (1,700 µg/g), and Valencia bunch (8,295 µg/g), however, the most hypoallergen line, ICG 1534 (4 µg/g) belongs to Spanish bunch (**Table 1** and **Figure 2C**). Likewise, Spanish bunch (17,887 µg/g) agronomic type had the lowest average for Ara h 6 as compared to other agronomic types but the most hypoallergen line was identified from Valencia bunch, ICGV 1328 (829 µg/g) (**Table 1**). Interestingly for Ara h 2 (6,139 µg/g), Ara h 3 (4,120 µg/g), and Ara h 8 (5 µg/g), the accessions with minimum average as well most hypoallergen line were detected in Spanish bunch agronomic type (**Table 1**).

Phenotypic Variation for Allergens Among Peanut Accessions of Cultivated Gene Pool With Different Biological Status

The reference set included 105 traditional cultivars/landraces, 52 breeding/research material, 67 advanced/improved cultivars, and 36 wild accessions. Except for Ara h 3 (6,857 µg/g), the average amount of Ara h 1 (404 µg/g), Ara h 2 (606 µg/g), Ara h 6 (13,420 µg/g), and Ara h 8 (2 µg/g) was lower in wild accessions as compared to all the biological status groups of cultivated gene pool (**Table 1** and **Figure 2D**). Interestingly, most hypoallergen lines were detected in different biological status groups of cultivated gene pool. For example, Ara h 1 in advanced/improved cultivars (13–19,235 µg/g), Ara h 2 in traditional cultivars/landraces (41–62,350 µg/g), Ara h 6 in breeding/research material (829–103,892 µg/g), and Ara h 8 in traditional cultivars/landraces (0.01–70 µg/g). Similarly, the average amount of Ara h 3 allergens was found lowest in advanced/improved cultivars (3,038 µg/g) but the hypoallergen line was identified in traditional cultivars/landraces (41–62,350 µg/g) (**Table 1**).

Phenotypic Variation for Allergens Among Peanut Accessions Representing Different Geographical Regions

The phenotyped set of 300 accessions represent different geographical regions namely South Asia (SA, 93 accessions), Southeast Asia (SEA, 23 accessions), West and Central Africa (WCA, 21 accessions), East and Southern Africa (ESA, 30 accessions), South America (72 accessions), North America (27 accessions), and 19 accessions from remaining South America. The average amount for Ara h 1 was lowest in accession representing South America (1,184 µg/g) followed by North America (1,514 µg/g), ESA (1,805 µg/g), WCA (1,962 µg/g), remaining South America (2,243 µg/g), SEA (2,283 µg/g), and SA (7,121). Likewise, the average amount of Ara h 2, Ara h 3, Ara h 6, and Ara h 8 were identified in SA (6,216 µg/g), North America (2,637 µg/), WCA (17,098),

and South America (4 μg/g), respectively (**Table 1** and **Figure 3A**). The genotypes available in North America region especially in the USA had low amount of Ara h 1 (12.5–12,470 μg/g) as compared to genotypes grown in other parts of the world while the average content for Ara h 2 was low in WCA (41–21,847 μg/g) (**Table 1** and **Figure 3A**). The genotypes from South America had low amount of Ara h 3 (68–63,720 μg/g). In general, the allergen content for Ara h 6 was high and low for Ara h 8 across geographical regions. The most hypoallergen line for Ara h 6 (829 μg/g) was identified from the SA region (**Table 1** and **Figure 3A**). Cluster analysis also revealed that the North America region having low allergen lines for Ara h 1 (4–21,999 μg/g) and Ara h 2 (40.5–20,600 μg/g) while South America region having hypoallergen lines for Ara h 3 (53.43–23,306 μg/g) (**Supplementary Table 2** and **Figure 1B**).

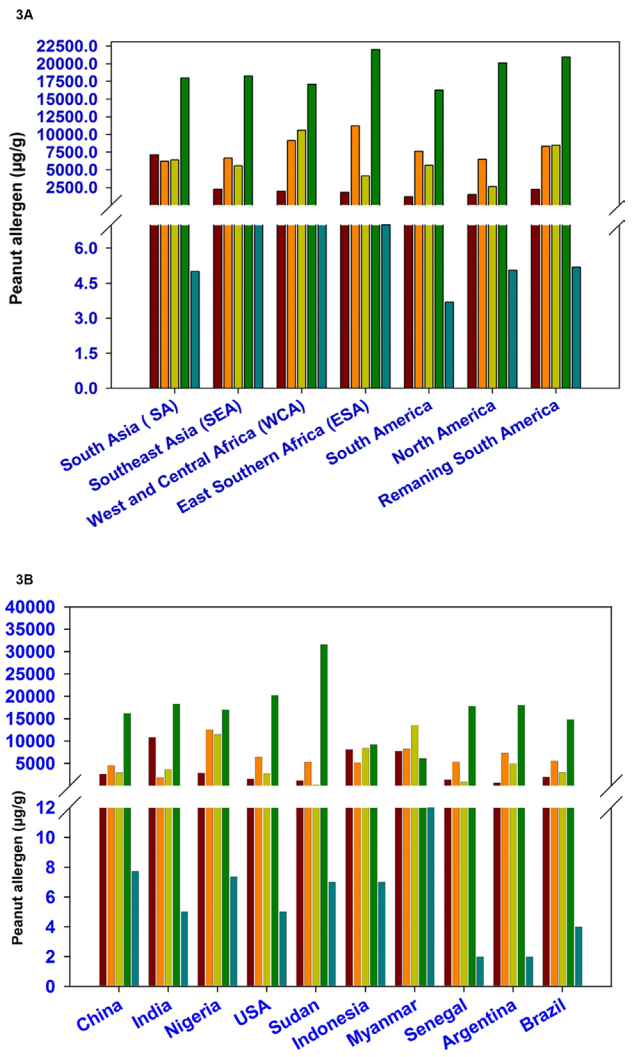

FIGURE 3 | Phenotypic variation for five major peanut allergens among peanut germplasm lines representing various geographical regions **(3A)** and top 10 peanut producing countries **(3B)**.

Phenotypic Variation for Allergens Among Peanut Accessions Representing Top 10 Peanut Producing Countries

Top 10 peanut producing countries include China, India, Nigeria, USA, Sudan, Indonesia, Myanmar, Senegal, Argentina, and Brazil which together contributed more than 81% of total global peanut production. Lowest average amount of Ara h 1 was found in Argentina (673 μg/g) followed by Sudan (1,174 μg/g), Senegal (1,383 μg/g), USA (1,554 μg/g), Brazil (1,968 μg/g), Nigeria (2,874 μg/g), Indonesia (4,037 μg/g), China (4,138 μg/g), Myanmar (7,714 μg/g), and India (**Table 2; Supplementary Table 3**). But the germplasm used in the USA had comparatively lower for Ara h 1 (13–5,023 μg/g) (**Table 2** and **Figure 3B**), followed by India (28–36,833 μg/g) and China (51–63,910 μg/g). Likewise for Ara h 2, USA (263–22,969 μg/g) (**Table 2** and **Figure 3B**) showed hypoallergen lines even the average content of Ara h 2 was lower in Indian germplasm line (1,876 μg/g). The hypoallergen lines for Ara h 3 (61–43,408 μg/g) and Ara h 6 (829–103,892 μg/g) were identified in Indian germplasm lines even though the average content of Ara h 3 (183 μg/g) and Ara h 6 (6,087 μg/g) was lower in Sudan and Myanmar germplasm lines. (**Table 2, Supplementary Table 3** and **Figure 3B**).

Hypoallergen Lines for Five Major Peanut Allergens (Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8)

A wide variation among the 300 germplasm lines was observed for the presence of five major allergens i.e., Ara h 1 (4–36,833 μg/g), Ara h 2 (41–77,041 μg/g), Ara h 3 (22–106,765 μg/g), Ara h 6 (829–103,892 μg/g), and Ara h 8 (0.01–70.12 μg/g). Some germplasm lines showed low allergen content for combinations of two to three allergens.

For Ara h 1, the phenotypic variation ranged from 4 μg/g (ICG 1534) to 36,833 μg/g (ICGV 02038) among cultivated accessions while it ranged from 28 μg/g (ICG 8124) to 1,293 μg/g (ICG 11555) among wild accessions (**Table 1** and **Figure 4A**). The most hypoallergen lines for Ara h 1 included ICG 1534 (4 μg/g) followed by ICG 311 (12.5 μg/g), ICG 81 (17.6 μg/g), ICG 442 (22.8 μg/g), and ICG 115 (27.9 μg/g) (**Table 3** and **Figure 4A**). All these hypoallergen lines belong to Spanish bunch types of *A. hypogaea* ssp. *fastigiata* except ICG 115 which belongs to Valencia bunch of the same subspecies (**Table 3** and **Figure 4A**). The best low allergen lines can be used for conducting further research on genomics and breeding.

A wide variation also observed in the presence of Ara h 2 and it ranged from 52 μg/g (ICG 3140) to 77,042 μg/g (ICGV 7000) among cultivated accessions (264) while it ranged from 516 μg/g (ICG 8144) to 15,757 μg/g (ICG 13218) among wild germplasm accessions (**Table 1** and **Figure 2A**). The most hypoallergen lines for Ara h 2 were ICG 13491 (41 μg/g), followed by ICG 3140 (52 μg/g), ICG 12189 (249 μg/g), ICG 311 (262 μg/g), and ICG 8144 (516 μg/g) (**Table 3** and **Figure 4B**). All these hypoallergen lines belonging to Spanish bunch types of *A. hypogaea* ssp. *fastigiata*. Interestingly ICG 311 belonging to Spanish bunch types of *A. hypogaea* ssp. *fastigiata* also had low Ara h 1 (12.5 μg/g) (**Table 3**).

TABLE 2 | Range and mean value for 5 major peanut allergens in peanut germplasm lines of the reference set representing top 10 peanut producing countries.

S. No	Country	Accession number	Ara h 1 (μg/g) Range and mean	Ara h 2 (μg/g) Range and mean	Ara h 3 (μg/g) Range and mean	Ara h 6 (μg/g) Range and mean	Ara h 8 (μg/g) Range and mean
1	China	10	51–6,391 4,138	660–6,331 4,362	89–11,897 2,921	5,196–39,660 16,201	0.1–19 8
2	India	90	28–36,833 10,796	581–20,734 1,876	61–43,408 3,672	829–103,892 18,285	0.2–18.5 5
3	Nigeria	5	799–5,023 2,874	8,146–21,847 12,533	841–19,819 11,464	6,467–25,401 16,970	0.1–20 7
4	USA	26	13–1,250 1,554	263–22,696 6,403	98–9,431 2,700	5,756–57,504 20,182	0.1–20 5
5	Sudan	3	599–2,109 1,174	3,984–6,618 5,317	105–225 183	19,825–47,041 31,612	3.6–8.7 7
6	Indonesia	3	432–11,014 4,037	2,997–6,306 4,807	1,371–20,574 9,357	3,910–19,197 9,247	3–5.7 7
7	Myanmar	2	1,614–13,813 7,714	7,211–9,371 8,291	7,197–19,902 13,550	6,087 6,087	12 12
8	Senegal	2	65–2,611 1,338	2,998–7,673 5,336	337–1,341 839	13,541–22,023 17,782	1.6–3 2
9	Argentina	22	78–4,177 673	516–15,757 7,297	131–20,456 4,975	4,381–56,850 17,951	0.3–8 2
10	Brazil	15	49–6,727 1,968	842–17,544 5,536	68–7,542 3,016	6,874–27,561 14,837	0.4–11 4

The phenotypic variation of Ara h 3 in cultivated genotypes ranged from 22 μg/g (ICG 6375) to 106,765 μg/g (ICG 12625) while in wild accessions ranged from 1,185 μg/g (ICG 8135) to 20,474 μg/g (ICG 1156) (**Table 1**). The most hypoallergen lines for Ara h 3 included ICG 6375 (22 μg/g) followed by ICG 4750 (41 μg/g), ICG 15779 (61 μg/g), ICG 6888 (68 μg/g), and ICG 4729 (89 μg/g) (**Table 3** and **Figure 4C**). All these hypoallergen lines belong to Spanish bunch types of *A. hypogaea* ssp. *fastigiata* except ICG 6888 (68 μg/g) which belongs to Valencia bunch of the same subspecies (**Table 3** and **Figure 4C**).

For Ara h 6, the phenotypic wide variation identified in cultivated genotypes, and it ranged from 829 μg/g (ICGV 01328) to 103,892 μg/g (ICGV 02194) compared to wild accessions ranged from 4,381 μg/g (ICG 8123) to 56,857 μg/g (ICG 8195) (**Table 1**). The most hypoallergen lines for Ara h 6 identified in cultivated genotypes namely ICG 01328 (829 μg/g) followed by ICG 15384 (1,577 μg/g), ICG 15403 (2,066 μg/g), ICGV 01276 (2,078 μg/g), and ICGV 02022 (2,298 μg/g) (**Table 4** and **Figure 4D**). All these hypoallergen lines belong to Valencia bunch type of *A. hypogaea* ssp. *fastigiata*.

The phenotypic variation for Ara h 8 was very narrow in cultivated accessions and it ranged from 0.01 μg/g (ICG 334) to 70 μg/g (ICG 12921) compared to wild accessions ranging from 0.3 μg/g (ICG 13206) to 11.27 μg/g (ICG 8973) (**Table 1**). The best low allergen lines for Ara h 8 included ICG 334 (.01 μg/g) followed by ICG 311(0.1 μg/g), ICG 405(0.1 μg/g), ICG 1487 (0.2 μg/g), and ICG 1569 (0.2 μg/g) (**Table 4** and **Figure 4D**). All these hypoallergen lines belong to Spanish bunch type of *A. hypogaea* ssp. *fastigiata*.

DISCUSSION

Peanut allergy is now a global health problem and so far no permanent solution is available to deal with this menace. More importantly, the peanut is consumed in the form of several peanut-based products, therefore, making the life of an allergic person more complicated and difficult. Hypoallergen lines provide an alternative approach to avoid these adverse reaction caused by IgE (Tscheppe and Breiteneder, 2017; Satitsuksanoa et al., 2018). The skin, the respiratory tract, and the gastrointestinal tract are allergic to the peanut and peanut-based product (Sicherer et al., 1998) and cute urticaria, acute vomiting, laryngeal oedema, hypotension, and dysrhythmia are the common symptoms (Bock et al., 2001; Sampson et al., 2017). Peanut-based allergy is very risky, and even the ingestion of trace amounts of peanut can cause life threats in minutes (Bock et al., 2001).

The Stand-Alone Effort for Phenotyping Large-Scale, Diverse Germplasm Set for Major Peanut Allergens Using Most Sensitive Enzyme-Linked Immunosorbent Assay Protocol

Not much efforts have been done toward phenotyping a large peanut germplasm collections in the world. This is majorly due to lack of robust and high-throughput analytical assays to quantify major allergen proteins in peanut seeds. Recently our lab developed an ELISA based protocol to estimate major peanut allergens (Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8) using peanut seeds (Pandey et al., 2019). By using this protocol, we phenotyped 300 germplasm lines to quantify major peanut allergens. This study successfully identified hypoallergen lines for all the five allergens and this genetic variation for allergens can be exploited in crop improvement for developing improved hypoallergen lines (**Figure 5**). Using a pool of human serum from patients, a sample ELISA protocol was used to identify antigens in the peanut seed (Dodo et al., 2002) which reported no significant difference in the allergen

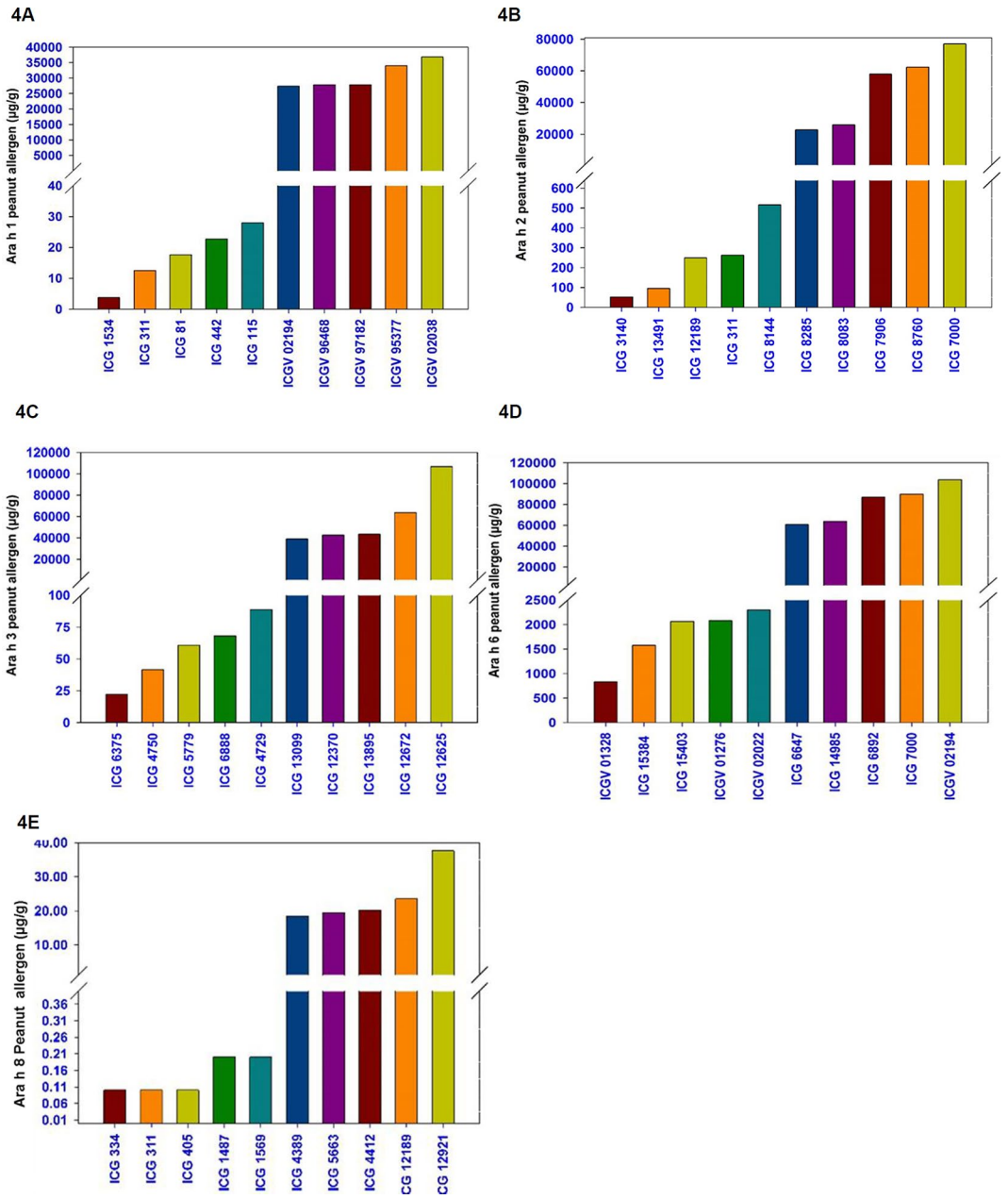

FIGURE 4 | Selected germplasm lines with minimum and the maximum amount of allergen protein. **(4A)** Five peanut lines with highest and five lines with lowest allergen content for Ara h1 **(4B)** Ara h 2, **(4C)** Ara h 3, **(4D)** Ara h 6, and **(4E)** Ara h 8.

content. Another such study on 53 Chinese peanut cultivars revealed that the allergenicity was caused by the allergen composition rather than a single allergen (Wu et al., 2016). This study also reported that the allergen content was high in all the peanut cultivars, however, the peanut allergen content could not be quantified in peanut seeds due to unavailability of antibodies. Hence our study is the first of its kind and identified a low/hypoallergen lines from peanut reference set to ensure food safety and security.

Diverse Germplasm Set Representing 51 Countries Showed Wide Phenotypic Variation for Allergens

Results confirmed a wide variation of five major peanut allergens in the peanut reference set. This study used monoclonal antibodies for each allergen for phenotyping of ICRISAT reference set representing global diversity. These monoclonal antibodies also used to observe differences in specific peanut allergen profile in peanut flour and peanut-based products such

TABLE 3 | List of accessions having minimum and maximum allergen protein content for Ara h 1, Ara h 2, and Ara h 3 from peanut germplasm lines.

S. No.	Accessions	Species	Sub-species	Botanical type	Agronomic type	Peanut allergen content (μg/g)
Low and high allergen lines for Ara h 1						
1	ICG 1534	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	4
2	ICG 311	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	12.5
3	ICG 81	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	17.6
4	ICG 442	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	22.8
5	ICG 115	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	27.9
6	ICGV 02194	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	27,375
7	ICGV 96468	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	27,831
8	ICGV 97182	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	27,831
9	ICGV 95377	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	34,034
10	ICGV 02038	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	36,833
Low and high allergen lines for Ara h 2						
1	ICG 13491	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	41.0
2	ICG 3140	*A. hypogaea*	*INA*	INA	INA	52.4
3	ICG 12189	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	249
4	ICG 311	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	262
5	ICG 8144	*A. villosa*	*INA*	Villosa	INA	516
6	ICG 8285	*A. hypogaea*	*hypogaea*	Hypogaea	Virginia bunch	22,695
7	ICG 8083	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	25,899
8	ICG 7906	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	57,915
9	ICG 8760	*A. hypogaea*	*hypogaea*	Hypogaea	Virginia runner	62,350
10	ICG 7000	*A. hypogaea*	*hypogaea*	Hypogaea	Virginia bunch	77,041
Low and high allergen lines for Ara h 3						
1	ICG 6375	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	22
2	ICG 4750	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	41
3	ICG 5779	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	61
4	ICG 6888	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	68
5	ICG 4729	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	89
6	ICG 13099	*A. hypogaea*	*hypogaea*	Hypogaea	Virginia runner	38,988
7	ICG 12370	*A. hypogaea*	*hypogaea*	Hypogaea	Virginia runner	42,508
8	ICG 13895	*A. hypogaea*	*hypogaea*	Hypogaea	INA	43,408
9	ICG 12672	*A. hypogaea*	*hypogaea*	Hypogaea	Virginia bunch	63,720
10	ICG 12625	*A. hypogaea*	*fastigiata*	Aequitoriana	INA	106,765

S. No. 1–5 are accessions with low allergen (hypoallergen) protein content while 6–10 are accessions with high allergen protein content. INA, information not available.

as peanut butter, flour, and other confectionary preparations for clinical use (Filep et al., 2018). Screening of ICRISAT peanut reference set showed wide range of variation for all the five allergens i.e., Ara h 1 (4–36,833 μg/g), Ara h 2 (41–77,041 μg/g), Ara h 3 (22–106,765 μg/g), Ara h 6 (829–103,892 μg/g), and Ara h 8 (0.01–70 μg/g). Similar wide variation was also identified for Ara h 1, Ara h 2, and Ara h 3 in peanut butter, peanut powder, and peanut flour (Filep et al., 2018). An earlier study reported screening of 34 peanut accessions through patient sera, but no significant difference was observed for allergen content (Dodo et al., 2002). The other study also reported not much variation among 53 Chinese peanut cultivars (Wu et al., 2016) which may be due to the use of human sera to estimate the allergen content in their cultivars. These circumstances encouraged us to develop ELISA based protocol which can be used for quantifying allergen content in peanut kernels. Furthermore, we used the most diverse panel "reference set" consists of 300 genotypes which geographically represents 51 countries (Upadhyaya et al., 2003; Upadhyaya et al., 2010) and showed wide variation for all the five major allergens. The sensitivity of peanut allergens varied among populations in different geographical regions (Vereda et al., 2011). In USA and Sweden, the Ara h 1, Ara h 2, and Ara h 3 cause majority of the peanut allergenic reactions leading to serious illnesses. Similar trend has also been observed in 11 European countries (Ballmer-Weber and Beye, 2018). In contrast, the Spanish patients have less sensitivity to Ara h 1, Ara h 2, and Ara h 3 allergens and have shown more sensitivity to Ara h 9, lipid transfer protein. Similarly, the Spanish patients had the highest level of sensitivity rate to birch pollen allergen, Ara h 8, a cross-reactive homolog Bet v 1. It is important to note that despite few reports, not much have been reported from different Asian and African countries. The above difference in allergen sensitivity among countries and continents may have resulted due to several factors, including genetic makeup, environmental factors, and food habits.

Hypoallergen Lines Identified for Major Peanut Allergens

Previous limited efforts in phenotyping closely related germplasm lines have not yielded in the identification of hypoallergen peanut lines. Keeping in mind this fact, we explored a large number of diverse germplasm lines for phenotyping using newly developed very precise protocol (Pandey et al., 2019). As a result, this study reports low or hypoallergen

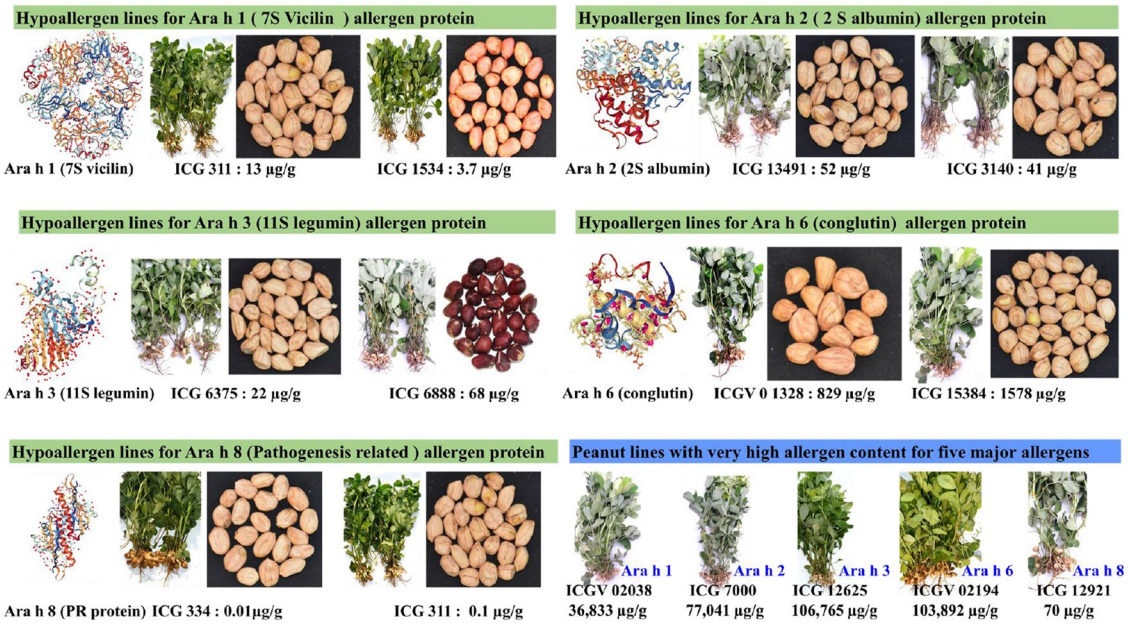

FIGURE 5 | Plant and seed features of hypoallergen and high allergen peanut lines identified for five major peanut allergens (Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8).

TABLE 4 | List of accessions having minimum and maximum allergen protein content for Ara h 6 and Ara h 8 from peanut germplasm lines.

S. No.	Genotypes	Species	Sub-species	Botanical Type	Agronomic type	Peanut allergen content (µg/g)
Low and high allergen lines for Ara h 6						
1	ICGV 01328	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	829
2	ICG 15384	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	1,577
3	ICG 15403	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	2,066
4	ICGV 01276	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	2,078
5	ICGV 02022	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	2,298
6	ICG 6647	*A. hypogaea*	*INA*	INA	INA	60,737
7	ICG 14985	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	63,867
8	ICG 6892	*A. hypogaea*	*INA*	INA	INA	87,074
9	ICG 7000	*A. hypogaea*	*INA*	INA	INA	89,782
10	ICGV 02194	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	103,892
Low and high allergen lines for Ara h 8						
1	ICG 334	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	0.01
2	ICG 311	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	0.1
3	ICG 405	*A. hypogaea*	*fastigiata*	Fastigiata	Valencia bunch	0.1
4	ICG 1487	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	0.2
5	ICG 1569	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	0.2
6	ICG 4389	*A. hypogaea*	*hypogaea*	Hypogaea	Virginia runner	18.5
7	ICG 5663	*A. hypogaea*	*hypogaea*	Hypogaea	Virginia bunch	19.4
8	ICG 4412	*A. hypogaea*	*hypogaea*	Hypogaea	Virginia runner	20.2
9	ICG 12189	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	41.26
10	ICG 12921	*A. hypogaea*	*fastigiata*	Vulgaris	Spanish bunch	70.12

S. No. 1–5 are accessions with low allergen protein content while 6–10 are accessions with high allergen protein content.

lines for five major peanut allergens for the first time. For Ara h 1, five hypoallergen lines ranged between 4 and 28 µg/g and the ICG 1534 (4 µg/g) and ICGV 02038 (36,833 µg/g) had minimum Ara h 1 allergen protein. ICG 1534 belongs to Spanish bunch while ICGV 02038 to Valencia bunch. Similarly, the best hypoallergen lines for Ara h 2 were ICG 3140 with just 52.4 µg/g allergen protein. The accessions ICG 6375 (Spanish bunch), ICGV 01328 (Valencia Bunch), and ICG 334 (Spanish Bunch) were identified as best hypoallergen lines for Ara h 3, Ara h 6, and Ara h 8, respectively. The

screening of 53 Chinese peanut cultivars through human sera, the Spanish bunch type having low peanut allergen content than the other agronomic type (Wu et al., 2016). They also reported that the Virginia type (Xinxiandahuasheng), Valencia type (Bangjihonghuasheng), Spanish type (Mangdou), and Peruvian type (Yaoshangxiaomake) are low allergen cultivars. Another study screened 35 US peanut cultivars using human antisera of the allergic patient but could not detect any significant variation (Dodo et al., 2002) which may be due to the narrow genetic base of these US cultivars derived from just two founder parents (Isleib and Wynne, 1992).

Landraces Conserve Higher Diversity for Major Peanut Allergens

The landraces have shown less allergen protein accumulation for Ara h 1 (28–13,977 μg/g), Ara h 2 (41–62,350 μg/g), Ara h 3 (22–106,765 μg/g), Ara h 6 (3,657–51,024 μg/g), and Ara h 8 (0.01–70.12 μg/g) as compared to other biological groups i.e., breeding/research material, advanced/improved cultivar, and wild accessions. The quantification of five major allergens through immunological assay showed that the landraces conserved hypoallergen feature. These accessions are ICG 442 (22.7 μg/g) for Ara h 1, ICG 13491 (41 μg/g) for Ara h 2, ICG 6375 (22 μg/g) for Ara h 3, ICG 15405 (3,657 μg/g), and ICG 334 (0.01 μg/g). These accessions mostly belong to *fastigiata* subspecies and Spanish bunch types and can be used for developing hypoallergen lines through marker-assisted selection (MAS) or clusters of regularly interspaced short palindromic repeats (CRISPR)/Cas9 approach. One previous study reported that the landraces conserved genetic variation for edible oil properties and also suitable for biodiesel production in Algerian peanut landraces (Giuffre et al., 2016). This finding provides hope to use either directly cultivating or further improvement through breeding for developing hypoallergen lines. Some of the hypoallergen lines identified in this study have also been reported having resistance to multiple stresses, e.g., ICG 442, a Spanish hypoallergen line for Ara h 1 was reported resistant to multiple abiotic stresses such as drought, salinity, and phosphorus deficiency (Upadhyaya et al., 2014).

A Sound Basis for Further Research and Cultivation of Hypoallergen Lines to Ensure Human Health From a Peanut Allergy

The development and release of several improved cultivars with high yield potential, biotic and abiotic stresses resistance, and enhanced/improved nutritional quality features in peanut has successfully been developed by combining the plant breeding techniques and efficient phenotyping methods. One of the previous studies reported that there are no significant differences in the allergen content among different peanut agronomic types consumed in western countries (Koppelman et al., 2016). However, that particular study involved very few numbers of genotypes representing various agronomic types. In our study, we used a large diverse peanut germplasm set and reporting that there are wide variation for allergen content among different agronomic types such as Spanish bunch, Valencia bunch, Virginia bunch, and Virginia runner. This study will provide hope to food industries to use hypoallergen lines in their food product preparations. Genetic improvement can be done using various modern tools and techniques through genomic research (Guo et al., 2012). Functional genomics and biotechnological techniques help discover and characterize agriculturaly important genes through deep analysis of the transcriptome, and their direct transfer to chosen cultivars (Brasileiro et al., 2014). Genes which encode storage protein, metabolic enzyme genes, genes involved in oil metabolism, and differentially expressed genes in response to pathogen stress, were identified and cloned in peanut by expressed sequence tag sequencing and are used to improve peanut production.

Wide varieties of peanut are grown to meet need of oil, food, and industries. The identified hypoallergen peanut lines can directly be used for cultivation and use in industry. Further, the identification of functional variation through genomics will facilitate the development of diagnostic markers for different allergens. The diagnostic markers can be used for improving varieties through MAS while the genes can be now edited through CRISPR/Cas9. CRISPR/Cas9 system has proven to be successful in various crop species over past years including wheat, tobacco, rice, potato, tomato sorghum, orange, and maize (Bortesi and Fischer, 2015). Although in peanut, there were no reports to implement genome editing, however, several reports of MAS and marker-assisted backcrossing (MABC) are available (Chu et al., 2011; Varshney et al., 2014; Janila et al., 2016; and Bera et al., 2018). CRISPR/Cas9 is able to introduce homozygous mutations into rice and tomato potentially accelerating crop improvement in the first generation of the transformants (Shen et al., 2014; Zhang et al., 2014). The elimination of allergen through genome editing technology would be useful for a specfiic group of customers. Silencing of Mal d 1 has decreased the allergenicity of apple, which may enhance the consumption without allergic reactions (Dubois et al., 2015). The immune dominant Ara h 2 peanut allergen successfully reduced the allergenicity in peanut through RNA interference technology (Dodo et al., 2008). All allergens coding genes should be silenced or removed in order to develop hypoallergen peanut that are safe for consumption by many patients, and the genome editing provide offers to do so effectively. The availability of hypoallergen lines will impact the peanut industry as well as contribute toward fighting the peanut allergy menace globally.

SUMMARY

The study identified several hypoallergen peanut lines for further study. These hypoallergen lines can be directly used for commercial cultivation in addition to further breeding research for developing improved peanut varieties by combining several other agronomic traits. The output of this study also encourages researchers to identify functional variation so that molecular breeding through MAS, MABC, and genome editing can be deployed for developing new hypoallergen lines in peanut. The results have shown great hope toward fighting peanut allergy and

ensuring enhanced food safety and security for humans as well as promises good opportunity for economic gains by producers, processors, and industry.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MP conceived the idea. MP, AP, HS, HU, and RV designed the experiments. AP performed the experiment. AP and HS analyzed the Data. AP, RV, HS, and MP wrote the manuscript.

ACKNOWLEDGMENTS

The authors are thankful to partial financial assistance from Indian Council of Agricultural Research (ICAR) and Department of Biotechnology (DBT), India. We acknowledge Sunil S. Gangurde for performing the cluster analysis. The work reported in this article was undertaken as part of the CGIAR Research Program on Grain Legumes and Dryland Cereals (CRP-GLDC). ICRISAT is a member of CGIAR Consortium.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01177/full#supplementary-material

REFERENCES

Ballmer-Weber, B. K., and Beye, K. (2018). Food challenges. *J. Allergy Clin. Immunol.* 141, 69–71. doi: 10.1016/j.jaci.2017.06.038

Ben-Shoshan, M., Kagan, R. S., Alizadehfar, R., Joseph, L., Turnbull, E., St Pierre, Y., et al. (2009). Is the prevalence of peanut allergy increasing? A 5-year follow-up study in children in Montreal. *J. Allergy Clin. Immunol.* 123, 783–788. doi: 10.1016/j.jaci.2009.02.004

Bera, S., Kamdar, J., Kasundra, S., Dash, P., Maurya, A., Patel, S., et al. (2018). Improving oil quality by altering levels of fatty acids through marker-assisted selection of ahfad2 alleles in peanut (Arachis hypogaea L.) varieties. *Euphytica* 214–162. doi: 10.1007/s10681-018-2241-0"10.1007/s10681-018-2241-0

Berger and Smith. (1998). Science commentary: why do some children grow out of peanut allergy? *Br. Med. J.* 316, 1275.

Bock, S. A., Muñoz-Furlong, A., and Sampson, H. A. (2001). Fatalities due to anaphylactic reactions to foods. *J. Allergy Clin. Immunol.* 107, 91–93. doi: 10.1067/mai.2001.112031

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi: 10.1016/j.biotechadv.2014.12.006

Brasileiro, A. C. M., Araujo, A. C. G., Leal-Bertioli, S. C., and Guimaraes, P. M. (2014). Genomics and genetic transformation in Arachis. *Int. J. Plant Biol. Res.* 2, 1017.

Cabanillas, B., Maleki, S. J., Rodríguez, J., Burbano, C., Muzquiz, M., Jiménez, M. A., et al. (2012). Heat and pressure treatments effects on peanut allergenicity. *Food Chem.* 132, 360–366. doi: 10.1016/j.foodchem.2011.10.093

Chu, Y., Wu, C. L., Holbrook, C. C., Tillman, B. L., Person, G., and Ozias-Akins, P. (2011). Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. *Plant Genome* 4, 110–117. doi: 10.3835/plantgenome2011.01.0001

de Jong, E. C., van Zijverden, M., Spanhaak, S., Koppelman, S. J., Pellegrom, H., and Penninks, A. H. (1998). Identification and partial characterization of multiple major allergens in peanut proteins. *Clin. Exp. Allergy* 28, 743–751. doi: 10.1046/j.1365-2222.1998.00301.x

Dodo, H., Marsic, D., Callender, M., Cebert, E., and Viquez, O. (2002). Screening 34 Peanut introductions for allergen content using ELISA. *Food Agric. Immunol.* 14, 147–154. doi: 10.1080/09540100220145179

Dodo, H., Konan, K., and Viquez, O. M. (2005). A genetic engineering strategy to eliminate peanut allergy. *Curr. Allergy Asthm. R.* 5, 67–73. doi: 10.1007/s11882-005-0058-0

Dodo, H. W., Konan, K. N., Chen, F. C., Egnin, M., and Viquez, O. M. (2008). Alleviating peanut allergy using genetic engineering: the silencing of the immuno dominant allergen Ara h 2 leads to its significant reduction and a decrease in peanut allergenicity. *Plant Biotechnol. J.* 6, 135–145. doi: 10.1111/j.1467-7652.2007.00292.x

Dubois, A. E. J., Pagliarani, G., Brouwer, R. M., Kollen, B. J., Dragsted, L. O., Eriksen, F. D., et al. (2015). First successful reduction of clinical allergenicity of food by genetic modification: Mal d 1-silenced apples cause fewer allergy symptoms than the wild-type cultivar. *Allergy* 70, 1406–1412. doi: 10.1111/all.12684

Filep, S., Block, D. S., Smith, B. R. E., King, E. M., Commins, S., Kulis, M., et al. (2018). Specific allergen profiles of peanut foods and diagnostic or therapeutic allergenic products. *J. Allergy Clin. Immunol.* 141, 626–663. doi: 10.1016/j.jaci.2017.05.049

Flinterman, A. E., van Hoffen, E., den Hartog Jager, C. F., Koppelman, S., Pasmans, S. G., Hoekstra, M. O., et al. (2007). Children with peanut allergy recognize predominantly Ara h2 and Ara h6, which remains stable over time. *Clin. Exp. Allergy* 37, 1221–1228. doi: 10.1111/j.1365-2222.2007.02764.x

Galili, T. (2015). Dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinf.* 31, 3718–3720. doi: 10.1093/bioinformatics/btv428

Giuffrè, A. M., Tellah, S., Capocasale, M., Zappia, C., Latati, M., Badiani, M., et al. (2016). Seed oil from ten Algerian peanut landraces for edible use and biodiesel production. *J. Oleo Sci.* 65, 9–20. doi: 10.5650/jos.ess15199

Grundy, J., Matthews, S., Bateman, B., Dean, T., and Arshad, S. H. (2002). Rising prevalence of allergy to peanut in children: Data from 2 sequential cohorts. *J. Allergy Clin. Immunol.* 110, 784–789. doi: 10.1067/mai.2002.128802

Guo, B. Z., Chen, C. Y., Chu, Y., Holbrook, C. C., Ozias-Akins, P., and Stalker, H. T. (2012). Advances in genetics and genomics for sustainable peanut production, in *Sustainable Agriculture and New Biotechnologies* (Boca Raton, Fla, USA: CRC Press), 341–367. doi: 10.1201/b10977-16

Hamilton, R. G., MacGlashan, D. W., and Saini, S. S. (2010). IgE antibody-specific activity in human allergic disease. *Immunol. Res.* 47, 273–284. doi: 10.1007/s12026-009-8160-3

Hefle, S. L., Nordlee, J. A., and Taylor, S. L. (1996). Allergenic foods. *Crit. Rev. Food Sci. Nutr.* 36, S6989. doi: 10.1080/10408399609527760

Holbrook, C. C., and Stalker, H. T. (2003). Peanut breeding and genetic resources. *Plant Breed. Rev.* 22, 297–356. doi: 10.1002/9780470650202.ch6

Iqbal, A., Shah, F., Hamayun, M., Ahmad, A., Hussain, A., Waqas, M., et al. (2016). Allergens of Arachis hypogaea and the effect of processing on their detection by ELISA. *Food Nutr. Res.* 60, 28945. doi: 10.3402/fnr.v60.28945

Isleib, T. G., and Wynne, J. C. (1992). Use of plant introduction in peanut improvement,” in *Use of Plant Introduction to Cultivar Development, Part 2* (MADISON, WI, USA: CSSA Spec. publication 20 of Crop Science Society of America).

Janila, P., Pandey, M. K., Shasidhar, Y., Variath, M. T., Sriswathi, M., Khera, P., et al. (2016). Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. *Plant Sci.* 242, 203–213. doi: 10.1016/j.plantsci.2015.08.013

Kagan, R. S., Joseph, L., Dufresne, C., ray-Donald, K., Turnbull, E., St Pierre, Y., et al. (2003). Prevalence of peanut allergy in primary-school children in Montreal, Canada. *J. Allergy Clin. Immuno.* 112, 1223– 1228. doi: 10.1016/j.jaci.2003.09.026

Kondo, Y., and Urisu, A. (2009). Oral allergy syndrome. *Allergol Int.* 58, 485–491. doi: 10.2332/allergolint.09-RAI-0136

Koppelman, S. J., Knol, E. F., Vlooswijk, R. A. A., Wensing, M., Knulst, A. C., Hefle, S. L., et al. (2003). Peanut allergen Ara h 3: Isolation from peanuts and biochemical characterization. *Allergy* 58, 1144–1151. doi: 10.1034/j.1398-9995.2003.00259.x

Koppelman, S. J., de Jong, G. A., Laaper-Ertmann, M., Peeters, K. A. B. M., Knulst, A. C., Hefle, S. L., et al. (2005). Purification and immunoglobulin E-binding properties of peanut allergen Ara h 6: Evidence for cross-reactivity with Ara h 2. *Clin. Exp. Allergy* 35, 490–497. doi: 10.1111/j.1365-2222.2005.02204.x

Koppelman, S. J., Jayasena, S., Luykx, D., Schepens, E., Apostolovic, D., De Jong, G. A., et al. (2016). Allergenicity attributes of different peanut market types. *Food Chem. Toxicol.* 1, 82–90. doi: 10.1016/j.fct.2016.02.016

Krause, S., Latendorf, T., Schmidt, H., Darcan-Nicolaisen, Y., Reese, G., Petersen, A., et al. (2010). Peanut varieties with reduced Ara h 1 content indicating no reduced allergenicity. *Mol. Nutr. Food Res.* 54, 3817. doi: 10.1002/mnfr.200900072

Lanser, B. J., Wright, B. L., Orgel, K. A., Vickery, B. P., and Fleischer, D. M. (2015). Current options for the treatment of food allergy. *Pediatr. Clin.* 62, 1531–1549. doi: 10.1016/j.pcl.2015.07.015

Leung, A. S. Y., Leung, N. Y. H., Wai, C. Y. Y., Leung, T. F., and Wong, G. W. K. (2018). Allergen immunotherapy for food allergy from the Asian perspective: key challenges and opportunities. *Expert Rev. Clin. Immunol.* 15, 153–164. doi: 10.1080/1744666X

Lieberman, J. A., and Sicherer, S. H. (2011). Diagnosis of food allergy: epicutaneous skin tests, in vitro tests, and oral food challenge. *Curr. Allergy Asthma Rep.* 11, 58– 64. doi: 10.1007/s11882-010-0149-4

Liew, W. K., Chiang, W. C., Goh, A. E., Lim, H. H., Chay, O. M., Chang, S., et al. (2013). Paediatric anaphylaxis in a Singaporean children cohort: changing food allergy triggers over time. *Asia Pac. Allergy* 3, 29–34. doi: 10.5415/apallergy.2013.3.1.29

Mahesh, P. A., Wong, G. W. K., Ogorodova, L., Potts, J., Leung, T. F., Fedorova, O., et al. (2016). Prevalence of food sensitization and probable food allergy among adults in India: the EuroPrevall INCO study. *Allergy* 71, 1010–1019. doi: 10.1111/all.12868

Maleki, S. J., Chung, S. Y., Champagne, E. T., and Raufmann, J. P. (2000). The effects of roasting on the allergenic properties of peanut proteins. *J. Allergy Clin. Immunol.* 106, 763–768. doi: 10.1067/mai.2000.109620

Mittag, D., Akkerdaas, J., Ballmer-Weber, B. K., Vogel, L., Wensing, M., Becker, W. M., et al. (2004). Ara h 8, a Bet v 1-homologous allergen from peanut, is a major allergen in patients with combined birch pollen and peanut allergy. *J. Allergy Clin. Immunol.* 114, 1410–1417. doi: 10.1016/j.jaci.2004.09.014

Morisset, M., Moneret-Vautrin, D. A., and Kanny, G. (2005). Prevalence of peanut sensitization in a population of 4,737 subjects: an Allergo-Vigilance Network enquiry carried out in 2002. *Eur. Ann. Allergy Clin. Immunol.* 37, 54–57.

Nicolaou, N., Murray, C., Belgrave, D., Poorafshar, M., Simpson, A., and Custovic, A. (2011). Quantification of specific IgE to whole peanut extract and peanut components in prediction of peanut allergy. *J. Allergy Clin. Immunol.* 127, 684–685. doi: 10.1016/j.jaci.2010.12.012

Osterballe, M., Hansen, T. K., Mortz, C. G., Host, A., and Bindslev-Jensen, C. (2005). The prevalence of food hypersensitivity in an unselected population of children and adults. *Pediatr. Allergy Immunol.* 16, 567–573. doi: 10.1111/j.1399-3038.2005.00251.x

Palmer, G. W., Dibbern, D. A.Jr., Burks, A. W., Bannon, G. A., Bock, S. A., Porterfield, H. S., et al. (2005). Comparative potency of Ara h 1 and Ara h 2 in immunochemical and functional assays of allergenicity. *Clin. Immunol.* 115, 302–312. doi: 10.1016/j.clim.2005.02.011

Pandey, M. K., and Varshney, R. K. (2018). Groundnut entered post-genome sequencing era: opportunities and challenges in translating genomic information from genome to field, in *Biotechnologies in Crop Improvement*, vol. 3 . Eds. Gosal, S. H., and Wani, S. S. (New York City: Springer International Publishing). doi: 10.1007/978-3-319-94746-4_9

Pandey, M. K., Monyo, E., Ozias-Akins, P., Liang, X., Guimarães, P., Nigam, S. N., et al. (2012). Advances in Arachis genomics for peanut improvement. *Biotechnol. Adv.* 30, 639–651. doi: 10.1016/j.biotechadv.2011.11.001

Pandey, A. K., Varshney, R. K., Sudini, H. K., and Pandey, M. K. (2019). An improved enzyme-linked immunosorbent assay (ELISA) based protocol using seeds for detection of five major peanut allergens Ara h 1, Ara h 2, Ara h 3, Ara h 6 and Ara h 8. *Front. Nutr.* 6, 68. doi: 10.3389/fnut.2019.00068

Peeters, K. A., Koppelman, S. J., Van Hoffen, E., Van Der Tas, C. W., den Hartog Jager, C. F., Penninks, A. H., et al. (2007). Does skin prick test reactivity to purified allergens correlate with clinical severity of peanut allergy? *Clin. Exp. Allergy* 37, 108–115. doi: 10.1111/j.1365-2222.2006.02628.x

Pele, M. (2010). Peanut allergens. *Romanian Biotechnol. Lett.* 15, 5205.

Porterfield, H. S., Murray, K. S., Schlichting, D. G., Chen, X., Hansen, K. C., Duncan, M. W., et al. (2009). Effector activity of peanut allergens: a critical role for Ara h 2, Ara h 6, and their variants. *Clin. Exp. Allergy* 39, 1099–1108. doi: 10.1111/j.1365-2222.2009.03273.x

Riecken, S., Lindner, B., Petersen, A., Jappe, U., and Becker, W. M. (2008). Purification and characterization of natural Ara h 8, the Bet v 1 homologous allergen from peanut, provides a novel isoform. *Biol. Chem.* 389, 415–423. doi: 10.1515/BC.2008.038

Sampson, H. A., Shreffler, W. G., Yang, W. H., Sussman, G. L., Brown-Whitehorn, T. F., Nadeau, K. C., et al. (2017). Effect of varying doses of epicutaneous immunotherapy vsplacebo on reaction to peanut protein exposure among patients with peanut sensitivity: a randomized clinical trial. *JAMA* 318, 1798–1809. doi: 10.1001/jama.2017.16591

Satitsuksanoa, P., Globinska, A., Jansen, K., van de Veen, W., and Akdis, M. (2018). Modified allergens for immunotherapy. *Curr. Allergy Asthma Rep.* 18, 9. doi: 10.1007/s11882-018-0766-x

Shek, L. P., Cabrera-Morales, E. A., Soh, S. E., Gerez, I., Ng, P. Z., Yi, F. C., et al. (2010). A population-based questionnaire survey on the prevalence of peanut, tree nut, and shellfish allergy in 2 Asian populations. *J. Allergy Clin. Immunol.* 126, 24–31. doi: 10.1016/j.jaci.2010.06.003

Shen, B., Zhang, W. S., Zhang, J., Zhou, J., Wang, J., Chen, L., et al. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Meth.* 11, 399–402. doi: 10.1038/nmeth.2857

Sicherer, H. S., and Sampson, A. H. (2007). Peanut allergy: emerging concepts and approaches for an apparent epidemic. *J. Allergy Clin. Immunol.* 120, 491–503. doi: 10.1016/j.jaci.2007.07.015

Sicherer, S.H., and Sampson, H.A. (2010). Food allergy. *J. Allergy Clin. Immunol.* 125, 116–125. doi: 10.1016/j.jaci.2009.08.028

Sicherer, H. S., and Sampson, A. H. (2014). Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *J. Allergy Clin. Immunol.* 133, 291–307. doi: 10.1016/j.jaci.2013.11.020

Sicherer, S. H., Burks, A. W., and Sampson, H. A. (1998). Clinical features of acute allergic reactions to peanut and tree nuts in children. *Pediatr.* 102, e6. doi: 10.1542/peds.102.1.e6

Sicherer, S. H., Munoz-Furlong, A., Godbold, J. H., and Sampson, H. (2010). US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. *J. Allergy Clin. Immunol.* 125, 1322–1326. doi: 10.1016/j.jaci.2010.03.029

Tscheppe, A., and Breiteneder, H. (2017). Recombinant allergens in structural biology, diagnosis, and immunotherapy. *Int. Arch. Allergy Immunol.* 172, 187–202. doi: 10.1159/000464104

Upadhyaya, H. D. (2009).*Reference Set of Groundnut. Grain Legumes (ICRISAT-Web Article)*. pp. 1–1.

Upadhyaya, H. D., Bramel, P. J., Ortiz, R., and Singh, S. (2002). Developing a mini core of peanut for utilization of genetic resources. *Crop Sci.* 42, 2 150–22156. doi: 10.2135/cropsci2002.2150

Upadhyaya, H. D., Dwivedi, S. L., Vadez, V., Hamidou, F., Singh, S., Varshney, R. K., et al. (2014). Multiple resistant and nutritionally dense germplasm identified from mini core collection in peanut. *Crop Sci.* 54, 679–693. doi: 10.2135/cropsci2013.07.0493

Upadhyaya, H. D., Ortiz, R., Bramel, P. J., and Singh, S. (2003). Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. *Genet. Resour. Crop Evol.* 50, 139–148. doi: 10.1023/A:1022945715628

Upadhyaya, H. D., Yadav, D., Dronavalli, N., Gowda, C. L. L., and Singh, S. (2010). Mini core germplasm collections for infusing genetic diversity in plant breeding programs. *Electron J. Plant Breed.* 1, 1294–1309.

Varshney, R. K., Pandey, M. K., Janila P., Nigam, S. N., Sudini, H., Gowda, M. V. C., et al. (2014). Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 127, 1771–1781. doi: 10.1007/s00122-014-2338-3

Varshney, R. K., Pandey, M. K., Bohra, A., Singh, V. K., Thudi, M., and Saxena, R. K. (2019). Toward sequence-based breeding in legumes in post-genome sequencing era. *Theor. Appl. Genet.* doi: 10.1007/s00122-018-3252-x

Varshney, R. K., Thudi, M., Pandey, M. K., Tardieu, F., Ojiewo, C., Vadez, V., et al. (2018). Accelerating genetic gains in legumes for prosperous smallholder agriculture: integrating genomics, phenotyping, systems modelling and agronomy. *J. Exp. Bot.* 69, 3293–3312. doi: 10.1093/jxb/ery088

Vereda, A., van Hage, M., Ahlstedt, S., Ibañez, M. D., Cuesta-Herranz, J., van Odijk, J., et al. (2011). Peanut allergy: Clinical and immunologic differences among patients from 3 different geographic regions. *J. Allergy Clin. Immunol.* 127, 603–607. doi: 10.1016/j.jaci.2010.09.010

Warner, J. O. (1999). Peanut allergy: A major public health issue. *Pediatr. Asthma Allergy Immunol.* 10, 14–20. doi: 10.1034/j.1399-3038.1999.101011.x

Warnes, G. R., Bolker, B., Lodewijk, B., Gentleman, R., Liaw, W. H. A., Lumley, T., et al. (2016). *Package 'gplots "Various R Programming Tools for Plotting Data" Repository CRAN*, Vol. 21. balloonplot pp. 04–18.

Wen, H. W., Borejsza-Wysocki, W., DeCory, T. R., and Durst, R. A. (2007). Peanut allergy, peanut allergens, and methods for the detection of peanut contamination in food products. *Compr. Rev. Food Sci. Food Saf.* 6, 47–58. doi: 10.1111/j.1541-4337.2007.00017.x

Wu, Z., Zhou, N., Xiong, F., Li, X., Yang, A., Tong, P., et al. (2016). Allergen composition analysis and allergenicity assessment of Chinese peanut cultivars. *Food Chem.* 196, 459–465. doi: 10.1016/j.foodchem.2015.09.070

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., et al. (2014). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotech. J.* 12, 797–807. doi: 10.1111/pbi.12200

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Pandey, Sudini, Upadhyaya, Varshney and Pandey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 13 March 2020
doi: 10.3389/fgene.2020.00229

The Genetic Basis and Nutritional Benefits of Pigmented Rice Grain

*Edwige Gaby Nkouaya Mbanjo[1,2], Tobias Kretzschmar[3], Huw Jones[4], Nelzo Ereful[4], Christopher Blanchard[5], Lesley Ann Boyd[4] and Nese Sreenivasulu[1]**

[1] International Rice Research Institute, Los Baños, Philippines, [2] International Institute for Tropical Agriculture, Ibadan, Oyo, Nigeria, [3] Southern Cross Plant Science, Southern Cross University, Lismore, NSW, Australia, [4] National Institute of Agricultural Botany, Cambridge, United Kingdom, [5] School of Biomedical Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia

Improving the nutritional quality of rice grains through modulation of bioactive compounds and micronutrients represents an efficient means of addressing nutritional security in societies which depend heavily on rice as a staple food. White rice makes a major contribution to the calorific intake of Asian and African populations, but its nutritional quality is poor compared to that of pigmented (black, purple, red orange, or brown) variants. The compounds responsible for these color variations are the flavonoids anthocyanin and proanthocyanidin, which are known to have nutritional value. The rapid progress made in the technologies underlying genome sequencing, the analysis of gene expression and the acquisition of global 'omics data, genetics of grain pigmentation has created novel opportunities for applying molecular breeding to improve the nutritional value and productivity of pigmented rice. This review provides an update on the nutritional value and health benefits of pigmented rice grain, taking advantage of both indigenous and modern knowledge, while also describing the current approaches taken to deciphering the genetic basis of pigmentation.

Keywords: pigmented rice grain, nutrition, flavonoids, metabolites, genetics

OPEN ACCESS

Edited by:
Manish Kumar Pandey, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

Reviewed by:
Larry Parnell, Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, United States
Yue-ie Caroline Hsing, Academia Sinica, Taiwan

***Correspondence:**
Nese Sreenivasulu
n.sreenivasulu@irri.org

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Genetics

Received: *20 May 2019*
Accepted: *26 February 2020*
Published: *13 March 2020*

Citation:
Mbanjo EGN, Kretzschmar T, Jones H, Ereful N, Blanchard C, Boyd LA and Sreenivasulu N (2020) The Genetic Basis and Nutritional Benefits of Pigmented Rice Grain. Front. Genet. 11:229. doi: 10.3389/fgene.2020.00229

INTRODUCTION

Rice is a staple food for over half of the world's population (World Rice Production, 2019). Meeting the demand of future rice supply for the growing population, which has been predicted to reach 9.7 billion by 2050[1], is central for ensuring food and nutritional security. In addition to its critical importance to Asian populations as a source of food, rice also features in a range of social, cultural, economic, and religious activities (Ahuja et al., 2007; Hedge et al., 2013; Sathya, 2013). In sub-Saharan Africa the consumption of rice is projected to grow from its current level of 27–28 Mt per year to around 36 Mt by the end of 2026 (Terungwa and Yuguda, 2014; Nigatu et al., 2017), replacing some of the current demand for cassava, yam, maize, millet, and sorghum.

Most of the nutrients found in rice grain accumulate in the outer aleurone layer and embryo, the endosperm being composed primarily of starch. The process of dehulling and milling discards most micronutrients, fatty acids, anti-oxidants, and fiber. As a result, diets over-reliant on white rice risk deficiencies for several nutritional factors (Verma and Shukla, 2011; Sharma et al., 2013; Saneei et al., 2016; Sarma et al., 2018). The focus of rice breeding has long been concentrated on improving the crop's productivity, although some emphasis has been given to improving the size, shape, and amylose content of the grain (Breseghello, 2013; Rao et al., 2014). The nutritional quality of the

[1] https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html

grain produced by certain traditional landraces has been shown to be higher than that of the grain produced by conventional, modern rice varieties, largely due to their more effective accumulation of bioactive compounds (Bhat and Riar, 2015; Berni et al., 2018). A growing consumer interest in health-promoting food products is generating a substantial market for more nutritionally valuable rice, creating health benefits for the large number of people for whom rice is a staple, while simultaneously generating economic benefits for the producers (Terungwa and Yuguda, 2014). As a result, the focus of a number of major rice research programs is turning to the issue of nutritional quality, encompassing an improved micronutrient and anti-oxidant content, along with a reduction in the grains' glycemic index.

This review explores the nutritional and health attributes of pigmented rice grain, based on both indigenous knowledge and current research, and discusses the potential of pigmented rice grain to address nutritional food security. In addition, it explores the genetic basis of grain pigmentation, and suggests the potential contribution which 'omics technologies can make to address the challenge of the double burden of malnutrition.

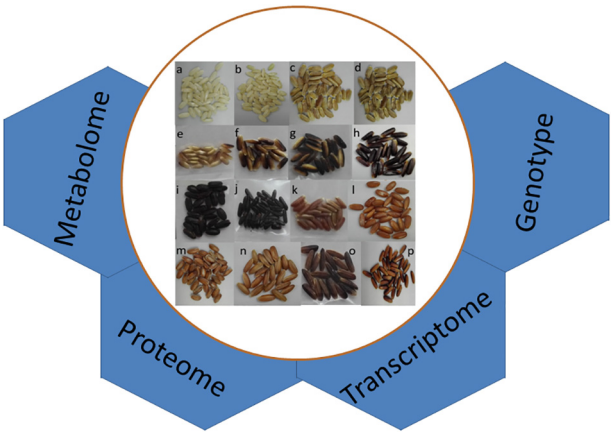

FIGURE 1 | Genetic variation for grain pigmentation in rice. Grains featuring **(a,b)** white, **(c,d)** brown, **(e–h)** purple, **(i,j)** dark purple or black, **(k–n)** red, and **(o,p)** mixed colored pericarp. The application of various genomic approaches to understand the genetic pathway of grain pigmentation is outlined.

INDIGENOUS KNOWLEDGE, COMPLEMENTED WITH CORROBORATED SCIENTIFIC EVIDENCE, INFORMS ON THE POTENTIAL OF PIGMENTED RICE GRAIN TO IMPROVE NUTRITION AND HEALTH

Indigenous diets have developed to meet the needs of local communities over a long period of time, and the knowledge associated with these should be viewed as a resource to inform the discussion concerning the place of rice in the modern diet (Berni et al., 2018; Khatri, 2018). The value of such landraces in the context of both human nutrition and health (Rahman et al., 2006; Chunta et al., 2014) can be exemplified by the proven advantages of consuming pigmented grain (**Figure 1**; Rahman et al., 2006; Umadevi et al., 2012; Sathya, 2013). In particular, pigmented rice has been associated with anti-inflammatory and diuretic activity (Umadevi et al., 2012). Based on native indigenous knowledge, it has also been recommended for the treatment of diarrhea, vomiting, fever, hemorrhaging, chest pain, wounds, burns, and gastrointestinal problems, as well as addressing various liver and kidney disorders (Hedge et al., 2013; Sathya, 2013). Certain pigmented rice varieties are still used to treat skin diseases, blood pressure, fever, paralysis, rheumatism, and leucorrhea, and even as the basis of a general health tonic (Ahuja et al., 2007). In the Philippines, "tiki tiki," derived from rice bran, has been used to cure thiamine deficiency (Umadevi et al., 2012). In India, the grain of pigmented rice landraces is offered to lactating mothers, and is used to both treat jaundice and cure paralysis. The rice variety "Laicha" was given its name because of its ability to prevent an eponymous skin disease (Das and Oudhia, 2000). For more than 2,000 years, grain of the South Indian landrace "Kavuni" has been reported to exhibit anti-oxidant, anti-arthritic, and anti-diabetic properties, and has been used to cure gastritis and peptic ulcers, as well as to enhance blood circulation (Valarmathi et al., 2014; Hemamalini et al., 2018).

A number of scientifically based studies have provided evidence to support the hypothesis that pigmented rice grain possesses anti-oxidant, anti-diabetic, anti-hyperlipidemic, and anti-cancer activity (Baek et al., 2015; Boue et al., 2016), which is reviewed below.

Anti-oxidant Activity

Dietary anti-oxidants represent an effective means of combating the accumulation of harmful reactive oxygen species and of balancing the redox status of the body (Krishnanunni et al., 2014). Analysis of extracts made from pigmented rice grain has shown that the phenolic compounds tocopherol and anthocyanin are efficient neutralizers of reactive oxygen species (Zhang et al., 2015; Ghasemzadeh et al., 2018), while animal tests have proven that these compounds are bioavailable (Tantipaiboonwong et al., 2017). Several studies have shown that the elevated anti-oxidation activity exhibited by pigmented rice grains (most markedly by black rice) can be used to mitigate the inflammatory response (Chakuton et al., 2012; Petroni et al., 2017).

Anti-diabetic Activity

The grain of some traditional pigmented rice varieties have proven to be effective in supporting glucose homeostasis, and are thus useful for the management of diabetes mellitus (Hemamalini et al., 2018). Unlike white rice grain consumption, which raises blood glucose levels, consuming pigmented grain can reduce blood glucose levels. Extracts of pigmented rice grain and bran have been shown to effectively inhibit the activity of endogenous α-amylase and α-glucosidase, thereby inhibiting the conversion of starch to glucose in the small intestine, which acts as a source of resistant starch to be utilized by

gut microbiota in the colon (Boue et al., 2016; Chiou et al., 2018). While extracts made from both red and purple grain have been reported to inhibit α-glucosidase activity, only the former was effective in also inhibiting α-amylase activity (Valarmathi et al., 2014; Boue et al., 2016). The anthocyanins found in the whole grain of black rice acted as a potent inhibitor of β-glucosidase, thus delaying the absorption of carbohydrates (Chandramouli et al., 2018). Extracts of black rice bran have also been shown to induce the repair and regeneration of pancreatic beta cells (Wahyuni et al., 2016). Overall, the anti-diabetic effects of pigmented rice seem to arise from a synergistic effect of anthocyanin, proanthocyanidin, vitamin E, γ-oryzanol, and various flavonoids (Tantipaiboonwong et al., 2017). Black rice extracts reduced blood glucose levels more quickly than did extracts from red rice, a difference which was attributed to the presence of cyanidin 3-glucoside, a compound which activates insulin sensitivity, glucose uptake, and adiponectin secretion (Matsukawa et al., 2016; Tantipaiboonwong et al., 2017). However, many of the black rice are low in its amylose content and upon milling most of the anthocyanins accumulated in aleurone will be lost, thus not necessarily would possess low GI property when consumed in the form of milled rice (Kang et al., 2011).

Anti-cancer Activity

A considerable body of evidence suggests that consumption of pigmented rice has a protective effect against certain cancers. Ghasemzadeh et al. (2018) demonstrated that extracts of both black and red rice inhibit the proliferation of breast cancer cells. The phenolic acids, flavonoids, anthocyanins, and phytic acid present in extracts of purple rice bran have been shown to act as anti-mutagens and potential suppressors of cancer. It has been proposed that these phytochemicals act by either blocking the carcinogenetic cytochromes P450 CYP1A1 and CYP1B1 and/or by effectively scavenging free radicals (Insuan et al., 2017). Bioactive compounds of pigmented grains can reduce the viability of cancer cells and even induce their apoptosis. The mechanistic basis of this effect has been found to be variety-dependent, reflecting differences in the spectrum of bioactive compounds present in each rice variety (Baek et al., 2015). The high anthocyanin content of purple rice has been associated with an inhibitory effect on the growth of human hepatocellular carcinoma cells (Banjerdpongchai et al., 2013), while extracts of purple rice bran were able to block the first stage of aflatoxin B1-initiated hepatocarcinogenesis by inhibiting key metabolic activating enzymes (Suwannakul et al., 2015). Extracts of red rice have been shown to limit the invasiveness of cancer cells in a dose-dependent manner (Pintha et al., 2014). The phytosterols 24-methylenecycloartanol, β-sitosterol, gramisterol, campesterol, stigmasterol, cycloeucalenol, 24-methylene-ergosta-5-en-3β-ol, and 24-methylene-ergosta-7-en-3β-ol, all of which are present in extracts of black rice bran, have also been reported to be effective as agents restricting the proliferation of murine leukemic cells (Suttiarporn et al., 2015). Consequently, one of the long-term strategies proposed by Luo et al. (2014) to prevent breast cancer metastasis relies on the inclusion of pigmented rice in the human diet.

THE BIOCHEMICAL PROPERTIES OF PIGMENTED RICE GRAIN

Phytosterols, Carotenoids, Vitamins, and Micronutrients in Pigmented Rice Grain

Phytosterols

Rice grains contain a wide range of secondary metabolites (**Table 1**). Pigmented grain appears to accumulate a higher level of γ-oryzanol than does non-pigmented grain (Chakuton et al., 2012). The grain accumulates the active anti-oxidant γ-oryzanol, which comprises a mixture of several phytosteryl ferulates (Chakuton et al., 2012), in particular 24-methylenecycloartanyl ferulate, cycloartenyl ferulate, campesteryl ferulate, and β-sitosteryl ferulate (Zubair et al., 2012; Pereira-Caro et al., 2013). The most important nutritional benefit of the phytosterols is their ability to both inhibit the absorption of cholesterol and to control the blood's content of undesirable lipoproteins (Jesch and Carr, 2017). The predominant phystosterols detected in commercial rice varieties are β-sitosterol, followed by campesterol, Δ^5-avenasterol, and stigmasterol (Zubair et al., 2012). The bran of the black rice variety "Riceberry" also harbors three additional sterols, namely 24-methylene-ergosta-5-en-3β-ol, 24-methylene-ergosta-7-en-3β-ol, and fucosterol (Suttiarporn et al., 2015).

Carotenoids

Carotenoids represent another class of nutritionally beneficial compounds (Roberts and Dennison, 2015). Lutein and zeaxanthin represent together >90% of the carotenoids produced by rice, with carotenes, lycopenes, and β-carotene present in trace amounts (Pereira-Caro et al., 2013; Melini and Acquistucci, 2017). Most of this class of compound is present in the bran, with little or no carotenoids being found in milled rice (Petroni et al., 2017). Grain carotenoid content is a genetically variable trait, and is strongly correlated with grain pigmentation (Ashraf et al., 2017). Red and black rice accumulate a particularly high carotenoid content, while white rice accumulates very little (Ashraf et al., 2017; Petroni et al., 2017).

Vitamins

Rice grain represents a good source of vitamin E, including both the tocopherols and the tocotrienols (Zubair et al., 2012). The β- and γ-tocotrienols are the most abundant forms present in rice (Irakli et al., 2016). According to Gunaratne et al. (2013), red rice grains harbor higher levels of total tocopherol and tocotrienol than do the grains of modern white rice varieties. Note, however, that dehulling and milling strongly reduce the tocopherol content of the grain (Zubair et al., 2012).

Micronutrients

Rice grain contains traces of a number of essential micronutrients, namely zinc, magnesium, iron, copper, potassium, manganese, and calcium (**Table 1**; Raghuvanshi et al., 2017; Shozib et al., 2017; Shao et al., 2018). Some genetic variation in mineral content has been reported; but in general, pigmented rice grain accumulates higher amounts than does white grain rice (Shozib et al., 2017). Other studies have

TABLE 1 | Bioactive and nutritional compounds identified in pigmented rice.

Compound	PubChem CID	Compound class	Figure 2[a]	References
Cyanine 3-glucoside	197081	Anthocyanin	✓	Pereira-Caro et al., 2013;
Peonidin-3-glucoside	443654	Anthocyanin	✓	Zhang et al., 2015;
Cyanidin	128861	Anthocyanin		Tantipaiboonwong et al., 2017;
Cyanidin-3,5-diglucoside	44256718	Anthocyanin	✓	Shao et al., 2018
Cyanidin-3-*O*-(6"-*O*-*p*-coumaroyl)glucoside		Anthocyanin	✓	
Pelargonidin-3-*O*-glucoside	443648	Anthocyanin		
Peonidin-3-*O*-(6"-*O*-*p*-coumaroyl)glucoside		Anthocyanin	✓	
Cyanidin-3-*O*-arabidoside		Anthocyanin	✓	
Flavone	10680	Flavone		
Luteolin-6/8-*C*-pentoside-6/8-*C*-hexoside (2 isomers)		Flavone		Pereira-Caro et al., 2013;
Apigenin-6/8-*C*-pentoside-8/6-*C*-hexoside (three isomers)		Flavone glycoside		Kim et al., 2014;
Apigenin-6-*C*-glucosyl-8-*C*-arabinoside		Flavone		Ghasemzadeh et al., 2018;
Tricin-*O*-rhamnoside-*O*-hexoside		Flavone		Poulev et al., 2019
Tricin	5281702	Flavone	✓	
Chrysoeriol	5280666	Flavone	✓	
Luteolin	5280445	Flavone	✓	
Apigenin	5280443	Flavone	✓	
Caffeic acid	689043	Hydrocinnamic acid	✓	Gunaratne et al., 2013;
p-Coumaric acid	637542	Hydrocinnamic acid	✓	Zhang et al., 2015;
Ferulic acid	445858	Hydrocinnamic acid	✓	Irakli et al., 2016;
Sinapic acid	637775	Hydrocinnamic acid	✓	Tantipaiboonwong et al., 2017;
Isoferulic acid	736186	Hydrocinnamic acid		Chiou et al., 2018;
Chlorogenic acid	1794427	Hydrocinnamic acid		Ghasemzadeh et al., 2018; Shao et al., 2018
2,5-Dihydroxybenzoic acid	3469	Hydroxybenzoic acid	✓	Kim et al., 2014;
p-Hydroxybenzoic acid	135	Hydroxybenzoic acid	✓	Valarmathi et al., 2014;
Gallic acid	370	Hydroxybenzoic acid	✓	Suwannakul et al., 2015;
Vanillic acid	8468	Hydroxybenzoic acid	✓	Huang and Lai, 2016;
Syringic acid	10742	Hydroxybenzoic acid	✓	Irakli et al., 2016;
Protocatechuic acid	72	Hydroxybenzoic acid	✓	Tantipaiboonwong et al., 2017;
Salicylic acid	338	Hydroxybenzoic acid	✓	Ghasemzadeh et al., 2018;
β-Resorcylic acid	1491	Hydroxybenzoic acid		Shao et al., 2018
Protocatechualdehyde	8768	Phenolic aldehyde		Huang and Lai, 2016
8-5′-Coupled diferulic acid		Phenolic dehydrodimer		Zhang et al., 2015
5-5′-Coupled diferulic acid		Phenolic dehydrodimer		
8-8′-Coupled diferulic acid benzofuran form		Phenolic dehydrodimer		
Proanthocyanidin dimer		Proanthocyanin	✓	Gunaratne et al., 2013
Proanthocyanidin trimer		Proanthocyanin	✓	
Catechin	73160	Flavanonol	✓	Tantipaiboonwong et al., 2017;
Epicatechin	72276	Flavanonol	✓	Ghasemzadeh et al., 2018
Quercetin	5280343	Flavonol		Pereira-Caro et al., 2013;
Quercetin-3-*O*-glucoside		Flavonol		Valarmathi et al., 2014;
Quercetin-3-*O*-rutinoside		Flavonol		Chiou et al., 2018;
Isorhamnetin-3-*O*-glucoside	5318645	Flavonol		Ghasemzadeh et al., 2018;
Myricetin	5281672	Flavonol		Poulev et al., 2019
Rutin	5280805	Flavonol		
Kaempferol	5280863	Flavonol		
Kaempferide	5281666	Flavonol		
Naringenin	932	Flavanone	✓	Chiou et al., 2018

(Continued)

TABLE 1 | Continued

Compound	PubChem CID	Compound class	Figure 2[a]	References
Cycloartenol ferulate	134695320	γ-Oryzanol		Chakuton et al., 2012;
24-Methylenecycloartenol ferulate	9920169	γ-Oryzanol		Gunaratne et al., 2013;
Campesteryl ferulate	15056832	γ-Oryzanol		Pereira-Caro et al., 2013
β-Sitosteryl ferulate	9938436	γ-Oryzanol		
Δ7-Campesteryl ferulate		γ-Oryzanol		
Campestanyl ferulate	13786591	γ-Oryzanol		
Sitostanyl ferulate	11227138	γ-Oryzanol		
Phytic acid	890	Phytic acid		Chakuton et al., 2012; Insuan et al., 2017
Tocotrienols (α-, β-, γ-, δ-forms)	9929901	Vitamin E		Zubair et al., 2012;
Tocopherols (α-, β-, γ-, δ-forms)	14986	Vitamin E		Gunaratne et al., 2013; Irakli et al., 2016
Riboflavin	493570	Vitamin B2		Valarmathi et al., 2014
Nicotinic acid	938	Vitamin B3		Kim et al., 2014
Lutein	5281243	Carotenoid		Pereira-Caro et al., 2013;
Zeaxanthin	5280899	Carotenoid		Valarmathi et al., 2014;
β-Carotene	5280489	Carotenoid		Irakli et al., 2016;
Lycopene	446925	Carotenoid		Melini and Acquistucci, 2017
β-Carotene-4,4′-dione		Carotenoid		
all-trans-3,3',4,4'-Tetrahydrospirilloxanthin	5366411	Carotenoid		
10′-Apo-β-carotenoic acid		Carotenoid		
24-Methylene-ergosta-5-en-3β-ol		Phytosterol		Suttiarporn et al., 2015
24-Methylene-ergosta-7-en-3β-ol		Phytosterol		
Fucosterol	5281326	Phytosterol		
Gramisterol	5283640	Phytosterol		
Campesterol	173183	Phytosterol		
Stigmasterol	5280794	Phytosterol		
β-Sitosterol	222284	Phytosterol		
Cycloeucalenol	101690	Triterpenoid		Suttiarporn et al., 2015
Lupenone	92158	Triterpenoid		
Lupeol	259846	Triterpenoid		
24-Methylenecycloartanol	94204	Triterpenoid		
LysoPC 14:0	460604	Phospholipid		Kim et al., 2014
LysoPC 18:2	11005824	Phospholipid		
LysoPC 16:0	460602	Phospholipid		
LysoPC 18:1	53480465	Phospholipid		
Histidine	6274	Essential amino acid		Kim et al., 2014;
Threonine	6288	Essential amino acid		Thomas et al., 2015
Valine	6287	Essential amino acid		
Methionine	6137	Essential amino acid		
Lysine	5962	Essential amino acid		
Isoleucine	6306	Essential amino acid		
Leucine	6106	Essential amino acid		
Phenylalanine	6140	Essential amino acid	✓	
L-Aspartate	5460294	Non-essential amino acid		Kim et al., 2014;
Serine	5951	Non-essential amino acid		Valarmathi et al., 2014;
Glutamine	5961	Non-essential amino acid		Thomas et al., 2015

(Continued)

TABLE 1 | Continued

Compound	PubChem CID	Compound class	Figure 2[a]	References
Glycine	750	Non-essential amino acid		
Arginine	6322	Non-essential amino acid		
Alanine	602	Non-essential amino acid		
Proline	614	Non-essential amino acid		
Tyrosine	6057	Non-essential amino acid		
α-Aminobutyric acid (AABA)	6657	Non-essential amino acid		
Potassium (K)	5462222	Mineral		Valarmathi et al., 2014;
Calcium (Ca)	5460341	Mineral		Thomas et al., 2015;
Magnesium (Mg)	5462224	Mineral		Raghuvanshi et al., 2017;
Sodium (Na)	5360545	Mineral		Shozib et al., 2017;
Chromium (Cr)	23976	Mineral		Hurtada et al., 2018
Iron (Fe)	23925	Mineral		
Manganese (Mn)	23930	Mineral		
Zinc (Zn)	23994	Mineral		
Copper (Cu)	23978	Mineral		
Phosphorus (P)	5462309	Mineral		
Caproic acid	8892	Fatty acid		Valarmathi et al., 2014;
Caprylic acid	379	Fatty acid		Thomas et al., 2015
Capric acid	2969	Fatty acid		
Lauric acid	3893	Fatty acid		
Tridecanoic acid	12530	Fatty acid		
Myristic acid	11005	Fatty acid		
Pentadecanoic acid	13849	Fatty acid		
Palmitic acid	985	Fatty acid		
Stearic acid	5281	Fatty acid		
Arachidic acid	10467	Fatty acid		
9-Octadecanoic acid	965	Fatty acid		
Undecanoic acid	8180	Fatty acid		
Oleanolic acid	10494			
Myristoleic acid	5281119	Mono-unsaturated fatty acid		Valarmathi et al., 2014;
cis-10-Pentadecenoic acid	5312411	Mono-unsaturated fatty acid		Thomas et al., 2015
Oleic acid	445639	Mono-unsaturated fatty acid		
cis-Vaccenic acid	5282761	Mono-unsaturated fatty acid		
Erucic acid	5281116	Mono-unsaturated fatty acid		
Hexadecadienoic acid		Polyunsaturated fatty acid		Valarmathi et al., 2014;
Hexadecatrienoic acid	6506600	Polyunsaturated fatty acid		Thomas et al., 2015
Linoleic acid	5280450	Polyunsaturated fatty acid		
Octadecatetraenoic acid	11778225	Polyunsaturated fatty acid		
cis-11,14,17-Eicosatrienoic acid	5312529	Polyunsaturated fatty acid		
cis-5,8,11,14-Eicosatetraenoate acid		Polyunsaturated fatty acid		
Eicosatetraenoic acid	21863049	Polyunsaturated fatty acid		
Pinellic acid	9858729	Oxylipin		Kim et al., 2014
Succinic acid	1110	Carboxylic acid		
Maleic acid	444266	Carboxylic acid		
Malonic acid	867	Carboxylic acid		
Citric acid	311	Carboxylic acid		
Cinnamic acid	444539	Carboxylic acid	✓	
D-Xylose	135191	Sugar		Kim et al., 2014;
D-Fructose	2723872	Sugar		Valarmathi et al., 2014
D-Glucose	5793	Sugar		
Maltose	439341	Sugar		
myo-Inositol	892	Sugar		

[a]*Compounds present in* ***Figure 2****.*

suggested that pigmented rice contains higher levels of zinc, iron, and manganese than does white grain, but a lower level of phosphorus (Raghuvanshi et al., 2017; Hurtada et al., 2018; Shao et al., 2018). Brown rice can provide as much as 75% of the recommended daily intake of zinc, copper, and iron, but this falls to just 37% for white rice (Hashmi and Tianlin, 2016).

Flavonoid Metabolism in Pigmented Rice Grain

The major flavonoids present in pigmented rice grain are proanthocyanidins and anthocyanins (**Table 1**). The synthesis of the flavonoids is initiated by the deamination of phenylalanine to form cinnamic acid, a reaction catalyzed by phenylalanine ammonia lyase. Cinnamate 4-hydroxylase catalyses the oxidation of cinnamic acid to 4-coumaric acid, which is in turn converted to 4-coumaroyl-CoA through the action of 4-coumaroyl-CoA ligase (Cheng et al., 2014). The rate limiting step is the conversion of cinnamic acid to *p*-coumaroyl-CoA, which affects the synthesis of phenolic acids, flavanones, proanthocyanidins, and anthocyanidins (**Figure 2**).

Phenolic Acids

Compared to white grain, pigmented grain contains a higher level of phenolic acids (Gunaratne et al., 2013; Zhang et al., 2015). Cinnamic acid serves as a precursor for the synthesis of various phenolic acids, including *p*-coumaric acid, ferulic acid, sinapic acid, isoferulic acid, and 2,5-dihydroxybenzoic acid (Zhang et al., 2015; Shao et al., 2018). The predominant phenolic acids present in white rice are *p*-coumaric acid and ferulic acid; these are largely utilized as building blocks for lignin synthesis (**Figure 2**). In an alternative pathway, particularly active in black rice, cinnamic acid is converted to vanillic acid and protocatechuic acid (Zhang et al., 2015; Shao et al., 2018). In red rice, caffeic acid has been identified as a minor phenolic acid, while this compound is not detectable in brown rice (Gunaratne et al., 2013; Zhang et al., 2015; Irakli et al., 2016; Shao et al., 2018). Additional phenolic acids identified include syringic acid in the extract of brown, red, and black rice (Ghasemzadeh et al., 2018; Shao et al., 2018), pinellic acids in red and white rice (Kim et al., 2014), hydroxybenzoic acid in black rice extracts (Tantipaiboonwong et al., 2017), and gallic acid in the extracts of the red rice mutant AM-425 (Chiou et al., 2018). Four diferulic acids (phenolic acid dehydrodimers) are present in the insoluble-bound (**Table 1**; Zhang et al., 2015).

Flavanones

The condensation and subsequent intramolecular cyclization of three molecules of malonyl CoA and one of 4-coumaroyl-CoA is then catalyzed by chalcone synthetase to produce naringenin chalcone. Naringenin chalcone is isomerized into naringenin by the action of chalcone isomerase to form the flavones (**Figure 2**). Small quantities of flavones and flavanol glycosides have been detected in the grain, notably luteolin-6/8-*C*-pentoside-6/8-*C*-hexoside and certain derivatives of apigenin (**Table 1**). In the tricin pathway, a flavone synthase II enzyme converts naringenin to apigenin, which is then converted first to luteonin by flavonoid 3′-hydroxylase, and then to tricin by *O*-methyltransferase and chrysoeriol 5′-hydroxylase (Park et al., 2016; **Figure 2**). Apigenin, luteolin, tricetin, tricin, quercetin, and myricetin have all been detected in extracts of red and brown rice bran (**Table 1**; Galland et al., 2014; Ghasemzadeh et al., 2018). The synthesis of C-glycosylated flavanones begins with the conversion of naringenin to 2-hydroxyflavanone by flavanone 2-hydroxylase, which is then C-glycosylated by C-glucosyltransferase and finally is dehydrated by an as yet unknown enzyme (**Figure 2**; Du et al., 2010; Galland et al., 2014; Sasaki et al., 2014; Park et al., 2016; Poulev et al., 2019). Other flavonoid-like compounds identified in rice include quercetin-3-*O*-glucoside, quercetin-3-*O*-rutinoside, methoxy-flavanol-3-*O*-glucoside, and isorhamnetin-3-*O*-glucoside (Pereira-Caro et al., 2013; Kim et al., 2014); tricin-*O*-rhamnoside-*O*-hexoside and apigenin-6-*C*-glucosyl-8-*C*-arabinoside are particularly predominant in white rice grains (Kim et al., 2014).

Proanthocyanidins

Proanthocyanidins are oligomers and polymers of flavan-3-ols (Gunaratne et al., 2013). Naringenin, the universal substrate for their synthesis, is 3′-hydroxylated by flavonoid 3′-hydroxylase, producing eriodictyol, which is then converted to dihydroquercetin by the action of flavone 3-hydroxylase (**Figure 2**). Dihydroflavonol 4-reductase catalyses the conversion of dihydroquercetin into leucoanthocyanidins. Leucocyanidin is converted into the flavan-3-ol catechin by leucoanthocyanidin reductase, while catechin monomers are polymerized by a yet unknown pathway to form proanthocyanidin (**Figure 2**; Zhao et al., 2010; Galland et al., 2014). Proanthocyanidins and catechins make up the bulk of the phenolic compounds found in red rice, being responsible for the red pigmentation of the pericarp (Pereira-Caro et al., 2013; Kim et al., 2014). No proanthocyanidins have been detected in white rice accessions (Gunaratne et al., 2013), while some black rice varieties have been reported to contain them (Vichit and Saewan, 2015).

Anthocyanidins

The oxidization of leucoanthocyanidin to form cyanidin, pelargonidin, and delphinidin is catalyzed by anthocyanin synthase (**Figure 2**; Cheng et al., 2014; Galland et al., 2014). Anthocyanins, which are responsible for purple to blue pigmentation, represent the bulk of the flavonoids present in black and purple rice (Pereira-Caro et al., 2013; Zhang et al., 2015). The compounds cyanidin-3-*O*-glucoside and peonidin-3-*O*-glucoside are the most prominent, but also represented are cyanidin-3,5-diglucoside, cyanidin-3-*O*-(6″-*O*-*p*-coumaroyl)glucoside, pelargonidin-3-*O*-glucoside, peonidin-3-*O*-(6″-O-*p*-coumaroyl)glucoside, and cyanidin-3-*O*-arabidoside. Red and white rice grains have been classified as lacking anthocyanin (Gunaratne et al., 2013; Xiongsiyee et al., 2018), but both Boue et al. (2016) and Ghasemzadeh et al. (2018) have been able to detect a low level in both red and brown rice accessions. Unstable anthocyanidins can be converted into the colorless flavan-3-ols epiafzelechin, epicatechin, and epigallocatechin through the action of anthocyanin reductase, and when glycosylated, a wide array of distinct molecules are generated (Ko et al., 2006; Sasaki et al., 2014; Kim et al., 2015;

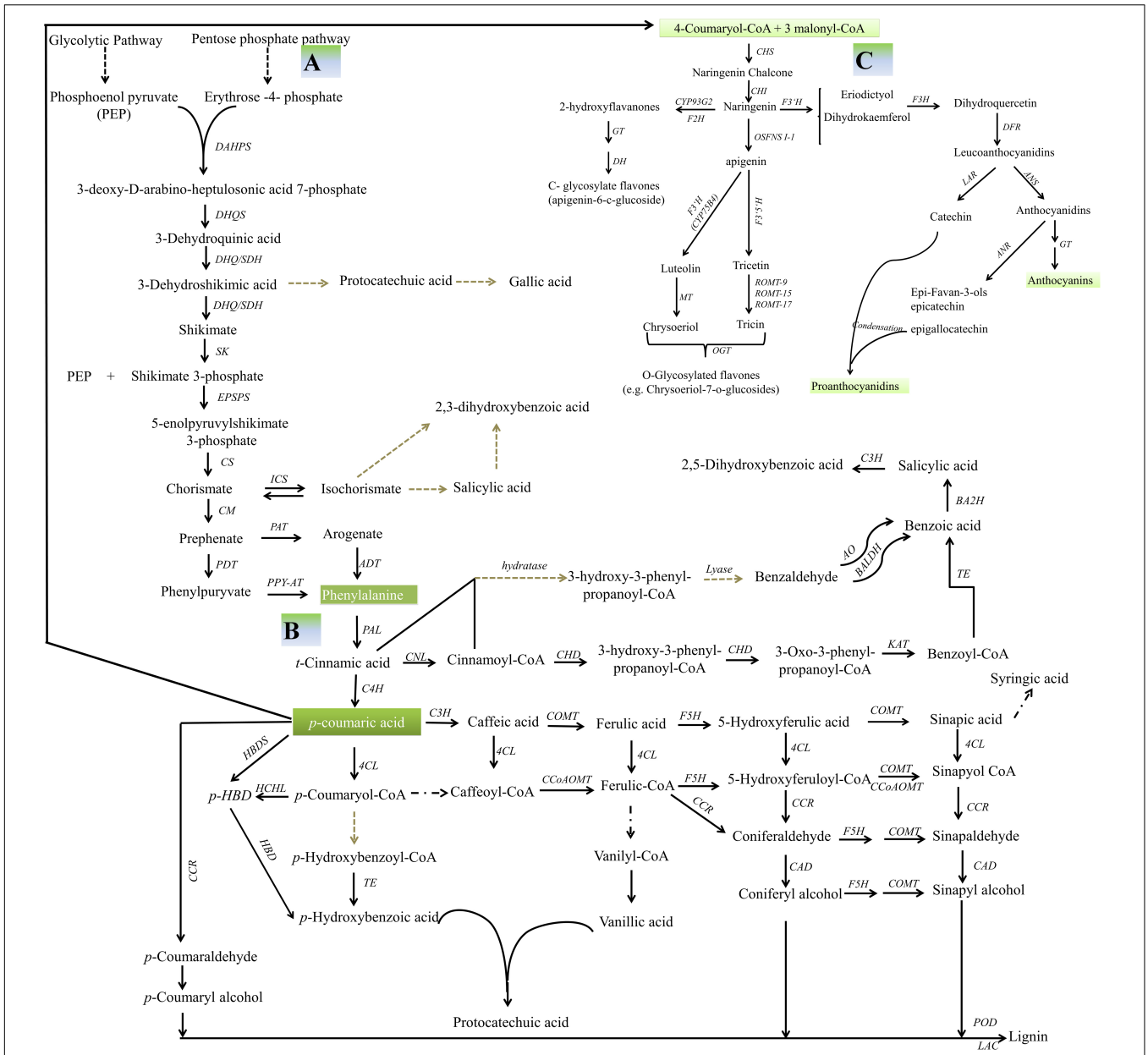

FIGURE 2 | Secondary metabolism in rice. **(A)** A schematic representation of the shikimic acid pathway. DAHP, 3-deoxy-D-arabino-heptulosonic acid 7-phosphate; DAHPS, 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHQ/SDH, 3-dehydroquinate dehydratase/shikimate 5 dehydrogenase; DHQS, 3-dehydroquinate synthetase; DHS, 3-dehydroshikimic acid; SDH, shikimate dehydrogenase; SK, shikimate kinase; S3P, shikimic acid 3-phosphate; EPSPS, 5-enolpyruvylshikimate 3-phosphate synthase; EPSP, 5-enolpyruvylshikimate 3-phosphate; CS, chorismate synthase; CM, chorismate mutase; PAT, prephenate aminotransferase; ADT, arogenate dehydratase; PDT, prephenate dehydratase; PPY-AT, phenylpyruvate aminotransferase (Tzin and Galili, 2010; Widhalm and Dudareva, 2015; Santos-Sánchez et al., 2019). **(B)** Possible routes to the production of benzoic acid, benzoic acid-derived compounds and lignin. CNL, cinnamate-CoA ligase; CHD, cinnamoyl-CoA-dehydrogenase/hydratase; KAT1, 3-ketoacyl-CoA thiolase; TE, CoA thioesterase; BA2H, benzoic acid 2-hydroxylase; BALDH, benzaldehyde dehydrogenase; AO, aldehyde oxidase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; ICS, isochorismate synthase; CCR, cinnamoyl-CoA reductase; CCoAOMT, caffeoyl-CoA *O*-methyltransferase; F5H, ferulate 5-hydroxylase; CSE, caffeoyl shikimate esterase; COMT, caffeic acid *O*-methyltransferase; CAD, cinnamyl alcohol dehydrogenase; LAC, laccase; POD, peroxidase; *p*-HBD, *p*-hydroxybenzaldehyde; HBDS, 4-hydroxybenzaldehyde synthase; HCHL, 4-hydroxycinnamoyl-CoA hydratase/lyase; HBD, 4-hydroxybenzaldehyde dehydrogenase (Qualley et al., 2012; Gallage and Møller, 2015; Widhalm and Dudareva, 2015; Liu et al., 2018). **(C)** Flavonoid metabolism. PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; CHS, chalcone synthetase; CHI, chalcone isomerase; F3′H, flavone 3-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanin synthase; ANR, anthocyanin reductase; GT, glucosyltransferase; LAR, leucoanthocyanidin reductase; MT, *O*-methyltransferase; F2H, flavanone 2-hydroxylase (Chen et al., 2013; Galland et al., 2014). The square dot arrows indicates steps which have not yet been fully elucidated, while the black arrows indicate steps supported by genetic evidence.

Figure 2). Although the major enzymes operating in the flavonoid pathway are well known and their encoding genes have been identified (**Table 2**), many aspects underlying the synthesis of these pigments in rice have yet to be fully elucidated.

THE GENETIC BASIS OF RICE GRAIN PIGMENTATION

The rice genome harbors at least two genes encoding chalcone synthetase: *CHS1* on chromosome 11 and *CHS2* on chromosome 7 (Shih et al., 2008; Han et al., 2009; Cheng et al., 2014), each contributing to flavanone biosynthesis. For the production of proanthocyanidins, three flavone 3-hydroxylase are relevant: namely *F3H-1* (chromosome 4), *F3H-2* (chromosome 10), and *F3H-3* (chromosome 4; Kim et al., 2008; Park et al., 2016). Two anthocyanin synthases are critical for the synthesis of anthocyanins, namely *ANS1* (chromosome 1) and *ANS2* (chromosome 6) (Shih et al., 2008; **Table 2**).

Rc Role in Red Pericarp in Ancestral Rice

White grained rice was selected during rice's domestication. The two complementary genes *Rc* (on chromosome 7), which encodes a basic helix-loop-helix (bHLH) transcription factor, and *Rd* (chromosome 1) encoding a form of dihydroflavonol 4-reductase, an enzyme which enhances the accumulation of proanthocyanidin, are together responsible for the red pericarp color. *Rc* is closely associated with shattering and grain dormancy (Sweeney et al., 2006), so therefore was selected against during domestication. *Rc-Rd* genotypes produce red grain, while *Rc-rd* genotypes produce brown grain (Furukawa et al., 2006). The three common *Rc* alleles are the wild type *Rc*, and mutant alleles *Rc-s* and *rc*. *Rc-s* differs from *Rc* due to the presence of a premature stop codon, while *rc* lacks a 14 bp stretch of the wild type sequence (Furukawa et al., 2006; Sweeney et al., 2006). Carriers of *rc* produce a colorless pericarp, while those of *Rc-s* produce a range of pericarp pigmentation (Sweeney et al., 2007). A number of variants have been identified as restoring the wild type (red) pericarp pigmentation: *Rc-g* carries a 1 bp deletion 20 bp upstream of the 14 bp *rc* deletion (Brooks et al., 2008), while Rc^r features a 44 bp deletion upstream of the 14 bp segment, which restores the wild type reading frame (Ferrari et al., 2015). Most varieties of African domesticated rice (*Oryza glaberrima*) produce a red pericarp, and white variants harbor a loss-of-function mutation in *Rc*. An exception is the *O. glaberrima* specific mutation *rc-gl*, which carries a premature stop codon 146 bp upstream of the site of the *Rc-s* point mutation (Gross et al., 2010).

Regulatory Cascade Influencing Purple Rice Color

Anthocyanins are responsible for the black-purple pigmentation in rice grain. The variation seen in pigmentation intensity has been taken to imply that the trait is under polygenic control, involving as yet unidentified genes (Ham et al., 2015). A number of publications report the identification of rice genes that regulate anthocyanin production, each adopting a different gene coding system, which only adds to the confusion. According to Hu et al. (1996), two classes of regulatory gene (*R/B* and *C1/Pl*) govern both the accumulation of anthocyanin and the regulation of its deposition. Two *R* genes have been characterized: *Ra* maps to chromosome 4 and *Rb* to chromosome 1. The former gene is thought to be a homolog of the maize *R/B* gene. Three alleles of *Pl* (chromosome 4) have been identified, namely Pl^w, Pl^i, and Pl^j, and each is responsible for a distinctive pattern of pigmentation. Pl^w activates anthocyanin synthesis in most of the aerial parts of the rice plant (although not in either the stem or the internode). The *Pl* locus harbors the two genes, *OSB1* and *OSB2*, each of which encodes a bHLH transcription factor (**Table 2**; Sakamoto et al., 2001). Other studies found the purple pericarp trait to be genetically determined by the dominant complementary genes *Pb* (synonym *Prp-b*) and *Pp* (synonym *Prp-a*), mapping to chromosomes 4 and 1, respectively (**Table 2**; Rahman et al., 2013; Ham et al., 2015). While the product of *Pb* appears to be responsible for the accumulation of pigment in the pericarp of brown grain, that of *Pp* increases the amount of the pigment, giving rise to purple grain. The number of copies of the *Pp* gene present is correlated with the intensity of the purple pigmentation (Rahman et al., 2013). In the absence of *Pp*, plants harboring *Pb* produce grain with a brown pericarp, while the pericarp of *Pp* carriers, lacking *Pb*, are white (Rahman et al., 2013). The *Pb* locus comprises of two genes, a myc transcription factor (*Ra*), along with *bHLH16*. The *bHLH16* has been shown to be involved in proanthocyanidin synthesis, while *Ra* is involved in anthocyanin synthesis. *Ra* and *OSB1* are believed to have synonymous functions (Hu et al., 1996; Sakamoto et al., 2001; Caixia and Qingyao, 2007). A 2 bp (GT) insertion in exon 7 of *Ra* abolishes purple pigmentation (Caixia and Qingyao, 2007; Lim and Ha, 2013; Rahman et al., 2013). Similarly, Sakulsingharoj et al. (2016) have found that a 2 bp (GT) insertion in exon 7 of *OSB1*, which along with a 1 bp deletion of a guanine nucleotide in exon 8, results in a threonine for methionine substitution at position 64, resulting in a white grain phenotype. Carriers of the three loci *Kala1* (chromosome 1), *Kala3* (chromosome 3), and *Kala4* (chromosome 4) express a black pericarp trait (Maeda et al., 2014). It has been suggested that *Kala4* is synonymous with *Pb*, and *Kala1* with *Pp*. *Kala4* encodes a bHLH transcription factor and corresponds to *OSB2* (**Table 2**). OSB2 regulates a number of genes encoding enzymes involved in anthocyanin synthesis, including *F3H*, *DFR*, and *ANS* (Sakulsingharoj et al., 2014). The chromosomal region harboring *Kala1* includes *Rd* (dihydroflavonol 4-reductase). *Kala3* is likely be a synonym of *MYB3* (Maeda et al., 2014). The black grain phenotype occurring in tropical *japonica* germplasm has been attributed to structural variants in the *Kala4* promoter sequence. Oikawa et al. (2015) have proposed that *Kala4* has been introgressed several times from *japonica* to *indica* germplasm.

The R2R3-MYB transcription factor *Os06g0205100* has been proposed as a candidate for the *C* gene, functioning as a possible activator of *DFR* and *ANS* (Saitoh et al., 2004; Rachasima et al., 2017; Sun et al., 2018). *Os01g0633500* (*A1*) is a dihydroflavonol

TABLE 2 | Regulatory and structural genes shown to be involved in the biosynthesis of anthocyanin and proanthocyanidin in rice.

Locus	Allelic locus	Gene name	Locus ID		CHRX[a]	Pericarp	References
			Rice Annotation Project (RAP)	Rice Genome Annotation Project (MSU)			
Regulatory genes							
Kala1	*Rd*	OsDFR	Os01g0633500	LOC_Os01g44260	1	Red and black	Furukawa et al., 2006; Shih et al., 2008; Maeda et al., 2014; Sun et al., 2018
	Pp				1	Red and black	Caixia and Qingyao, 2007
Kala3		OsMYB3	Os03t0410000	LOC_Os03g29614	3	Black	Maeda et al., 2014
Kala4	*Pl*^{w}	*OSB1/Pb/Ra*	Os04g0557800	LOC_Os04g47080	4		Hu et al., 1996; Sakamoto et al., 2001; Caixia and Qingyao, 2007; Sakulsingharoj et al., 2016
		OSB2	Os04g0557500	LOC_Os04g47059	4		Sakamoto et al., 2001; Sakulsingharoj et al., 2014; Oikawa et al., 2015
Rc	*Rc*-s	bHLH	Os07g0211500	LOC_Os07g11020	7	Red	Sweeney et al., 2006
	Rc				7	Red	Furukawa et al., 2006
	rc				7	White	Furukawa et al., 2006; Sweeney et al., 2006
	Rc-g				7	Red	Brooks et al., 2008
	Rc^{r}				7	Red	Ferrari et al., 2015
	Rc-gl				7	White	Gross et al., 2010
Chromogen		*OsC1*	Os06g0205100	LOC_Os06g10350	6	Black	Saitoh et al., 2004; Rachasima et al., 2017; Sun et al., 2018
Structural genes							
Chalcone synthetase (CHS)		*OsCHS1*	Os11g0530600	LOC_Os11g32650	11	Common intermediate	Shih et al., 2008
		OsCHS2	Os07g0214900	LOC_Os07g11440	7	Common intermediate	
Chalcone isomerase (CHI)		*OsCHI*	Os03g0819600	LOC_Os03g60509	3	Common intermediate	Shih et al., 2008
Flavanone 3-hydroxylase (F3H)		*OsF3H-1*	Os04g0662600	LOC_Os04g56700	4	Common intermediate	Kim et al., 2008
		OsF3H-2	Os10g0536400	LOC_Os10g39140	10		
		OsF3H-3	Os04g0667200	LOC_Os04g57160	4		
Flavanone 3′-hydroxylase (F3′H)		*OsF3′H*	Os10g0320100	LOC_Os10g17260	10	Common intermediate	Shih et al., 2008
Leucoanthocyanidin reductase (LAR)		*OsLAR*	Os03g0259400	LOC_Os03g15360	3	Black rice	Kim et al., 2015
Anthocyanidin synthase (ANS)		*OsANS1*	Os01g0372500	LOC_Os01g27490	1	Black rice	Shih et al., 2008
		OsANS2	Os06g0626700	LOC_Os06g42130	6		
UDP-glycosyltransferase (UF3GT)		OsUGT	Os06g0192100	LOC_Os06g09240	6	Black rice	Yoshimura et al., 2012
			Os07g0148200	LOC_Os07g05420	7	Black rice	
Anthocyanin reductase (ANR)		*OsANR*	Os04g0630800	LOC_Os04g53850	4	Black rice	Kim et al., 2015

[a]*CHRX: chromosome.*

reductase gene involved in anthocyanin synthesis (**Table 2**). Thus, *A1* and *C1* determine the purple color of grain. The *S1* gene (*Os04g0557500*) encodes a bHLH protein, and contributes to hull-specific pigmentation. The presence of a functional copy of both *C1* and *S1* has been shown to be required for hull pigmentation, while the product of *A1* acts as a catalyst for the development of purple hulls (Sun et al., 2018). The pattern of anthocyanin pigmentation is determined by the allelic status of *A1*, *C1*, and *S1* (Sun et al., 2018). Several authors have attempted to correlate sequence variants of a number of regulatory genes, e.g., *C1* and *OSB2*, with phenotypic variation in rice grain pigmentation (Sakulsingharoj et al., 2016; Rachasima et al., 2017; Sun et al., 2018). Lachagari et al. (2019) conducted comparative genomics in 108 rice lines and identified novel allelic variants in a number of genes belonging to the flavonoid pathway, cytokinins glucoside, and betanidin degradation biosynthesis that were associated with purple pigmentation. Although a number of genes responsible for grain pigmentation have already been identified (**Table 2**), there is still a possibility that additional genes and variants thereof, remain to be discovered.

The Genetic Basis of Grain Pigmentation Inferred From Quantitative Trait Loci Analysis or Genome Wide Association Studies

A number of attempts have been made to exploit the quantitative trait loci (QTL) mapping approach as a means of inferring the genetic basis of grain pigmentation (**Table 3**). Tan et al. (2001) identified nine QTL in an analysis of flour pigmentation in a recombinant inbred line (RIL) population. Three QTL reflected variation in the CIE 1976 color parameter L* (lightness), two in a* (red-green), and four in b* (yellow-blue). In a backcross RIL population, made from a cross between the rice varieties "Kasalath" (red pericarp) and "Koshihikari" (white pericarp), Dong et al. (2008) identified four QTL underlying variation in red pigmentation, with the two largest effect QTL co-locating with *Rc* and *Rd*, the two minor effect QTL being novel. An analysis carried out by Matsuda et al. (2012) suggested that flavonoid content was governed by genetic factors which control flavone glycosylation. In a recent study 21 QTL, responsible for variation in the content and composition of anthocyanin and proanthocyanidin, were identified (Xu et al., 2017). While some mapped to locations occupied by already known genes, others mapped to genomic regions not previously identified as harboring genes involved in rice grain pigmentation.

The genome wide association study (GWAS) approach, which has certain advantages over QTL mapping (Korte and Farlow, 2013), has been applied in a few cases to determine the genetic basis of grain pigmentation (**Table 3**). Shao et al. (2011) used GWAS to identify 25 marker–trait associations for grain pigmentation: some related to pigment intensity, others to hue angle, L*, a*, or b*. Their analysis confirmed the importance of *Ra* and *Rc*. Butardo et al. (2017) used GWAS to uncover a number of single nucleotide polymorphism loci (SNPs) linked to *Rc*. The 763 SNPs associated with pericarp pigmentation uncovered by Yang et al. (2018) mapped to 6 of the 12 rice chromosomes (chromosomes 1, 3, 4, 8, 10, and 11); some of the most significantly associated SNPs lying close to previously identified structural or regulatory genes, but others map to regions not previously associated with variation in rice grain pigmentation.

'Omics Approaches Taken to Unraveling the Mechanistic Basis of Grain Pigmentation

High throughput genomics, including transcriptomics, proteomics, and metabolomics, have contributed to the unraveling of biochemical pathways underlying target traits. By combining genetics with systems biology tools the target genomic regions were narrowed down to identify candidate genes and proteins influencing key nutritional traits of interest in rice (Butardo et al., 2017; Anacleto et al., 2019).

Differential transcriptomic analyses between pigmented and non-pigmented rice grains identified regulators and downstream targets of flavonoid pathway genes (Oh et al., 2018). The high anthocyanin content of black rice was associated with enhanced transcription of genes encoding anthocyanidin synthase, while high proanthocyanin content, characteristic of red rice, was accompanied by a notable abundance of transcript for a gene encoding leucoanthocyanidin reductase (Chen et al., 2013). Transcript abundance of genes encoding chalcone synthetase, chalcone isomerase, flavanone 3-hydroxylase, dihydroflavonol 4-reductase, and anthocyanin synthetase was compared in white, black, and red rice grain by Lim and Ha (2013). Four genes were markedly up-regulated in pigmented grain varieties, while the gene encoding chalcone isomerase displayed a similar level of transcription in both white and pigmented varieties. The enhanced abundance of transcripts of chalcone synthetase, flavanone 3-hydroxylase, and anthocyanin synthetase seen in some black varieties implied a strong correlation between transcription and pigment content. Sun et al. (2018) found that flavonoid pathway genes were regulated by ternary MYB-bHLH-WD40 transcriptional complexes (Xu et al., 2013; Zhang et al., 2018). A microarray-based comparison of black and white rice identified nearly 1,300 differentially transcribed genes, of which 137 were predicted to encode transcription factors belonging to 1 of 10 different classes (Kim et al., 2011). When Kim et al. (2018) applied RNA-seq to analyze differential transcription, it was concluded that the B-box protein encoded by *BBX14* was a key regulator of the anthocyanin synthesis pathway. Anthocyanin production in pigmented grain appeared to be induced and fine-tuned by BBX14 in conjunction with the basic leucine zipper transcription factor HY5. Both irradiation at a high light intensity and the plant's sugar content can influence anthocyanin and proanthocyanin synthesis (Xu et al., 2013; Ma et al., 2018; Zhang et al., 2018). Therefore it would be of value to search for linkages between photoreceptor and light signal transduction elements associated with anthocyanin/proanthocyanidin synthesis in pigmented rice (Teng et al., 2005). While transcriptomic

TABLE 3 | Quantitative trait loci identified for colored related traits, anthocyanin and proanthocyanidin.

No.	Population	Size	Markers	Trait category	QTLs/QTNs	Closest structural and/or regulatory genes	Chromosome	Effect (%)	References
1	RILs	238	162 RFLP and 48 SSRs	Flour color	9		1, 3, 4, 5,6, 7, 8	4.3–25.4	Tan et al., 2001
					L* (3)		5,6, 8	4.5–15.7	
					a* (2)		4, 7	6.9–10.5	
					b* (4)		1, 3, 6, 8	4.3–25.4	
2	BRILs	182	162 RFLP	Degree of red coloration	4		1, 7, 9, 11	2.1–83.7	Dong et al., 2008
					*q*DRC-1*	*Rd*	1	3.6–3.7	
					*q*DRC-7*	*Rc*	7	75.9 -83.7	
					*q*DRC-9		9	2.1–3.2	
					*q*DRC-11		11	3.3–3.4	
3	RILs	182	126 SSRs	Anthocyanin and proanthocyanidin	21		1, 2, 3, 7, 8, 10, 12	3.8–34.8	Xu et al., 2017
					ANC (8)		1, 2, 3, 7, 10	8.8–34.8	
					PAC (13)		1, 2, 7, 8, 10, 12	3.8–17.0	
4	Diversity panel	416	100 SSRs and 10 gene markers	Grain color	25		1, 4, 6, 7, 8, 9, 10, 11, 12	1.39–86.68	Shao et al., 2011
					L* (3)	*Ra*, *Rc*	4, 7, 10	4.96–31.23	
					a* (8)	*Ra*, *Rc*	1,4, 7, 8, 9, 11, 12	1.51–19.65	
					b* (6)	*Ra*	4, 6, 8, 9, 10	4.38–49.82	
					c (3)	*Ra*	4, 8, 10	1.39–3.99	
					H° (5)	*Ra*	4, 6, 8, 9	5.4–86.68	
				Phenolic and flavonoid content	10		4, 7, 8, 9, 10	2.64–39.67	
					PC (4)	*Ra*, *Rc*	4, 7, 8, 9	5.87–39.67	
					FC (6)	*Ra*, *Rc*	4, 7, 8, 9, 10	2.64–35.35	
5	Diversity panel	203	sequencing data	Pericarp color (PC)	4		7, 10		Wang et al., 2016
					Rc-s	Rc-s	7		
					qPc10	F3H	10		
6	Diversity panel	244	122,785 SNPs	Red seed color	snp_07_6067391	bHLH			Butardo et al., 2017
7	Diversity panel	419	208,993 SNPs	Pericarp color_whole panel	763		1,3, 4, 7, 8, 10, 11		Yang et al., 2018
						Rd	1		
						MYB family transcription factors	10, 11		

(Continued)

TABLE 3 | Continued

No.	Population	Size	Markers	Trait category	QTLs/QTNs	Closest structural and/or regulatory genes	Chromosome	Effect (%)	References
						WD domain, G-beta repeat domain containing protein	8		
				Pericarp color_*indica*	99				
						MYB family transcription factors	10, 11		
						OsCHI	3		
						Kala4	4		
						Rc	7		
						WD domain, G-beta repeat domain containing protein	8		
						OsCHI	3		
						Kala4	4		
						Rc	7		

RILs, recombination inbred lines; BRILs, backcross-recombinant inbred lines.

analyses have succeeded in shedding some light on the transcriptional regulation of secondary metabolites, unraveling post-transcriptional and post-translation processes may well provide further insights into the identity of the rate-limiting steps of grain pigmentation (Merchante et al., 2017; Spoel, 2018).

The abundance of a given transcript and its translation product are not always linearly related, due to post-transcriptional regulation, translation and post-translational processing, and peptide modification (Chen et al., 2017; Zhang et al., 2017). Thus, a proteomic analysis can give a more nuanced picture of the differences between pigmented and non-pigmented grains, than is possible from a transcriptomic analysis. A total of 230 differentially abundant proteins, involved in various metabolic processes, were identified by Chen et al. (2016) from a comparison between black and white grains sampled at five stages of grain development. A number of proteins involved in the synthesis of flavonoids and sugars were found to be more abundant in the black grain, while proteins associated with signal transduction, redox homeostasis, photosynthesis, nitrogen metabolism, and tocopherol synthesis were less abundant. In particular, chalcone synthetase was pinpointed as a key component required for the synthesis of anthocyanin.

Metabolomic analyses have also been successfully used to characterize the cellular composition of rice (**Table 1**). Comparative metabolome studies of black, red, and non-colored rice revealed that various anthocyanins, tocopherol, fatty acid methyl esters, free sugars, and fatty acids were found to be significantly different (Frank et al., 2012). de Guzman et al. (2017) were able to monitor more than 1,000 metabolites in a screen of several rice varieties differing with respect to their nutritional quality and glycemic response (de Guzman et al., 2017). Comparisons of the grain metabolomes of diverse rice accessions have revealed that a substantial degree of variation exists at this level (Gong et al., 2013; Pereira-Caro et al., 2013). However, the grain metabolome is highly dynamic, responding strongly to the plant's external environment, so this variation is as a consequence of genetic and environmental variation and the interaction there-of. Correlation analyses carried out between individual metabolites have nevertheless revealed the regulation of the grain metabolome, with clusters of co-accumulated metabolites appearing to be under the control of shared genetic factors (Matsuda et al., 2012, 2015).

The application of 'omics-based platforms have begun to reveal the genetic and biochemical basis of the grain pigment parameters a*, b*, L*, hue, and chroma. While transcriptomic and metabolomic analyses have identified certain important structural and regulatory genes influencing core components of flavonoid synthesis (Lee et al., 2015; Oh et al., 2018), what is still lacking is a comprehensive understanding of the molecular machinery underlying key metabolic processes such as the polymerization and transport of tannins. A more integrated approach, focusing on identifying linkages between regulatory networks is needed (**Figure 2**). Combining diverse datasets facilitates the reconstruction of regulatory networks and the

identification of key modulators (Butardo et al., 2017; Wambugu et al., 2018). As an example, an exploration of the key genetic influences affecting grain amylose–amylopectin composition has implicated two genomic regions, one on chromosome 6 and the other on chromosome 7 (Butardo et al., 2017). The genetic region on chromosome 7 is in the vicinity of *Rc* and includes a haplotype associated with increased amylose and reduced accumulation of short chain amylopectin. The bHLH transcription factor encoded by *Rc* activates the gene encoding dihydroflavonol 4-reductase, thereby influencing the formation of red pigmentation. The same transcription factor has also been proposed to act as a regulator of starch structure, as it engages within the network regulating granule-bound starch synthase activity.

Conserving and Utilizing Pigmented Rice Landraces for Future Breeding

Although rice landraces with pigmented grain represent an important genetic reservoir for rice improvement, these populations are rapidly being lost as a result of the introduction of more productive, modern white rice varieties. Some of these materials have been safeguarded in *ex situ* gene banks, such as the major gene bank curated by the International Rice Research Institute[2]. Few of these landraces have been systematically characterized in terms of their grain end-use quality, their nutritional features and potential health benefits. Therefore, there is an urgent need to validate the traditional knowledge associated with these materials with scientific-based analyses. While the productivity of the landrace materials is undoubtedly lower than that of modern white rice varieties, their market value is potentially quite high, given the growing consumer preference for nutritious foods (Islam et al., 2018). Looking forward, there is a major opportunity for breeding programs to develop productive pigmented varieties (Voss-Fels et al., 2019).

Understanding the mode of inheritance of grain pigmentation, identifying beneficial alleles of the key genes underlying these traits, and developing trait-specific markers, will contribute to accelerating efforts to breed high yielding pigmented rice varieties. Advanced generation breeding lines of pigmented lines have been developed (Bhuiyan et al., 2011; Arbelaez et al., 2015). A black rice line has been developed in the genetic background of a leading Japanese white rice variety (Koshihikari); which has eating quality superior to that of the widely cultivated black rice variety "Okunomurasaki" (Maeda et al., 2014). Crosses have been initiated between pigmented and non-pigmented varieties to develop pigmented varieties adapted to the growing conditions in Kazakhstan (Rysbekova et al., 2017). The Thai aromatic, deep purple *indica*-type rice variety "Riceberry" has developed a reputation for its health-promoting properties. Riceberry combines the desirable features of two prominent rice varieties, one a local, non-glutinous purple rice and the other an aromatic white jasmine rice (Waiyawuththanapoom et al., 2015; Gene Discovery Rice and Rice Science Center, 2017). Two improved pigmented varieties (the red rice "Rubi" and the black rice "Onix") have been released in Brazil (Wickert et al., 2014).

CONCLUSION

Pigmented rice varieties are gaining popularity among consumers, and demand is only expected to rise. The seed supply chain of pigmented rice is weak and thus rice value chain opportunities have to evolve to meet the current nutritional demand. Production of pigmented rice using landraces is unable to meet market demand, emphasizing the need to genetically improve these landrace materials. Systematic nutritional characterization of the 130,657 accessions curated by International Rice Research Institute's gene bank and Africa Rice[3], which including pigmented entries, will create new avenues for nutritional diversification that reaches lower income target countries. These, as well as other, national *ex situ* collections, represent a valuable source of genetic variation for the improvement of pigmented rice, providing materials to elucidate the genetic basis of grain pigmentation and associated nutrition-related traits. The process of identifying as yet unknown genes influencing flavonoid metabolism and grain pigmentation could be accelerated by whole genome re-sequencing, allowing novel allelic variants to be harnessed for use as markers. Fine mapped genetic regions associated with proanthocyanidins and anthocyanin needs to be undertaken to develop quality markers to support marker-assisted-selection breeding of these nutritional traits into high yielding rice backgrounds. A systems approach to study implication of diet based health benefits would require holistic understanding of the molecular basis of human health benefits of consuming grain pigmentation, enabling the identification of the modulators involved to overcome the prevailing double burden malnutrition and communicable diseases in the target communities. While several health benefits were shown to possess to consume pigmented rice, its texture and palatability is found to be poor and thus its acceptance rate is lower. To address this limitation, we need to explore the genetic variation for the retention of flavonoids in the milled endosperm.

AUTHOR CONTRIBUTIONS

EM and NS drafted the manuscript. TK, HJ, NE, CB, and LB edited the part of the sections.

FUNDING

The authors thank the UK Biotechnology and Biological Sciences Research Council's Newton Fund Sustainable Rice Research Initiative (Project BB/N013603/1) for its financial support.

[2]www.irri.org/international-rice-genebank

[3]www.genebanks.org/genebanks/africarice/

REFERENCES

Ahuja, U., Ahuja, S., Chaudhary, N., and Thakrar, R. (2007). Red rices - past, present, and future. *Asian Agri Hist.* 11, 291–304.

Anacleto, R., Badoni, S., Parween, S., Butardo, V. M., Misra, G., Cuevas, R. P., et al. (2019). Integrating a genome wide association study with a large scale transcriptome analysis to predict genetic regions influencing the glycaemic index and texture in rice. *Plant Biotechnol. J.* 17, 1261–1275. doi: 10.1111/pbi.13051

Arbelaez, J. D., Moreno, L. T., Singh, N., Tung, C. W., Maron, L. G., Ospina, Y., et al. (2015). Development and GBS-genotyping of introgression lines (ILs) using two wild species of rice. *O. meridionalis and O. rufipogon, in a common recurrent parent, O. sativa cv. Curinga. Mol. Breed.* 35, 81. doi: 10.1007/s11032-015-0276-7

Ashraf, H., Murtaza, I., Nazir, N., Wani, A. B., Naqash, S., and Husaini, A. M. (2017). Nutritional profiling of pigmented and scented rice genotypes of Kashmir Himalayas. *J. Pharmacogn. Phytochem.* 6, 910–916.

Baek, J. A., Chung, N. J., Choi, K. C., Hwang, J. M., and Lee, J. C. (2015). Hull extracts from pigmented rice exert antioxidant effects associated with total flavonoid contents and induce apoptosis in human cancer cells. *Food Sci. Biotechnol.* 24, 241–247. doi: 10.1007/s10068-015-0032-0

Banjerdpongchai, R., Wudtiwai, B., and Sringarm, K. (2013). Cytotoxic and apoptotic-inducing effects of purple rice extracts and chemotherapeutic drugs on human cancer cell lines. *Asian Pacific J. Cancer Prev.* 14, 6541–6548. doi: 10.7314/apjcp.2013.14.11.6541

Berni, R., Id, C. C., Romi, M., Hausman, J., Guerriero, G., and Cai, G. (2018). Agrobiotechnology goes wild?: ancient local varieties as sources of bioactives. *Int. J. Mol. Sci.* 19:E2248. doi: 10.3390/ijms19082248

Bhat, F., and Riar, C. (2015). Health benefits of traditional rice varieties of temperate regions. *Med. Aromat. Plants* 4, 3–5. doi: 10.4172/2167-0412.1000198

Bhuiyan, M. A. R., Narimah, M. K., Rahim, H. A., Abdullah, M. Z., and Wickneswari, R. (2011). Transgressive variants for red pericarp grain with high yield potential derived from *Oryza rufipogon*×*Oryza sativa*: field evaluation, screening for blast disease, QTL validation and background marker analysis for agronomic traits. *Field Crops Res.* 121, 232–239. doi: 10.1016/j.fcr.2010.12.012

Boue, S. M., Daigle, K. W., Chen, M.-H., Cao, H., and Heiman, M. L. (2016). Antidiabetic potential of purple and red rice (*Oryza sativa* L.) bran extracts. *J. Agric. Food Chem.* 64, 5345–5353. doi: 10.1021/acs.jafc.6b01909

Breseghello, F. (2013). Traditional and modern plant breeding methods with examples in rice (*Oryza sativa* L.). *J. Agric. Food Chem.* 61, 8277–8286. doi: 10.1021/jf305531j

Brooks, S. A., Yan, W., Jackson, A. K., and Deren, C. W. (2008). A natural mutation in *rc* reverts white-rice-pericarp to red and results in a new, dominant, wild-type allele: *Rc-g. Theor. Appl. Genet.* 117, 575–580. doi: 10.1007/s00122-008-0801-8

Butardo, V. M. Jr., Anacleto, R., Parween, S., Samson, I., Guzman, K., De, et al. (2017). Systems genetics identifies a novel regulatory domain of amylose synthesis. *Plant Physiol.* 173, 887–906. doi: 10.1104/pp.16.01248

Caixia, W., and Qingyao, S. (2007). Fine mapping and candidate gene analysis of purple pericarp gene *Pb* in rice (*Oryza sativa* L.). *Chinese Sci. Bull.* 52, 3097–3104. doi: 10.1007/s11434-007-0472-x

Chakuton, K., Puangpronpitag, D., and Nakornriab, M. (2012). Phytochemichal content and antioxidant activity of colored and non-colored Thai rice cultivars. *Asian J. Plant Sci.* 6, 285–293. doi: 10.3923/ajps.2012.285.293

Chandramouli, B., Latha, M. M., Narendra, K., and Mallikarjuna, K. (2018). Phytochemical and antimicrobial investigations of methanolic extract & ethyl acetate extract of rice husk (Oryza sativa) mentioned in an ancient palm leaf manuscript (Talapatra). *World J. Pharm. Res.* 7, 598–616. doi: 10.20959/wjpr20183-10797

Chen, L., Huang, Y., Xu, M., Cheng, Z., Zhang, D., and Zheng, J. (2016). iTRAQ-based quantitative proteomics analysis of black rice grain development reveals metabolic pathways associated with anthocyanin biosynthesis. *PLoS One* 11:e0159238. doi: 10.1371/journal.pone.0159238

Chen, L., Huang, Y., Xu, M., Cheng, Z., and Zheng, J. (2017). Proteomic analysis reveals coordinated regulation of anthocyanin biosynthesis through signal transduction and sugar metabolism in black rice leaf. *Int. J. Mol. Sci.* 18:E2722. doi: 10.3390/ijms18122722

Chen, X., Itani, T., Wu, X., Chikawa, Y., and Irifune, K. (2013). Physiological factors affecting transcription of genes involved in the flavonoid biosynthetic pathway in different rice varieties. *Plant Signal. Behav.* 8:e27555. doi: 10.4161/psb.27555

Cheng, A., Han, X., Wu, Y., and Lou, H. (2014). The Function and catalysis of 2-oxoglutarate-dependent oxygenases involved in plant flavonoid biosynthesis. *Int. J. Mol. Sci.* 15, 1080–1095. doi: 10.3390/ijms15011080

Chiou, S. Y., Lai, J.-Y., Liao, J.-A., Sung, J.-M., and Lin, S.-D. (2018). In vitro inhibition of lipase, α-amylase, α-glucosidase, and angiotensin-converting enzyme by defatted rice bran extracts of red-pericarp rice mutant. *Cereal Chem.* 95, 167–176. doi: 10.1002/cche.10025

Chunta, S., Prathepha, P., and Jongdee, B. (2014). Nuances of traditional knowledge in utilization of rice landraces by a farming community in North-Eastern Thailand. *Indian J. Tradit. Knowl.* 13, 473–483.

Das, D. K., and Oudhia, P. (2000). Rice as a medicinal plant in Chhattisgarh, India. *NBPGR Newsl.* 122:46.

de Guzman, M. K., Parween, S., Butardo, V. M., Alhambra, C. M., Anacleto, R., Seiler, C., et al. (2017). Investigating glycemic potential of rice by unraveling compositional variations in mature grain and starch mobilization patterns during seed germination. *Sci. Rep.* 7:5854. doi: 10.1038/s41598-017-06026-0

Dong, Y., Xu, J., Xiao, K., Zhang, Y., Zhang, J., Luo, L., et al. (2008). Genomic regions associated with the degree of red coloration in pericarp of rice (*Oryza sativa* L.). *J. Cereal Sci.* 48, 556–560. doi: 10.1016/j.jcs.2007.11.011

Du, Y., Chu, H., Chu, I. K., and Lo, C. (2010). CYP93G2 Is a flavanone 2-hydroxylase required for c -glycosylflavone biosynthesis in rice. *Plant Physiol.* 154, 324–333. doi: 10.1104/pp.110.161042

Ferrari, B., Gianinetti, A., Finocchiaro, F., and Terzi, V. (2015). *Rc* gene sequence and expression evaluation in a red-kernel rice. *J. Rice Res.* 3:145. doi: 10.4172/2375-4338.1000145

Frank, T., Reichardt, B., Shu, Q., and Engel, K.-H. (2012). Metabolite profiling of colored rice (*Oryza sativa* L.) grains. *J. Cereal Sci.* 55, 112–119. doi: 10.1016/j.jcs.2011.09.009

Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., et al. (2006). The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* 49, 91–102. doi: 10.1111/j.1365-313X.2006.02958.x

Gallage, N. J., and Møller, B. L. (2015). Vanillin-bioconversion and bioengineering of the most popular plant flavor and its de novo biosynthesis in the vanilla orchid. *Mol. Plant* 8, 40–57. doi: 10.1016/j.molp.2014.11.008

Galland, M., Boutet-Mercey, S., Lounifi, I., Balzergue, S., Grandjean, O., Morin, H., et al. (2014). Compartmentation and Dynamics of flavone metabolism in dry and germinated rice seeds. *Plant Cell Physiol.* 55, 1646–1659. doi: 10.1093/pcp/pcu095

Gene Discovery Rice and Rice Science Center (2017). *Riceberry.* Available online at: http://dna.kps.ku.ac.th/index.php/news-articles-rice-rsc-rgdu-knowledge/rice-breeding-lab/riceberry-variety (accessed May 13, 2019)

Ghasemzadeh, A., Karbalaii, M. T., Jaafar, H. Z. E., and Rahmat, A. (2018). Phytochemical constituents, antioxidant activity, and antiproliferative properties of black, red, and brown rice bran. *Chem. Cent. J.* 12:17. doi: 10.1186/s13065-018-0382-9

Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., et al. (2013). Genetic analysis of the metabolome exemplified using a rice population. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20320–20325. doi: 10.1073/pnas.1319681110

Gross, B. L., Steffen, F. T., and Olsen, K. M. (2010). The molecular basis of white pericarps in African domesticated rice: novel mutations at the *Rc* gene. *J. Evol. Biol.* 23, 2747–2753. doi: 10.1109/TMI.2012.2196707

Gunaratne, A., Wu, K., Li, D., Bentota, A., Corke, H., and Cai, Y. (2013). Antioxidant activity and nutritional quality of traditional red-grained rice varieties containing proanthocyanidins. *Food Chem.* 138, 1153–1161. doi: 10.1016/j.foodchem.2012.11.129

Ham, T., Kwon, S. W., Ryu, S., and Koh, H. (2015). Correlation analysis between grain color and cyanidin-3-glucoside content of rice grain in segregate population. *Plant Breed. Biotechnol.* 3, 160–166. doi: 10.9787/pbb.2015.3.2.160

Han, Y. Y., Wang, J. W., Han, N., Liu, Q. J., Liu, T. M., Guan, F. M., et al. (2009). Duplication and sequence divergence of rice chalcone synthase genes. *Russ. J. Plant Physiol.* 56, 417–422. doi: 10.1134/s1021443709030169

Hashmi, M. I., and Tianlin, J. S. (2016). Minerals contents of some indigenous rice varieties of Sabah Malaysia. *Int. J. Agric. For. Plant.* 2, 31–34.

Hedge, S., Yenagi, N. B., and Kasturiba, B. (2013). Indigenous knowledge of the traditional and qualified ayurveda practitioners on the nutritional significance and use of red rice in medications. *Indian J. Tradit. Knowl.* 12, 506–511.

Hemamalini, S., Umamaheswari, D. S., Lavanya, D. R., and Umamaheswara, R. D. C. (2018). exploring the therapeutic potential and nutritional properties of 'Karuppu Kavuni' variety rice of Tamil Nadu. *Int. J. Pharma Bio Sci.* 9, 88–96.

Hu, J., Anderson, B., and Wessler, S. R. (1996). Isolation and characterization of rice *R* genes: evidence for distinct evolutionary paths in rice and maize. *Genetics* 142, 1021–1031.

Huang, Y. P., and Lai, H. M. (2016). Bioactive compounds and antioxidative activity of colored rice bran. *J. Food Drug Anal.* 24, 564–574. doi: 10.1016/j.jfda.2016.01.004

Hurtada, W. A., Barrion, A. S. A., and Nguyen-Orca, M. F. R. (2018). Mineral content of dehulled and well-milled pigmented and non-pigmented rice varieties in the Philippines. *Int. Food Res. J.* 25, 2063–2067.

Insuan, O., Chariyakornkul, A., Rungrote, Y., and Wongpoomchai, R. (2017). Antimutagenic and antioxidant activities of Thai rice brans. *J. Cancer Prev.* 22, 89–97. doi: 10.15430/jcp.2017.22.2.89

Irakli, M. N., Samanidou, V. F., Katsantonis, D. N., Billiaderis, C. G., and Papadoyannis, I. N. (2016). Phytochemical profiles and antioxidant capacity of pigmented and non-pigmented genotypes of rice (*Oryza sativa* L.). *Cereal Res. Commun.* 44, 98–110. doi: 10.1556/0806.43.2015.033

Islam, M. Z., Khalequzzaman, M., Prince, M. F. R. K., Siddique, M. A., Rashid, E. S. M. H., Ahmed, M. S. U., et al. (2018). Diversity and population structure of red rice germplasm in Bangladesh. *PLoS One* 13:e0196096. doi: 10.1371/journal.pone.0196096

Jesch, E. D., and Carr, T. P. (2017). Food ingredients that inhibit cholesterol absorption. *Prev. Nutr. Food Sci.* 22, 67–80. doi: 10.3746/pnf.2017.22.2.67

Kang, M.-Y., Kim, J.-H., Rico, C. W., and Nam, S.-H. (2011). A comparative Study on the physicochemical characteristics of black rice varieties. *Int. J. Food Property* 14, 1241–1254. doi: 10.1080/10942911003637350

Khatri (2018). *The Protection of Indigenous Knowledge for Peoples Health Rose*. Available online at: https://www.ghwatch.org/sites/www.ghwatch.org/files/indig_kph.pdf (accessed May 13, 2019).

Kim, C., Cho, M., Choi, Y., and Kim, J. (2011). Identification and characterization of seed-specific transcription factors regulating anthocyanin biosynthesis in black rice. *J. Appl. Genet.* 52, 161–169. doi: 10.1007/s13353-011-0027-3

Kim, C. K., Seol, Y. J., Shin, Y., Lim, H. M., Lee, G. S., Kim, A. R., et al. (2015). Whole-genome resequencing and transcriptomic analysis to identify genes involved in leaf-color diversity in ornamental rice plants. *PLoS One* 10:e0124071. doi: 10.1371/journal.pone.0124071

Kim, D. H., Park, S., Lee, J. Y., Ha, S. H., Lee, J. G., and Lim, S. H. (2018). A rice B-Box protein, OsBBX14, finely regulates anthocyanin biosynthesis in rice. *Int. J. Mol. Sci.* 9:E2190. doi: 10.3390/ijms19082190

Kim, G. R., Jung, E. S., Lee, S., Lim, S., Ha, S., and Lee, C. H. (2014). Combined mass spectrometry-based metabolite profiling of different pigmented rice (*Oryza sativa* L.) seeds and correlation with antioxidant activities. *Molecules* 19, 15673–15686. doi: 10.3390/molecules191015673

Kim, J. H., Lee, Y. J., Kim, B. G., Lim, Y., and Ahn, J.-H. (2008). Flavanone 3ß-hydroxylases from rice: key enzymes for favonol and anthocyanin biosynthesis. *Mol. Cells* 25, 312–316.

Ko, J. H., Kim, B. G., Hur, H. G., Lim, Y., and Ahn, J. H. (2006). Molecular cloning, expression and characterization of a glycosyltransferase from rice. *Plant Cell Rep.* 25, 741–746. doi: 10.1007/s00299-006-0119-4

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS?: a review. *Plant Methods* 9:29. doi: 10.1186/1746-4811-9-29

Krishnanunni, K., Ramaiah, S., and Anbarasu, A. (2014). Total phenolic content and " in-vitro " antioxidant assay of two medicinal rice varieties - Karungkavuni and Kuzhiadichan. *Int. J. Pharma Bio Sci.* 5, 540–548. doi: 10.1007/s13197-014-1292-z

Lachagari, V. B. R., Gupta, R., Lekkala, S. P., Mahadevan, L., Kuriakose, B., Chakravartty, N., et al. (2019). Whole genome sequencing and comparative genomic analysis reveal allelic variations unique to a purple colored rice landrace (*Oryza sativa* ssp. *Indica cv. Purpleputtu). Front. Plant Sci.* 10:513. doi: 10.3389/fpls.2019.00513

Lee, J., Seol, Y., Hahn, J., Won, S., Won, Y., Kim, Y., et al. (2015). Transcriptomics analyses of genes regulating anthocyanin production in black rice. *BioChip J.* 9, 59–66. doi: 10.1007/s13206-014-9108-9

Lim, S. H., and Ha, S. H. (2013). Marker development for the identification of rice seed color. *Plant Biotechnol. Rep.* 7, 391–398. doi: 10.1007/s11816-013-0276-1

Liu, Q., Luo, L., and Zheng, L. (2018). Lignins: biosynthesis and biological functions in plants. *Int. J. Mol. Sci.* 19:E335. doi: 10.3390/ijms19020335

Luo, L. P., Han, B., Yu, X. P., Chen, X. Y., Zhou, J., Chen, W., et al. (2014). Anti-metastasis activity of black rice anthocyanins against breast cancer: analyses using an ErbB2 positive breast cancer cell line and tumoral xenograft model. *Asian Pacific J. Cancer Prev.* 15, 6219–6225. doi: 10.7314/APJCP.2014.15.15.6219

Ma, Y., Id, H. D., Zhang, F., Li, Y., Yang, H., Tian, F.-P., et al. (2018). Transcriptomic analysis of *Lycium ruthenicum* Murr . during fruit ripening provides insight into structural and regulatory genes in the anthocyanin biosynthetic pathway. *PLoS One* 13:e0208627. doi: 10.1371/journal.pone.0208627

Maeda, H., Yamaguchi, T., Omoteno, M., Takarada, T., Fujita, K., Murata, K., et al. (2014). Genetic dissection of black grain rice by the development of a near isogenic line. *Breed. Sci.* 64, 134–141. doi: 10.1270/jsbbs.64.134

Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J. I., Ebana, K., et al. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J.* 81, 13–23. doi: 10.1111/tpj.12681

Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, J., et al. (2012). Dissection of genotype – phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J.* 70, 624–636. doi: 10.1111/j.1365-313X.2012.04903.x

Matsukawa, T., Villareal, M. O., and Isoda, H. (2016). The Type 2 diabetes-preventive effect of cyanidin-3-glucoside on adipocytes. *J. Dev. Sustain. Agric.* 11, 31–35.

Melini, V., and Acquistucci, R. (2017). Health-promoting compounds in pigmented Thai and wild rice. *Foods* 6:E9. doi: 10.3390/foods6010009

Merchante, C., Stepanova, A. N., and Alonso, J. M. (2017). Translation regulation in plants?: an interesting past, an exciting present and a promising future. *Plant J.* 90, 628–653. doi: 10.1111/tpj.13520

Nigatu, G., Hansen, J., Childs, N., and Seeley, R. (2017). *Sub-Saharan Africa is Projected to be the Leader in Global Rice Imports Highlights*. Available online at: https://ideas.repec.org/a/ags/uersaw/266026.html (accessed May 13, 2019).

Oh, H. J., Lee, Y.-J., Byeon, E.-J., Kang, B.-C., Kyeong, D.-S., and Kim, C.-K. (2018). Whole - genome resequencing and transcriptomic analysis of genes regulating anthocyanin biosynthesis in black rice plants. *3 Biotech* 8:115.

Oikawa, T., Maeda, H., Oguchi, T., Yamaguchi, T., Tanabe, N., Ebana, K., et al. (2015). The birth of a black rice gene and its local spread by introgression. *Plant Cell* 27, 2401–2414. doi: 10.1105/tpc.15.00310

Park, S., Choi, M. J., Lee, J. Y., Kim, J. K., Ha, S. H., and Lim, S. H. (2016). Molecular and biochemical analysis of two rice flavonoid 3'-hydroxylase to evaluate their roles in flavonoid biosynthesis in rice grain. *Int. J. Mol. Sci.* 17, 1549–1562. doi: 10.3390/ijms17091549

Pereira-Caro, G., Watanabe, S., Crozier, A., Fujimura, T., and Yokota, T. (2013). Phytochemical profile of a Japanese black – purple rice. *Food Chem.* 141, 2821–2827. doi: 10.1016/j.foodchem.2013.05.100

Petroni, K., Landoni, M., Tomay, F., and Calvenzani, V. (2017). Proximate composition, polyphenol content and anti-inflammatory properties of white and pigmented Italian rice varieties. *Univ. J. Agric. Res.* 5, 312–321. doi: 10.13189/ujar.2017.050509

Pintha, K., Yodkeeree, S., Pitchakarn, P., and Limtrakul, P. (2014). Anti-invasive activity against cancer cells of phytochemicals in red jasmine rice (*Oryza sativa* L.). *Asian Pacific J. Cancer Prev.* 15, 4601–4607. doi: 10.7314/apjcp.2014.15.11.4601

Poulev, A., Heckman, J. R., Raskin, I., and Belanger, F. C. (2019). Tricin levels and expression of flavonoid biosynthetic genes in developing grains of purple and brown pericarp rice. *PeerJ* 7:e6477. doi: 10.7717/peerj.6477

Qualley, A. V., Widhalm, J. R., Adebesin, F., Kish, C. M., and Dudareva, N. (2012). Completion of the core-oxidative pathway of benzoic acid biosynthesis in plants. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16383–16388. doi: 10.1073/pnas.1211001109

Rachasima, L. N., Sukkasem, R., Pongjaroenkit, S., Sangtong, V., Chowpongpang, S., and Sakulsingharoj, C. (2017). Expression analysis and nucleotide variation of *OsC1* gene associated with anthocyanin pigmentation in Thai rice cultivars. *Genomics Genet.* 10, 46–53.

Raghuvanshi, R. S., Dutta, A., Tewari, G., and Suri, S. (2017). Qualitative characteristics of red rice and white rice procured from local market of Uttarakhand?: a comparative study. *J. Rice Res.* 10, 49–53.

Rahman, M. M., Lee, K. E., Lee, E. S., Matin, M. N., Lee, D. S., Yun, J. S., et al. (2013). The genetic constitutions of complementary genes *Pp* and *Pb* determine the purple color variation in pericarps with cyanidin-3-O-glucoside depositions in black rice. *J. Plant Biol.* 56, 24–31. doi: 10.1007/s12374-012-0043-9

Rahman, S., Sharma, M. P., and Sahai, S. (2006). Nutritional and medicinal values of some indigenous rice varieties. *Indian J. Tradit. Knowl.* 5, 454–458.

Rao, Y., Li, Y., and Qian, Q. (2014). Recent progress on molecular breeding of rice in China. *Plant Cell Rep.* 33, 551–564. doi: 10.1007/s00299-013-1551-x

Roberts, J. E., and Dennison, J. (2015). The photobiology of lutein and zeaxanthin in the eye. *J. Ophtalmol.* 2015:687173. doi: 10.1155/2015/687173

Rysbekova, A. B., Kazkeyev, D. T., Usenbekov, B. N., Mukhina, Z. M., Zhanbyrbaev, E. A., Sartbaeva, I. A., et al. (2017). Prebreeding selection of rice with colored pericarp based on genotyping Rc and Pb genes. *Russ. J. Genet.* 53, 49–58. doi: 10.1134/s1022795416110119

Saitoh, K., Onishi, K., Mikami, I., Thidar, K., and Sano, Y. (2004). Allelic diversification at the C (*OsC1*) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. *Genetics* 168, 997–1007. doi: 10.1534/genetics.103.018390

Sakamoto, W., Ohmori, T., Kageyama, K., Miyazaki, C., Saito, A., Noda, K., et al. (2001). The purple leaf (*Pl*) Locus of Rice?: the *Plw* allele has a complex organization and includes two genes encoding basic helix-loop-helix proteins involved in anthocyanin biosynthesis. *Plant Cell Physiol.* 42, 982–991. doi: 10.1093/pcp/pce128

Sakulsingharoj, C., Inta, P., Sukkasem, R., Pongjaroenkit, S., Chowpongpang, S., and Sangtong, V. (2014). Overexpression of *OSB2* gene in transgenic rice up-regulated expression of structural genes in anthocyanin biosynthesis pathway. *Thai J. Genet.* 7, 173–182. doi: 10.14456/tjg.2014.24

Sakulsingharoj, C., Inta, P., Sukkasem, R., Pongjaroenkit, S., Chowpongpang, S., and Sangtong, V. (2016). Cloning and characterization of *OSB1* gene controlling anthocyanin biosynthesis from Thai black rice. *Genomics Genet.* 9, 7–18. doi: 10.14456/gag.2016.2

Saneei, P., Larijani, B., and Esmaillzadeh, A. (2016). Rice consumption, incidence of chronic diseases and risk of mortality?: meta-analysis of cohort studies. *Public Health Nutr.* 20, 233–244. doi: 10.1017/S1368980016002172

Santos-Sánchez, N. F., Salas-Coronado, R., Hernández-Carlos, B., and Villanueva-Cañongo, C. (2019). "Shikimic Acid pathway in biosynthesis of phenolic compounds," in *Plant Physiological Aspects of Phenolic Compounds*, eds M. Soto-Hernández, R. García-Mateos, and M. Palma-Tenango (London: IntechOpen), doi: 10.5772/intechopen.83815

Sarma, R., Swamy, H. V. V. K., and Shashidhar, H. E. (2018). Dealing with zinc and iron deficiency in rice?: combine Strategies to fight hidden hunger in developing countries. *Int. J. Curr. Microbiol. Appl. Sci.* 7, 1887–1895. doi: 10.20546/ijcmas.2018.703.224

Sasaki, N., Nishizaki, Y., Ozeki, Y., and Miyahara, T. (2014). The role of acyl-glucose in anthocyanin modifications. *Molecules* 19, 18747–18766. doi: 10.3390/molecules191118747

Sathya, A. (2013). Are the Indian rice landraces a heritage of biodiversity to reminisce their past or to reinvent for future?? *Asian Agrihist.* 17, 221–232.

Shao, Y., Hu, Z., Yu, Y., Mou, R., Zhu, Z., and Beta, T. (2018). Phenolic acids, anthocyanis, proanthocyanidins, antioxidant activity, minerals and their correlations in non-pigmented, red, and black rice. *Food Chem.* 239, 733–741. doi: 10.1016/j.foodchem.2017.07.009

Shao, Y., Jin, L., Zhang, G., Lu, Y., Shen, Y., and Bao, J. (2011). Association mapping of grain color, phenolic content, flavonoid content and antioxidant capacity in dehulled rice. *Theor. Appl. Genet.* 122, 1005–1016. doi: 10.1007/s00122-010-1505-4

Sharma, A., Patni, B., Shankhdhar, D., and Shankhdhar, S. C. (2013). Zinc – an indispensable micronutrient. *Physiol. Mol. Biol. Plants* 19, 11–20. doi: 10.1007/s12298-012-0139-1

Shih, C. H., Chu, H., Tang, L. K., Sakamoto, W., Maekawa, M., Chu, I. K., et al. (2008). Functional characterization of key structural genes in rice flavonoid biosynthesis. *Planta* 228, 1043–1054. doi: 10.1007/s00425-008-0806-1

Shozib, H. B., Jahan, S., Das, S. C., Alam, S., Amin, R. B., Hasan, M., et al. (2017). Mineral profiling of HYV rice in Bangladesh. *Vitam. Miner.* 6:164. doi: 10.4172/2376-1318.1000164

Spoel, S. H. (2018). Orchestrating the proteome with post-translational modifications. *J. Exp. Bot.* 69, 4499–4503. doi: 10.1093/jxb/ery295

Sun, X., Zhang, Z., Chen, C., Wu, W., Ren, N., Jiang, C., et al. (2018). The C–S–A gene system regulates hull pigmentation and reveals evolution of anthocyanin biosynthesis pathway in rice. *J. Exp. Bot.* 69, 1485–1498. doi: 10.1093/jxb/ery001

Suttiarporn, P., Chumpolsri, W., Mahatheeranont, S., Luangkamin, S., Teepsawang, S., and Leardkamolkarn, V. (2015). Structures of phytosterols and triterpenoids with potential anti-cancer activity in bran of black non-glutinous rice. *Nutrients* 7, 1672–1687. doi: 10.3390/nu7031672

Suwannakul, N., Punvittayagul, C., Jarukamjorn, K., and Wongpoomchai, R. (2015). Purple rice bran extract attenuates the aflatoxin B1-induced initiation stage of hepatocarcinogenesis by alteration of xenobiotic metabolizing enzymes. *Asian Pacific J. Cancer Prev.* 16, 3371–3376. doi: 10.7314/APJCP.2015.16.8.3371

Sweeney, M. T., Thomson, M. J., Cho, Y. G., Park, Y. J., Williamson, S. H., Bustamante, C. D., et al. (2007). Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet.* 3:e133. doi: 10.1371/journal.pgen.0030133

Sweeney, M. T., Thomson, M. J., Pfeil, B. E., and Mccouch, S. (2006). Caught red-handed?: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18, 283–294. doi: 10.1105/tpc.105.038430.1

Tan, Y. F., Sun, M., Xing, Y. Z., Hua, J. P., Sun, X. L., Zhang, Q. F., et al. (2001). Mapping quantitative trait loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. *Theor. Appl. Genet.* 103, 1037–1045. doi: 10.1007/s001220100665

Tantipaiboonwong, P., Pintha, K., Chaiwangyen, W., Chewonarin, T., Pangjit, K., Chumphukam, O., et al. (2017). Anti-hyperglycaemic and anti-hyperlipidaemic effects of black and red rice in streptozotocin-induced diabetic rats. *ScienceAsia* 43, 281–288.

Teng, S., Keurentjes, J., Bentsink, L., Koornneef, M., and Smeekens, S. (2005). Sucrose-specific induction of anthocyanin biosynthesis in arabidopsis requires the MYB75 / PAP1 Gene. *Plant Physiol.* 139, 1840–1852. doi: 10.1104/pp.105.066688.1840

Terungwa, T. I., and Yuguda, M. A. (2014). The impact of rice production, consumption and importation in Nigeria?: the political economy perspectives. *Int. J. Sustain. Dev. World Policy* 3, 90–99.

Thomas, R., Bhat, R., and Kuang, Y. T. (2015). Composition of amino acids, fatty acids, minerals and dietary fiber in some of the local and import rice varieties of Malaysia. *Int. Food Res. J.* 22, 1148–1155.

Tzin, V., and Galili, G. (2010). New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Mol. Plant* 3, 956–972. doi: 10.1093/mp/ssq048

Umadevi, M., Pushpa, R., Sampathkumar, K. P., and Bhowmik, D. (2012). Rice-traditional medicinal plant in India. *J. Pharmacogn. Phytochem.* 1, 6–12.

Valarmathi, R., Raveendran, M., and Senthil, N. (2014). Unraveling the nutritional and therapeutic properties of 'Kavuni' a traditional rice variety of Tamil Nadu. *J. Plant Biochem. Biotechnol.* 24, 305–315. doi: 10.1007/s13562-014-0274-6

Verma, D. K., and Shukla, K. (2011). Nutritional value of rice and their importance. *Indian Farmers Dig.* 44, 21–35.

Vichit, W., and Saewan, N. (2015). antioxidant activities and cytotoxicity of Thai pigmented rice. *Int. J. Pharm. Pharm. Sci.* 7, 329–334.

Voss-Fels, K. P., Stahl, A., and Hickey, L. T. (2019). Q&A: modern crop breeding for future food security. *BMC Biol.* 17:18.

Wahyuni, A. S., Munawaroh, R., and Da'i, M. (2016). Antidiabetic mechanism of ethanol extract of black rice bran on diabetic rats. *Natl. J. Physiol. Pharm. Pharmacol.* 6, 106–110. doi: 10.5455/njppp.2015.5.1111201590

Waiyawuththanapoom, P., Waiyawuththanapoom, W., and Tirastittam, P. (2015). Social media as a channel for Thailand's Rice Berry Product. *Int. J. Econ. Manag. Eng.* 9, 904–907.

Wambugu, P., Ndjiondjop, M., Furtado, A., and Henry, R. (2018). Sequencing of bulks of segregants allows dissection of genetic control of amylose content in rice. *Plant Biotechnol. J.* 16, 100–110. doi: 10.1111/pbi.12752

Wang, H., Xu, X., Vieira, F. G., Xiao, Y., Li, Z., Wang, J., et al. (2016). The power of inbreeding: NGS-based GWAS of rice reveals convergent evolution during rice domestication. *Mol. Plant* 9, 975–985. doi: 10.1016/j.molp.2016.04.018

Wickert, E., Schiocchet, M. A., Noldin, J. A., Raimondi, J. V., De Andrade, A., Scheuermann, K. K., et al. (2014). Exploring variability?: New Brazilian varieties SCS119 Rubi and SCS120 Onix for the specialty rices market. *Open J. Genet.* 4, 157–165. doi: 10.4236/ojgen.2014.42016

Widhalm, J. R., and Dudareva, N. (2015). A familiar ring to it: Biosynthesis of plant benzoic acids. *Mol. Plant* 8, 83–97. doi: 10.1016/j.molp.2014.12.001

World Rice Production (2019). *World Rice Production 2019/2020*. Available online at: http://www.worldagriculturalproduction.com/crops/rice.aspx/ (accessed May 17, 2019).

Xiongsiyee, V., Rerkasem, B., Veeradittakit, J., Saenchai, C., Lordkaew, S., and Prom-U-Thai, C. T. (2018). Variation in grain quality of upland rice from Luang Prabang province, Lao PDR. *Rice Sci.* 25, 94–102. doi: 10.1016/j.rsci.2018.02.002

Xu, T. Y., Sun, J., Chang, H. L., Zheng, H. L., Wang, J. G., Liu, H. L., et al. (2017). QTL mapping for anthocyanin and proanthocyanidin content in red rice. *Euphytica* 213:243.

Xu, W., Grain, D., Bobet, S., Le Gourrierec, J., Thevenin, J., Kelemen, Z., et al. (2013). Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB – bHLH – WDR complexes and their targets in Arabidopsis seed. *New Phytol.* 202, 132–144. doi: 10.1111/nph.12620

Yang, X., Xia, X., Zeng, Y., Nong, B., Zhang, Z., Wu, Y., et al. (2018). Identification of candidate genes for gelatinization temperature, gel consistency and pericarp color by GWAS in rice based on SLAF-sequencing. *PLoS Genet.* 13:e0196690. doi: 10.1371/journal.pone.0196690

Yoshimura, Y., Zaima, N., Moriyama, T., and Kawamura, Y. (2012). Different localization patterns of anthocyanin species in the pericarp of black rice revealed by imaging mass spectrometry. *PLoS One* 7:e31285. doi: 10.1371/journal.pone.0031285

Zhang, H., Shao, Y., Bao, J., and Beta, T. (2015). Phenolic compounds and antioxidant properties of breeding lines between the white and black rice. *Food Chem.* 172, 630–639. doi: 10.1016/j.foodchem.2014.09.118

Zhang, L., Dong, Y., Wang, Q., Du, C., Xiong, W., Li, X., et al. (2017). iTRAQ-based proteomics analysis and network integration for kernel tissue development in maize. *Mol. Sci.* 18:1840. doi: 10.3390/ijms18091840

Zhang, Y., Xu, S., Cheng, Y., and Peng, Z. (2018). Transcriptome profiling of anthocyanin- related genes reveals effects of light intensity on anthocyanin biosynthesis in red leaf lettuce. *PeerJ* 6:e4607. doi: 10.7717/peerj.4607

Zhao, J., Pang, Y., and Dixon, R. A. (2010). The mysteries of proanthocyanidin transport and polymerization. *Plant Physiol.* 153, 437–443. doi: 10.1104/pp.110.155432

Zubair, M., Anwar, F., Ashraf, M., and Uddin, K. (2012). Characterization of high-value bioactives in some selected varieties of Pakistani rice (*Oryza sativa* L.). *Int. J. Mol. Sci.* 13, 4608–4622. doi: 10.3390/ijms13044608

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Mbanjo, Kretzschmar, Jones, Ereful, Blanchard, Boyd and Sreenivasulu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Nutrition

ORIGINAL RESEARCH
published: 07 April 2020
doi: 10.3389/fnut.2020.00026

Zinc Biofortified Rice Varieties: Challenges, Possibilities, and Progress in India

D. Sanjeeva Rao[1], C. N. Neeraja[1], P. Madhu Babu[1], B. Nirmala[1], K. Suman[1], L. V. Subba Rao[1], K. Surekha[1], P. Raghu[2], T. Longvah[2], P. Surendra[3], Rajesh Kumar[4], V. Ravindra Babu[1] and S. R. Voleti[1]*

[1] ICAR-Indian Institute of Rice Research, Hyderabad, India, [2] ICMR-National Institute of Nutrition, Hyderabad, India, [3] Agricultural Research Station, University of Agricultural Sciences-D, Bangalore, India, [4] Department of Plant Breeding and Genetics, AICRIP (Rice), Rajendra Agricultural University, Samastipur, India

Zinc malnutrition is a major issue in developing countries where polished rice is a staple food. With the existing significant genetic variability for high zinc in polished rice, the development of biofortified rice varieties was targeted in India with support from HarvestPlus, Department of Biotechnology, and Indian Council of Agricultural Research of Government of India. Indian Institute of Rice Research (IIRR) facilitates rice varietal release through All India Coordinated Rice Improvement Project (AICRIP) and also supports rice biofortification program in India. Various germplasm sets of several national institutions were characterized at IIRR for their zinc content in brown rice using energy-dispersive X-ray fluorescence spectroscopy indicating the range of zinc to be 7.3 to 52.7 mg/kg. Evaluation of different mapping populations involving wild germplasm, landraces, and varieties for their zinc content showed the feasibility of favorable recombination of high zinc content and yield. Ninety-nine genotypes from germplasm and 344 lines from mapping populations showed zinc content of ≥28 mg/kg in polished rice meeting the target zinc content set by HarvestPlus. Through AICRIP biofortification trial constituted since 2013, 170 test entries were nominated by various national institutions until 2017, and four biofortified rice varieties were released. Only the test entry with target zinc content, yield, and quality parameters is promoted to the next year; thus, each test entry is evaluated for 3 years across 17 to 27 locations for their performance. Multilocation studies of two mapping populations and AICRIP biofortification trials indicated the zinc content to be highly influenced by environment. The bioavailability of a released biofortified rice variety, *viz*., DRR Dhan 45 was found to twice that of control IR64. The technology efficacy of the four released varieties developed through conventional breeding ranged from 48 to 75% with zinc intake of 38 to be 47% and 46 to 57% of the RDA for male and female, respectively. The observations from the characterization of germplasm and mapping populations for zinc content and development of national evaluation system for the release of biofortified rice varieties have been discussed in the context of the five criteria set by biofortification program.

Keywords: rice, biofortification, high zinc, germplasm, varieties, AICRIP, RIL's

OPEN ACCESS

Edited by:
Sachiko Narita Isobe, Kazusa DNA Research Institute, Japan

Reviewed by:
Nimai Prasad Mandal, Indian Council of Agricultural Research, India
Chikara Kuwata, Chiba Prefectural Agriculture and Forestry Research Center, Japan

****Correspondence:***
C. N. Neeraja
cnneeraja@gmail.com

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Nutrition

Received: *27 August 2019*
Accepted: *25 February 2020*
Published: *07 April 2020*

Citation:
Sanjeeva Rao D, Neeraja CN, Madhu Babu P, Nirmala B, Suman K, Rao LVS, Surekha K, Raghu P, Longvah T, Surendra P, Kumar R, Babu VR and Voleti SR (2020) Zinc Biofortified Rice Varieties: Challenges, Possibilities, and Progress in India. Front. Nutr. 7:26. doi: 10.3389/fnut.2020.00026

INTRODUCTION

Among various food crops, rice is the world's most important crop, and more than half of the global population is dependent on rice. Greater than 90% of rice is produced and consumed in Asia (1). India is the second largest producer of rice with a production of 112.76 million metric tons during 2017–2018 (2). Imbalanced supply of essential amino acids, micronutrients, and vitamins leads to their deficiency or accumulation affecting the human metabolism. According to World Health Organization, zinc deficiency is the fifth important factor for illness and diseases in developing countries and 11th in world (3). Worldwide, the prevalence of zinc deficiency has been estimated to be ∼20% in soils (4). Zinc deficiency causes diarrhea and respiratory diseases, leading to 400,000 deaths annually across the world (5, 6). Zinc deficiency is also associated with poor growth, loss of appetite, skin lesions, impaired taste acuity, delayed wound healing, hypogonadism, delayed sexual maturation, and impaired immune response (7, 8). Every year 1.31 million disability-adjusted life-years (DALYs) are lost in India because of zinc malnutrition. An ex-ante analysis of zinc biofortification of rice in India revealed that of the 1.31 million DALYs lost, 0.142 and 0.456 million DALYs could be saved in pessimistic and optimistic assumptions, if zinc-biofortified rice is introduced (9).

As a "global public good," international agriculture research in the form of International Rice Research Institute and Center for Improvement of Wheat and Maize merged into Consultative Group on International Agricultural Research (CGIAR) initially implemented "Green Revolution" during 1960s, successfully leading to enhanced grain production through the development of high-yielding varieties (HYVs). However, grains of HYVs contain lesser amounts of nutrients; in case of rice, polishing further reduces the nutrients content, *viz.*, iron and zinc. Subsequently in 1991, considering concern raised by the international nutrition community about micronutrient deficiency as a global problem, CGIAR initiated studies on developing "micronutrient-dense" staple crops (10) in the name of biofortification. In 2003, CGIAR launched HarvestPlus Challenge Program as a global program, with the objective of developing biofortified staple crops such as wheat, rice, maize, cassava, and so on, through plant breeding (11, 12). Biofortification of rice grain with iron was initiated in 1992 and zinc in 1995 (13). Improving nutrition is also being targeted under 12 of the 17 Sustainable Development Goals set by United Nations (https://www.undp.org/content/undp/.../sustainable-development-goals.html).

The word "biofortification" refers to enhancing bioavailable micronutrient content of food crops through genetic selection via plant breeding. It is a promising strategy to address nutrition security and was initiated with five criteria, *viz.*, (1) crop productivity (i.e., yield) must be maintained or increased to ensure farmer acceptance; (2) the enhanced micronutrient level must have significant impact on human health; (3) the enhanced micronutrient trait must be relatively stable across various edaphic environments and climatic zones; (4) the bioavailability of micronutrients in enriched lines must be tested in humans to ensure that they improve the micronutrient status of people preparing and eating them in traditional ways within normal household environments; and (5) consumer acceptance must be tested (taste and cooking quality must be acceptable to household members) to ensure maximum impact on nutritional health (14).

Although a few independent groups have initiated research on biofortification since 2000 in India, the programs kick-started after HarvestPlus program implementation in India during 2007. Department of Biotechnology (DBT) and Indian Council of Agricultural Research (ICAR) of Government of India have also initiated support to biofortification projects leading to the consorted national and international research efforts toward the development of biofortified rice varieties. With the negative impact of zinc deficiency on human metabolism, especially in countries with rice as a major staple food, development of zinc-biofortified varieties has become one of the important objectives of crop improvement programs. The success of biofortification relies on the existence of diversity for the target trait in the germplasm of crop, successful recombination of the target trait with yield, and identification of suitable recombinants with yield and target trait through multilocation evaluation. To phenotype zinc content of rice grains across thousands of germplasm lines or mapping populations, X-ray fluorescence spectroscopy (XRF) was found to be promising because of its high throughput, cost-effectiveness, and rapid analyses over atomic absorption spectrometry and inductively coupled plasma mass spectrometry (15). HarvestPlus has supported rice biofortification research of several laboratories in India through sponsoring XRF equipment and providing preliminary biofortified rice breeding lines.

ICAR–Indian Institute of Rice Research (IIRR) is a national institute that facilitates rice varietal release through All India Coordinated Rice Improvement Project (AICRIP) (http://www.icar-iirr.org/); IIRR is also translating the benefits of the research on zinc biofortification of rice to society through release of biofortified rice varieties through AICRIP and coordination of a Consortia Research Platform on biofortification of the cereals. In the present study, we summarize our observations from the studies on characterization of germplasm and mapping populations for zinc content and development of national evaluation system, that is, AICRIP, for the release of biofortified rice varieties and its successful implementation in India.

MATERIALS AND METHODS

Two datasets were analyzed for characterization of zinc content in rice, *viz.*, (1) data of zinc content from the field experiments of genotypes and recombinant inbred lines (RILs) populations

Abbreviations: ED-XRF, Energy Dispersive X-ray fluorescence spectroscopy; RIL, Recombinant Inbred Line; ICAR, Indian Council of Agricultural Research; ICMR, Indian Council of Medical Research; IIRR, Indian Institute of Rice Research; AICRIP, All India Coordinated Rice Improvement Programme; NIN, National Institute of Nutrition; CRP, Consortia Research Platform; ARS, Agricultural Research Station; IVT, Initial Variety Trial; AVT, Advance Varietal Trial.

developed by our group at IIRR and (2) data from analysis of the zinc content in the germplasm and RIL populations of various national institutions including IIRR at XRF facility of IIRR.

Conduct of Field Experiments at IIRR

The field experimental material included 170 rice genotypes comprising landraces, released varieties, and derivatives of various crosses (**Table 1**) and 13 RIL populations developed from the six identified donors for high zinc in polished rice (Chittimuthyalu, Jalpriya, BR2655, Type-3, Suraksha, and Ranbir Basmati) and five mega varieties (MTU1010, IR64, Swarna, RPBio226, and PR116) using a single seed descent method (**Table 3**). The germplasm experiment was laid out in randomized block design with two replications. The 13 mapping populations were grown in 13 blocks in an augmented block design with IR64 and Chittimuthyalu as replicated checks. The germplasm and mapping populations were evaluated at the research farm of IIRR, Hyderabad, during the wet season following the recommended package of rice production and protection practices. The experimental soil characteristics were as follows: pH 8.2; non-saline (EC 0.7l dS/m); calcareous (free $CaCO_3$ 5.01%); CEC 44.1 C mol (p+)/kg soil and medium soil organic carbon (0.69%); low soil available nitrogen (228 kg ha^{-1}); high available phosphorus (105 kg $P_2O_5 ha^{-1}$); high available potassium (530 kg K_2O ha^{-1}); and high available zinc (12.5 mg/kg). A RIL population, *viz*., MTU1010/Jalpriya, was also grown at Agriculture Research Station (ARS), Mugad and Rajendra Agricultural University (RAU), Pusa, and another RIL population, MTU1010/Suraksha was evaluated at ARS, Mugad. For each genotype and RIL, three representative plants from the middle row were harvested at maturity, and single plant yield (g) adjusted to 14% grain moisture content was recorded. The single plant yield from the three plants was pooled and divided into three parts to be analyzed as three replicates for zinc content.

Evaluation of Rice Germplasm and Mapping Populations at IIRR

Seventeen sets of rice genotypes, several mapping populations, and the selected lines from mapping populations from various national institutions including IIRR were evaluated for their zinc content at the XRF facility of IIRR.

Constitution of Biofortification Trial Under AICRIP

To evaluate the breeding lines developed with high zinc by various institutions of India, a national evaluation study was constituted under AICRIP as Biofortification trial during 2013 with two kinds of check genotypes for zinc content and yield. Initially, during 2013, to identify the promising lines in biofortification trial, a modest threshold/target zinc value of 20 mg/kg was set, as most of the cultivated rice varieties showed grain zinc content of ∼15 mg/kg. From 2015, optimum threshold/target content of zinc in polished rice was increased to 24 mg/kg to match the threshold value of international target set earlier by HarvestPlus. The landraces, Kalanamak and Chittimuthyalu, with high zinc are the two check genotypes for zinc content in the trial (**Supplementary Table 1**). BPT5204, also known as Samba Mahsuri (BPT), popular with farmers and consumers across India because of its high yield and preferred cooking quality was included as yield check genotype. Because of the late crop duration of BPT, a mid–early duration popular variety IR64 was also included as another yield check during 2014. A hybrid (DRRH3) check was also added as another yield

TABLE 1 | Range and classification based on zinc content in brown rice of 170 rice genotypes comprising landraces, released varieties, and derivatives from various crosses.

Name of the genotype/cross derivative	Zinc range (mg/kg)	No. of genotypes
Chandrahasini, Manipuri, Satabdi, Bharathy, ARB-47, Varsha, Govindbhog, Jhelum, ARB-35, NDR6279 and KalluRandaikar	12–15	12
ARB-43, Karma mahsuri, Sankar, Prafulla, NDR359, PR116/NDR359, Karna, Radhi, NLR145/Sheshi, TKM-9, PR116/Sheshi, Badshabhog, Samalei, R133-968-2-1, Pushpaka, ARB-36, Sarathi, Sarabasti, PB-1, Vandana, SKAU-338, PS-4, Chandarhasini, ARB-44, Moti, NDR2008, ARB-34, ARB-40, KHP-2, Pratikhya, K-116, K-39, Jeerigesanna, Kamod, ARB-39, ARB-30, Ketakijoha, Ranbir Basmati, NDR2026, ShushkSamrat, ARB-52, ARB-54, 28(B), ARB-51, Karidaddi, ARB-46, ARB-37, 117(B), BPT5204/Chittimuthyalu-74, Naveen, Basmati 370, NC365, CharthalaiPokkali, Jodumani, ARB-48, 142(S), ChampaKhushi, ARB-50, Shalimar Rice-1, ADT45, Heera, Makom, Khitish, Type-3,Jyothi, MattaTriveni, PR116/Ranbir Basmati-63, Seetabhog, CSAR 840, PR116/Ranbir Basmati-117, Sneha and ARB-33	15.1–20.0	72
ARB-31, NLR145/Sheshi, ARB-53, Kalanamak-Birdapur, ARB-38, ARB-45, BPT5204/Vikas, BPT5204/Ranbir Basmati-72, DL-184, VytillaAnakondan, Varalu, Mahalaxmi, Pathara, Savitri, Gouri, China 1007, Panvel-3, Nuadhusara, Pratikhya, Tara, Njavara, Kalinga-III, KadamakudyPokkali, ARB-56, Sebati, Pant Sugandhdhan-17, Ranjit, Manika, Karthika, 14(S), 14-3, Kanchana, NLR 145/Type-3−42,Swarna, BPT5204/Kalanamak, Azucena, BPT5204/Jalmagna, Ambemohr, 140(M), 196(M), ARB-41, 236(K), 176(M), MSE-9, Vasumati, Pathat-23, Pimpudipasa, Tilakkachari, Aghonibora, Vasumati, 51(B), NLR 145/Type-3-70, 166(M), Moirangphou, ARB-49, 185(M), Birupa, BPT5204/Basmati 370, ARB-32, BPT5204/Chittimuthyalu-MS, Swarna Sub-1, RadhuniPagal, BR2655, IET-23427, 65-(S), ARB-42, Akutphou, Seetasail, PB-164, Bindli, EdavankudiPokkali, Jalpriya and Suraksha	20.1–25.0	71
Subhadra,Chittimuthyalu,ARB-55, 182(M), Sabita, Kalanamak, Taroari Basmati, 233(K), BPT5204/ Chittimuthyalu-SB, BPT5204/Sheshi, Kasturi and 24(K)	25.1–30.0	12
High Iron Rice and Improved Chittimuthyalu	30.1–35.0	2

check during 2015 to compare the hybrids evaluated under this trial, but discontinued from 2016 because of the lack of hybrids nominated as test entries for evaluation. Under AICRIP, seed of potential lines gets nominated by the developers from various national institutions to IIRR as test entries. IIRR assigns an initial evaluation trial number to each test entry and constitutes Initial Variety Trial (IVT) with the nominated test entries along with check genotypes for yield and zinc content and sends seed along with details of the conduct of experiment to 10 to 30 locations across the country each year. After completion of the harvest, the agro-morphological and yield data of each location were sent to IIRR for compilation and identification of the promising lines with high yield and zinc. The seed is also sent to IIRR for estimation of zinc through XRF. The quality of test entries is also studied at IIRR as 16 traits/parameters (http://www.aicrip-intranet.in/PlantBreeding.aspx). The promising test entries of IVT are promoted to next year as advanced varietal trial 1 (AVT1) and subsequently to AVT2 based on their zinc content, yield, and quality. The consistent promising test entry during the 3 years of evaluation is identified as a biofortified variety.

Zinc Estimation

The seed samples were dehusked and polished as per the standardized protocol for analyzing samples with XRF (16). Seed was cleaned manually and dehusked using JLGJ4.5 rice husker with non-ferrous components (Jingjian Huayuan International Trade Co., Ltd., Jiangsu Sheng, China) sponsored by HarvestPlus. Dehusked brown rice was sieved to remove the broken rice grains, and the full brown rice grains were cleaned with soft tissue paper. Sample was gently rubbed for 1 min against paper in the hands of a trained person wearing gloves to ensure removal of particles other than rice. Each brown rice sample was milled in a specially designed K-710 Non-Ferrous Rice Polisher (Krishi International India Ltd., Hyderabad, India), and the percentage of polishing was calculated based on weight of the polished (white) rice to the brown rice. As polished rice is the most preferred form of consumption, brown rice samples were subjected to polishing for 90 to 120 s until the desirable whiteness was reached. Cleaning of the polished rice was repeated like that of brown rice. Time gap between polishing and cleaning was minimized to avoid sticking of the bran particles to the polished grain. Each sample of brown or polished rice (5 g) was subjected to energy-dispersive XRF (ED-XRF) (OXFORD Instruments X-Supreme 8000, Highwycombe Bucks, England) sponsored by HarvestPlus at IIRR. Samples were scanned against the prestandardized method, which converts the fluorescence intensity of each sample into zinc content (mg/kg). The percentage loss of zinc was calculated using the zinc content in brown and polished rice.

Statistical Analysis

Statistical analysis was done using R software, R foundation for Statistical computing, Vienna, Austria and all the graphs were made using ggplot2. Correlation between single plant yield and zinc content was done by performance analytics package.

Technology Efficacy of Zinc Biofortification in Rice

The current zinc intake from rice with the popular varieties was calculated based on the per-capita rice consumption of 220 g/day (17). Under assumption that the current rice consumption pattern is maintained in India and considering the adoption of biofortified rice varieties as technology, improved intake of zinc through biofortified rice varieties was calculated. The technology efficacy (E) of zinc biofortification in rice was derived as follows:

$$\text{Technology efficacy (E)} = \frac{\ln\left[\frac{IZ}{CZ}\right] - \left[\frac{IZ-CZ}{RDA}\right]}{\ln\left[\frac{RDA}{CZ}\right] - \left[\frac{RDA-CZ}{RDA}\right]} * 100$$

where IZ is improved zinc intake, CZ is current zinc intake, and RDA is the Recommended Dietary Allowance of zinc.

RESULTS AND DISCUSSION

Country-wide food supplies including those from India were found to be zinc-deficient for ∼15% to 20% of the world, based on the FAO food balance sheets (18). Biofortification strategy is suggested to be the most feasible and cost-effective means of delivering micronutrients to populations with limited access to diverse diets and other micronutrient interventions (15). Of various strategies for biofortification, development of biofortified varieties is often considered the most sustainable, especially if major staple cereal food crops are targeted. Once developed, the biofortified varieties can easily be adopted by the farmers with minimal cost of cultivation, unlike agronomic biofortification with additional expenditure on external nutrient application. Biofortification of rice with zinc in polished rice is envisaged as a promising approach to address zinc deficiency in countries where rice is the major staple food crop.

Genetic Variability for Zinc in Rice Grains

For developing varieties with high zinc in polished rice through conventional breeding, genetic variability for grain zinc content should exist in the germplasm. Since 1992, various sets of germplasm comprising varieties including *indica* and *japonica*, landraces from various countries and wild accessions, have been analyzed, and a wide variability of zinc content was reported (13, 19–21). In our study, zinc content in brown rice of the sample set of 170 rice germplasm ranged from 12 to 31.7 mg/kg (**Table 1**). Most of the genotypes showed zinc content in the range of 15 to 25 mg/kg, and four promising donors with >25 mg/kg zinc content were selected as donors for the development of mapping populations at IIRR. Zinc content in the germplasm comprising landraces, breeding lines, and varieties from 16 national institutions evaluated at XRF facility of IIRR ranged from 7.3 to 52.7 mg/kg (**Table 2**). Unlike reported mutants of IR64 with high zinc content in polished rice, higher zinc content was not observed in the mutants of BPT (ICAR-IIRR-SM-MUT) in our study (22). The mean values of zinc content in brown rice of germplasm sets ranged from 15.9 to 27.3 mg/kg. Of 3,177 germplasm evaluated, 3.1% showed zinc content of ≥35

TABLE 2 | Zinc content in brown rice of germplasm sets analyzed at the XRF facility of IIRR and the number of promising lines with zinc ≥35 mg/kg.

Institution	No. of genotypes	Zinc content (mg/kg)		No. of genotypes with zinc content ≥35 mg/kg
		Range	Mean	
ANGRAU	27	11.2–23.7	18.7	Nil
ARS-BRL	20	19.2–34.0	24	Nil
BSKKV-ARS-RTN	25	18.6–34.5	26.7	Nil
ICAR-IARI	221	17.0–50.2	26.9	20
ICAR-IIRR	1,317	9.6–52.7	22.7	33
ICAR-IIRR-SM-MUT	64	11.2–31.1	18.8	Nil
ICAR-NRRI	66	13.6–42.0	27.3	8
ICAR-RCNEH	51	17.4–34.2	24.8	Nil
IGKV-Raipur	840	13.1–48.3	25.1	20
Liberian-African	89	15.1–39.0	23.9	3
PJTSAU-ARS-JGL	52	7.3–28.0	15.9	Nil
RARS-AAU	73	17.0–34.8	23.1	Nil
RRS-Chinsurah	2	16.0–16.5	16.3	Nil
SKUAST-MRCFC	130	13.9–35.5	23.3	2
TNAU-ADT	47	13.9–44.3	26.8	9
UAS-Bangalore	134	9.7–36.1	22.8	2
UAS-Dharwad	19	17.4–38.7	25.7	2
Total	3,177	NA	NA	99

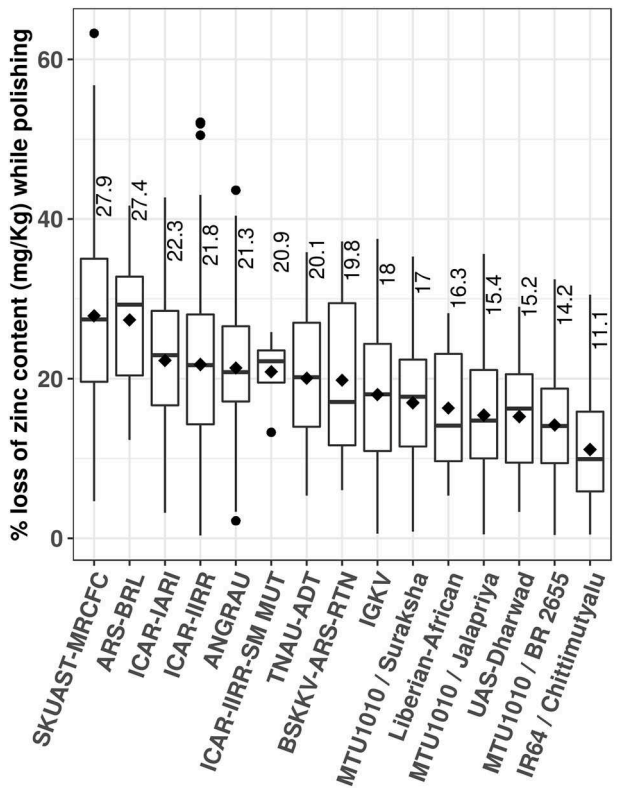

FIGURE 1 | The range of loss in zinc content (%) during polishing in the germplasm sets and mapping populations. ANGRAU, Acharya N. G. Ranga Agricultural University, ARS-BRL, Agricultural Research Station, breeding lines, Shirgoan, BSKKV-ARS-RTN, Dr. BalasahebSawantKonkanKrishiVidyapeeth, Agricultural Research Station, Ratnagiri, ICAR-IARI, Indian Council of Agricultural Research–Indian Agricultural Research Institute, ICAR-IIRR, ICAR–Indian Institute of Rice Research, ICAR-IIRR-SM-MUT, Sambha Mahsuri Mutants, IGKV, Indira Gandhi Krishi Vishwavidyalaya-Raipur, SKUAST-MRCFC, Share e Kashmir University of Agricultural Science & Technology–Mountain Research Center for Field Crops, TNAU-ADT, Tamil Nadu Agricultural University, Aduthurai, UAS-Dharwad, University of Agricultural Sciences–Dharwad.

mg/kg, and only 0.9% had zinc content of ≥40 mg/kg. Some of the reported ranges of grain zinc content include 13.5 to 58.4 mg/kg in 1,138 germplasm (13), 24.0 to 35.0 mg/kg in aromatic genotypes (14), 25.5 to 37.0 mg/kg in eight highly cultivated varieties (23), 12 to 27.6 mg/kg in 60 Iranian rice genotypes (24), and 13.3 to 34.2 mg/kg in 18 indigenous cultivars of Tripura state of India (25). In case of polished rice, zinc content ranged from 4.8 to 40.9 mg/kg in 170 rice germplasm of the present study. The reported ranges of grain zinc content in polished rice were 16.0 to 26.5 mg/kg in eight highly cultivated varieties (23), and up to 33 mg/kg zinc content in polished rice of 246 germplasm (26). Across the analyzed samples, a mean percentage loss of zinc content was found to be 19.0%, that is, 1.9 mg/kg loss in every 10 mg/kg of brown rice during polishing. The mean percentage loss of zinc varied from 11.1 to 27.9% in mapping populations and various germplasm sets (**Figure 1**). The reported losses of zinc content due to polished ranged from 5 to 30% (26, 27). The variation in the polishing could be due to the difference in the thickness of aleurone layer across rice genotypes as reported earlier (23, 28).

The zinc content in brown rice of varieties widely grown by the farmers ranged from <5 to 25 mg/kg (16, 20). The zinc content in polished rice of the popular rice varieties widely grown by the farmers ranges from (<12–14 mg/kg). Considering the threshold value of 28 mg/kg by HarvestPlus and 19.0% overall mean loss of zinc during polishing, germplasm with zinc content ≥35 mg/kg in brown rice can be considered as promising as donors to meet the target zinc.

Among 16 mapping populations grown at IIRR, the mean zinc content was least (18.6 mg/kg) in IR64/Jalpriya and highest (36.3 mg/kg) in RP Bio 226/Jalpriya (**Figure 2**). Of four mapping populations with IR64 as one of the parents, the mean zinc content was least (18.6 mg/kg) in cross with Jalpriya and highest (26.9 mg/kg) in cross with Chittimuthyalu. In the five mapping populations derived from MTU1010 as one of the parents, the mean zinc content was least (20.6 mg/kg) in cross with Chittimuthyalu and highest (29.0 mg/kg) in cross with Suraksha. Of the five mapping populations with Swarna as one of the parents, mean zinc content was least (19.2 mg/kg) in cross with Type-3 and highest (28.3 mg/kg) in cross with Ranbir Basmati followed closely by BR 2655. The mean zinc content of mapping populations and number of promising lines per mapping population varied despite common donors and recipients, suggesting the complexity of the trait and effect of interactions of parents for zinc content in progeny as per the cross combinations. Hence, screening of rice germplasm for higher zinc content should be a continuous process to identify compatible donors. Approximately 389 lines with zinc content

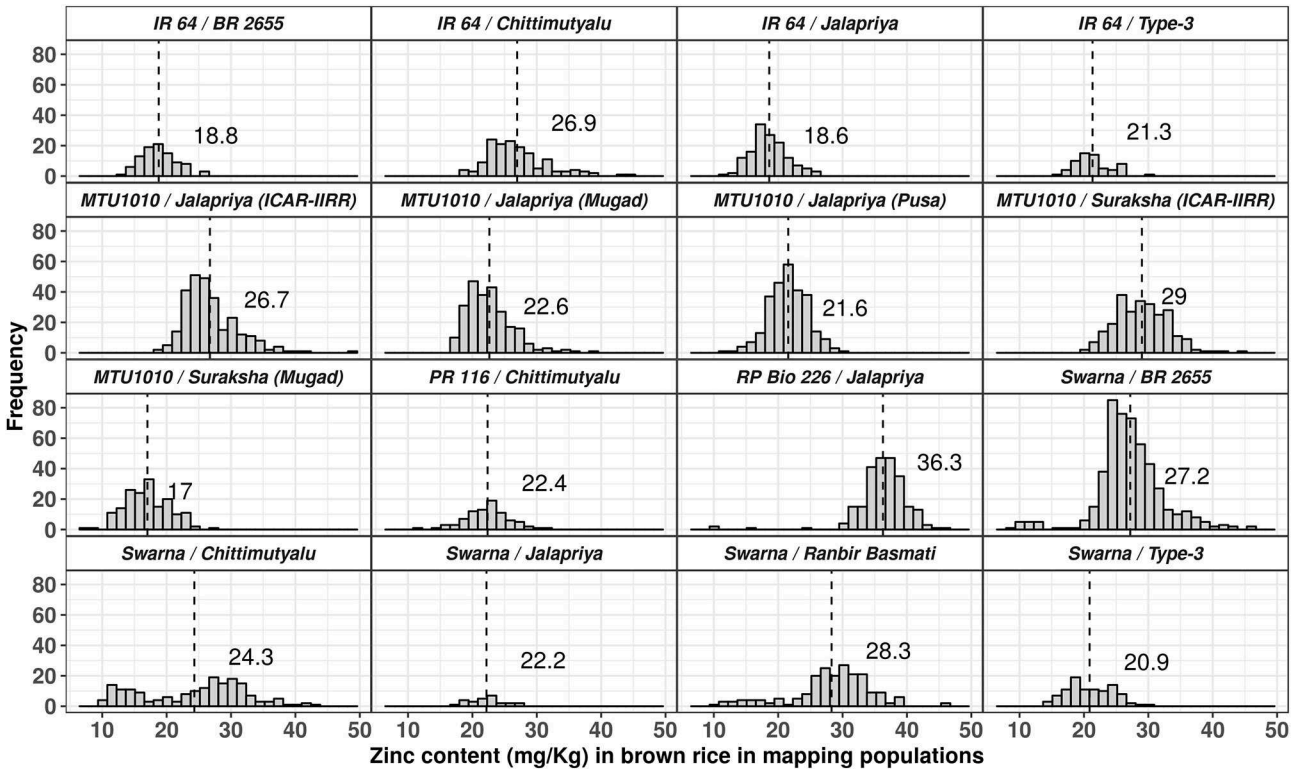

FIGURE 2 | Distribution of zinc content in brown rice of 13 mapping populations evaluated at IIRR. MTU1010/Jalpriya also evaluated at ARS, Mugad, and RAU, Pusa. MTU1010/Suraksha also evaluated at ARS, Mugad.

TABLE 3 | Zinc content in brown rice of recombinant inbred lines (RILs) and their parents developed and evaluated at IIRR and the number of promising lines with zinc ≥35 mg/kg.

Cross	No. of RILs	Zinc content (mg/kg)				No. of lines with zinc content ≥35 mg/kg
		P1	P2	Range	Mean	
IR64/BR2655	95	9.1	24	12.4–25.3	18.8	Nil
IR64/Chittimuthyalu	151		25	18.5–44.9	26.9	11
IR64/Jalpriya	141		25	11.8–26.6	18.6	Nil
IR64/Type-3	62		19	16.3–29.9	21.3	Nil
MTU1010/Jalpriya (IIRR)	277	12	25	18.6–48.5	26.7	10
MTU1010/Jalpriya (Mugad)*	245		25	16.7–39.2	22.6	2
MTU1010/Jalpriya (Pusa)*	273		25	12.1–30.0	21.6	Nil
MTU1010/Suraksha (IIRR)	252		22	20.7–44.0	29	18
MTU1010/Suraksha (Mugad)*	169		22	6.8–27.8	17	Nil
PR116/Chittimuthyalu	85	12	25	11.5–31.8	22.4	Nil
RP Bio 226/Jalpriya	241	14	25	9.9–46.5	36.3	174
Swarna/BR 2655	495	21	24	8.2–46.2	27.2	38
Swarna/Chittimuthyalu	185		25	10.1–43.1	24.3	13
Swarna/Jalpriya	26		25	17.7–27.4	22.2	Nil
Swarna/Ranbir Basmati	202		18	9.6–45.9	28.3	20
Swarna/Type-3	98		19	14.4–29.5	20.9	Nil
Total	3,054					325

*P1, Maternal parent; P2, Pollen parent. *Evaluated at the Agriculture Research Station (ARS), Mugad and Rajendra Agricultural University (RAU), Pusa.*

TABLE 4 | Zinc content in brown rice of mapping populations and selected lines of mapping populations of various national institutions analyzed at the XRF facility of IIRR and the number of promising lines with zinc ≥35 mg/kg.

Mapping populations (MP)/Selected lines (SL)	No. of lines	Zinc content (mg/kg)		No. of lines with zinc content ≥35 mg/kg
		Range	Mean	
NS-MP	250	10.7–37.0	20.0	1
NSA-SL	30	21.3–46.5	36.3	18
NSB-MP	178	13.8–33.6	21.5	Nil
NSZ-MP	376	11.6–30.1	19.4	Nil
AMB-SL	55	12.9–31.7	19.4	Nil
IR681444/BAS-1-SL	5	13.7–31.7	19.8	Nil
SS-SL	37	11.3–31.6	18.7	Nil
NSJ, NSK, NSM, NSS, NST-MP	857	8.1–27.3	11.5	Nil
Abhaya/Moro, Abhaya/Swarna, Abhaya/TN1-SL	36	11.9–21.4	15.8	Nil
Cure-1/IR681444, Cure-10/MTU1010, Cure-2/IR681444, Cure-3/IR681444, Cure-4/IR681444, Cure-7/IR681444-SL	62	10.8–27.1	18.4	Nil
CHIR, GC, HP, IR68144/Moro-MP	317	8.3–28.8	15.7	Nil
IR681444/Kranti, IR 681444/Moro, IR681444/Abhaya, IR681444/BAS-8, IR681444/HMT, IR681444/IR64, IR681444/Kranti, IR681444/Mahamaya, IR681444/MTU1010, IR681444/Shyamala, IR681444/Swarna	73	13.8–24.5	17.7	Nil
BAS-1/IR681444, Kranthi/IR681444, Mahamaya/IR681444, Moro/IR681444, Safari-17/IR68144	25	13.5–26.9	18.8	Nil
Swarna/Bas-1, Swarna/Danteswari, Swarna/HMT, Swarna/IR64, Swarna/IR68144, Swarna/IR681444, Swarna/Moro, Swarna/Moro, Swarna/MTU1010	51	11.6–27.3	16.2	Nil
Total	2,352	NA	NA	19

≥35 mg/kg were observed in 16 mapping populations studied (**Table 3**). Among 2,352 lines of mapping populations from other research groups of various national institutions analyzed using XRF facility at IIRR, 24 lines (1%) showed ≥35 mg/kg zinc content (**Table 4**). Thus, the development of rice breeding lines with target zinc content appears to be feasible from the recombination and selection, although the frequency of RILs with high zinc content appears to be less.

Multilocation studies of two mapping populations, *viz*., MTU1010/Suraksha and MTU1010/Jalpriya, in two to three locations indicated that the zinc content to be highly influenced by environment. Earlier studies also reported significant environmental variance for zinc in rice germplasm, transgenics developed for high zinc content, and mapping populations (29–31).

Significant negative correlation ($P < 0.001$) was observed between single plant yield and zinc content in Swarna/Chittimuthyalu and IR64/Chittimuthyalu RILs, whereas no correlation was observed in MTU1010/Suraksha and MTU1010/Jalpriya (**Figure 3**). Overall, lines with highest zinc content showed poor yield and vice versa. While most studies report negative associations between grain zinc content and yield, a few contradicting results of positive associations were also reported (32). The dilution effect of decrease of nutrient concentrations in plant tissues with the dry matter increase is generally observed in cereals, thus explaining the inverse relationship of zinc content and yield (33). Although negative correlations are observed when the associations were studied for complete set of mapping populations, individual recombinants of high yield and zinc were also obtained. In addition to higher zinc content, the biofortified rice varieties should have yield on a par with the cultivated varieties. Until now, there is no special price for biofortified rice grains in India; thus, there is no incentive for the farmer for growing the biofortified varieties. Thus, the adoption of the biofortified varieties by the farmers is possible only when the yield is on a par with the existing popular cultivated rice varieties. Cooking quality is also important for the rice varieties for their release and adoption. Thus, for biofortified rice varieties to be released and adopted, they should possess zinc content ≥35 mg/kg without yield penalty and desired cooking quality. Nevertheless, this combination happens rarely in germplasm, and thus, conscious breeding efforts are needed to develop biofortified rice varieties with desirable attributes.

Evaluation of Zinc Biofortified Rice Under AICRIP

With the breeding lines developed for high zinc in polished rice across the country from various projects sponsored by HarvestPlus, DBT, and ICAR, a biofortification breeding trial was initiated by AICRIP during 2013 to evaluate high zinc breeding lines without compromising yield, that is, yield must be on par the respective yield check or greater (IR64 or BPT5204). Up to 2013, the promising lines in the biofortification trial were identified with a modest threshold/target zinc value of 20 mg/kg. The optimum threshold content of zinc in polished rice since 2015 was increased to 24 mg/kg to match the international target set earlier by HarvestPlus. Considering the baseline zinc content as 16.25 mg/kg, target levels were recently revised to

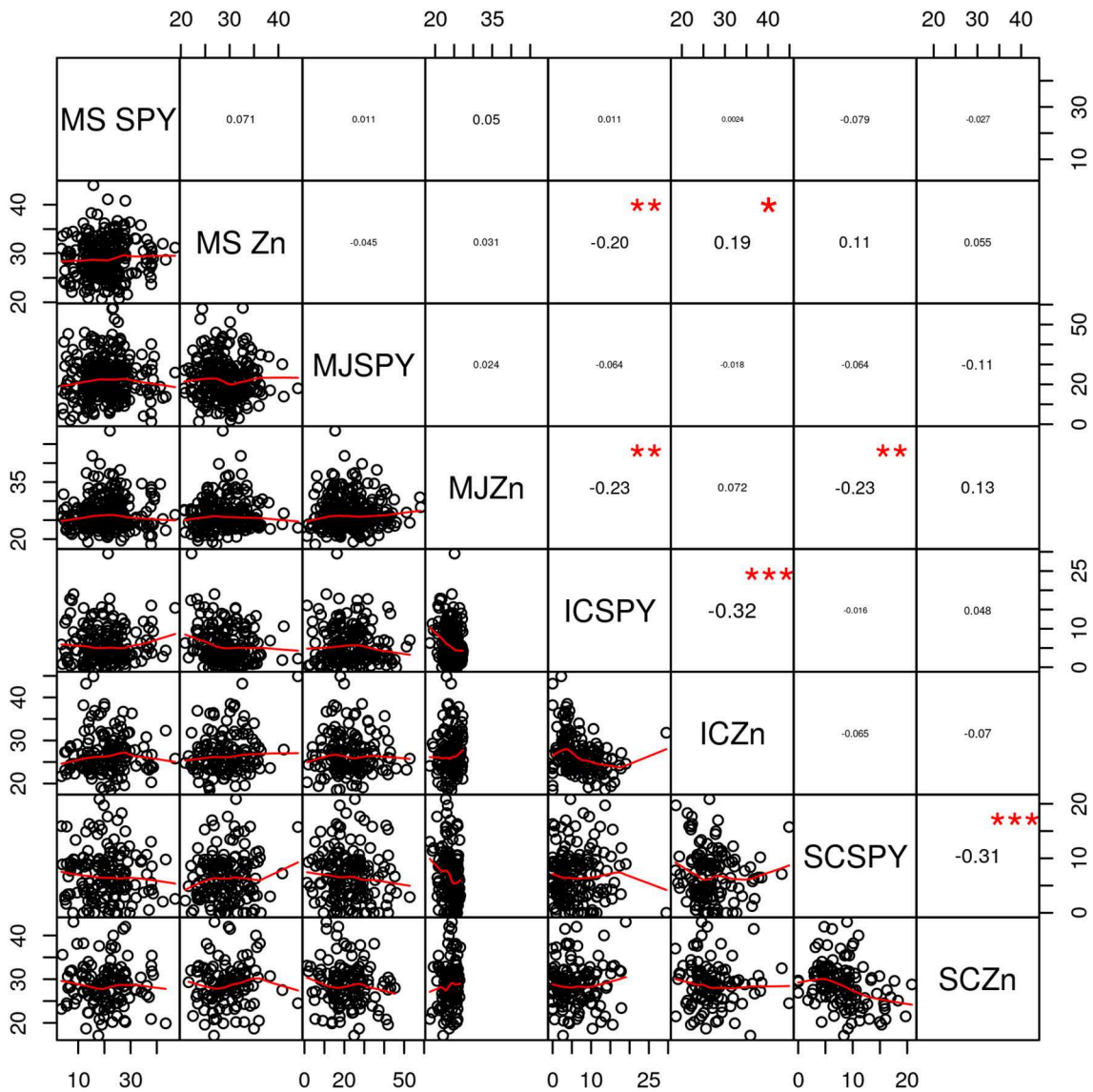

FIGURE 3 | Correlation between zinc content in brown rice and single plant yield (SPY) in RILs. MS, MTU1010/Suraksha; MJ, MTU1010/Jalpriya; IC, IR64/Chittimuthyalu; SC, Swarna/Chittimuthyalu. *, **, and *** significant at the 0.05, 0.01, and 0.001 level.

28 mg/kg by HarvestPlus based on the daily requirement of the women, the consumption of rice along with absorption and retention of zinc in the body (15). However, the threshold level of 28 mg/kg still needs to be adopted by AICRIP. Since 2013, the biofortification trial has completed 5 years by April 2018 with five IVTs, four AVT 1s, and three AVT 2s (**Table 5**). A total of 170 entries were nominated at IVT level. Among these, 83 entries contain threshold zinc content ≥24 mg/kg, and 40 entries were found to contain desirable zinc and yield and exhibited desirable cooking quality. The 170 nominated lines represented several grain types, viz., short bold, short slender, long slender, long bold, medium slender, and medium bold and scented. From IVT, 32.3% of the test entries were promoted to AVT 1 and from AVT 1, 50.9% of the entries were promoted to AVT 2. As the AICRIP locations were categorized into seven zones, performance of each test entry is checked state-wise, zone-wise, and at country level every year. Thus, of the 170 breeding lines tested for 3 consecutive years across 5 years for their consistency in zinc content, yield, and cooking quality, only four promising test entries were recommended for their release based on their performance in one or more than one state or zones or more than one zone for 2 years to the Central Variety Release Committee (CVRC), India (**Table 6**, **Supplementary Table 1**). The promising lines in the state/s for two test entries were recommended to the State Variety Release Committee.

TABLE 5 | Overview of performance of test entries in AICRIP biofortification trials of 2014–2018.

References	Trial	No. of test entries	No. of test locations	No. of entries ≥threshold zinc value*	Entries with yield < yield checks† and poor grain quality‡	Entries with yield > yield check and acceptable grain quality	IET no. of released entries
Directorate of Rice Research (DRR) (34)	IVT	26	17	12	23,814, 23,191, 23,826, 23,825, 22,624, 23,824, 23,833, and 23,834, and 23,829	23,830, 23,831, and 23,832	Nil
Indian Institute of Rice Research (IIRR) (35)	AVT 1	6	26	6	23,830, 23,831, and 23,834	23,824, 23,832, and 23,829	Nil
	IVT	41		31	23,814 (R), 24,754, 24,788, 24,768, 24,784, 24,762, 24,786, 23,826, 24,757, 24,763, 23,191 (R), 24,776, 24,756, 24,764, 24,761, 24,769, 24,755, 24,751, 24,753, 24,773	24,774, 24,766, 24,777, 24,772, and 24,783, 24,760, 24,779, 24,787, 24,771, 24,775, and 24,765	
Indian Institute of Rice Research (IIRR) (36)	AVT 2	5	22	4	23,834	23,824, 23,829, and 23,832	23,832
	AVT 1	24	19	12	Nil	24,438, 24,440, 24,779, 24,555, and 24,557, 24,760, 24,774, 24,775, 24,783, 24,360, 24,772, and 24,336	
	IVT	40		14	25,472, 25,462, and 25,466	25,475, 25,446, 25,461, 25,477, and 25,450, 25,458, 25,467, and 25,451, 25,465, 25,443, and 25,470	
Indian Institute of Rice Research (IIRR) (37)	AVT 2	12	23	8	24,779, 24,440, 24,438, 24,760, and 24,360	24,336, 24,555, 24,774, and 24,557	24,555 24,557 24,760
	AVT 1	13	24	12	25,476, 25,450, and 25,465	25,443, 25,446, 25,451, 25,458, 25,461, 25,467, 25,470, 25,475, and 25,477	
	IVT	32	21	12	Nil	25,253, 26,398, 26,385, 26,373, 26,376, 26,386, 26,396, 26,383, 26,375, 26,378, 26,380, and 26,377	
Indian Institute of Rice Research (IIRR) (38)	AVT 2	11	31	3	25,477, 25,443, and 25,475	Nil	NIL
	AVT 1	12		4	26,373, and 26,376	26,383, and 26,375	
	IVT	31	23	14	27,157, 27,158, and 27,169, 27,159, 27,155, 27,171, 27,172, 27,176, 27,177, 27,180, and 27,183	27,170, 27,169, and 27,179	

Threshold zinc value of 20 mg/kg was set during 2013. From 2015, the optimum threshold/target content of zinc in polished rice was increased to 24 mg/kg. The same threshold values were maintained from IVT to AVT 2.

†Yield checks: IR64/BPT5204/DRRH3.

‡Acceptable grain quality—HRR% ≥60 with intermediate amylose content/high amylose with soft gel consistency.

Effect of Environment on Grain Zinc Content of Rice

Although biofortified rice varieties have high grain zinc, the zinc content in the grains is highly variable, depending on the soils, seasons, agronomic practices according to AICRIP, and reported studies in India. The zinc content of two micronutrient check genotypes varied within and across locations and even within the locations across the experimental plots where IVT, AVT 1, and AVT 2 were conducted. The minimum ranges of 1.0 to 17.0 mg/kg for Kalanamak and 1.0 to 21.4 mg/kg for Chittimuthyalu were observed across 27 locations for 5 years of AICRIP studies. And maximum ranges of 9.5 to 21.4 mg/kg for Kalanamak and 12.4 to 36.3 mg/kg for Chittimuthyalu were noted (**Supplementary Table 2**). The grain zinc content of the four released varieties was also found to be variable with locations, as shown in **Figure 4**. Therefore, it is essential to identify suitable locations with favorable soils for large-scale seed production of biofortified varieties; otherwise, the potential of biofortified rice variety might not be realized because of zinc deficiency in the soil. Several countries including India, Pakistan, China, Iran, and Turkey come under zinc-deficient soils, and availability of zinc will further decrease if the soil pH is alkaline and if there is less organic matter and less moisture (3). And HYVs also take up zinc from the soil each cropping season, making the soil further deficient in zinc. Therefore, external application of zinc is essential to realize and maintain the nutrient performance potential of a biofortified rice variety.

Biofortified Rice Varieties and Their Anticipated Impact

Considering the zinc content of the popular varieties of rice as 13 mg/kg and RDA of zinc as 12 mg/kg for male and 10 mg/kg for

TABLE 6 | Details of high-zinc rice varieties evaluated under AICRIP released through CVRC.

IET no. and name	Year of release	Developer Institution	Pedigree	Grain type	Zinc content in mg/kg*	States
23832 DRR Dhan 45	2016	ICAR-IIRR	IR 7307-45-3-2-3/IR 77080-B-34-3	Long slender	22.3	Andhra Pradesh, Telangana, Karnataka and Tamil Nadu
24555 DRR Dhan 48	2017	ICAR-IIRR	RP Bio 226*1/CSR 27	Medium slender	20.91	Uttar Pradesh, West Bengal, Kerala and Punjab
24557 DRR Dhan 49	2017	ICAR-IIRR	RP Bio 226*1/CSR 27	Medium slender	26.13	Gujarat and Kerala
24760 Surabhi	2017	Nuziveedu Seeds Limited	PRN-19045/PRN-14	Medium slender	22.84	Maharashtra and Gujarat

Threshold zinc value of 20 mg/kg was set during 2013. From 2015, the optimum threshold/target content of zinc in polished rice was increased to 24 mg/kg. The same threshold values were maintained from IVT to AVT 2 for test entries.

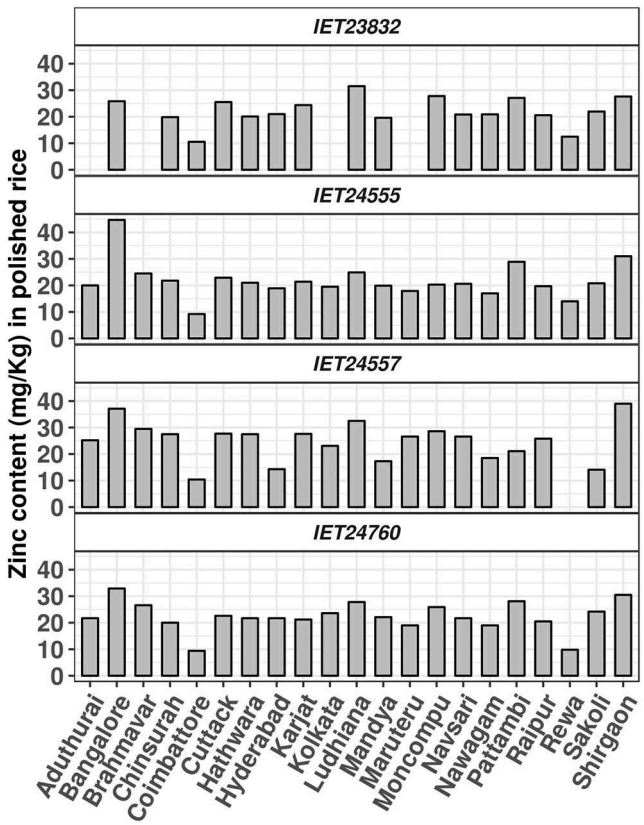

FIGURE 4 | Zinc content mg/kg of zinc-biofortified varieties, released through CVRC evaluated across 21 locations under AICRIP Biofortification trial.

female (39), the zinc intake through rice consumption is 2.9 g/day per person. With the consumption of rice of existing popular varieties, the % RDA of zinc intake accounted for 24% for males and 29% for females. On the assumption that the consumption of biofortified varieties with higher zinc in polished rice will result in higher zinc intake, the effect of zinc biofortification through higher zinc intake and the extent of the improved intake to meet the RDA of zinc because of biofortified varieties in comparison to popular varieties are summarized in **Table 7**. With zinc-biofortified varieties, the zinc intake could account for 38 to 47% of the RDA for male and 46 to 57% of the RDA for female. The technology efficacy of zinc biofortification ranged from 48 to 67% for male and 55 to 75% for female.

With rice as the sole diet component, in order to meet the RDA of zinc, polished grains should have 54.5 to 68.2 mg/kg zinc (without including bioavailability status) (40). Transgenic rice lines with soybean ferritin gene were developed for enhanced iron content in polished rice grains, but the same lines also accumulated higher amount of zinc ranging from 34.9 to 55.5 mg/kg (41). NASFer-274 event of IR64 transgenic rice with nicotianamine synthase (OsNAS2) and soybean ferritin (SferH-1) genes showed 45.7 mg/kg zinc content in polished rice without yield penalty or altered grain quality (31). Although zinc content of transgenic rice lines can meet the RDA of zinc, the transgenic rice lines are still to be accepted and adopted in India.

Our learning experiences since 2013 from the studies on characterization of germplasm for zinc content, development of national evaluation system for the release of biofortified rice varieties, and evaluation of biofortified rice breeding lines have been consolidated with reference to the five criteria set by the biofortification program (14).

As per criterion 1 (crop productivity, i.e., yield must be maintained or increased to ensure farmer acceptance), the recombination of zinc content and yield appears to be feasible as evidenced from the nominations made by several national research groups to biofortification trials and the release of four biofortified rice varieties through AICRIP.

Criterion 2 (impact of the enhanced micronutrient level on human health) has been indirectly confirmed through *in vitro Caco-2* cell model studies of biofortified rice varieties. Because *in vivo* human clinical tests are expensive, labor-intensive, and subject-dependent response with, results of *in vitro* bioavailability studies can be explored as a possible alternative as proof of concept for zinc-biofortified rice varieties popularization. Bioavailability of biofortified rice variety with higher zinc in

TABLE 7 | Technology efficacy and potential impact of biofortified zinc varieties calculated based on the zinc content of the popular varieties of rice as 13 mg/kg as baseline; RDA of zinc as 12 mg/kg for male and 10 mg/kg for female; zinc intake through rice consumption as 2.9 g/day per person and considering per-capita consumption of rice in India as 220 g/day.

Name of the biofortified rice variety	Additional zinc over baseline zinc content	Gender	RDA with popular varieties (%)	Additional zinc with Biofortification (g/d)	RDA (%)	Efficacy (%)
DRR Dhan 45	9.3	Male	24	4.9	41	54
		Female	29		49	61
DRR Dhan 48	7.9	Male	24	4.6	38	48
		Female	29		46	55
DRR Dhan 49	13.1	Male	24	5.7	47	67
		Female	29		57	75
Surabhi	9.8	Male	24	5.0	42	56
		Female	29		50	63

polished rice showed higher absorbed zinc through *in vitro Caco-2* cell model and suckling rat pups studies (42). Preliminary studies of Indian Council of Medical Research–National Institute of Nutrition (ICMR-NIN) Hyderabad demonstrated that the zinc content and bioavailability from DRR Dhan 45 rice variety were almost twice those of the control IR64 (43). The clinical trial of one of the released biofortified rice varieties is under consideration by the Government of India.

According to criterion 3 (the enhanced micronutrient trait must be relatively stable across various edaphic environments and climatic zones), only the consistent genotypes across the states/zones and years are being released as biofortified zinc varieties under AICRIP in India. Multilocation data of AICRIP or independent research groups (31) indicated that the zinc content across the locations is highly variable and dependent on soil factors (**Figure 4**). Considering the wide diversity of agroclimatic regions of India, for the identification of the promising lines and their release as varieties, the data collected under AICRIP are being analyzed as performance state and zone-wise for the identification of stable varieties.

Criterion 4 addresses the bioavailability of micronutrients and is being carried out at ICMR-NIN. The coupled *in vitro* digestion/*Caco-2* cell model studies with extrinsic ^{65}Zn isotopic labeling demonstrated higher absorption of zinc from DRR Dhan 45 rice in intestinal cells. The proximate composition of DRR Dhan 45 and IR64 varieties remained similar, except that DRR Dhan 45 rice had a little higher protein. Detailed studies of zinc absorption of rice genotypes with differential zinc using coupled *in vitro* digestion/*Caco-2* cell model studies are in progress. The overall food composition dictates the fate of intestinal absorption of zinc in diet. For zinc, phytate, fiber (cellulose, hemicellulose, lignin), oxalic acid, lectins, and heavy metals act as antinutrients, whereas organic acids (ascorbate, fumarate, malate, and citrate), long-chain fatty acids, β-carotene, and some of the amino acids act as pronutrients (44). As per criterion 5, consumer acceptance by taste and cooking quality is also taken care through a set of approved laboratory experiments in AICRIP coupled with panel test at IIRR.

CONCLUSIONS

Zinc deficiency, especially in the developing countries with rice as a major staple food, is being addressed by the development of biofortified rice varieties with high zinc in polished rice with the national and international support in India. Germplasm screening for their zinc content in brown and polished rice using ED-XRF revealed wide genetic variability, and evaluation of mapping populations indicated the possibility of favorable recombination of high zinc content and yield. Although only 2.4% biofortified varieties were released out of the total test entries until 2017, the rigorous screening through AICRIP biofortification indicated the achievability of combining target zinc content, yield, and quality parameters. A preliminary study of bioavailability of a released biofortified rice variety was found to be twice that of control variety IR64. The technology efficacy and the enhancement of RDA calculated with the four released varieties were found to be promising. Thus, the benefits of biofortification research are being translated to the society as biofortified rice varieties with high zinc, hoping for their adoption and impact in achieving the nutritional security of the needy.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

DS: analyses and writing of the manuscript. CN: planning and writing of the manuscript. PM, PS, and RK: evaluation of mapping populations. BN: economic impact of biofortification. KSum: writing of the manuscript. LR: AICRIP analyses. KSur: editing of the manuscript. PR and TL: bioavailability studies. VB and SV: Coordination of the study.

FUNDING

This work was supported by ICAR funded project ICAR plan CRP Biofortification, India.

ACKNOWLEDGMENTS

The authors acknowledge to ICAR-CRP-Biofortification project, ICAR-Indian Institute of Rice Research, Hyderabad for financial support and providing facilities and HarvestPlus for providing ED-XRF. Authors acknowledge AICRIP for the Biofortification trial reports and cooperating centers.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2020.00026/full#supplementary-material

REFERENCES

1. Timsina J, Connor DJ. Productivity and management of rice-wheat cropping systems: issues and challenges. *F Crop Res*. (2001) 69:93–132. doi: 10.1016/S0378-4290(00)00143-X
2. United States Department of Agriculture (USDA). *World Agricultural Production*. (2019) Available online at: https://apps.fas.usda.gov/psdonline/circulars/production.pdf (accessed July 2019).
3. Cakmak I. Enrichment of cereal grains with zinc: Agronomic or genetic biofortification? *Plant Soil*. (2008) 302:1–17. doi: 10.1007/s11104-007-9466-3
4. Hotz C, Brown KH. Contents international zinc nutrition consultative group (IZiNCG) technical document. *Food Nutr Bull*. (2004) 25:S94–200.
5. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, et al. Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet*. (2008) 371:243–60. doi: 10.1016/S0140-6736(07)61690-0
6. Tan Z, Ma G, Lin L, Liu C, Liu Y, Jiang J, et al. Prevalence of subclinical vitamin A deficiency and its affecting factors in 8,669 children's of China. *Zhonghua Yu Fang Yi XueZaZhi*. (2002) 6:161–3.
7. Prasad AS. Metabolism of zinc and its deficiency in human subjects. In: Prasad AS, Thomas CC, editors. *Zinc Metabolism*. Springfield, IL (1966). p. 250–302.
8. Paul W. Iron and zinc interactions in humans. *Am J Clic Nutr*. (1998) 68:442S—6S. doi: 10.1093/ajcn/68.2.442S
9. Nirmala B, Babu VR, Neeraja CN, Waris A, Muthuraman P, Rao DS. Linking agriculture and nutrition: an ex-ante analysis of zinc biofortification of rice in india. *Agric Econ Res Rev*. (2016) 29:171. doi: 10.5958/0974-0279.2016.00044.6
10. Sally Brooks. Is international agricultural research a global public good? The case of rice biofortification. *J Peasant Stud*. (2011). 38:67–80. doi: 10.1080/03066150.2010.538581
11. HarvestPlus. *Breeding Crops for Better Nutrition: Harnessing Agricultural Technology to Improve Micronutrient Deficiencies*. Washington, DC: International Food Policy Research Institute. (2004).
12. Bouis HE, Graham RD, Welch RM. The consultative group on international agricultural research (CGIAR) micronutrients project: justification and objectives. *Food Nutr Bull*. (2000) 21:374–81. doi: 10.1177/156482650002100406
13. Gregorio GB. Progress in breeding for trace minerals in staple crops. Symposium: plant breeding: a new tool for fighting micronutrient malnutrition. *J. Nutr*. (2002) 132:500S—2S. doi: 10.1093/jn/132.3.500S
14. Welch RM, Graham RD. Breeding for micronutrients in staple food crops from a human nutrition perspective. *J Exp Bot*. (2004) 55:353–64. doi: 10.1093/jxb/erh064
15. Bouis HE, Saltzman A. Improving nutrition through biofortification: a review of evidence from harvestplus, 2003 through (2016). *Glob Food Sec*. (2017) 12:49–58. doi: 10.1016/j.gfs.2017.01.009
16. Babu VR, Neeraja CN, SanjeevaRao D, Sundaram RM, Longvah T, Usharani G, et al. *Bioforotification in Rice*. DRR Technical Bulletin No. 81/2014. Hyderabad: Directorate of Rice Research. (2014) pp. 86.
17. Wailes E, Chavez E. International rice baseline with deterministic and stochastic projections, 2012-2021. *Staff Pap*. (2012)
18. Wessells KR, Brown KH. Estimating the global prevalence of zinc deficiency: results based on zinc availability in national food supplies and the prevalence of stunting. *PLoS ONE*. (2012) 7:e50568. doi: 10.1371/journal.pone.0050568
19. Chandel G, Banerjee S, See S, Meena R, Sharma DJ, Verulkar SB. Effects of different nitrogen fertilizer levels and native soil properties on rice grain Fe, Zn and protein contents. *Rice Sci*. (2010) 17:213–27. doi: 10.1016/S1672-6308(09)60020-2
20. Anandan A, Rajiv G, Eswaran R, Prakash M. Genotypic variation and relationships between quality traits and trace elements in traditional and improved rice (*Oryza sativa*L.) genotypes. *J Food Sci*. (2011) 76:H122–30. doi: 10.1111/j.1750-3841.2011.02135.x
21. Nachimuthu VV, Robin S, Sudhakar D, Raveendran M, Rajeswari S, Manonmani S. Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. *Indian J Sci Technol*. (2014) 7:1555–62.
22. Jeng TL, Lin YW, Wang CS, Sung JM. Comparison and selection of rice mutents with high and zinc contents in their polished grains that were mutated from the *indica* type cultiver IR64. *J Food Compost Anal*. (2012) 28:149–54. doi: 10.1016/j.jfca.2012.08.008
23. Sellappan K, Datta K, Parkhi V, Datta SK. Rice caryopsis structure in relation to distribution of micronutrients (iron, zinc, β-carotene) of rice cultivars including transgenic indica rice. *Plant Sci*. (2009) 177:557–62. doi: 10.1016/j.plantsci.2009.07.004
24. Zazoli MA, Shokrzadeh M, Bazerafshan E, Hazrati M, Tavakkouli A. Investigation of zinc content in Iranian rice *(Oryza sativa*) and its weekly intake. *Am Eurasian J Agric Environ Sci*. (2006) 1:156–9.
25. Thongbam PD, Raychaudhury M, Durai A, Das SP, Ramesh T, Ramya KT et al. Studies on grain and food quality traits of some indigenous rice cultivars of north-eastern hill region of India. *J Agric Sci*. (2011) 4:259–70. doi: 10.5539/jas.v4n3p259
26. Jeomho L, Kyuseong L, Hunggoo H, Changihn Y, Sangbok L, Younghuran C, et al. Evaluation of iron and zinc content in rice germplasm. *Korean J Breed Sci*. (2008) 40:101–5.
27. Franz KB, Kennedy BM, Fellers DA. Relative bioavailability of zinc from selected cereals and legumes using rat growth. *J Nutr*. (1980) 110:2272–83. doi: 10.1093/jn/110.11.2272
28. Gregorio GB, Senadhira D, Htut H, Graham RD. Breeding for trace mineral density in rice. *Food Nutr Bull*. (2000) 21:382–6. doi: 10.1177/156482650002100407
29. Huang Y, Tong C, Xu F, Chen Y, Zhang C, Bao J. Variation in mineral elements in grains of 20 brown rice accessions in two environments. *Food Chem*. (2016) 192:873–8. doi: 10.1016/j.foodchem.2015.07.087
30. Hu BL, Huang DR, Xiao YQ, Fan YY, Chen DZ, Zhuang JY. Mapping QTLs for mineral element contents in brown and milled rice using an *Oryza sativa* × *O. rufipogon* backcross inbred line population. *Cereal Res Commun*. (2016) 44:57–68. doi: 10.1556/0806.43.2015.044
31. Trijatmiko KR, Duenãs C, Tsakirpaloglou N, Torrizo L, Arines FM, Adeva C, et al. Biofortified indica rice attains iron and zinc nutrition dietary targets in the field. *Sci Rep*. (2016) 6:19792. doi: 10.1038/srep19792
32. Swamy BPM, Rahman MA, Inabangan-Asilo MA, Amparado A, Manito C, Chadha-Mohanty P, et al. Advances in breeding for high grain zinc in rice. *Rice*. (2016) 9:49. doi: 10.1186/s12284-016-0122-5
33. Mcdonald G, Genc Y. A simple method to evaluate genetic variation in grain zinc concentration by correcting for differences in grain yield. *Plant Soil*. (2008) 306:49–55. doi: 10.1007/s11104-008-9555-y

34. Directorate of Rice Research (DRR). *Progress Report 2013. Varietal Improvement. All India Coordinated Rice Improvement Programme (ICAR).* Vol. 1. Hyderabad: Directorate of Rice Research (2014).
35. Indian Institute of Rice Research (ICAR). *Progress Report 2014. Varietal Improvement All India Coordinated Rice Improvement Programme*. Vol. 1. Hyderabad: ICAR- Indian Institute of Rice Research (formerly Directorate of Rice Research) (2015).
36. Indian Institute of Rice Research (ICAR). *Progress Report 2015. Varietal Improvement All India Coordinated Rice Improvement Programme*. Vol. 1. Hyderabad: ICAR- Indian Institute of Rice Research (formerly Directorate of Rice Research) (2016).
37. Indian Institute of Rice Research (ICAR). *Progress Report 2016. Varietal Improvement All India Coordinated Rice Improvement Programme*. Vol. 1. Hyderabad: ICAR- Indian Institute of Rice Research (formerly Directorate of Rice Research) (2017). Available online at: http://www.icar-iirr.org/AICRIP/Volume_1Progress%20Report%20Revised-2016.pdf (accessed July 2019).
38. Indian Institute of Rice Research (ICAR). *Progress Report-2017, Varietal Improvement All India Coordinated Rice Improvement Programme*. Vol. 1. Hyderabad: ICAR- Indian Institute of Rice Research (formerly Directorate of Rice Research) (2018). Available online at: http://www.icariirr.org/AICRIP/Varietal%20Improvement%20Progress%20Report%202017.pdf (accessed July 2019).
39. National Institute of Nutrition (NIN). *Nutritional Intake in India*. NSS Report No. 560 (2011). Available online at: http://www.indiaenvironmentportal.org.in/files/file/nutritional%20intake%20in%20India%202011-12.pdf
40. Sanjeeva Rao D, Babu PM, Swarnalatha P, Kota S, Bhadana VP, Surekha K, et al. Assessment of grain zinc and iron variability in rice germplasm using energy dispersive X-ray fluorescence spectrophotometer (ED - XRF). *J Rice Res*. (2014) 7:45–52.
41. Vasconcelos M, Datta K, Oliva N, Khalekuzzaman M, Torrizo L, Krishnan S, et al. Enhanced iron and zinc accumulation in transgenic rice with the ferritin gene. *Plant Sci*. (2003) 164:371–8. doi: 10.1016/S0168-9452(02)00421-1
42. Jou MY, Du X, Hotz C, Lönnerdal B. Biofortification of rice with zinc: assessment of the relative bioavailability of zinc in a caco-2 cell model and suckling rat pups. *J Agric Food Chem*. (2012) 60:3650–7. doi: 10.1021/jf202338t
43. Neeraja CN Voleti SR. *Consortia Research Plotform (CRP) Biofortification in Selected Crops for Nutritional Security-Status.*, Vol. 104. Technical bulletin. Hyderabad: ICAR-Indian Institute of Rice Research (2019) . p. 1–84.
44. Graham RD, Welch RM, Bouis HE. Addressing micronutrient malnutrition through enhancing the nutritional quality of staple foods. *Adv Agron*. (2001) 70:77–142. doi: 10.1016/s0065-2113(01)70004-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sanjeeva Rao, Neeraja, Madhu Babu, Nirmala, Suman, Rao, Surekha, Raghu, Longvah, Surendra, Kumar, Babu and Voleti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Nutrition

ORIGINAL RESEARCH
published: 15 April 2020
doi: 10.3389/fnut.2020.00045

Profiling of Nutraceuticals and Proximates in Peanut Genotypes Differing for Seed Coat Color and Seed Size

Spurthi N. Nayak[1], Viresh Hebbal[1], Pushpa Bharati[2], Hajisab L. Nadaf[3], Gopalkrishna K. Naidu[3] and Ramesh S. Bhat[1]*

[1] Department of Biotechnology, University of Agricultural Sciences, Dharwad, India, [2] Department of Food Science and Nutrition, University of Agricultural Sciences, Dharwad, India, [3] Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad, India

A total of 60 genotypes of peanut comprising 46 genotypes selected from ICRISAT mini core collection and 14 elite cultivars with differing kernel color and size were used to profile the nutritional parameters such as proximates (moisture, fat, ash, crude protein, crude fiber, carbohydrate content) and nutraceuticals (total polyphenol content and total antioxidant activity). The genotypes showed varied kernel color ranging from white to purple. Kernel skin color was quantified using colorimetry, and the color parameters were expressed as CIELAB color parameters. In total, nine morphological traits, six yield related traits, eight nutritional traits and eleven color parameters were observed across 60 genotypes. The sixty genotypes were grouped into ten clusters based on the color strength. Among them, Cluster-III with dark red seeds had the maximum fat content and total polyphenol content (TPC). Cluster-VI with light pink colored seeds had high antioxidant activity (AOA) and Cluster-X with white colored seeds had highest moisture and crude protein content. Color strength (K/S) was found to be positively correlated with TPC. Another color parameter, redness/greenness (a*) was found to be positively correlated with AOA. However, seed size was positively correlated with the crude protein content, but not with any other nutritional traits under study. The population studies based on the genotypic data indicated two distinct groups pertaining to botanical types of peanut. The marker-trait association (MTA) using single marker analysis indicated 75 major MTAs for most of the nutritional traits except for moisture content. The markers associated with nutritional parameters and other important yield related traits can further be utilized for genomics-assisted breeding for nutrient-rich peanuts.

Keywords: groundnut, nutrient profile, polyphenol, antioxidants, proximates, peanut skin color, AhTE markers, marker-trait association

OPEN ACCESS

Edited by:
Kannan R. R. Rengasamy, Konkuk University, South Korea

Reviewed by:
Ren-You Gan, Institute of Urban Agriculture (CAAS), China
Sachiko Narita Isobe, Kazusa DNA Research Institute, Japan

****Correspondence:***
Spurthi N. Nayak
nayaksn@uasd.in

Specialty section:
This article was submitted to Nutrition and Food Science Technology, a section of the journal Frontiers in Nutrition

Received: *07 August 2019*
Accepted: *23 March 2020*
Published: *15 April 2020*

Citation:
Nayak SN, Hebbal V, Bharati P, Nadaf HL, Naidu GK and Bhat RS (2020) Profiling of Nutraceuticals and Proximates in Peanut Genotypes Differing for Seed Coat Color and Seed Size. Front. Nutr. 7:45. doi: 10.3389/fnut.2020.00045

INTRODUCTION

Peanut (*Arachis hypogaea* L.) widely known as groundnut is a cultivated allotetraploid (2n = 4x = 40), particularly valued for its protein content (28%), oil content (50%) and is rich source of calcium, phosphorus, potassium, magnesium, iron, zinc, and boron. The peanuts also contain vitamin E and small amounts of vitamin B complex (1). Due to these properties, it is rightly

called as "king of oilseeds" and "poor man's almond" and provides 567 kcal of energy per 100 g of kernels. Peanut is derived from a natural hybridization process between *A. duranensis* (AA genome progenitor) and *A. ipaensis* (BB genome progenitor) resulted in amphidiploid *A. monticola*. Later in the evolutionary process fertile, now cultivated allotetraploid *A. hypogaea* was evolved. The genome sequence of cultivated peanut revealed a genome size of 2.7 GB (2, 3). The global production of peanut is estimated to be 47.09 million tons from 27.94 million hectare area with productivity of 1685.6 kg ha^{-1} (4). More than 50% of the peanut produced are used for oil extraction. Rest are either consumed as kernels or processed into various products like peanut butter, milk, cheese analogs, beverages, plumpy nut (a ready to use therapeutic food), and chocolate additives. Peanut is considered as both legume (botanically) and nut (nutritionally). As a plant based protein, it also contains wholesome amount of fiber and other bioactive functional compounds that improve health in humans including heart health, weight loss, gallstone prevention, malnutrition, etc. (5–9). For instance, "Plumpy nuts," a ready-to-use therapeutic food (RUFT) is being used for the treatment of severe acute malnutrition (SAM) which affect adversely about 20 million children globally, with approximately 8 million cases in India alone. RUTF use is very common in Africa and other developing countries where the incidence of SAM is high. Peanut has several functional components with variety of health benefits like coenzyme Q10 which helps the heart protection under low oxygen areas like high altitude and clogged arteries, rich arginine helps in blood circulation, resveratrol protects against cancer, cognitive disorder and Alzheimer's disease, phytosterols of peanut butter, oil reduces the absorption of cholesterol from the diet, magnesium and dietary fiber reduce the diabetes, reduces LDL cholesterol and helps in weight management [reviewed by (5)]. The foods with different colors including fruits, vegetables, legumes and cereals are found to have functional components that are beneficial for the human health (10). Peanuts have different tinge of kernel colors varying from the white to dark purple with solid or variegated color patterns. The peanut germplasm includes peanut cultivars with varying kernel colors. For selection of appropriate lines that retain the diversity persisting in the global collection becomes essential in utilizing in crop breeding programs. In this regard, the "core collection" (i.e., 10% of the entire germplasm collection) presents a manageable and cost-effective entry point into germplasm collections for identifying candidate genotypes for various traits (11). In peanut, the core collection were further narrowed down to mini core collections and U. S. peanut mini core collection (12), ICRISAT mini core collection (13), and Chinese peanut mini core collection (14). Traits related to abiotic stresses (drought, heat, salinity, low temperature, and P deficiency, calcium induced iron chlorosis), biotic stresses (early leaf spot, late leaf spot, rust, tobacco cut worm, *Aspergillus flavus*, peanut bud necrosis disease, and bacterial wilt) and seed quality (oil, protein, oleic/linoleic ratio, Fe and Zn) have been screened by many researchers in ICRISAT mini core collection (15–17). GWAS for 50 different agronomic traits was carried out in reference set of core collection having 300 genotypes which included the 184 mini core genotypes indicated several markers associated with major agronomic traits (18). The efforts toward elucidating the marker-trait association related to quality or nutritional parameters are limited to some crops like sorghum (19), wheat (20), rice, and barley (21). Evidence of metabotype-phenotype linkage was explained by using parallel metabolite and phenotypic GWAS in traits such as grain color and size in rice and maize kernels (22).

Estimation of proximates like moisture content (MC), fat, crude protein (CP), ash, crude fiber (CF), and carbohydrates (CHO) provide the distribution of major components in any food or processed product. The peanut is rich in nutraceuticals like total polyphenol content (TPC) and total antioxidant activity (AOA). Proximates and nutraceuticals content in the kernel determine the quality of peanut. Peanuts must be dried or cured to guarantee that, the moisture content does not surpass 10.5%, to guarantee quality and to avoid the development of microorganisms. Low moisture percentage of peanut seed prevents it from the susceptibility to the aflatoxin producing fungal pathogens like *Aspergillus flavus* (23). Fatty acid composition in peanut is heart-friendly. Oleic acid (monounsaturated omega-9 fatty acid) is important seed quality parameter and has inverse association with systolic blood pressure. Digestibility of peanut protein is similar to that of animal protein (24). Crude fiber in peanut has little food value but provide the proper help in the intestinal tract for adequate peristaltic action. It contains low glycemic index (GI) and glycemic load (GL) and addition of peanut butter to high glycemic load food can stabilize the blood glucose level of the body. Nutraceuticals of peanut gives health advantages, including the aversion or potential treatment for the diseases [reviewed by (5)].

Peanut pod is made up of external shell (21–29%), kernel (69–73%), germ (2–3.5%), and testa or seed coat (2–3%). Due to papery like structure, seed coat is also called as peanut skin or peanut kernel skin (PKS). Several studies showed that PKS are rich in polyphenol and antioxidant [reviewed by (25)]. The kernel skin is removed through blanching or roasting before the preparation of snack food, groundnut butter and other groundnut-based foods and PKS remained as by-product. High fiber and tannin content of PKS make them usable only for ruminants (and possibly rabbits) but, due to chance of aflatoxin contamination, utilization of PKS as feed remains limited with inclusion rates lower than 10% of the diet as feed for ruminants (26). The reason for the variation in skin color was reported to be due to flavonoid content in cereals, legumes, oilseeds, and several other colored foods (27–29). Due to growing evidence of the versatile health benefits of dietary flavonoids through epidemiological studies (30), the PKS emerges as the potential product to be utilized in diet.

Peanut skin color was found to be strongly associated with total phenolic content and hue angle was proposed to be a biomarker for total polyphenol content and antioxidant capacity when 27 cultivars were screened (31). Similarly total anthocyanin content in the peanut skin was found to be associated with the skin color when 22 selected genotypes from US mini core collection along with 4 Israeli Virginia type cultivars were screened (32). The relation between TPC, flavonoids and AOA

was studied with grain color in 481 accessions of rice with white, red and black colored genotypes and correlation studies indicated a negative correlation between a* and antioxidant capacity (33). The grain color was also found to be associated with anthocyanin and Zn content when 156 rice accessions varying in grain color were screened and analyzed through genome-wide association studies (34). Similar findings were revealed from transcriptomics and proteomics studies in colored rice (22). Similarly, GWAS studies have been carried out to dissect the candidate genomic regions for flour color using colorimetric approaches (35).

Different sized peanuts are preferred for different purposes. For instance, peanuts with large seed, low oil but with high oleic acid/linoleic acid (O/L) ratio are preferred for direct consumption, while medium seeded peanuts are preferred by the industries. Studies related to kernel size are limited to inheritance studies (36). A study indicated that kernel size was not significantly correlated with the oil content. Protein content was higher in small seeded compared to the bold seeded peanuts (37). However, in soybean no significant correlation was found between protein and oil content with seed size (38). In peanut the studies on kernel color are limited to the estimation of TPC and flavonoids through biochemical assays and correlation studies thereof. There were limited efforts to identify the genomic regions underlying these traits. In this regard, the molecular markers are considered to be powerful genomic tools to characterize the genetic variation present in the population. The molecular markers can be used in trait mapping and molecular breeding programs [reviewed in (39)]. Low level of genetic variation was observed in peanut. Transposable elements (TE) provide an important source of variation and are highly dynamic in diverse species. Due to the positive correlation of transposable element content with the genome size of the organism, it has been widely known as a genomic parasite and found to be source of variation in plants and animals (40). Use of DNA transposon markers in peanut was proposed by Bhat et al. (41) and presence or absence of *AhMITE1* at a predetermined site in the genome was confirmed using PCR (42). There are two types of transposable elements viz., class I TE (move within genomes via RNA intermediates, using a copy-and-paste mechanism) and class II TE (DNA of a DNA transposon moves by a cut-and-paste mechanism). Class II TEs entail autonomous and non-autonomous elements. Miniature inverted-repeat transposable elements (MITEs) are non-autonomous elements of less than 600 bp in length (43). It contains 10-15 bp terminal inverted repeats and two flanking target site duplications. MITEs are inserted mainly in the gene-rich regions and can affect gene expression (44). In peanut, due to the availability of genome sequence information, several TE markers commonly known as *Arachis hypogaea* transposable element (AhTE) have been identified and are made available in a database at Kazusa Peanut Database (45). AhTE markers have shown greater potential to differentiate the genotypes in groundnut (42, 43, 46). As AhTE markers can be easily screened on agarose gel through electrophoresis, they have been utilized in trait mapping using linkage and association mapping approaches (41–43, 47–51). The candidate genes or genomic regions governing the nutritional traits are presently not available. Hence an effort was made to identify genomic regions that are associated with the nutritional traits including proximates and nutraceuticals like TPC and AOA genotypes that vary in kernel skin color and size.

MATERIALS AND METHODS

Plant Materials and Evaluation of Population

A total of 60 genotypes (46 from ICRISAT mini core collection and 14 elite cultivars) were selected based on variation in kernel skin color varying from white to purple color and size (**Figure 1**). The passport details suggested that they are collected from 25 different countries across the globe (**Table 1**). Sixty genotypes were evaluated for different morphological and yield related traits by growing them in post-rainy season 2017–18 and rainy season of the year 2018 in a randomized block design with two replications. Each replication consisted of 2 rows of 1.5 m length with a spacing of 30 × 10 cm for the bunch type cultivars (53 genotypes) and 60 × 10 cm for the runners (7 genotypes). Three representative plants were selected randomly from each genotype for recording the phenotyping data on taxonomic and productivity traits. Observations on morphological traits (flower color, stem color, leaf color, leaf shape, growth habit, branching pattern, leaflet length and leaflet width, and plant height) and productivity traits (pod weight per plant, number of pods per plant, shelling percentage, test weight, sound mature kernel weight percentage, and haulm weight) were recorded as per the groundnut descriptor (52).

Estimation of Color Parameters

For identifying the color difference, the color space and coordinates were determined using CIE L*, a*, and b* (CIELAB) values for these genotypes using Spectrophotometer (Fluoro Spectrophotometer-SS5100A, Premier Colorscan) available at AICRP Home Science (CT), UAS, Dharwad. Peanut skin color was quantified in terms of color strength (K/S), reflectance (RFL) and CIELAB (or CIEL*a*b*) color space values (as defined by International Commission on Illumination). CIELAB values include lightness/darkness (L*), redness/greenness (a*), and yellowness/blueness (b*). To identify the color difference, white colored genotype ICG9418 was used as standard and color difference parameters (ΔE, ΔL, Δa, Δb) were derived from CIELAB values. hue angle (h°) an attribute of color perception measure of distinguishing red from green and blue from yellow (expressed in degrees), and chroma (C*) indicating the saturation of the color was determined by using the formulas.

$$\text{hue angle} = \tan^{-1}\left(\frac{b^*}{a^*}\right) \quad (1)$$

$$\text{chroma} = \sqrt{a^{*2} + b^{*2}}. \quad (2)$$

Estimation of Proximates and Nutraceuticals

The proximates were estimated from the seeds as per the standard protocols given by the Association of Official Analytical Chemists (AOAC), 2005. Seeds were powdered in pestle and mortar and

FIGURE 1 | Seeds of sixty peanut genotypes with varying skin color.

oven dry method was used to determine the moisture content of the sample and stored in desiccator to avoid absorption of moisture from outside. The dry powder was used to estimate the ash content in a muffle furnace. The dry powder was also used to determine the fat content (FC) by using Soxhlet extraction apparatus (SOCS PLUS six place automatic solvent extraction system model SCS 6 AS DLS, Pelican). The defatted moisture free samples were then used to determine the crude protein (CP) by micro-Kjeldahl method (KEL PLUS automatic micro six sample digestion system, Pelican and Kjeldhal automatic nitrogen distillation system (Classic DX VA, Pelican). The crude fiber (CF) was assessed from the defatted samples using alkali and acid digestion followed by ash estimation. The carbohydrate content was determined by differential method. Total polyphenol content (TPC) was determined by using Folin-Ciocalteau reagent (FCR) and catechol (1, 2-dihydroxybenzene) was used as standard. The values were expressed in catechol equivalent (CE) per 100 gram of seeds. Total antioxidant activity (AOA) was determined by using 2, 2-diphenyl-1-picrylhydrazyl (DPPH) reagent and expressed in % DPPH activity or per cent inhibition.

Genotyping of the Population

The genomic DNA from 60 genotypes was isolated from healthy young leaves using modified CTAB method (53). A total of 100

TABLE 1 | List of 60 peanut genotypes with differing skin color and seed size.

Sl. No.	Genotypes	Country	Sub species	Botanical varieties	Seed size	Color clusters
1	ICG334	China	*fastigiata*	*vulgaris*	Medium	Cluster-VIII
2	ICG1274	Indonesia	*fastigiata*	*fastigiata*	Small	Cluster-I
3	ICG1668	USA	*hypogaea*	*hypogaea*	Bold	Cluster-VI
4	ICG3053	India	*hypogaea*	*hypogaea*	Medium	Cluster-V
5	ICG3421	India	*fastigiata*	*vulgaris*	Medium	Cluster-VII
6	ICG5475	Kenya	*fastigiata*	*fastigiata*	Medium	Cluster-II
7	ICG5051	USA	*hypogaea*	*hypogaea*	Medium	Cluster-V
8	ICG5221	Argentina	*fastigiata*	*fastigiata*	Bold	Cluster-I
9	ICG5286	Zambia	*hypogaea*	*hypogaea*	Medium	Cluster-II
10	ICG5327	USA	*hypogaea*	*hypogaea*	Small	Cluster-V
11	ICG4670	Sudan	*fastigiata*	*fastigiata*	Small	Cluster-VI
12	ICG5662	China	*hypogaea*	*hypogaea*	Small	Cluster-V
13	ICG1711	Bolivia	*fastigiata*	*fastigiata*	Small	Cluster-III
14	ICG5236	Chile	*fastigiata*	*vulgaris*	Bold	Cluster-III
15	ICG5609	Sri Lanka	*fastigiata*	*fastigiata*	Medium	Cluster-II
16	ICG6201	Cuba	*fastigiata*	*fastigiata*	Medium	Cluster-II
17	ICG6375	Unknown	*fastigiata*	*vulgaris*	Small	Cluster-II
18	ICG6407	Zimbabwe	*fastigiata*	*vulgaris*	Medium	Cluster-VII
19	ICG6646	Unknown	*fastigiata*	*fastigiata*	Small	Cluster-VIII
20	ICG6703	Paraguay	*hypogaea*	*hypogaea*	Small	Cluster-II
21	ICG7181	India	*fastigiata*	*fastigiata*	Small	Cluster-III
22	ICG7190	Brazil	*fastigiata*	*vulgaris*	Small	Cluster-I
23	ICG7963	USA	*hypogaea*	*hypogaea*	Small	Cluster-VIII
24	ICG8106	Peru	*fastigiata*	*fastigiata*	Small	Cluster-I
25	ICG8285	USA	*hypogaea*	*hypogaea*	Medium	Cluster-V
26	ICG9418	Martinique	*fastigiata*	*vulgaris*	Small	Cluster-X
27	ICG9507	Philippines	*fastigiata*	*vulgaris*	Small	Cluster-VII
28	ICG6667	USA	*hypogaea*	*hypogaea*	Medium	Cluster-VI
29	ICG11855	Republic of Korea	*hypogaea*	*hypogaea*	Medium	Cluster-V
30	ICG11862	Republic of Korea	*hypogaea*	*hypogaea*	Small	Cluster-IV
31	ICG12625	Ecuador	*fastigiata*	*aequatoriana*	Medium	Cluster-I
32	ICG12988	India	*fastigiata*	*vulgaris*	Small	Cluster-VIII
33	ICG13982	USA	*hypogaea*	*hypogaea*	Medium	Cluster-VIII
34	ICG14475	Nigeria	*hypogaea*	*hypogaea*	Medium	Cluster-VII
35	ICG12276	Bolivia	*hypogaea*	*hypogaea*	Small	Cluster-IV
36	ICG7243	USA	*hypogaea*	*hypogaea*	Bold	Cluster-IX
37	ICG14710	Cameroon	*fastigiata*	*fastigiata*	Medium	Cluster-X
38	ICG15287	Brazil	*fastigiata*	*vulgaris*	Medium	Cluster-I
39	ICG15309	Brazil	*fastigiata*	*fastigiata*	Small	Cluster-V
40	ICG13723	Niger	*hypogaea*	*hypogaea*	Medium	Cluster-VIII
41	ICG11457	India	*hypogaea*	*hypogaea*	Small	Cluster-VI
42	ICG2381	Brazil	*hypogaea*	*hypogaea*	Bold	Cluster-I
43	ICG5827	USA	*hypogaea*	*hypogaea*	Medium	Cluster-V
44	ICG3992	India	*hypogaea*	*hypogaea*	Medium	Cluster-III
45	ICG8760	Zambia	*hypogaea*	*hypogaea*	Medium	Cluster-VI
46	ICG9037	Cote d'Ivoire	*hypogaea*	*hypogaea*	Small	Cluster-VII
47	ICGV86031	India	*fastigiata*	*vulgaris*	Small	Cluster-VII
48	A30b	India	*fastigiata*	*vulgaris*	Medium	Cluster-VII
49	ICGV06146	India	*fastigiata*	*vulgaris*	Medium	Cluster-VI
50	GPBD5	India	*fastigiata*	*vulgaris*	Small	Cluster-VIII

(Continued)

TABLE 1 | Continued

Sl. No.	Genotypes	Country	Sub species	Botanical varieties	Seed size	Color clusters
51	GPBD4	India	*fastigiata*	*vulgaris*	Medium	Cluster-VIII
52	TMV2	India	*fastigiata*	*vulgaris*	Small	Cluster-VII
53	R9227	India	*fastigiata*	*vulgaris*	Small	Cluster-IX
54	DSG1	India	*hypogaea*	*hypogaea*	Medium	Cluster-I
55	ICGV06189	India	*fastigiata*	*fastigiata*	Bold	Cluster-VII
56	Dh40	India	*fastigiata*	*vulgaris*	Small	Cluster-IV
57	Dh245	India	*fastigiata*	*vulgaris*	Small	Cluster-IX
58	Kadari6	India	*fastigiata*	*vulgaris*	Small	Cluster-VII
59	Kadari9	India	*fastigiata*	*vulgaris*	Small	Cluster-VI
60	TGLPS3	India	*hypogaea*	*hypogaea*	Bold	Cluster-V

The seed size is based on the test weight. The genotypes with <40 g test weight were categorized as "Small," 40–60 g are categorized as "Medium" and > 60 g were grouped as "Bold" seeded.

Color clustering was made as per DMRT analysis, Cluster-I, dark purple colored seeds; Cluster- II, light purple seeds; Cluster-III, dark red; Cluster-IV, light red; Cluster-V, dark pink; Cluster-VI, light pink; Cluster-VII, dark tan; Cluster-VIII, medium tan; Cluster-IX, light tan and Cluster-X, white colored seeds.

Arachis hypogaea transposable element (AhTE) markers were used to screen these genotypes (**Supplementary Table 1**). The PCR was carried out in a reaction volume of 10 μl with 50 ng of template DNA, 5 pmol of each primer, 10X of Taq polymerase buffer [500 mM KCl, 100 mM Tris-HCI (pH 8.5)], 0.25 mM of dNTPs and 0.1 U of Taq polymerase. The PCR reaction was carried out in 96 well plates using Mastercycler-PCR (MC Nexus Gradient, Eppendorf AG, Germany) with the temperature profile of initial denaturation of 5 min at 95°C and then 35 cycles at 95°C for 1 min, 53°C for 1 min, and 72°C for 1 min 30 s and 72°C for 8 min for final extension. The amplicons were separated by gel electrophoresis on 1% agarose gels in Bio-Rad gel electrophoresis unit using 1X TAE in buffer tank with electric voltage of 80 V for 1 h.

Data Analysis

The phenotypic data was analyzed to determine the analysis of variance (ANOVA), frequency distribution and genetic variability for nutritional and yield related traits of sixty peanut genotypes sown during post rainy-2017 and rainy season of 2018 and were calculated at 0.05 alpha value. A null hypothesis was rejected if the F-value > F critical value. Correlation was computed using SPSS 16.0 software. DMRT (Duncan multiple range test) was used for grouping the genotypes based on mean values of color strength. The genotypic data was used to estimate the major allele frequency, heterozygosity, and polymorphic information content (PIC) (54).

The population structure and number of subpopulations were assessed by model-based clustering algorithms using STRUCTURE Version 2.3.4 (55). The number of subpopulations (K) was set from 1 to 10, and at least 500 runs per K were conducted separately with 100,000 generations of "burn-in" and 200,000 Markov chain Monte Carlo (MCMC). The best K value was determined based on delta K analysis (56). Molecular diversity was assessed using NTSYSpc Version 2.0 and DARwin Version 6.0. In order to check if any of the AhTE marker screened on the population is associated with the traits, a marker-trait association study was carried out by using single marker analysis (SMA) using single factor ANOVA. Those significant and major marker-trait associations showing >10% R^2 were analyzed for their position in the genome and functional annotation using the gene prediction data from diploid progenitor genomes (available at https://peanutbase.org).

RESULTS

Variability in Morphological, Yield and Nutritional Traits

In this study, sixty genotypes varying in kernel skin color and size were screened phenotypically for 23 traits including nine morphological, eight nutritional and six yield related traits. Of the 23 traits, six traits (flower color, stem color, leaf color, leaf shape, growth habit, branching pattern) were qualitative in nature and rest 17 traits (leaflet length (cm), leaflet width (cm), plant height (cm), pod weight per plant (g), number of pods per plant, shelling percentage (%), test weight (g), sound mature kernel weight percentage (%), haulm weight (g), and nutritional traits including contents of moisture (%), fat (%), crude protein (%), ash (%), crude fiber (%), carbohydrate (%), and nutraceuticals like TPC (catechol equivalent per 100 g fresh weight) and AOA (% DPPH activity) were quantitative.

The 17 quantitative traits were used to determine the variance value among the sixty genotypes using ANOVA. The results revealed that there were highly significant differences among sixty genotypes for the all the quantitative traits except for haulm weight at 5% probability (**Supplementary Table 2**). High GCV coupled with PCV was observed for the traits like moisture content (23.13 and 25.52%), total antioxidant activity (20.11 and 20.12%), pod weight per plant (21.9 and 30.84%), and number of pods per plant (20.59 and 34.88%). Low level of GCV and PCV (<10%) was observed for traits like fat, crude protein, crude fiber, leaflet width, shelling percentage, and sound mature kernel weight percentage. Rest of the traits showed moderate GCV and PCV (10–20%). All the nutritional traits

except for carbohydrate content (30.32%) studied showed high heritability (>60%). Majority of morphological and yield related parameters showed moderate heritability while leaflet length and width showed high heritability (92.68 and 66.48%). The lowest heritability was observed for haulm weight (11.95%). The highest GA and GAM was observed for TPC (129.35%) and moisture content (43.19%), respectively (**Supplementary Table 3**).

Variability in Peanut Skin Color and Size

Among the sixty genotypes, varying degree of kernel skin color was observed that ranged from white to purple. In order to quantitatively determine the peanut skin color, a total of 11 color estimates like K/S, RFL, ΔE, L, a*, b*, ΔL, Δa, Δb, h°, and C* were used. The mean value of the estimates is showed in **Supplementary Table 4**. Reflectance of the surface of a material can be defined as its effectiveness in reflecting the radiant energy. Color strength is derived from reflectance and it has ability to represent the intensity of color. It can differentiate high intensity color with faded colors having low intensity. The genotypes were found to be highly differing with respect to intensity of skin color. For grouping the genotypes with similar color, we used color strength as reference parameter using DMRT analysis. A total of 10 clusters differing significantly for color strength at 5% probability were formed. The proximates and nutraceuticals were assessed according to the groups differing with color strength. The Cluster-I consisted of majority of dark purple colored seeds, Cluster-II with light purple seeds, Cluster-III dark red, Cluster-IV light red, Cluster-V dark pink, Cluster-VI light pink, Cluster-VII dark tan, Cluster-VIII medium tan, Cluster-IX light tan, and Cluster-X with white colored seeds.

The color strength decreased from Cluster-I (dark purple) to Cluster-X (white). Cluster-I having dark purple colored seeds showed the lowest crude proteins and moisture content and high range of fat content and crude fiber. Cluster-III with dark red colored seeds showed highest fat and TPC content and lowest crude fiber. Cluster- IV with light red colored seeds showed highest moisture, fat, ash, crude fiber contents. Cluster- VI with light pink colored seeds showed highest AOA and lowest fat content. Cluster-X with white colored genotypes showed the lowest fat content, TPC and AOA and highest crude protein content and moisture content. Other clusters showed intermediate range of nutritional parameters (**Table 2**).

Assessment of kernel size was based on the test weight of the seeds. Genotypes with test weight less than 40 g were considered as small seeded, those with the range of 40–50 g were considered as medium sized and with more than 50 g of test weight were recorded as bold seeded genotypes (**Table 1**).

Phenotypic Correlation

The degree of relationship between kernel size along with nutritional traits, color parameters and yield related traits were computed using Pearson's correlation test. Coefficients of correlation were computed to assess the magnitude and direction of relation between the traits. Among the 420 possible correlations from 29 quantitative traits (kernel size, 8 nutritional, 11 color parameters, 3 morphological and 6 yield related traits), a total of 113 trait pairs were found to be significant at 5% probability (**Supplementary Table 5**). Kernel size showed positive correlation with crude protein and test weight. Carbohydrate content showed negative correlation with moisture, fat, ash, crude fiber, and TPC. Fat content showed negative correlation with crude protein content. Crude protein content showed negative correlation with K/S, ΔE and h° while it showed positive correlation with RFL, L*, ΔL, b*, Δb, and chroma value. Fat content showed positive correlation with ΔE and h° while it showed negative correlation with b* value. TPC was found to be positively correlated with K/S and ΔE while it showed negative correlation with RFL, ΔL, L*, and b* value. AOA showed positive correlation with a* and Δa. Number of pods per plant showed negative correlation with plant height and positive correlation with haulm weight and pod weight per plant. Sound mature kernel weight percentage is positively correlated with pod weight per plant and test weight.

Molecular Diversity and Population Studies

Sixty genotypes of groundnut were genotyped with 100 AhTE markers. A total of 59 markers were found to be monomorphic across 60 genotypes while 41 AhTE markers showed polymorphism. The polymorphic markers were utilized to study the molecular diversity, population structure and marker-trait associations. To understand the ability of markers to differentiate the genotypes, major allele frequency, heterozygosity, and PIC were computed. Major allele frequency ranged from 0.51 (AhTE1542) to 0.98 (AhTE1438, AhTE1587) with a mean of 0.84. Heterozygosity of the markers ranged from 0.03 (AhTE1587) to 0.47 (AhTE0446) with a mean of 0.23. The PIC value ranged from 0.03 (AhTE1438 and AhTE1587) to 0.37 (AhTE0205, AhTE0474, and AhTE1542) with mean PIC value of 0.19 (**Supplementary Table 6**).

Population structure analysis showed two major groups based on delta K value when genotyping data from 41 polymorphic markers were used in analysis. The principal coordinate analysis (PCoA) and the dendrogram also showed the presence of two major groups among the genotypes (**Figure 2**). These two groups clearly showed to be consisting of genotypes with different botanical varieties. The genotypes exhibited significant phenotypic differences with respect to morphological, nutritional, colorimetric and yield related parameters but with moderate genetic diversity explained by 41 polymorphic markers among them.

Marker-Trait Association Analysis

The genotypic data from 41 polymorphic AhTE markers across 60 genotypes along with phenotypic data on 28 traits including 8 nutritional, 11 color parameters, 3 morphological and 6 yield related traits were subjected for single marker analysis (SMA) to study the marker-trait associations (MTAs). A total of 36 markers (110 MTAs) showed the significant association at 5% probability level. In total 75 MTAs (contributed by 30 markers) showed major marker-trait associations with PVE value more than 10% (**Table 3**) and about 35 MTAs showed the PVE less than 10%. An MTA, AhTE1542-Haulm weight, had the highest PVE value (27.68 %). Maximum number of marker-trait associations (43 MTAs) were detected for 9 morphological and yield related

TABLE 2 | Clusters based on color strength of the kernel skin color using Duncan's multiple range test (DMRT).

Traits	MC	Fat	CP	Ash	CF	CHO	TPC	AOA
Cluster-I	3.85[cd]	45.04[ab]	25.98[e]	2.53[d]	11.19[ab]	11.28[a]	359.76[e]	68.94[i]
Cluster-II	3.89[cd]	44.73[ab]	27.58[cd]	2.45[f]	11.02[abc]	10.35[a]	380.70[c]	71.90[h]
Cluster-III	4.74[abc]	45.31[a]	27.63[bcd]	2.50[e]	10.63[c]	9.19[a]	429.77[a]	86.27[b]
Cluster-IV	5.18[a]	45.51[a]	26.07[e]	3.11[a]	11.42[a]	8.70[a]	392.85[b]	76.03[g]
Cluster-V	4.52[abcd]	43.84[bc]	28.03[bc]	2.39[g]	10.59[c]	10.62[a]	344.36[f]	78.41[e]
Cluster-VI	4.91[ab]	42.62[c]	28.17[b]	2.50[e]	11.09[abc]	10.71[a]	385.62[bc]	86.68[a]
Cluster-VII	3.76[d]	44.71[ab]	27.28[d]	2.57[c]	10.72[bc]	10.96[a]	335.16[f]	82.24[c]
Cluster-VIII	4.19[bcd]	43.74[bc]	27.25[d]	2.30[h]	10.96[abc]	11.55[a]	299.82[g]	81.53[d]
Cluster-IX	4.19[bcd]	44.97[ab]	27.93[bc]	2.74[b]	11.48[a]	8.69[a]	369.62[d]	77.36[f]
Cluster-X	5.28[a]	42.56[c]	29.58[a]	2.39[g]	10.96[abc]	9.22[a]	269.99[h]	25.64[j]

Within columns, mean values are followed by the same letters (like a, b, c, etc.) are not significantly different ($P < 0.05$) according to Duncan's multiple range test (DMRT).
MC, Moisture content; Fat, Fat content; CP, Crude protein content; Ash, Ash content; CF, Crude fiber content; CHO, Carbohydrate content; TPC, Total polyphenol content; AOA, Total antioxidant activity; K/S: Color strength.

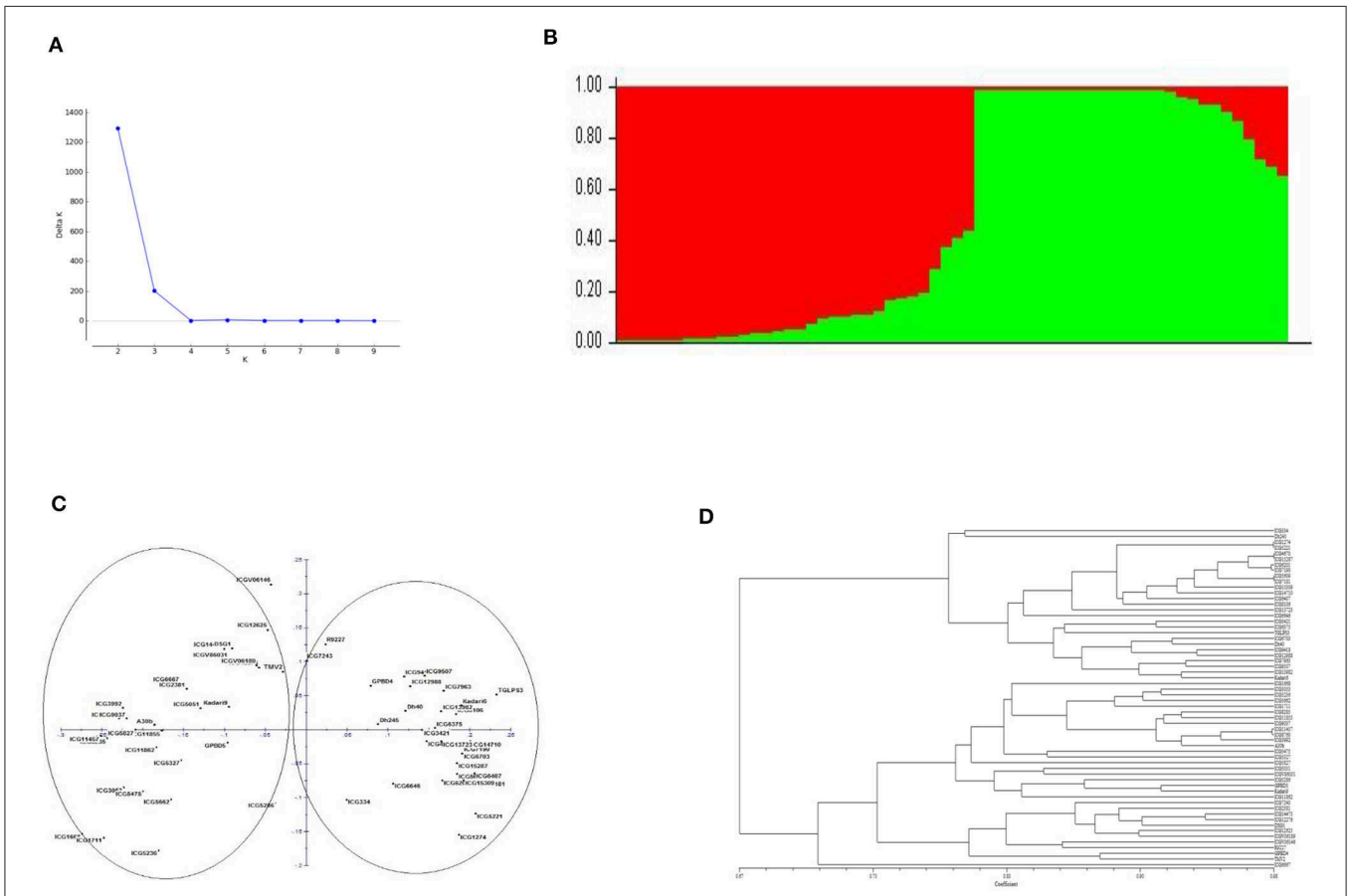

FIGURE 2 | (A) Population structure of the sixty genotypes of peanut. **(B)** Bar plot representing the number of subpopulations. **(C)** Principal co-ordinate analysis (PCoA). **(D)** Dendrogram representing the grouping of genotypes based on AhTE markers genotyping data.

traits (leaflet length, leaflet width, plant height, number of pods per plant, pod weight per plant, shelling percentage, test weight, sound mature kernel weight percentage, and haulm weight), followed by nutritional (24 MTAs) for 7 traits (fat, crude protein, ash, crude fiber, carbohydrate, TPC, and total AOA) and 8 MTAs were detected for 6 color parameters (color strength, reflectance, L*, a*, Δa, hue angle). Out of 75 major MTAs, two MTAs (AhTE0474-AOA and AhTE0362-leaflet length) were significant at 0.1% P-value and six MTAs (AhTE1542-haulm weight, AhTE0296-plant height and leaflet length, AhTE0189-TPC and

leaflet length) were significant at 0.05% P-value. Two highly significant markers (AhTE0474 and AhTE0189) were strongly associated with total AOA and TPC, respectively. The marker-trait associations for some of the important nutritional traits (fat content, crude protein content, ash content, crude fiber content, TPC, and AOA) is represented in Manhattan plot constructed by plotting position on chromosome against negative logarithm to the base 10 of observed p value of all the polymorphic AhTE markers (**Supplementary Figure 1**).

Among the 30 AhTE markers showing significant major MTA, a few were associated with multiple traits. For example, AhTE1542 showed the strong association with two nutritional traits like crude protein (11.06%) and AOA (12.63%) and five yield related traits like leaflet length (11.26%), plant height (16.05%), number of pods per plant (15.36%), shelling percentage (20.81%), and haulm weight (27.68%). AhTE1277 showed association with two nutritional traits like fat content (10.5%) and carbohydrate content (13.78%) and four yield related traits like leaflet length (18.87%), leaflet width (11.62%), plant height (13.38%), and shelling percentage (16.12%). AhTE1581 showed association for 5 traits including color parameters like color strength (10.79%), reflectance (16.13%), nutritional traits like ash content (11.39%), carbohydrate content (11.30%) and yield related trait like haulm weight (13.46%). Five markers, AhTE0143 (plant height, leaflet length, TPC, and AOA), AhTE0189 (a*, AOA, TPC, and leaflet length), AhTE0205 (haulm weight, shelling percentage, plant height, and AOA), AhTE1907 (plant height, shelling percentage, test weight, and haulm weight), and AhTE2000 (fat, crude protein, crude fiber and shelling percentage) showed the association for 4 traits each. AhTE2000 showed maximum number of association with 3 nutritional traits like fat content (14.67%), crude protein content (12.85%) and crude fiber content (15.30%). Among the quality traits AhTE0189-TPC, showed the highest PVE (24.56%). Maximum number of marker-trait associations was observed for markers on chromosome number B04 (5 AhTE markers with 10 MTAs) (**Table 3**).

Prediction of Candidate Genes using Associated Markers

The 30 AhTE markers that showed major MTAs were used to predict the probable candidate genes. Except for A09, B03, and B10, the associated markers were distributed on all other chromosomes. Out of 30 AhTE markers, 3 contains *AhMITE1* insertion site in intron, 6 in upstream, 4 in downstream, 1 UTR, 1 exon and 15 in intergenic regions (**Table 2**, **Supplementary Table 7**). For instance, the AhTE1542 that showed insertion at the exonic region of *Aradu.S8151* [codes for U3 small nucleolar ribonucleoprotein (MPP10-like protein)] on chromosome A10 showed significant association with seven traits including crude protein, total AOA, leaflet length, plant height, number of pods per plant, shelling percentage, and haulm weight. Similarly, the marker AhTE0442 showed significance association with TPC. *AhMITE1* insertion was at intronic region of the gene *Aradu.UKZ71* on chromosome A03 that codes for catalytic activity or oxidoreductase activity.

TABLE 3 | Association of molecular markers for nutritional, color, morphological, and yield related traits using single marker analysis in sixty peanut genotypes.

Sl. No.	Chromosome	Marker	Traits	*P*-value	F value	PVE
1	A01	AhTE0233	PH	0.0256*	3.92	12.69
2			AOA	0.0158*	4.48	14.23
3	A02	AhTE0465	LL	0.0077**	7.64	12.01
4			LW	0.0078**	7.63	11.98
5			SP	0.0011**	11.83	17.44
6		AhTE1098	SP	0.0037**	9.16	13.84
7	A03	AhTE0205	HW	0.0156*	6.27	11.14
8			SP	0.0036**	9.33	15.72
9			PH	0.0099**	7.19	12.57
10			AOA	0.0162*	6.19	11.01
11		AhTE0442	TPC	0.0430*	3.34	11.39
12		AhTE0474	AOA	0.0008***	12.49	18.25
13	A04	AhTE1253	PWPP	0.0056**	8.33	13.57
14			TW	0.0082**	7.55	12.48
15	A05	AhTE1271	PH	0.0461*	3.25	10.57
16		AhTE1277	Fat	0.0417*	3.36	10.50
17			CHO	0.0146*	4.55	13.78
18			LL	0.0025**	6.63	18.87
19			LW	0.0295*	3.74	11.62
20			PH	0.0166*	4.40	13.38
21			SP	0.0066**	5.47	16.12
22	A06	AhTE1337	CHO	0.0059**	8.18	12.55
23		AhTE1363	Ash	0.0047**	8.73	14.87
24			SMKW	0.0216*	5.62	10.10
25		AhTE1379	LL	0.0146*	6.34	10.35
26	A08	AhTE1437	PH	0.0042**	6.06	18.32
27	A10	AhTE1542	CP	0.0447*	3.29	11.06
28			AOA	0.0278*	3.83	12.63
29			LL	0.0422*	3.36	11.26
30			PH	0.0097**	5.06	16.05
31			NPPP	0.0120*	4.81	15.36
32			SP	0.0020**	6.96	20.81
33			HW	0.0002***	10.14	27.68
34		AhTE1564	HW	0.0369*	3.50	11.30
35	B01	AhTE1581	Ash	0.0095**	7.20	11.39
36			CHO	0.0098**	7.14	11.30
37			K/S	0.0117*	6.77	10.79
38			RFL	0.0017**	10.77	16.13
39			HW	0.0046**	8.71	13.46
40	B02	AhTE0296	SP	0.0036**	9.24	14.62
41			PH	0.0003***	14.58	21.25
42			LL	0.0003***	14.58	21.25
43	B04	AhTE0189	a*	0.0195*	5.86	11.29
44			AOA	0.0073**	7.89	14.65
45			TPC	0.0003***	14.97	24.56
46			LL	0.0001***	16.60	22.85
47		AhTE1744	HW	0.0137*	4.65	14.92
48		AhTE1761	Fat	0.0095**	5.08	15.84
49			CP	0.0088**	5.16	16.06
50		AhTE1777	Δa	0.0088**	7.37	11.82

(Continued)

TABLE 3 | Continued

Sl. No.	Chromosome	Marker	Traits	*P*-value	F value	PVE
51			h°	0.0068**	7.92	12.59
52		AhTE1780	SP	0.0361*	3.53	11.57
53	B05	AhTE0446	AOA	0.0350*	3.57	11.67
54			PH	0.0080**	8.08	23.05
55			SP	0.0425*	3.35	11.04
56	B06	AhTE1872	CHO	0.0069**	5.48	17.40
57		AhTE1907	PH	0.0105*	6.70	10.93
58			SP	0.0035**	9.30	14.02
59			TW	0.0068**	7.89	12.15
60			HW	0.0037**	9.19	13.88
61		AhTE1916	RFL	0.0073**	7.78	12.59
62	B07	AhTE0143	PH	0.0002***	16.09	22.96
63			LL	0.0090**	12.35	18.61
64			AOA	0.0038**	9.15	14.49
65			TPC	0.0144*	6.38	10.58
66	B08	AhTE2000	Fat	0.0162*	4.47	14.67
67			CP	0.0280*	3.83	12.85
68			CF	0.0133*	4.67	15.3
69			SP	0.0124*	4.78	15.52
70		AhTE2011	K/S	0.0116*	4.87	16.04
71			L*	0.0078**	5.34	17.32
72			TW	0.0011**	7.85	23.53
73	B09	AhTE0362	TPC	0.0046**	8.75	14.17
74			LL	0.0008***	12.57	19.17
75			SP	0.0175*	6.01	10.18

****Significant level at 0.5%, **Significant level at 1%, and *Significant level at 5%.*
PVE, Phenotypic variance; Fat, Fat content; CP, Crude protein; Ash, Ash content; CF, Crude fiber content; CHO, Carbohydrate content; TPC, Total polyphenol content; AOA, Total antioxidant activity; K/S, Color strength; RFL, Reflectance; L, Lightness value; a*, Redness-greenness of the color; Δa, Redness or greenness difference between sample and standard colors; h°, Hue angle; LL, Leaflet length; LW, Leaflet width; PH, Plant height; PWPP, Pod weight per plant; NPPP, Number of pods per plant; SP, Shelling percentage; TW, Test weight; SMKW, Sound mature kernel weight percentage; HW, Haulm weight.*

AhTE1437 showed significant association with plant height. *AhMITE1* insertion was found at intronic region of the gene *Aradu.Z3TSR* on chromosome A06. The gene codes for serine type carboxypeptidase activity. The marker AhTE0205 showed the significant association with haulm weight, shelling percentage, plant height, and AOA. The *AhMITE1* insertion was found at the intronic region of gene *Aradu.SV32V* on chromosome A03, that codes for uncharacterized protein LOC100811541 isoform X2 (*Glycine max*).

Among the 30 markers associated with nutritional, color parameters, morphological, and yield related traits, majority (15) were found to be reside on intergenic regions. The marker AhTE1761 corresponded to intergenic region (between *Araip.A758G* and Araip. *7QQ9G*) on chromosome B04 showed significant association with fat and crude protein content. The candidate genes showed function of stress induced protein. Similarly, AhTE1363 with intergenic position (*Aradu.ULX7G* and *Aradu.SDQ6B*) on chromosome A06 showed significant association with ash content and sound mature kernel weight percentage. The candidate genes showed the function zinc ion transmembrane transporter activity. The intergenic marker AhTE2000 located between genes *Araip.TVQ3P* and *Araip.H5SDY* on chromosome B08, was associated with fat content, crude protein, crude fiber and shelling percentage. The candidate genes showed the function of structural constituent of cell wall protein. Another intergenic marker AhTE0189 located between genes *Aradu.WFR54* and *Aradu.05MV1* on chromosome B04 showed significant associated with color parameter a* value, AOA, TPC, and leaflet length. The candidate genes showed the function of receptor-like kinase activity (**Supplementary Table 7**).

DISCUSSION

Genetic improvement of peanut for quality and yield traits is essential for combating the malnutrition and hunger in the developing countries. Major focus of the peanut breeding programs hovers around increasing the yield potential of the crop to increase the productivity under biotic and abiotic stresses. Advances in genomics technologies in peanut through next generation sequencing, have deciphered tremendous genomic resources that can be utilized in genomics-assisted breeding. In this regard, efforts were made to identify molecular markers or genomic regions associated with traits such as late leaf spot and rust (57–61), aflatoxin contamination (62), drought (63), yield related traits (50), and some quality traits like protein content, oil content, and oil quality (64–66). Another important trait that needs attention in peanut is its skin color- more commonly called as peanut kernel skin (PKS) color. The association of skin color with various nutritional parameters is well established in various vegetables and fruits (30). In food crops, the relationship between flavonoids and antioxidant activity was studied with grain color in colored rice (33, 34). However, in peanut the studies are limited to studying the relation between PKS with nutraceuticals like total polyphenol content and antioxidant capacity (31, 32). In this context this, present study reports the profiling of nutritional parameters in sixty genotypes of peanut that differ in skin color and size.

Colorimetric method was used for quantification of skin color that had estimated 11 color parameters. For color estimation, there are many color models like munsell color space, RGB/CMYK, YIQ/YUV/YCbCr, HIS/HSV/HSL, Hunter Lab color space, and CIEXYZ/CIEL*U*V*/CIEL*a*b*. Among these models CIE model and Hunter Lab color space are commonly used for evaluation of color difference while other have their application area in computer graphics, image analysis, and processing, etc. (67, 68). Both the color space are based on opponent-colors theory (69). In the present study CIELAB values were derived by using the colorimetric approach to quantify the peanut skin color. Similar approach was used to study the skin color in 27 peanut cultivars (31), 481 colored rice accessions (33). Similarly hunter scale was used for quantification of color estimation of skin color of 22 US peanut mini core collection and 4 Israeli cultivars (32), 17 colored chickpea lines (70), and 33 cool season legumes (71). These color parameters were

also considered as the biomarkers for estimating the TPC and AOA in peanut skin (31). Hence color strength was used for clustering the 60 genotypes used in the study by DMRT analysis and 10 distinct groups were made. Cluster-I consisted of dark purple while cluster-X with white colored genotypes. Cluster-X with white colored genotypes showed the lowest fat content, TPC, and AOA indicating light skin colored genotypes were poor in the nutraceuticals while dark colored seeds are rich in nutraceuticals compared to white seeded genotypes. Similar observations were recorded in peanut (31, 32), rice (33), chickpea (70). In this study light pink colored seeds showed highest antioxidant activity rather than expected dark purple colored seeds, which may be due to colorless flavonoid that exhibit strong antioxidant activity. Similar finding i.e. high TPC was observed in light pink colored peanut kernel skin (32). Cluster-X with white skin colored genotypes showed highest protein content and lowest fat content, further the study also showed the negative correlation between fat content and protein content. This clearly suggests that lighter skin color is an indicator of high protein and low fat content. The flavonoids were found to have inhibitory action on protein production in eukaryotes (72, 73). In the present study peanut skin was used to assess the correlation with nutritional parameters in the whole kernel (including skin), that provides better understanding of the nutritional status of the peanut seeds. There are some reports where nutraceuticals were measured in peanut skin devoid of kernel (32). Presently, there are no published reports to correlate proximates with skin color in peanut; hence this study will provide the prime information on proximate profiling. As expected, white colored genotypes showed lowest reflectance value and highest moisture content. The moisture content was lowest in dark tan (Cluster-VII) followed by dark and light purple colored (Cluster-I and II) seeds which contain lower moisture content maybe due to their high light absorbance nature. Generally, colored/darker grains have higher nutritional value compared to their light counterparts. For instance, the dark colored rice were nutrient rich compared white rice (74). In this study also colored seeds showed higher nutritional value compared to white and light colored seeds except crude protein content. The AOA in the present study was comparatively higher in colored seeds than the white ones; this kind of observation was also evident in rice (33). The TPC content was found to be highest in dark red (Cluster-III) colored seeds followed by light red (Cluster-IV) and lowest in white seeds (Cluster-X). But, the findings from other peanut skin studies found that light pink colored skins followed by red had larger amount of TPC content (32). The crude fiber content was maximum in light red (Cluster-IV) and light tan (Cluster-IX) and minimum in dark red and dark pink seeds (Cluster III and V).

The inheritance studies of peanut skin color carried out way back by Higgins (75) shows that the flesh-colored seed testa is dominant over white color with bigenic difference, whereas the red testa color is dominant to flesh-colored testa with monogenic difference. It also showed the necessity of flesh pigment for the expression of red color. These results were further confirmed on different white colored peanuts by Hammons (76). However, Norden *et al.* (77) revealed the presence of single dominant gene that was epistemic to the previously described genes for testa color in peanut. A contradictory observation where, white testa color was found to be dominant in one of the genotypes (78). Further, three different alleles were identified for the red testa color and among them; R1 allele was to be dominant over the other two (r2 and r3) recessive alleles (79). In another study, the purple and purple striped- testa in peanut showed monogenic and bigenic inheritance respectively (80). The observations made on the inheritance studies in peanut showed debatable results, indicating the complex nature of seed coat color. Since then, there are limited efforts to elucidate the mechanisms involved in seed coat color. The present study provides some insight into seed coat color and its association with AhTE markers like, AhTE1581-color strength, AhTE0189-reddness-greenness of the color, AhTE2011-ligthness-darkness of the color, AhTE1777-hue angle of the kernel skin color. Further investigations are necessary to understand the molecular mechanisms underlying testa color determination using holistic genomics and metabolomics approaches.

Generally, Virginia peanut cultivars are bold seeded, majority of Valencia are medium sized and Spanish types are small seeded genotypes. In the present study based on test weight genotypes were classified into small, medium, and bold seeded types. The kernel size had no effect on any of the nutritional parameters except for crude protein according to the correlation studies. A positive correlation of kernel size with protein content was observed which was contrary to the previous observation in peanut by Prathiba and Reddy (37). However, no significant correlations between seed size with protein content was observed in soybean (38). Hence the relation between kernel size and protein content is debatable and needs further investigation.

Correlation studies showed the negative association of fat content and protein which was in accordance with the findings by Dwivedi et al. (81). Number of pods per plant showed negative correlation with plant height and positive correlation with pod weight per plant. As the plant height increases, the peg penetrating the soil to make pod decreases. Hence the number of pods per plant as well as pod weight per plant decreases (82). In the present study, number of pods per plant and pod weight per plant showed high GCV coupled with PCV which was also in accordance with the findings of Hake et al. (50).

The studies on seed color and their relation with other nutrients in peanut and many other crops was restricted to the biochemical analysis and lacked the molecular approach to identify probable candidate/ genomic regions. Hence for molecular characterization, AhTE markers that show higher polymorphism were utilized in this study with PIC value ranging from 0.03 to 0.37. The heterozygosity ranged from 0.03 to 0.47, with mean 0.23 across the genotypes. This may be due to the biallelic nature of AhTE markers and also the type of population which consist of diverse genotypes. This finding was in accordance with previous peanut studies (50, 51). About 41 out of 100 AhTE markers screened across 60 genotypes were found to be polymorphic. The genotypic data was used to study the structure of the population. The study revealed that there were two subpopulations that were grouped based on the botanical types. Though variability was found with respect to phenotype,

genotypically the population showed less variation. Population structure was analyzed in different populations in peanut (like mini core collections, reference collections, mutant populations) and found that the grouping was made as per the botanical varieties of *Arachis hypogaea*, which include *hypogaea*, *hirsuta*, *fastigiata*, *vulgaris* etc. (18, 50, 83).

Most of these traits are quantitative in nature and are influenced by number of minor genes. The single marker analysis showed about 110 MTAs with 75 major and 35 minor MTAs. Out of which, 43 MTAs were detected for 9 morphological and yield related traits (leaflet length, leaflet width, plant height, number of pods per plant, pod weight per plant, shelling percentage, test weight, sound mature kernel weight percentage, and haulm weight). AhTE0465 (leaflet length, leaflet width, and shelling percentage), AhTE1253 (pod weight per plant and test weight), and AhTE1907 (plant height, shelling percentage, test weight and haulm weight) showed association with only morphological and yield related traits. The marker AhTE0205 on chromosome number A03 showed association with plant height. In another study, a SNP marker with location A03-26481539 was found to be associated with plant height in peanut using GWAS approach (84).

Few MTAs were also observed for color parameters like AhTE1581 (K/S and RFL), AhTE0189 (a*), AhTE1777 (hue angle and Δa), AhTE2011 (L*), and localized on chromosomes B01, B04, and B08. A genome wide association study in peanut also indicated a SNP with location B03-22076736 position had significant association with seed coat color (44). The occurrence of associated markers on B genome might indicate the genome dominance for color parameter that needs to be further investigated.

Out of 75 major MTAs, 24 were related to nutritional traits. Markers associated with AOA include AhTE0233, AhTE0205, AhTE0474, AhTE1542, AhTE0189, AhTE0446, AhTE0143, and highest PVE (18.25%) was explained for AhTE0474 which is localized on chromosome A03. Markers associated with TPC include AhTE0442, AhTE0189, AhTE0362, AhTE0143 and highest PVE (24.56%) was explained for AhTE0189. AhTE0189 and AhTE0143 were fund to be associated with both the nutraceuticals i.e., TPC and AOA. Trait mapping in sorghum through GWAS found candidate genes that are associated with TPC (19). Except for moisture content, markers associated with all the proximates were identified in this study. For example, AhTE2000 marker localized at chromosome B08 was associated with fat content, crude protein and crude fiber. Similarly, AhTE1761 on chromosome B04 was found be associated with fat content and crude protein content. In peanut, association studies on quality parameters using molecular markers were limited to oil content, protein, Fe and Zn content and no published records are available for proximates and nutraceuticals. Previous reports on genome-wide association on reference set of peanut consisting of 300 genotypes showed MTAs for quality traits like, oil (25 MTAs), oleic acid (2 MTAs) protein (11 MTAs), oleic/linoleic acid (22 MTAs), and zinc (1 MTA) (18). Also, highly significant 38 MTAs for protein and oil related traits were observed (50).

The associated markers were checked for their location on the peanut genome from PeanutBase. The chromosome B04 had five markers that were found to be associated with a*, AOA, TPC, leaflet length, haulm weight, fat, crude protein, Δa, hue angle, and shelling percentage. Chromosomes A03, A06, and B06 had 3 markers each that were found to be associated with 5 traits (haulm weight, shelling percentage, plant height, AOA, and TPC), 4 traits (carbohydrate, ash, sound mature kernel weight percentage, and leaflet length) and 6 traits (carbohydrate, plant height, shelling percentage, test weight, haulm weight, reflectance), respectively. Further gene annotation for the candidate genes of some of the marker like AhTE0442, AhTE0205, and AhTE0474 on chromosome number A03 showed functions related to oxidoreductase activity. Several reports in peanut showed markers associated with disease resistance were also localized on A03 that was evident from QTL mapping and sequencing (59, 60, 85). The occurrence of genes encoding for antioxidant activity at A03 might provide disease resistance in peanut. However, large number of markers needs to be screened on a larger population to identify the candidate genes/markers using genome-wide association analysis to validate these findings. Further, the associated markers after validation and identified genotypes with favorable alleles can be utilized in molecular breeding for nutritionally rich peanuts.

CONCLUSIONS

In the current scenario, improving the peanut crop with quality parameters is equally important along with increasing yield and pest resistance. In order to combat the malnutrition and hunger, we need to work toward "more nutrition per bite." This study provides relation of peanut skin color with various nutritional parameters (proximates, nutraceuticals), morphological and yield related traits. The markers associated with important traits after validation and genotypes with favorable alleles can be utilized in genomics-assisted peanut improvement.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

SN and HN conceived the presented idea. VH performed the experiment. GN provided the seed material. SN, PB, and HN designed the experiments. RB provided the molecular markers. PB facilitated lab for nutrition estimation and helped in conducting nutritional experiments. SN and VH analyzed the data, and wrote the manuscript. SN supervised the project and arranged for funding.

FUNDING

Authors acknowledge the funding from Staff Research Project by University of Agricultural Sciences, Dharwad and Department of Biotechnology, Government of India to carry out this research. Authors also acknowledge the instrumentation facility

at Department of Biotechnology, UAS Dharwad developed by funding from Karnataka Innovation and Technology Society (KITS), Department of IT and BT, Government of Karnataka.

ACKNOWLEDGMENTS

Authors acknowledge the technical help from Mr. Paramesh S. and Ms. Pallavi Badami for estimating the nutrients. Authors are thankful to Dr. Sannapapamma Kudligi for facilitating the colorimetric estimation of peanut kernel skin color and Dr. Uma Kulkarni for technical advice during nutrient estimation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2020.00045/full#supplementary-material

REFERENCES

1. Griel AE, Eissenstat B, Juturu V, Hsieh G, Kris-Etherton PM. Improved diet quality with peanut consumption. *J Am Coll Nutr.* (2004) 23:660–8. doi: 10.1080/07315724.2004.10719408
2. Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat Genet.* (2019) 51:865. doi: 10.1038/s41588-019-0402-2
3. Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat Genet.* (2019) 51:877. doi: 10.1038/s41588-019-0405-z
4. FAOSTAT (2017). Available online at: www.fao.org/faostat/en/#data. QC (accessed January, 2018).
5. Arya SS, Salve AR, Chauhan S. Peanuts as functional food: a review. *J Food Sci Technol.* (2016) 53:31–41. doi: 10.1007/s13197-015-2007-9
6. Kris-Etherton PM, Hu FB, Ros E, Sabaté J. The role of tree nuts and peanuts in the prevention of coronary heart disease: multiple potential mechanisms. *J Nutri.* (2008) 138:1746S—51S. doi: 10.1093/jn/138.9.1746S
7. Awad AB, Chan KC, Downie AC, Fink CS. Peanuts as a source of beta-sitosterol, a sterol with anticancer properties. *Nutr Cancer.* (2000) 36:238–41. doi: 10.1207/S15327914NC3602_14
8. Tsai J, Leitzmann MF, Hu FB, Willett WC, Giovannucci EL. A prospective cohort study of nut consumption and the risk of gallstone disease in men. *Am J Epidemiol.* (2004) 160:961–8. doi: 10.1093/aje/kwh302
9. De Pee S, Bloem MW. Current and potential role of specially formulated foods and food supplements for preventing malnutrition among 6-to 23-month-old children and for treating moderate malnutrition among 6-to 59-month-old children. *Food Nutr Bull.* (2009) 30:S434–63. doi: 10.1177/15648265090303S305
10. Boyd W. Natural colors as functional ingredients in healthy foods. *Cereal Foods World.* (2000) 45:221–2.
11. Brown A. Core collections: a practical approach to genetic resources management. *Genome.* (1989) 31:818–24. doi: 10.1139/g89-144
12. Holbrook CC, Dong W. Development and evaluation of a mini core collection for the US peanut germplasm collection. *Crop Sci.* (2005) 45:1540–44. doi: 10.2135/cropsci2004.0368
13. Upadhyaya HD, Bramel PJ, Ortiz R, Singh S. Developing a mini core of peanut for utilization of genetic resources. *Crop Sci.* (2002) 42:2150–6. doi: 10.2135/cropsci2002.2150
14. Jiang H, Ren X, Huang J, Liao B, Lei Y. Establishment of peanut mini core collection in China and exploration of new resource with high oleate. *Chinese J Oil Crop Sci.* (2008) 30:2942299.
15. Upadhyaya H, Dwivedi S, Vadez V, Hamidou F, Singh S, Varshney R, et al. Multiple resistant and nutritionally dense germplasm identified from mini core collection in peanut. *Crop Sci.* (2014) 54:679–93. doi: 10.2135/cropsci2013.07.0493
16. Pattanashetti SK, Naidu GK, Singh OK, Biradar BD. Identification of iron deficiency chlorosis tolerant sources from mini-core collection of groundnut (*Arachis hypogaea* L.). *Plant Genetic Resour.* (2018) 16:446–58. doi: 10.1017/S1479262117000326
17. Saleem M, Naidu GK, Nadaf HL, Tippannavar PS. Assessment of groundnut (*Arachis hypogaea* L.) mini core for resistance to multiple biotic stresses under hot spot location. *Int J Curr Microbiol Appl Sci.* (2018) 7:1599–614. doi: 10.20546/ijcmas.2018.711.182
18. Pandey MK, Upadhyaya HD, Rathore A, Vadez V, Sheshshayee M, Sriswathi M, et al. Genomewide association studies for 50 agronomic traits in peanut using the 'reference set' comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. *PLoS ONE.* (2014) 9:e105228. doi: 10.1371/journal.pone.0105228
19. Rhodes DH, Hoffmann L Jr, Rooney WL, Ramu P, Morris GP, Kresovich SJ, et al. Genome-wide association study of grain polyphenol concentrations in global sorghum [*Sorghum bicolor* (L.) Moench]. *Germplasm.* (2014) 62:10916–27. doi: 10.1021/jf503651t
20. Kumar J, Saripalli G, Gahlaut V, Goel N, Meher PK, Mishra KK, et al. Genetics of Fe, Zn, β-carotene, GPC and yield traits in bread wheat (*Triticum aestivum* L.) using multi-locus and multi-traits GWAS. *Euphytica.* (2018) 214:219. doi: 10.1007/s10681-018-2284-2
21. Han Z, Zhang J, Cai S, Chen X, Quan X, Zhang G. Association mapping for total polyphenol content, total flavonoid content and antioxidant activity in barley. *BMC Gen.* (2018) 19:81. doi: 10.1186/s12864-018-4483-6
22. Chen X, Tao Y, Ali A, Zhuang Z, Guo D, Guo Q, et al. Transcriptome and proteome profiling of different colored rice reveals physiological dynamics involved in the flavonoid pathway. *Int J Mol Sci.* (2019) 20:2463. doi: 10.3390/ijms20102463
23. Bowen K, Hagan A. Temperature and moisture conditions that affect aflatoxin contamination of peanuts. *Pean Sci.* (2015) 42:121–7. doi: 10.3146/0095-3679-42.2.121
24. Singh B, Singh U. Peanut as a source of protein for human foods. *Plant Foods Hum Nutr.* (1991) 41:165–77. doi: 10.1007/BF02194085
25. Francisco LD, Resurreccion A. Functional components in peanuts. *Crit Rev Food Sci Nutr.* (2008) 48:715–46. doi: 10.1080/10408390701640718
26. Hill GM. Peanut by-products fed to cattle. The Veterinary Clinics of North America. *Food An Pract.* (2002) 18:295–315. doi: 10.1016/S0749-0720(02)00019-1
27. Todd JJ, Vodkin LO. Pigmented soybean (Glycine max) seed coats accumulate proanthocyanidins during development. *Plant Physiol.* (1993) 102:663–70. doi: 10.1104/pp.102.2.663
28. Liu Z, Liu Y, Pu Z, Wang J, Zheng Y, Li Y, et al. Regulation, evolution, and functionality of flavonoids in cereal crops. *Biotechnol. Lett.* (2013) 35:1765–80. doi: 10.1007/s10529-013-1277-4
29. Nair S, Nagar R, Gupta R. Antioxidant phenolics and flavonoids in common Indian foods. *J Assoc Phys India.* (1998) 46:708–10.
30. Yao LH, Jiang YM, Shi J, Tomas-Barberan FA, Datta N, Singanusong R, et al. Flavonoids in food and their health benefits. *Plant Foods Hum Nutr.* (2004) 59:113–22. doi: 10.1007/s11130-004-0049-7
31. Chukwumah Y, Walker LT, Verghese MJ. Peanut skin color: a biomarker for total polyphenolic content and antioxidative capacities of peanut cultivars. *Int J Mol Sci.* (2009) 10:4941–52. doi: 10.3390/ijms10114941
32. Shem-Tov, Badani H, Segev A, Hedvat I, Galili S, Hovav R. Determination of total polyphenol, flavonoid and anthocyanin contents and antioxidant capacities of skins from peanut (*Arachis hypogaea*) lines with different skin colors. *J Food Biochem.* (2012) 36:301–8. doi: 10.1111/j.1745-4514.2011.00539.x
33. Shen Y, Jin L, Xiao P, Lu Y, Bao J. Total phenolics, flavonoids, antioxidant capacity in rice grain and their relations to grain color,

size and weight. *J Cer Sci.* (2009) 49:106–11. doi: 10.1016/j.jcs.2008.07.010
34. Descalsota-Empleo GI, A.Noraziyah AS, Navea IP, Chung C, Dwiyanti MS, R.Labios JD, et al. Genetic dissection of grain nutritional traits and leaf blight resistance in rice. *Genes.* (2019) 10:30. doi: 10.3390/genes10010030
35. Zhai S, Liu J, Xu D, Wen W, Yan J, Zhang P, et al. A genome-wide association study reveals a rich genetic architecture of flour color-related traits in bread wheat. *Front Plant Sci.* (2018) 9:1136. doi: 10.3389/fpls.2018.01136
36. Venuprasad R, Aruna R, Nigam S. Inheritance of traits associated with seed size in groundnut (*Arachis hypogaea* L.). *Euphytica.* (2011) 181:169–177. doi: 10.1007/s10681-011-0390-5
37. Prathiba K, Reddy MU. Nutrient composition of groundnut cultures (*Arachis hypogaea* L.) in relation to their kernel size. *Plant Foods Hum Nutr.* (1994) 45:365–9. doi: 10.1007/BF01088086
38. Maestri DM, Guzmán GA, Giorda LM. Correlation between seed size, protein and oil contents, and fatty acid composition in soybean genotypes. *Grasas Aceites.* (1998) 49:450–3. doi: 10.3989/gya.1998.v49.i5-6.757
39. Nayak SN, Singh VK, Varshney RK. Marker-assisted selection. In: Thomas B, Murray BG, Murphy DJ, editors. *Encyclopedia of Applied Plant Sciences (Second Edition),* Oxford: Academic Press. (2017) p. 183–97. doi: 10.1016/B978-0-12-394807-6.00192-1
40. Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci USA.* (1997) 94:7704–11. doi: 10.1073/pnas.94.15.7704
41. Bhat RS, Patil VU, Chandrashekar TM, Sujay V, M.Gowda VC, Kuruvinashetti MS. Recovering flanking sequence tags of miniature inverted-repeat transposable element by thermal asymmetric interlaced-PCR in peanut. *Current Science.* (2008) 95:452–3.
42. Gowda VC, Bhat RS, Motagi BN, Sujay V, Varshakumari, Bhat S. Association of high-frequency origin of late leaf spot resistant mutants with AhMITE1 transposition in peanut. *Plant Breeding.* (2010) 129:567–9. doi: 10.1111/j.1439-0523.2009.01704.x
43. Shirasawa K, Hirakawa H, Tabata S, Hasegawa M, Kiyoshima H, Suzuki S, et al. Characterization of active miniature inverted-repeat transposable elements in the peanut genome. *Theor Appl Genet.* (2012) 124:1429–38. doi: 10.1007/s00122-012-1798-6
44. Zhang Q, Arbuckle J, Wessler SR. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize. *Proc Natl Acad Sci USA.* (2000) 97:1160–5. doi: 10.1073/pnas.97.3.1160
45. Shirasawa K, Isobe S, Tabata S, Hirakawa H. Kazusa Marker DataBase: a database for genomics, genetics, and molecular breeding in plants. *Breed Sci.* (2014) 64:264–71. doi: 10.1270/jsbbs.64.264
46. Gowda VC, Bhat RS, Sujay V, Kusuma P, Varshakumari, Bhat S, et al. Characterization of AhMITE1 transposition and its association with the mutational and evolutionary origin of botanical types in peanut (Arachis spp.). *Plant Syst Evol.* (2011) 291:153–8. doi: 10.1007/s00606-010-0373-3
47. Kolekar RM, Sujay V, Shirasawa K, Sukruth M, Khedikar YP, M.Gowda VC, et al. QTL mapping for late leaf spot and rust resistance using an improved genetic map and extensive phenotypic data on a recombinant inbred line population in peanut (*Arachis hypogaea* L.). *Euphytica.* (2016) 209:147–56. doi: 10.1007/s10681-016-1651-0
48. Mondal S, Hande P, Badigannavar AM. Identification of transposable element markers for a rust (Puccinia arachidis Speg.) resistance gene in cultivated peanut. *J Phytopathol.* (2013) 162:548–52. doi: 10.1111/jph.12220
49. Gayathri M, Shirasawa K, Varshney RK, Pandey MK, Bhat RS. Development of new AhMITE1 markers through genome-wide analysis in peanut (*Arachis hypogaea* L.). *BMC Res Notes.* (2018) 11:10. doi: 10.1186/s13104-017-3121-8
50. Hake AA, Shirasawa K, Yadawad A, Nayak SN, Mondal S, Badigannavar AM, et al. Identification of transposable element markers associated with yield and quality traits from a large population of independent mutants in peanut (*Arachis hypogaea* L.). *Euphytica.* (2017) 213:283. doi: 10.1007/s10681-017-2070-6
51. Hake AA, Shirasawa K, Yadawad A, Sukruth M, Patil M, Nayak SN, et al. Mapping of important taxonomic and productivity traits using genic and non-genic transposable element markers in peanut (*Arachis hypogaea* L.). *PLoS ONE.* (2017) 12:e0186113. doi: 10.1371/journal.pone.0186113
52. IBPGR and ICRISAT. *Descriptors for Groundnut.* Rome; Patancheru: International Board for Plant Genetic Resources; International Crops Research Institute for the Semi-Arid Tropics (1992).
53. Mace ES, Buhariwalla KK, Buhariwalla HK, Crouch JH. A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol Biol Rep.* (2003) 21:459–460. doi: 10.1007/BF02772596
54. Hildebrand CE, Torney DC, Wagner RP. Informativeness of polymorphic DNA markers. *Los Alamos Sci.* (1992) 20:100–2.
55. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* (2000) 155:945–59.
56. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* (2005) 14:2611–20. doi: 10.1111/j.1365-294X.2005.02553.x
57. Khedikar Y, Gowda M, Sarvamangala C, Patgar K, Upadhyaya H, R.Varshney JT, et al. A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea* L.). *Theor Appl Genet.* (2010) 121:971–84. doi: 10.1007/s00122-010-1366-x
58. Sujay V, Gowda MV, Pandey MK, Bhat RS, Khedikar YP, Nadaf HL, et al. Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). *Mol Breed.* (2012) 30:773–88. doi: 10.1007/s11032-011-9661-z
59. Han S, Yuan M, Clevenger JP, Li C, Hagan A, Zhang X, et al. A SNP-based linkage map revealed QTLs for resistance to early and late leaf spot diseases in peanut (*Arachis hypogaea* L.). *Front Plant Sci.* (2018) 9:1012. doi: 10.3389/fpls.2018.01012
60. Pandey MK, Khan AW, Singh VK, Vishwakarma MK, Shasidhar Y, Kumar V, et al. QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea* L.). *Plant Biotechnol J.* (2017) 15:927–941. doi: 10.1111/pbi.12686
61. Yeri SB, Bhat RS. Development of late leaf spot and rust resistant backcross lines in JL 24 variety of groundnut (*Arachis hypogaea* L.). *Electr J Plant Breed.* (2016) doi: 10.5958/0975-928X.2016.00005.3
62. Yu B, Huai D, Huang L, Kang Y, Ren X, Chen Y, et al. Identification of genomic regions and diagnostic markers for resistance to aflatoxin contamination in peanut (*Arachis hypogaea* L.). *BMC Genet.* (2019) 20:32. doi: 10.1186/s12863-019-0734-z
63. Ravi K, Vadez V, Isobe S, Mir R, Guo Y, Nigam S, et al. Identification of several small main-effect QTLs and a large number of epistemic QTLs for drought tolerance related traits in groundnut (Arachishypogaea L.). *Theor Appl Genet.* (2011) 122:1119–1132. doi: 10.1007/s00122-010-1517-0
64. Sarvamangala C, Gowda M, Varshney R. Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis hypogaea* L.). *Field Crops Res.* (2011) 122:49–59. doi: 10.1016/j.fcr.2011.02.010
65. Pandey MK, Wang ML, Qiao L, Feng S, Khera P, Wang H, et al. Identification of QTLs associated with oil content and mapping FAD2 genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* L.). *BMC Genet.* (2014) 15:133. doi: 10.1186/s12863-014-0133-4
66. Huang L, He H, Chen W, Ren X, Chen Y, Zhou X, et al. Quantitative trait locus analysis of agronomic and quality-related traits in cultivated peanut (*Arachis hypogaea* L.). *Theor Appl Genet.* (2015) 128:1103–115. doi: 10.1007/s00122-015-2493-1
67. Ibraheem NA, Hasan MM, Khan RZ, Mishra PK. Understanding color models: a review. *J Sci Tech.* (2012) 2:265–75.
68. Hunter RS, Harold RW. *The Measurement of Appearance*. New Jersey, NJ: John Wiley & Sons (1987).
69. Krantz DH. Color measurement color theory: II. Opponent-colors theory. *J Math Psychol.* (1975) 12:304–27. doi: 10.1016/0022-2496(75)90027-9
70. Segev A, Badani H, Kapulnik Y, Shomer I, Oren-Shamir M, Galili S. Determination of polyphenols, flavonoids, and antioxidant capacity in colored chickpea (Cicer arietinum L.). *J Food Sci.* (2010) 75:S115–S9. doi: 10.1111/j.1750-3841.2009.01477.x
71. Xu B, Yuan S, Chang S. Comparative analyses of phenolic composition, antioxidant capacity, and color of cool season legumes and other selected food legumes. *J. Food Sci.* (2007) 72:S167–S77. doi: 10.1111/j.1750-3841.2006.00261.x

72. Ito T, Warnken SP, May WS. Protein synthesis inhibition by flavonoids: roles of eukaryotic initiation factor 2α kinases. *Biochem Biophys Res Commun.* (1999) 265:589–94. doi: 10.1006/bbrc.1999.1727
73. Hosokawa N, Hirayoshi K, Nakai A, Hosokawa Y, Marui N, Yoshida M, et al. Flavonoids inhibit the expression of heat shock proteins. *Cell Struct Funct.* (1990) 15:393–401. doi: 10.1247/csf.15.393
74. Kushwaha UKS. *Black Rice: Research, History and Development*. New York, NY: Springer (2016). doi: 10.1007/978-3-319-30153-2
75. Higgins B. Inheritance of seed-coat color in peanuts. *J Agric Res.* (1940) 61:745–52.
76. Hammons RO. White testa inheritance in the peanut. *J. Hered.* (1963) 54:139–42. doi: 10.1093/jhered/54.4.139
77. Norden A, Knauft D, Gorbet D. A dominant gene for white seed coat in peanut (*Arachis hypogaea* L.). *J. Hered.* (1988) 79:212–4. doi: 10.1093/oxfordjournals.jhered.a110495
78. Branch W. Inheritance of dominant white peanut testa color. *J Hered.* (1989) 80:155–6. doi: 10.1093/oxfordjournals.jhered.a110817
79. Branch W, Holbrook C. Genic relationship between R1, R2, and R3 for red peanut testa color. *Peanut Sci.* (1988) 15:13–4. doi: 10.3146/i0095-3679-15-1-4
80. Branch W. Inheritance of purple purple-stripe testa colors in the peanut. *J Hered*. (1985) 76:225–6. doi: 10.1093/oxfordjournals.jhered.a110080
81. Dwivedi S, Jambunathan R, Nigam S, Raghunath K, Shankar KR, Nagabhushanam G. Relationship of seed mass to oil and protein contents in peanut (*Arachis hypogaea* L.). *Peanut Sci.* (1990) 17:48–52. doi: 10.3146/i0095-3679-17-2-1
82. Gomes LF, de Almeida Lopes ÂC. Correlations and path analysis in peanut. *Crop Breed Appl Biotechnol.* (2005) 5:14. doi: 10.12702/1984-7033.v05n01a14
83. Wang ML, Sukumaran S, Barkley NA, Chen Z, Chen CY, Guo B, et al. Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. *Theor Appl Genet.* (2011) 123:1307–17. doi: 10.1007/s00122-011-1668-7
84. Zhang X, Zhang J, He X, Wang Y, Ma X, Yin D. Genome-wide association study of major agronomic traits related to domestication in peanut. *Front Plant Sci.* (2017) 8:1611. doi: 10.3389/fpls.2017.01611
85. Shirasawa K, Bhat RS, Khedikar YP, Sujay V, Kolekar RM, Yeri SB, et al. Sequencing analysis of genetic loci for resistance for late leaf spot and rust in peanut (*Arachis hypogaea* L.). *Front Plant Sci.* (2018) 9:1727. doi: 10.3389/fpls.2018.01727

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Nayak, Hebbal, Bharati, Nadaf, Naidu and Bhat. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 10 June 2020
doi: 10.3389/fgene.2020.00514

Combining High Oleic Acid Trait and Resistance to Late Leaf Spot and Rust Diseases in Groundnut (*Arachis hypogaea* L.)

*Dnyaneshwar B. Deshmukh[1,2], Balram Marathi[1], Hari Kishan Sudini[2], Murali T. Variath[2], Sunil Chaudhari[2], Surendra S. Manohar[2], Ch V. Durga Rani[1], Manish K. Pandey[2] and Janila Pasupuleti[2]**

[1] Professor Jayashankar Telangana State Agricultural University (PJTSAU), Hyderabad, India, [2] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

OPEN ACCESS

Edited by:
Jianping Wang,
University of Florida, United States

Reviewed by:
Ramesh S. Bhat,
University of Agricultural Sciences,
Dharwad (ICAR), India
Charles Y. Chen,
Auburn University, United States

***Correspondence:**
Janila Pasupuleti
P.Janila@cgiar.org

Specialty section:
This article was submitted to
Nutrigenomics,
a section of the journal
Frontiers in Genetics

Received: 26 July 2019
Accepted: 27 April 2020
Published: 10 June 2020

Citation:
Deshmukh DB, Marathi B, Sudini HK, Variath MT, Chaudhari S, Manohar SS, Rani CVD, Pandey MK and Pasupuleti J (2020) Combining High Oleic Acid Trait and Resistance to Late Leaf Spot and Rust Diseases in Groundnut (Arachis hypogaea L.). Front. Genet. 11:514. doi: 10.3389/fgene.2020.00514

High oleic trait, resistance to rust and late leaf spot (LLS) are important breeding objectives in groundnut. Rust and LLS cause significant economic loss, and high oleic trait is an industry preferred trait that enhances economic returns. This study reports marker-assisted selection to introgress high oleic content, resistance to LLS and rust into Kadiri 6 (K 6), a popular cultivar. The alleles for target traits were selected using linked allele-specific, simple sequence repeats and single nucleotide polymorphic markers. The F_1s (384), intercrossed F_1s (441), BC_1F_1s (380), BC_1F_2s (195), and BC_1F_3s (343) were genotyped to obtain desired allelic combination. Sixteen plants were identified with homozygous high oleic, LLS and rust resistance alleles in BC_1F_2, which were advanced to BC_1F_3 and evaluated for disease resistance, yield governing and nutritional quality traits. Phenotyping with Near-Infrared Reflectance Spectroscopy identified three lines (BC_1F_3-76, BC_1F_3-278, and BC_1F_3-296) with >80% oleic acid. The identified lines exhibit high levels of resistance to LLS and rust diseases (score of 3.0–4.0) with preferred pod and kernel features. The selected lines are under yield testing trials in multi-locations for release and commercialization. The lines reported here demonstrated combining high oleic trait with resistance to LLS and rust diseases.

Keywords: groundnut, marker-assisted backcrossing, high oleic acid, late leaf spot, rust, disease resistance

INTRODUCTION

Groundnut or peanut (*Arachis hypogaea* L.) has diversified uses; it is used for food, feed, fodder, and industrial purposes (Sahdev, 2015). It is cultivated across 118 countries covering 27.94 million hectares (Mha) area, with an annual production of 47.09 million tons (MT) of pods. India ranks first with an area of 5.30 Mha, and second in production with 9.17 MT of pods (FAOSTAT, 2017). The demand for groundnut in the food industry is increasing due to its multiple health benefits. The global vegetable oil consumption is expected to double by 2040 including groundnut oil (FAOSTAT, 2015). The groundnut kernel fat consists of monounsaturated fatty acid, oleic acid (36–81.3%), and polyunsaturated fatty acid, linoleic acid (3.9–40.2%) (Janila et al., 2016b). Groundnut kernels with

≥78% oleic acid are considered as high oleic, which contains an oleic to linoleic (O/L) acid fraction of ≥10:1. The consumption of oleic acid-rich diet enhances the ratio of high-density lipoprotein (HDL) to low-density lipoprotein (LDL), thus minimizes the risk of cardiovascular diseases (CVD) by 15% (Wang and Hu, 2017). Palmitic acid, another major fatty acid in groundnut kernels is associated with risk of CVD (Hu et al., 2001). Unlike oleic acid, higher linoleic acid is vulnerable to oxidation causing off-flavors, rancidity, and negatively impacts the oil stability (Patel et al., 2004). Whereas, high oleic groundnut and its products have 5–10 times elevated shelf life than that of normal groundnut, making high oleic groundnut beneficial for oil, food and processing industries.

Spontaneous groundnut mutants (F435-1 and F435-2 with ∼80.0% oleic acid, and 2.0% linoleic acid) were identified at the University of Florida, USA (Norden et al., 1987). Subsequently, the first high oleic line, SunOleic 95R, was bred in the USA (Gorbet and Knauft, 1997). The high oleic trait in groundnut is a result of two homoeologous mutant alleles of *FAD2A* and *FAD2B* on linkage group A09, and B09, respectively, which encode Δ12 fatty acid desaturase (FAD) (Moore and Knauft, 1989; Yu et al., 2008; Chu et al., 2009). In normal groundnut (∼45.0% oleic acid), oleic acid gets converted to linoleic acid by the catalytic activity of FAD. The *FAD2A* mutation involves the substitution of G:C→A:T in the open reading frame (ORF) at 448 base pair (bp) positions on 9th chromosome of A genome, thereby replacing the aspartic acid with asparagine. In the case of *FAD2B* mutation, insertion of adenine at 442 bp in the ORF on 9th chromosome of B genome shifts the ORF resulting in a truncated inactive Δ12 FAD (Yu et al., 2008). These mutations result in the elevated accumulation of high oleic acid as the conversion of oleic to linoleic is stalled and thus decreases linoleic acid. Until now, more than 80 high oleic groundnut varieties are registered globally, which are developed through conventional breeding, marker-assisted selection (MAS), marker-assisted backcrossing (MABC), and mutagenesis (Wang M. L. et al., 2015; Janila et al., 2016b; Bera et al., 2018). In Asian and African market high oleic groundnut has been recently commercialized, however, combining must-have traits such as late leaf spot (LLS), and rust resistance with high oleic are limited (Shasidhar et al., 2020).

Co-occurrence of LLS and rust is common in rainfed ecologies world over, and studies show that in India they together cause 50–70% reduction in pod yield, and reduce the quality and digestibility of haulms up to 22% (Pande et al., 2003; Singh et al., 2011). The fungal pathogens *Phaeoisariopsis personata* (Berk. and Curtis) causes LLS, and *Puccinia arachidis* (Spegazzini) causes rust disease. Although preventative fungicidal spray keeps LLS and rust at bay (Subrahmanyam et al., 1984), fungicides are not feasible in low-input agro-ecologies of Asia and Africa and are not environmentally sustainable. Groundnut crop benefits the farmers through pod yield and adds value through haulms (vines with leaves) as a fodder source for livestock especially in Asia and Africa. In India, groundnut haulms contribute to more than 30% to the fodder obtained from leguminous residues (Nigam and Blummel, 2010). Breeding for disease-resistant varieties is most feasible and environmentally sound approach (Pande et al., 2003).

Host-resistance has been described in cultivated and wild *Arachis* spp. for LLS, and rust (Pande and Rao, 2001; Sudini et al., 2015; Chaudhari et al., 2019). The interspecific derivatives, ICGV 86687 (CS 16–B2–B2), ICGV 86855 (CS 16), ICRISAT (1987) and VG 9514 (Varman, 1999) were developed with high-levels of resistance to LLS and rust. The cultivar GPBD 4 which was bred using CS 16 (interspecific derivative, *A. hypogaea* × *A. cardenasii*) is often used as a donor parent to introgress LLS and rust-resistant quantitative trait loci (QTLs) into popular groundnut varieties. The QTLs governing LLS, and rust resistance were mapped in the recombinant inbred line (RIL) population derived from GPBD 4 (resistant) and TAG 24 (susceptible). The genomic region on chromosome A03 consists of three QTLs for LLS representing about 67.98% phenotypic variation explained (PVE) and rust up to 82.96% PVE (Sujay et al., 2012). The identified QTLs were resolved to 3.06 Mb (131.6–134.6 Mb) on chromosome A03 regions which consisted of ∼25 putative candidate-genes governing LLS and rust resistance (Pandey et al., 2017). The molecular markers for LLS and rust resistance were validated (Sukruth et al., 2015; Yeri and Bhat, 2016) and used to develop LLS and rust-resistant lines using MABC in three susceptible, yet popular varieties *viz.,* TAG 24, ICGV 91114 and JL 24 (Varshney et al., 2014). Six introgression lines with 39–79% higher mean pod yield, and 8–25% higher haulm yield than their recurrent parents were selected based on field evaluation trials (Janila et al., 2016a). More recently three popular varieties from Gujarat, namely GJG 9, GG 20, and GJG-HPS 1 have also been improved for foliar disease resistance in addition to high oleic acid (Shasidhar et al., 2020).

In earlier reports, MABC has been successful in improving groundnuts for nematode resistance (Simpson et al., 2003), rust resistance (Varshney et al., 2014), LLS and rust resistance (Janila et al., 2016a; Yeri and Bhat, 2016; Kolekar et al., 2017; Shasidhar et al., 2020) and enhancing the oleic acid content in groundnut (Janila et al., 2016b; Bera et al., 2018; Shasidhar et al., 2020). Similarly, MABC was deployed to develop lines in groundnut for high oleic content and nematode resistance (Chu et al., 2011); tomato spotted wilt virus (TSWV), peanut root-knot nematode resistance (Holbrook et al., 2017).

Modern crop breeding and improved management practices have enriched 0.8–1.2% of annual gain in crop productivity (Li et al., 2018). In essence, the genetic gain (ΔG) in cultivar improvement program can be elevated by increasing selection intensity (i), additive genetic variation (σ_a) existing in the population, selection accuracy (r) and reducing the breeding cycle time (L) (Cobb et al., 2019).

$$\Delta G = [(\sigma_a)\ \ (i)\ \ \ (r)]/L$$

Deploying molecular markers for selection in early generation increases the selection intensity in the breeding population. Rapid generation advancement and resource allocation reduces the breeding cycle time. The selection accuracy can be increased using high-throughput phenotyping tools like Near-Infrared Reflectance Spectroscopy (NIRS) and detached leaf technique.

Kadiri 6 (K 6) is a popular Spanish Bunch groundnut variety has high pod yield, uniform pod size, preferred pod and kernel

features, and attractive light tan kernels. K 6 matures in 100–105 days with 70% shelling outturn and 39.0 g sound mature 100-kernel weight. The K 6 kernel contains ∼48.0% oil consist of ∼40.0% oleic acid, and 38.0% linoleic acid. It was developed at Agricultural Research Station, Kadiri, Acharya N. G. Ranga Agricultural University of Andhra Pradesh, India and released for cultivation in the Indian states of Andhra Pradesh, Telangana, and Karnataka in 2005. At present, K 6 covers over 46% Breeder Seed demand in India through public sector seed distribution (AICRP-G, 2018) and is the most popular variety. However, K 6 is highly susceptible to LLS and rust diseases and this study was attempted to combine high oleic trait, LLS and rust resistance in K 6 using the MABC approach.

MATERIALS AND METHODS

Plant Material

A popular Spanish Bunch groundnut cultivar, K 6 was used as recipient parent (RP). The advanced breeding lines- ICGV 15033 [(ICGV 06420 × Sun Oleic 95R) F_2P402-P1-P1-B1-B1] (Janila et al., 2016b) with an oleic acid content of ca. 82% was used as a donor parent for high oleic trait, while ICGV 13193 {ICGV 91114-P1 × [ICGV 91114-P1 × (ICGV 91114-P1 × GPBD 4-P1_13-1)]} (Varshney et al., 2014; Janila et al., 2016a) with rust and LLS score of 2.0 and 1.0, respectively, at 75 days after sowing (DAS) was used as a donor parent for rust and LLS resistance.

Hybridization and Generation Advancement

Two independent crosses were made with K 6 as a pistillate parent and ICGV 15033, and ICGV 13193 as pollen parents (**Figure 1**). The unopened well-developed basal buds from the female parent were emasculated in the evening (15:00–18:30 h) and pollination was carried out in the next morning (06:00–08:00 h). For pollination, the fully bloomed flower from the male parent was plucked and pollens were gently deposited on the stigma of emasculated flower (Nigam et al., 1990). The population size of F_1s (384), intercrossed F_1s (441), BC_1F_1s (380), BC_1F_2s (195), and BC_1F_3s (343) was developed and genotyped to select desired allelic combination.

DNA Isolation and Genotyping With Linked Markers

The genomic DNA was isolated from the young leaf (50–100 mg) of F_1s, backcross progenies and parents using modified cetyl trimethyl ammonium bromide (CTAB) method (Mace et al., 2003). DNA was quantified by loading 1 μl on the 0.8% agarose gel containing 10 μl ethidium bromide (10 mg/ml) and run at 80V for 30–45 min. The agarose gel was documented under UV transilluminator. The final DNA quality and quantity were estimated using Nanodrop (Shimadzu UV160A, Japan) and diluted to 5 ng/μl concentration for polymerase chain reaction (PCR). The allele-specific molecular markers positioned on linkage group A09 and B09 were used to select *ahFAD2A* and *ahFAD2B* mutant alleles in F_1 generation during post-rainy 2016–17 (**Table 1** and **Figure 1**). To track QTL region for disease resistance, earlier reported simple sequence repeat (SSR) markers SEQ8D09 and GMLQ975 were used for LLS, and IPAHM103, GM2301, GMRQ786, and GMRQ843 for rust during post-rainy 2016–17 (Khedikar et al., 2010; Sujay et al., 2012; Pandey et al., 2017).

Genotyping for LLS and Rust-Resistant Alleles

The touchdown PCR profile was used for screening of LLS and rust-resistant alleles using linked markers. The reaction mixture consisted of 5 ng of DNA, 2 pmol of M13 labeled forward primer, 5 pmol of reverse primer, 2 mM $MgCl_2$, 2 mM dNTPs, 0.1 U of Taq DNA polymerase (Kappa Taq) and 1X PCR buffer. A standardized touch down PCR program with 5 min initial denaturation, followed by 5 cycles of 94°C for 20 s, 65°C for 20 s and 72°C for 30 s with 1°C decrement for every cycle, followed by 40 cycles of 94°C for 20 s, constant annealing temperature of 59°C and 72°C for 30 s ending with extension for 20 min at 72°C. The PCR products were resolved on 2% agarose gel for confirmation of amplification. In SSRs, forward primers were

TABLE 1 | List of markers used for genotyping of high oleic, late leaf spot, and rust resistance traits.

Trait	Marker name	Linkage group	Sequence (5′-3′)	Allele size (bp)	References
High oleic acid	F435-F (*ahFAD2A*)	A09	Forward: ATCCAAGGCTGCATTCTCAC Reverse (SUB): TGGGACAAACACTTCGTT	M- 203 W- 290	Chen et al., 2010
High oleic acid	F435 (*ahFAD2B*)	B09	Forward: ATCCAAGGCTGCATTCTCAC Reverse (INS): AACACTTCGTCGCGGTCT	M- 190, 290 W- 290	Chen et al., 2010
LLS	GMLQ975	A03	FP: GGTATCATGATGAATTTTTAGAAGACTAGG RP: GAAATTTGGCTTTGGGTTCA	R- 150, 280 S- 280	Sujay et al., 2012
Rust	IPAHM103	A03	FP: GCATTCACCACCATAGTCCA RP: TCCTCTGACTTTCCTCCATCA	R- 154 S- 130	Khedikar et al., 2010
Rust	GM2301	A03	FP: GTAACCACAGCTGGCATGAAC RP: TCTTCAAGAACCCACCAACAC	R- 235 S- 260	Sujay et al., 2012
Rust	GMRQ786 (Gene specific marker)	A03	FP: AACATTGTAACACTCACCTGGCTA RP: TCATGCTTGAACTGTGCCTC	R- 200	Pandey et al., 2017

FP: Forward primer, RP: Reverse primer, R: resistant, S: susceptible, M: mutant type, W: wild type.

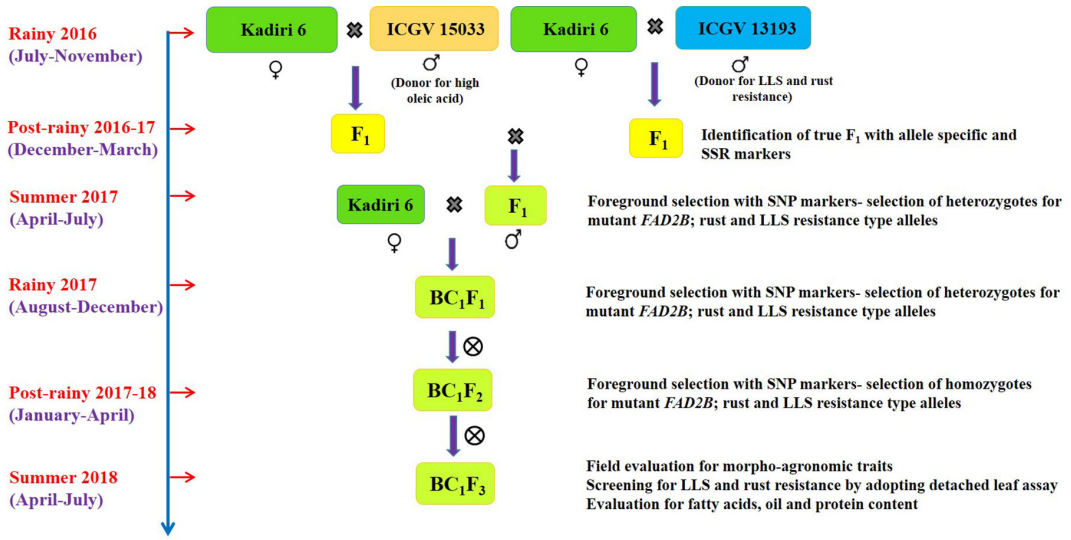

FIGURE 1 | Marker-assisted backcross breeding scheme for improvement of oil quality and resistance to late leaf spot and rust in groundnut *cv*. Kadiri 6.

TABLE 2 | Selected SNPs associated with late leaf spot, rust resistance and high oleic acid traits used for high throughput genotyping using KASP assay.

SNP ID	SNP code	Trait category	Genomic position (bp)	Target allele (donor type)	Alternate allele
snpAH0002	GKAMFAD2B	Oleic acid content	B sub-genome	A	-:-*
snpAH0015	GKAMA03QR786	LLS and rust resistance	133497786	T	A
snpAH0017	GKAMA03QR517	LLS and rust resistance	131739517	A	C
snpAH0018	GKAMA03QR796	LLS and rust resistance	131937796	G	A
snpAH0021	GKAMA03QR661	LLS and rust resistance	133527661	C	G
snpAH0026	GKAMA03GR173	LLS and rust resistance	134613173	G	C

-:- Null allele.

dye-labeled with FAM, VIC and NED which were detected as blue, green and black color peaks, respectively, upon capillary electrophoresis. The PCR products were denatured and capillary separated with ABI 3700 automatic DNA sequencer (Applied Biosystems, United States) and GeneMapper Software V (Applied Biosystems) was used to analyze the peak patterns.

Genotyping for Mutant *ahFAD2* Alleles Controlling High Oleic Trait

The touchdown PCR program described above was used for allele-specific markers. The PCR products were resolved on 2% agarose gel for confirming the amplification. For *ahFAD2A* allele, 203 bp fragment was amplified by F435-F and F435SUB-R in mutant alleles (substitution of G:C to A:T). For *ahFAD2B*, F435-F and F435INS-R amplified 195 bp fragment with insertion mutation (A:T insertion). The primer combination of F435-F and F435-IC-R was used to amplify the 250 bp wild type allele of *ahFAD2A* and *ahFAD2B*.

Genotyping Using KASP Based SNP Markers

After post-rainy 2016–17, the diagnostic SNPs were utilized for genotyping the breeding population. The SNP markers used in the study consisted one SNP for *ahFAD2B* positioned at +442 bp in the ORF on 9th chromosome of B genome associated with high oleic trait (Pandey et al., unpublished data). The validated five diagnostic SNPs present on A03 (**Table 2** and **Figure 2**) developed by Pandey et al. (2017) associated with LLS and rust resistance were used. The two leaf disks of 5 mm each were collected from each plant for genotyping using Kompetitive allele-specific PCR (KASP ^{TM}) assay.

Phenotypic Evaluation for LLS and Rust Resistance Using Detached Leaf Assay

A set of BC_1F_3's along with resistant checks GPBD 4 and ICGV 13193 and susceptible checks Kadiri 6, and TMV 2 were selected for the foliar fungal disease screening using the detached leaf assay (Foster et al., 1980). The fully expanded quadrifoliate third/fourth leaves from the top of a 40-days old plant were excised and washed with deionized water to use as an explant for detached leaf assay. The mixture of sand: vermiculite mixture (1:1 v/v) was prepared and steam-sterilized at 15 lbs pressure and 121°C for 1 hr in two cycles of the autoclave. The experiment was conducted in two replications following a completely randomized design (CRD). The excised explants were implanted in sterile culture (4 cm

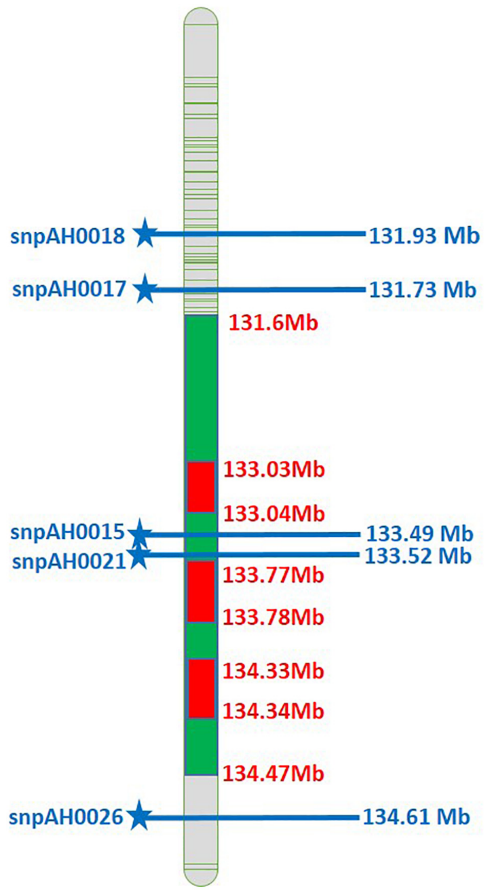

FIGURE 2 | Physical map of A03 comprising the SNPs associated with LLS and rust resistance (Green color represents the QTLs for rust and LLS resistance; Blue color SNPs indicates the physical position of SNPs associated with LLS and rust resistance).

thick sand) in plastic trays (40 × 30 cm) and sprayed with LLS and rust inoculum (30,000 spores ml^{-1}). Transparent polythene sheets were used to cover the trays and incubated in the humidified growth chamber at 24°C with higher relative humidity (∼85%). The leaflets were sprayed once in a day with deionized water and the process was continued up to 8 days after inoculation (DAI). The progress of disease was monitored from 5 to 35 DAI. For LLS resistance, incubation period (IP), latent period (LP), lesion per leaflet (LPL), leaf area damaged in percentage (%LAD), sporulation period (SP) and lesion diameter (LD) were recorded. The parameters *viz*., IP, LP, leaf area damage (LAD), sporulation index and disease scores were recorded to quantify rust resistance. The rust and LLS disease scores were evaluated using modified 1–9 scale at 35 DAI (Subrahmanyam et al., 1995).

Phenotyping for Morpho-Agronomic and Kernel Quality Traits

The field evaluation of the backcrossed lines was carried out during the post-rainy season of 2018 at red campus West 18 field-block (17.51° North latitude and 78.27° East longitude) under Alfisols (pH 7.0–7.5). Seeds were treated with mancozeb @ 2 g/kg seed and imidachloprid @ 2 ml/kg seed. The air temperature ranged between 22 and 36°C and relative humidity varied between 34 and 87% during post-rainy 2018. The backcrossed lines were planted in 2 m rows. Row to row spacing was maintained at 30 cm while the plant to plant spacing in a row was 10 cm. The basal fertilizer dose of 60 kg phosphorus pent-oxide (P_2O_5) was applied. The gypsum was applied @ 400 kg/ha at peak flowering stage of the crop. The lines were harvested at physiological maturity and pods were dried in sunlight for 3 days. The pods were stripped separately from each progeny. The mature-pods from each line were shelled and obtained kernels were used for quality analysis.

The morpho-agronomic traits *viz*., days to 50% flowering, plant height, pod yield per plant, kernel yield per plant, number of pods per plant, shelling outturn, pod length, pod width, pod constriction, pod reticulation, pod beak and pod ridges were recorded by following the groundnut descriptor (IBPGR and ICRISAT, 1992). The oil, protein, and fatty acids were estimated using near-infrared reflectance spectroscopy (NIRS) (XDS monochromator, FOSS Analytical AB, Sweden) for simple, non-destructive, economic and fast screening of progenies. The calibration equation used in prediction had high values of coefficient of determination in external validation (r^2) for oleic acid (r^2 = 0.96), linoleic acid (r^2 = 0.96), and moderate for oil (r^2 = 0.89), protein (r^2 = 0.83) and palmitic acid (r^2 = 0.80) (Murali et al., unpublished data). Seeds obtained from the single plant were scanned twice and data are represented in percentage (%).

RESULTS

Marker-Assisted Breeding Scheme

Two independent crosses were attempted with K 6 as a pistillate parent and ICGV 15033 (oleic acid ca. 82.0%), and ICGV 13193 (LLS and rust-resistant line) as pollen parents to generate F_1s during rainy 2016. The seeds obtained from the high oleic cross (cross I) were used to raise 210 F_1s (**Table 3**) and screened for hybridity using *ahFAD2A* and *ahFAD2B* allele-specific makers. A total of 151 (72.0%) plants showed heterozygous mutant alleles for *ahFAD2A* and *ahFAD2B*. For LLS and rust resistance cross (cross II), 128 F_1 plants (74.0%) were confirmed with marker GMLQ975 linked to LLS resistance (**Figure 3A**) and marker GM2301 linked to rust resistance. The amplified PCR product for marker GMLQ975 showed LLS resistant alleles of 150 bp and a susceptible allele of 280 bp. The PCR products amplified using SSR marker GM2301 were separated using capillary electrophoresis followed by electropherogram assay (**Figure 3B**) which could differentiate recipient and donor type alleles with 260 bp and 235 bp, respectively, and true hybrids showed both the alleles. Positive F_1s carrying mutant *ahFAD2* alleles, and QTLs for LLS and rust resistance were intercrossed to combine high oleic, LLS and rust resistance. A total of 580 intercrossed F_1 seeds were obtained of which 441 seeds germinated. The

TABLE 3 | Summary of hybridization program and marker assisted selection for high oleic trait, LLS and rust resistance in Kadiri 6 derived backcross lines.

Crop season	Cross details	Seedling raised	Confirmed plants for target alleles	Hybridity test output (%)
Post-rainy 2016-17	K6 × ICGV 15033 (Cross I)*	210	151	72
	K6 × ICGV 13193 (Cross II)*	174	128	74
Summer 2017	Cross I × Cross II (Intercross- IC)†	441	44	10
Rainy 2017	K 6 × IC-F_1 (Backcross 1)†	380	41	10
Post-rainy 2017–18	$BC_1F_2^{\dagger}$	195**	16	8

Plants selected based on gel based markers; †Plants selected based on SNP call.

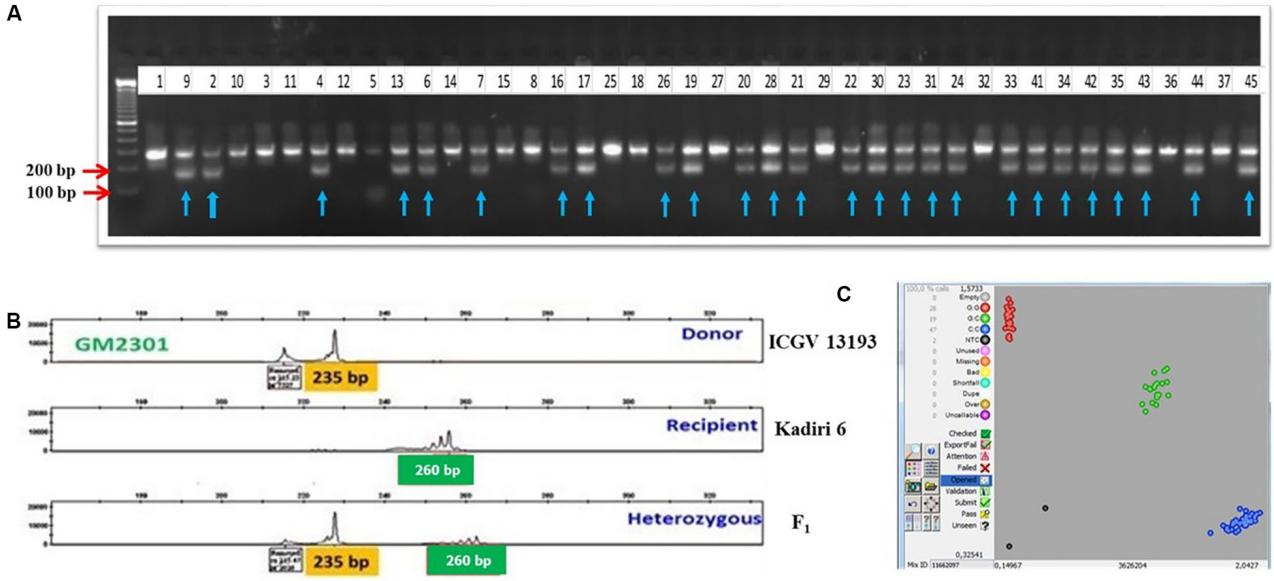

FIGURE 3 | (A) Confirmation of zygosity in F_1 plants derived from a cross between Kadiri 6 and ICGV 13193 using marker GMLQ975 for LLS resistance; Lane 1: Recipient (Kadiri 6), Lane 2: Donor (ICGV 13193), 3–45: F_1 plants, with arrows indicated for true hybrids. **(B)** Electropherogram showing the resistant type allele (235 bp), susceptible type allele (260 bp) and heterozygous with both the alleles amplified using SSR marker GM2301 for rust resistance. **(C)** Kompetitive Allele Specific PCR (KASP) graphs of SNPs genotyping with snpAH0021 for LLS and rust resistance; (The scatter plot with axes *x* and *y* represents allelic discrimination. Red, green and blue dots represent the susceptible homozygous, heterozygous, and resistant homozygous alleles, respectively).

441 seedlings were genotyped using SNP markers which revealed 44 (10%) true hybrid plants for the high oleic, LLS and rust resistance alleles. **Figure 3C** represents the differentiation of homozygous resistant (blue dots), heterozygous types (green dots) and homozygous susceptible types (red dots) for snpAH0021 using SNP markers. The positive F_1 plants (44) were used as a pollen parent to make the backcross with the pistillate parent K 6 and 415 BC_1F_1 seeds were harvested.

The hybridity test carried out with SNP markers for 380 BC_1F_1s resulted in the identification of 41 positive BC_1F_1s, which produced 396 BC_1F_2 seeds after self-pollination in rainy 2017. In post-rainy 2017–18, 195 BC_1F_2 seedlings were raised and screened for zygosity using SNP markers. The assay revealed 16 plants with homozygous alleles for LLS, rust, and *ahFAD2B* allele which were advanced through selfing to generate BC_1F_3 seeds. In post-rainy 2018, these 16 BC_1F_3s were raised in the field and evaluated for oil quality, LLS and rust resistance, pod and kernel features. The summary of the data is presented in **Table 4**.

Evaluation for LLS and Rust Resistance

The results on the screening of fully expanded quadrifoliate leaves using detached leaf assay distinguished resistance, moderate resistance, and susceptibility for LLS and rust (**Figure 4A**).

Late Leaf Spot

The components of LLS resistance were recorded to quantify the resistance of BC_1F_3s along with parents and resistant (GBPD 4) and susceptible (TMV 2) checks. The incubation period (IP) varied from 9.5 to 12.8 DAI in backcross lines with highest of 12.8 DAI was recorded in the line, BC_1F_3-327, while recurrent parent K 6 recorded 7.0 DAI (**Table 5**). The sporulation period (SP) varied from 21.5 to 24.7 DAI. The BC_1F_3 lines, 144, 186, and 269 showed longer SP of 24.5 to 24.7 DAI as compared to recipient parent, K 6 (15.5 DAI) and susceptible check, TMV 2 (13.5 DAI). The average number of lesions per leaflet (LPL) varied from 19.5 to 78.5 at 35 DAI; the lines BC_1F_3-4 (19.5), BC_1F_3-5 (27.0), BC_1F_3-120 (38.0), and BC_1F_3-34 (44.0) recorded a LPS comparable with the donor parent, ICGV 13193 with LPL of 32.5. The recurrent parent recorded an average of 140 LPL.

TABLE 4 | Summary of Kompetitive allele-specific PCR (KASP) assay for target traits.

SNP combinations	Total genotypes in BC_1F_2	Rust score (in BC_1F_3)*	LLS score (in BC_1F_3)†	Oleic acid (%) (in BC_1F_3)¥
All six target SNPs	16	3–4	3–4	61.2–83.4
High oleic SNP	71	4–7	4–7	60.4–83.4
LLS and rust resistance associated SNPs	45	3–4	3–5	40.6–55.2
Null alleles for LLS and rust resistance and high oleic trait	211	5–8	5–8	38.4–52.6

**-Range of rust score at 35 days after inoculation (DAI), †-range of LLS score at 35 DAI, ¥-Range oleic acid content in selected genotypes based on presence of SNPs for target traits.*

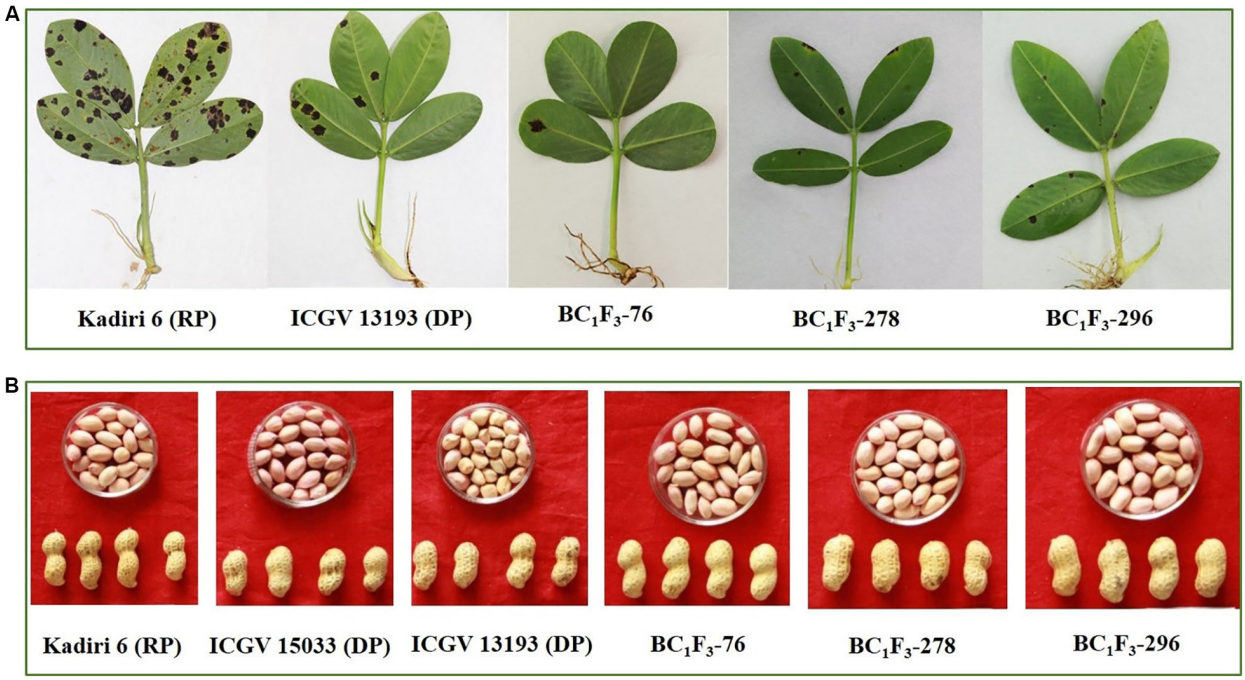

FIGURE 4 | (A) Late leaf spot and rust reaction of the recurrent parent (Kadiri 6), donor parent (ICGV 13193) and backcrossed lines at 35 DAI (days after inoculation). **(B)** Pod and kernel features of selected Kadiri 6 derived backcrossed lines.

The lesion diameter (LD) in 16 BC_1F_3 lines varied from 2.1 to 4.6 mm, while it was 5.7 mm in K 6 and 1.6 mm in ICGV 13193. The lines, BC_1F_3-76 (2.1 mm), BC_1F_3-296 (2.2 mm), and BC_1F_3-225 (2.3 mm) recorded low LD. The percent leaf area damaged by LLS (%LADL) was lower in BC_1F_3-296 (15.5%) and BC_1F_3-216 (19.5%). The parental lines, K 6 and ICGV 13193 recorded 61.0 and 21.0% LADL, respectively. The disease score of LLS in 16 BC_1F_3 lines varied from 3.0 to 5.0. The lines BC_1F_3s-76, 216, 225, and 311 were found to be resistant for LLS with a score of $\leq$3.0 which is at par with donor parent ICGV 13193 (3.0), whereas K 6 recorded a disease score of 8.0 at 35 DAI.

Rust

The components of rust resistance were recorded to quantify the level of rust resistance in 16 backcrossed lines. The data of each BC_1F_3 lines for various resistance components in comparison with donor and recipient parent are presented in **Table 5**. The parental lines K 6 and ICGV 13193 recorded IP of 9.5 and 14 DAI, respectively. The IP for rust among BC_1F_3s varied from 10.0 to13.5 DAI; long IP of 13.5 DAI was observed in the BC_1F_3s-5, 120 and 186. The sporulation period (SP) varied from 15.5 to19.5 DAI; four BC_1F_3 lines, 5, 120, 225, and 249 recorded an SP of 19.5 DAI, comparable to the SP of donor parent, ICGV 13193 (19 DAI), whereas the recurrent parent, K 6 recorded an SP of 12.0 DAI. The number of rust pustules per leaflet (PPL) varied from 19.5 to 68.5. Two lines, BC_1F_3 225 (19.5) and BC_1F_3-5 (21) recorded the lowest rust PPL comparable with donor parent ICGV 13193 (19.5 PPL), while the RP, K 6 recorded 98.5 rust PPL. The percent leaf area damaged by rust (%LADR) varied from 9.0 to 24.5%; three lines, BC_1F_3-249 (9.0%), BC_1F_3-144 (9.5%), and BC_1F_3-269 (11.0%) recorded low% LADR. The parental lines K 6 and ICGV 13193 recorded a%LADR of 25.0 and 8.5%, respectively. The rust disease score of 3.0 and 4.0 were recorded in BC_1F_3 lines, comparable with the score of DP, ICGV 13193 (4.0), whereas the RP, K 6 recorded a score of 7.0 at 35 DAI.

TABLE 5 | Components of resistance and disease score of Kadiri 6 derived backcross lines (BC_1F_3) against late leaf spot (LLS) and rust.

		Late leaf spot						Rust				
S. No.	Genotypes	IP (DAI)	SP (DAI)	LPL (35 DAI)	LD (mm)	% LADL	LLS score (35 DAI)	IP (DAI)	SP (DAI)	PPL (35 DAI)	%LADR	Rust score (35 DAI)
1.	BC_1F_3-4	11.5	23.5	19.5	3.3	21.0	4.0	12.5	18.5	23.0	15.5	3.0
2.	BC_1F_3- 5	11.5	22.9	27.0	3.5	24.5	4.0	13.5	19.5	21.0	13.0	4.0
3.	BC_1F_3- 34	10.0	21.9	44.0	4.5	24.0	4.0	11.5	17.5	21.5	19.0	4.0
4.	BC_1F_3-76	10.6	22.7	74.0	2.1	37.7	3.0	11.0	17.0	66.5	24.5	4.0
5.	BC_1F_3-120	12.0	23.5	38.0	2.8	24.5	5.0	13.5	19.5	25.0	11.5	4.0
6.	BC_1F_3-144	11.5	24.5	72.5	3.5	22.5	4.0	12.5	17.5	22.5	9.5	4.0
7.	BC_1F_3-186	9.9	24.5	62.0	2.5	22.7	4.0	13.5	18.0	25.0	21.0	3.0
8.	BC_1F_3-214	9.5	22.5	61.5	3.1	24.0	4.0	12.5	16.0	30.5	18.5	4.0
9.	BC_1F_3-216	11.5	23.4	76.5	3.4	19.5	3.0	10.0	17.5	26.0	14.5	4.0
10.	BC_1F_3-225	11.7	22.7	78.5	2.3	32.5	3.0	12.5	19.5	19.5	12.5	4.0
11.	BC_1F_3-249	12.6	23.5	68.6	3.4	21.0	4.0	11.5	19.5	21.5	9.0	4.0
12.	BC_1F_3-269	11.6	24.7	63.5	4.6	21.0	4.0	10.5	16.0	37.5	11.0	4.0
13.	BC_1F_3-278	9.9	23.5	61.0	2.5	22.5	4.0	12.0	18.0	68.5	20.5	4.0
14.	BC_1F_3-296	10.8	21.5	61.5	2.2	15.5	4.0	11.5	15.5	52.0	11.5	4.0
15.	BC_1F_3-311	11.9	22.5	55.0	3.1	25.5	3.0	12.5	18.0	29.5	12.5	4.0
16.	BC_1F_3-327	12.8	21.5	44.5	4.2	23.0	4.0	12.5	16.5	24.0	13.5	3.0
Checks	TMV 2	7.1	13.5	153.0	6.0	64.5	9.0	8.5	10.5	122.5	40.0	9.0
	GPBD 4	15.6	25.7	13.5	1.2	20.7	3.0	16.5	22.5	12.5	7.5	3.0
DP for LLS & rust	ICGV 13193	11.7	23.5	32.5	1.6	21.0	3.0	14.0	19.0	19.5	8.5	4.0
DP for high O/L	ICGV 15033	9.9	18.5	93.0	4.3	33.5	5.0	12.5	15.5	22.5	10.5	4.0
Recurrent parent	Kadiri 6	7.0	15.5	140.0	5.7	61.0	8.0	9.5	12.0	98.5	25.0	7.0
	SEm ±	1.3	0.4	2.7	0.3	1.2	0.4	0.4	1.1	2.3	1.3	0.4
	LSD (5%)	1.1	1.3	8.0	0.8	3.5	1.2	1.3	3.2	6.8	3.8	1.2

IP: Incubation Period (days after inoculation), SP: Sporulation Period (days after inoculation), LPL: Lesion Per Leaflet, LD: Lesion Diameter, % LADL: percent Leaf Area Damaged by LLS, PPL: Pustule Per Leaf, %LADR: percent Leaf Area Damaged by Rust, DAI: Days After Inoculation, SEm: Standard Error of the Mean, LSD: Least Significant Difference, DP: Donor parent, LLS: late leaf spot.

Evaluation of Backcrossed Lines for Morpho-Agronomic and Kernel Quality Traits

Field screening identified individuals resembling recurrent parent K 6 for morpho-agronomic and kernel features which resulted in the selection of recombinants with at par to superior performance for the selected traits.

Morpho-Agronomic Traits

The morpho-agronomic performance of 16 BC_1F_3 lines of K 6 is presented in **Table 6**. Days to 50% flowering (DFF) ranged from 31–38 days with an average of 34 days whereas recurrent parent K 6 recorded 35 DFF. The mean plant height was 26.29 cm; it varied from 17.0 cm (BC_1F_3-186) to 38.4 cm (BC_1F_3-214) and K 6 recorded plant height of 34.6 cm. The mature pods per plant (NPP) varied from 24 to 58; BC_1F_3-269 (58) followed by BC_1F_3-120 (55) and BC_1F_3-144 (54) recorded highest NPP. The weight of mature pods per plant (PYP) varied from 17.2 (BC_1F_3-214) to 40.8 g (BC_1F_3-269) in comparison with K 6 (26.8 g).

The mean seed weight per plant (SYP) was 21.3 g with a range of 11.8–32.2 g. The highest SYP was recorded in BC_1F_3-269 (32.2 g) as compared to K 6 (19.8 g). Shelling out-turn (SH) varied from 65.3 to 78.9% with a mean of 72.4%. The higher SH was recorded in BC_1F_3-269 (78.9%) followed by BC_1F_3-186 (78.6%) and BC_1F_3-34 (78.3%). The length of mature pods (PL) in BC_1F_3s-214 and 327 (27.4 mm) was comparable to K 6 (24.5 mm). The width of mature pods (PW) was measured 8.7 mm (BC_1F_3-76) to 11.9 mm (BC_1F_3-34). Recurrent parent K 6 recorded an average PW of 11.5 mm. The recurrent parent had pods with moderate constriction. In the study, a total of 14 lines were observed with slight pod constriction whereas two lines (BC_1F_3-34 and BC_1F_3-327) were observed with moderate constriction similar to K 6. Presence of slight constriction is preferred over moderate and/or prominent constriction, hence the selection of pods with lesser constriction is preferred to further improve K 6 for preferred pod features along with target traits. Pod reticulation was slight in the selected lines. Pod beak was slight in two lines (BC_1F_3s-76 and 296) and medium in BC_1F_3-278. Similarly, pod ridges were absent in BC_1F_3s- 278 and 296 and, whereas slight in BC_1F_3-76.

Kernel Quality Traits

Oil Content (%)

The kernel oil content in back cross lines varied from 46.0 to 54.9% and the recurrent parent, K 6 recorded an oil content of 49.5%. The higher oil content (>50.0%) was observed in 11 lines (BC_1F_3s-5, 76, 120, 144, 186, 214, 225, 249, 278, 296, 327).

TABLE 6 | Morpho-agronomic and pod traits of backcross lines of K 6 evaluated during the post-rainy of 2018 at ICRISAT, Patancheru.

S. No.	Genotypes	DFF	PH (cm)	NPP	PYP (g)	SYP (g)	SH (%)	PL (mm)	PW (mm)	PCN	PR	PB	PRG
1.	BC_1F_3-4	32.0	29.5	29.0	20.7	14.6	72.0	25.1	10.7	S	S	S	M
2.	BC_1F_3-5	32.0	18.5	30.0	28.7	21.5	74.9	25.8	10.6	S	S	S	A
3.	BC_1F_3-34	35.0	27.5	30.0	24.0	18.8	78.3	25.9	11.9	M	M	M	S
4.	BC_1F_3-76	35.0	21.0	38.0	28.1	20.5	73.0	18.8	8.7	S	S	S	S
5.	BC_1F_3-120	34.0	24.5	55.0	33.8	25.1	74.3	19.5	9.3	S	S	S	A
6.	BC_1F_3-144	31.0	19.0	54.0	38.0	24.8	65.3	26.4	11.7	S	S	S	S
7.	BC_1F_3-186	32.0	17.0	24.0	19.6	15.4	78.6	24.9	10.5	S	S	S	A
8.	BC_1F_3-214	34.0	38.4	27.0	17.2	11.8	68.6	27.4	8.9	S	M	S	M
9.	BC_1F_3-216	35.0	22.0	38.0	29.1	21.9	74.0	19.8	11.7	S	M	S	S
10.	BC_1F_3-225	31.0	29.2	40.0	31.4	24.2	70.0	20.5	10.8	S	S	S	S
11.	BC_1F_3-249	32.0	30.4	31.0	25.9	22.0	68.3	21.4	9.9	S	S	S	S
12.	BC_1F_3-269	38.0	18.0	58.0	40.8	32.2	78.9	23.0	10.1	S	S	S	A
13.	BC_1F_3-278	34.0	29.0	45.0	39.7	29.2	73.6	20.7	9.1	S	S	M	A
14.	BC_1F_3-296	36.0	32.5	28.0	21.5	15.7	73.0	25.1	10.5	S	S	S	A
15.	BC_1F_3-311	32.0	31.2	42.0	32.2	23.2	72.0	25.5	9.8	S	S	S	A
16.	BC_1F_3-327	34.0	33.0	30.0	30.2	20.8	68.9	27.4	8.9	M	S	S	S
17.	ICGV 13193 (DP for LLS & rust)	37.0	29.5	41.0	32.7	28.2	74.0	21.4	11.0	S	S	S	S
18.	ICGV 15033 (DP for high O/L)	36.0	30.0	33.0	25.3	19.3	76.3	20.5	9.8	S	S	S	A
19.	Kadiri 6 (recurrent parent)	35.0	34.6	30.0	26.8	19.8	73.9	24.5	11.5	M	M	P	M
	Mean	34.0	26.2	38.5	29.4	21.3	72.4	23.9	9.9	–	–	–	–
	Range	31.0–38.0	17.0–38.4	24.0–58.0	17.2–40.8	11.8–32.2	65.3–78.9	18.8–27.4	8.7–11.9	–	–	–	–

DFF: Days to 50% Flowering, PH: Plant Height (cm), NPP: Number of Pods per Plant, PYP: Pod Yield per Plant (g), SYP: Seed Yield per Plant (g), SH (%): Shelling outturn (%), PL: Pod Length (mm); PW: Pod Width (mm), PCN: Pod Constriction, PR: Pod Reticulation, PB: Pod Beak, PRG: Pod Ridges, S: Slight, M: Medium, P: Prominent, A: Absent, DP: donor parent, LLS: late leaf spot.

Fatty Acid Content (%)

The oleic, linoleic, and palmitic acid were estimated using NIRS. The oleic acid content varied from 62.5 to 83.4% and the linoleic acid varied from 25.4 to 43.5%. The oleic acid content of >80% was observed in three lines BC_1F_3-76 (81.3%), BC_1F_3-278 (82.8%) and BC_1F_3-296 (83.4%). The recurrent parent, K6 recorded an oleic acid content of 39.8%. The selected lines recorded lower linoleic acid of 9.3% compared to the recurrent parent, K 6 (37.4%). The oleic: linoleic acid (O/L) ratio was 5.2 to 17.1; the ratio in the selected lines, BC_1F_3-76, 278 and 296 is >10.

The palmitic acid varied from 7.5 to 11.1% with an average of 9.6% indicating a reduction to an extent of 2.2–5.8% compared to the recurrent parent, K 6 (13.3%). The high oil content with elevated oleic acid and reduced linoleic and palmitic acid is desirable for the improvement in oil quality of groundnut. The present study identified three lines BC_1F_3s-76, 278 and 296 with high oleic acid (≥80.0%) and high oil (>50.0%) content indicating that these segregants can be used to develop a high oil and oleic variant of K 6.

Protein Content (%)

The protein content of the BC_1F_3 lines varied between 23.3 and 29.5%; the protein content of selected lines is BC_1F_3-278 (29.5%), BC_1F_3-216 (28.9%) and BC_1F_3-327 (28.1%), while that of RP, K 6 is 26.6%.

Superior Segregants for Late Leaf Spot and Rust Resistance, Agronomic, and Kernel Quality Traits

The combined approach MABC and phenotypic-based screening in BC_1F_3s was found to be effective in developing and selecting superior segregants for LLS and rust resistance, desired pod features, and improved oil quality traits. The lines BC_1F_3-76, BC_1F_3-278 and BC_1F_3-296 were found at par to marginally superior to the recurrent parent (**Table 7** and **Figure 4B**). The LLS and rust score of the selected lines at 35 DAI varied from 3.0 to 4.0, and 4.0, respectively, indicating good level of resistance. The day to 50% flowering in lines was 35 days in BC_1F_3-76, 34 days in BC_1F_3-278 and 36 days in BC_1F_3-296 comparable to the recurrent parent, K 6 (35 days). The number of pods per plant was 38 and 45 for BC_1F_3-76 and BC_1F_3-278, respectively, as against 30 for K 6. The selected lines had pod yields ranging from 21.5 to 39.7 g per plant, while it was 26.8 g for K 6. The pod characteristics including pod constriction and pod reticulation were slight in all three selected backcross lines, pod beak was slight in BC_1F_3s-76 and 296 and medium in BC_1F_3-278, pod ridges were absent in

TABLE 7 | Late leaf spot and rust reaction, agronomic performance, and pod characteristics of selected superior lines.

Genotypes	Disease score (35 DAI)		Agronomic traits and oleic acid content							Pod characteristics					
	LLS score	Rust score	DFF	PH	NPP	PYP	SYP	SH	OA	PL	PW	PCN	PR	PB	PRG
BC_1F_3-76	3.0	4.0	35.0	21.0	38.0	28.1	20.5	73.0	81.3	18.8	8.7	S	S	S	S
BC_1F_3-278	4.0	4.0	34.0	29.0	45.0	39.7	29.2	73.6	82.8	20.7	9.1	S	S	M	A
BC_1F_3-296	4.0	4.0	36.0	32.5	28.0	21.5	15.7	73.0	83.4	25.1	10.5	S	S	S	A
ICGV 13193	3.0	4.0	37.0	29.5	41.0	32.7	28.2	74.0	44.6	21.4	11.0	S	S	S	S
ICGV 15033	5.0	4.0	36.0	30.0	33.0	25.3	19.3	76.2	82.4	20.5	9.8	S	S	S	A
Kadiri 6	8.0	7.0	35.0	34.6	30.0	26.8	19.8	73.8	39.8	24.5	11.5	M	M	P	M

DAI: Days After Inoculation, LLS: Late Leaf Spot, DFF: Days to 50% Flowering, PH: Plant Height (cm), NPP: Number of Pods Per Plant, PYP: Pod Yield per Plant (g), SYP: Seed Yield per Plant (g), SH: Shelling Outturn (%), OA (%): Oleic acid content, PL: Pod Length (mm); PW: Pod Width (mm), PCN: Pod Constriction, PR: Pod Reticulation, PB: Pod Beak, PRG: Pod Ridges, S: Slight, M: Medium, P: Prominent, A: Absent.

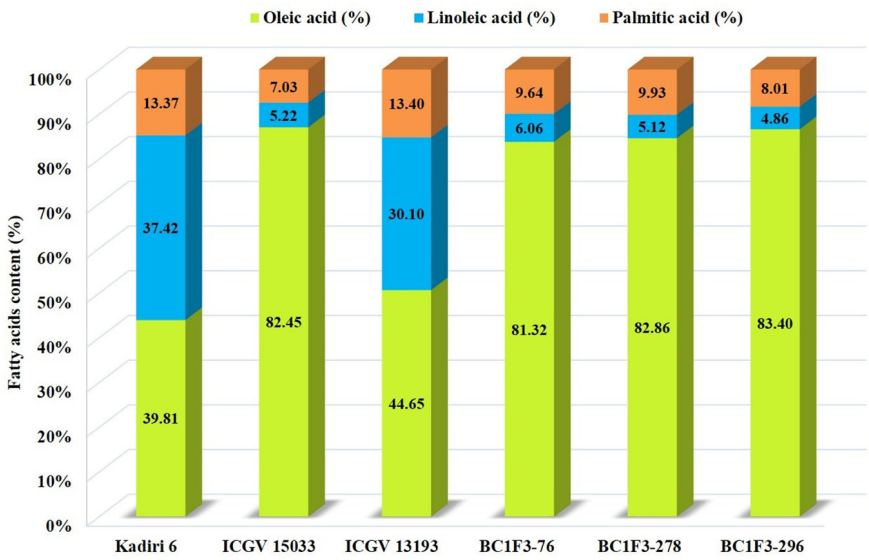

FIGURE 5 | Fatty acid profiles of parental lines and selected backcrossed lines.

BC_1F_3s-278 and 296 whereas slightly present in BC_1F_3-76. The oleic acid content in the selected lines was high and ranged from 81.3 to 83.4% (**Figure 5**). All the selected lines showed higher oil content (>50.0%), protein content (>26.0%) (**Figure 6**), oleic acid (>80.0%) and reduced linoleic acid content (<10.0%).

DISCUSSION

To meet the market needs of groundnut, kernel quality is an important trait and combining it with productivity traits such as biotic and abiotic stress tolerance is an important objective of groundnut breeding program. The high oleic trait is preferred by the food industry for increased shelf-life and consumer health benefits. The co-occurrence of *P. personata* and *P. arachidis* pathogens causes necrotic lesions, defoliation, and rust pustules, respectively, reducing the pod yield by 50–70%, and indirectly affecting kernel and haulm quality.

In the present study, the allele-specific markers were utilized to distinguish mutant and wild-type alleles of *ahFAD2A* and *ahFAD2B* for the high oleic trait (Chen et al., 2010). These markers were successfully deployed in groundnut breeding programs in the USA, Argentina, China and India to develop high oleic groundnut varieties (Chu et al., 2011; Mienie and Pretorius, 2013; Xiuzhen et al., 2016; Janila et al., 2016b; Bera et al., 2018; Shasidhar et al., 2020). For LLS and rust resistance, SSR markers (Sujay et al., 2012) were utilized during the first season of the crossing program. The KASP assay facilitates the plant breeding program with the use of a few SNPs to screen many samples and has an advantage of cost, simplicity, and speed (Chopra et al., 2018). The KASP assay has been successfully deployed in groundnut for rust resistance (Leal-Bertioli et al., 2015), nematode resistance (Chu et al., 2016), high oleic trait (Chopra et al., 2015; Zhao et al., 2017), bacterial wilt resistance (Zhao et al., 2016), LLS resistance (Clevenger et al., 2018), plant architectural traits (Chopra et al., 2018), early leaf spot and LLS, and TSWV (Agarwal et al., 2018) and for LLS, rust and high oleic traits at ICRISAT (Murali et al., 2019).

The gene introgression using MAS/MABC has been deployed in many crops to improve an elite cultivar through introgression

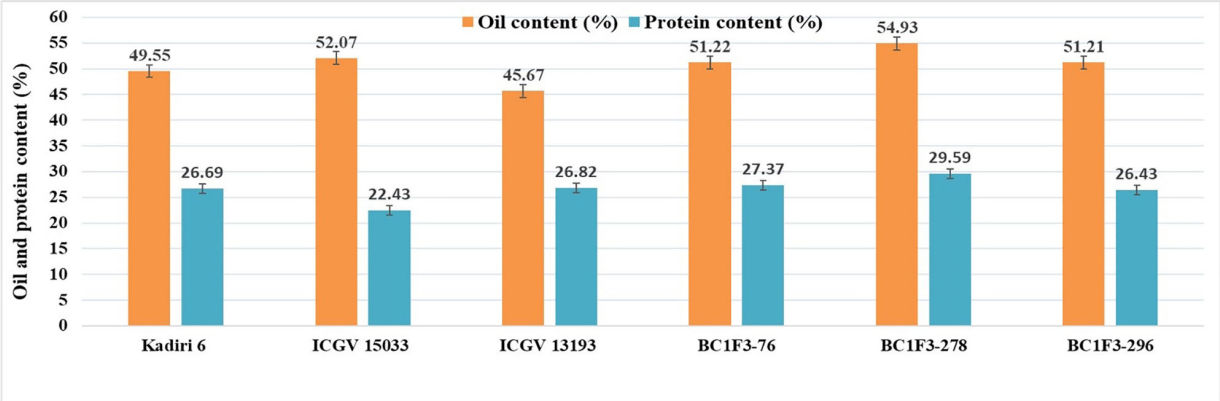

FIGURE 6 | Oil and protein contents of selected Kadiri 6 derived backcrossed lines.

of target traits from one or many donors (Tyagi et al., 2014). The parallel crossing of a recurrent parent with donor parents and intercrossing F_1s followed by backcross with a recurrent parent is a promising way of combine multiple traits. This scheme was utilized in combining traits like bacterial leaf blight (BLB) and lepidopteran insect pest tolerance in rice (Jiang et al., 2004), BLB in rice (Shanti et al., 2010), wheat stripe rust and leaf rust resistance genes (Santra et al., 2006; Yadawad et al., 2017), barley leaf rust, net blotch and spot blotch resistance (Hickey et al., 2017).

This is the most reliable method as it reduces the time period to combine genes/QTLs and assures the fixation of genes/QTLs (Joshi and Nayak, 2010). The normal breeding of an elite cultivar typically takes 6–8 breeding cycles and another 3–5 years of testing for release or commercialization. The use of molecular markers, and improved high-throughput phenotyping methods help the breeders to develop new cultivars in a shorter time span thereby enhancing the rate of genetic gain. We successfully introgressed high oleic, and resistance to LLS and rust trait into popular cultivar K 6 within a 2-year period using high-throughput genotyping and phenotyping coupled with rapid generation advancement. In this study, the breeding cycles per year has been increased from two to three. This resulted in a reduction in breeding cycle time (L) from 0.5 (two cycles per year) to 0.33 (3 cycles per year), thus significantly enhancing the rate of genetic gain.

In MABC, up to four backcrosses are suggested to retain the yield and quality features of the recurrent parent (Sundaram et al., 2008). Moreover, extra cycles of backcrossing have leverage when the adaptation of donor genotypes (wild relatives/germplasm/landraces) is poor. In the present study, the donor parents, ICGVs 15033 and 13193 possess good agronomic characters with a higher yield potential (Janila et al., 2016a,b). Even though repeated cycles of backcrossing can increase the recurrent parent genome recovery, each additional backcrossing could result in the reduction of minor QTLs/genes that leads to moderate expression of target traits in introgressed lines (Wang et al., 2009). Kolekar et al. (2017) observed more recovery of resistant progenies for rust over the LLS, especially when the target trait is controlled by a combination of large and small effect QTLs. A strategy of single backcrossing has been successfully used to introgress rust-resistant genes in wheat (Wang et al., 2009; Yadawad et al., 2017), leaf rust, net blotch, and spot blotch resistance in barley (Hickey et al., 2017), and LLS and rust resistance in groundnut (Kolekar et al., 2017). These studies indicated that one to two cycles of backcross is sufficient in improving elite cultivars while transferring the desired genes from an adapted donor parent. A single backcross can recover an average of 75% of the recipient genome.

The KASP genotyping platform enabled the early selection of plants for crossing thus facilitated generation advancement. The application of MAS in early cycles of breeding helps to detect the favorable introgressions from the donor parents and helps to reduce the population size (Collard and Mackill, 2007). We generated 380 BC_1F_1 plants from which 41 were selected based on presence all the three alleles. Even though the population of 195 BC_1F_2 plants was low to generate desirable segregants, hence we had a backup program where we have advanced heterozygotes too. Nonetheless, we could identify promising lines from the limited F_2 population which can be attributed to the use of two elite lines as parents in the crossing.

Detached leaf assay is a rapid and simple method used to screen groundnut genotypes for LLS resistance (Foster et al., 1980), and offers an advantage of off-season screening. In the present study, detached leaf assay was used to evaluate BC_1F_3s for LLS and rust resistance under controlled conditions. The detached leaves were placed in a culture of sand: vermiculite (1:1v/v) which has the advantages of maintenance of optimum moisture and relative humidity ($\geq$85%), with minimum contamination of saprophytes and secondary infection. Other commonly used methods include detached leaf culture in Hoagland's solution (Foster et al., 1980), whole seedling assay planted in pots (Dwivedi et al., 2002) and sand culture (Janila et al., 2013). For artificial inoculations, the spore concentration of ~30,000 conidia ml^{-1} was maintained. Motagi (2001) suggested the spore suspension of $\geq$20,000 conidia ml^{-1} is sufficient to build the proper inoculum load to screen groundnut entries for

LLS, and rust resistance. Dwivedi et al. (2002) reported that the disease reaction of detached leaflets inoculated with the artificial inoculum determined using disease incidence and severity values were similar to the reaction of the intact plant. This indicates the reliability of the resistant lines identified in the study.

The phenotyping for pod and kernel characteristics including kernel quality helps to select the progenies with a desirable combination of traits as most of these features are monogenically inherited (Bera and Das, 1998; Reddy et al., 2017). Pod and kernel features in groundnut are important for overall appearance and for fetching higher market value, hence considered as selection criteria along with higher pod yield. In the present study, three lines (BC_1F_3s-76, 278 and 296) were obtained with superior pod shape over the recurrent parent. Fourteen backcross lines featured with slight beak whereas two lines (BC_1F_3s-34 and 278) showed moderate beak. Selection of pods without beaks is an important feature as pods with beaks are susceptible to cracking in soil paving the way for the entry of *Aspergillus* group of fungi and subsequent aflatoxin contamination. After drying, pods sometimes crack toward the beak region making the kernels vulnerable for entry of storage pest and molds (Okello et al., 2010). The recipient parent K 6 had a prominent beak (**Figure 4B**), an undesirable feature that can be a potential source for the entry pathogens such as, *Aspergillus* Sp., hence priority was given to select the lines with pods with the slight beak.

The kernel oil, protein, and fatty acid content are important criteria in trade, food, and seed industries of groundnut. The wet chemical methods for the analysis of oil, protein and fatty acids content are accurate but operationally complex, time-consuming, destructive and expensive (Sundaram et al., 2009). The relative proportion of three fatty acids namely, palmitic, oleic, and linoleic acid determine the kernel quality and shelf-life. Janila et al. (2016b) and Bera et al. (2018) observed that the introgression of mutant *ahFAD2* alleles into elite breeding lines leads to increase in oleic acid (≥80.0%) and reduced accumulation of linoleic acid (≤8.0%) and palmitic acid (≤10.0%). The palmitic acid (C16:0), a saturated fatty acid contributes to ca. 10.0% of the kernel fat, and the remaining is shared by other minor fatty acids (Wang X. Z. et al., 2015). Normal groundnut oil (∼45% linoleic acid) is vulnerable to oxidation if stored for longer period, resulting in unpleasant smell, and taste of the oil and groundnut products. Hence, groundnut oil and its products with a high O/L are preferable (Vassiliou et al., 2009).

The high oleic acid content of ca. 80% is conferred by a combination of two mutant alleles, *ahFAD2A* and *ahFAD2b*. In the present study, the BC_1F_3 lines were selected based on the homozygosity for *ahFAD2B* mutant allele, hence the selected lines recorded an oleic acid content variation of 61.2–83.4%. Thus, the lines that carried *ahFAD2A* mutant allele along with *ahFAD2B* recorded high oleic acid content of ca. 80%. Selection for *ahFAD2A* allele was done by phenotype using NIRS in the present study. NIRS is a non-destructive technique of estimating fatty acids, which includes the standard error in prediction (of up to 5%), due to the lesser quantity and poor quality of seeds used in estimation. Insufficient quantity and poor quality of seeds in early generations can reduce the prediction accuracies, hence confirmation in the later generation when the seed quantity is large is desirable. The accurate estimations are necessary for the commercialization purpose and it is recommended to use gas chromatography (GC). The selected lines with >80% high oleic acid content recorded a variation to an extent of 2% and similar results were also reported by Janila et al. (2016b). Hamdan et al. (2009) reported the loss of minor effect QTLs or modifying genes resulted in variation in the oleic acid content even toward the higher side. The variation in oleic content could also relate to incomplete dysfunctional mutant *ahFAD2B* allele. There could be possible involvement of other unidentified genetic factors or enhancers conditioning oleic acid content in groundnut (Wang X. Z. et al., 2015).

The benefit of phenotypic screening coupled with MAS in early generations facilitates selection for a trait/s (Hickey et al., 2017). Even though, SNP markers confirmed the presence of LLS and rust governing major effect QTLs, information on genes with minor effects and genetic modifiers necessary to achieve higher levels of LLS resistance is limited. The LLS resistance is a genetically complex mechanism, governed by a combination of component traits such as reduced sporulation, latent period, and smaller lesion diameters (Dwivedi et al., 2002). The rust resistance in groundnut was reported due to the accumulation of phytoalexins and elevated proteinases 1,3-α-glucanases, and 1,3-β-glucanases (Sankara et al., 1996). The insights on physiological and biochemical components of resistance to LLS and rust and yield trade-off, if any in resistant cultivars can be future avenues of research. Broadening the genetic base for LLS and rust resistance can provide the broad-spectrum resistance against LLS and rust. This can contribute to improved selection for LLS and rust resistance and modeling growth and yield responses of groundnut.

The backcross-derived populations bred in the present study showed improved resistance to LLS and rust. These lines were higher or at par for yield governing traits in comparison to K 6. Adoption of more efficient selection strategies such as high throughput genotyping, NIRS for oil quality traits is critical for selection in early generations. This approach could be useful for the rapid transfer of disease resistance and high oleic trait into popular groundnut cultivars grown by farmers.

Development and release of high oil and high oleic groundnut varieties have tremendous market potential both in India as well as globally. The promising breeding lines identified in the study will be advanced and evaluated in replicated field trials for disease resistance, high oleic acid and yield traits. Advanced superior breeding lines derived from K 6 for LLS, rust resistance and high oleic traits will be a valuable resource either for use in crossing program or for release as a variety. The above results demonstrated the effectiveness of the molecular marker for early generation selection, and to enhance selection intensity as the number of selected candidates for the same number of selected individuals is higher when we used markers. Selection intensity is directly proportional to genetic gain and thus the use of MABC resulted in an increase in genetic gain. The process also demonstrated enhanced selection accuracy (r) by using high throughput phenotyping, such as NIRS for assessing

quality parameters and detached leaf technique for studying disease reaction.

DATA AVAILABILITY STATEMENT

The genotype data is uploaded in institutional repository with the DOI number: https://doi.org/10.21421/D2/FWIWWC.

AUTHOR CONTRIBUTIONS

DD conducted the experiment and drafted the manuscript. BM, CR, and JP conceived and designed the experiment. HS helped with disease screening and manuscript editing. MP performed foreground selection using linked markers for oleic acid, and rust, and LLS, and helped in the manuscript editing. JP, MV, SC, and SM helped with the experiment and critical review of the manuscript.

FUNDING

This work is supported by the OFID, Department of Agriculture and Co-operation (DOAC), Ministry of Agriculture, Government of India, and CGIAR Research Program on Grain Legumes and Dry Land Cereals (CRP-GLDC).

ACKNOWLEDGMENTS

Authors are thankful to Mr. Papaiah V., Ms. Chandrakala, Ms. Aparna, Mr. Ranjeet, Mr. Abhishek for the technical assistance during the hybridization. Dr. Y. Shasidhar and Mr. S. Gangurde for providing assistance in genotyping. Ms. Mangala is acknowledged for extending support during disease screening. Authors also acknowledge the OPEC Fund for International Development (OFID) for the fellowship to Dnyaneshwar during his doctoral research.

REFERENCES

Agarwal, G., Clevenger, J., Pandey, M. K., Wang, H., Shasidhar, Y., Chu, Y., et al. (2018). High-density genetic map using whole-genome resequencing for fine mapping and candidate gene discovery for disease resistance in peanut. *Plant Biotechnol. J.* 16, 1954–1967. doi: 10.1111/pbi.12930

AICRP-G. (2018). "All india coordinated research project on groundnut," in *Proceedings in Annual Groundnut Workshop, April 18-20th at Professor Jayashankar Telangana State Agricultural University*, Hyderabad.

Bera, S. K., and Das, P. K. (1998). Inheritance of pod beak and pod constriction in groundnut (*Arachis hypogaea* L.). *Curr. Agric. Res.* 11, 1–3.

Bera, S. K., Kamdar, J. H., Kasundra, S. V., Dash, P., Maurya, A. K., Jasani, M. D., et al. (2018). Improving oil quality by altering levels of fatty acids through marker-assisted selection of ahfad2 alleles in peanut (*Arachis hypogaea* L.). *Euphytica* 214:162. doi: 10.1007/s10681-018-2241-0

Chaudhari, S., Khare, D., Patel, S. C., Subramaniam, S., Variath, M. T., Sudini, H. K., et al. (2019). Genotype× Environment studies on resistance to late leaf spot and rust in Genomic Selection training population of peanut (*Arachis hypogaea* L.). *Front. Plant Sci.* 10:1338. doi: 10.3389/fpls.2019.01338

Chen, Z., Wang, M. L., Barkley, N. A., and Pittman, R. N. (2010). A simple allele-specific PCR assay for detecting FAD2 alleles in both A and B genomes of the cultivated peanut for high-oleate trait selection. *Plant Mol. Biol. Rep.* 28, 542–548. doi: 10.1007/s11105-010-0181-5

Chopra, R., Burow, G., Farmer, A., Mudge, J., Simpson, C. E., Wilkins, T. A., et al. (2015). Next-generation transcriptome sequencing, SNP discovery and validation in four market classes of peanut, *Arachis hypogaea* L. *Mol. Genet. Genomics* 290, 1169–1180. doi: 10.1007/s00438-014-0976-4

Chopra, R., Simpson, C. E., Hillhouse, A., Payton, P., Sharma, J., and Burow, M. D. (2018). SNP genotyping reveals major QTLs for plant architectural traits between A-genome peanut wild species. *Mol. Genet. Genomics* 293, 1477–1491. doi: 10.1007/s00438-018-1472-z

Chu, Y., Gill, R., Clevenger, J., Timper, P., Holbrook, C. C., and Ozias-Akins, P. (2016). Identification of rare recombinants leads to tightly linked markers for nematode resistance in peanut. *Peanut Sci.* 43, 88–93. doi: 10.3146/PS16-12.1

Chu, Y., Holbrook, C. C., and Ozias-Akins, P. (2009). Two alleles of ahFAD2B control the high oleic acid trait in cultivated peanut. *Crop Sci.* 49, 2029–2036. doi: 10.2135/cropsci2009.01.0021

Chu, Y., Wu, C. L., Holbrook, C. C., Tillman, B. L., Person, G., and Ozias-Akins, P. (2011). Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. *Plant Genome* 4, 110–117. doi: 10.3835/plantgenome2011.01.0001

Clevenger, J., Chu, Y., Chavarro, C., Botton, S., Culbreath, A., Isleib, T. G., et al. (2018). Mapping late leaf spot resistance in peanut (*Arachis hypogaea*) using QTL-seq reveals markers for Marker-Assisted Selection. *Front. Plant Sci.* 9:83. doi: 10.3389/fpls.2018.00083

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., et al. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theor. Appl. Genet.* 132, 627–645. doi: 10.1007/s00122-019-03317-0

Collard, B. C., and Mackill, D. J. (2007). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170

Dwivedi, S. L., Pande, S., Rao, J. N., and Nigam, S. N. (2002). Components of resistance to late leaf spot and rust among interspecific derivatives and their significance in a foliar disease resistance breeding in groundnut (*Arachis hypogaea* L.). *Euphytica* 125, 81–88. doi: 10.1023/A:1015707301659

FAOSTAT (2015). *FAO Statistical Database*. Available online at: http://faostat.fao.org/ (accessed April 5, 2019).

FAOSTAT (2017). *FAO Statistical Database*. Available online at: http://faostat.fao.org/ (accessed April 5, 2019).

Foster, D. J., Wynne, J. C., and Beute, M. K. (1980). Evaluation of detached leaf culture for screening peanuts for leafspot resistance. *Peanut Sci.* 7, 98–100. doi: 10.3146/i0095-3679-7-2-10

Gorbet, D. W., and Knauft, D. A. (1997). Registration of 'SunOleic 95R'peanut. *Crop Sci.* 37, 1392–1392. doi: 10.2135/cropsci1997.0011183X003700040081x

Hamdan, Y. A., Pérez-Vich, B., Velasco, L., and Fernández-Martínez, J. M. (2009). Inheritance of high oleic acid content in safflower. *Euphytica* 168, 61–69. doi: 10.1007/s10681-008-9879-y

Hickey, L. T., Germán, S. E., Pereyra, S. A., Diaz, J. E., Ziems, L. A., Fowler, R. A., et al. (2017). Speed breeding for multiple disease resistance in barley. *Euphytica* 213:64. doi: 10.1007/s10681-016-1803-2

Holbrook, C. C., Ozias-Akins, P., Chu, Y., Culbreath, A. K., Kvien, C. K., and Brenneman, T. B. (2017). Registration of 'TifNV-High O/L' Peanut. *J. Plant Regist.* 11, 228–230. doi: 10.3198/jpr2016.10.0059crc

Hu, F. B., Van, R. M., and Liu, S. (2001). Diet and risk of type II diabetes: the role of types of fat and carbohydrate. *Diabetologia* 44, 805–817. doi: 10.1007/s001250100547

IBPGR, and ICRISAT (1992). *Descriptors for Groundnut*. Rome: International Crops Research Institute for the Semi-Arid Tropics.

ICRISAT (1987). *International Crops Research Institute for the Semi-Arid Tropics Annual Report*. Patancheru: ICRISAT, 217–231.

Janila, P., Pandey, M. K., Manohar, S. S., Variath, M. T., Nallathambi, P., Nadaf, H. L., et al. (2016a). Foliar fungal disease-resistant introgression lines of groundnut (*Arachis hypogaea* L.) record higher pod and haulm yield in multilocation testing. *Plant Breed.* 135, 355–366. doi: 10.1111/pbr.12358

Janila, P., Pandey, M. K., Shasidhar, Y., Variath, M. T., Sriswathi, M., Khera, P., et al. (2016b). Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. *Plant Sci.* 242, 203–213. doi: 10.1016/j.plantsci.2015.08.013

Janila, P., Ramaiah, V., Rathore, A., Rupakula, A., Reddy, R. K., Waliyar, F., et al. (2013). Genetic analysis of resistance to late leaf spot in interspecific groundnuts. *Euphytica* 193, 13–25. doi: 10.1007/s10681-013-0881-7

Jiang, G. H., Xu, C. G., Tu, J. M., Li, X. H., He, Y. Q., and Zhang, Q. F. (2004). Pyramiding of insect-and disease-resistance genes into an elite indica, cytoplasm male sterile restorer line of rice, 'Minghui 63'. *Mol. Breed.* 123, 112–116. doi: 10.1046/j.1439-0523.2003.00917.x

Joshi, R. K., and Nayak, S. (2010). Gene pyramiding-A broad spectrum technique for developing durable stress resistance in crops. *Biotechnol. Mol. Biol. Rev.* 5, 51–60.

Khedikar, Y. P., Gowda, M. V. C., Sarvamangala, C., Patgar, K. V., Upadhyaya, H. D., and Varshney, R. K. (2010). A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 121, 971–984. doi: 10.1007/s00122-010-1366-x

Kolekar, R. M., Sukruth, M., Shirasawa, K., Nadaf, H. L., Motagi, B. N., Lingaraju, S., et al. (2017). Marker-assisted backcrossing to develop foliar disease-resistant genotypes in TMV 2 variety of peanut (*Arachis hypogaea* L.). *Plant Breed.* 136, 948–953. doi: 10.1111/pbr.12549

Leal-Bertioli, S. C., Cavalcante, U., Gouveia, E. G., Ballen-Taborda, C., Shirasawa, K., Guimarães, P. M., et al. (2015). Identification of QTLs for rust resistance in the peanut wild species *Arachis magna* and the development of KASP markers for marker assisted selection. *Genes Genom. Genet.* 5, 1403–1413. doi: 10.1534/g3.115.018796

Li, H., Rasheed, A., Hickey, L. T., and He, Z. (2018). Fast-forwarding genetic gain. *Trends Plant Sci.* 23, 184–186. doi: 10.1016/j.tplants.2018.01.007

Mace, E. S., Buhariwalla, K. K., Buhariwalla, H. K., and Crouch, J. H. (2003). A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol. Biol. Rep.* 21, 459–460. doi: 10.1007/BF02772596

Mienie, C. M. S., and Pretorius, A. E. (2013). Application of marker-assisted selection for ahFAD2A and ahFAD2B genes governing the high-oleic acid trait in South African groundnut cultivars (*Arachis hypogaea* L.). *Afr. J. Biotechnol.* 12, 4283–4289. doi: 10.5897/AJB2012.2976

Moore, K. M., and Knauft, D. A. (1989). The inheritance of high oleic acid in peanut. *J. Hered.* 80, 252–253. doi: 10.1093/oxfordjournals.jhered.a110845

Motagi, B. N. (2001). *Genetic Analyses of Resistance to Late Leaf Spot and Rust Vis-A-Vis Productivity in Groundnut (Arachis hypogaea L.)*. Ph.D. thesis, University of Agricultural Sciences, Dharwad.

Murali, T. V., Manohar, S. S., Chaudhari, S., Varshney, R. K., Pandey, M. K., Deshmukh, D., et al. (2019). "Genomics-enabled early generation selection in peanut breeding pipeline," in *Proceedings of the 9th ICLGG*, Dijon.

Nigam, S. N., and Blummel, M. (2010). Cultivar-dependent variation in food-feed-traits in groundnut (*Arachis hypogaea* L.). *Anim. Nutr. Feed Technol.* 10, 39–48.

Nigam, S. N., Rao, M. V., and Gibbons, R. W. (1990). *Artificial Hybridization in Groundnut. Information Bulletin no. 29*. Patancheru: International Crops Research Institute for the Semi-Arid Tropics.

Norden, A. J., Gorbet, D. W., Knauft, D. A., and Young, C. T. (1987). Variability in oil quality among peanut genotypes in the Florida breeding program. *Peanut Sci.* 14, 7–11. doi: 10.3146/i0095-3679-14-1-3

Okello, D. K., Kaaya, A. N., Bisikwa, J., Were, M., and Oloka, H. K. (2010). *Aflatoxins in Groundnuts. Management of Aflatoxins in Groundnuts: A Manual for Farmers, Processors, Traders and Consumers in Uganda*. Entebbe: National Agricultural Research Organisation.

Pande, S., Bandyopadhyay, R., Blümmel, M., Rao, J. N., Thomas, D., and Navi, S. S. (2003). Disease management factors influencing yield and quality of sorghum and groundnut crop residues. *Field Crops Res.* 84, 89–103. doi: 10.1016/S0378-4290(03)00143-6

Pande, S., and Rao, J. N. (2001). Resistance of wild *Arachis* species to late leaf spot and rust in greenhouse trials. *Plant Dis.* 85, 851–855. doi: 10.1094/PDIS.2001.85.8.851

Pandey, M. K., Khan, A. W., Singh, V. K., Vishwakarma, M. K., Shasidhar, Y., Kumar, V., et al. (2017). QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea* L.). *Plant Biotechnol. J.* 15, 927–941. doi: 10.1111/pbi.12686

Patel, M., Jung, S., Moore, K., Powell, G., Ainsworth, C., and Abbott, A. (2004). High-oleate peanut mutants result from a MITE insertion into the FAD2 gene. *Theor. Appl. Genet.* 108, 1492–1502. doi: 10.1007/s00122-004-1590-3

Reddy, S., Vishwakarma, M. K., Pandey, M. K., Janila, P., Variath, M. T., Manohar, S. S., et al. (2017). Molecular mapping of oil content and fatty acids using dense genetic maps in groundnut (*Arachis hypogaea* L.). *Front. Plant Sci.* 8:794. doi: 10.3389/fpls.2017.00794

Sahdev, R. K. (2015). Present status of peanuts and progression in its processing and preservation techniques. *A J. Agric. Eng.* 17, 309–327.

Sankara, P., SubbaRao, P. V., and Strange, R. N. (1996). Differential accumulation of phytoalexins in leaves of susceptible and resistant genotypes of groundnut (*Arachis hypogaea* L.) inoculated with Puccinia arachidis Speg. *J. Phytopathol.* 144, 527–532. doi: 10.1111/j.1439-0434.1996.tb00294.x

Santra, D., DeMacon, V. K., Garland-Campbell, K., and Kidwell, K. (2006). "Marker assisted backcross breeding for simultaneous introgression of stripe rust resistance genes yr5 and yr15 into spring wheat (*Triticum aestivum* L.)," in *Proceedings of the International Meeting of ASA-CSSA-SSSA*, Salt Lake City, UT, 74–75.

Shanti, M. L., Shenoy, V. V., Devi, G. L., Kumar, V. M., Premalatha, P., Kumar, G. N., et al. (2010). Marker-assisted breeding for resistance to bacterial leaf blight in popular cultivar and parental lines of hybrid rice. *J. Plant Pathol.* 92, 495–501. doi: 10.4454/jpp.v92i2.194

Shasidhar, Y., Variath, M. T., Vishwakarma, M. K., Manohar, S. S., Gangurde, S. S., Sriswathi, M., et al. (2020). Improvement of three Indian popular groundnut varieties for foliar disease resistance and high oleic acid using SSR markers and SNP array in marker-assisted backcrossing. *Crop J.* 8, 1–15. doi: 10.1016/j.cj.2019.07.001

Simpson, C. E., Starr, J. L., Church, G. T., Burow, M. D., and Paterson, A. H. (2003). Registration of 'NemaTAM' peanut. *Crop Sci.* 43, 1561–1562. doi: 10.2135/cropsci2003.1561

Singh, M. P., Erickson, J. E., Boote, K. J., Tillman, B. L., Jones, J. W., and Van Bruggen, A. H. (2011). Late leaf spot effects on growth, photosynthesis, and yield in peanut cultivars of differing resistance. *Agron. J.* 103, 85–91. doi: 10.2134/agronj2010.0322

Subrahmanyam, P., McDonald, D., Waliyar, F., Reddy, L. J., Nigam, S. N., et al. (1995). *Screening Methods and Sources of Resistance to Rust and Late Leaf Spot of Groundnut. Information Bulletin no 47*. Patancheru: ICRISAT, 1–26.

Subrahmanyam, P., Williams, J. H., McDonald, D., and Gibbons, R. W. (1984). The influence of foliar diseases and their control by selective fungicides on a range of groundnut (*Arachis hypogaea* L.) genotypes. *Ann. Appl. Biol.* 104, 467–476. doi: 10.1111/j.1744-7348.1984.tb03029.x

Sudini, H., Upadhyaya, H. D., Reddy, S. V., Mangala, U. N., Rathore, A., and Kumar, K. V. K. (2015). Resistance to late leaf spot and rust diseases in ICRISAT's mini core collection of peanut (*Arachis hypogaea* L.). *Austral. Plant Pathol.* 44, 557–566. doi: 10.1007/s13313-015-0368-1

Sujay, V., Gowda, M. V. C., Pandey, M. K., Bhat, R. S., Khedikar, Y. P., Nadaf, H. L., et al. (2012). Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). *Mol. Breed.* 30, 773–788. doi: 10.1007/s11032-011-9661-z

Sukruth, M., Paratwagh, S. A., Sujay, V., Kumari, V., Gowda, M. V. C., Nadaf, H. L., et al. (2015). Validation of markers linked to late leaf spot and rust resistance, and selection of superior genotypes among diverse recombinant inbred lines and backcross lines in peanut (*Arachis hypogaea* L.). *Euphytica* 204, 343–351. doi: 10.1007/s10681-014-1339-2

Sundaram, J., Kandala, C. V., and Butts, C. L. (2009). Application of near infrared spectroscopy to peanut grading and quality analysis: overview. *Sens. Instr. Food Qual. Saf.* 3, 156–164. doi: 10.1007/s11694-009-9081-5

Sundaram, R. M., Vishnupriya, M. R., Biradar, S. K., Laha, G. S., Reddy, G. A., Rani, N. S., et al. (2008). Marker assisted introgression of bacterial blight resistance in Samba Mahsuri, an elite indica rice variety. *Euphytica* 160, 411–422. doi: 10.1007/s10681-007-9564-6

Tyagi, S., Mir, R. R., Kaur, H., Chhuneja, P., Ramesh, B., Balyan, H. S., et al. (2014). Marker-assisted pyramiding of eight QTLs/ genes for seven different traits in common wheat (*Triticum aestivum* L.). *Mol. Breed.* 34, 167–175. doi: 10.1007/s11032-014-0027-1

Varman, P. V. (1999). A foliar disease resistant line developed through interspecifc hybridization in groundnut (*Arachis hypogaea*). *Indian J. Agr. Sci.* 69, 67–68.

Varshney, R. K., Pandey, M. K., Janila, P., Nigam, S. N., Sudini, H., Gowda, M. V. C., et al. (2014). Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 127, 1771–1781. doi: 10.1007/s00122-014-2338-3

Vassiliou, E. K., Gonzalez, A., Garcia, C., Tadros, J. H., Chakraborty, G., and Toney, J. H. (2009). Oleic acid and peanut oil high in oleic acid reverse the inhibitory effect of insulin production of the inflammatory cytokine TNF-α both in vitro and in vivo systems. *Lipids Health Dis.* 8:25. doi: 10.1186/1476-511X-8-25

Wang, D. D., and Hu, F. B. (2017). Dietary fat and risk of cardiovascular disease: recent controversies and advances. *Annu. Rev. Nutr.* 37, 423–446. doi: 10.1146/annurev-nutr-071816-064614

Wang, J., Singh, R. P., Braun, H. J., and Pfeiffer, W. H. (2009). Investigating the efficiency of the single backcrossing breeding strategy through computer simulation. *Theor. Appl. Genet.* 118, 683–694. doi: 10.1007/s00122-008-0929-6

Wang, M. L., Khera, P., Pandey, M. K., Wang, H., Qiao, L., Feng, S., et al. (2015). Genetic mapping of QTLs controlling fatty acids provided insights into the genetic control of fatty acid synthesis pathway in peanut (*Arachis hypogaea* L.). *PLoS One* 10:e0119454. doi: 10.1371/journal.pone.0119454

Wang, X. Z., Wu, Q., Tang, Y. Y., Sun, Q. X., and Wang, C. T. (2015). "FAD2B from a peanut mutant with high oleic acid content was not completely dysfunctional," in *Advances in Applied Biotechnology. Lecture Notes in Electrical Engineering*, Vol. 332, eds T. C. Zhang and M. Nakajima (Berlin: Springer), 265–271.

Xiuzhen, W., Yueyi, T., Qi, W., Ligui, Z., Quanxi, S., and Zhiwei, W. (2016). Development of a new high-oleic peanut cultivar Huayu 662 by using molecular marker aided selection and NIRS. *J. Nuclear Agri. Sci.* 30, 1054–1058. doi: 10.11869/j.issn.100-8551.2016.06.1037

Yadawad, A., Gadpale, A., Hanchinal, R. R., Nadaf, H. L., Desai, S. A., Biradar, S., et al. (2017). Pyramiding of leaf rust resistance genes in bread wheat variety DWR 162 through marker assisted backcrossing. *Indian J. Genet. Plant Br.* 77, 251–257. doi: 10.5958/0975-6906.2017.00033.5

Yeri, S. B., and Bhat, R. S. (2016). Development of late leaf spot and rust resistant backcross lines in JL 24 variety of groundnut (*Arachis hypogaea* L.). *Electron. J. Plant Breed.* 7, 37–41. doi: 10.5958/0975-928X.2016.00005.3

Yu, S., Pan, L., Yang, Q., Min, P., Ren, Z., and Zhang, H. (2008). Comparison of the Δ12 fatty acid desaturase gene between high-oleic and normal-oleic peanut genotypes. *J. Genet. Genomics* 35, 679–685. doi: 10.1016/S1673-8527(08)60090-9

Zhao, S., Li, A., Li, C., Xia, H., Zhao, C., Zhang, Y., et al. (2017). Development and application of KASP marker for high throughput detection of AhFAD2 mutation in peanut. *Electron J. Biotechnol.* 25, 9–12. doi: 10.1016/j.ejbt.2016.10.010

Zhao, Y., Zhang, C., Chen, H., Yuan, M., Nipper, R., Prakash, C. S., et al. (2016). QTL mapping for bacterial wilt resistance in peanut (*Arachis hypogaea* L.). *Mol. Breed.* 36:13. doi: 10.1007/s11032-015-0432-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Deshmukh, Marathi, Sudini, Variath, Chaudhari, Manohar, Rani, Pandey and Pasupuleti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 June 2020
doi: 10.3389/fgene.2020.00656

Genome-Wide SNP Identification and Association Mapping for Seed Mineral Concentration in Mung Bean (*Vigna radiata* L.)

*Xingbo Wu[1], A. S. M. Faridul Islam[1†], Naransa Limpot[2], Lucas Mackasmiel[1], Jerzy Mierzwa[1], Andrés J. Cortés[3,4] and Matthew W. Blair[1]**

[1] Department of Agricultural and Environmental Sciences, Tennessee State University, Nashville, TN, United States, [2] Agro-Biotechnology Institute, Selangor, Malaysia, [3] Corporación Colombiana de Investigación Agropecuaria AGROSAVIA, C.I. La Selva, Rionegro, Colombia, [4] Departamento de Ciencias Forestales, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia – Sede Medellín, Medellín, Colombia

Mung bean (*Vigna radiata* L.) quality is dependent on seed chemical composition, which in turn determines the benefits of its consumption for human health and nutrition. While mung bean is rich in a range of nutritional components, such as protein, carbohydrates and vitamins, it remains less well studied than other legume crops in terms of micronutrients. In addition, mung bean genomics and genetic resources are relatively sparse. The objectives of this research were three-fold, namely: to develop a genome-wide marker system for mung bean based on genotyping by sequencing (GBS), to evaluate diversity of mung beans available to breeders in the United States and finally, to perform a genome-wide association study (GWAS) for nutrient concentrations based on a seven mineral analysis using inductively coupled plasma (ICP) spectroscopy. All parts of our research were performed with 95 cultivated mung bean genotypes chosen from the USDA core collection representing accessions from 13 countries. Overall, we identified a total of 6,486 high quality single nucleotide polymorphisms (SNPs) from the GBS dataset and found 43 marker × trait associations (MTAs) with calcium, iron, potassium, manganese, phosphorous, sulfur or zinc concentrations in mung bean grain produced in either of two consecutive years' field experiments. The MTAs were scattered across 35 genomic regions explaining on average 22% of the variation for each seed nutrient in each year. Most of the gene regions provided valuable candidate loci to use in future breeding of new varieties of mung bean and further the understanding of genetic control of nutritional properties in the crop. Other SNPs identified in this study will serve as important resources to enable marker-assisted selection (MAS) for nutritional improvement in mung bean and to analyze cultivars of mung bean.

Keywords: association mapping, nutritional improvement, mineral nutrients, biofortification, nutritional breeding, nutrigenomics, micronutrients

OPEN ACCESS

Edited by:
Manish Kumar Pandey, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

Reviewed by:
Rohit Joshi, Institute of Himalayan Bioresource Technology (CSIR), India
Pei Xu, China Jiliang University, China
Aditya Pratap, Indian Institute of Pulses Research (ICAR), India

***Correspondence:**
Matthew W. Blair
mblair@tnstate.edu

†Present address:
A. S. M. Faridul Islam, Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, United States

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Genetics

Received: *26 July 2019*
Accepted: *29 May 2020*
Published: *24 June 2020*

Citation:
Wu X, Islam ASMF, Limpot N, Mackasmiel L, Mierzwa J, Cortés AJ and Blair MW (2020) Genome-Wide SNP Identification and Association Mapping for Seed Mineral Concentration in Mung Bean (Vigna radiata L.). Front. Genet. 11:656. doi: 10.3389/fgene.2020.00656

INTRODUCTION

Mung bean (*Vigna radiata*) is an important pulse crop for most of the highly populated countries of Asia, such as Bangladesh, China, India, Indonesia, Laos, Malaysia, Myanmar, Pakistan, Thailand and Vietnam with growing presence in other parts of the world too (Nair et al., 2013). The cultivation area for the crop is over 6 million hectares worldwide (8.5% of the global pulse area), and over 3 million tons of grain are produced (5% of global pulse production) (Schafleitner et al., 2016). High nitrogen fixation gives mung bean a vital role in intercropping with corn or millet in East Asia or for crop rotation with cereals, such as rice or wheat in South and Southeast Asia (Blair et al., 2016c). Production of mung bean usually increases soil fertility for subsequent cereal crop yields through symbiotic Nitrogen fixation. High market demands make mung bean a profitable crop for export from Australia, Europe, and United States to East Asia (Pandey et al., 2018). South Asia is also a developing market for imports from Eastern Africa. Mung bean is the third most important legume in China after soybean and common bean (Li et al., 2017).

Nutritionally, mung bean along with other Asian *Vigna* species, are considered highly nutritious, easy to cook and palatable legumes (Day, 2013; Rehman et al., 2019). Mung bean contains easily digestible proteins and various micronutrients (Akaerue and Onwuka, 2010; Kollárová et al., 2010). The seeds can be consumed as a dry bean or as a vegetable bean sprouts (Kim et al., 2015). They can be shelled from green pods, decorticated when dry and processed into a flour that is used commercially for products such as noodles, flatbreads, drinks and sweets (Dahiya et al., 2014, 2015). Mung bean sprouts are an important vegetable in several Asian cuisines (Lambrides and Godwin, 2007; Somta and Srinives, 2007). The mung bean is less well studied for biofortification traits than some other legumes even though it has a role to play in preventing micronutrient deficiencies (Dwivedi et al., 2012).

Human nutritional deficiencies of some minerals can lead to stunted growth and development in children, lower resistance to disease, and increased mortality rates (Caulfield et al., 2006). Over two-thirds of humans are experiencing inadequate intake of one or more mineral nutrients, with more than half having iron deficiency and over one third zinc deficiency (White and Broadley, 2009). Mineral deficiencies are, therefore, a major concern in global health (Soetan et al., 2010). Nutritional deficiencies are especially prevalent in developing countries where people do not have diverse diet of vegetables, fruits and meat or fish so grain crops are by necessity the major source of essential nutrients for humans (Dwivedi et al., 2012).

Food legumes usually contain higher concentrations of mineral nutrients than do cereals or root crops, but still need improvement for total concentration and bioavailability (Blair, 2013). Biofortification is the process of genetic improvement for increasing mineral acquisition, reducing anti-nutritional factors and balancing the appropriate mineral concentrations in the edible plant parts and seeds (Pfeiffer and McClafferty, 2007; Dwivedi et al., 2012). Quantitative trait loci (QTL) analyses through mapping populations or genome-wide association studies (GWAS) using molecular markers and high throughput sequencing techniques provide valuable ways of identifying the genes underlying nutritional traits (Blair, 2013). Previous QTL studies for mineral nutrients in legumes have been with mapping populations as reported on common bean, *Phaseolus vulgaris* (Blair et al., 2010a, 2011, 2013, 2016c), *Lotus japonicus* (Klein and Grusak, 2009), lentil, *Lens culinaris* (Khazaei et al., 2017; Ates et al., 2018), *Medicago truncatula* (Sankaran et al., 2009), and pea, *Pisum sativum* (Cheng et al., 2015; Diapari et al., 2015; Ma et al., 2017). GWAS for micronutrient concentration has been used for peas (Kwon et al., 2012) but has not been used in most other legume crops.

Despite all the economic and nutritional value of mung beans, the expansion of the crop is limited by a lack of genetic resources that are important for crop adaptation (Das et al., 2018). Breeding of mung beans for wider adaptation is a priority and modern marker tools are needed for selection of new varieties (Burlyaeva et al., 2019). Genetic markers are an important part of modern plant breeding for the legumes. In this regard, mung bean was one of the first legumes to have a genetic linkage map (Fatokun et al., 1992). However, genetic studies of mung bean have been sparse although the release of a whole genome sequence is now helping researchers to study genes from this important legume (Kang et al., 2014).

Few germplasm studies have been published for mung bean because the genetic diversity based on microsatellite (SSR) markers is limited and few applied SSR markers have been developed (Humphry et al., 2002; Chen et al., 2015; Liu et al., 2017) because of low polymorphism (Gwag et al., 2006; Somta et al., 2008; Tangphatsornruang et al., 2009). Single Nucleotide Polymorphism (SNP) markers have emerged as an interesting option for mung bean with over 300,000 discovered by Van et al. (2013). However, only 43 and 20 SNPs have been validated as competitive allele specific polymorphism (KASP) markers in the two studies conducted to date (Van et al., 2013; Islam and Blair, 2018), respectively.

An alternative to single marker assays for SNPs, is Genotyping by Sequencing (GBS) based on the fast development of next generation sequencing (NGS) technology (Poland and Rife, 2012). Compared to other SNP marker assays, GBS can allow genotyping to be completed rapidly and is a simultaneous method for polymorphism discovery (Elshire et al., 2011). The technique has been used extensively for a wide range of applications and crops (He et al., 2014). To date, GBS studies have been undertaken in mung bean for genetic mapping and diversity study (Schafleitner et al., 2016; Noble et al., 2018), but not for association mapping.

In summary, mung bean can effectively contribute to the alleviation of iron, zinc and protein deficiency in human populations of Asia (Singh R. et al., 2017), but so far there is no genetic study of mineral nutrients in the crop. GBS, being a whole-genome method of surveying polymorphism across a collection of genotypes is ideal for gene discovery in mung bean and for finding SNP polymorphisms that can be used to select for biofortified, nutrient dense genotypes. With this in mind, the objectives of this study were to (1) develop genome-wide SNPs for mung bean using a GBS approach, (2) investigate

the genetic diversity and population structure of mung bean as it can account for spurious genomic association, and (3) perform genome-wide association studies with seed mineral concentrations in mung bean.

MATERIALS AND METHODS

Plant Material

This study began in 2015 with the evaluation of a core collection of 408 genotypes of mung bean provided by the United States Department of Agriculture (USDA), Plant Genetic Resources Conservation Unit, in Griffin, GA, United States. As not all genotypes were adapted, a smaller set of 140 genotypes were grown in 2016. From these a subset of 95 higher yielding genotypes were selected for molecular and GWAS analyses. The subset had the following geographical representation: 3 genotypes were from China, 6 from the Philippines, 7 from South Korea, 2 from Thailand, 37 from India, 32 from Iran, 2 from Pakistan, and one each from Afghanistan, Argentina, Mexico, Turkey, United States, and Uzbekistan. Therefore, the total from each continent skewed toward Asia, the center of origin for mung bean, over those from the Americas. The genotypes tended to have sub-tropical rather than temperate adaptation because they were selected for growth ability in the hot summer season in the Tennessee State University (TSU) field station as described below.

Field Site and Experiment Design

The experiments were carried across two consecutive summer seasons of 2015 and 2016 at the Agricultural Research and Education Center (AREC) field station of TSU in Nashville TN, United States of America which is situated at 36° 10′ 1″ N/86° 49′ 45″ W with an elevation of 54.3 m above sea level. The climate conditions were hot and humid (Koppen zone: Cfa, humid subtropical). The experiments were conducted in an open field and only watered once to induce germination. Thereafter, the crop relied upon rainfall for water supply during both years. An augmented design was used in 2015 and a randomized complete block design with two repetitions was used in 2016. A total of 30 seeds per genotype were sown into each plot, and the plots were 2.4 m long with l.0 m row centers and blocks were separated by 0.8 m alleys.

Soil Analysis of the Site

The soil was a Byler type Silt Loam with 2–5 percent slope toward the West in the direction of the Cumberland River which surrounds the farm of which the field is part of the valley. Water carrying capacity was estimated to be 50 mm of available water per cubic foot. The natural fertility and micronutrient content of the site was tested with a soils analysis at the University of Tennessee – Extension Soil, Plant and Pest Center (Nashville, TN) based on soil collected in the late winter prior to planting of the mung bean experiments. In addition to macronutrients nitrogen (N), phosphorus (P) and Potassium (K), the soils test evaluated Boron (B), Calcium (Ca), Iron (Fe), Magnesium (Mg), Manganese (Mn), Sodium (Na), Sulfur (S), and Zinc (Zn) along with organic matter and pH readings. No fertilizer was used as that could bias micro-nutrient accumulation.

Seed Mineral Analysis

The seeds from each genotype were harvested from pods, cleaned and dried at room temperature. To analyze mineral concentrations, 20 g of seeds were ground using zirconium oxide grinding balls and Teflon coated contained using a Retsch mix mill 400 (Verder Scientific Inc., Newtown, PA, United States) to grind each sample. A 0.5 g sub-sample for each genotype was measured out for each genotype were digested in CEM tubes containing 10 mL of ultra-pure nitric acid using a MARS 5 Microwaving Digestion Oven (CEM Corporation, Matthews, NC, United States) set at 200°C for 15 min. Reagent blanks were included in each series of digestions. Subsequently, the digested samples were diluted in 50 mL volumetric flasks by using ultrapure water and filtered before analysis. The sample solutions were stored in polyethylene vials at 4°C until analysis. Each sample was then analyzed for seven different elements, Ca, Fe, K, Mn, P, S, and Zn, using inductively coupled plasma optical emission spectroscopy (ICP-OES) with a CIROS Model FCE12 (Spectro Inc., Kleve, Germany) housed in the Soils and Tissue Analysis laboratory of TSU. Mineral concentration was calculated by multiplying the average sample elemental concentration and reported in $\mu g\ g^{-1}$ (ppm).

DNA Extraction

High molecular weight DNA that was also of high quality and purity was needed for the GBS method. In this study, we used the shoot apex and young leaves of mung bean sprouts to obtain soft etiolated leaf tissue that were easily ground to a powder in the Retsch mill using stainless steel balls and grinding chambers cooled to below freezing with liquid Nitrogen. In this method, seeds were sterilized in bleach for 5 min followed by 2 5-min rinses in autoclaved ddH_2O. Sterilized seeds were then transferred into individual magenta box (Sigma-Aldrich Co. St. Louis, MO, United States) with a sterilized paper tower and placed in a growth chamber (Thermo Fisher Scientific, Inc., Waltham, MA, United States) for 7 days until seeds had sprouted. The ground tissue (approximately 50 mg) was enough for DNA extraction with a DNeasy 96 Plant Kit (QIAGEN, Valencia, CA, United States). Quantity and quality verification for DNA was made with a FLUOstar Omega spectrometer (BMG LABTECH Inc., Cary, NC, United States). A threshold value of 1.8 for 260/280 nm absorbance ratio was considered a threshold to select for high DNA purity and was confirmed by electrophoresis on a 1% agarose gel run at 110V in TAE buffer. A lyophilized aliquot consisting of 1.5 μg of DNA was prepared for each accession and was placed in a 96 well plate for GBS.

Genotyping by Sequencing, SNP Identification and Physical Mapping

GBS library construction was done at the Cornell University Biotechnology Institute according to Elshire et al. (2011). Briefly, the genomic DNA was digested with *Ape*KI (New England BioLabs Inc., Ipswich, MA, United States), followed by ligation

with a 95-plex barcode adaptor and a common Illumina adaptor. DNA fragments selection was done after digestion/ligation using magnetic beads for a size range of 300–400 nt. (Rohland and Reich, 2012). Single-end sequencing was done on a HiSeq 2000 sequencer (Illumina Inc. San Diego CA, United States). Raw sequence data were analyzed with the GBS discovery pipeline in TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) software (Glaubitz et al., 2014). The FASTQ raw files with information of plate layout and bar codes for each genotype, were used to construct a GBS database for SNP identification. The reads containing bar code sequences followed by *Ape*KI sticky end restriction cut sites (CWGC) were trimmed to 64 bases and stored in the database. Reads that had no matching barcode or cut site overhangs were excluded, as well as reads containing unidentified bases (N) and reads with adapter dimers. Subsequently, the bar-coded and tagged sequence reads were sorted and collapsed into unique sequence tags that were aligned with the *Vigna radiata* genome v1.0 from Legume Information System (LIS) at https://legumeinfo.org/genomes, using the Burrows-Wheeler Aligner (BWA) algorithm (Li and Durbin, 2009) for confirmation of SNP loci. The aligned positions of the SNP markers were used in the construction of a physical map for all 11 chromosomes of the mung bean genome based on R v.3.3.1 software showing distance (in Million base pair, Mb) to the left of the map and SNP loci to the right. All newly discovered SNPs were scored for coverage, depth and genotypic information using vcftools. A quality threshold score of 10 was applied for the validation of any given SNP locus. Each new locus was given a name corresponding to SNP for *Vigna radiata* (SVR) identifying its chromosome and position.

Statistical Analysis, GWAS Evaluation, and Candidate Gene Identification

Histograms and correlation coefficients for mineral concentrations measured in 2 years were calculated and plotted using customized codes in the software R v.3.3.1 (R Core Team). During the marker-based diversity analysis, we examined subpopulation structure with all the SNP markers found across all the mung bean accessions using principal coordinate analysis (PCoA) carried out and diagramed with the same program. Analyses of variance (ANOVAs) were conducted for each experiment independently, distribution normality was checked and correlations between means across years and amongst minerals were calculated also in R v.3. TASSEL was also used to perform association analyses between the SNP markers and the seed minerals Ca, Fe, K, Mn, P, S, and Zn across the mung bean panel in 2 years. Generalized linear models (GLMs) and mixed linear models (MLMs) were compared for each of the nutrients for both years. Combinations of co-variables were used within each model as follows: (1) model with the country of origin and the first two PCoA axes as covariates; (2) model with the first two PCoA axes scores as covariates; (3) model with the country of origin as covariate; and (4) model without covariates. All four MLMs used a centered IBS kinship matrix as a random effect to control for genomic background implementing the EMMA (Efficient Mixed-Model Association) and P3D (Particle-Particle, Particle-Density) algorithms to reduce computing time (Zhang et al., 2010). QQ-plots (quantile-quantile plots) of the P-values were inspected to assess whether excessive numbers of false positives were generated and choose the optimum models. Significant associations were determined using a strict Bonferroni correction of *P*-values at alpha = 0.05, leading to a significance threshold of 7.7×10^{-6} (0.05 divided by the number of markers, which was 6,486 SNP loci) or $-\log_{10}(7.7 \times 10^{-6}) = 5.11$.

GWAS evaluations were done on a per year basis with construction of customized Manhattan diagrams carried out with the R v.3.3.1 software. For genome diversity inspection in the light of associated regions, we used a sliding window approach to determine patterns of overall variation across the genome, using a window size of 1 Mb and a step size of 200 kilo base pairs (Kb). We computed per-window averages of SNP density, Tajima's D (Tajima, 1989) and nucleotide diversity as measured by π (Nei, 1987). Results were plotted against window midpoints in mega base pairs (Mb). For any SNPs significantly associated with seed mineral concentration we performed a candidate gene search within 10 Kb around the locus using annotated genes for the mung bean genome from LIS. Genes regions with significant SNPs were listed per mineral.

RESULTS

Mineral Concentration Variability

The entire USDA mung bean core collection consisting of 408 genotypes was the starting point for our work and was planted in an augmented design in 2015 after a winter rye grass/summer corn rotation. Early planting was done in April of that year to accommodate the variable maturity dates, photoperiod responses and differential growth habits of the diverse set of genotypes. Most genotypes were harvested by September, but some never produced seed. The best adapted genotypes from the initial trial were re-planted with multiplied seed from that year and put into another section of the same field to avoid diseases or pest carry over. The 2016 trial was implemented from June to October with 140 genotypes. Soils analysis based on a sampling and mixing topsoil from across the fields, showed that pH varied only slightly from 6.0 to 5.9, Na salt was low and average organic matter content was 2%. The soil concentration of minerals showed that no liming was necessary and as a result no extra Zn or Fe micro-fertilization was required.

The seed mineral concentrations were measured across both years and resulted in normally distributed data for each year and for most of the seven minerals analyzed for the genotypes used in this GWAS study (**Figure 1**). Notably, the mean mineral concentrations of mung beans for Fe and Zn were higher in 2015 compared to 2016 (**Table 1**) and variability was correspondingly higher in the first year for these two micronutrients (Fe = 44.69 ± 17.66 and Zn = 71.75 ± 41.69) than in the second year (Fe = 27.11 ± 9.47 and Zn = 31.27 ± 4.11). This was to be expected, considering that a larger number of genotypes were evaluated in the first year (n = 408) compared to the second year (n = 95) and the many higher yielding genotypes

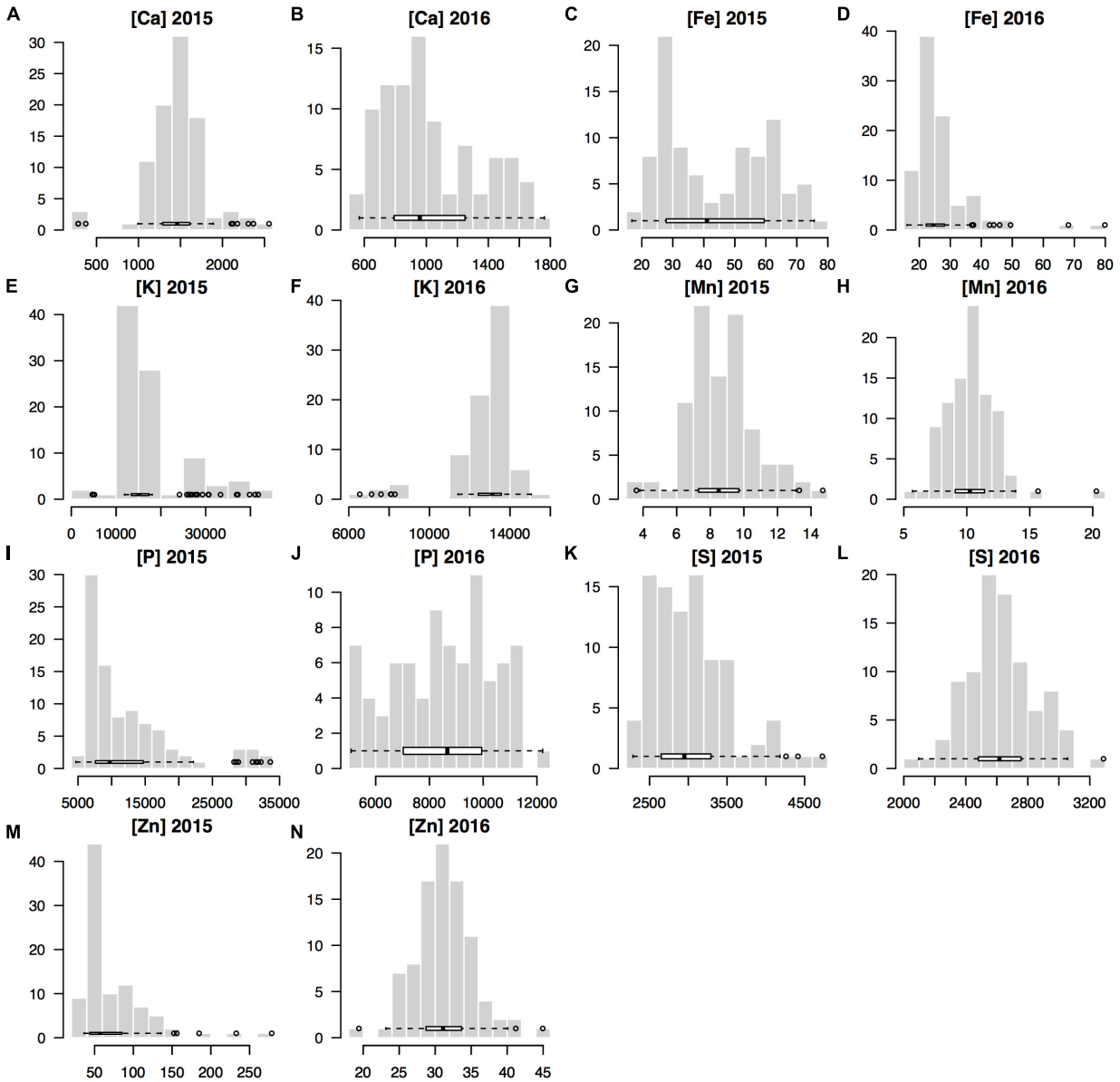

FIGURE 1 | Histograms for the concentrations of seven minerals measured for the seed produced by 95 mung bean (*Vigna radiata*) genotypes harvested across 2 years (2015 and 2016) used in genome wide association studies (GWAS) for Ca, calcium **(A,B)**; Fe, iron **(C,D)**; K, potassium **(E,F)**; Mn, manganes **(G,H)**; P, phosphorus **(I,J)**; S, sulfur **(K,L)**; and Zn, zinc **(M,N)**.

of the second year compared to the low yielding genotypes which were screened during the first year.

By comparison, the means for the seed macronutrient concentrations for K and P were higher in 2016 than in 2015 primarily due to the within field locations used since the experiments were planted in the same field across years; however, the smaller RCBD experiment in 2016 was planted in the lower half and higher nutrient section of that field while the augmented design experiment occupied the entire field in 2015. The mean concentrations of the seed minerals Ca Mn and S were similar in both years given that the sections of the field used for the two trials were of the same soil series and had the same overall agronomic management regime. The variability in concentration for these minerals was still high with ten-fold differences for Ca and Mn found among genotypes' seed mineral concentration, although S concentrations were less variable.

Pairwise correlation coefficients (R^2) calculated for the seven minerals within year showed different patterns depending on the mineral. Some correlations were high as with K and P measured in 2016 (R^2 = 1) or P and S (R^2 = 0.84), Mn and Ca (R^2 = 0.78), and Mn and K (R^2 = 0.75) measured in 2015 (**Figure 2A**); showing relationships between uptake of the macronutrients into seeds harvested from mung beans on the soil type for this location. Meanwhile, Fe, Zn, and Mn were positively and highly

TABLE 1 | Descriptive statistics and *t*-test comparison for seed mineral concentrations in parts per million (ppm) analyzed in the entire USDA core collection and high yielding selection of mung bean genotypes planted in 2015 (n = 408) and 2016 (n = 95), respectively.

Mineral trait	Year	Population Mean (ppm)	Population SD (ppm)	Population range (ppm)	*p* (*t*-test)
K	2015	17783.48	7774.60	4699.20–41737.90	0.47
	2016	19358.62	19623.72	6542.70–77746.81	
P	2015	12499.70	6852.32	4678.90–33625.30	<0.001
	2016	49124.34	119193.25	5085.90–403256.86	
S	2015	3033.60	510.02	2286.90–4730.10	<0.003
	2016	2613.83	231.78	1978.42–3288.20	
Ca	2015	1463.47	349.40	281.20–2554.90	<0.001
	2016	1038.87	311.80	567.51–1761.96	
Fe	2015	44.69	17.66	16.80–90.10	<0.001
	2016	27.11	9.47	16.04–79.87	
Mn	2015	8.93	2.63	3.60–24.30	0.003
	2016	11.10	6.70	5.69–72.50	
Zn	2015	71.75	41.69	36.30–278.60	<0.001
	2016	31.27	4.11	19.40–44.95	

correlated with each other in the same year. The correlation coefficients of Fe-Mn, Fe-Zn, and Mn-Zn measured in 2015 were 0.55, 0.44, and 0.27, compared to 0.66, 0.34, and 0.12 measured in 2016 (**Figure 2B**). Correlation was not observed within the seven minerals or between the same mineral measured in two different years. The lack of correlation between years may again have been due to climate variability and were not explored further.

SNP Discovery, Principal Component Analysis, and Physical Mapping

In our GBS technique, tissues from small cells harvested from etiolated, mung bean sprouts were better for DNA extraction compared to expanded cells from fully grown seedlings (data not shown). We used the shoot tips of sprouts that not only contained more cells than expanded leaves, but also reduced the time and space for genetic experiment preparation. ApeK1 was found to be an ideal enzyme for GBS library preparation and after filtering, a total of 6,486 SNP markers were validated for the analyses of the mung bean accessions (**Supplementary Table S1**) and for the diversity assessment as represented in the PCoA diagram (**Figure 3**). That analysis grouped the mung bean genotypes from the GBS panel into three main clusters which were consistent with their origins; (1) the first group from India had the most diverse genetic spectrum; (2) a second group from Southeast Asia overlapped slightly with the first group; and (3) a third group consisted of less diverse mung beans were from Central (Uzbekistan), East (China and Korea), and West (Afghanistan, Iran, and Pakistan) Asia or other parts of the world. Many grouping were by country or by region as shown in the dendogram (**Figure 4**). For example, mung beans from Iran were grouped with accessions from Afghanistan, Pakistan and Uzbekistan. The genotypes from the Americas (Argentina, Mexico, and United States) grouped together and were all probably derived from breeding lines.

In physical mapping of the SNP loci, the markers were well distributed across the chromosomes of mung bean and marker density did not decay toward presumed centromeres (**Figure 5**). Overall, average marker density was found to be 19 SNPs per Mb of sequence ranging from 1 to 200 loci. Confidence intervals at 95% probability around the mean (95% CI) for number of SNPs per Mb was at least one up to 55 (**Figure 6A**). Knowing that markers were found rather evenly across the entire genome, a sliding window analysis (window size = 1×10^6 bp, step size = 200 kb) was conducted to understand the patterns of genome wide diversity. Average nucleotide diversity as measured by π was found to be 0.33 per Mb (95% CI, 0.17–0.46, **Figure 6B**). Average Tajima's D was 0.77 per Mb (95% CI, -0.17–1.59, **Figure 6C**).

Marker x Trait Associations (MTAs) per Mineral

Based on the GWAS analysis for each of the seven minerals, a total of 43 SNP markers were found to be highly associated in MTA analysis with seed mineral concentrations measured in either of two different years (**Figure 6** and **Table 2**). These MTA loci were found in 20 different 1 Mb regions (**Figure 6D**) and explained on average 22% of the overall variation (**Supplementary Table S2**). It is worth noting the comparative value of two association models used, one being a general linear model (GLM) and the other a mixed linear model (MLM). In each case, the GWAS results were optimized by considering population structure and germplasm origin by country as fixed effect covariates with kinship additionally used also as a random effect in the MLM analysis. In the GLM analysis, significant associations were observed for K (**Figure 7B**), P (**Figure 7E**), S (**Figure 7G**), and Zn (**Figure 7I**) measured in 2015, and K (**Figure 7C**), P (**Figure 7F**), Fe (**Figure 7A**), and Zn (**Figure 7J**) measured in 2016. In MLM, associations were only detected for S (**Figure 7H**) and Mn (**Figure 7D**) measured in 2016.

The MTAs could be divided into those pertinent to seed macronutrients and those for seed micronutrients. In terms of the first category, 5 MTA loci in three regions were associated with K concentration in 2015, one on chromosome Vr07 (with 2 positive SNPs) and two on chromosome Vr08 (with 1 and 2 SNPs, respectively), explaining on average 30%

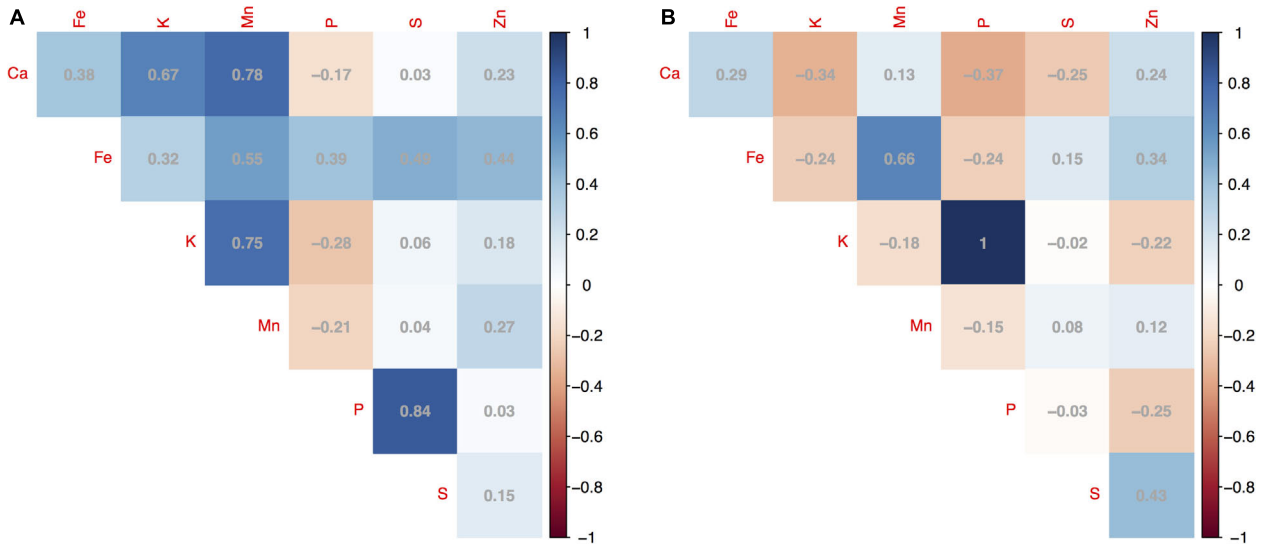

FIGURE 2 | Pairwise correlation coefficients (R^2) for concentrations of seven seed minerals measured for 95 mung bean (*Vigna radiata*) genotypes harvested across 2 years showing **(A)** correlations between minerals in 2015; and **(B)** between the same minerals in 2016. The magnitude of significance shown with a gradient color scale.

of the variation. However, four different SNP markers over two other regions, one on chromosome Vr01 (with 3 SNPs) and the other on chromosome Vr05 (with 1 SNP), were associated with K concentration measured in 2016, explaining on average 27% of the variation. The same four SNP markers and chromosomal regions were also associated with seed P measured in 2016, agreeing with the high correlation between K and P measured that year.

On the other hand, 4 different MTA loci were located in two other genomic regions, both on chromosome Vr07 (with 1 and 3 SNPs, respectively), were associated with seed P concentration measured in the previous year, 2015, explaining on average 23% of the variation. In comparison, 3 MTA loci in only two regions, one on chromosome Vr06 (with 2 SNPs) and the other in chromosome Vr09 (with 1 SNP), were associated with seed P measured in 2016, explaining on average 22% of the variation.

In terms of one of the most important micronutrients for biofortification, 5 SNP markers were associated with seed Fe measured in 2016. These were in a single region in chromosome Vr06 and explained on average 13% of the variation. Interestingly, this MTA region coincided with values of lower nucleotide diversity (π) and Tajima's D profiles (**Figures 6B,C**), suggesting a selective sweep in this region given that the genomic average already accounted for demographic peculiarities such as bottlenecks according to previously used criteria (Cortés et al., 2012a,b; Blair et al., 2016a).

For seed Mn concentration in 2016, 5 SNP markers were associated with three genomic regions, one MTA on chromosome I (with 1 SNP), another MTA on chromosome Vr08 and a further MTA on chromosome Vr09 (both with 2 SNPs each) explaining on average 39% of the variation. Interestingly, the associated region for seed Mn concentration on chromosome Vr08 coincided with the peak in the SNP density profile (**Figure 6D**) of 200 SNPs/Mb with these extreme cases of high diversity usually matching genomic features such as inversions. Most other regions contained less than 60 SNPs per Mb and averaged near 20 SNPs per Mb, in other words tenfold less than for this region.

For seed S concentration as measured in 2015, there were associations with 4 additional SNP markers found in four different regions, three on chromosome Vr07 and one on chromosome Vr09, explaining on average 19% of the variation. In addition, five different SNP markers were associated with S concentration in 2016. These were found in two other regions, both on chromosome Vr08 (with 4 and 1 SNPs, respectively), but only explained very low percent the variation despite having LOD values above 8.5. Four of the five associated markers fell within the same 1 Mb region on Vr08, suggesting that the explanatory power in this region emerges from various SNPs together rather than single marker associations.

Finally, seed Zn concentration in 2015 was associated with 4 markers over three regions, one on chromosome Vr01 (with 1 SNP), one on chromosome Vr07 (with 2 SNPs) and the other one on chromosome Vr08 (with 1 SNP), and explained on average 19% of the variation. One further associated region at the end of chromosome Vr09, though not involving the same associated markers, was associated with both Mn and Zn measured in 2016 (explaining on average 39 and 19% of the variation, respectively).

Overlap of significant SNP for different mineral concentrations was observed for all 4 markers that were associated with both K and P measured in 2016 (explaining on average 27% of the variation), 1 marker at the end of chromosome Vr07 was associated with both P and S measured in 2015 (explaining 23 and 18% of the variation, respectively) and associated with high correlation for these two elements in that year (R^2 = 0.84, **Figure 2A**). On the other hand, the

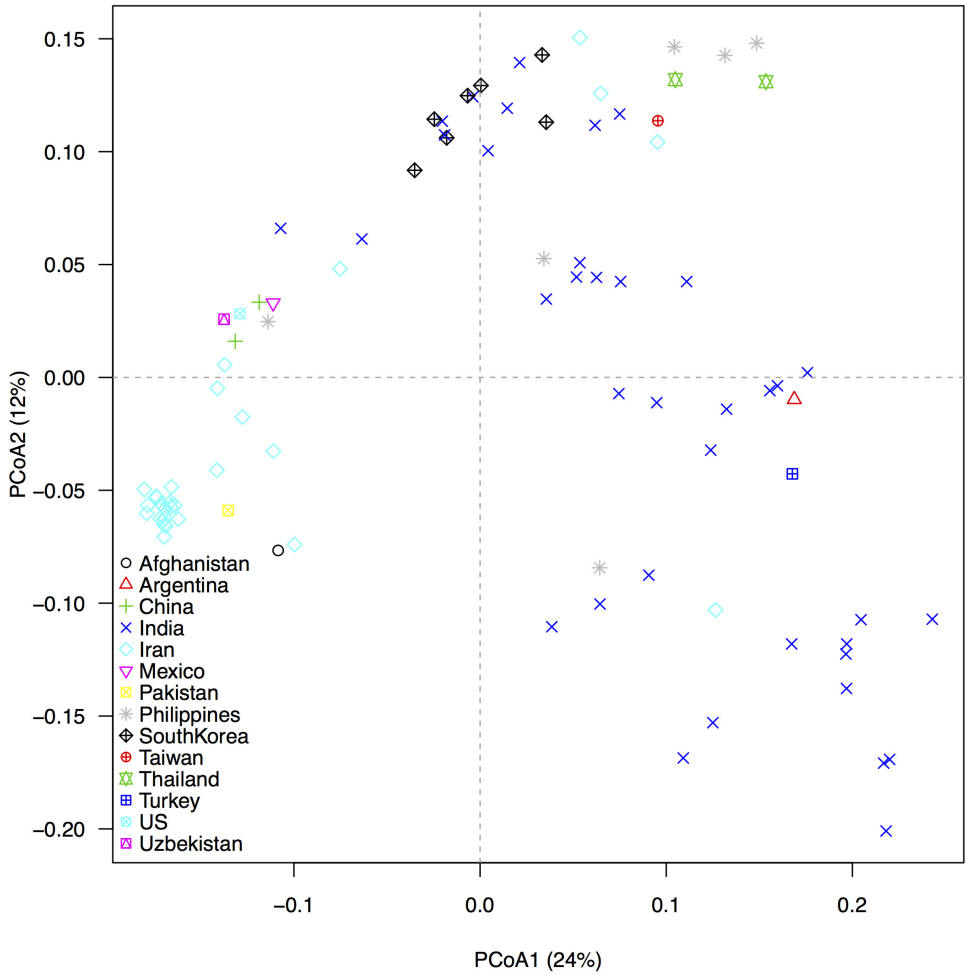

FIGURE 3 | Genetic associations between 95 mung bean (*Vigna radiata*) accessions from the USDA core collection based on 6,486 SNP markers developed from Genotyping by Sequencing (GBS), as revealed by a principal coordinate analysis (PCoA). Different colored symbols identify the country of origin of each accession. The percentage of variation explained by each axis is shown within parenthesis in the label of the corresponding axes.

optimum models for Ca measured in both years (**Supplementary Figure S1**), Fe measured in 2015 (**Supplementary Figure S2**), and Mn measured in 2015 (**Supplementary Figure S4**) did not have any significant SNP associations and therefore were not reported on in the sections above. This might be because of the alluvial nature and slightly acid to neutral pH of the river valley soil used and might not extrapolate to other locations or soils but is fairly typical of production zones for legumes in Tennessee.

Candidate Gene Identification

A total of 35 regions were identified within the windows of 43 significant SNP loci associated with the different mineral concentrations. Of these, 11 regions were associated with seed macronutrients, 12 with micronutrients and 12 with other elements (**Supplementary Table S2**). Totals of 10, 9, and 7 genes were found for each category. For macronutrients, these included three genes on chromosome Vr01 (*Vradi01g00820*, *Vradi01g00830,* and *Vradi01g00840*) and one on Vr05 (*Vradi05g16350*) for K and P concentrations measured in 2016, respectively. Two genes (*Vradi07g26320* and *Vradi07g26340*), both on chromosome Vr07, were near SNPs associated with P concentration in 2015. Three genes (*Vradi07g14180* on chromosome Vr07, and Vradi*08g22740* and *Vradi08g17100* on chromosome Vr08), were identified as associated with K in 2015. These genes that identified with macronutrients were involved in cell proliferation regulation, oxidation-reduction process, ATP biosynthesis pathways, vacuolar transporting pathway, and transcription and translation regulation during plant development. Two genes, *Vradi05g16350* and *Vradi08g15760*, were uncharacterized genes with unknown protein function.

For the micronutrients, a total of 12 regions were found to be associated with Fe and Zn, five with the former and seven with the latter. Genes related to Fe accumulation were all located on chromosome Vr06 and participate in various biological process such as metal iron binding (*Vradi06g09900*), metal translocation (*Vradi06g10020*), uptake of minerals in plants (*Vradi06g10060*), membrane transport (*Vradi06g10120*)

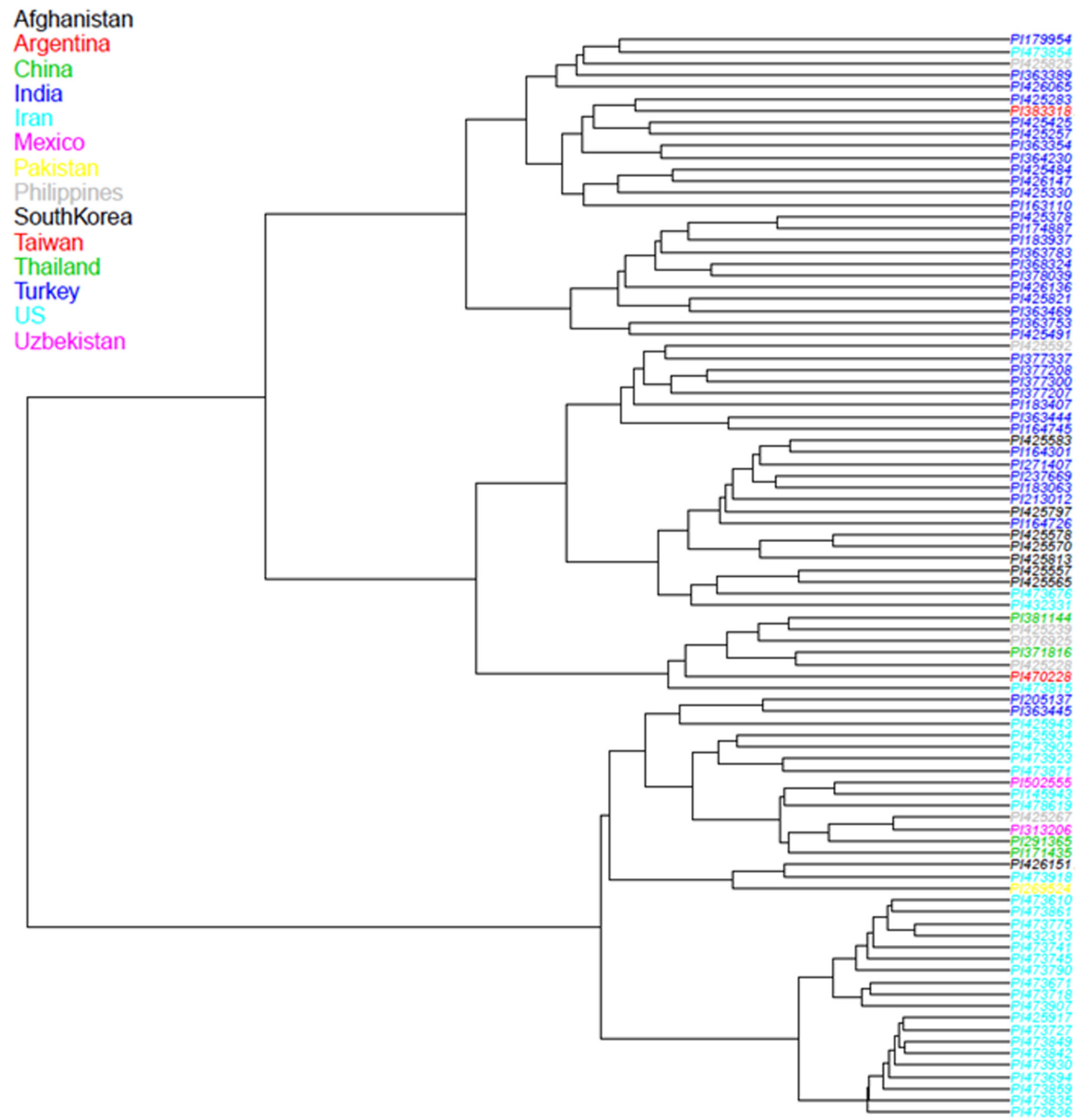

FIGURE 4 | Neighbor-Joining (NJ) dendrogram showing the relationships between 95 mung bean (*Vigna radiata*) accessions from the USDA core collection based on 6,486 SNP markers developed from Genotyping by Sequencing (GBS) with color coding of branches by country and shaded by region in Asia or the Americas. The end of each branch provides the Plant Introduction (PI) number for the genotype.

and ATP binding activities (*Vradi06g10210*). Meanwhile, of the four genes identified for Zn measured in 2015, some were found on chromosome Vr01 (*Vradi01g05570*) and others on chromosome Vr07 (*Vradi07g05950* and *Vradi07g06200*).

These four genes were involved in gene regulation (*Vradi01g05570*), metal translocation (*Vradi07g05950*), metal iron binding (*Vradi07g06200*), and membrane signal transduction pathways (*Vradi06g02380*). One additional gene (*Vradi06g02380*) was identified on chromosome Vr06 that was associated with Zn in 2016; while more were identified for S over two years. S concentration might reflect the accumulation of sulfur containing amino acids such as cysteine and methionine, another nutritionally important trait for legumes. Therefore, we were interested in the loci associated with this mineral even if they explained moderate and low average variability (R^2) in 2015 and 2016, respectively. A total of six genes (*Vradi07g21720*,

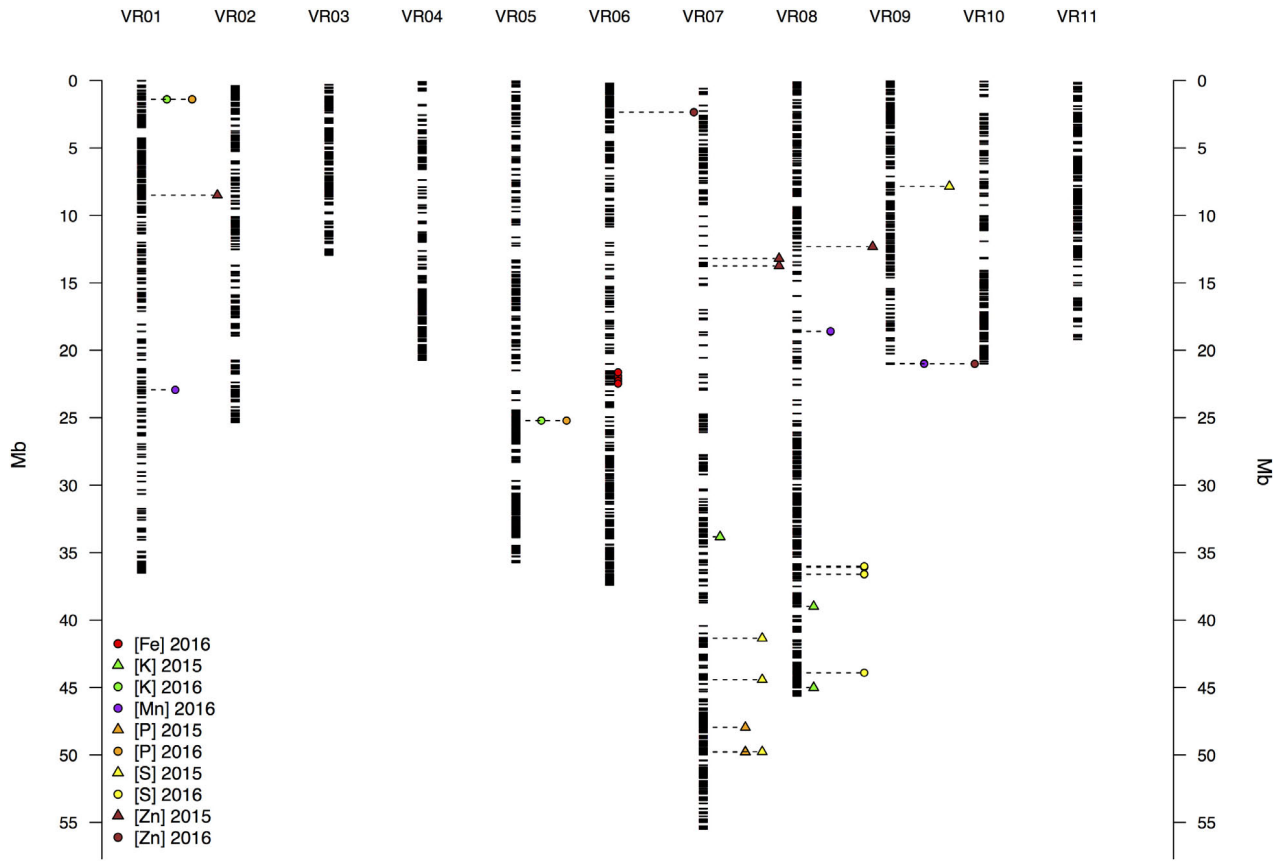

FIGURE 5 | Physical map of the 6,486 SNP markers identified by Genotyping by Sequencing (GBS) of 95 mung bean (*Vigna radiata*) accessions showing all 11 mung bean (VR) chromosomes. Physical position is shown in millions of base pairs (Mb) based on genome sequence in Phytozome. Each black hyphen corresponds to a SNP marker. The other symbols indicate SNP loci significantly associated with various seed mineral concentrations in different years according to the legend and coloring indicating each element analyzed by inductively coupled plasma (ICP) spectroscopy.

Vradi09g05410, *Vradi07g19370*, *Vradi08g16110*, *Vradi08g15850,* and *Vradi08g15760*) with three each per year were identified. As is evident from the gene numbering that reflects the chromosome base position, the last of these three genes, all on Vr08 were very close together in position and this might explain the lower R^2 values as the genes might have been conferring a string association for the region as a whole more so than for each marker individually. One of the genes had unknown function while the other two had cell functions. Finally, a total of four SNPs was associated with Mn measured in either of the two years but only one gene (*Vradi01g11650*) was found in their vicinity. This gene was involved in a carbohydrate metabolic process unrelated to Mn accumulation.

DISCUSSION

The variability found for seed mineral concentration in mung bean agreed with that found in most legumes, especially for the better studied micronutrients, Fe and Zn which have been the subject of biofortification breeding (Dwivedi et al., 2012). Our results showed averages between 30 and 40 ppm of seed Fe and wide variation for individual genotypes in mung beans, which agrees with results for other small seeded legumes such as Mesoamerican common beans (Blair et al., 2013). A narrower range for values and averages between 25 and 30 ppm seed Zn is also typical of small seeded legumes (Gelin et al., 2007; Blair et al., 2009) but we found mung bean to have higher average Zn values, perhaps due to selection rice-based, vegetarian diets.

Large seeded legumes such as Andean common beans or Chickpeas tend to have more Fe but about the same Zn perhaps because of the influence of embryo size differences and the proportion of seed coat to cotyledonary tissues (Ariza-Nieto et al., 2007; Cvitanich et al., 2010). It seems that mung bean is intermediate between these extremes represented by large and small seeded legumes. The distinction is important as small seeded legumes such as mung bean, cowpea and rice bean, all *Vigna* species, are preferred in the diets of Africa and Asia due to their quick cooking time and ease of de-hulling or other processing, compared to large seeded legumes found in European diets and in areas with more fuel sources used for longer cooking times.

Differences in seed macro-nutrient concentrations for Ca, K, and P were significant but variable from year to year and site to

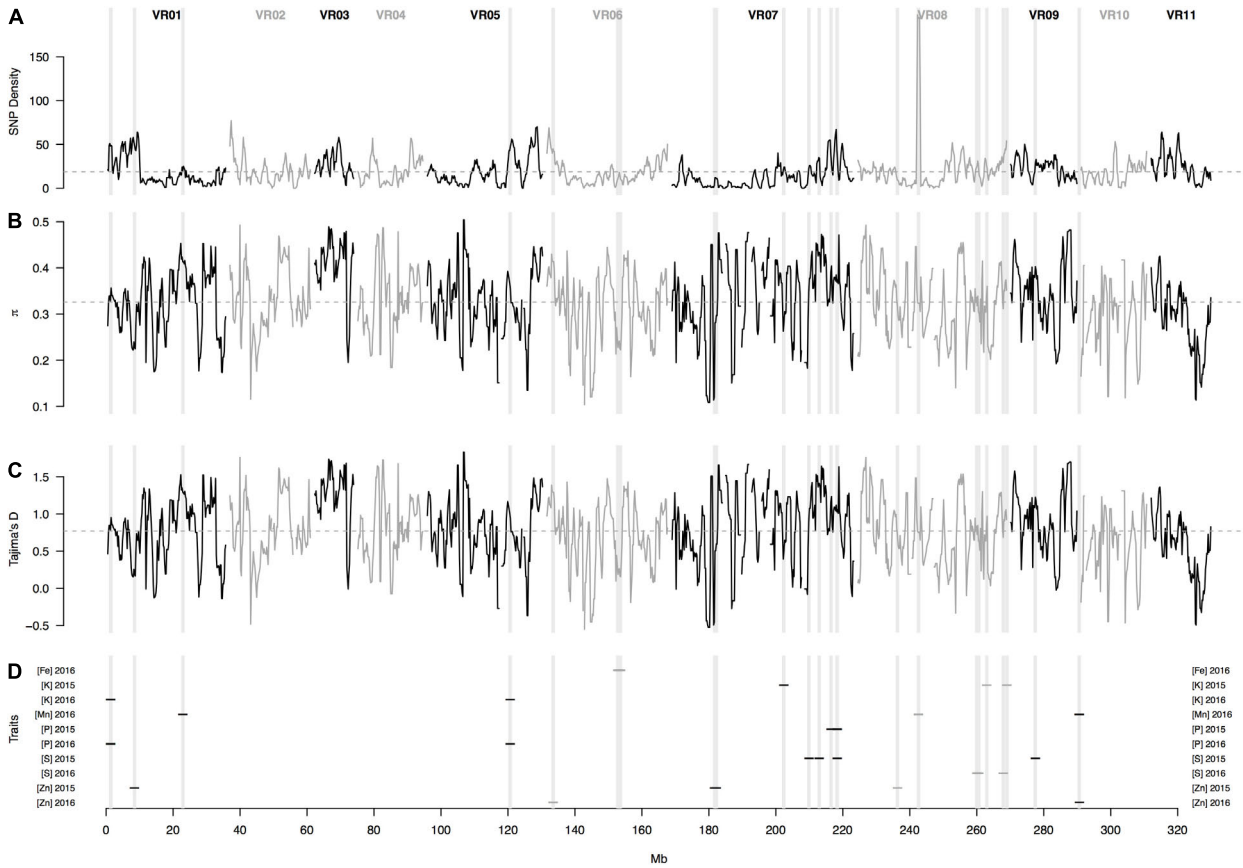

FIGURE 6 | Patterns of genome-wide diversity based on 6,486 SNP markers developed from Genotyping by Sequencing (GBS) of 95 mung bean (*Vigna radiata*) accessions. A sliding window analysis (window size = 1 Mb, step size = 200 Kb) was used to compute **(A)** SNP density, **(B)** nucleotide diversity as measured by π, and **(C)** Tajima's D. **(D)** shows the SNP loci significantly associate with various seed mineral concentrations in two different years. Results of all windowed analyses were plotted against window midpoints in millions of base pairs (Mb). Black and gray colors highlight different mung bean (VR) chromosomes. Gray dashed horizontal lines indicate genome-wide averages. Gray vertical boxes indicate the 1 Mb flanking region of each marker that was associated with a seed mineral locus.

TABLE 2 | Summary statistics for the 43 SNP markers associated with seven seed mineral nutrients measured in 92 mung bean accessions harvested across 2 years.

Nutrient[a]	Year	Number of associated SNPs	Average $-\log_{10}$ (*P*-value)	Number of 1 Mb regions with associated markers	Average R^2 (%)	Number of 1 Mb regions with more than one associated SNP	Number of 1 Mb regions that explained on average more than 10% of the variation
Fe	2016	5	5.11	1	13.4	1	1
K	2016	5	7.08	3	30.4	2	3
K	2016	4	6.04	2	27.2	1	2
Mn	2016	5	6.46	3	38.7	2	3
P	2015	4	5.33	2	22.9	1	2
P	2016	4	6.23	2	27.9	1	2
S	2015	4	4.48	4	19.0	0	4
S	2016	5	8.81	2	1.0	1	0
Zn	2015	4	4.21	3	19.2	1	3
Zn	2016	3	4.61	2	21.8	1	2
Total	NA	43	5.84	24	22.0	11	22

Results are from the optimum genome-wide association analyses and regions are defined by 1 Mb sequences that flanked associated markers. [a]Cases for which there were no associated markers (Ca measured in 2016 and 2017, Fe measured in 2016 and Mn measured in 2016) are not in this table but are found described in ***Supplementary Figures S1, S2, S4****.*

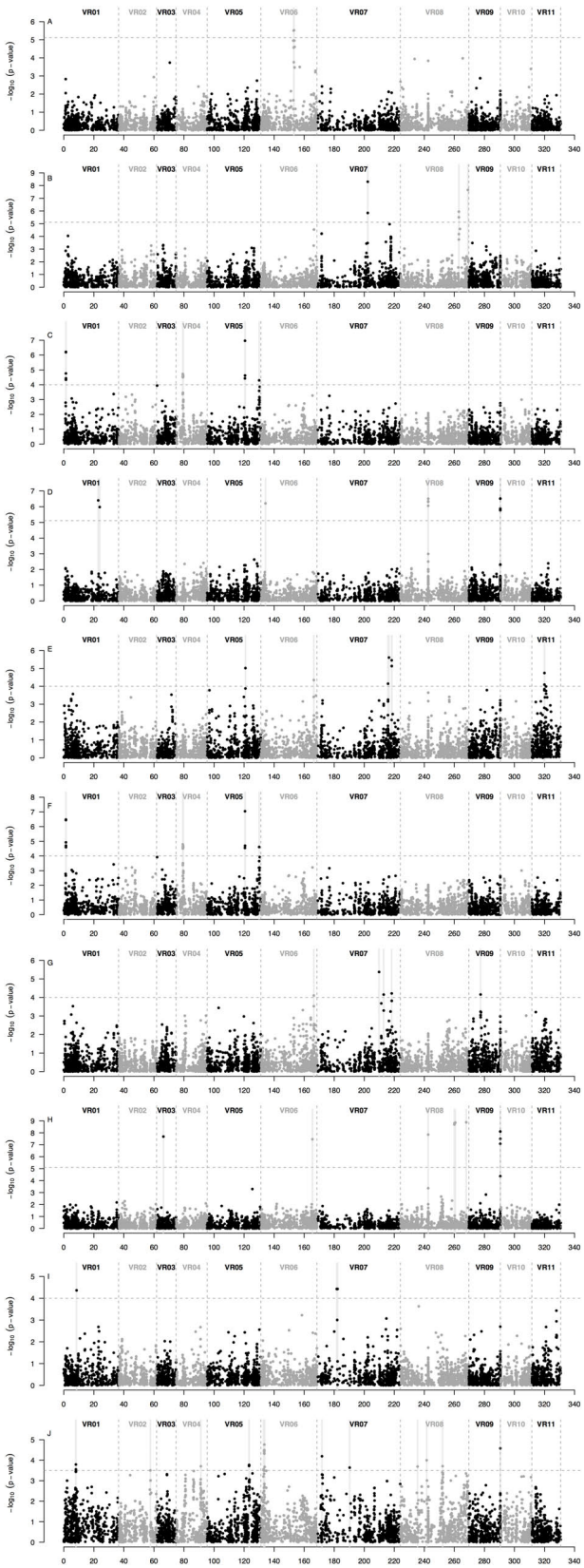

FIGURE 7 | Continued

FIGURE 7 | Manhattan plots of the SNP loci that were significantly associated with seed mineral concentrations in 95 accessions of mung bean (*Vigna radiata*) measured in two different years across a total of 6,486 loci from genotyping by sequencing (GBS) ordered according to physical map position along the 11 mung bean (VR) chromosomes of the Phytozome v.1.0 genome sequence. The plots display probability (*P*)-values per-marker on a -$\log_{10}$ scale for the associations between each SNP locus in order down each chromosome and each seed mineral concentration in each year. Significant marker x trait associations (MTAs) shown for: **(A)** Fe measured in 2016; **(B)** K measured in 2015; **(C)** K measured in 2016; **(D)** Mn measured in 2016; **(E)** P measured in 2015; **(F)** P measured in 2016; **(G)** S measured in 2015; **(H)** S measured in 2016; **(I)** Zn measured in 2015; and **(J)** Zn measured in 2016. The gray dashed horizontal line marks the P-value threshold after Bonferroni-correction for multiple tests. Black and gray colors highlight different mung bean (VR) chromosomes. Gray vertical boxes indicate the 1 Mb flanking region of each marker that was associated. Optimum models include: **(A,G)** a general linear model (GLM) with the country of origin and the first two PCoA axes scores (from **Figure 3**) as covariates; **(B,E,I)** a GLM with the first two PCoA axes scores (from **Figure 3**) as covariates; **(C,F,J)** a GLM with the country of origin as covariate; **(H)** a mixed linear model (MLM) with a centered IBS kinship matrix as a random effect, and the country of origin and the first two PCoA axes scores (from **Figure 3**) as covariates; and **(D)** a MLM with a centered IBS kinship matrix as a random effect and the first two PCoA axes scores (from **Figure 3**) as covariates. Cases for which there were no significant associations between SNP loci and seed mineral concentrations in either year (as for Ca measured in 2015 and 2016, Fe measured in 2015 and Mn measured in 2015) are not depicted but are shown in **Supplementary Figures 1–7**.

site as was Mn. P uptake efficiency is known to vary between legume accessions based on P use efficiency and microbial interactions (Ramaekers et al., 2010; Lynch, 2011). The study of Ca, K, and Mn in legumes is less exhaustive but does show that soil pH is important for uptake. Since we used a neutral soil but relied on previous fertilization, we may have had an effect of field variability. However the field was all part of a rotation of corn/beans.

A second point of our study was that SNP discovery was highly effective by GBS and was useful in distinguishing the mung bean genotypes from the USDA core collection. In general, population diversity analyses found country-based grouping in accordance with isolation by distance model, well-described for self-pollinating legumes such as common bean (Cortés et al., 2012a). In physical mapping of the SNPs, markers were well distributed across the chromosomes of mung bean and marker density did not decay toward presumed centromeres in contrast to what has been observed for other species (Ellegren and Galtier, 2016).

Indeed, GBS has been shown to be an effective method that allows simultaneous discovery and genotyping of a large number of novel SNPs in a lot of species (Davey et al., 2011) as first suggested when the method was conceived and tested (Elshire et al., 2011). The first GBS study in mung bean was conducted by Schafleitner et al. (2016) with a total of 9,289 and 6,463 SNPs mapped in a mung bean genetic map in two RIL populations for bruchid resistance. A year later, an additional 1,321 SNPs were incorporated into 11 linkage groups for mapping of flowering genes using GBS in mung bean (Hwang et al., 2017). Now that the whole genome sequence is available for mung bean (Kang et al., 2014), an added benefit of GBS evaluation is the physical mapping of all the SNPs found.

For GBS to be commonplace in the legumes proper DNA extraction methods are needed. Regular CTAB methods with tissue from fully grown plants (Doyle and Doyle, 1987) usually take 4–6 weeks in legumes which during early growth stages contain many bioactive phytochemicals and an abundance of starch and protein, that remain a major obstacle for high quality genomic DNA isolation, especially in *Vigna* species. In this study, we developed a sprout-based DNA extraction technique that provided enough high-quality DNA for GBS.

A third benefit of the large number of SNPs found in our study was the highly informative PCoA results which classified genotypes accurately based on countries and reflected diversity within the regions. The first axis explained 24% of the overall variation and mostly resolved differences between the temperate and tropical accessions. The second axis explained 12% of the overall variation and uncovered further differences between countries-of-origin as well as within-country variation. Based on both axes, the Indian accessions were the most diverse. The high diversity of mung beans from India is consistent with the hypothesis of domestication there or in an adjacent region but with additional diversity in Eastern, Southwestern and Southeast Asia.

Moderate diversity of mung beans in other countries can be explained by it being mostly selected for vegetable or grain and not being adjacent to wild relatives for inter-crossing. Proximity between regions could have explained the overlap between the cultivated mung bean accessions we analyzed from India and Southeast Asia. Since mung bean is one of the most important food legume crops in South, East and Southeast Asia, where 90% of global production currently occurs, historical germplasm exchange has been extensive for the crop. The unique genetic diversity of Iran mung bean accessions has been reported in other diversity studies (Islam and Blair, 2018; Noble et al., 2018), which could be due to the geographic isolation and breeding history of mung beans in West Asia, where a genetic bottleneck for dryland adaptation could have occurred.

In the fourth and most important of our results, we found that mung beans have a large amount of diversity for nutritional quality and are a tractable genetic system for the study of nutritional genomics. The social impact of this finding for biofortification of crops cannot be overemphasized as mung bean is important in the highly populous countries in Asia where a rice-based diet tends to be low in the micronutrients such as Fe and Zn provided by legumes.

Nutrient-enriched crops offer an important approach to fight nutrient deficiency known as "hidden hunger." Legumes are important as nutrient dense crops and are often considered as poor man's meat in many developing countries. Mung bean seeds exhibit wide variability of macro- and micro-nutrient traits in germplasm collections and inheritance was quantitative in nature governed by many genes and influenced by environment in this study. The concentration of macronutrients in mung bean seeds varied by site and year, with differences probably due to excess precipitation in year 2, suggesting the sensitivity of the genotypes to environmental variations from trial to trial. The water stress could trigger the plant response system for abiotic stress, causing the unstable accumulation of macronutrients in plant tissues.

The micronutrients, Fe, Mn, and Zn, are three of the essential minerals for the growth and development of all higher organisms. In the present study, all three nutrients were positively correlated in each year. The correlation of seed Fe and Zn levels in seeds was reported in many crops and has been extensively studied in other crops such as common bean (Blair et al., 2009), rice (Sperotto et al., 2010), and wheat (Morgounov et al., 2007). The level of Fe and Zn in this study was concurrent with that of previous researchers (Nair et al., 2013, 2015) but was more variable due to the genotype × year interaction seen and perhaps due to within field soil variability. The Mn concentration appeared stable with less variation compared to other micronutrients.

Although Fe and Zn deficiencies are among the most prevalent micronutrient deficiencies in humans, Mn deficiency is less common compare to Fe and Zn, lack of Mn can also cause a serious problem and can lead to asthma and severe birth defects. The transport of Mn is believed to share an entry route with Fe, thus the transportation of Fe, Zn, and Mn overlaps in plant biological systems, leading to joint accumulation in some circumstances (Bashir et al., 2013).

Forty-three SNPs showed significant MTAs with minerals in the seeds of current genotypic panel. Of them, 26 SNPs were associated with macronutrients (K, P, and S). Four markers were found to be associated with both P and K, one marker was found to be associated with both P and S. In total, 17 SNPs were found to be associated with micronutrients Fe, Zn, and Mn. In term of SNP locations for the MTAs within the mung bean genome, macronutrient-associated SNPs were located on chromosomes Vr01, Vr05, Vr07, Vr08, and Vr09, while micronutrient-associated SNPs spanned chromosomes Vr01, Vr06, V07, V08, and Vr09.

The multigenic control of nutrient accumulation in mung bean seeds agrees with previous of Singh V. et al. (2017) showing quantitative inheritance. These authors identified 17 QTLs for Fe and Zn in a mung bean RIL population, with seven QTL on linkage groups LG 6 and LG 7 for Zn with one QTL shared with Fe. In our study, nine out of twelve SNPs associated with Fe and Zn were located on chromosome Vr06 and Vr07. Even though different methods were used to locate genomic regions/genes for Fe and Zn concentration in mung bean, the overlapping regions indicated common genomic regions that are responsible for Fe and Zn accumulation in mung bean. Correlation between minerals was observed for mung bean by Singh R. et al. (2017) and similar results observed in common bean, where the majority of QTLs for Fe and Zn accumulation were located together (Izquierdo et al., 2018). An association study of Fe, Se, and Zn in the garden pea have also indicated that loci related to Fe and Zn are clustered (Diapari et al., 2015).

Among the candidate genes found in this study of mung beans, those MTAs associated with macronutrients were involved with biological process and plant development. This would be expected since Ca, P, K, and S are essential elements for cell proliferation and differentiation. However, candidate genes for micronutrients tended to be more specific.

For seed Fe breeding in mung bean, an interesting genomic region with MTA loci on chromosome Vr06 had the most promising candidate genes for immediate selection. Of the

five genes identified within the 4 Mb significant region, three were related to iron metabolism and movement and have been extensively studied. The first gene included the SNP locus SVR06_21630747 and corresponded to a zinc finger CCCH protein which is related to Fe accumulation in Arabidopsis (Le et al., 2016). The second gene associated with Fe accumulation included the SNP locus SVR06_22114556 and corresponded to a homolog of a bZIP transcription factor, which have been shown to play a role in the uptake of minerals in plants (Assuncao et al., 2010). Finally, the nearby SNP locus SVR06_21903341 corresponded to a homolog of the Yellow Stripe1-Like (*YSL*) gene, which has often been implicated in Fe uptake and transport as well as tolerance to Fe-deficient soils (Curie et al., 2001; Kobayashi and Nishizawa, 2012).

The YSL1 proteins are members of the oligopeptide transporter family and acting as a transporter of iron- and metal-nicotianamine chelates responsible of iron loading of the seeds (Jean et al., 2005). The protein was reported in several crops including rice, maize and soybean, and expressed in several plant tissues such as root and shoot tissues and are important along with Ferric Reductase genes (Waters et al., 2002). We also found a gene similar to *YLS7* on chromosome Vr07 but this one was associated with Zn accumulation. Joint association of Fe and Zn accumulation has been identified in other legumes and a functional SNP marker called "MBkSNP_38" from Islam and Blair (2018) was also discovered to be within a *YSL* gene. As this marker was validated as a KASP assay for mung bean it could be useful for mung bean nutritional breeding even if it was evaluated on a different genotype panel. Although there is no meta-QTL analysts in mung bean yet, a multi population comparison in common bean found YSL to be one of the gene families affecting iron and zinc uptake (Izquierdo et al., 2018).

Similar analyses have been done in other legumes. In *Medicago truncatula*, genes from the Zinc/Iron-regulated transporter-related protein family (*YSL3*) were identified to be associated with seed Zn accumulation (Sankaran et al., 2009). In soybean, a homolog of the *YSL7* gene was found to be related with seed Fe accumulation (Ramamurthy et al., 2014). In chickpea, a gene similar to *YSL1* was also identified to be associated with seed Fe accumulation, along with genes from ZIP and HMA families (Upadhyaya et al., 2016). The implications for breeding across the legumes are obvious. Although mung bean biofortification is still beginning, breeding programs have been shown to produce new micronutrient rich varieties (Blair et al., 2010b).

Large variations of Fe, Zn and the correlation between Fe, Zn, and Mn in mung bean accessions reported here could be useful for breeding biofortified, mung bean varieties to reduce micronutrient deficiency in developing countries. The current study provides the first association study of mineral concentrations in seeds of mung bean. Given the multiple methods of mung bean consumption and their wide popularity especially in India (Dahiya et al., 2014), seed mineral concentration should be studied in relationship to micronutrients levels in food products that can be derived from this legume. These include whole, split or dehulled mung bean dahls; sweets such as burfi and halwa; and snacks such as papad and pakore. Our research demonstrated that the genotypic variability of mung beans and its diversity for seed mineral concentration is sufficient for a biofortification program using available genetic resources in the United States. Testing of bred mung bean varieties for Fe and Zn in India found similar levels of micronutrients with older varieties but suggested that phytic acid must be considered when breeding (Dahiya et al., 2013). Wider diversity for Fe and Zn is found when considering other release varieties from South Asia (Nair et al., 2015). In conclusion, larger numbers of genotypes, multi-site testing, utilization of landraces and consideration of phytate levels must be part of future studies.

CONCLUSION

The major achievements of our study were finding variability for micronutrient and macronutrient traits and a total of 6,484 high-quality SNPs by GBS of 95 mung bean accessions. The SNP loci could be placed on all eleven chromosomes throughout the genome and the overall cost was economical especially as the modified extraction protocol provided enough quality DNA for sequencing. Given the benefits of using GBS in our study, we expect this technology to become widely adopted in mung bean research for marker discovery and application in marker-assisted breeding. Association of genotypic and phenotypic data in our GWAS models and based on our GBS dataset, identified a total of 43 SNP markers that were found to be highly associated with mineral concentration. These SNP loci were found in 20 different 1 Mb regions and explained on average 22% of the overall variation. Further variation could be explained in the future, by additional genotype x environment tests needed both locally and in other climates to evaluate the stability of the GWAS associations found. For biofortification one of the most important associations was for one SNP related to the *YSL1* gene that has been responsible for iron transportation to seeds in other plants. This SNP would show potential to be used in marker-assisted selection (MAS) for higher Fe concentration in mung bean breeding programs. This would be important for South and Southeast Asian diets which are mostly cereal and legume based, and where mung bean can be a very important food for the balanced nutrition of people in these regions. The genotyping by sequencing (GBS) technology is highly efficient for producing large numbers of SNP markers across the genome. Together with the available whole genome information of mung bean, the new loci discovered could greatly facilitate genome-wide association studies in this crop.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in Dryad repository https://doi.org/10.5061/dryad.k0p2ngf5s.

AUTHOR CONTRIBUTIONS

MB, XW, and AI planned the study. XW, AI, and LM did field harvesting and laboratory work. AI, NL, JM, and XW analyzed the seed mineral concentrations. XW, AC, and AI organized the data and prepared graphs.

FUNDING

The following sources of funding included Evans Allen project TENX-07 to Tennessee State University from the United States Department of Agriculture (USDA) for M.S. and Ph.D. student assistantship for AI and XW as well as research funds for LM, JM, and MB. We also acknowledge additional funds for this project from (1) China Scholarship Council (CSC) scholarship for funding of XW student establishment; (2) Fulbright/ICETEX travel funding for MB; (3) Vetenskapsrådet (VR) grant (BS2017-0036) to AC of Uppsala University in Sweden for mobility; and (4) USDA Borlaug Fellows (Asia Biotechnology) scholarship for NL and study leave from the National Institute of Biotechnology of Malaysia. TSU funds for farm work was also gratefully recognized.

ACKNOWLEDGMENTS

We are grateful to USDA station in Griffin for providing the mung bean core collection germplasm used in this study. We are grateful to TSU staff involved in AREC (Agricultural Research and Education Center) at TSU's farm facility, faculty involved in ICP facilities management, and to core TSU funds for Argon gas supply.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00656/full#supplementary-material

REFERENCES

Akaerue, B., and Onwuka, G. (2010). Evaluation of the yield, protein content and functional properties of mungbean [*Vigna radiata* (L.) Wilczek] protein isolates as affected by processing. *Pak. J. Nutr.* 9, 728–735. doi: 10.3923/pjn.2010.728.735

Ariza-Nieto, M., Blair, M. W., Welch, R. M., and Glahn, R. P. (2007). Screening of iron bioavailability patterns in eight bean (*Phaseolus vulgaris* L.) genotypes using the *Caco*-2 cell in vitro model. *J. Agric. Food Chem.* 55, 7950–7956. doi: 10.1021/jf070023y

Assuncao, A. G. L., Herrero, E., Lin, Y.-F., Huettel, B., Talukdar, S., Smaczniak, C., et al. (2010). *Arabidopsis thaliana* transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10296–10301. doi: 10.1073/pnas.1004788107

Ates, D., Aldemir, S., Yagmur, B., Kahraman, A., Ozkan, H., Vandenberg, A., et al. (2018). QTL mapping of genome regions controlling manganese uptake in lentil seed. *G3 Genes Genomes Genet.* 8, 1409–1416. doi: 10.1534/g3.118.200259

Bashir, K., Takahashi, R., Nakanishi, H., and Nishizawa, N. K. (2013). The road to micronutrient biofortification of rice: progress and prospects. *Front. Plant Sci.* 4:15. doi: 10.3389/fpls.2013.00015

Blair, M. W. (2013). Mineral biofortification strategies for food staples: the example of common bean. *J. Agric. Food Chem.* 61, 8287–8294. doi: 10.1021/jf400774y

Blair, M. W., Astudillo, C., Grusak, M. A., Graham, R., and Beebe, S. E. (2009). Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 23, 197–207. doi: 10.1007/s11032-008-9225-z

Blair, M. W., Astudillo, C., Rengifo, J., Beebe, S. E., and Graham, R. (2011). QTL analyses for seed iron and zinc concentrations in an intra-genepool population of Andean common beans (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 122, 511–521. doi: 10.1007/s00122-010-1465-8

Blair, M. W., Izquierdo, P., Astudillo, C., and Grusak, M. A. (2013). A legume biofortification quandary: variability and genetic control of seed coat micronutrient accumulation in common beans. *Front. Plant Sci.* 4:275. doi: 10.3389/fpls.2013.00275

Blair, M. W., Medina, J. I., Astudillo, C., Rengifo, J., Beebe, S. E., Machado, G., et al. (2010a). QTL for seed iron and zinc concentration and content in a mesoamerican common bean (*Phaseolus vulgaris* L.) population. *Theor. Appl. Genet.* 121, 1059–1070. doi: 10.1007/s00122-010-1371-0

Blair, M. W., Monserrate, F., Beebe, S. E., Restrepo, J., and Flores, J. O. (2010b). Registration of high mineral common bean germplasm lines NUA35 and NUA56 from the red-mottled seed class. *J. Plant Regist.* 4, 55–59. doi: 10.3198/jpr2008.09.0562crg

Blair, M. W., Cortés, A. J., and This, D. (2016a). Identification of an Erecta gene and its drought adaptation associations with wild and cultivated common bean. *Plant Sci.* 242, 250–259. doi: 10.1016/j.plantsci.2015.08.004

Blair, M. W., Wu, X., Bhandari, D., and Astudillo, C. (2016b). Genetic dissection of ICP-detected nutrient accumulation in the whole seed of common bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* 7:219. doi: 10.3389/fpls.2016.00219

Blair, M. W., Wu, X., Bhandari, D., Zhang, X., and Hao, J. (2016c). "Role of legumes for and as horticultural crops in sustainable agriculture," in *Organic Farming for Sustainable Agriculture*, ed. D. Nandwani (Cham: Springer), 185–211. doi: 10.1007/978-3-319-26803-3_9

Burlyaeva, M., Vishnyakova, M., Gurkina, M., Kozlov, K., Lee, C. R., Ting, C. T., et al. (2019). Collections of Mungbean [*Vigna radiata* (L.) R. Wilczek] and urdbean [*V. mungo* (L.) Hepper] in Vavilov Institute (VIR): traits diversity and trends in the breeding process over the last 100 years. *Genet. Resour. Crop Evol.* 66:767. doi: 10.1007/s10722-019-00760-2

Caulfield, L. E., Richard, S. A., Rivera, J. A., Musgrove, P., and Black, R. E. (2006). "Stunting, wasting, and micronutrient deficiency disorders," in *Disease Control Priorities in Developing Countries. The International Bank for Reconstruction and Development*, 2nd Edn, eds D. T. Jamison, J. G. Breman, A. R. Measham, et al. (Washington, DC: The International Bank for Reconstruction and Development/The World Bank).

Chen, H., Qiao, L., Wang, L., Wang, S., Blair, M. W., and Cheng, X. (2015). Assessment of genetic diversity and population structure of mung bean (*Vigna radiata*) germplasm using EST-based and genomic SSR markers. *Gene* 566, 175–183. doi: 10.1016/j.gene.2015.04.043

Cheng, P., Holdsworth, W., Ma, Y., Coyne, C. J., Mazourek, M., Grusak, M. A., et al. (2015). Association mapping of agronomic and quality traits in USDA pea single-plant collection. *Mol. Breed.* 35:75. doi: 10.1111/jipb.12118

Cortés, A. J., Chavarro, M. C., Madriñán, S., This, D., and Blair, M. W. (2012a). Molecular ecology and selection in the drought-related Asr gene polymorphisms in wild and cultivated common bean (*Phaseolus vulgaris* L.). *BMC Genet.* 13:58. doi: 10.1186/1471-2156-13-58

Cortés, A. J., This, D., Chavarro, C., Madriñán, S., and Blair, M. W. (2012b). Nucleotide diversity patterns at the drought-related Dreb2 encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 125, 1069–1085. doi: 10.1007/s00122-012-1896-5

Curie, C., Panaviene, Z., Loulergue, C., Dellaporta, S. L., Briat, J. F., and Walker, E. L. (2001). Maize yellow stripe1 encodes a membrane protein directly involved in Fe (III) uptake. *Nature* 409:346. doi: 10.1038/35053080

Cvitanich, C., Przybyłowicz, W. J., Urbanski, D. F., Jurkiewicz, A. M., Mesjasz-Przybyłowicz, J., Blair, M. W., et al. (2010). Iron and ferritin accumulate in separate cellular locations in phaseolus seeds. *BMC Plant Biol.* 10:26. doi: 10.1186/1471-2156-13-26

Dahiya, P. K., Linnemann, A. R., Nout, M. J. R., van Boekel, M. A., and Grewal, R. B. (2013). Nutrient composition of selected newly bred and established mung bean varieties. *Food Sci. Technol.* 54:249e256.

Dahiya, P. K., Linnemann, A. R., Van Boekel, M. A. J. S., Khetarpaul, N., Grewal, R. B., and Nout, M. J. R. (2015). Mung bean: technological and nutritional potential. *Crit. Rev. Food Sci. Nutr.* 55, 670–688. doi: 10.1080/10408398.2012.671202

Dahiya, P. K., Nout, M. J. R., van Boekel, M. A., Khetarpaul, N., Grewal, R. B., and Linnemann, A. (2014). Nutritional characteristics of mung bean foods. *Br. Food J.* 116, 1031–1046. doi: 10.1108/bfj-11-2012-0280

Das, A., Singh, P., Kushwah, N. S., Thakur, S., Rathore, M., Pratap, A., et al. (2018). "Genetic and genomic approaches for improvement in mungbean (*Vigna radiata* L.)," in *Pulse Improvement*, eds S. Wani and M. Jain (Cham: Springer).

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Gene.* 12:499. doi: 10.1038/nrg3012

Day, L. (2013). Proteins from land plants–potential resources for human nutrition and food security. *Trends Food Sci. Technol.* 32, 25–42. doi: 10.1016/j.tifs.2013.05.005

Diapari, M., Sindhu, A., Warkentin, T. D., Bett, K., and Tar'an, B. (2015). Population structure and marker-trait association studies of iron, zinc and selenium concentrations in seed of field pea (*Pisum sativum* L.). *Mol. Breed.* 35:30.

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11–15.

Dwivedi, S. L., Sahrawat, K. L., Rai, K. N., Blair, M. W., Andersson, M. S., and Pfeiffer, W. H. (2012). *Nutritionally Enhanced Staple Food Crops*. Hoboken, NJ: John Wiley & Sons Inc.

Ellegren, H., and Galtier, N. (2016). Determinants of genetic diversity. *Nat. Rev. Gene.* 17:422. doi: 10.1038/nrg.2016.58

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Fatokun, C. A., Menancio-Hautea, D. I., Danesh, D., and Young, N. D. (1992). Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics* 132, 841–846.

Gelin, J. R., Forster, S., Grafton, K. F., McClean, P. E., and Rojas-Cifuentes, G. A. (2007). Analysis of seed zinc and other minerals in a recombinant inbred population of navy bean (*Phaseolus vulgaris* L.). *Crop Sci.* 47, 1361–1366. doi: 10.2135/cropsci2006.08.0510

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9:e90346. doi: 10.1371/journal.pone.0090346

Gwag, J. G., Chung, J. W., Chung, H. K., Lee, J. H., Ma, K. H., Dixit, A., et al. (2006). Characterization of new microsatellite markers in mung bean, *Vigna radiata* (L.). *Mol. Ecol. Resour.* 6, 1132–1134. doi: 10.1111/j.1471-8286.2006.01461.x

He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, H., and Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* 5:484. doi: 10.3389/fpls.2013.00484

Humphry, M., Konduri, V., Lambrides, C., Magner, T., Mcintyre, C., Aitken, E., et al. (2002). Development of a mungbean (*Vigna radiata*) RFLP linkage map and its comparison with lablab (*Lablab purpureus*) reveals a high level of colinearity between the two genomes. *Theor. Appl. Genet.* 105, 160–166. doi: 10.1007/s00122-002-0909-1

Hwang, W. J., Ha, J., Lee, T., Jeong, H., Kim, M. Y., Kim, S. K., et al. (2017). A candidate flowering gene in mungbean is homologous to a soybean phytochrome A gene. *Euphytica* 213, 1–9.

Islam, A., and Blair, M. W. (2018). Molecular characterization of mung bean germplasm from the USDA core collection using newly developed KASP-based SNP markers. *Crop Sci.* 58, 1659–1670. doi: 10.2135/cropsci2018.01.0044

Izquierdo, P., Astudillo, C., Blair, M. W., Iqbal, A. M., Raatz, B., and Cichy, K. A. (2018). Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 131, 1645–1658. doi: 10.1007/s00122-018-3104-8

Jean, M. L., Schikora, A., Mari, S., Briat, J. F., and Curie, C. (2005). A loss-of-function mutation in AtYSL1 reveals its role in iron and nicotianamine seed loading. *Plant J.* 44, 769–782. doi: 10.1111/j.1365-313x.2005.02569.x

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B.-K., et al. (2014). Genome sequence of mungbean and insights into evolution within Vigna species. *Nat. Commun.* 5:5443.

Khazaei, H., Podder, R., Caron, C. T., Kundu, S. S., Diapari, M., Vandenberg, A., et al. (2017). Marker-trait association analysis of iron and zinc concentration in lentil (Medik.) seeds. *Plant Genome* 10:7.

Kim, S. K., Nair, R. M., Lee, J., and Lee, S. H. (2015). Genomic resources in mungbean for future breeding programs. *Front Plant Sci.* 6:626. doi: 10.3389/fpls.2013.00626

Klein, M. A., and Grusak, M. A. (2009). Identification of nutrient and physical seed trait QTL in the model legume *Lotus japonicus*. *Genome* 52, 677–691. doi: 10.1139/g09-039

Kobayashi, T., and Nishizawa, N. K. (2012). Iron uptake, translocation, and regulation in higher plants. *Annu. Rev. Plant Biol.* 63, 131–152. doi: 10.1146/annurev-arplant-042811-105522

Kollárová, K., Vatehová, Z., Slováková, L. U., and Lišková, D. (2010). Interaction of galactoglucomannan oligosaccharides with auxin in mung bean primary root. *Plant Physiol. Biochem.* 48, 401–406. doi: 10.1016/j.plaphy.2010.03.009

Kwon, S.-J., Brown, A. F., Hu, J., Mcgee, R., Watt, C., Kisha, T., et al. (2012). Genetic diversity, population structure and genome-wide marker-trait association analysis emphasizing seed nutrients of the USDA pea (*Pisum sativum* L.) core collection. *Genes Genom.* 34, 305–320. doi: 10.1007/s13258-011-0213-z

Lambrides, C., and Godwin, I. (2007). "Mungbean," in *Pulses, Sugar And Tuber Crops*, ed. C. Kole (Berlin: Springer).

Le, C. T. T., Brumbarova, T., Ivanov, R., Stoof, C., Weber, E., Mohrbacher, J., et al. (2016). Zinc finger of *Arabidopsis thaliana*12 (ZAT12) interacts with FER-like iron deficiency-induced transcription factor (FIT) linking iron deficiency and oxidative stress responses. *Plant Phys.* 170, 540–557. doi: 10.1104/pp.15.01589

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, L., Yang, T., Liu, R., Redden, B., Maalouf, F., and Zong, X. (2017). Food legume production in China. *Crop J.* 5, 115–126. doi: 10.1016/j.cj.2016.06.001

Liu, C., Wu, J., Wang, L., Fan, B., Cao, Z., Su, Q., et al. (2017). Quantitative trait locus mapping under irrigated and drought treatments based on a novel genetic linkage map in mungbean (*Vigna radiata* L.). *Theor. Appl. Genet.* 130, 2375–2393. doi: 10.1007/s00122-017-2965-6

Lynch, J. P. (2011). Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. *Plant Phys.* 156, 1041–1049. doi: 10.1104/pp.111.175414

Ma, Y., Coyne, C. J., Grusak, M. A., Mazourek, M., Cheng, P., Main, D., et al. (2017). Genome-wide SNP identification, linkage map construction and QTL mapping for seed mineral concentrations and contents in pea (*Pisum sativum* L.). *BMC Plant Biol.* 17:43. doi: 10.1186/1471-2156-13-43

Morgounov, A., Gómez-Becerra, H. F., Abugalieva, A., Dzhunusova, M., Yessimbekova, M., Muminjanov, H., et al. (2007). Iron and zinc grain density in common wheat grown in Central Asia. *Euphytica* 155, 193–203. doi: 10.1007/s10681-006-9321-2

Nair, R. M., Thavarajah, D., Thavarajah, P., Giri, R. R., Ledesma, D., Yang, R. Y., et al. (2015). Mineral and phenolic concentrations of mungbean [*Vigna radiata* (L.) R. Wilczek var. radiata] grown in semi-arid tropical India. *J. Food Composit. Anal.* 39, 23–32. doi: 10.1016/j.jfca.2014.10.009

Nair, R. M., Yang, R. Y., Easdown, W. J., Thavarajah, D., Thavarajah, P., Hughes, J. D. A., et al. (2013). Biofortification of mungbean (*Vigna radiata*) as a whole food to enhance human health. *J. Sci. Food Agric.* 93, 1805–1813. doi: 10.1002/jsfa.6110

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia university press.

Noble, T. J., Tao, Y., Mace, E. S., Williams, B., Jordan, D. R., Douglas, C. A., et al. (2018). Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. *Front. Plant Sci.* 8:2102. doi: 10.3389/fpls.2013.02012

Pandey, A. K., Burlakoti, R. R., Kenyon, L., and Nair, R. M. (2018). Perspectives and challenges for sustainable management of fungal diseases of mungbean (*Vigna radiata* (L.) R. Wilczek var. radiata): a review. *Front. Environ. Sci.* 6:53. doi: 10.3389/fpls.2013.00053

Pfeiffer, W. H., and McClafferty, B. (2007). HarvestPlus: breeding crops for better nutrition. *Crop Sci.* 47, S–88–S–105.

Poland, J. A., and Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5, 92–102. doi: 10.3835/plantgenome2012.05.0005

Ramaekers, L., Remans, R., Rao, I. M., Blair, M. W., and Vanderleyden, J. (2010). Strategies for improving phosphorus acquisition efficiency of crop plants. *Field Crop. Res.* 117, 169–176. doi: 10.1016/j.fcr.2010.03.001

Ramamurthy, R. K., Jedlicka, J., Graef, G. L., and Waters, B. M. (2014). Identification of new QTLs for seed mineral, cysteine, and methionine concentrations in soybean [*Glycine max* (L.) Merr.]. *Mol. Breed.* 34, 431–445. doi: 10.1007/s11032-014-0045-z

Rehman, H. M., Cooper, J. W., Lam, H. M., and Yang, S. H. (2019). Legume biofortification is an underexploited strategy for combatting hidden hunger. *Plant Cell Environ.* 42, 52–70. doi: 10.1111/pce.13368

Rohland, N., and Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* 22, 939–946. doi: 10.1101/gr.128124.111

Sankaran, R. P., Huguet, T., and Grusak, M. A. (2009). Identification of QTL affecting seed mineral concentrations and content in the model legume *Medicago truncatula*. *Theor. Appl. Genet.* 119, 241–253. doi: 10.1007/s00122-009-1033-2

Schafleitner, R., Huang, S. M., Chu, S. H., Yen, J. Y., Lin, C. Y., Yan, M. R., et al. (2016). Identification of single nucleotide polymorphism markers associated with resistance to bruchids (*Callosobruchus* spp.) in wild mungbean (*Vigna radiata* var. sublobata) and cultivated V. radiata through genotyping by sequencing and quantitative trait locus analysis. *BMC Plant Bio.* 16:159. doi: 10.1186/1471-2156-13-159

Singh, R., Van Heusden, A. W., and Visser, R. K. (2017). Genetic variation and correlation studies between micronutrient (Fe and Zn), protein content and yield attributing traits in mungbean (*Vigna radiata* L.). *Legume Res.* 41, 167–174.

Singh, V., Yadav, R. K., Yadav, N. R., Yadav, R., Malik, R. S., and Singh, J. (2017). Identification of genomic Regions/genes for high iron and zinc content and cross transferability of SSR markers in mungbean (*Vigna radiata* L.). *Legume Res. Intern. J.* 40:6.

Soetan, K., Olaiya, C., and Oyewole, O. (2010). The importance of mineral elements for humans, domestic animals and plants-A review. *Afr. J. Food Sci.* 4, 200–222.

Somta, P., Musch, W., Kongsamai, B., Chanprame, S., Nakasathien, S., Toojinda, T., et al. (2008). New microsatellite markers isolated from mungbean (*Vigna radiata* (L.) *Wilczek). Mol. Ecol. Resour.* 8, 1155–1157. doi: 10.1111/j.1755-0998.2008.02219.x

Somta, P., and Srinives, P. (2007). Genome research in mungbean (*Vigna radiata* (L.) Wilczek) and blackgram (*V. mungo* (L.) Hepper). *Science Asia* 33, 69–74.

Sperotto, R. A., Boff, T., Duarte, G. L., Santos, L. S., Grusak, M. A., and Fett, J. P. (2010). Identification of putative target genes to manipulate Fe and Zn concentrations in rice grains. *J. Plant Phys.* 167, 1500–1506. doi: 10.1016/j.jplph.2010.05.003

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.

Tangphatsornruang, S., Somta, P., Uthaipaisanwong, P., Chanprasert, J., Sangsrakru, D., Seehalak, W., et al. (2009). Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) *Wilczek). BMC Plant Biol.* 9:137. doi: 10.1186/1471-2156-13-137

Upadhyaya, H. D., Bajaj, D., Das, S., Kumar, V., Gowda, C. L. L., Sharma, S., et al. (2016). Genetic dissection of seed-iron and zinc concentrations in chickpea. *Sci Rep.* 6:24050.

Van, K., Kang, Y. J., Han, K. S., Lee, Y. H., Gwag, J. G., Moon, J. K., et al. (2013). Genome-wide SNP discovery in mungbean by Illumina HiSeq. *Theor. Appl. Genet.* 126, 2017–2027. doi: 10.1007/s00122-013-2114-9

Waters, B. M., Blevins, D. G., and Eide, D. J. (2002). Characterization of FRO1, a pea ferricchelate reductase involved in root iron acquisition. *Plant Physiol.* 129, 85–94. doi: 10.1104/pp.010829

White, P. J., and Broadley, M. R. (2009). Biofortification of crops with seven mineral elements often lacking in human diets–iron, zinc, copper, calcium, magnesium, selenium and iodine. *New Phytol.* 182, 49–84. doi: 10.1111/j.1469-8137.2008.02738.x

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wu, Islam, Limpot, Mackasmiel, Mierzwa, Cortés and Blair. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 01 September 2020
doi: 10.3389/fgene.2020.00744

Association Mapping of Seed Quality Traits Under Varying Conditions of Nitrogen Application in *Brassica juncea* L. Czern & Coss

Javed Akhatar, Mohini Prabha Singh, Anju Sharma, Harjeevan Kaur, Navneet Kaur, Sanjula Sharma, Baudh Bharti, V. K. Sardana and Surinder S. Banga*

DBT Centre of Excellence on Brassicas, Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University,
United States

Reviewed by:
Harsh Raman,
New South Wales Department of Primary Industries, Australia
Jitendra Kumar,
University of Minnesota Twin Cities,
United States

***Correspondence:**
Surinder S. Banga
nppbg@pau.edu

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Genetics

Received: *04 February 2020*
Accepted: *22 June 2020*
Published: *01 September 2020*

Citation:
Akhatar J, Singh MP, Sharma A, Kaur H, Kaur N, Sharma S, Bharti B, Sardana VK and Banga SS (2020) Association Mapping of Seed Quality Traits Under Varying Conditions of Nitrogen Application in Brassica juncea L. Czern & Coss. Front. Genet. 11:744. doi: 10.3389/fgene.2020.00744

Indian mustard (*Brassica juncea*) is a major source of vegetable oil in the Indian subcontinent. The seed cake left after the oil extraction is used as livestock feed. We examined the genetic architecture of oil, protein, and glucosinolates by conducting a genome-wide association study (GWAS), using an association panel comprising 92 diverse genotypes. We conducted trait phenotyping over 2 years at two levels of nitrogen (N) application. Genotyping by sequencing was used to identify 66,835 loci, covering 18 chromosomes. Genetic diversity and phenotypic variations were high for the studied traits. Trait performances were stable when averaged over years and N levels. However, individual performances differed. General and mixed linear models were used to estimate the association between the SNP markers and the seed quality traits. Population structure, principal components (PCs) analysis, and discriminant analysis of principal components (DAPCs) were included as covariates to overcome the bias due to the population stratification. We identified 16, 23, and 27 loci associated with oil, protein, and glucosinolates, respectively. We also established LD patterns and haplotype structures for the candidate genes. The average block sizes were larger on A-genome chromosomes as compared to the B- genome chromosomes. Genetic associations differed over N levels. However, meta-analysis of GWAS datasets not only improved the power to recognize associations but also helped to identify common SNPs for oil and protein contents. Annotation of the genomic region around the identified SNPs led to the prediction of 21 orthologs of the functional candidate genes related to the biosynthesis of oil, protein, and glucosinolates. Notable among these are: *LACS5* (A09), *FAD6* (B05), *ASN1* (A06), *GTR2* (A06), *CYP81G1* (B06), and *MYB44* (B06). The identified loci will be very useful for marker-aided breeding for seed quality modifications in *B. juncea*.

Keywords: Indian mustard, seed oil, seed protein, glucosinolates, genome-wide association study, SNP, quantitative trait loci

INTRODUCTION

Crop Brassicas are comprised of six economically important species belonging to the family *Brassicaceae*. These plants are cultivated as vegetables, fodder, edible oilseeds, or biofuel crops. Of these, Indian mustard (*B. juncea* L. Czern & Coss) is the most widely cultivated oilseed crop in India, with very high acreage (6.9 million hectares) and production (7.2 million metric tonnes) (USDA, 2018–2019). It is also cultivated in China, Southern Russia, and the Caspian steppes as a condiment, vegetable, and oilseed crop. *B. juncea* is an allotetraploid (AABB, 2n = 36) and it evolved through multiple hybridization events between *B. rapa* (AA, 2n = 20) and *B. nigra* (BB, 2n = 16) (Nagaharu, 1935). Breeding programs in this crop are focused on improving productivity and seed quality. Oil, protein, and glucosinolates (GSLs) determine seed quality in Indian mustard. The oil content in brassicas range from 45 to 50% and is mainly made up of unsaturated fatty acids (Smooker et al., 2011). Palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2), linolenic (C18:3), eicosenoic (C20:1), and erucic (C22:1) are the key fatty acids in this crop (Abbadi and Leckband, 2011). Breeding programs emphasize on providing varieties with both high (for industrial applications) and low (suitable for human consumption) erucic acid content (Beare-Rogers et al., 1971; Kaur et al., 2019). High oleic acid, low linoleic acid, or a combination of these have been developed in rapeseed-mustard crops (Appelqvist, 1971). Complex biochemical metabolic reactions with corresponding enzymes in fatty acid biosynthesis have been reviewed by Barker et al. (2007). *FAE1* plays a key role in the synthesis of erucic acid (James and Dooner, 1990) and two orthologs – *BnaA.FAE1.a* and *BnaC.FAE1.a* – are present in *B. napus*, as confirmed by associative transcriptomics (Wu et al., 2008; Harper et al., 2012; Havlickova et al., 2018). A high amount of protein is stored in the seeds of *B. juncea*, and it is reported to be negatively correlated with the oil content (Grami and Stefansson, 1977). Seed cake left after the extraction of oil is used as livestock feed. Two classes of seed storage proteins are prevalent: legumin-type globulins (11S or 12S or cruciferin) and napin-type albumins (2S or napins) (Wanasundara, 2011). The primary structural proteins of *Brassica* oilseeds are oleosins or oil body proteins. Considerable metabolic proteins, such as lipid transfer proteins (LTP), protease inhibitors (Ceciliani et al., 1994), Ca^{2+} dependent-calmodulin binding proteins (Neumann et al., 1996), and dehydrins (Svensson et al., 2000), are well-documented in *Brassica* seeds. Extensive knowledge of the protein structure, structure-function relationships, and genetic control is important for optimal utilization of *Brassica* proteins and for developing new protein-based applications.

Glucosinolates (β-thioglucoside-*N*-hydroxysulfates), a endogenous allelochemical group and a subset of secondary metabolites, are only present in the family *Brassicaceae* (Zhang and Hamauzu, 2004). Glucosinolates are anticancer, antibacterial, anti-fungal, anti-oxidative, and allelopathic compounds (Latté et al., 2011). However, these are considered anti-nutritional components in food as they can affect the thyroid function (Walker and Booth, 2001). Therefore, it is advisable to reduce GSL content in the seeds, and maintain high GSL content in other tissues to prevent herbivore damage and pathogenic microbes. Over 200 GSLs are known in *Brassica* crops. Depending on the precursor amino acid involved, glucosinolates are categorized into aliphatic, aromatic, and indole types (Fahey et al., 2001; Clarke, 2010). Glucoiberin, progoitrin, epiprogoitrin, glucoraphanin, sinigrin, glucoraphenin, gluconapin, glucobarbarin, glucobrassicanapin, glucoerucin, glucobrassicin, and gluconasturtin are the most prevalent GSLs (Wittstock and Halkier, 2002). *B. juncea* contains significant amounts of aliphatic glucosinolates (Yadav and Rana, 2018). Most of the genes responsible for biosynthetic steps are now known in *Arabidopsis thaliana* (Sonderby et al., 2010). A group of R2R3 MYB transcription factors belonging to a single gene family within *Arabidopsis* is involved in the direct transcriptional regulation of GSLs' biosynthesis (Sonderby et al., 2010; Frerigmann and Gigolashvili, 2014). Also, MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis by directly interacting with glucosinolate-related MYB24. Limited knowledge is available for the chain-elongated homophenylalanine-derived aromatic GSLs (Bhandari et al., 2015). Genes controlling aromatic GSL biosynthesis are still uncharacterized. An extensive study and regulation of GSL natural variations in *Brassica* requires a better understanding of its genetic system. All these traits are under the control of complex regulatory mechanisms, strongly affected by the environment. So, phenotyping in replicated and multi-environment trials are important (Becker and Léon, 1988; Wang et al., 2015). Quantitative trait loci (QTL) control complex traits in crop plants and these are better understood by genome wide association studies (GWAS). Associations between phenotypes and markers can be deduced from linkage disequilibrium (LD). GWAS requires the use of numerous genome-wide markers and it provides high mapping resolution by exploiting ancestral recombination events present in a plant species (Soto-Cerda and Cloutier, 2012). It is an alluring approach for the discovery of QTLs in plant genomes associated with phenological, morphological (Honsdorf et al., 2010; Akhatar and Banga, 2015), and seed quality traits (Li et al., 2014). Many studies have been conducted to understand the genetics underlying variations for oil, protein, and glucosinolate contents in *B. napus* and *B. rapa*. However, previous studies were primarily focused on phenotypic evaluations in greenhouse or growth chambers under uniform fertilizer applications. So, there is always a theoretical possibility of missing important loci if the crop growth conditions fail to include limiting environmental variables such as Nitrogen (N), a major limiting factor for *Brassica* production (Danesh-Shahraki et al., 2008). There is a positive correlation between soil N level and seed quality (Ahmadi and Bahrani, 2009). Both protein and oil content are impacted strongly by nitrogen availability as the plants produce particular stress proteins in response to sub-optimal conditions (Bettey and Finch-Savage, 1996). For the present studies, we phenotyped 100 genotypes of *B. juncea* under real farm conditions. The crop was sown under optimal versus limited N fertilization over 2 years to generate N stress and G × N interactions. We then carried out GWAS to investigate the genetic architecture of seed quality-related traits and their stability across environments.

TABLE 1 | Descriptive statistics of the quality traits in *B. juncea* diversity panel during years 2015–2016 and 2016–2017 at two N-level.

Trait	Values	N0		N100	
		Y1	Y2	Y1	Y2
Oil (%)	Mean ± SE	38.3 ± 0.2	37.4 ± 0.4	37.7 ± 0.3	37.6 ± 0.3
	Range	(30.6–43.7)	(26.1–44.5)	(29.7–44.0)	(29.5 to 48.1)
	CV	7.0	9.3	6.8	7.4
	H^2 (%)	44.68	91.27	47.51	63.09
PTN (%)	Mean ± SE	29.3 ± 0.2	29.7 ± 0.2	29.2 ± 0.2	29.2 ± 0.2
	Range	(25.0–34.8)	(25.2–34.4)	(25.5–33.8)	(23.0–33.1)
	CV	6.9	7.4	5.6	5.6
	H^2 (%)	53.36	85.56	34.74	54.17
GSL (μmolg^{-1})	Mean ± SE	93.1 ± 1.5	82.1 ± 1.6	92.4 ± 1.3	78 ± 1.3
	Range	(45.5–130.1)	(38.0–125.3)	(42.8–124.5)	(27.7–107.9)
	CV	16.0	19.1	14.3	16.5
	H^2 (%)	60.24	95.39	16.15	18.34

N0, with no added nitrogen; N100, with added nitrogen @100 kg/ha; Y1, year 1 (2015–2016); Y2, year 2 (2016–2017); SE, standard error; CV, coefficient of variation; PTN, proteins; GSL, glucosinolates; H^2, broad sense heritability.

MATERIALS AND METHODS

Plant Material, Field Trial, and Phenotyping

B. juncea diversity fixed foundation set (*Bj*DFFS), comprising 100 accessions (S_7/S_8 inbred lines), constituted the experimental materials for the present investigations. These accessions were evaluated in the field trials conducted during the winters of 2015–2016 (Y1) and 2016–2017 (Y2) as per the alpha lattice design with two doses of nitrogen (N) application (N0: no added N and N100: added N @100 kg/ha). The experiment was conducted in the oilseeds research area, Punjab Agricultural University, Ludhiana, India (30.9010° N, 75.8573° E). Nitrogen was added in the form of urea (46% N). For N100 treatment, urea was applied in two split doses – half at the time of sowing and the remaining half at the stem elongation stage.

Oil, protein, and glucosinolates content in intact seeds were estimated by Near Infrared Reflectance Spectroscopy (NIRS), Model 6500 spectrophotometer, Foss-NIR Systems, Inc., Silver Spring, MD, United States. Existing NIRS-based calibration models (Sen et al., 2018) were used to estimate seed quality parameters. All the trait values were adjusted for seed moisture content. We used an average of five readings per replication for each test trait.

Statistical and Correlation Network Analysis

Statistical software SAS v9.3 was used for analyses of variance (ANOVA) to assess the significance of variance due to genotype, replication, years, N-levels, and all possible interactions between these (year × genotype, N-level × genotype, year × N-level and replication × N-level). The following model was used for this purpose.

$$P_{ijkl} = \mu + G_i + R_j + Y_k + N_l + (Y \times G)_{kj} + (N \times G)_{lj} + (Y \times N)_{kl} + (R \times N)_{jl} + e_{ijkl}$$

where P_{ijkl} is the phenotypic value of accession, i noted by the lth N-level in year k, μ the overall mean, G_i the effect of the accession i, R_j the effect of replication j, Y_k the effect of year k, N_l the effect of N-level l, $(Y \times G)_{kj}$ the interaction term between year k and genotype j, $(N \times G)_{lj}$ the interaction between N-level l and genotype j, $(Y \times N)_{kl}$ the interaction between year k and N-level l, $(R \times N)_{jl}$ the interaction term between replication j and N-level l, and e_{ijkl} the random independent and identically distributed residual term. This approach was also used for characterizing the influence of the N-levels on the trait observations: multi-year and within-year models through the use of linear mixed models. Best unbiased linear predictors (BLUPs) were assessed to obtain datasets across the years and N-levels. These estimates and predictors extracted from both year and both N-level models were then used as input data for GWAS. Pairwise Pearson's correlation coefficients (r) between the traits were estimated by using the R package "Hmisc"[1]. We visualized r network value with the R package "q-graph."

SNP Genotyping

Total genomic DNA was extracted from young leaves of 100 genotypes using a standard genomic DNA extraction procedure (Doyle and Doyle, 1990), with minor modifications. DNA samples were quantified by visual comparison to λ-DNA standards on ethidium bromide-stained agarose gels. The purity and concentration of the samples were calculated using a spectrophotometer at 260 and 280 nms. High quality DNA samples were genotyped by sequencing (GBS) on the Illumina® HiSeq platform, which was outsourced to Novogene (HK) Company Limited, Hong Kong. GBS data was carried out for only 92 genotypes, as DNA from the remaining eight samples failed the quality test. SNP data calling was carried out using NGSEP (Next Generation Sequencing Experience Platform) GBS pipeline[2] (Duitama et al., 2014). *B. juncea* reference genome v.1.5[3] was used for aligning 25× whole genome sequence of

[1]http://biostat.mc.vanderbilt.edu/wiki/Main/Hmisc
[2]https://github.com/NGSEP/NGSEPcore
[3]http://brassicadb.org/brad

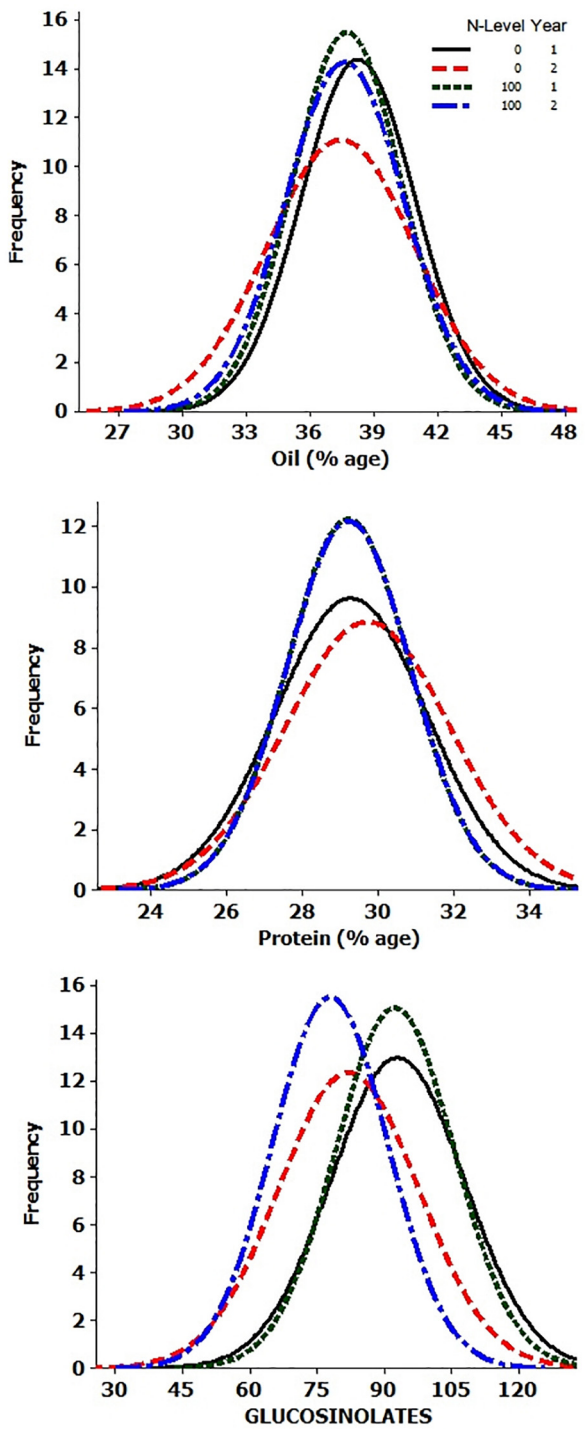

FIGURE 1 | Frequency histograms of 92 genotypes for oil, protein, and glucosinotes.

a commercial *B. juncea* genotype, PBR357, using the software Bowtie2[4]. We preferred to use a mock-up pseudomolecule reference based on the oilseed type mustard cultivar as the available *B. juncea* reference genome assembly was based on a Chinese vegetable mustard genotype, Tumida. To construct a mock-up pseudomolecule reference, total SNPs were replaced in the reference genome using a perl script, PseudoMaker, as implemented in SEG-Map (Zhao et al., 2010). All 92 inbred lines were then aligned using the pseudomolecule genome reference and, using NGSEP-GBS pipeline, SNPs were identified. SNPs with a minor allele frequency (MAF) of >0.05, minimum allele proportion of 0.7, and minimum quality score of 30 were selected for GWAS. After this filtration, a total of 66,835 SNPs were left for association studies. These are uniformly distributed over 18 chromosomes. Imputation of SNPs was performed using fcGENE v1.0.7 (Roshyara and Scholz, 2014) and BEAGLE v3.3.2 (Browning and Browning, 2007). We used R package "GAPIT" (genome association and prediction integrated tool) to transform SNP data into a numeric format (Lipka et al., 2012).

Covariates Analysis

Different covariates, i.e., population structure, principal components (PCs) analysis, and discriminant analysis of principal components (DAPCs), are often used to overcome the bias due to the population stratification on outcomes from GWAS. These allow for adjustments for population stratification. Population structure analysis was performed by using STRUCTURE v2.3.4 (Evanno et al., 2005), with the subgroups (K) ranging from 1 to 10. The Markov Chain Monte Carlo (MCMC) repetitions were set to 10,000. An optimum number of subgroups (Q) was selected based on the log probability of the data [lnP(D)] and *ad hoc* statistic ΔK method. R package, GAPIT was used for the selection of the number of significant PCs with the largest eigenvalues based on all pairs of SNPs. Discriminant Analysis of Principal Components (DAPC) analysis was also used to reduce false positives by reducing the effects of population stratification which were implemented in the R package "adegenet."

Genome-Wide Association Analysis Based on SNP Genotyping

Data for protein, oil, and glucosinolates content were normalized by arcsine transformation. Genome-wide association analysis was performed on normalized values of 92 genotypes and 66,835 SNPs. Phenotypic data were pooled across 2 years for N0, N100, and NP (average values of N0 and N100) levels. Kinship matrix data was generated using the MVP-package of the software R. Q, PCs, and DAPCs analysis were used as covariates in different GWAS analysis algorithms to reduce false positives by reducing the effects of population stratification. Marker trait associations (MTA's) were estimated by MVP (A Memory-efficient Visualization-enhanced and Parallel-accelerated Tool)[5] with default settings to identify marker trait associations. Three methods (GLM, MLM, and FarmCPU) were used with Q, PCs, and DAPCs as covariates. Quantile-Quantile (Q-Q) plots were performed with $-\log_{10}(P)$ of each observed SNP and the expected P value. These allowed us to select best fit algorithm and only

[4]http://bowtie-bio.sourceforge.net/bowtie2

[5]https://github.com/XiaoleiLiuBio/MVP

corresponding SNPs were retained for further analysis. The significance of the association between SNPs and traits was assessed based on an arbitrary threshold of $-\log_{10}(P) \geq 3$.

Functional Annotation and Linkage Disequilibrium (LD)

25 kb region, upstream or downstream of peak SNPs were used for gene finding. To facilitate that, we used the software Blast2GO Pro (Götz et al., 2008) to BLAST against *A. thaliana* database. Annotation allowed us to identify candidates in the vicinity of peak SNPs. SNPs, common across N treatments, were also depicted. LD was estimated between annotated SNPs, calculating the square value of correlation coefficient (r^2) between all pairs of markers by software TASSEL v5.2 (Bradbury et al., 2007) with LD type "sliding window" and LD window size "50."

Haplotype and Linkage Disequilibrium Analysis

Haplotypes were generated from the annotated SNPs. The haplotype and linkage disequilibrium (LD) analysis was carried out by using the software Haploview v4.2 (Barrett et al., 2005). It uses an expectation maximization (EM) algorithm to calculate measures of LD and create a graphical representation of block definitions (Gabriel et al., 2002) to partition the region into segments of strong LD.

Meta-Analysis

The meta-analysis was conducted to test for differences in trait associations between two N-levels and 2 years. This test is based on comparing the differences between the two regression coefficients. We used the meta-analysis software METASOFT26[6]. The diversity panel was first stratified by using software PLINK v1.9 (Purcell et al., 2007). The genome-wide annotated significant SNPs *P*-values from an association test of PLINK results were used in a Binary Effects (BE) model, optimized to detect associations when some studies have an effect and others do not have any effect (Han and Eskin, 2012). ForestPMPlot, an open-source python-interfaced R package, was used to analyze the heterogeneous studies in the meta-analysis by visualizing the effect size differences between N-levels and years. The resulting plot(s) facilitated a better understanding of the heterogeneous genetic effects on the phenotypes at the different N-levels.

RESULTS

Analysis of Variation for Seed Quality Traits

Analysis of variance showed significant differences for the genotypes, years, and N-level × genotype interactions for OIL, PTN, and GSLs (**Supplementary Table 1**). ANOVA also revealed highly significant differences in PTN and GSLs over N levels. Variation due to Y × N interactions was significant for oil content alone. The summary data for oil, protein, and glucosinolates contents are presented in **Table 1** and **Figure 1**. Average values for OIL remained stable over N levels and years. However, individual performances varied for both OIL and PTN contents over N and years. DJ-1-2 DT5, PLM-4, and DJ-27 DTA18 showed high OIL (≥43%) at N0Y1. In comparison, DJ-1-2 DT5 and JC-1359-23-558 accumulated more OIL during N0Y2. T-26-15C-R1631 and JC-1359-23-558 performed better for N100Y1. The same was true for JLM-96 and JC-1359-23-558 at N100Y2. Heritability (H^2) values were generally high, but these were lower during Y1 as compared to Y2 at both N levels. PTN ranged from 23 to 35% over environments. TM117 (35%) and MCP-12-227 (34%) revealed maximum seed PTN for N0Y1, while MCP-12-227, TM117, DT 70, and PBR-357 had higher PTN (≥34%) for N0Y2. DJ-1-2 DT2 and BAUSAM-2 had high PTN (≥34%) at N100Y1. In comparison, DJ-1-2 DT2 and TM117 had maximum PTN (≥33%) at N100Y2. H^2 for PTN ranged from 35 to 86% over years and N-levels. H^2 was lower during Y1 as contrasted to Y2 at both N-levels. Average GSL values varied and these were lower by∼11 and 14 μmol/g during Y2 compared to Y1. The lowest GSL values were noted for the genotypes, JM-06006 (38 μmol/g) and JM-06026 (27.7 μmol/g) for N0Y2 and N100Y2 respectively. H^2 values were lower at N100 compared to N0. Similar to OIL and PTN, H^2 values for GSLs were lower for Y1 as compared to Y2. Oil and PTN were inversely correlated. GSL and PTN were positively associated at N0 (**Supplementary Table 2** and **Supplementary Figure 1**).

GWAS Analysis

We used Q-Q plots to identify a best fit model (**Supplementary Figure 2**).

Oil

We recognized fifteen MTA's involving chromosomes A04, A06, A09, B05, B06, and B08 (**Table 2** and **Figure 2**). The phenotypic variation explained by these loci ranged from 8.82 to 20.37%. Functional annotation predicted seven genes between 0.38 and 13.96 kb on either side of the peak SNPs. We envisaged *CER26* at a distance of 2.08 kb from the SNP A04_6252658. Annotation also called GDSL lipase gene at a distance of 13.96 kb from a group of three SNPs (A06_26456373, A06_26456397, and A06_26456466). This gene was also envisaged close (0.38 kb) to A09_14737522. Another gene, *LACS5* (*long-chain acyl-coA synthetase 5)*, encoding a long-chain-fatty-acid-CoA ligase, was identified near SNPs A09_24082148, A09_24082198, and A09_24083336. We annotated *FAD6* at a distance of 1.64 kb from the nearest of the associated SNPs: B05_5272794, B05_5272795, and B05_5281484. *At1g06090*, a gene encoding delta-9 desaturase-like 1 protein (SNP in the intron), and *ADS1* (2.31 kb) were called on chromosome B06 (B06_19315729). GWAS also allowed for recognition of four SNPs (B08_34645905, B08_34645977, B08_34645991, and B08_34646006) for association with *DIR1*, a gene encoding bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein, annotated at a distance of 3.83 kb from peak SNP.

[6]http://genetics.cs.ucla.edu/meta/

TABLE 2 | Summary of significant SNPs observed for quality traits of the diversity panel in GWAS and meta-analysis.

Trait	Candidate gene (distance from peak SNP in kb)	SNP IDs	NCBI ID	Chr.	Position	GWAS				Meta-analysis	
						$-\log_{10}(P)$	PVE%	A.E.	MAF	Meta *P* value	N-level
Oil	*CER26* (2.08)	A04_6252658	15236357	A04	6252658	3.09	14.82	1.04	0.18	0.00×10^{-35}	N100
	GDSL esterase/lipase gene (13.96)	A06_26456373, 97, 466	15228051	A06	26456373-6466	3.17	16.29	0.47	0.17	4.33×10^{-6}	N0, NP
	GDSL esterase/lipase gene (0.38)	A09_14737522	15228051	A09	14737522	3.19	17.1	-1.12	0.3	2.44×10^{-6}	N0, NP
	LACS5 (within gene, in intronic region)	A09_24082148, 2198, 3336	1032284411	A09	24082148-3336	3.39	15.88	0.45	0.3	1.44×10^{-6}	N0, NP
	FAD6 (1.64)	B05_5272794, 95	15235766	B05	5272794-2795	4.1	20.37	0.82	0.09	3.63×10^{-6}	N0, NP
	At1g06090 (within gene, in intronic region) *ADS1 (2.31)*	B06_19315729	1032298998 15221393	B06	19315729	3.03	11.92	-0.93	0.12	1.67×10^{-7}	N0, N100, NP
	DIR1 (3.83)	B08_34645905, 5977, 5991, 6006	18413820	B08	34645905-6006	3.34	8.82	-0.34	0.4	4.92×10^{-5}	N0, NP
PTN	*PASTICCINO 1* (1.46)	A04_12420732, 5239	3080740	A04	12420732-5239	3.08	16.73	0	0.26	1.13×10^{-6}	N0, N100, NP
	AtP4H3 (within gene, in exonic region)	A06_9257535, 68	18394842	A06	9257535-7568	3.74	16.11	0	0.12	9.18×10^{-5}	N100, NP
	ASN1 (1.08)	A06_15534622, 4718, 4750	15232775	A06	15534622-4750	3.02	14.09	-0.23	0.12	0.00×10^{-96}	N100, NP
	GTR2 (6.84)	A06_17974981, 5007, 5009,5010	15241927	A06	17974981-5010	3.3	15.04	0.35	0.12	0.00×10^{-79}	N100, NP
	At2g41640 (9.24)	A06_27490668, 71, 93	186499036	A06	27490668-0693	4.03	23.21	0.83	0.25	8.72×10^{-6}	N0, NP
	PSP1 (3.71)	A09_30208138	30692244	A09	30208138	3.08	15.94	-1.05	0.11	6.28×10^{-6}	N0, NP
	PSP1 (1.91)	B03_3166770, 6905, 6907, 6926, 6935, 6980, 7018, 7058	42571535	B03	3166770-7058	3.3	17.84	0.28	0.23	2.74×10^{-6}	N0, NP
GSL	*HB16* (2.29) *SK1* (3.35)	A04_11732900, 2912, 2989, 3020, 3023, 3024, 3028, 3029	189303601 189303601	A04	11732900-3029	3.99	15.48	2.37	0.19	7.99×10^{-7}	NP*
	AT2G35450 (1.20)	A05_5070339	79594244	A05	5070339	3.08	13.17	7.37	0.05	1.34×10^{-5}	NP*
	CM1 (1.70)	A06_24763290	9293913	A06	24763290	3.05	13.75	2.73	0.5	6.59×10^{-7}	NP*
	CM1 (1.70)	A06_24773388		A06	24773388	3.34	13.86	-3.05	0.4	2.47×10^{-7}	NP*
	JMT (1.92)	B03_19283625, 26	13676829	B03	19283625-3626	3.05	15.28	0	0.08	0.00×10^{-7}	–
	CM1 (6.51)	B04_2285588, 589, 789	18406100	B04	2285588-5789	3.11	15.53	3.38	0.04	0.00×10^{-73}	N0*
	LINC4 (within gene, in exonic region)	B06_8990634, 55, 70	240256486	B06	8990634-0670	3.55	17.46	-1.51	0.49	6.61×10^{-5}	N0*
	CYP81G1 (12.97) *MYB44* (15.71)	B06_9279577, 727, 731	30698292 1263095	B06	9279577-9731	3.19	15.88	-4.8	0.28	1.35×10^{-5}	N0*

**These had highest m-values and were non-significant for meta-analysis over N-levels. PTN, proteins; GSL, glucosinolates; Chr., chromosome; PVE, phenotypic variation explained; A.E., allelic effect; MAF, minor allele frequency.*

Protein

We found 23 MTA's on chromosomes A04, A06, A09, and B03 (**Table 2** and **Figure 2**). The defined phenotypic variation ranged from 14.09 to 23.21%. Functional annotation predicted seven genes within 1.08–9.24 kb from respective peak SNPs. *PASTICCINO-1 (pas1)* was anticipated at 1.46 kb from SNPs A04_12420732 and A04_12425239. This gene encodes peptidylprolyl isomerase. We also envisage *AtP4H3*

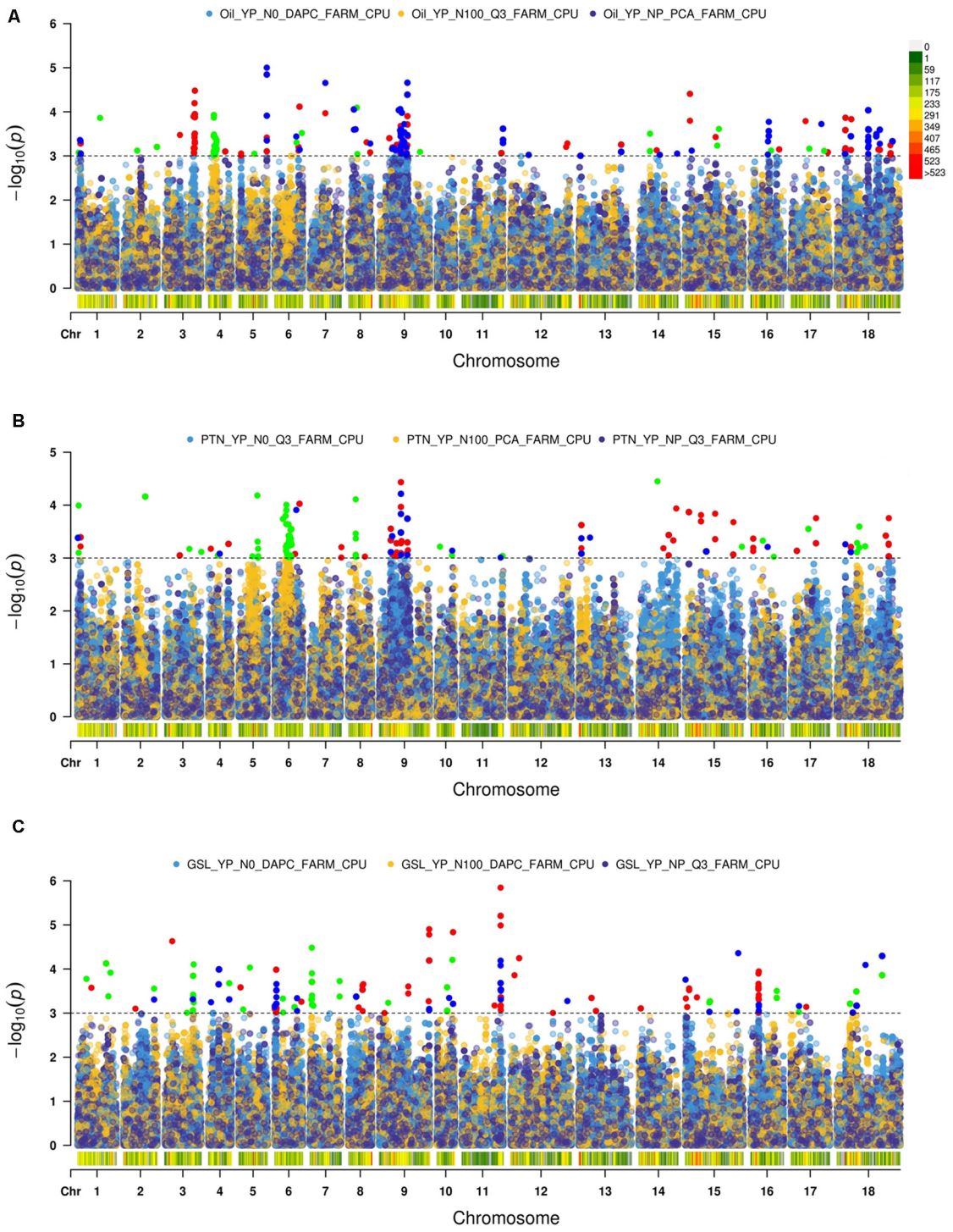

FIGURE 2 | Manhattan plots for Association Analysis of **(A)** oil, **(B)** protein, and **(C)** glucosinolates.

with SNPs (A06_9257535 and A06_9257568) within the exon. This gene codes for procollagen proline 4-dioxygenase. The associated SNPs were recognized in the exonic region of the gene. Three significant SNPs – A06_15534622, A06_15534718, and A06_15534750 – were present at a distance of 1.08 kb from *ASN1*, a gene encoding glutamine dependent-asparagine synthetase. *GTR2* was envisioned at a distance of 6.84 kb from the closest of the associated

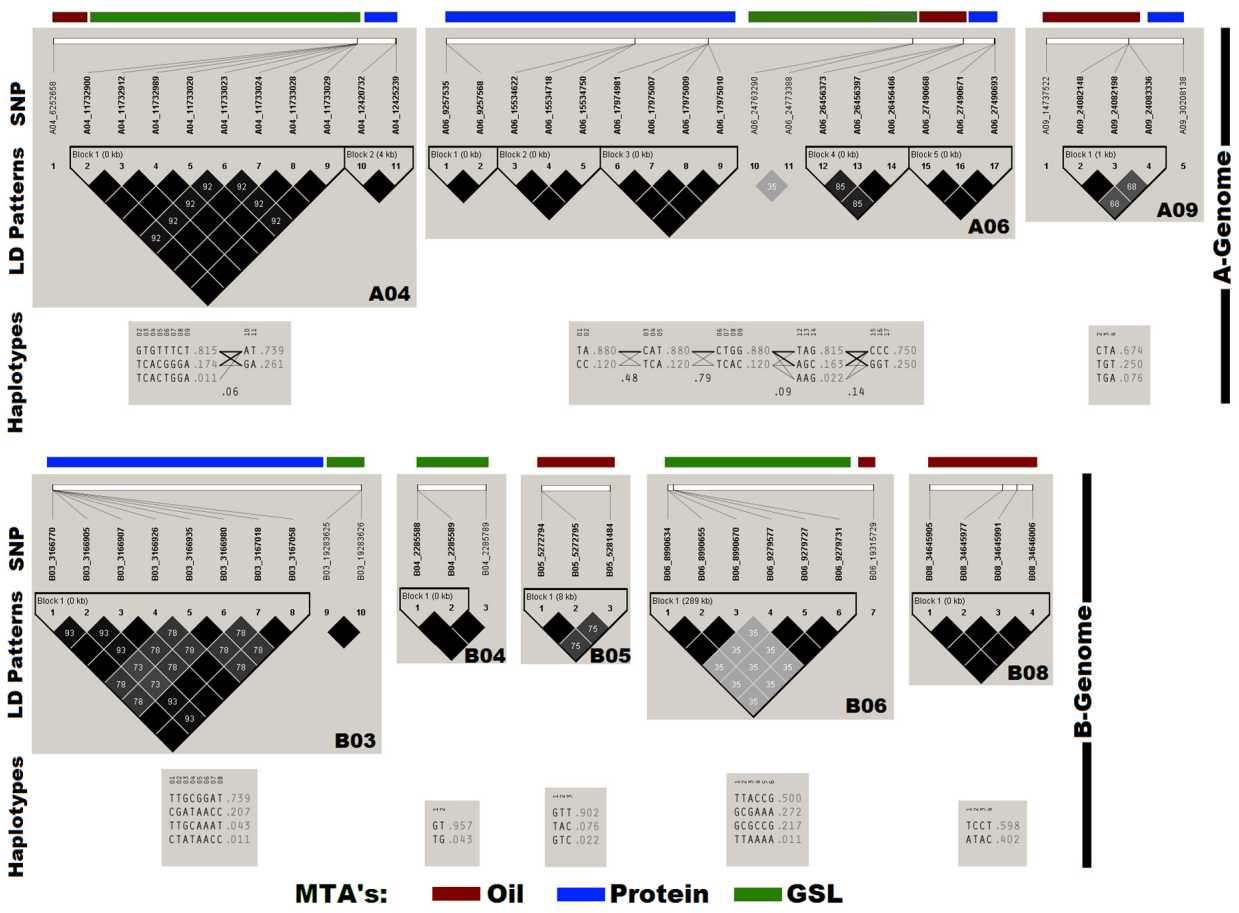

FIGURE 3 | Chromosome-wise linkage disequilibrium (LD) plot generated using Haploview. The plot also depicts the haplotypes block containing the annotated SNPs for OIL, PTN, and GSL. Dark gray colors in LD plots indicate strong LD between markers as estimated by r^2 values.

SNPs: A06_17974981, A06_17975007, A06_17975009, and A06_17975010. *At2g41640*, encoding glycosyltransferase family 61 protein, was predicted at a distance of 9.24 kb from SNPs A06_27490668, A06_27490671, and A06_27490693. We also detected *PSP1* (*PHOSPHOSERINE PHOSPHATASE 1)* on the chromosome A09 at a distance of 3.71 kb from SNP (A09_30208138). This gene was also envisioned next to the peak SNP on the chromosome B03.

Glucosinolates

We identified 22 MTA's involving chromosomes A04, A05, A06, B03, B04, and B06 (**Table 2** and **Figure 2**). The phenotypic variation explained varied from 13.17 to 17.46%. Functional annotation predicted 10 genes in the vicinity (1.20–15.71 kb) of the significant SNPs. Of these, eight SNPs (A04_11732900, 2912, 2989, 3020, 3023, 3024, 3028, and 3029), were located close to the predicted genes *HB16* and *SK1* at the respective distances of 2.29 and 3.35 kb from the peak SNPs. Both these encode shikimate kinase, which catalyzes a step of shikimate pathway for the synthesis of aromatic amino acids (tryptophan, tyrosine, and phenylalanine). We also established a gene *AT2G35450* on chromosome A05 at a distance of 1.2 kb from SNP A05_5070339. Chorismate mutase1 (*CM1*) was predicted close to two SNPs on chromosome A06 at a distance of 1.7 kb and three SNPs for chromosome B04 at a distance of 6.51 kb from the respective peak SNPs. Two significant SNPs, B03_19283625 and B03_19283626, were associated with JMT, encoding jasmonate O-methyltransferase. Three SNPs (B06_8990634, 55, and 70) were detected within the exon of the predicted gene, *LINC4*. This gene encodes branched chain-amino acid-transaminase. *CYP81G1 and MYB44* were envisioned at distances of 12.97 and 15.71 kb from the respective significant SNPs present on chromosome B06.

LD Plot of Annotated SNPs

Haplotype analysis and pair-wise LD estimation was performed in the panel using 60 SNPs annotated for OIL, PTN, and GSL (**Figure 3**). Thirteen out of sixteen SNPs associated with OIL generated four haplotype blocks on chromosomes A06, A09, B05, and B08. We found six haplotype blocks on chromosomes A04, A06, and B03 for PTN with strong r^2 values. The remaining 16 SNPs produced three haplotype blocks on the chromosomes A04, B04, and B06 for GSLs.

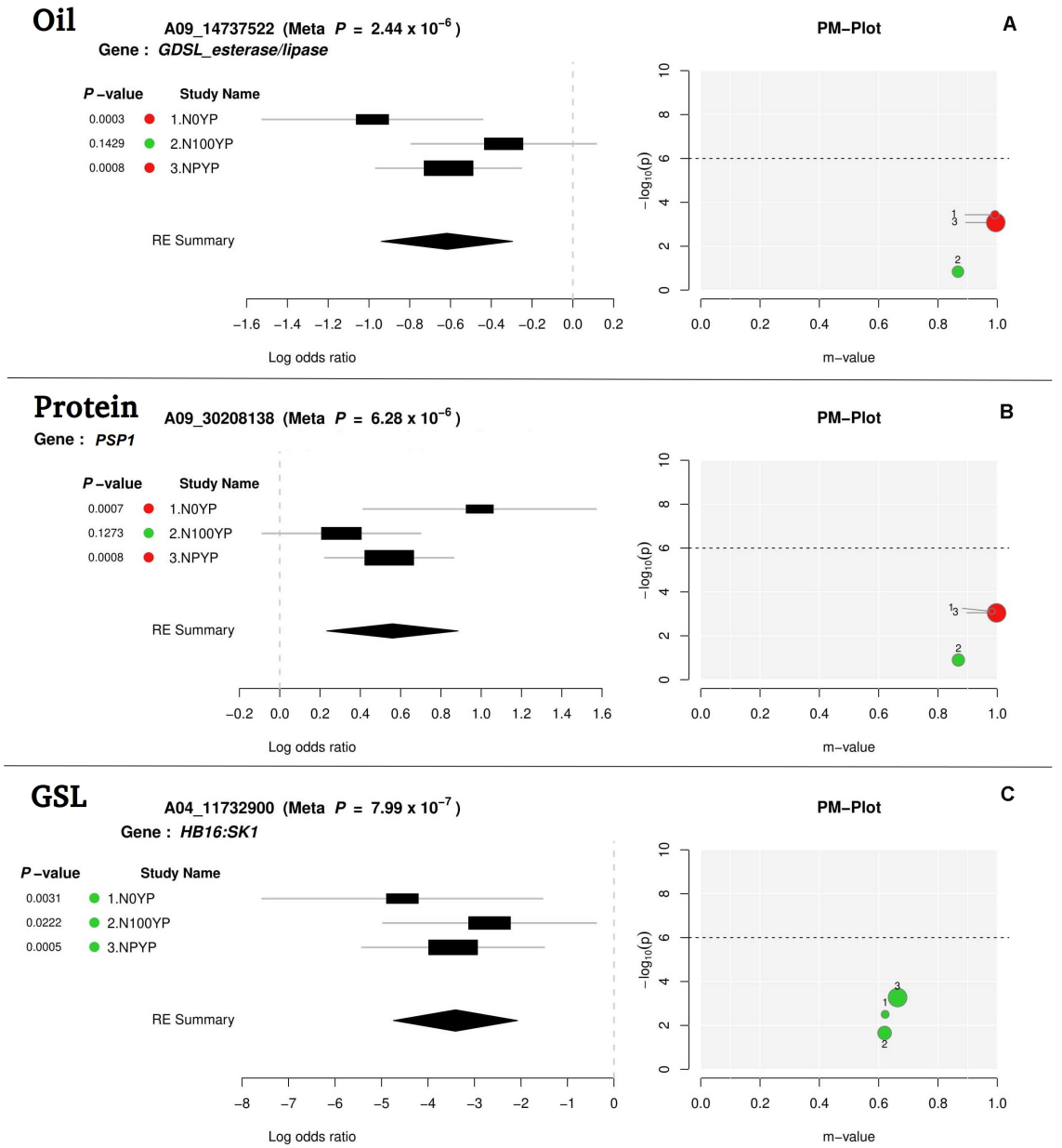

FIGURE 4 | Meta-analysis output file for oil **(A)**, protein **(B)**, and glucosinolates content **(C)** (one gene for each trait).

Meta-Analysis

For meta-analysis, we looked into 21 candidate genes that were significantly associated with OIL, PTN, and GSL. Meta-analysis allowed reconfirmation of most GWAS-SNPs for OIL and PTN, but with higher mapping resolution and an improved number of significant variants (over N levels) (**Table 2** and **Figure 4**). *At1g06090* and *ADS1* were predicted repeatedly at both N levels (meta $p = 1.67 \times 10^{-7}$) with B06_19315729. *DIR1*, predicted on chromosome B08, was significant at both N0 and NP in meta-analysis while this was significant at NP in GWAS. For PTN (**Table 2**), *PASTICCINO1* was called twice on chromosome A04 at meta p value of 1.13×10^{-6}. *AtP4H3*, *ASN1*, and *GTR2* (all on A06) were significant at N0 and NP following meta-analysis. These were significant at N100 alone for GWAS. SNP linked to *At2g41640*, a gene encoding glycosyltransferase family 61 protein, was significant at N0 and NP in meta-analysis (meta $p = 8.72 \times 10^{-6}$). It was significant only at N0 level following GWAS. *PSP1*, with A09_30208138 as associated SNP, was significant at N0 and NP during meta-analysis (meta $p = 6.28 \times 10^{-6}$). For glucosinolates, *HB16* and *SK1* were present close to eight SNPs associated with chromosome A04. These were significant at NP during GWAS. We recorded the highest

TABLE 3 | Details of identified gene copies in A genome.

Trait	AT ID	Gene/protein	Chromosome(s)	
			B. juncea	***B. napus***
Oil	*AT1G06080*	*ADS1*	A09, A10	A6, A10
	AT1G06090	Delta-9 desaturase-like 1 protein	A08, A09, A10	A6, A8, A10
	AT2G03980	*GDSL* esterase/lipase	A02, A06, A09	A2, A2, A9
	AT4G11030	*LACS5*	A02, A09	A6, A8
	AT4G12510	*DIR1*	A03	A3
	AT4G13840	*CER26*	A04	A4
	AT4G30950	*FAD6*	A01, A08	A1, A8
	AT1G31070	*GlcNAc1pUT1*	A05, A08, A09	A8, A9
	AT2G03360	Glycosyltransferase family 61 protein	A06, A09	–
	AT3G47340	GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE	A02, A06	A6
	AT3G54010	*PASTICCINO-1*	A04, A07	A4
	AT5G62680	*GTR2*	A02, A03, A06, A09	A1, A2, A6, A9
GSL	*AT1G19640*	*S*-adenosyl-L-methionine-dependent methyltransferases superfamily protein	A06, A07, A08	A6, A7
	AT2G21940	*SK1* (AT2G21940)	A04	A6
	AT2G35450	AT2G35450	A05	–
	AT3G29200	*CM1*	A06, A09	A6, A9, A9
	AT4G40060	*HB16*	A01, A08	A1, A8
	AT5G65770	*LINC4*	A06, A09	A2, A2
	AT5G67310	*CYP81G1*	A07	A7, A7

but non-significant m values (meta-analysis) for NP (**Table 2**). A similar trend was observed for *CM1* envisaged on chromosome A06 twice. We also recognized *CM1* on chromosome B04 with an m-value of 10^{-73}. It was significant at N0 for GWAS but non-significant following meta-analysis (highest m values at N0). Almost similar trends were recorded for *LINC4*, *CYP81G1*, and *MYB44* on the chromosome B06.

In silico analysis of the genome sequences of *B. napus* and *B. juncea* allowed identification of multiple copies of the genes that predicted seed quality traits on the A genome common to both *B. juncea* and *B. napus* (**Table 3**). A fair degree of parallelism existed, but these were differences in their location on specific A genome chromosomes.

DISCUSSION

The development of high-yielding strains with increased oil and protein contents coupled with low meal glucosinolates in seeds is a major crop improvement goal in *B. juncea*. Simultaneous improvement for these interrelated traits is, however, challenging due to complex genetics and large environmental influences (Banga et al., 2015). Many investigations have been undertaken to understand the genetics of seed oil content in *B. juncea* (Mahmood et al., 2006; Ramchiary et al., 2007; Yadava et al., 2012; Rout et al., 2018), *B. napus* (Qiu et al., 2006; Cao et al., 2010; Zhao et al., 2012; Wang et al., 2013; Jiang et al., 2014; Körber et al., 2016; Fu et al., 2017), and *B. carinata* (Zhang et al., 2017). GWAS was also combined with transcriptome analyses to predict seven functional candidate genes for the seed oil content in *B. napus* (Xiao et al., 2019). However, there are only a few studies involving association mapping for seed quality traits in *B. juncea*. We investigated the genetics of seed quality traits through GWAS based on an association panel. Trait phenotyping was carried out at two doses of N application. ANOVA revealed highly significant differences in PTN and GSLs over N levels, but OIL was relatively stable. Relatively stable expression of oil content is largely in confirmation with an earlier report in mustard (Chauhan et al., 2012). GWAS and meta-GWAS allowed us to examine environment (N) specific SNPs and candidate genes related to OIL, PTN, and GSL. We identified several genes assigned to oil content and fatty acid synthesis. These were placed on both A-(A04, A06, A09) and B-(B05, B06, B08) genome chromosomes. Notable among these were *CER26*, GDSL lipase gene, and *LACS5*. *CER26* has a role in fatty acid elongation (Pascal et al., 2013), while GDSL esterase/lipase (GLIP) gene(s) encode hydrolytic enzymes involved in development and morphogenesis as reported in rice, *Arabidopsis*, and maize (Akoh et al., 2004). GDSL lipase gene was first isolated from *B. napus* (Ling et al., 2006) and QTLs for esterase and lipases family protein have been reported in *B. napus* (Gu et al., 2017). *LACS5* is a floral tissue-specific gene that contributes to oil biosynthesis (Shockey et al., 2002). *FAD6* (called on chromosome B05) encodes delta-12 desaturase that aids in fatty acid desaturation (Mikkilineni and Rocheford, 2003). We envisaged delta9 desaturase (Δ9 DS) and *ADS1* on chromosome B06 at all N levels. Δ9 DS, an intrinsic membrane protein, regulates the catalytic desaturation of saturated fatty acids at the C9 and C10 positions to form unsaturated fatty acids (Stukey et al., 1990). *ADS1* has been used widely for genetically modifying saturated fatty acids in oilseed crops (Yao et al., 2003). Plant lipid-transfer proteins (LTPs) are able to reversibly bind and transport lipids *in vitro* (Kader et al., 1984). We predict gene encoding LTPs on chromosome B08 in contrast to their reported presence on chromosomes A04, A05, and A06 in *B. napus* (Wang et al., 2018). A fair degree of parallelism existed for the genes associated with oil and fatty acid identified for A-genomes of *B. juncea* and *B. napus*. Candidate genes predicted in our studies were predominantly located on chromosomes A04, A06, and A09. Chromosome A09 was also important for oil content in *B. napus* (Xu et al., 2017; Xiao et al., 2019).

Seed meal of *B. juncea* is protein-rich (Grami and Stefansson, 1977; Das et al., 2009) and it is used as livestock feed depending on its nutritional value (Wanasundara, 2011). We anticipated many genes associated with plant development, transport, and protein synthesis on chromosomes A04, A06, A09, and B03. Our

in silico analysis also predicted corresponding orthologs for A03, A04, A06, and A09 for A genome of *B. napus*. *PASTICCINO1*, important for coordinating cell division and differentiation during plant development (Harrar et al., 2003), was envisioned on chromosome A04 at all N levels. Glutamine-dependent asparagine synthetase (*ASN1*) was envisaged on chromosome A06. It is required in nitrogen storage and transport (Canales et al., 2012). Transgenic *Arabidopsis* lines over-expressing *ASN1* produced high seed-soluble protein (Lam et al., 2003). We also identified that *GTR2* is a glucosinolate-specific transporter. It regulates the loading of glucosinolates from the apoplast into the phloem in *Arabidopsis* (Nour-Eldin et al., 2012). Two copies of a functional nucleotidyltransferase gene, *PSP1*, were annotated on chromosomes A09 and B03. *PSP1* encodes UDP-N-acetylglucosamine diphosphorylase 1, a fundamental precursor for glycoprotein and glycolipid synthesis (Yang et al., 2010).

Glucosinolates are the most extensively studied defense-related secondary metabolites, produced exclusively in the family *Brassicaceae* (Halkier and Gershenzon, 2006; Yatusevich et al., 2010). Very large numbers of genes encoding various steps of glucosinolates biosynthesis have been predicted in *B. rapa* (Wang et al., 2011) and *B. napus* (Chalhoub et al., 2014). Expression of these genes is subject to influence by temperature, nitrogen, and sulfur (Barthet and Daun, 2011). We recorded eight genes on chromosomes A04, A05, A06, B03, B04, and B06. These include *HB16* and *SK1*, which encodes shikimate kinase (SK). It is an enzyme of shikimate pathway that directs carbon from the central metabolism pool to a broad range of secondary metabolites (Fucile et al., 2008; Tzin and Galili, 2010; Tohge et al., 2013; Averesch and Krömer, 2018). *CM1*, predicted on chromosomes A06 and B04, is associated with the biosynthesis of phenylalanine and tyrosine. It also participates in tryptophan biosynthesis (Westfall et al., 2014; Qian et al., 2019). *JMT* was annotated on chromosome B03. It is critical for jasmonate-dependent induction of indole glucosinolates in *Arabidopsis* (Brader et al., 2001; Ku et al., 2016). *LINC4* was predicted on B06. It encodes branched-chain-amino-acid transaminase. This enzyme catalyzes the conversion of branched-chain amino acids and α-ketoglutarate into branched chain α-keto acids and glutamate (De Kraker et al., 2007). We identified *CYP81G1* on chromosome B06. It is involved in the metabolic processes of indole glucosinolates (Halkier, 1999). *MYB44* (*AT4G37260*) was envisioned on chromosome B06. *MYB44* regulates the expression of most GSL biosynthesis genes in partnership with *EIN2* (Lü et al., 2013). *In silico* analysis of *B. juncea* genome assembly revealed four orthologous copies (A02, A07, A09, and B06) with a coding sequence comparable to *MYB44*. Three major QTLs are known for GSL content in *B. napus* (Howell et al., 2003; Li et al., 2014; Lu et al., 2014; Qu et al., 2015). These colocalized with three orthologs of the Arabidopsis *MYB28* on chromosomes A09, C02, and C09 (Chalhoub et al., 2014; Lu et al., 2014; Wang et al., 2018).

To summarize, we identified 21 orthologs of the functional candidate genes related to the biosynthesis of OIL, PTN, and GSL. As was expected for a polyploid crop, *in silico* analysis of the reference genome sequence revealed multiple copies of predicted genes on different A- and B- genome chromosomes. We were also able to establish LD patterns and haplotype structures for the candidate genes. The average block sizes were larger on A-genome chromosomes as compared to the B- genome chromosomes. Genetic associations differed over N levels and meta-analysis of GWAS datasets not only improved the power to detect associations but also helped to identify common SNPs. N0 proved better to unravel subtle variations for OIL. In contrast, evaluation at N100 appeared suitable for investigating PTN.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI under bioproject PRJNA639209.

AUTHOR CONTRIBUTIONS

SB developed genetic resources and designed and supervised the research, edited the manuscript. VS helped in the conduct of field trials. SS performed biochemical analysis. JA conducted statistical and bioinformatics analysis. BB helped with statistical analysis. MS and AS annotated the results. MS, AS, HK, JA, and NK interpreted the results and wrote the manuscript. All authors have read and approved the version of manuscript.

FUNDING

The studies were financially supported by the Department of Biotechnology, Government of India in the form of the Centre of Excellence and Innovation in Biotechnology "Germplasm enhancement for crop architecture and defensive traits in *Brassica juncea* L. Czern. and Coss." SB also acknowledges salary support from the Indian Council of Agricultural Research under ICAR National Professor Project "Broadening the genetic base of Indian mustard (*Brassica juncea)* through alien introgressions and germplasm enhancement."

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00744/full#supplementary-material

TABLE S1 | Analysis of Variance (ANOVA).

TABLE S2 | Correlation (*r*) of the quality traits across years (Y1 & Y2) at two N-levels: N0 and N100.

FIGURE S1 | A correlation network plot based on Pearson's correlations (*r*) for seed quality traits across years and N-levels. Black connecting lines denote positive and red connecting lines denote negative correlations (G = GSL = Glucosinolates; P = Protein; O = Oil; N0Y1 = N-level (N0) at Year1; N0Y2 N-level (N0) at Year2; N100Y1 N-level (N100) at Year1 and N100Y2 = N-level (N100) at Year2.

FIGURE S2 | Q-Q plots for selection of the best fitted GWAS algorithm.

REFERENCES

Abbadi, A., and Leckband, G. (2011). Rapeseed breeding for oil content, quality, and sustainability. *Eur. J. Lipid Sci. Technol.* 113, 1198–1206. doi: 10.1002/ejlt.201100063

Ahmadi, M., and Bahrani, M. J. (2009). Yield and Yield Components of Rapeseed as Influenced by Water Stress at Different Growth Stages and Nitrogen Levels. *Am. Eur. J. Agric. Environ. Sci.* 5, 755–761.

Akhatar, J., and Banga, S. S. (2015). Genome-wide association mapping for grain yield components and root traits in *Brassica juncea* (L.) Czern & Coss. *Mol. Breed.* 35, 48–54. doi: 10.1007/s11032-015-0230-8

Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H., and Shaw, J. F. (2004). GDSL family of serine esterases/lipases. *Prog. Lipid Res.* 43, 534–552. doi: 10.1016/j.plipres.2004.09.002

Appelqvist, L. A. (1971). Lipids in Cruciferae: VIII. The fatty acid composition of seeds of some wild or partially domesticated species. *J. Am. Oil Chem. Soc.* 48, 740–744. doi: 10.1007/BF02638533

Averesch, N. J. H., and Krömer, J. O. (2018). Metabolic engineering of the shikimate pathway for production of aromatics and derived compounds-Present and future strain construction strategies. *Front. Bioeng. Biotechnol.* 6:32. doi: 10.3389/fbioe.2018.00032

Banga, S. K., Kumar, P., Bhajan, R., Singh, D., and Banga, S. S. (2015). "Genetics and Breeding," in *Brassica Oilseeds: Breeding and Management*, eds P. R. Kumar, A. Banga, S. S. Meena, and P. D. Kumar, (United Kingdom: CABI), 11–41.

Barker, G. C., Larson, T. R., Graham, I. A., Lynn, J. R., and King, G. J. (2007). Novel insights into seed fatty acid synthesis and modification pathways from genetic diversity and quantitative trait Loci analysis of the Brassica C genome. *Plant Physiol.* 144, 1827–1842. doi: 10.1104/pp.107.096172

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457

Barthet, V. J., and Daun, J. K. (2011). "5 - Seed Morphology, Composition, and Quality," in *Canola*, eds J. K. Daun, N. A. M. Eskin, and D. Hickling, (Urbana, IL: AOCS Press), 119–162. doi: 10.1016/B978-0-9818936-5-5.50009-7

Beare-Rogers, J. L., Nera, E. A., and Heggtvert, H. A. (1971). Cardiac lipid changes in rats fed oils containing long-chain fatty acids. *Can. Inst. Food Technol. J.* 4, 120–124. doi: 10.1016/s0008-3860(71)74194-4

Becker, H. C., and Léon, J. (1988). Stability analysis in plant breeding. *Plant Breed.* 101, 1–23. doi: 10.1111/j.1439-0523.1988.tb00261.x

Bettey, M., and Finch-Savage, W. E. (1996). Respiratory enzyme activities during germination in Brassica seed lots of differing vigour. *Seed Sci. Res.* 6, 165–174. doi: 10.1017/s0960258500003226

Bhandari, S. R., Jo, J. S., and Lee, J. G. (2015). Comparison of glucosinolate profiles in different tissues of nine brassica crops. *Molecules* 20, 15827–15841. doi: 10.3390/molecules200915827

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Brader, G., Tas, É, and Palva, E. T. (2001). Jasmonate-dependent induction of indole glucosinolates in Arabidopsis by culture filtrates of the nonspecific pathogen *Erwinia carotovora*. *Plant Physiol.* 126, 849–860. doi: 10.1104/pp.126.2.849

Browning, B. L., and Browning, S. R. (2007). Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. *Genet. Epidemiol.* 13, 365–375. doi: 10.1002/gepi.20216

Canales, J., Rueda-López, M., Craven-Bartle, B., Avila, C., and Cánovas, F. M. (2012). Novel insights into regulation of asparagine synthetase in conifers. *Front. Plant Sci.* 3:100. doi: 10.3389/fpls.2012.00100

Cao, Z., Tian, F., Wang, N., Jiang, C., Lin, B., Xia, W., et al. (2010). Analysis of QTLs for erucic acid and oil content in seeds on A8 chromosome and the linkage drag between the alleles for the two traits in *Brassica napus*. *J. Genet. Genomics* 37, 231–240. doi: 10.1016/S1673-8527(09)60041-60042

Ceciliani, F., Bortolotti, F., Menegatti, E., Ronchi, S., Ascenzi, P., and Palmieri, S. (1994). Purification, inhibitory properties, amino acid sequence and identification of the reactive site of a new serine proteinase inhibitor from oil-rape (*Brassica napus*) seed. *FEBS Lett.* 342, 221–224. doi: 10.1016/0014-5793(94)80505-9

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 80, 950–953. doi: 10.1126/science.1253435

Chauhan, J. S., Meena, S. S., Singh, K. H., and Meena, M. L. (2012). Environmental effects on genetic parameters for oil and seed meal quality components of Indian mustard (*Brassica juncea* L.). *Indian J. Genet. Plant Breed.* 81, 648–653.

Clarke, D. B. (2010). Glucosinolates, structures and analysis in food. *Anal. Methods* 2, 310–325. doi: 10.1039/b9ay00280d

Danesh-Shahraki, A., Nadian, H., Bakhshandeh, A., Fathi, G., Alamisaied, K., and Gharineh, M. (2008). Optimization of irrigation and nitrogen regimes for rapeseed production under drought stress. *J. Agron.* 7, 321–326. doi: 10.3923/ja.2008.321.326

Das, R., Bhattacherjee, C., and Ghosh, S. (2009). Preparation of mustard (*Brassica juncea* L.) protein isolate and recovery of phenolic compounds by ultrafiltration. *Ind. Eng. Chem. Res.* 48, 4939–4947. doi: 10.1021/ie801474q

De Kraker, J. W., Luck, K., Textor, S., Tokuhisa, J. G., and Gershenzon, J. (2007). Two arabidopsis genes (IPMS1 and IPMS2) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. *Plant Physiol.* 143, 970–986. doi: 10.1104/pp.106.085555

Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15. doi: 10.2307/4119796

Duitama, J., Quintero, J. C., Cruz, D. F., Quintero, C., Hubmann, G., Foulquié-Moreno, M. R., et al. (2014). An integrated framework for discovery and genotyping of genomic variants from high-throughput sequencing experiments. *Nucleic Acids Res.* 42:e44. doi: 10.1093/nar/gkt1381

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Fahey, J. W., Zalcmann, A. T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56, 5–51. doi: 10.1016/S0031-9422(00)00316-312

Frerigmann, H., and Gigolashvili, T. (2014). MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Mol. Plant* 7, 814–828. doi: 10.1093/mp/ssu004

Fu, Y., Zhang, D., Gleeson, M., Zhang, Y., Lin, B., Hua, S., et al. (2017). Analysis of QTL for seed oil content in *Brassica napus* by association mapping and QTL mapping. *Euphytica* 213:17. doi: 10.1007/s10681-016-1817-1819

Fucile, G., Falconer, S., and Christendat, D. (2008). Evolutionary diversification of plant shikimate kinase gene duplicates. *PLoS Genet.* 4:e1000292. doi: 10.1371/journal.pgen.1000292

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. (2002). The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229. doi: 10.1126/science.1069424

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176

Grami, B., and Stefansson, B. R. (1977). Paternal and maternal effects on protein and oil content in summer rape. *Can. J. Plant Sci.* 57, 945–949. doi: 10.4141/cjps77-135

Gu, J., Chao, H., Wang, H., Li, Y., Li, D., and Xiang, J. (2017). Identification of the Relationship between Oil Body Morphology and Oil Content by Microstructure Comparison Combining with QTL Analysis in *Brassica napus*. *Front. Plant Sci.* 7:1989. doi: 10.3389/fpls.2016.01989

Halkier, B. (1999). "glucosinolates," in *Naturally Occurring Glucosides*, ed. R. Ikan, (New York, NY: John Wiley & Sons Ltd), 193–223.

Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 57, 303–333. doi: 10.1146/annurev.arplant.57.032905.105228

Han, B., and Eskin, E. (2012). Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* 8:e1002555. doi: 10.1371/journal.pgen.1002555

Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., et al. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat. Biotechnol.* 30, 798–802. doi: 10.1038/nbt.2302

Harrar, Y., Bellec, Y., Bellini, C., and Faure, J. D. (2003). Hormonal control of cell proliferation requires PASTICCINO genes. *Plant Physiol.* 132, 1217–1227. doi: 10.1104/pp.102.019026

Havlickova, L., He, Z., Wang, L., Langer, S., Harper, A. L., Kaur, H., et al. (2018). Validation of an updated Associative Transcriptomics platform for the polyploid crop species *Brassica napus* by dissection of the genetic architecture of erucic acid and tocopherol isoform variation in seeds. *Plant J.* 93, 181–192. doi: 10.1111/tpj.13767

Honsdorf, N., Becker, H. C., and Ecke, W. (2010). Association mapping for phenological, morphological, and quality traits in canola quality winter rapeseed (*Brassica napus* L.). *Genome* 53, 899–907. doi: 10.1139/G10-049

Howell, P. M., Sharpe, A. G., and Lydiate, D. J. (2003). Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (*Brassica napus* L.). *Genome* 46, 454–460. doi: 10.1139/g03-028

James, D. W., and Dooner, H. K. (1990). Isolation of EMS-induced mutants in Arabidopsis altered in seed fatty acid composition. *Theor. Appl. Genet.* 80, 241–245. doi: 10.1007/BF00224393

Jiang, C., Shi, J., Li, R., Long, Y., Wang, H., Li, D., et al. (2014). Quantitative trait loci that control the oil content variation of rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 127, 957–968. doi: 10.1007/s00122-014-2271-2275

Kader, J. C., Julienne, M., and Vergnolle, C. (1984). urification and characterization of a spinach leaf protein capable of transferring phospholipids from liposomes to mitochondria or chloroplasts. *Eur. J. Biochem* 139, 411–416. doi: 10.1111/j.1432-1033.1984.tb08020.x

Kaur, H., Wang, L., Stawniak, N., Sloan, R., van Erp, H., Eastmond, P., et al. (2019). The impact of reducing fatty acid desaturation on the composition and thermal stability of rapeseed oil. *Plant Biotechnol. J.* 18, 983–991. doi: 10.1111/pbi.13263

Körber, N., Bus, A., Li, J., Parkin, I. A. P., Wittkop, B., Snowdon, R. J., et al. (2016). Agronomic and seed quality traits dissected by genome-wide association mapping in *Brassica napus*. *Front. Plant Sci.* 7:386. doi: 10.3389/fpls.2016.00386

Ku, K. M., Becker, T. M., and Juvik, J. A. (2016). Transcriptome and metabolome analyses of glucosinolates in two broccoli cultivars following jasmonate treatment for the induction of glucosinolate defense to *Trichoplusia ni* (Hübner). *Int. J. Mol. Sci.* 17:1135. doi: 10.3390/ijms17071135

Lam, H. M., Wong, P., Chan, H. K., Yam, K. M., Chen, L., Chow, C. M., et al. (2003). Overexpression of the ASN1 gene enhances nitrogen status in seeds of *Arabidopsis*. *Plant Physiol.* 132, 926–935. doi: 10.1104/pp.103.020123

Latté, K. P., Appel, K. E., and Lampen, A. (2011). Health benefits and possible risks of broccoli - An overview. *Food Chem. Toxicol.* 49, 3287–3309. doi: 10.1016/j.fct.2011.08.019

Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., et al. (2014). Genome-Wide Association study dissects the genetic architecture of seed weight and seed quality in Rapeseed (*Brassica napus* L.). *DNA Res.* 21, 355–367. doi: 10.1093/dnares/dsu002

Ling, H., Zhao, J., Zuo, K., Qiu, C., Yao, H., Qin, J., et al. (2006). Isolation and expression analysis of a GDSL-like lipase gene from Brassica napus L. *J. Biochem. Mol. Biol.* 39, 297–303. doi: 10.5483/bmbrep.2006.39.3.297

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Lü, B. B., Li, X. J., Sun, W. W., Li, L., Gau, R., Zhu, Q., et al. (2013). AtMYB44 regulates resistance to the green peach aphid and diamondback moth by activating EIN2-affected defences in Arabidopsis. *Plant Biol.* 15, 841–850. doi: 10.1111/j.1438-8677.2012.00675.x

Lu, G., Harper, A. L., Trick, M., Morgan, C., Fraser, F., O'Neill, C., et al. (2014). Associative transcriptomics study dissects the genetic architecture of seed glucosinolate content in *Brassica napus*. *DNA Res.* 21, 613–625. doi: 10.1093/dnares/dsu024

Mahmood, T., Rahman, M. H., Stringam, G. R., Yeh, F., and Good, A. G. (2006). Identification of quantitative trait loci (QTL) for oil and protein contents and their relationships with other seed quality traits in *Brassica juncea*. *Theor. Appl. Genet.* 113, 1211–1220. doi: 10.1007/s00122-006-0376-371

Mikkilineni, V., and Rocheford, T. R. (2003). Sequence variation and genomic organization of fatty acid desaturase-2 (fad2) and fatty acid desaturase-6 (fad6) cDNAs in maize. *Theor. Appl. Genet.* 106, 1326–1332. doi: 10.1007/s00122-003-1190-1197

Nagaharu, U. (1935). Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn. J. Bot.* 7, 389–452.

Neumann, G. M., Condron, R., and Polya, G. M. (1996). Purification and sequencing of napin-like protein small and large chains from *Momordica charantia* and *Ricinus communis* seeds and determination of sites phosphorylated by plant Ca2+-dependent protein kinase. *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 1298, 223–240. doi: 10.1016/S0167-4838(96)00133-131

Nour-Eldin, H. H., Andersen, T. G., Burow, M., Madsen, S. R., Jørgensen, M. E., Olsen, C. E., et al. (2012). NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature* 488, 531–534. doi: 10.1038/nature11285

Pascal, S., Bernard, A., Sorel, M., Pervent, M., Vile, D., Haslam, R. P., et al. (2013). The Arabidopsis cer26 mutant, like the cer2 mutant, is specifically affected in the very long chain fatty acid elongation process. *Plant J.* 73, 733–746. doi: 10.1111/tpj.12060

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qian, Y., Lynch, J. H., Guo, L., Rhodes, D., Morgan, J. A., and Dudareva, N. (2019). Completion of the cytosolic post-chorismate phenylalanine biosynthetic pathway in plants. *Nat. Commun.* 10:15. doi: 10.1038/s41467-018-07969-7962

Qiu, D., Morgan, C., Shi, J., Long, Y., Liu, J., Li, R., et al. (2006). A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. *Theor. Appl. Genet.* 114, 67–80. doi: 10.1007/s00122-006-0411-412

Qu, C. M., Li, S. M., Duan, X. J., Fan, J. H., Jia, L. D., Zhao, H. Y., et al. (2015). Identification of candidate genes for seed glucosinolate content using association mapping in *Brassica napus* L. *Genes* 6, 1215–1229. doi: 10.3390/genes6041215

Ramchiary, N., Bisht, N. C., Gupta, V., Mukhopadhyay, A., Arumugam, N., Sodhi, Y. S., et al. (2007). QTL analysis reveals context-dependent loci for seed glucosinolate trait in the oilseed *Brassica juncea*:importance of recurrent selection backcross scheme for the identification of "true" QTL. *Theor. Appl. Genet.* 116, 77–85. doi: 10.1007/s00122-007-0648-644

Roshyara, N. R., and Scholz, M. (2014). fcGENE: a versatile tool for processing and transforming SNP datasets. *PLoS One* 9:e97589. doi: 10.1371/journal.pone.0097589

Rout, K., Yadav, B. G., Yadava, S. K., Mukhopadhyay, A., Gupta, V., Pental, D., et al. (2018). QTL landscape for oil content in *Brassica juncea*: analysis in multiple bi-parental populations in high and "0" erucic background. *Front. Plant Sci.* 9:1448. doi: 10.3389/fpls.2018.01448

Sen, R., Sharma, S., Kaur, G., and Banga, S. S. (2018). Near-infrared reflectance spectroscopy calibrations for assessment of oil, phenols, glucosinolates and fatty acid content in the intact seeds of oilseed Brassica species. *J. Sci. Food Agric.* 98, 4050–4057. doi: 10.1002/jsfa.8919

Shockey, J. M., Fulda, M. S., and Browse, J. A. (2002). Arabidopsis contains nine long-chain acyl-coenzyme A synthetase genes that participate in fatty acid and glycerolipid metabolism. *Plant Physiol.* 129, 1710–1722. doi: 10.1104/pp.003269

Smooker, A. M., Wells, R., Morgan, C., Beaudoin, F., Cho, K., Fraser, F., et al. (2011). The identification and mapping of candidate genes and QTL involved in the fatty acid desaturation pathway in *Brassica napus*. *Theor. Appl. Genet.* 122, 1075–1090. doi: 10.1007/s00122-010-1512-1515

Sonderby, I. E., Burow, M., Rowe, H. C., Kliebenstein, D. J., Halkier, B. A., Sønderby, I. E., et al. (2010). A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in *Arabidopsis*. *Plant Physiol.* 153, 348–363. doi: 10.1104/pp.109.149286

Soto-Cerda, J. B., and Cloutier, S. (2012). "Association mapping in plant genomes," in *Genetic Diversity in Plants*, eds A. Muhammed, R. Aksel, R. Aksel, and R. C. Von Borstel, (London: InTech), doi: 10.5772/33005

Stukey, J. E., McDonough, V. M., and Martin, C. E. (1990). The OLE1 gene of Saccharomyces cerevisiae encodes the $Δ$9 fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase gene. *J. Biol. Chem.* 265, 20144–20149.

Svensson, J., Palva, E. T., and Welin, B. (2000). Purification of recombinant Arabidopsis thaliana dehydrins by metal ion affinity chromatography. *Protein Expr. Purif.* 20, 169–178. doi: 10.1006/prep.2000.1297

Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A. R. (2013). Shikimate and phenylalanine biosynthesis in the green lineage. *Front. Plant Sci.* 4:62. doi: 10.3389/fpls.2013.00062

Tzin, V., and Galili, G. (2010). The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in *Arabidopsis thaliana*. *Arab. B.* 8, e0132. doi: 10.1199/tab.0132

Walker, K. C., and Booth, E. J. (2001). Agricultural aspects of rape and other *Brassica* products. *Eur. J. Lipid Sci. Technol.* 103, 441–446. doi: 10.1002/1438-9312(200107)103:7(441:AID-EJLT441(3.0.CO;2-D

Wanasundara, J. P. (2011). Proteins of brassicaceae oilseeds and their potential as a plant protein source. *Crit. Rev. Food Sci. Nutr.* 51, 635–677. doi: 10.1080/10408391003749942

Wang, B., Wu, Z., Li, Z., Zhang, Q., Hu, J., Xiao, Y., et al. (2018). Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in *Brassica napus*. *Plant Biotechnol. J.* 16, 1336–1348. doi: 10.1111/pbi.12873

Wang, H., Wu, J., Sun, S., Liu, B., Cheng, F., Sun, R., et al. (2011). Glucosinolate biosynthetic genes in *Brassica rapa*. *Gene.* 487, 135–142. doi: 10.1016/j.gene.2011.07.021

Wang, X., Long, Y., Yin, Y., Zhang, C., Gan, L., Liu, L., et al. (2015). New insights into the genetic networks affecting seed fatty acid concentrations in *Brassica napus*. *BMC Plant Biol.* 15:91. doi: 10.1186/s12870-015-0475-478

Wang, X., Wang, H., Long, Y., Li, D., Yin, Y., Tian, J., et al. (2013). Identification of QTLs associated with oil content in a high-oil *Brassica napus* cultivar and construction of a high-density consensus map for QTLs comparison in *B. napus*. *PLoS One* 8:e80569. doi: 10.1371/journal.pone.0080569

Westfall, C. S., Xu, A., and Jez, J. M. (2014). Structural evolution of differential amino acid effector regulation in plant chorismate mutases. *J. Biol. Chem.* 289, 28619–28628. doi: 10.1074/jbc.M114.591123

Wittstock, U., and Halkier, B. A. (2002). Glucosinolate research in the *Arabidopsis era*. *Trends Plant Sci.* 7, 263–270. doi: 10.1016/S1360-1385(02)02273-2272

Wu, G., Wu, Y., Xiao, L., Li, X., and Lu, C. (2008). Zero erucic acid trait of rapeseed (*Brassica napus* L.) results from a deletion of four base pairs in the fatty acid elongase 1 gene. *Theor. Appl. Genet.* 116, 491–499. doi: 10.1007/s00122-007-0685-z

Xiao, Z., Zhang, C., Tang, F., Yang, B., Zhang, L., Liu, J., et al. (2019). Identification of candidate genes controlling oil content by combination of genome-wide association and transcriptome analysis in the oilseed crop *Brassica napus*. *Biotechnol. Biofuels* 12:206. doi: 10.1186/s13068-019-1557-x

Xu, P., Cao, S., Hu, K., Wang, X., Huang, W., Wang, G., et al. (2017). Trilocular phenotype in *Brassica juncea* L. resulted from interruption of CLAVATA1 gene homologue (BjMc1) transcription. *Sci. Rep.* 7:3498.

Yadav, M., and Rana, J. S. (2018). Quantitative analysis of Sinigrin in *Brassica juncea*. *J. Pharmacogn. Phytochem.* 7, 948–954.

Yadava, S. K., Arumugam, N., Mukhopadhyay, A., Sodhi, Y. S., Gupta, V., Pental, D., et al. (2012). QTL mapping of yield-associated traits in *Brassica juncea*: meta-analysis and epistatic interactions using two different crosses between east European and Indian gene pool lines. *Theor. Appl. Genet.* 125, 1553–1564. doi: 10.1007/s00122-012-1934-1933

Yang, T., Echols, M., Martin, A., and Bar-Peled, M. (2010). Identification and characterization of a strict and a promiscuous N-acetylglucosamine-1-P uridylyltransferase in *Arabidopsis*. *Biochem. J.* 430, 275–284. doi: 10.1042/BJ20100315

Yao, K., Bacchetto, R. G., Lockhart, K. M., Friesen, L. J., Potts, D. A., Covello, P. S., et al. (2003). Expression of the Arabidopsis ADS1 gene in *Brassica juncea* results in a decreased level of total saturated fatty acids. *Plant Biotechnol. J.* 1, 221–229. doi: 10.1046/j.1467-7652.2003.00021.x

Yatusevich, R., Mugford, S. G., Matthewman, C., Gigolashvili, T., Frerigmann, H., Delaney, S., et al. (2010). Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in Arabidopsis thaliana. *Plant J.* 62, 1–11. doi: 10.1111/j.1365-313X.2009.04118.x

Zhang, D., and Hamauzu, Y. (2004). Phenolics, ascorbic acid, carotenoids and antioxidant activity of broccoli and their changes during conventional and microwave cooking. *Food Chem.* 88, 503–509. doi: 10.1016/j.foodchem.2004.01.065

Zhang, W., Hu, D., Raman, R., Guo, S., Wei, Z., Shen, X., et al. (2017). Investigation of the genetic diversity and quantitative trait loci accounting for important agronomic and seed quality traits in Brassica carinata. *Front. Plant Sci.* 8:307. doi: 10.3389/fpls.2017.00615

Zhao, Q., Huang, X., Lin, Z., and Han, B. (2010). SEG-Map : a novel software for genotype calling and genetic Map construction from next-generation sequencing. *Rice* 3, 98–102. doi: 10.1007/s12284-010-9051-x

Zhao, Y. T., Wang, M., Fu, S. X., Yang, W. C., Qi, C. K., and Wang, X. J. (2012). Small RNA profiling in two *Brassica napus* cultivars identifies MicroRNAs with oil production- and development-correlated expression and new small RNA classes. *Plant Physiol.* 158, 813–823. doi: 10.1104/pp.111.187666

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Akhatar, Singh, Sharma, Kaur, Kaur, Sharma, Bharti, Sardana and Banga. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Nutrition

ORIGINAL RESEARCH
published: 14 October 2020
doi: 10.3389/fnut.2020.00134

Enhancing β-Carotene Concentration in Parental Lines of CO6 Maize Hybrid Through Marker-Assisted Backcross Breeding (MABB)

Senthil Natesan[1], Thirusenduraselvi Duraisamy[2], Bharathi Pukalenthy[3], Sarankumar Chandran[3], Jagadeeshselvam Nallathambi[4], Karthikeyan Adhimoolam[5], Dhasarathan Manickam[5], Vellaikumar Sampathrajan[5], Samuel Jeberson Muniyandi[6], Laishram Joykumar Meitei[7], Nepolean Thirunavukkarasu[8], Ganesan Kalipatty Nalliappan[9] and Ravikesavan Rajasekaran[10]*

[1] Department of Plant Molecular Biology and Bioinformatics, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India, [2] Department of Seed Technology, Tamil Nadu Agricultural University, Coimbatore, India, [3] Department of Plant Breeding and Genetics, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai, India, [4] Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India, [5] Department of Biotechnology, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai, India, [6] AICRP-MULLARP, Directorate of Research, Central Agricultural University, Imphal, India, [7] Department of Genetics and Plant Breeding, College of Agriculture, Central Agricultural University, Imphal, India, [8] Department of Genomics and Molecular Breeding, ICAR-Indian Institute of Millets Research, Hyderabad, India, [9] Department of Plant Breeding and Genetics, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, India, [10] Department of Millets, Center for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, India

OPEN ACCESS

Edited by:
Manish Kumar Pandey,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

Reviewed by:
Vinod K. K,
Indian Agricultural Research Institute (ICAR), India
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

***Correspondence:**
Senthil Natesan
senthil_natesan@tnau.ac.in

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Nutrition

Received: 23 November 2019
Accepted: 13 July 2020
Published: 14 October 2020

Citation:
Natesan S, Duraisamy T, Pukalenthy B, Chandran S, Nallathambi J, Adhimoolam K, Manickam D, Sampathrajan V, Muniyandi SJ, Meitei LJ, Thirunavukkarasu N, Kalipatty Nalliappan G and Rajasekaran R (2020) Enhancing β-Carotene Concentration in Parental Lines of CO6 Maize Hybrid Through Marker-Assisted Backcross Breeding (MABB). Front. Nutr. 7:134. doi: 10.3389/fnut.2020.00134

Vitamin A deficiency (VAD) is a global health problem; many people around the world, especially children and pregnant women, are VAD deficient or insufficient. Maize is known as an important source of provitamin A for humans. Hence, enhancement of provitamin A carotenoids (pVAC) in maize varieties through breeding or biofortification is a good option for alleviating VAD in developing countries, especially India. So far, numerous maize hybrids have been developed in India. Among them, CO6, derived from UMI1200 × UMI1230, is a popular maize hybrid and adapted to different agro-climatic zones of India, especially Tamil Nadu, a southern state of India. However, CO6 is deficient for pVAC carotenoid β-carotene. Thus, the objectives of this study were to increase the β-carotene concentration in UMI1200 and UMI1230 and generate the β-carotene enriched hybrids through marker-assisted backcross breeding (MABB). For this purpose, the maize genotype HP467-15 was used as the donor for transferring the β-carotene gene, *crtRB1,* into UMI1200 and UMI1230. In the MABB scheme, we used one gene-specific marker (*crtRB1* 3′TE) and 214 simples sequence repeat (SSR) markers for foreground and background selection, respectively. As a result, six improved lines with recurrent parent genome recovery (RPGR) ranging from 90.24 to 92.42%, along with good agronomic performance, were generated. The β-carotene concentration of the improved lines ranged from 7.056 to 9.232 μg/g. Furthermore, five hybrid combinations were generated using improved lines and evaluated in a comparative yield trial (CYT) and multi-location trials (MLT) along with the original hybrid CO6 and commercial hybrids. It was revealed that ACM-M13-002 was a superior hybrid with a 7.3-fold increase in

β-carotene concentration and with a comparable yield to CO6. In summary, the improved maize inbreds can be used as possible donors for the development of β-carotene-rich cultivars in maize breeding programs and the β-carotene enriched hybrid developed in this study will hold great promise for food and nutritional security.

Keywords: β-Carotene, Co6 maize hybrid, crtRB1, vitamin A deficiency, MABB

INTRODUCTION

Vitamin A deficiency (VAD) is one of the most prevalent micronutrient deficiencies and poses a serious health problem to children and pregnant women around the world, especially in developing countries. The consequences of VAD include night blindness, reduced growth in children, and increased morbidity and mortality (1). Maize (*Zea mays* L.) is one of the major sources of provitamin A for humans. Thus, the improvement of provitamin A carotenoids (pVAC) in maize varieties through breeding or biofortification is a successful and cost-effective approach for alleviating VAD in developing countries, especially India (2). The preference of maize for biofortification also relies on higher bioavailability and bioconversion of pVAC into vitamin A. α-carotene, β-carotene, β-cryptoxanthin, zeaxanthin, and lutein are important carotenoid compounds that exist in maize grains (3). Among them, β-carotene has the highest provitamin A activity and can be converted easily to vitamin A by the human metabolism (4, 5). However, the success of this approach depends on the identification of superior maize genotypes with a high β-carotene concentration by characterizing the maize genotypes and subsequently transferring the high β-carotene from the superior genotypes to the popular maize cultivars through introgression breeding. However, β-carotene is present at a comparatively low level in maize kernels of popular cultivated lines, and only a very small percentage of maize varieties have naturally high β-carotene levels. Muthusamy et al. screened 105 maize inbreds of both Indian and CIMMYT origins for kernel β-carotene content. The results showed that kernel β-carotene concentration varied from 0.02 to 16.50 μg/g in maize inbreds (6). Hence, there is a need to understand the major genes involved in the carotenoid biosynthesis pathway that could be modulated for enhancing β-carotene concentration.

Recent advancements in maize genomics have reduced the breeding cycles and significantly increased the precision and effectiveness of breeding methods, and are also helping to identify the genes and biosynthesis pathways involved in β-carotene accumulation. Maize breeding programs through marker-assisted selection (MAS) have become more standard with the availability of reasonably dense molecular marker linkage maps, and many researchers have been pursuing the association of markers with known genes. Lycopene epsilon cyclase (*lcyE*) located at chromosome 8 and β-carotene hydroxylase enzyme (*crtRB1*) located at chromosome 10 are the two major genes associated with β-carotene accumulation in maize (3, 7). Among them, the gene *crtRB1* was stated to have a much greater effect on the concentrations of β-carotene than that of *LcyE* (8). The favorable allele, namely *crtRB1* 3'TE (allele 1, 543 bp), is accountable for causing a 2- to 10-fold upsurge in β-carotene concentration (8). Polymerase chain reaction (PCR) based co-dominant markers from the 3'TE region of the *crtRB1* gene provided opportunities to identify and develop the higher β-carotene concentration genotypes through MAS. Several studies showed the feasibility of this allele-based marker to detect the *crtRB1* allele in diverse maize genotypes (9–11).

In India, maize is one of the major cereal crops after rice and wheat, and is utilized as a major source of both food and feed. To date, several maize inbreds and hybrids have been developed for their economically important traits and are commercially available in India. Among them, CO6, derived from UMI1200 × UMI1230, is a popular single-cross hybrid and adapted to different agro-climatic zones of India, especially Tamil Nadu, a southern state of India. CO6 is well-suited for cultivation in both under irrigate and rainfed conditions and also exhibits multiple disease resistances. It shows superior performance in grain yield and quality, and ease in hybrid seed production (12). These salient features make CO6 a popular maize hybrid among farmers. However, CO6 is deficient in nutritional traits, especially β-carotene. Thus, it is necessary to improve the CO6 hybrid with a higher β-carotene concentration. Marker-assisted backcross breeding (MABB) is an effective approach to improve the nutritional traits in maize. Recently, several studies showed success with improving the nutritional traits in maize hybrids (Pusa Vivek QPM9 Improved, Pusa HM4 Improved, Pusa HM8 Improved, and Pusa HM9 Improved) using MABB (9, 13). With this background information, the present study aimed to: (i) introgress the *crtRB1* allele into the parental lines (UMI1200 and UMI1230) of CO6 hybrid using MABB; (ii) evaluate the improved lines agronomic performance and β-carotene concentration; and (iii) generate the hybrids using the improved lines and evaluate the hybrids in diverse maize growing regions of Tamil Nadu, a southern state of India.

MATERIALS AND METHODS

Plant Genetic Materials

CO6 is a popular maize hybrid well-adapted to different agro-climatic zones of India, especially Tamil Nadu, a southern state of India. Maize inbreds viz., UMI1200 and UMI1230 are the parental lines of the CO6 hybrid. UMI1200 is a female parent for the CO6 hybrid and is suitable for tropical regions; it is characterized as a dent grain type. UMI1230 is a pollen parent for the CO6 hybrid and is suitable for both tropical and semi-arid regions. It is characterized as a flint type. Both have diverse genetic backgrounds and are placed in different subspecies in

accordance with the starch type in the endosperm. Due to their excellent combining ability, these two inbreds were used to develop the CO6 hybrid. The donor parent was HP467-15 from CIMMYT (International Maize and Wheat Improvement Center, Mexico), which carries the *crtRB1* allele responsible for the high β-carotene concentration.

Outline of the Conversion of Maize Inbreds to Its High β-Carotene Version

The conversion process included crossing, two generations of backcrossing, and two generations of selfing. All the field experiments were carried out on the Central farm, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India from 2011 to 2015. Foreground selection was done using a *crtRB1* gene-specific marker (*crtRB1* 3′TE) (7) to select the plants at F_1, backcross, and selfed generations. Foreground selection targeted heterozygous plants with allele 1, 543 bp (Favorable allele) and allele 3, 296 bp (Unfavorable allele), or homozygous plants with allele 1, 543 bp (Favorable allele). Background selection was done to select the foreground positive plants with maximum recurrent parent genome recovery (RPGR) using polymorphic simple sequence repeat (SSR) markers in backcross and selfed generations. Two independent crosses, namely UMI1200 × HP467-15 and UMI1230 × HP467-15, were made during Rabi 2011-12 to produce F_1 seeds. In Kharif 2012, the F_1 plants were confirmed for their heterozygosity and backcrossed with the recurrent parent to produce BC_1F_1 progenies. In Rabi, 2012-13, the BC_1F_1 plants that were found to be heterozygous were backcrossed with the recurrent parent to produce BC_2F_1 progenies. Further, BC_2F_1 plants were screened to identify heterozygous and the selected plants were selfed to produce BC_2F_2 progenies during Kharif 2013. In Rabi 2013-14, BC_2F_2 progenies were screened before flowering to identify plants with allele 1, 543 bp (Favorable allele) and the selected plants were selfed to produce BC_2F_3 lines, whereas the homozygous plants with allele 3, 296 bp (Unfavorable allele) and heterozygous plants with a'llele 1, 543 bp (Favorable allele) and allele 3, 296 bp (Unfavorable allele) were excluded. Agro-morphological character evaluation and β-carotene estimation were carried out in BC_2F_3 improved lines.

Genomic DNA Isolation and Marker Genotyping

The genomic DNA was isolated from the young leaves using the cetyl trimethyl ammonium bromide (CTAB) method (14). The DNA was checked for its quantity and quality on 0.8% agarose gel. The *crtRB1* gene-specific marker (*crtRB1* 3′TE) (65F: ACACCACATGGACAAGTTCG, 62R: ACACTCTGGCCCATG AACAC, 66R: ACAGCAATACAGGGGACCAG) was used for foreground selection. A total of 214 simple sequence repeat (SSR) markers spanning uniformly across the maize genome were used for background selection. The primer sequences for SSR markers were obtained from the maize genome database (www.maizegdb.org) and were synthesized by Eurofins Ltd, Bangalore, India. The polymerase chain reaction (PCR) for *crtRB1* 3'TE/SSR markers and gel electrophoresis were carried out following the procedure of Muthusamy et al. and Pukalenthy et al. (9, 15). All the obtained genotyping results were tested for goodness of fit using the chi-square analysis. The amplicons of the markers were scored as "A" for the recurrent parent, "B" for the donor parent, and "H" for heterozygotes. Chi-square analysis was done using PROC MIXED of SAS and LSMEANS and ADJUST options were used to find the least square mean values and for the comparison of the *p*-value. The percentage of RPGR was calculated following the procedure of Sundaram et al. (16).

Evaluation of Agro-Morphological Characters in Improved Lines

The recurrent and donor parents and the improved lines were evaluated for eight agro-morphological characters [plant height (cm), days to 50% tasselling (days), days to 50% silking (days), cob length (cm), number of kernel rows/ cob, number of kernels/row, grain yield/ plant (g), and 100-kernel weight (g)] which were recorded according to the plant stage based on the standard maize descriptors formulated by IBPGR (17).

Estimation of β-Carotene Concentration Using HPLC Analysis

The seeds obtained from the recurrent and donor parents and the improved lines were subjected to β-carotene estimation. Since the carotenoid compounds are sensitive to light, the seeds were harvested at 14% moisture content and stored at 22–26°C. Estimation of β-carotene was carried out by adopting the Harvest plus protocol (18). The carotenoid compounds were extracted by grinding the seed samples using ice-cold acetone until a fine powder was obtained. Since the carotenoid compounds undergo photo-oxidation, further processing was done under yellow light (19, 20). The samples extracted were concentrated in a rotary evaporator at 45°C and made up to 2 ml using methanol prior to separation. The β-carotene concentration was quantified by high-performance liquid chromatography (HPLC). Samples were eluted by C18G 120A column (250 × 4.6 mm) and measured with a photodiode array detector set at 450 nm. The mobile phase comprised of Acetonitrile: Methanol: Ethyl acetate (80:10:10) at high pressure by the column with the flow rate of 1 mL min^{-1}. β-carotene (M/s. Sigma Aldrich, India) reconstituted in methanol to five different concentrations (0.1 μg/g; 1 μg/g; 10 μg/g; 50 μg/g; 100 μg/g) was used to make the standard curve. The β-carotene concentration was identified by their characteristic spectra and comparison of their retention times with known standard solutions.

Generation of Hybrids Using Improved Lines and Their Evaluation

Four improved lines from UMI1200 × HP467-15 and two improved lines from UMI1230 × HP467-15 were crossed during Kharif 2014 to generate the five hybrids. These hybrids were evaluated along with the commercial hybrid NK6240 and CO6 during the Rabi 2014-15 and Kharif 2015 at the Experimental Farm, Agricultural Research Station, Vaigai dam, Tamil Nadu Agricultural University, India and Maize Research Station, Tamil Nadu Agricultural University, Vagarai, India under comparative

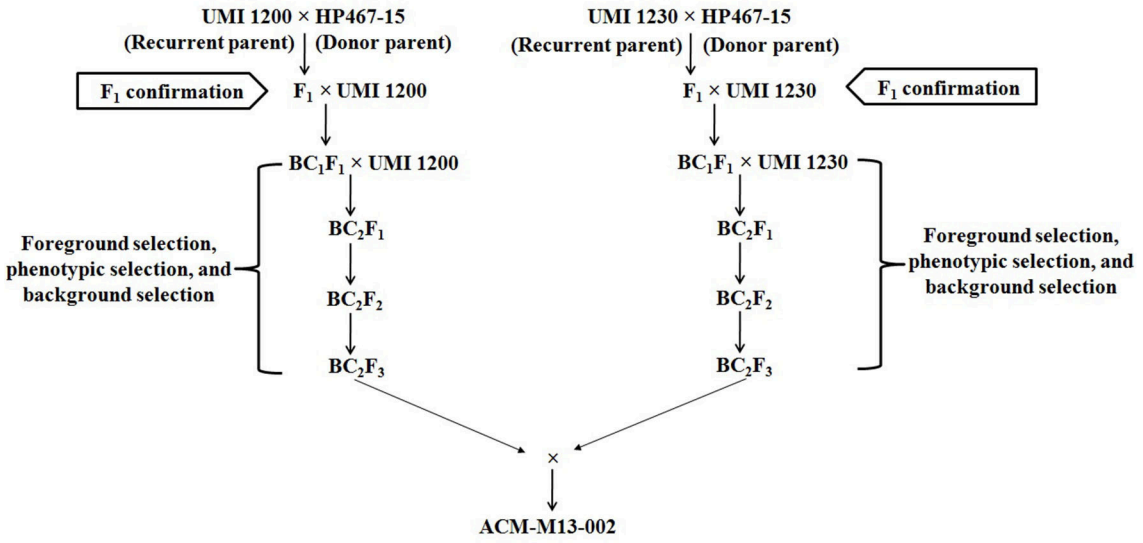

FIGURE 1 | Scheme of marker assisted backcross breeding (MABB) for improving β-carotene concentration in parental lines of CO6 maize hybrid.

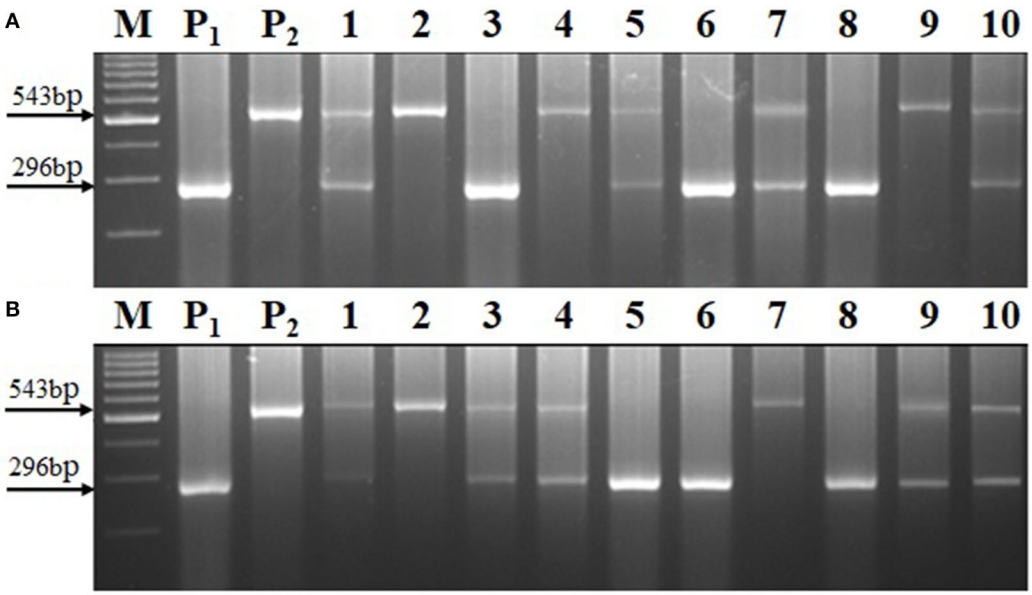

FIGURE 2 | Segregation of allele1 and allele 3 in BC_2F_2 generation using the *crtRB1* gene specific marker (i.e., *crtRB1* 3′TE). **(A)** UMI1200 × HP467-15, **(B)** UMI1230 × HP467-15, **(M)** Ladder (100 bp), **(P_1)** UMI1200/UMI1230, **(P_2)** HP467-15, and **(1–10)** BC_2F_2 progenies.

yield trial (CYT). Further, one superior hybrid identified from CYT was tested in multi-location trials (MLT) during Kharif 2016 and 2017 in diverse maize growing environments of Tamil Nadu along with eight maize hybrids including commercial hybrids. Kharif 2016 included four environments viz., Coimbatore (E1), Vagarai (E2), Vridhachalam (E3), and Bhavanisagar (E4) and Kharif 2017 included six environments viz., Coimbatore (E5), Vagarai (E6), Vridhachalam (E7), Bhavanisagar (E8), Vaigai Dam (E9), and Athiyandal (E10). Hybrids [namely NK6240 (G1), CO 6 (G2), ACM-M13-002 (G3), ACM-M13-007 (G4), CMH 11-583 (G5), VaMH 14020 (G6), VaMH 12014 (G7), 900 M GOLD (G8), and CMH 11-586 (G9)] were used in MLT. All the field trials were conducted in the irrigated condition. The spacing followed was 60 × 25 cm with a plot size of 5 × 3.6 m (6 rows each) and a fertilizer dosage of 250:75:75 kg/ha. Trials were carried out using a randomized complete block design (RCBD) with two replications. Grain yield data were subjected to a combined analysis of variance (ANOVA) using GEA-R statistical software (21). Additive Main Effect and Multiplicative Interaction (AMMI) model (22) and Genotype

TABLE 1 | Segregation pattern of alleles of the *crtRB1* gene in backcrossed and selfed progenies.

	Cross	Generation	Population size	No of homozygotes (allele1/allele1)	No of heterozygotes (allele1/allele3)	No of homozygotes (allele3/allele3)	χ^2	*P*-value
1	UMI1200 × HP467-15	BC_1F_1	100	–	24	76	27.04	0.0001*
		BC_2F_1	53	–	11	42	18.13	0.0001*
		BC_2F_2	146	34	46	66	33.99	0.0004*
2	UMI1230 × HP467-15	BC_1F_1	100	–	37	63	6.76	0.0093*
		BC_2F_1	87	–	18	69	29.89	0.0005*
		BC_2F_2	126	31	67	28	14.25	0.0008*

*Significant at $P < 0.05$.

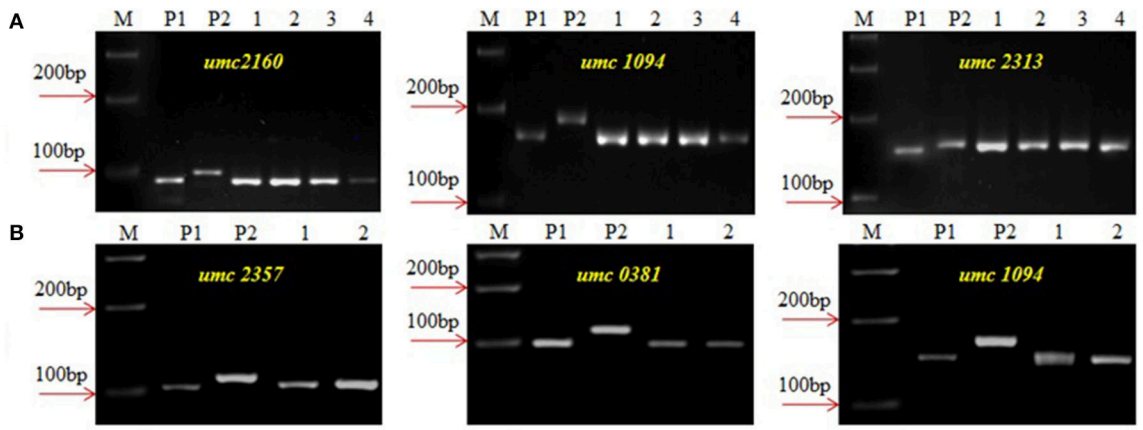

FIGURE 3 | Background screening of improved lines using SSR markers. **(A)** UMI1200 × HP467-15, **(B)** UMI1230 × HP467-15, **(M)** Marker 100 bp, **(P_1)** UMII1200/UMI1230, **(P_2)** HP467-15, **(1–4)** Improved lines from UMI11200 × HP467-15 (1-UMI1200β$^+$-1, 2-UMI1200β$^+$-2, 3-UMI1200β$^+$-3, 4-UMI1200β$^+$-4, and **(1 and 2)** Improved lines from UMI1230 × HP467-15, (1-UMI1230β$^+$-1, 2-UMI1230β$^+$-2).

and Genotype by Environmental Interaction (GGE) effects biplot (23) were employed to analyse the Genotype by Environment (GE) interaction and to assess grain yield stability based on the principal component analysis (PCA).

RESULTS

Marker-Assisted Introgression of the *crtRB1* Gene Into UMI1200 and UMI1230

The *crtRB1* allele was introgressed into UMI1200 and UMI1230 using a recurrent backcrossing procedure, combined with foreground and background selection (**Figure 1**). Parental polymorphism screening was conducted between the recurrent and donor parents using a set of 214 SSR markers. In the cross combinations UMI1200 × HP467-15 and UMI1230 × HP467-15, among the 214 SSR markers, 117 and 109 SSR markers were found to be polymorphic markers. These polymorphic SSR markers were used for background selection and RPGR analysis. The F_1 plants were produced from crosses UMI1200 × HP467-15 and UMI1230 × HP467-15. After confirming heterozygosity with a *crtRB1* 3'TE marker, the true F_1 plants were backcrossed to their respective recurrent parent. The resulting BC_1F_1 population were screened with a *crtRB1* 3'TE marker and 24 heterozygous plants (allele 1/allele 3) in UMI1200 × HP467-15 and 37 heterozygous plants (allele 1/allele 3) in UMI1230 × HP467-15 were identified. The percentage of RPGR of positive plants ranged from 66.52–78.75% to 66.41–78.43%. The best BC_1F_1 plant showed a maximum RPGR of 78.75 and 78.43% was further selected to produce the BC_2F_1 population. In the BC_2F_1 population, a total of 53 and 87 plants were screened with a *crtRB1* 3'TE marker. Eleven heterozygous (allele 1/allele 3) plants in UMI1200 × HP467-15 and 18 heterozygous plants (allele 1/allele 3) in UMI1230 × HP467-15 were identified in the BC_2F_1 population. The percentage of RPGR of positive plants ranged from 81.39–88.12% to 82.59–89.65%. The best three BC_2F_1 plants from each cross having maximum RPGR were further selected and selfed to produce the BC_2F_2 population. A total of 146 and 126 BC_2F_2 plants were screened with a *crtRB1* 3'TE marker to identify homozygous plants (allele 1) (**Figure 2**). It revealed 34 homozygous plants in UMI1200 × HP467-15 and 31 homozygous plants in UMI1230×HP467-15. The percentage of RPGR of positive plants ranged from 87.26–92.42% to 88.36–92.21%. The population details and their segregation pattern are presented in **Table 1.** The numbers of selected plants were further reduced to four and two based on the RPGR and phenotype data, and selfed to produce BC_2F_3 lines. As a result, four BC_2F_3

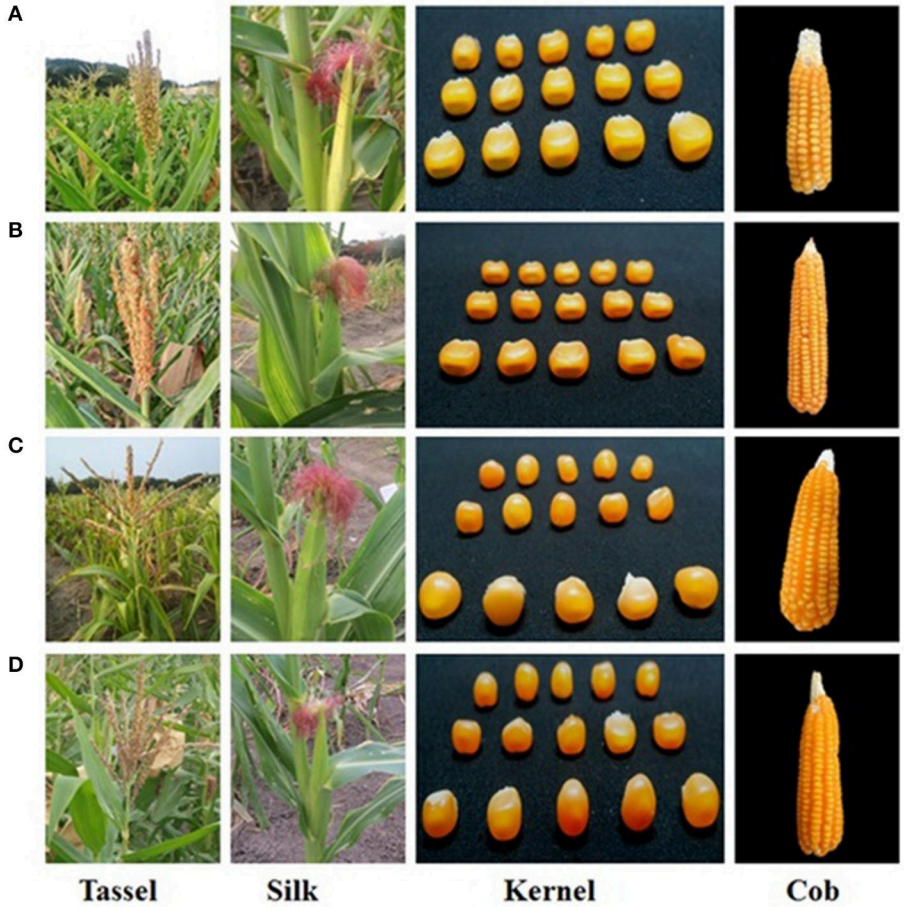

FIGURE 4 | Morphological resemblance of the parents and the improved lines. **(A)** UMI1200, **(B)** UMI1200β$^+$-1, **(C)** UMI1230, **(D)** UMI1230β$^+$-1.

improved lines (UMI1200β$^+$-1, UMI1200β$^+$-2, UMI1200β$^+$-3, and UMI1200β$^+$-4) from the cross UMI1200 × HP467-15 and two BC_2F_3 improved lines (UMI1230β$^+$-1 and UMI1230β$^+$-2) from the cross UMI1230 × HP467-15 were developed. UMI1200 × HP467-15 based four BC_2F_3 lines showed 90.24–92.42% RPGR with an average of 91.33% and UMI1230 × HP467-15 based two BC_2F_3 lines showed 90.41% to 92.21% RPGR with an average of 91.31%**.** Among the improved lines, UMI1200β$^+$-1 and UMI1230β$^+$-1 showed the highest RPGR of 92.42% and 92.21% from their respective crosses (**Figure 3**). The postive plant's RPGR details are presented in **Supplementary Tables 1**, **2**.

Introduction of *crtRB1* Gene Had No Effects on Agro-Morphological Characters

Six improved lines from both crosses were evaluated for the agro-morphological characters (**Figure 4** and **Table 2**). It was revealed that all the improved lines recorded more than 80% of phenotypic resemblance, except for the number of kernel rows per cob (71.43%), to their respective recurrent parent. Among them, UMI1200β$^+$-1, UMI1200β$^+$-2, and UMI1200β$^+$-4 from UMI1200×HP467-15 and UMI1230 β$^+$-1 from UMI1230× HP467-15 recorded high phenotypic resemblance percentage (>90%) for more than six and five traits *viz.,* days to 50% tasselling, days to 50% silking, plant height, cob length, number of kernels per row, 100 kernel weight, and grain yield per plant. The overall performance showed that the traits recorded among the improved lines were on par with their recurrent parents and its phenotypic resemblance percentage ranging from 71.43% (number of kernel rows per cob) to 99.89% (grain yield per plant).

β-Carotene Concentration in Improved Lines

The β-carotene concentration was estimated in six improved lines; it showed an increased level of β-carotene ranging from 7.056 to 9.232 μg/g with an average of 8.322 μg/g. The β-carotene concentration for the recurrent parents was 0.893 μg/g (UMI1200) and 1.610 μg/g (UMI1230). Introgression of allele1 resulted in an increased level of the kernel β-carotene concentration with a maximum of 9.073 μg/g in UMI1200β$^+$-1 from UMI1200 × HP467-15 and 9.232μg/g in UMI1230β$^+$-1 from UMI1230 × HP467-15. We also recorded a minimum level of the kernel β-carotene concentration 7.056 μg/g in UMI1200β$^+$-3 and 9.021 μg/g in UMI1230β$^+$-2. However, it was higher than their respective recurrent parent.

TABLE 2 | β-carotene concentration and agronomic performance of the improved lines developed through MABB.

	Genotypes	β-carotene (μg/g)	Plant height (cm)	Days to 50% tasseling (days)	Days to 50% silking (days)	Cob length (cm)	Number of kernel rows/cob	No of kernels/ row	100 kernel weight (g)	Mean grain yield/plant (g)
1	[a]UMI1200β$^+$-1	9.073 ± 0.557	126.50	58.00	61.00	13.10	12.00	10.00	24.10	87.12
2	[a]UMI1200β$^+$-2	8.303 ± 0.364	126.10	59.00	59.00	13.40	10.00	10.00	24.10	87.32
3	[a]UMI1200β$^+$-3	7.056 ± 0.562	126.80	57.00	61.00	13.20	12.00	09.00	24.20	87.61
4	[a]UMI1200β$^+$-4	7.251 ± 0.453	126.30	58.00	60.00	13.50	12.00	10.00	24.60	88.12
5	[a]UMI1230β$^+$-1	9.232 ± 0.508	133.90	57.00	59.00	12.10	10.00	09.00	23.89	76.41
6	[a]UMI1230β$^+$-2	9.021 ± 0.471	134.10	56.00	58.00	12.30	10.00	08.00	23.78	76.39
7	[b]UMI1200 (RP)	0.893 ± 0.070	127.00	60.00	62.00	13.70	14.00	11.00	25.00	88.21
8	[b]UMI1230 (RP)	1.610 ± 0.083	135.00	58.00	60.00	12.50	12.00	10.00	24.10	76.71
9	[b]HP467-15 (DP)	10.525 ± 0.263	110.00	56.00	59.00	13.00	10.00	10.00	23.00	54.80
10	Test inbred mean	8.322 ± 0.485	128.95	57.50	59.67	12.93	11.00	9.33	24.11	83.83
11	Check inbred mean	4.342 ± 0.138	124.00	58.00	60.33	13.07	12.00	10.33	24.03	73.24

[a]*Test inbreds and* [b]*Check inbreds.*

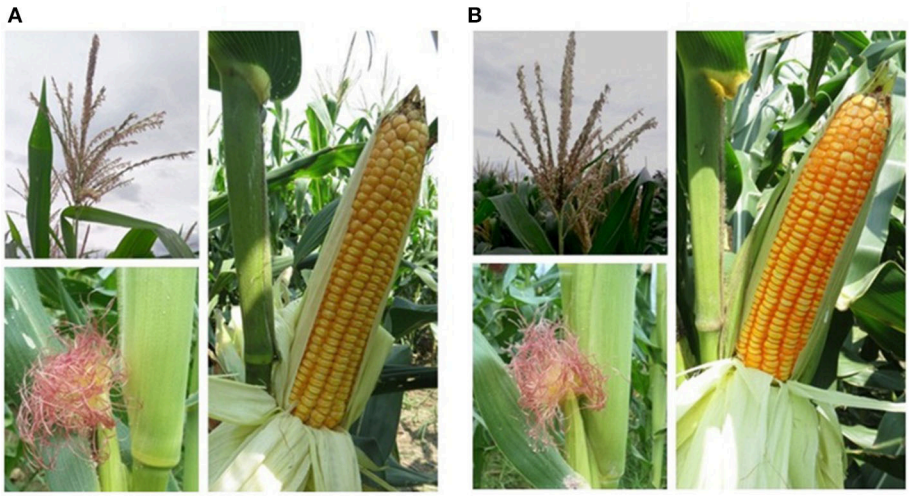

FIGURE 5 | Morphological characteristics of original version and reconstituted hybrid with high β-carotene. **(A)** CO6, **(B)** ACM-M13-002.

The β-carotene concentration of improved lines is presented in **Table 2**. In addition, the β-carotene concentration of the improved lines at different environments are summarized in **Supplementary Table 3**.

Evaluation of Hybrid Combination Under Comparative Yield Trial (CYT) and Multi-Location Trial (MLT)

Five hybrid combinations were developed using the improved lines. The agronomic performance and the β-carotene concentration of the five hybrids were evaluated along with their corresponding original hybrid (CO6) and a commercial hybrid (NK6240) under CYT (**Figure 5**). Among the five hybrids, ACM-M13-002 showed a 13.38% yield increase and a 7.3-fold β-carotene increase over its original hybrid and 15.59% yield increase over a commercial hybrid NK6240 (**Table 3**). Thus, ACM-M13-002 forwarded to MLT along with eight hybrids to study the GE interaction over the different maize growing regions in Tamil Nadu. The yield performance of MLT showed that the hybrid ACM-M13-002 (G3) recorded a value that is comparable to that of its original hybrid (CO6) in all the locations. An average yield performance across 10 environments recorded 7,256.2 kg/ha, which was comparable with the original hybrid CO6 (G2) (7,358.1 kg/ha).

Per se performance of the yield for the developed five hybrids ranged from 10.73 to 11.86 (t/ha) with an average of 11.02 (t/ha). Analysis of variation for the grain yield (kg/ha) of maize hybrids showed a significant difference between the components under study (Genotypes, Environments and Genotype × Environment) with the percent accounting for the highest total variation sum of the square by 80% for environments, 1.33%, and 18.68% for GE (**Tables 4**, **5** and **Supplementary Table 4**). The effect of the genotype and environment is 14.04 times higher compared

TABLE 3 | Agronomic performance and β-carotene value of hybrids developed through MABB under comparative yield trail (CYT).

Hybrid	Plant height (cm)			Days to 50% tasseling			Days to 50% silking			1,000 grain weight (g)			Grain yield/plant (g)					Yield (t/ha)	β-carotene (μg/g)
	VGD[1] Rabi 2014–15	VGD[1] Kharif 2015	VGI[2] Rabi 2014–15	VGD[1] Rabi 2014–15	VGD[1] Kharif 2015	VGI[2] Rabi 2014–15	VGD[1] Rabi 2014–15	VGD[1] Kharif 2016	VGI[2] Rabi 2014–15	VGD[1] Rabi 2014–15	VGD[1] Kharif 2015	VGI[2] Rabi 2014–15	VGD[1] Rabi 2014–15	VGD[1] Kharif 2015	VGI[2] Rabi 2014–15	% increase over CO6	% increase over NK6240		
[a]ACM-M13-001[3]	185.00	184.00	186.00	46.00	47.00	46.00	48.00	49.00	49.00	345.00	343.00	344.00	163.00	164.00	164.00	4.49	6.53	10.93	8.92
[a]ACM-M13-002[3]	187.00	186.00	188.00	45.00	46.00	45.00	48.00	48.00	47.00	330.00	329.00	332.00	179.00	176.00	178.00	13.38	15.59	11.86	9.05
[a]ACM-M13-003[3]	238.00	237.00	239.00	50.00	49.00	49.00	53.00	52.00	52.00	400.00	401.00	400.00	160.00	162.00	162.00	2.58	4.58	10.73	8.24
[a]ACM-M13-004[3]	206.00	205.00	206.00	48.00	49.00	48.00	51.00	52.00	51.00	420.00	420.00	418.00	164.00	162.00	163.00	3.82	5.84	10.86	8.47
[a]ACM-M13-005[3]	208.00	207.00	208.00	49.00	49.00	48.00	52.00	52.00	50.00	340.00	339.00	341.00	161.00	162.00	160.00	2.58	4.58	10.73	8.83
SE[4]	9.55	9.55	9.53	0.93	0.63	0.73	1.03	0.87	0.86	18.00	18.39	17.49	3.47	2.73	3.22	2.04	2.08	0.21	0.15
SD[5]	21.35	21.35	21.30	2.07	1.41	1.64	2.30	1.95	1.92	40.25	41.13	39.12	7.77	6.10	7.20	4.55	4.64	0.48	0.33
[b]CO6[6]	189.00	188.00	190.00	53.00	53.00	52.00	56.00	56.00	55.00	400.00	401.00	400.00	157.00	159.00	157.00			10.46	1.24
[b]NK6240[6]	200.00	198.00	200.00	49.00	50.00	48.00	52.00	53.00	51.00	330.00	331.00	330.00	168.00	169.00	167.00			10.26	0.71
Test hybrid mean	204.80	203.80	205.40	47.60	48.00	47.20	50.40	50.60	49.80	367.00	366.40	367.00	165.40	165.20	165.40	5.37	7.42	11.02	8.7
Check hybrid mean	194.50	193.00	195.00	51.00	51.50	50.00	54.00	54.50	53.00	365.00	366.00	365.00	162.50	164.00	162.00			10.36	0.97

[1]*Vaigai dam,* [2]*Vagarai,* [3]*High beta carotene hybrids,* [4]*Standard error,* [5]*Standard deviation,* [6]*Standard checks.* [a]*Test hybrids,* [b]*Check hybrids. Note: Yield (t/ha) given is the average of the hybrids.*

to the component genotype. Hence, the variation that exists among the genotypes across the environments paves the way to understanding the yield stability of the hybrids over the varying environments and specific adaption of the hybrids. Based on the Interaction Principal Component Axis 1 (IPCA1) of the AMMI model, the genotypes G1, G2, and G5 recorded values nearer to the zero on the biplot (0.056, −0.154, and 0.002) which represents that the genotypes are highly stable with good yield potential (**Figure 6**). From the "Which won where" polygon view of GGE analysis (**Figure 7**) four genotypes - G3, G4, G5, and G8 - were located at the corner, which denotes that the genotypes are specifically adapted to the particular environment for the trait under study. Average Environment Coordination (AEC) recorded that G3 is placed at the second concentric circle and is ideal in terms of high yielding ability and stability, whereas other genotypes, G2, G6, and G7, placed at the third and fourth concentric circle where the stability gradient is low compared to the genotypes closer to the center axis. Based on the vertices of the Average Environment Axix (AEA), the environments E9 and E4 were considered as highly interactive environments from where the wider adaption of the genotypes are selected.

DISCUSSION

CO6 is one of the popular maize hybrids of India, especially in Tamil Nadu. CO6 is known for its high yield, multiple disease resistances, and its suitablity for rainfed and irrigated conditions. However, CO6 is low in pVAC β-carotene, thus, the present investigation was taken up to improve the β-carotene in CO6 through MABB. Previously, Muthusamy et al. (9) and Zunjare et al. (10) used an accelerated MABB scheme to generate the improved version of VQL1-K10-40-11-53 × VQL2-K10-08-14, HQPM7-B, HQPM 5-C hybrids for β-carotene. We also used the same MABB scheme to enrich the β-carotene in CO6. Marker-assisted foreground selection using gene-linked markers permits the transfer of the gene of interest with high precision in MABB. The *crtRB1* gene-specific marker (*crtRB1* 3′TE) used in this study allows us to detect the positive plants exactly without any false positives at any stage of MABB. The advantages of using a linked marker for the selection of single or multiple genes simultaneously in MABB have been described in many studies for nutritional traits in maize (7, 8, 13). The backcross and selfed-generations from both crosses, UMI1200 × HP467-15 and UMI1230 × HP467-15, showed segregation distortion. The segregation pattern of allele 1 and allele 3 deviated from the expected Mendelian ratio (1:1 and 1:2:1) and allele 1 was diminished. These results are similar to the reports of Babu et al. (8), who described the segregation distortion of allele 1 in five out of eight populations. There are many facts, such as embryo-specific mutation (24), segregation distortion regions in the maize genome (25), mutants like the defective kernel (24), gametophytic factors (24, 26), and genetic background of the target allele (8), that are reported to be the reason for the segregation distortion. In this case, the evaluation of a large population for attaining enough foreground positives is essential. Marker-assisted background selection using SSR markers helped

TABLE 4 | Analysis of variance of yield data of nine maize hybrids tested across environments during Kharif 2016 and 2017.

S.No	Source	DF	SS	MS	Total variation explained (%)
1	Environment (E)	9	391935207	43548356**	80.00
2	Genotype (G)	8	6505899	813237**	1.33
3	Genotype × Environment Interaction (GE)	72	91495843	1270776**	18.68
4	PC1	16	54822906	3426432**	59.76
5	PC2	14	15431914	1102280**	16.82
6	PC3	12	8193433	682786**	8.93
7	Residuals	90	2241107	24901**	0.00

***Represents highly significant at 1% level.*

TABLE 5 | Multi location trial (MLT) conducted in Kharif 2016 and 2017.

S. No	Hybrid	Parents	Code for hybrids	Grain yield (Kg/ha) kharif 2016				Grain yield (Kg/ha) kharif 2017						
				CBE[E1]	VGI[E2]	VRI[E3]	BSR[E4]	CBE[E5]	VGI[E6]	VRI[E7]	BSR[E8]	VGD[E9]	CEM[E10]	Mean
1	NK6240	–	G1	8,705.0	8,598.5*	5,508.0	6,564.5	10,447.5	5,023.5	8,220.5	6,141.5	5,671.5	8,347.0	7,322.8
2	CO6	UMI1200×UMI1230	G2	8,796.5	9,466.5	5,912.0	6,008.5	10,125.0	6,12.0	8,381.5	5,554.0	5,794.0*	7,531.0	7,358.1
3	ACM-M13-002	UMI1200β$^+$ × UMI1230β$^+$	G3	8,794.5	8,824.0	6,875.5*	7,138.5*	8,309.0	6,452.5	7,615.5	6,393.5*	6,393.5*	5,765.0	7,256.2
4	ACM-M13-007	UMI285lpa × UMI1200-1	G4	8,477.5	8,474.5	6,681.5*	6,493.5	10,525.0	5,715.0	9,499.0*	5,225.5	6,005.5*	4,693.5	7,179.1
5	CMH 11-583	N09-153-3A × N10-65-3	G5	8,984.5*	8,264.0	5,918.0	5,563.0	10,349.5	5,478.5	8,892.0*	5,745.0	4,730.0	7,622.0	7,154.7
6	VaMH 14020	UMI1230 × VIM153	G6	8,147.0	8,850.0	6,882.5*	6,557.0	10,258.0	6,002.0*	9,008.0*	5,702.0	5,837.5*	9,664.0*	7,690.8
7	VaMH12014	UMI1200×VIM357	G7	7,507.0	8,574.5	6,324.0	6,695.0*	9,766.5	6,070.5*	8,583.5	6,041.5	5,305.0	9,876.0*	7,474.4
8	900 M GOLD	-	G8	8,989.0*	8,127.5	5,110.0	6,255.0	9,249.5	6,118.0*	8,913.5*	6,393.5*	5,094.0	9,263.5*	7,351.4
9	CMH 11-586	N09-164-2×N148	G9	9,435.5	8,319.5	7,333.5*	6,416.5	10,618.0*	5,550.5	8,235.5	5,978.5	5,237.5	9,924.0*	7,704.9
10	[a]SEd			140.37	115.25	87.36	116.95	257.63	65.31	110.11	133.76	82.26	138.49	
11	[b]CD ($p = 0.05$)			323.71	265.76	201.46	269.68	594.11	150.60	253.92	308.45	189.70	319.37	
12	Mean			8,648.5	8,611.0	6,282.8	6,410.8	9,960.9	5,824.7	8,594.3	5,908.3	5,563.8	8,076.2	

(E1) *Coimbatore,* ***(E2)*** *Vagarai,* ***(E3)*** *Vridhachalam,* ***(E4)*** *Bhavanisagar,* ***(E5)*** *Coimbatore,* ***(E6)*** *Vagarai,* ***(E7)*** *Vridhachalam,* ***(E8)*** *Bhavanisagar,* ***(E9)*** *Vaigai dam, and* ***(E10)*** *Athiyandal,* [a]*Standard error deviation,* [b]*Critical difference. *Represents significant at 5% level.*

in selecting the foreground positive plants with a high RPGR. We performed the background selection starting from the BC_1F_1 generation; it hastened the RPGR in advanced lines. Following this approach, after only two generations of backcrossing, it was possible to identify the plants carrying RPGR ranging from 90.24 to 92.42% (UMI1200 × HP467-15) and 90.41–92.21% (UMI 1230 × HP467-15) in BC_2F_3 improved lines. These results are in accordance with the previous studies (9, 27, 28).

The study aims to introgress the *crtRB1* gene without troubling the recurrent parental genome by MABB, thus the improved line will be suitable for further use. For confirming the suitability of the new genotypes, it is essential to characterize and choose the progenies that are closer to the parent for both morphological and nutritional traits as this selection adds value to the MABB program. Thus, we investigated the agro-morphological characters of the improved lines. It revealed that the agro-morphological characters of the improved lines were on par with the recipient parent. The agro morphological characters among the six improved lines ranged from 71.43 to 99.89%. Among them, UMI1200β$^+$-1, UMI1200β$^+$-2, and UMI1200β$^+$-4 from UMI1200 × HP467-15, and UMI1230β$^+$-1 from UMI1230 × HP467-15 recorded more than 90% recovery for the traits *viz.,* days to tasseling, days to silking, plant height, cob length, number of kernels per row, 100 kernel weight, and grain yield per plant. Moreover, the β-carotene content of the improved lines ranged from 7.056 to 9.073 μg/g (UMI1200 × HP467-15) and 9.021–9.232 μg/g (UMI1230 × HP467-15). The β-carotene concentration of the improved lines is on par with the β-carotene parent. Similar results were obtained in other studies (10, 29).

The improved lines can be considered as the candidate parents for developing β-carotene hybrids adapted to irrigated and rainfed conditions. They were successfully used to develop the five hybrid combinations. Among them, ACM-M13-002 is a superior hybrid with increased yield and β-carotene concentration over its original hybrid and a commercial hybrid NK6240 under CYT. Hence, this hybrid was forwarded for MLT along with eight hybrids to study the GE interaction over the different maize growing regions in Tamil Nadu. The yield performance of MLT showed that ACM-M13-002 (G3) recorded a value that is comparable to that of its original hybrid CO6 (G2) in all the locations. Based on the stability ANOVA, the GE toward the yield was lesser compared to the environment

(E) and it is 14.04 times higher than that of genotype (G). These results are similar to the earlier study on maize (20). We observed significant genotype by environment interaction (GE) effect, which revealed that the yield performance of all the hybrids over the environments under study was not consistent. This situation necessitates understanding the nature and magnitude of genotype by environment effect which may not be possible through standard ANOVA (21, 22). IPCA1 from the AMMI model gives an idea about the association of the genotypes with the environments under study. A higher IPCA1 value reflects the specific adaptation of the genotypes irrespective of the direction in which it falls (Positive/ Negative) and the values nearer to the zero reflect the genotypes widely adapted to all the environments tested. From the above concept the genotype G1, G2, and G5 recorded values closer to the zero, which could be promising genotypes for the yield trait over all the environments. GGE biplot analysis reflects the similarity existing between the genotypes, environment, and their corresponding response to each other, as well as the interaction between them, making it the best tool to graphically represent the stability of the genotype (23). From the polygon view of GGE biplot, the genotypes G3, G4, G5, and G8 are recorded as a specific adaptation to the respective environments which are present nearer to the vertex of the polygon. High *per se* performance, higher vector length, and high stability across the environments are the properties of ideal genotype and the genotypes places closed to the ideal genotype are called desirable genotypes. These are most probably at the second concentric circle of the GGE biplot (23), with the greatest ability to discriminate genotypes, favoring the selection of superior genotypes. The GGE biplot revealed that the genotype G3 (ACM-M13-002) could be recommended for yield trait to all the environments studied since it is placed nearer to the highly interactive environment (E4) and also at the second concentric circle on the GGE biplot.

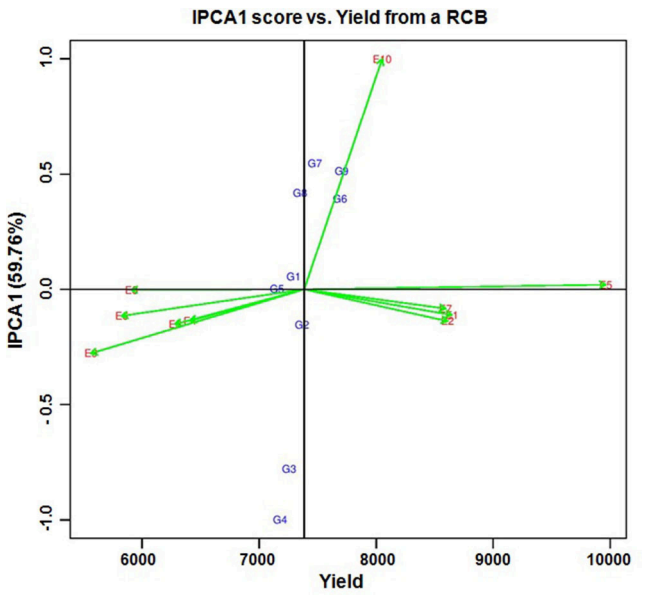

FIGURE 6 | AMMI biplot for IPCA1 score vs. yield in nine hybrids of maize from 10 environments in the state of Tamil Nadu. ***(G1)*** *NK6240,* ***(G2)*** *CO 6,* ***(G3)*** *ACM-M13-002,* ***(G4)*** *ACM-M13-007,* ***(G5)*** *CMH 11-583,* ***(G6)*** *VaMH 14020,* ***(G7)*** *VaMH 12014,* ***(G8)*** *900 M GOLD, and* ***(G9)*** *CMH 11-586.* ***(E1)*** *Coimbatore,* ***(E2)*** *Vagarai,* ***(E3)*** *Vridhachalam,* ***(E4)*** *Bhavanisagar,* ***(E5)*** *Coimbatore,* ***(E6)*** *Vagarai,* ***(E7)*** *Vridhachalam,* ***(E8)*** *Bhavanisagar,* ***(E9)*** *Vaigai dam, and* ***(E10)*** *Athiyandal.*

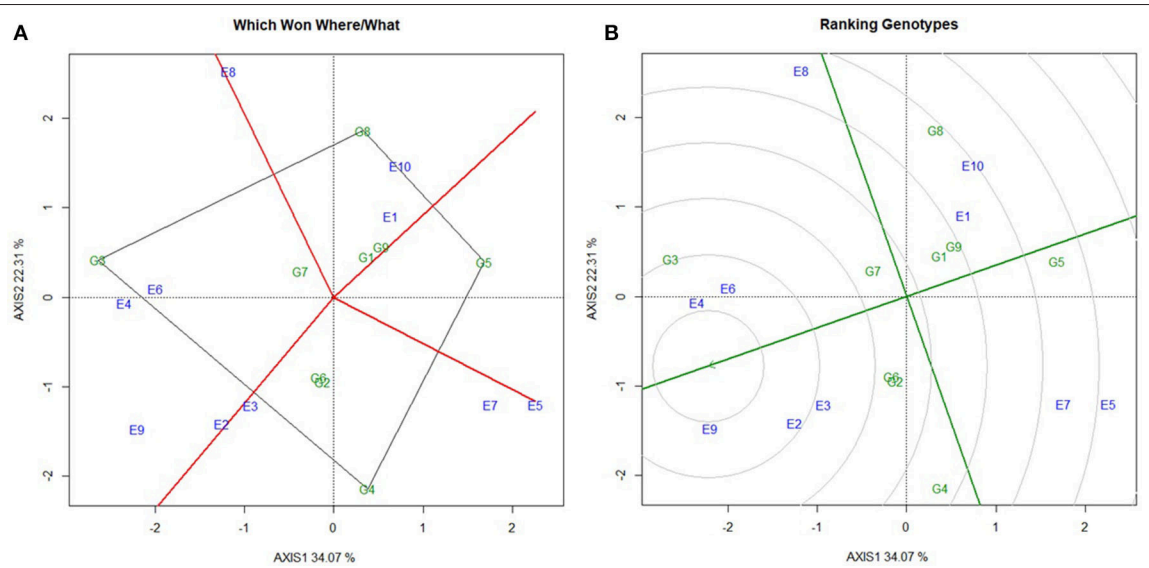

FIGURE 7 | Stability analysis to evaluate the mean grain yield of nine maize hybrids. **(A)** Polygon view of GGE biplot for grouping the environments, **(B)** GGE biplot for comparing the test hybrids with ideal environment. ***(G1)*** *NK6240,* ***(G2)*** *CO 6,* ***(G3)*** *ACM-M13-002,* ***(G4)*** *ACM-M13-007,* ***(G5)*** *CMH 11-583,* ***(G6)*** *VaMH 14020,* ***(G7)*** *VaMH 12014,* ***(G8)*** *900 M GOLD,* ***(G9)*** *CMH 11-586,* ***(E1)*** *Coimbatore,* ***(E2)*** *Vagarai,* ***(E3)*** *Vridhachalam,* ***(E4)*** *Bhavanisagar,* ***(E5)*** *Coimbatore,* ***(E6)*** *Vagarai,* ***(E7)*** *Vridhachalam,* ***(E8)*** *Bhavanisagar,* ***(E9)*** *Vaigai dam, and* ***(E10)*** *Athiyandal.*

In summary, we have successfully improved the β-carotene concentration in the CO6 hybrid through MABB. The β-carotene enriched hybrid developed in this study will hold great promise for nutritional and food security. Also, the improved maize inbreds with β-carotene will serve as a potential genetic material for the development of β-carotene rich cultivars in maize breeding programs.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

SN conceived and designed the methods and experiments. SN, TD, GK, NT, and RR constructed backcross progenies and managed field work. TD, BP, SC, JN, DM, KA, and VS conducted phenotype, genotype, and biochemical analyses. SN, BP, SC, and DM analyzed the data. LM and SM provided suggestions on experiments and data analysis. SN, KA, DM, and SC drafted the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Financial support from the Department of Biotechnology (DBT), Government of India (GOI), through project entitled Enrichment of nutritional quality in maize through molecular breeding (BT/PR10922AGII/106/9442014 dt.25.3.2015) and Ministry of Science and Technology, Department of Biotechnology, Government of India DBT's Twinning programme for the NE for the project entitled Marker assisted introgression of LyCE gene for enhanced ProA in maize (BT/166/NE/TBP/2011 dt 12.12.2011) are acknowledged. The funders had no role in the work design, data collection and analysis, or decision and preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2020.00134/full#supplementary-material

REFERENCES

1. Nuss ET, Tanumihardjo SA. A paramount staple crop in the context of global nutrition. *Comp Rev Food Sci.* (2010) 9:417–36. doi: 10.1111/j.1541-4337.2010.00117.x
2. Maqbool MA, Aslam M, Beshir A, Khan MS. Breeding for provitamin A biofortification of maize (*Zea mays* L.). *Plant Breed.* (2018) 137:451–69. doi: 10.1111/pbr.12618
3. Harjes CE, Torbert R, Rocheford L, Bai TP, Brutnell CB, Kandianis SG, et al. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science.* (2008) 319:330–3. doi: 10.1126/science.1150255
4. Haskell MJ. The challenge to reach nutritional adequacy for vitamin A: β-carotene bioavailability and conversion evidence in humans. *Am J Clin Nutr.* (2012) 96:1193S−203. doi: 10.3945/ajcn.112.034850
5. Weber D, Grune T. The contribution of β-carotene to vitamin A supply of humans. *Mol Nutr Food Res.* (2012) 56:251–8. doi: 10.1002/mnfr.201100230
6. Muthusamy V, Hossain F, Nepolean T, Saha S, Agrawal PK, Guleria SK, et al. Genetic variability for kernel β-carotene and utilization of crtRB1 3'TE gene for biofortification in maize (*Zea mays* L.). *Indian J Genet Plant Breed.* (2012) 72:189–94.
7. Yan J, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang X, et al. Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain. *Nat Genet.* (2010) 42:322–7. doi: 10.1038/ng.551
8. Babu R, Rojas NP, Gao S, Yan J, Pixley K. Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. *Theor App Genet.* (2013) 126:389–99. doi: 10.1007/s00122-012-1987-3
9. Muthusamy V, Hossain F, Thirunavukkarasu N, Choudhary M, Saha S, Bhat JS, et al. Development of β-carotene rich maize hybrids through marker-assisted introgression of β-carotene hydroxylase allele. *PLoS ONE.* (2014) 10:e0122130. doi: 10.1371/journal.pone.0113583
10. Zunjare RU, Hossain F, Muthusamy V, Baveja A, Chauhan HS, Bhat JS, et al. Development of biofortified maize hybrids through marker-assisted stacking of β-carotene hydroxylase, lycopene-ε-cyclase and opaque2 genes. *Front Plant Sci.* (2018) 9:178. doi: 10.3389/fpls.2018.00178
11. Sagare DB, Shetti P, Surender M, Reddy SS. Marker-assisted backcross breeding for enhancing β-carotene of QPM inbreds. *Mol Breed.* (2019) 39:31. doi: 10.1007/s11032-019-0939-x
12. Nallathambi G, Ganesan KN, Tamilarasi PM, Sain D, Thiyagarajan K, Veerabadhiran P, et al. A high yielding multiple disease resistant TNAU maize hybrid CO 6 for Tamil Nadu. *Madras Agric J.* (2012) 99:677–80.
13. Hossain F, Muthusamy V, Pandey N, Vishwakarma AK, Baveja A, Zunjare RU, et al. Marker-assisted introgression of opaque2 allele for rapid conversion of elite hybrids into quality protein maize. *J genet.* (2018) 97:287–98. doi: 10.1007/s12041-018-0914-z
14. Dellaporta SL, Wood J, Hicks JB. A plant DNA mini-preparation: version 2. *Plant Mol Biol Rep.* (1983) 1:19–21. doi: 10.1007/s11033-008-9362-9
15. Pukalenthy B, Manickam D, Chandran S, Adhimoolam K, Sampathrajan V, Rajasekaran R, et al. Incorporation of opaque-2 into 'UMI 1200', an elite maize inbred line, through marker-assisted backcross breeding. *Biotechnol Biotechnological Equip.* (2019) 33:144–53. doi: 10.1080/13102818.2018.1556121
16. Sundaram RM, Vishnupriya RM, Biradar SK, Laha GS, Ashok Reddy G, Shobha Rani N, et al. Marker assisted introgression of bacterial blight resistance in Samba Mahsuri, an elite indica rice variety. *Euphytica.* (2008) 160:411–22. doi: 10.1007/s10681-007-9564-6
17. Anonymous. *Descriptors for Maize.* Mexico City: International Board for Plant Genetic Resources. (1991). p. 86
18. Rodriguez-Amaya DB, Kimura M. *Harvest Plus Handbook for Carotenoid Analysis*. Washington DC: International Food Policy Research Institute. (2004). p. 10.
19. Quackenbush FW. Corn carotenoids: effects of temperature and moisture on losses during storage. *Cereal Chem.* (1963) 40:266–9.
20. Weber EJ. Carotenoids and tools of corn grain determined by HPLC. *J Am Oil Chem Soc.* (1987) 64:1129–34.
21. Pacheco A, Vargas M, Alvarado G, Rodríguez F, Lopez M, Crossa J, et al. (2016) *GEA-R (Genotype× Environment Analysis with R for Windows.) Version 4.0. CIMMYT Research Data and Software Repository Network*. Mexico: CIMMYT.
22. Gauch HG. A simple protocol for AMMI analysis of yield trials. *Crop Sci.* (2013) 53:1860–9. doi: 10.2135/cropsci2013.04.0241
23. Yan W, Hunt LA, Sheng Q, Szlavnics Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* 40:597–605. doi: 10.2135/cropsci2000.403597x
24. Neuffer MG, Coe EH, Wessler SR. *Mutants of Maize*. Long Island, NY: Cold Spring Harbor Laboratory Press. (1997). p. 640.

25. Lu H, Romero-Severson J, Bernardo R. Chromosomal regions associated with segregation distortion in maize. *Theor Appl Genet.* (2002) 105:622–8. doi: 10.1007/s00122-002-0970-9
26. Mangelsdorf PC, Jones DF. The expression of mendelian factors in the gametophyte of maize. *Genetics.* (1926) 11:423–55.
27. Chandran S, Pukalenthy B, Adhimoolam K, Manickam D, Sampathrajan V, Chocklingam V, et al. Marker-assisted selection to pyramid the opaque-2 (o2) and β-carotene (crtRB1) genes in maize. *Front Gene.* (2019) 10:859. doi: 10.3389/fgene.2019.00859
28. Shetti P, Sagare DB, Surender M, Reddy SS. Development of lysine and tryptophan rich maize (Zea mays) inbreds employing marker assisted backcross breeding. *Plant Gene.* (2020) 23:100236. doi: 10.1016/j.plgene.2020.100236
29. Goswami R, Zunjare RU, Khan S, Baveja A, Muthusamy V, Hossain F. Marker-assisted introgression of rare allele of β-carotene hydroxylase (crtRB1) gene into elite quality protein maize inbred for combining high lysine, tryptophan and provitamin A in maize. *Plant Breed.* (2019) 138:174–83. doi: 10.1111/pbr.12676

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Natesan, Duraisamy, Pukalenthy, Chandran, Nallathambi, Adhimoolam, Manickam, Sampathrajan, Muniyandi, Meitei, Thirunavukkarasu, Kalipatty Nalliappan and Rajasekaran. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Nutrition

ORIGINAL RESEARCH
published: 15 October 2020
doi: 10.3389/fnut.2020.559120

Genome-Wide SNP Discovery and Mapping QTLs for Seed Iron and Zinc Concentrations in Chickpea (*Cicer arietinum* L.)

Syed Sab[1,2], Ramappa Lokesha[1], D. M. Mannur[1], Somasekhar[1], Kisan Jadhav[1], Bingi Pujari Mallikarjuna[2], Laxuman C[3], Sharanbasappa Yeri[3], Vinod Valluri[2], Prasad Bajaj[2], Annapurna Chitikineni[2], AnilKumar Vemula[2], Abhishek Rathore[2], Rajeev Kumar Varshney[2], I. Shankergoud[1] and Mahendar Thudi[2]**

[1] Department of Genetics and Plant Breeding, University of Agricultural Sciences - Raichur (UAS-R), Raichur, India, [2] Center of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, [3] Zonal Agricultural Research Station, University of Agricultural Sciences - Raichur, Kalaburagi, India

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University,
United States

Reviewed by:
Anupam Singh,
Shriram Bioseed Genetics, India
Debasis Chattopadhyay,
National Institute of Plant Genome Research (NIPGR), India

***Correspondence:**
Mahendar Thudi
t.mahendar@cgiar.org
I. Shankergoud
ishankergoud@rediffmail.com

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Nutrition

Received: 05 May 2020
Accepted: 21 August 2020
Published: 15 October 2020

Citation:
Sab S, Lokesha R, Mannur DM, Somasekhar, Jadhav K, Mallikarjuna BP, C L, Yeri S, Valluri V, Bajaj P, Chitikineni A, Vemula A, Rathore A, Varshney RK, Shankergoud I and Thudi M (2020) Genome-Wide SNP Discovery and Mapping QTLs for Seed Iron and Zinc Concentrations in Chickpea (Cicer arietinum L.). Front. Nutr. 7:559120. doi: 10.3389/fnut.2020.559120

Biofortification through plant breeding is a cost-effective and sustainable approach towards addressing micronutrient malnutrition prevailing across the globe. Screening cultivars for micronutrient content and identification of quantitative trait loci (QTLs)/genes and markers help in the development of biofortified varieties in chickpea (*Cicer arietinum* L.). With the aim of identifying the genomic regions controlling seed Fe and Zn concentrations, the $F_{2:3}$ population derived from a cross between MNK-1 and Annigeri 1 was genotyped using genotyping by sequencing approach and evaluated for Fe and Zn concentration. An intraspecific genetic linkage map comprising 839 single nucleotide polymorphisms (SNPs) spanning a total distance of 1,088.04 cM with an average marker density of 1.30 cM was constructed. By integrating the linkage map data with the phenotypic data of the $F_{2:3}$ population, a total of 11 QTLs were detected for seed Fe concentration on CaLG03, CaLG04, and CaLG05, with phenotypic variation explained ranging from 7.2% (*CaqFe3.4*) to 13.4% (*CaqFe4.2*). For seed Zn concentration, eight QTLs were identified on CaLG04, CaLG05, and CaLG08. The QTLs individually explained phenotypic variations ranging between 5.7% (*CaqZn8.1*) and 13.7% (*CaqZn4.3*). Three QTLs for seed Fe and Zn concentrations (*CaqFe4.4*, *CaqFe4.5*, and *CaqZn4.1*) were colocated in the "*QTL-hotspot*" region on CaLG04 that harbors several drought tolerance-related QTLs. We identified genes in the QTL regions that encode iron–sulfur metabolism and zinc-dependent alcohol dehydrogenase activity on CaLG03, iron ion binding oxidoreductase on CaLG04, and zinc-induced facilitator-like protein and ZIP zinc/iron transport family protein on CaLG05. These genomic regions and the associated markers can be used in marker-assisted selection to increase seed Fe and Zn concentrations in agronomically superior chickpea varieties.

Keywords: chickpea, seed, iron, zinc, SNP, QTL

INTRODUCTION

Micronutrients are vital for healthy growth and development of the human body, and their adequate intake through diet is a prerequisite for various metabolic functions. However, a majority of the world's population face difficulties in meeting the recommended daily allowances (RDA). This has led to micronutrient malnutrition (also known as "hidden hunger") that affects one in three people worldwide (1). Iron (Fe) and zinc (Zn) are important micronutrients that play a major role in the normal growth and development of humans by acting as cofactors for several proteins including hemoglobin, cytochrome, and transcription factors (2). Fe and Zn deficiency causes anemia, tissue hypoxia, dwarfism, lowered disease immunity, stunting, poor cognitive development, and orificial and acral dermatitis (3, 4). It is estimated that by 2050, around 1.4 billion women and children would be Fe deficient and 175 million zinc deficient (5). This risk will be high in Southeast Asia, Africa, and the Middle East. Micronutrient malnutrition has received increased global attention in recent decades, with efforts to address it by various strategies such as mineral supplementation and dietary diversification. Biofortification of staple crops through plant breeding is a highly cost-effective and long-term strategy to enhance micronutrient density in crop plants (6). Efforts in this direction have involved understanding the genetic architecture of Fe and Zn concentrations in seeds of cereals crops (7–10). More recently, the focus has been on pulses that serve as secondary staples, mainly because of their increased protein content (11). Among legumes, common bean (*Phaseolus vulgaris*), lentil (*Lens culinaris*), Pea (*Pisum sativum*), and mung bean (*Vigna radiata*) have been targeted for micronutrient studies (6).

Chickpea, a highly nutritious diploid ($2n = 2x = 16$) legume crop plant having a genome size of 738 Mbp (12) ranks second after dry beans in terms of global production among pulses, with over 14.77 million tons produced in 2017 (FAOSTAT, accessed on 17 February 2020). Chickpea is considered an excellent whole food and a source of carbohydrate (52–71%), protein (18–24%), fiber (10–23%), fat (4–10%), minerals and vitamins, dietary fiber, folate, β-carotene, and some important fatty acids (13, 14). Genomic research during the last decade enabled the decoding of the genome sequence (12) and sequencing of several germplasm lines (15–17) and provided insights into the genetics of complex abiotic and biotic stresses (18). Besides simplifying these complex traits, molecular breeding lines with enhanced drought tolerance and disease resistance were developed and released for commercial cultivation (19–21). Nevertheless, very few genomic studies have attempted to understand the variability for nutritionally important traits. The variability study (22) for physicochemical, nutritional and cooking quality traits among 30 chickpea germplasm lines indicated that higher 100-seed weight reduced cooking time. Using principal component analysis, variation for seed and flour characteristics among 79 chickpea germplasm accessions used in European breeding programs revealed higher protein and fiber in desi and higher fat content in kabuli types (23). Recently, studies on genotype and environment effects on seed quality traits demonstrated that environment plays an important role in variation in amylose and amylopectin content (24). None of these studies utilized genomics tools to understand the genetics of these traits.

Marker trait associations for protein content have been reported using simple sequence repeat markers (25). In the case of chickpea, there are very few studies on Fe and Zn accumulation in seeds. Genetic variations in the concentration of Fe (36.2–86.4 mg kg^{-1}) and Zn (18.6–62.2 mg kg^{-1}) were reported in 94 diverse accessions of chickpea (26). Fe and Zn in seeds appear to be a complex trait governed by a number of major genes/quantitative trait loci (QTLs), and the concentration of these minerals is highly influenced by edaphic and environmental factors (2). Recently, besides the variation in seed mineral concentration among cultivars and breeding lines (27, 28), markers linked with seed Fe and Zn content were reported in a select set of germplasm collections (2, 26). The identification of new sources of Fe and Zn concentrations, their genetic dissection, and using them to diversify the breeding population is a continuous process in developing high-yielding chickpea varieties with enhanced nutrition content. Cost-effective and high-throughput discovery and genotyping of single nucleotide polymorphism (SNP) markers using genotyping-by-sequencing (GBS) approach (29) have been widely used in many crop species including chickpea to assess diversity, in genome-wide association studies, in constructing high-density genetic linkage maps, and QTL identification (2, 30–33).

This study reports on high-throughput genome-wide marker discovery and genotyping using GBS approach and QTLs for Fe and Zn concentrations in chickpea, using an $F_{2:3}$ intraspecific mapping population developed from MNK-1 × Annigeri 1.

MATERIALS AND METHODS

Mapping Population and DNA Extraction

A mapping population consisting of 188 F_2 plants was developed from a cross between MNK-1 and Annigeri 1. MNK-1 is a kabuli type, extra-large-seeded chickpea variety having Fe and Zn concentrations of 72.88 and 30.98 mg kg^{-1}, respectively. Annigeri 1 is a desi-type, medium seed size variety having Fe and Zn concentrations of 98.67 and 37.07 mg kg^{-1}, respectively. Fresh, young leaves were collected from parental lines, and each F_2 individual and high-quality genomic DNA was extracted using high throughput NucleoSpin® 96 Plant II Kit (Macherey-Nagel, Düren, Germany) following the manufacturer's protocol. The DNA was quantified using a spectrophotometer (Shimadzu UV160A, Japan) and used to prepare the GBS library for SNP discovery. Individual F_2 plants were selfed, and seeds harvested were advanced to develop an $F_{2:3}$ population for phenotyping seed Fe and Zn contents.

Genotyping-by-Sequencing and SNP Calling

Genotyping by sequencing was carried out for SNP calling between the parental genotypes and genotyping the 188 F_2 progenies following the protocol described by Elshire et al. (29). GBS libraries were prepared by restriction digestion of DNA of each of the F_2s as well as the parents with *ApeKI*

endonuclease (recognition site: G/CWCG), followed by ligation with uniquely barcoded adapters using T4 DNA ligase. An equal proportion of barcoded adapters ligated DNA fragments from each sample were mixed to construct the GBS libraries, which were amplified and purified in order to remove excess adapters, followed by sequencing on the HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA). The sequence reads from fastq files were used for SNP identification using GBSv2 (34) pipeline implemented in TASSEL v5 (35). In short, the sequencing reads were first searched for barcode information and reads with no Ns. The barcode containing reads was then demultiplexed according to the barcode sequence and trimmed. The remaining good quality and distinct reads (tags) were retrieved in the form of fastq and aligned against the draft genome sequence of chickpea (36) using Bowtie2 (37). The alignment file was then parsed through the remaining GBSv2 pipeline for SNP calling and genotyping. The SNPs identified were then filtered in order to retain data with contrasting alleles in parental genotypes and those having <30% missing data for further analysis.

Phenotyping for Seed Fe and Zn

During the crop season of 2019, $F_{2:3}$ progenies (plant:progeny rows) were planted following augmented design at the Zonal Agricultural Research Station, Kalaburagi, Karnataka (latitude, 17.36 and longitude, 76.82). Each experimental block consisted of a single 2-m row plot with 30 cm spacing between rows and 10 cm interplant spacing. The experimental plot was divided into six blocks to reduce heterogeneity, and each occupied 27 lines plus 4 checks (MNK-1, Annigeri 1, KAK-2, and BGD-103) replicated twice in each block. Standard cultivation practices were followed to ensure healthy crop growth.

Mature seeds of each parental line and mapping individual (five to six representative plants from each $F_{2:3}$ progeny line) were collected separately during harvesting stage. Care was taken to avoid the contamination of seeds with dust and inert particles during harvest and while sample preparation. Five grams of dry seed sample (with 5.6% moisture content) from each accession was ground into fine powder, and 0.5 g of fine ground powder was put in a reaction vessel to which was added 10 ml of di-acid mixture (nitric acid and hydrochloric acid in a 9:1 ratio). The vessel was sealed and loaded into the chamber and microwave of the Microwave Digestion System (Anton Paar Company, USA), which was set to run for 40 min at 180°C. After the vessels cooled down to <50°C, they were removed from the chamber and moved to an exhaust clean hood to vent excess pressure slowly. The digested samples from the vessel were then placed in a clean container and diluted with distilled water to make up a volume of 50 ml in a conical flask. They were then analyzed for Fe and Zn concentrations. An atomic absorption spectrophotometer (Analytik Jena AG, contrAA 700, Germany) was used to estimate Fe and Zn concentrations based on the absorption of light at a wavelength of 370 and 388 nm, respectively. Both Fe and Zn concentrations were estimated and expressed as mg kg^{-1} (ppm) seed.

Statistical Analysis

Variance analysis was done using PROC MIXED of SAS v9.4 (38), considering blocks and checks as fixed and $F_{2:3}$ progeny lines as random. Repeated checks were used to estimate error mean square. Best linear unbiased predictors (BLUPs) and best linear unbiased estimates (BLUEs) were estimated for random and fixed factors, respectively. Genetic variability parameters including mean, range, frequency distribution, phenotypic coefficient of variance (PCV), genotypic coefficient of variance (GCV), broad sense heritability [$h^2_{(bs)}$], and genetic advance per mean (GAM) were computed using SAS.

Genetic Linkage Map Construction

The allele calls for all genotypes were compiled and used for genetic linkage analysis by employing JoinMap V4.1 program (39). In total, 185 F_2s were used to construct the genetic map after discarding three individuals with insufficient data. Chi square goodness-of-fit test was performed ($p < 0.05$) to test the segregation distortion for each marker. Highly distorted and unlinked markers were excluded from the analysis. Regression mapping algorithm with maximum recombination frequency of 0.3 at minimum logarithm of odds (LODs) of 3 and jump threshold of 5 were used to group the markers into eight linkage groups (LGs). Ripple command was used after adding each marker locus to confirm marker order. Map distance was estimated using Kosambi's mapping function (40). The final high-resolution linkage map was generated using LinkageMapView package in R software (41).

QTL Analysis

The genotyping data of the intraspecific linkage map and phenotypic data for iron and zinc content from the $F_{2:3}$ population were used for QTL analysis using QTL Cartographer V2.5 software (42). Composite interval mapping (CIM) method was used by selecting Model 6 with the default window size of 10 cM, control marker number 5, and using backward stepwise regression. For each trait, the LOD score threshold was determined by performing 1,000 permutations at significance level of $p \leq 0.05$. The amount of phenotypic variation explained was determined using the coefficient of determination (R^2) value and expressed as phenotypic variation explained (PVE) percentage.

Identification of Genes in QTL Regions

The QTL region's coordinates were compared to the reference genome's gff file using bedtools (43) to identify the genes lying in the QTL region and functionally categorized using UniProt KB database (http://www.uniprot.org/).

RESULTS

Genetic Variability for Seed Fe and Zn Concentrations

The parental lines of the mapping population had wide variation for seed Fe and Zn concentrations. Parent MNK-1 had seed Fe and Zn concentrations of 72.88 and 30.98 mg kg^{-1}, respectively, while parental line Annigeri 1 had seed

TABLE 1 | Phenotypic variation for seed iron and zinc concentrations in $F_{2:3}$ generation of the cross MNK-1 × Annigeri 1.

Traits	Mean	Range		GCV (%)	PCV (%)	$h^2_{(bs)}$ (%)	GAM (%)
		Min.	Max.				
Seed Fe	82.15	51.60	118.74	21.89	22.17	97.50	44.45
Seed Zn	36.10	17.89	45.60	14.07	15.11	86.68	26.92

GCV, genotypic coefficient of variance; PCV, phenotypic coefficient of variance; $h^2_{(bs)}$, broad sense heritability; GAM, genetic advance per mean.

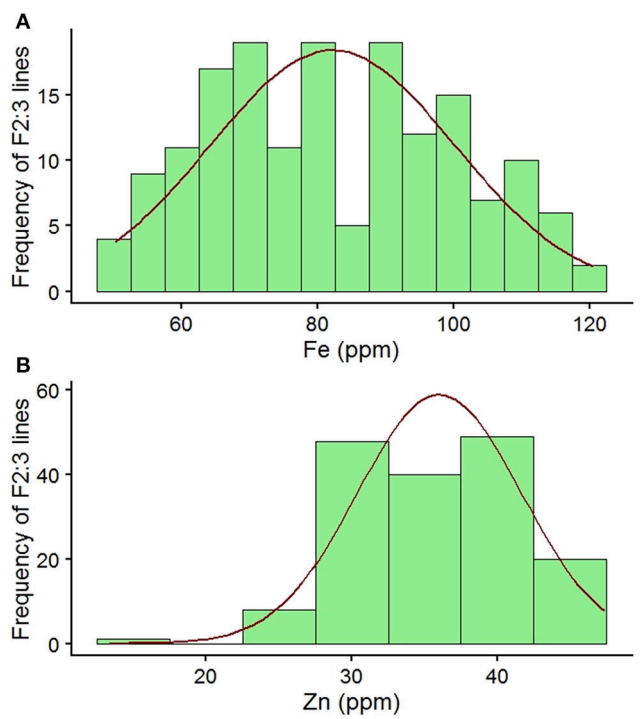

FIGURE 1 | (A, B) Frequency distribution of seed Fe and Zn concentrations among $F_{2:3}$ individuals of an intraspecific mapping population (MNK-1 × Annigeri 1) along with checks and parental lines.

TABLE 2 | Variance component for $F_{2:3}$ progeny lines and F-statistic for block, checks, and check vs. progeny lines for seed Fe and Zn concentrations ANOVA for augmented design in $F_{2:3}$ generation of the cross MNK-1 × Annigeri 1.

Source of variation		Seed Fe	Seed Zn
Random effect	**Df**	**VC**	**VC**
$F_{2:3}$ progeny lines	161	323.52**	25.78**
Fixed Effect	**Df**	**F-statistic**	**F-statistic**
Block	5	0.37	1.69
$F_{2:3}$ progeny lines vs. Checks	1	0.002	0.53
Checks	3	92.59**	24.25**
Error	15	8.30	3.96

VC, variance component for random factors.
***Significant at prob <0.01.*

Fe and Zn concentrations of 98.67 and 37.07 mg kg^{-1}, respectively. The $F_{2:3}$ population showed wide variability for both the traits (**Table 1**, **Figure 1**). Seed Fe concentration ranged from 51.60 to 118.74 mg kg^{-1} with a mean of 82.15 mg kg^{-1}. Seed Zn concentration ranged from 17.89 to 45.60 mg kg^{-1} with a mean of 36.10 mg kg^{-1}. Analysis of variance (ANOVA) results indicated highly significant genotypic effects (**Table 2**). A high GCV (21.89%) and PCV (22.17%) for seed Fe concentration and moderate GCV (14.07%) and PCV (15.11%) for seed Zn concentration were observed among the $F_{2:3}$ progenies. Very high broad sense heritability values of 97.50 and 86.68% were observed for seed Fe and Zn concentrations, respectively.

Identification of SNPs Using GBS Approach and Construction of Intraspecific Linkage Map

Genotype by sequencing of the parental genotypes generated ∼211.46 Mb for MNK-1 and ∼144.44 Mb for Annigeri1. For 188F_2s, ∼46.6 Gb was generated with average of ∼257.76 Mb per sample. A total of 31,311 SNPs were identified, of which 7,204 (23%) were polymorphic between MNK-1 and Annigeri 1, and the remaining 17,034 (54.40%) were monomorphic. A set of 7,204 polymorphic SNPs was used for linkage map analysis. As a result, a total of 839 SNPs were mapped on the eight chickpea chromosomes postfiltration of SNPs' set (**Supplementary Table 1**). A total of 7,204 SNPs from GBS data were used to generate the intraspecific linkage map. Of these, 839 SNPs were grouped into eight linkage groups (CaLG), whereas the remaining 6,365 SNPs were unlinked to any of the linkage groups. Marker density per linkage group varied from 24 (CaLG05) to 318 (CaLG04), with an overall average of 104.8 markers per linkage group (**Table 3**, **Figure 2**). The final linkage map had a total length of 1,088.04 cM, with CaLG05 (164.17 cM) being the largest and CaLG08 (116.98 cM) being the smallest. The average intermarker distance in the linkage map was 1.30 cM, with CaLG04 being the densest (0.49 cM) and CaLG05 being the least dense (6.84 cM). The final linkage map is shown in **Figure 2**.

QTL Mapping for Seed Fe and Zn Concentrations

Genotyping data of mapped 839 SNPs was integrated with the phenotypic data of the Fe and Zn concentrations and analyzed using composite interval mapping (CIM) method.

TABLE 3 | Features of an intraspecific genetic map based on the F_2 population of the cross MNK-1 × Annigeri 1.

Linkage group	Polymorphic markers	Markers mapped	Map length (cM)	Intermarker distance (cM)
CaLG01	1,126	76	134.56	1.77
CaLG02	625	28	132.68	4.74
CaLG03	866	138	126.92	0.92
CaLG04	1,081	318	155.80	0.49
CaLG05	917	24	164.17	6.84
CaLG06	1,184	75	137.32	1.83
CaLG07	1,179	108	119.60	1.11
CaLG08	226	72	116.98	1.62
Total	**7,204**	**839**	**1,088.04**	**1.30**

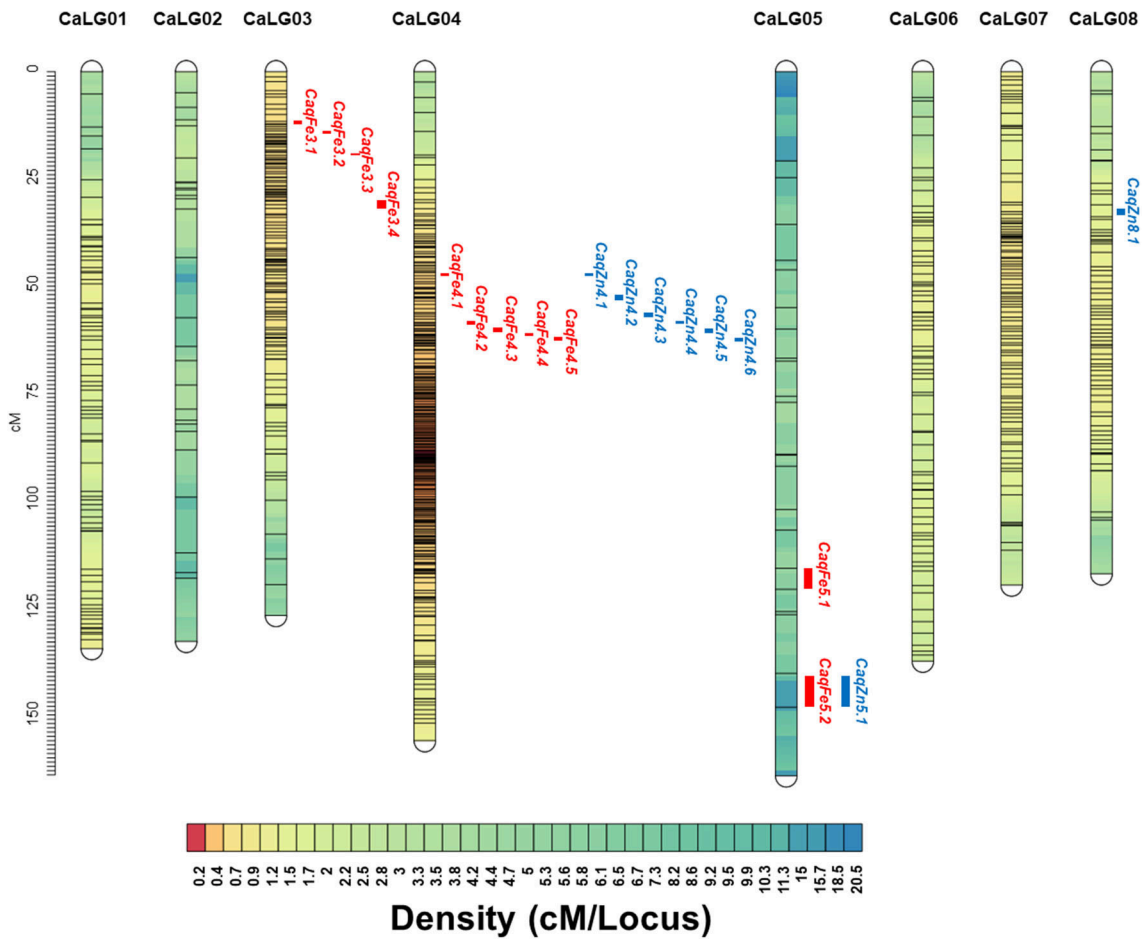

FIGURE 2 | Intraspecific genetic linkage map of chickpea constructed using the F_2 population derived from the cross MNK-1 × Annigeri 1. The linkage map with eight linkage groups (CaLG01–CaLG08) shows the quantitative trait loci (QTLs) for seed Fe and Zn concentrations; 11 QTLs for seed Fe and 8 QTLs for seed Zn concentration are shown on the left side of the corresponding linkage group.

For seed Fe concentration, a total of 11 QTLs were mapped on CaLG03, CaLG04, and CaLG05 with LOD thresholds ranging from 3.0 (*CaqFe3.4* and *CaqFe5.2*) to 4.8 (*CaqFe4.2*) (**Table 4**, **Figure 2**). PVE by individual QTLs ranged from 7.2% (*CaqFe3.4*) to 13.4% (*CaqFe4.2*). At all the five loci on CaLG04 (*CaqFe4.1*, *CaqFe4.2*, *CaqFe4.3*, *CaqFe4.4*, and *CaqFe4.5*), alleles from parent MNK-1 favored seed Fe concentration. For seed Zn concentration, eight QTLs were detected on

TABLE 4 | Quantitative trait loci (QTLs) for seed Fe and Zn concentrations in $F_{2:3}$ population.

Trait	QTL	Linkage group	Left marker	Right marker	Position (cM)	LOD	Additve variance	Phenotypic variation explained (PVE; %)
Seed Fe	*CaqFe3.1*	CaLG03	SCA3_38112971	SCA3_31331006	12.81	4.2	21.2	9.3
	CaqFe3.2	CaLG03	SCA3_38103368	SCA3_38060814	15.51	4.1	22.4	8.0
	CaqFe3.3	CaLG03	SCA3_36227565	SCA3_36549464	20.41	4.5	19.8	9.7
	CaqFe3.4	CaLG03	SCA3_39612119	SCA3_37949465	32.91	3.0	16.1	7.2
	CaqFe4.1	CaLG04	SCA4_11398699	SCA4_11244334	48.91	3.1	−12.9	9.8
	CaqFe4.2	CaLG04	SCA4_23943928	SCA4_13840227	59.21	4.8	−17.0	13.4
	CaqFe4.3	CaLG04	SCA4_13787448	SCA4_13845581	61.11	4.0	16.6	11.6
	CaqFe4.4	CaLG04	SCA4_13724666	SCA4_4477846	62.61	3.3	−14.3	11.6
	CaqFe4.5	CaLG04	SCA4_13840191	SCA4_3223962	63.81	3.4	−15.5	11.2
	CaqFe5.1	CaLG05	SCA5_40736291	SCA5_47154065	119.81	4.2	4.7	7.4
	CaqFe5.2	CaLG05	SCA5_37174786	SCA5_35312501	145.31	3.0	15.2	10.5
Seed Zn	*CaqZn4.1*	CaLG04	SCA4_11398699	SCA4_11244334	48.91	4.4	−7.0	11.4
	CaqZn4.2	CaLG04	SCA4_12558541	SCA4_11441673	53.41	4.6	−6.8	6.5
	CaqZn4.3	CaLG04	SCA4_13660846	SCA4_13787649	57.61	6.3	−8.1	13.7
	CaqZn4.4	CaLG04	SCA4_23943928	SCA4_13840227	59.21	6.0	−7.9	13.0
	CaqZn4.5	CaLG04	SCA4_13787448	SCA4_13845581	61.11	5.5	−8.0	12.1
	CaqZn4.6	CaLG04	SCA4_13840191	SCA4_3223962	63.81	4.6	−7.7	10.7
	CaqZn5.1	CaLG05	SCA5_37174786	SCA5_35312501	146.31	3.8	5.2	10.2
	CaqZn8.1	CaLG08	SCA8_4106644	SCA8_16027922	33.81	3.8	−6.8	5.7

CaLG04, CaLG05, and CaLG08 (**Table 4**, **Figure 2**). QTLs *CaqZn5.1* and *CaqZn8.1* had the lowest LOD threshold of 3.0, whereas QTL *CaqZn4.3* had the highest LOD score of 6.3. The PVE of these QTLs ranged between 5.7% (*CaqZn8.1*) and 13.7% (*CaqZn4.3*). At all the loci except *CaqZn5.1*, the alleles from parent MNK-1 favored seed Zn concentration.

Colocalization of Seed Fe and Zn Concentration QTLs and Genes in the QTL Regions

Based on the physical position of the SNP markers linked with the seed Fe and Zn QTLs, we found that three QTLs (*CaqFe4.4*, *CaqFe4.5*, and *CaqZn4.1*) are colocalized in the "*QTL-hotspot*" region that harbors several drought tolerance related QTLs on CaLG04, as reported earlier (**Table 5**). We identified a total of 7,496 genes in the QTL regions reported in the study (**Supplementary Table 2**). Among these, we identified three genes—Ca_20872, Ca_00798, and Ca_01283—in the *CaqFe3.1* QTL region that encode for iron–sulfur assembly protein IscA-like 1, iron–sulfur cluster assembly protein IscA, and Zn-dependent alcohol dehydrogenase family class III protein, respectively. Further, gene Ca_04513 that encodes for iron ion binding oxidoreductase was identified in the *CaqFe4.4* QTL region. Interestingly, we identified gene Ca_01633 that encodes for ZIP zinc/iron transport family protein in the QTL region of *CaqZn5.1*.

DISCUSSION

Growing hunger and high levels of different forms of malnutrition are major challenges to achieving food and nutritional security. Effective interventions like biofortification aimed at guaranteeing access to nutritious foods are needed to address micronutrient malnutrition or hidden hunger. In several pulse crops, biofortification through plant breeding has gained momentum in the past decade (6). In chickpea, knowledge of genetic variation underlying the concentration of important seed minerals such as Fe and Zn and their association with yield-related traits is vital for accelerating breeding for biofortified varieties. In the present study, an $F_{2:3}$ population of chickpea was used to identify QTLs for seed Fe and Zn concentrations.

The parental lines MNK-1 and Annigeri 1 had seed Fe concentrations of 72.88 and 98.67 mg kg^{-1} and seed Zn concentrations of 30.98 and 37.07 mg kg^{-1}, respectively. The seed Fe concentration values of the cultivars are comparatively higher than the range of 49–56 mg kg^{-1} in chickpea cultivars reported by Vandemark et al. (28). With respect to seed Zn concentration, the parental values are comparable (35–43 mg kg^{-1}) to Vandemark et al. (28) but higher than the range of 21–28 mg kg^{-1} as reported by Ray et al. (27) for chickpea cultivars. Jayalaxmi et al. (45) reported that Fe concentration varied from 26 to 146 mg kg^{-1} and Zn concentration from 35 to 77 mg kg^{-1} in 56 chickpea varieties. Analysis of variance (ANOVA) results inferred a highly significant difference among $F_{2:3}$ individuals for both seed Fe and Zn concentrations (**Table 2**). The seed Fe and Zn concentrations were in the range of 51.60–118.74

TABLE 5 | Colocalization of Fe and Zn Quantitative trait loci (QTLs) in the *"QTL-hotspot"* region reported by Varshney et al. (44).

Fe/Zn QTL reported in this study				"*QTL-hotspot*" region				
Pseudomolecule	**Start**	**End**	**QTL**	**Pseudomolecule**	**Start**	**End**	**QTL**	**Overlap**
Ca4	4477846	13724666	*CaqFe 4.4*	Ca4	13239546	13378761	*"QTL-hotspot_a"*	139215
Ca4	3223962	13840191	*CaqFe 4.5*	Ca4	13239546	13378761	*"QTL-hotspot_a"*	139215
Ca4	3223962	13840191	*CaqZn4.1*	Ca4	13239546	13378761	*"QTL-hotspot_a"*	139215
Ca4	4477846	13724666	*CaqFe4.4*	Ca4	13393647	13547009	*"QTL-hotspot_b"*	153362
Ca4	3223962	13840191	*CaqFe4.5*	Ca4	13393647	13547009	*"QTL-hotspot_b"*	153362
Ca4	3223962	13840191	*CaqZn4.1*	Ca4	13393647	13547009	*"QTL-hotspot_b"*	153362

and 17.89–45.60 mg kg^{-1}, respectively (**Table 1**). It is evident from the distribution graph that $F_{2:3}$ individuals recorded values beyond the checks for both seed Fe and Zn concentrations (**Figure 1**). In addition, a very high broad sense heritability was observed among the $F_{2:3}$ individuals for both seed Fe [$h^2_{(bs)} = 97.50$] and Zn [$h^2_{(bs)} = 86.68$] concentrations. This indicates the feasibility of early generation selection that may result in improving these traits. Given that there are reports of significant environment effects on the concentration of minerals (26, 27) and this study was conducted only at a single location, better genotype × environment can be realized by further evaluating the advanced generation population across different locations or years.

Identification of genetic variants like SNPs is a key step in several genetic studies, including linkage mapping and QTL identification. Among the different genetic variants, SNP markers have become the markers of choice due to their high abundance and cost effectiveness. GBS provides increased efficiency in SNP discovery and genotyping by enabling high multiplexing of samples and simple library preparation procedures (29). This method involves only a targeted sequencing approach and does not require the complete genome sequencing of all individuals of the mapping population and parental genotypes. Due to its high-throughput efficiency, GBS has been used in recent studies in SNP identification and QTL mapping for important traits in chickpea (33, 46, 47). This study employed the GBS approach to identify SNPs in an intraspecific F_2 mapping population derived from two popular cultivars of chickpea. As a result, 7,204 novel SNPs were identified and used for genetic linkage analysis. The SNPs identified in this study were higher compared to other GBS studies in chickpea (33, 46) but lower than those reported by Deokar et al. (47). Of the 7,204 SNPs used, 839 were mapped on to eight linkage groups covering a total map distance of 1,088.04 cM and average marker density of 1.30 cM. The map density is comparable with that of Jaganathan et al. (33) and less dense compared to that of Hiremath et al. (48) and Deokar et al. (47, 49). Of these 839 SNPs, a majority (37%) were mapped on CaLG04. Earlier, Jaganathan et al. (33) made similar observations that may be attributed to the presence of high repeat-rich regions in Ca4 pseudomolecule of chickpea as reported by Varshney et al. (36).

The intraspecific linkage map was utilized to identify QTLs linked to seed Fe and Zn concentrations. A total of 11 QTLs for seed Fe concentration were detected on CaLG03, CaLG04, and CaLG05, which explained phenotypic variance in the range of 7.2–13.4%. For seed Zn concentration, eight QTLs were identified on CaLG03, CaLG05, and CaLG08. The phenotypic variance explained by these QTLs ranged between 5.7 and 13.7%. Previous studies have reported QTLs for seed Fe on chromosomes 1, 4, and 6 (26) and on chromosomes 1, 3, and 4 (2) and for seed Zn on chromosomes 1, 4, and 7 (26) and on chromosomes 2 and 3 (2). In all these studies, the distribution of QTLs on multiple chromosomes clearly indicates the quantitative nature and complexity of seed Fe and Zn concentration traits.

In this study, several colocated QTLs for seed Fe and Zn concentrations were identified on CaLG04 and CaLG05. For instance, QTLs *CaqFe*4.1-*CaqZn*4.1, *CaqFe*4.2-*CaqZn*4.2, *CaqFe*4.3-*CaqZn*4.1, and *CaqFe*4.5-*CaqZn*4.1 were colocalized on CaLG04 in the same marker interval SCA4_11398699-SCA4_11244334, SCA4_23943928-SCA4_13840227, SCA4_13787448-SCA4_13845581, and SCA4_13840191-SCA4_3223962, respectively (**Table 4**, **Figure 2**). On CaLG05, QTLs *CaqFe*5.1-*CaqZn*5.1 were colocalized within the SNP marker interval of SCA5_37174786-SCA5_35312501. Previous studies have reported QTL colocalization for micronutrient concentration in chickpea on chromosomes 4, 5, and 7 (2) and also in rice (10). Hence, these genomic regions on chromosome 4 may contain conserved QTLs for seed Fe and Zn concentrations across diverse genetic backgrounds. Pleiotropy or tight linkage of genes controlling multiple traits may be the reason for the colocalization of QTLs (50). These regions with colocated QTLs will be useful for simultaneous improvement using marker-assisted selection (10).

Interestingly, QTLs for seed Fe and Zn were also mapping in the "*QTL-hotspot region*" reported earlier (44). This genomic region was successfully introgressed into different elite genetic backgrounds, and molecular breeding lines with enhanced climate resilience have been reported (36). The molecular breeding lines may also be rich in seed Fe and Zn, as the QTLs are in the genomic region that was introgressed. By comparing the QTL region's coordinates to the kabuli reference genome of chickpea (12), the identified genes were found to be involved in Fe- and Zn-related activities. The $F_{2:3}$ population exhibited wide variation for seed Fe and Zn concentrations. A high heritability was obtained for both the traits studied.

This study was successful in using the GBS approach to develop a linkage map for an intraspecific population in chickpea. The map consisted of 839 SNPs spanning 1,088.04 cM. Genomic

regions for seed Fe and Zn concentrations were mapped on multiple linkage groups, and the linked SNPs were reported. These QTLs are good candidates for further validation, functional characterization, and gene cloning to clarify the role of these genes in Fe and Zn homeostasis in chickpea. The SNPs linked to seed Fe and Zn concentrations have potential for use as informative molecular tags in marker-assisted breeding to improve seed Fe and Zn concentrations in chickpea varieties.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

IS and MT conceived the idea. SS conducted the experiments, analyzed the data, and prepared the manuscript along with BM. AV and AR performed statistical analysis. VV and PB did the GBS analysis. RL, DM, LC, SY, and KJ contributed to the germplasm characterization and mapping population development and phenotyping. AC and RKV contributed consumables and to the generation of genotyping data. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

This work has been carried out as part of the CGIAR Research Program on GLDC. ICRISAT is a member of the CGIAR System Organization.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2020.559120/full#supplementary-material

REFERENCES

1. Ritchie H, Reay DS, Higgins P. Quantifying, projecting, and addressing India's hidden hunger. *Front Sustain Food Syst*. (2018) 2:11. doi: 10.3389/fsufs.2018.00011
2. Upadhyaya HD, Bajaj D, Narnoliya L, Das S, Kumar V, Gowda CLL, et al. Genome-wide scans for delineation of candidate genes regulating seed-protein content in chickpea. *Front Plant Sci*. (2016) 7:302. doi: 10.3389/fpls.2016.00302
3. Tulchinsky TH. Micronutrient deficiency conditions: global health issues. *Public Health Rev*. (2010) 32:243. doi: 10.1007/BF03391600
4. Bailey RL, West KP, Black RE. The epidemiology of global micronutrient deficiencies. *Ann Nutr Metab*. (2015) 66:2233. doi: 10.1159/000371618
5. Smith MR, Myers SS. Impact of anthropogenic CO_2 emissions on global human nutrition. *Nat Clim Chang*. (2018) 8:834–9. doi: 10.1038/s41558-018-0253-3
6. Jha AB, Warkentin TD. Biofortification of pulse crops: status and future perspectives. *Plant J*. (2020) 9:73. doi: 10.3390/plants9010073
7. Tiwari VK, Rawat N, Chhuneja P, Neelam K, Aggarwal R, Randhawa GS. Mapping of quantitative trait Loci for grain iron and zinc concentration in diploid A genome wheat. *J Hered*. (2009) 100:771–6. doi: 10.1093/jhered/esp030
8. Anuradha K, Agarwal S, Venkateswara RY, Rao KV, Viraktamath BC, Sarla N. Mapping QTLs and candidate genes for iron and zinc concentrations in unpolished rice of Madhukar × Swarna RILs. *Gene*. (2012) 508:233–40. doi: 10.1016/j.gene.2012.07.054
9. Jin T, Zhou J, Chen J, Zhu L, Zhao Y, Huang Y. The genetic architecture of zinc and iron content in maize grains as revealed by QTL mapping and meta-analysis. *Breed Sci*. (2013) 63:317–24. doi: 10.1270/jsbbs.63.317
10. Calayugan MIC, Formantes AK, Amparado A, Descalsota-Empleo GI, Nha CT, Inabangan-Asilo MA, et al. Genetic analysis of agronomic traits and grain iron and zinc concentrations in a doubled haploid population of rice (*Oryza sativa* L.). *Sci Rep*. (2020) 10:2283. doi: 10.1038/s41598-020-59184-z
11. Iqbal A, Khalil IA, Ateeq N, Khan MS. Nutritional quality of important food legumes. *Food Chem*. (2006) 97:331–5. doi: 10.1016/j.foodchem.2005.05.011
12. Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, et al. Draft genome sequence of chickpea (*Cicer arietinum* L.). *provides a resource for trait improvement. Nat Biotechnol*. (2013) 31:240–6. doi: 10.1038/nbt.2491
13. Yadav RDS, Srivastava JP. DUS characteristics of chickpea varieties. *Seed Tech News*. (2002) 32:29–30.
14. Jukanti AK, Gaur PM, Gowda CLL, Chibbar RN. Nutritional quality and health benefits of chickpea (*Cicer arietinum* L.): a review. *Br J Nutr*. (2012) 108:S11–26. doi: 10.1017/S0007114512000797
15. Thudi M, Khan AW, Kumar V, Gaur PM, Katta K, Garg V, et al. Whole genome re-sequencing reveals genome wide variations among parental lines of mapping populations in chickpea (*Cicer arietinum* L.). *BMC Plant Biol*. (2016) 16:10. doi: 10.1186/s12870-015-0690-3
16. Thudi M, Chitikineni A, Liu X, He W, Roorkiwal M, Yang W, et al. Recent breeding programs enhanced genetic diversity in both desi and kabuli varieties of chickpea (*Cicer arietinum* L.). *Sci Rep*. (2016) 6:38636. doi: 10.1038/srep38636
17. Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, Yang W, et al. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat Genet*. (2019) 51:857–64. doi: 10.1038/s41588-019-0401-3
18. Roorkiwal M, Bharadwaj C, Barmukh R, Dixit GP, Thudi M, Gaur PM, et al. Integrating genomics for chickpea improvement: achievements and opportunities. *Theor Appl Genet*. (2020) 133:1703–20. doi: 10.1007/s00122-020-03584-2
19. Varshney RK, Mohan SM, Gaur PM, Chamarthi SK, Singh VK, Srinivasan S, et al. Marker-assisted backcrossing to introgress resistance to *Fusarium* wilt (FW) race 1 and *Ascochyta* blight (AB) in C 214, an elite cultivar of chickpea. *Plant Genome*. (2014) 7:1–11. doi: 10.3835/plantgenome2013.10.0035
20. Pratap A, Chaturvedi SK, Tomar R, Rajan N, Malviya N, Thudi M, et al. Marker-assisted introgression of resistance to fusarium wilt race 2 in Pusa 256, an elite cultivar of desi chickpea. *Mol Genet Genomics*. (2017) 292:1237–45. doi: 10.1007/s00438-017-1343-z
21. Mannur DM, Babbar A, Thudi M, Sabbavarapu MM, Roorkiwal M, Sharanabasappa BY, et al. Super Annigeri 1 and improved JG 74: two Fusarium wilt-resistant introgression lines developed using marker-assisted backcrossing approach in chickpea (*Cicer arietinum* L.). *Mol Plant Breed*. (2019) 39:2. doi: 10.1007/s11032-018-0908-9
22. Saxena AK, Sharma S, Sandhu JS, Gupta SK, Kaur J. Variability and correlation studies for nutritional and cooking quality traits in chickpea (*Cicer arietinum* L.). *Indian J Genet*. (2013) 73:51–6. doi: 10.5958/j.0019-5200.73.1.007
23. Serrano C, Carbas B, Castanho A, Soares A, Patto MCV, Brites C. Characterization of nutritional quality traits of a chickpea (*Cicer arietinum* L.). germplasm collection exploited in chickpea breeding in Europe. *Crop Pasture Sci*. (2017) 68:1031–40. doi: 10.1071/CP17129
24. Cobos MJ, Izquierdo I, Sanz MA, Tomás A, Gil J, Flores F, et al. Genotype and environment effects on sensory, nutritional, and physical traits in chickpea (*Cicer arietinum* L.). *Span J Agric Res*. (2016) 14:e0709. doi: 10.5424/sjar/2016144-8719
25. Jadhav AA, Rayate SJ, Mhase LB, Thudi M, Chitikineni A, Harer PN, et al. Marker-trait association study for protein content

in chickpea (*Cicer arietinum* L.). *J Genet.* (2015) 94:279–86. doi: 10.1007/s12041-015-0529-6
26. Diapari M, Sindhu A, Warkentin TD, Bett K, Tar'an B. Genetic diversity and association mapping of iron and zinc concentrations in chickpea (*Cicer arietinum* L.). *Genome Res.* (2014) 57:459–68. doi: 10.1139/gen-2014-0108
27. Ray H, Bett KE, Tar'an B, Vandenberg A, Thavarajah D, Warkentin T. Mineral micronutrient content of cultivars of field pea, chickpea, common bean, and lentil grown in Saskatchewan, Canada. *Crop Sci.* (2014) 54:1698–708. doi: 10.2135/cropsci2013.08.0568
28. Vandemark GJ, Grusak MA, McGee RJ. Mineral concentrations of chickpea and lentil cultivars and breeding lines grown in the U.S. Pacific Northwest. *Crop J.* (2018) 6:253–62. doi: 10.1016/j.cj.2017.12.003
29. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE.* (2011) 6:e19379. doi: 10.1371/journal.pone.0019379
30. Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE.* (2012) 7:e32253. doi: 10.1371/journal.pone.0032253
31. Huang YF, Poland JA, Wight CP, Jackson EW, Tinker NA. Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. *PLoS ONE.* (2014) 9:e102448. doi: 10.1371/journal.pone.0102448
32. Bajaj D, Das S, Badoni S, Kumar V, Singh M, Bansal KC. Genome-wide high-throughput SNP discovery and genotyping for understanding natural (functional) allelic diversity and domestication patterns in wild chickpea. *Sci Rep.* (2015) 5:12468. doi: 10.1038/srep12468
33. Jaganathan D, Thudi M, Kale S, Azam S, Roorkiwal M, Gaur P, et al. Genotyping-by-sequencing based intra-specific genetic map refines a "QTL-hotspot" region for drought tolerance in chickpea. *Mol Gen Genomics.* (2015) 290:559–71. doi: 10.1007/s00438-014-0932-3
34. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE.* (2014) 9:e90346. doi: 10.1371/journal.pone.0090346
35. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TE, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples, *J Bioinform.* (2007) 23:2633–5. doi: 10.1093/bioinformatics/btm308
36. Varshney RK, Gaur PM, Chamarthi SK, Krishnamurthy L, Tripathi S, Kashiwagi J, et al. Fast-Track introgression of "QTL-hotspot" for root traits and other drought tolerance traits in JG 11, an elite and leading variety of chickpea. *Plant Genome.* (2013) 6:1–9. doi: 10.3835/plantgenome2013.07.0022
37. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* (2012) 9:357–9. doi: 10.1038/nmeth.1923
38. SAS Institute Inc. *SAS/STAT® 15.1 User's Guide.* Cary, NC: SAS Institute Inc. (2018).
39. Van Ooijen W, Jansen, J. *JoinMap 4.1 Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Cambridge: Kyazma ® (2013).
40. Kosambi DD. The estimation of map distances from recombination values. *Ann Hum Genet.* (1943) 12:172–5. doi: 10.1111/j.1469-1809.1943.tb02321.x
41. Lisa AO, Robert WR, Steven GB, Cory RB. Linkage map view-rendering high-resolution linkage and QTL maps. *J Bioinform.* (2018) 34:306–7. doi: 10.1093/bioinformatics/btx576
42. Wang S, Basten CJ, Zeng ZB. *Windows QTL Cartographer 2.5*. Department of Statistics, North Carolina State University, Raleigh, NC (2012).
43. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *J Bioinform.* (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033
44. Varshney RK, Thudi M, Nayak SN, Gaur PM, Kashiwagi J, Krishnamurthy L, et al. Genetic dissection of drought tolerance in chickpea (Cicer arietinum L.). *Theor Appl Genet.* (2014) 127:445–62. doi: 10.1007/s00122-013-2230-6
45. Jayalakshmi V, Reddy AT, Nagamadhur KV. Genetic diversity and variability for protein and micro nutrients in advance breeding lines and chickpea varieties grown in Andhra Pradesh. *Legum Res.* (2018) 42:768–72. doi: 10.18805/LR-3933
46. Verma S, Gupta S, Bandhiwal N, Kumar T, Bharadwaj C, Bhatia S, et al. High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.). using Genotyping-by-Sequencing (GBS). *Sci Rep.* (2015) 5:17512. doi: 10.1038/srep17512
47. Deokar A, Sagi M, Tar'an, B. Genome-wide SNP discovery for development of high-density genetic map and QTL mapping of ascochyta blight resistance in chickpea (*Cicer arietinum* L.). *Theor Appl Genet.* (2019) 132:1861. doi: 10.1007/s00122-019-03322-3
48. Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK, et al. Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol J.* (2012) 10:716–32. doi: 10.1111/j.1467-7652.2012.00710.x
49. Deokar AA, Ramsay L, Sharpe AG, Diapari M, Sindhu A, Bett K, et al. Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. *BMC Genomics.* (2014) 15:708. doi: 10.1186/1471-2164-15-708
50. Chen L. High-resolution QTL mapping for grain appearance traits and co-localization of chalkiness-associated differentially expressed candidate genes in rice. *Rice Sci.* (2016) 9:48. doi: 10.1186/s12284-016-0121-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sab, Lokesha, Mannur, Somasekhar, Jadhav, Mallikarjuna, C, Yeri, Valluri, Bajaj, Chitikineni, Vemula, Rathore, Varshney, Shankergoud and Thudi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Nutrition

REVIEW
published: 26 November 2020
doi: 10.3389/fnut.2020.533453

Enhancing the Nutritional Quality of Major Food Crops Through Conventional and Genomics-Assisted Breeding

Kiran B. Gaikwad[1], Sushma Rani[2], Manjeet Kumar[1], Vikas Gupta[3], Prashanth H. Babu[1], Naresh Kumar Bainsla[1] and Rajbir Yadav[1]*

[1] Division of Genetics, Indian Council of Agricultural Research (ICAR)-Indian Agricultural Research Institute, New Delhi, India, [2] Indian Council of Agricultural Research (ICAR)-National Institute for Plant Biotechnology, New Delhi, India, [3] Division of Genetics, Indian Council of Agricultural Research (ICAR)-Indian Institute of Wheat and Barley Research, Karnal, India

OPEN ACCESS

Edited by:
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

Reviewed by:
Chao-Qiang Lai,
Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, United States
Sachin Kumar,
Chaudhary Charan Singh University, India

***Correspondence:**
Kiran B. Gaikwad
kiran.gaikwad@icar.gov.in;
gaikwadkb@gmail.com

Specialty section:
This article was submitted to Nutrigenomics, a section of the journal Frontiers in Nutrition

Received: 08 February 2020
Accepted: 03 September 2020
Published: 26 November 2020

Citation:
Gaikwad KB, Rani S, Kumar M, Gupta V, Babu PH, Bainsla NK and Yadav R (2020) Enhancing the Nutritional Quality of Major Food Crops Through Conventional and Genomics-Assisted Breeding. Front. Nutr. 7:533453. doi: 10.3389/fnut.2020.533453

Nutritional stress is making over two billion world population malnourished. Either our commercially cultivated varieties of cereals, pulses, and oilseed crops are deficient in essential nutrients or the soils in which these crops grow are becoming devoid of minerals. Unfortunately, our major food crops are poor sources of micronutrients required for normal human growth. To overcome the problem of nutritional deficiency, greater emphasis should be laid on the identification of genes/quantitative trait loci (QTLs) pertaining to essential nutrients and their successful deployment in elite breeding lines through marker-assisted breeding. The manuscript deals with information on identified QTLs for protein content, vitamins, macronutrients, micro-nutrients, minerals, oil content, and essential amino acids in major food crops. These QTLs can be utilized in the development of nutrient-rich crop varieties. Genome editing technologies that can rapidly modify genomes in a precise way and will directly enrich the nutritional status of elite varieties could hold a bright future to address the challenge of malnutrition.

Keywords: nutritional quality, genomics assisted breeding, QTLs, food crops, malnutrition, genome editing, CRISPR/Cas9, plant breeding

INTRODUCTION

Over two billion of the world's population is at the risk of micronutrient deficiency which resulted due to an inadequate supply of micronutrients in daily diet (1). The principal reason behind this nutrient stress is our food crops which are inadequate to supply essential nutrients as they are grown on the soils which are deficient in minerals. The agricultural production scenario in most developing countries does not address the issues related to malnutrition; rather, it focuses on increasing grain yield and crop productivity. Malnutrition is also a vital threat to the population's health and productivity and is reviewed as the largest adverse health impact of climate change. In the last few decades, the climate had altered a lot since the temperature is rising in many parts of the world, coupled with low or unexpectedly high rainfall. According to the World Health Organization and the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, malnutrition is considered as one of the five largest adverse health impacts of climate change (2).

The situation of continuous degrading natural resources, farm resource constraints, and agriculture affected by changing climate in the developing countries is making a rapid rise in micronutrient deficiency in food grains, thereby increasing micronutrient malnutrition among

the population, but thanks to the innovative breeding efforts by plant breeders, the agriculture in developing countries is shifting toward producing nutrient-rich food crops. This will eliminate micronutrient malnutrition in these countries, where daily diets are dominated by micronutrient-poor staple food crops. Biofortification is a strategy that can overcome nutrient deficiency more sustainably. It is a one-time investment and offers a cost-effective, long-term, and sustainable approach in fighting hidden hunger because, once the biofortified crops are developed, there are no additional costs of buying the fortificants and adding them to the food supply during processing (3). Biofortification is a process of enriching the content of vitamins and minerals in crops through genetic and agronomic interventions. Biofortification not only makes our food nutrient-rich but also reduces the cost of external amendments required to make food rich in nutrients. Approaches such as conventional plant breeding, molecular breeding, transgenic techniques, or agronomical practices provide a new avenue for the development of nutrient-rich crops. Biofortification is mainly focused on important diet-dominant foods such as cereal crops, *viz*., rice, wheat, maize, sorghum, millet, and legumes, and starchy food, *viz*., potato and sweet potatoes.

We, humans, require around 40 known nutrients in optimum amounts to live a healthy life. These nutrients are classified as macro nutrients, micronutrients, and minerals. The micronutrients, *viz*., iron, zinc, copper, manganese, iodine, selenium, molybdenum, cobalt, nickel, and vitamin A, are extremely important for a wide range of metabolic processes which ultimately lead to normal growth and development. The mineral elements such as sodium, magnesium, calcium, potassium, phosphorous, chlorine, and sulfur are classified as essential nutrients that are required in trace amounts in the body. Overall, these nutrients play a pivotal role in our physical and mental development (4). Apart from micronutrients and minerals, protein and oil content are the chief nutritional factors in food grains. Therefore, it is foremost essence to incorporate essential micronutrients, minerals, and proteins into the diets of resource-poor populations whose diet is based on cereals such as rice, wheat, cassava, and maize which contain insufficient amounts of several nutrients.

The adverse effect of climate change on nutritional food security will be primarily seen in the developing countries of Africa and South Asia. Climate change will affect not only the number of food crops but also nutritional quality. Overall, hundreds of millions of people are expected to be placed at risk of zinc, iron, and/or protein deficiencies as a result of rising CO_2 concentrations. Recent studies proved that elevated CO_2 concentration in the atmosphere reduces the concentrations of iron, zinc, and proteins in staple cereals, *viz*., rice, wheat, barley, and pulses. Nowadays, CO_2 concentrations of 550 ppm, which is higher than normal, can lead to 3–11% decreases of zinc and iron concentrations in cereal grains and legumes (5). Further increase in CO_2 concentration by 690 ppm will lead to 5–10% reductions in the concentration of iron, zinc, potassium, calcium, phosphorus, sulfur, magnesium, copper, and manganese in the majority of crops (6). The decline in zinc content due to the continuously increasing level of CO_2 is expected to place 150–200 million people at risk of zinc deficiency (7). Mitigating the adverse effects of climate change on the nutritional quality of food crops can be achieved through breeding biofortified food crops.

Breeding biofortified crops require the identification of genetic resources with high micronutrient and mineral content from available germplasm. Most of the crops' wild relatives, landraces, and local cultivars are a rich source of these nutrients, which provide their effective utilization in the breeding program. The HarvestPlus biofortification program was started by International Food Policy Research Institute and International Center for Tropical Agriculture in collaboration with the Consultative Group on International Agricultural Research (CGIAR) centers which focused on making crops rich in vitamin A, iron, and zinc. The target crops were beans and pearl millet for iron, maize, cassava, and sweet potato for vitamin A, and wheat, rice, and maize for zinc content. The sole aim of this project is to develop micronutrient-rich crops through breeding and delivery of these nutrient-rich crops in the developing world which are mainly affected by micronutrient malnutrition. The quality traits are polygenic and are quantitatively controlled; therefore, it is difficult to improve these traits by conventional breeding. The rapid development of advanced genomic tools like molecular markers provides an effective means for improving the efficiency of plant breeding in transferring these quantitatively inherited traits. The adoption of molecular markers can remarkably facilitate the breeding programs by identifying the exact location of the genomic region/QTL controlling the trait for nutrient content. The identified QTLs can then be easily transferred to elite breeding materials if these lines have a low level of nutrient content. The usefulness of these markers based on QTL mapping was not that significant because these are based on bi-parental mapping populations. However, genome-wide association mapping studies (GWAS) offer unique opportunities to use diverse germplasm which had accumulated a large number of meiotic events as opposed to only one or a few meiotic recombinations in bi-parental mapping populations. The ability to resolve marker/trait associations depends upon the extent of linkage disequilibrium present in the panel. The markers identified to be linked with QTL have the potential to be used across breeding material for identification and introgression. In this review, we aimed to compile information on QTLs pertaining to nutrient content in important food crops, which once biofortified, may ensure good nutrition and eliminate hunger among populations, especially food-insecure children who are majorly impacted by climate-related devastations.

QTLs for Pro-vitamin A

Vitamin A deficiency is one of the serious health problems in developing countries, leading to irreversible blindness. This situation encouraged the researchers to attempt the biofortification of plant-derived foods including pro-vitamin A carotenoids (mainly β-carotene), which led to "golden" crops. The conversion rates of β-carotene into vitamin A are reported to be high in golden rice, maize, and cassava, demonstrating that these staple crops may incorporate a higher level of nutritional

impact (8, 9). Some QTLs such as *crtRB13′TE*, *crtRB1-5′TE-2*, *crtRB1-3′TE-1*, *crtRB1-5′TE-2*, *crtRB1-3′TE-1*, *LCYE*, *PSY1*, and *crtRb1* for pro-vitamin A have been identified on chromosome 10 in maize (10–13). In addition to this, QTL for β-carotene content has been reported in recombinant inbred line (RIL) populations through SSR markers in maize (14) (**Table 1**).

QTLs for Fe and Zn Content

Iron (Fe) and zinc (Zn) are among the essential micronutrients that are often lacking in human diet (4). According to WHO, over 30% of the population is anemic. These micronutrients play a wide range of metabolic functions required for normal growth and brain development. Zinc is an essential nutrient in the proper functioning of the immune system. Infants, small children, and pregnant and lactating women are at a higher risk of Fe and Zn depletion. Therefore, it is important to raise the level of Fe and Zn in the daily diet of children and women to save them from malnutrition. Cereals biofortified with Fe and Zn can eliminate the problem of malnutrition. These traits (Zn and Fe) are polygenic in nature, and genetic variation for these traits exists in landraces and progenitor species of important food crops. Breeding strategies are therefore focusing on novel approaches to broaden the genetic base using wild/related species and landraces and identifying genetic control and their effects (107). Several genetic mapping populations have been developed in the past few years to dissect Fe- and Zn-related traits. In wheat, various studies have reported QTL for high grain Fe and Zn concentrations on chromosomes 1A, 1D, 1B, 2A, 2B, 3B, 3D, 4B, 5A, 6A, 6B, 7A, 7B, and 7D in hexaploid wheat (16, 18, 20–24) and on chromosome 2 in barley (25) (**Table 1**). Recently, Velu et al. (19) evaluated the HarvestPlus Association Mapping panel across a range of environments in India and Mexico. GWAS analysis revealed two larger QTL regions on chromosomes 2 and 7 for large grain Zn content.

In rice, QTLs (*qFe 1.2*, *qFe 2*, and *qFe 6*) were identified on chromosomes 1, 2, and 6 for Fe content (37, 38, 41) and on chromosomes 5, 8, and 7 (*qZn5, qZn 8*) for Zn content (37, 41, 42). Most of the QTLs for Fe and Zn content were identified through single-nucleotide polymorphism (SNP) and DArT-SSR markers in RIL and BIL populations, but some of them were also from DH lines (**Table 1**). Recently, many new QTLs for Fe and Zn content have been mapped mainly through SSR markers in DH and BIL mapping populations (26, 27, 29, 34–36) (**Table 1**). Maize hybrids and varieties having a high yield potential along with 25–30% more Fe and Zn than common cultivars have been developed as part of the HarvestPlus program (108). Pearl millet is an important nutri-cereal rich in Fe and Zn. Recently, Kumar et al. (46) and Anuradha et al. (47) have identified QTLs for grain Fe and Zn content in RILs and 130 pearl millet germplasm lines (**Table 1**).

Pulses have a uniformly higher amount of Fe and Zn compared to the cereals and are a better source of micronutrients. Izquierdo et al. (48), using a meta-QTL analysis in 7 populations, identified three major QTL regions governing seed Fe and Zn content and concentration in seeds. Many other QTLs for Fe and Zn content (**Table 1**) have also been reported in common bean using SSR markers (49, 51), in lentil using SNP markers (54, 55, 58), and in peas using SSR (60) and SNP markers (59). In chickpea, Upadhyaya et al. (62) and Diapari et al. (64) identified QTLs for Fe and Zn content using SNP markers in germplasm lines (**Table 1**). Similarly in mungbean, two QTLs, *viz*., *qFe4.1* and *qZn6.4* were mapped on chromosomes 4 and 6, respectively, using SSR markers in a RIL population (68). In soybean, Ning et al. (77) reported two QTLs, *qZn11.1* and *qFe3*, on chromosomes 11 and 3, respectively, for seed Zn and Fe content. In lentils, the diversity of genes relating to iron and zinc contents in seeds was studied using molecular markers, with implications for biofortification (109–111).

QTLs for Oil Content

Oil content is a classic but significant yield trait in groundnut, peanut, soybean, and various *Brassica* species. The fatty acid composition of various edible oil crops has various parameters such as nutrition, oxidative stability, and shelf-life that correlate with overall oil quality. So, modifying the oil content of the majority of vegetable oil crops is one of the breeding objectives, while breeding vital, healthy, and desired fatty acids should be envisaged for improving oil quality in the specific genotypes. Generally, a higher content of oleic acid and a lower content of linoleic acid (high oleic/linoleic acid ratio) in cooking/edible oil is reported to be healthier for human consumption. Many QTLs (**Table 1**) have been reported for oil content in diploid and amphiploid species of *Brassica* such as *Brassica napus* (87, 89–91, 94, 99) and *Brassica carinata* (102). In *Brassica juncea*, QTLs for glucosinolate, gluconapin, sinigrin, glucoiberin, gluconeobrasscin, gluconasturtin, and tocopherol content were mapped in RILs on chromosomes (J18, J1, J4, J9, J7, and J6, respectively) using SSR markers (103). Researchers have developed "canola-grade" mustard varieties containing low levels of erucic acid and glucosinolates. The erucic acid content in *Brassica*s is controlled by the *FAE1* gene. CAPS markers for genes *FAE1.1* and *FAE1.2* were developed and successfully used in breeding programs in the improvement of *B. juncea* (104) and *B. rapa* (105), but when compared to simple PCR-based markers, CAPS markers are labor- and cost-intensive, which restrict their routine use by breeders. Recently, Saini et al. (112) assessed polymorphism in the upstream region of FAE1.1 and FAE1.2 genes, across the low erucic acid and high erucic acid genotypes, and developed PCR-based markers (FAE1.1P and FAE1.2P) based on this variability. These markers can be effectively used in marker-assisted selection for the development of low-erucic-acid varieties in B. juncea. In *B. Oleracea*, gene *FAD3-2* for low α-linolenic acid has been tagged by a gene-specific marker (106) (**Table 1**). Peanut and soybean are extensively cultivated in many parts of the world and are an important source of edible oil. Researchers are identifying QTLs for the development of peanut genotypes with good oil quality and desired fatty acid composition. Huang and co-workers (65) reported QTL *qOCb3* for oil content in the F_2 population, whereas Pandey and co-workers (67) reported a meta-QTL (*mqOC*) for oil content in the RIL population (**Table 1**). In soybean, seed oil content is under polygenic control and is sensitive to environmental effects. QTLs for oil content flanked

by SSR marker (76, 78) and SNP markers (83) have been reported in the RIL population.

QTLs for Fatty Acid Content

The trait associated with oil quality, measured in terms of its fatty acid composition, is an important agronomic trait that can eventually be tracked using molecular markers. Oils provide essential fatty acids along with an array of vitamins. Researchers have identified QTLs for major saturated (palmitic acid, arachidic acid, stearic acid, behenic acid, *etc.*) and unsaturated fatty acids (oleic acid, linoleic acid, linolenic acid, erucic acid, etc.) in major oilseed crops. The QTLs for major fatty acids will be discussed below.

Palmitic acid (C16:0) is the most common saturated fatty acid found in animals, plants, and microorganisms and is also the precursor to longer fatty acids (113). Many QTLs have been reported in major oilseed crops such as peanut (65, 66), soybean (83), and *B. napus* (87, 89, 95, 96, 100) for this trait. The details of the marker linked to palmitic acid content are given in **Table 1**.

Stearic acid (C18:0) is a saturated fatty acid, and its ester is one of the most common saturated fatty acids found in nature following palmitic acid. Many QTLs controlling stearic acid content have been reported in several oil crops. In soybean, a major QTL (*qSTEL001*) has been reported using SNP markers on chromosome 16 in RIL population (83). In *B. napus*, a QTL (qC18:0-C3e) was reported on chromosomes C8 and C3 using DH lines (89). In peanuts, QTLs such as *ASAA4*, *S_mqSAa07_1*, and *T_mqSA_b04_1* have been reported in F_2 and RILs on chromosomes A4, A7, and A9 (65, 66).

Arachidic acid (C20:0) is a saturated fatty acid present in peanut seeds. Several QTLs have been identified in soybean, peanut, and *Brassica* species pertaining to it. In *B. napus*, QTL qC18:3-C3c was present on chromosomes C3 and C8 in DH lines (87). QTLs such as *qAAB9*, *S_mqAA_b098*, and *T_mqAA_b04_1* in peanut were identified on chromosome B9 using F2 and RILs population (65, 66). The details of the marker linked to arachidic acid content are given in **Table 1**.

Oleic acid (C18:1) is a naturally occurring fatty acid among various animal and vegetable oils. Oleic acid accounts for about 80% of peanut oil. Norden et al. (114) have identified two high oleic acid peanut lines (F435-2-1 and F435-2-2) through a systematic breeding program in Florida. In the same year, genetics of this trait came before the world. Moore and Knauft (115) identified two homozygous recessive mutant genes FAD2A and FAD2B for high oleic acid content, but this was at the cost of lower linoleic acid. As time progressed, many reports have come out on mapping of this trait. In peanut, QTLs for high oleic acid content have been reported by many researchers (**Table 1**). Through the DArTseq marker, a QTL (*OLE-WH14*) was reported on chromosome B2 in inbred lines and another QTL (*qOLE.C4-1*) was also reported on chromosome C4 in DH lines in *B. carinata* (102). Similarly, in *B. Napus*, QTLs (*BnaA05g 23520*, *BnaA05g 23930*, *Gene BrFAD5*, and *Gene BrFAD7*) were detected for oleic acid using DH lines as mapping populations (**Table 1**). In soybean, a QTL (*qOLE003*) was reported using SNPs in the RIL population. In peanut, a QTL (*qOAA3*) was identified in F_2 and another QTL (*mqQA181*) in RILs population on chromosomes A3 and B9 (**Table 1**).

Linoleic acid (C18:2) is about 40% in normal peanut lines. Pandey and co-workers (67) reported the relationship of FAD2 genes with peanut oil quality and suggested that FAD2B contributed higher phenotypic variance for linoleic acid than the FAD2A alleles. Recently, Hu et al. (116) detected the two main QTLs for linoleic acid content located in linkage groups A09 and B09 in peanut. They reported that SNP markers, *viz.*, Marker2575339 and Marker2379598 in B09, were associated with oleic acid and linoleic acid in seven environments and Marker-4391589 and Marker-4463600 in A09 in six environments. The QTLs for linoleic acid were identified in other crops, too, *viz.*, soybean (*qLINN007*), *B. carinata* (LEN-WH14, *qLEI.B8-4*), and *B. napus* (*uqFA-C3-3*, *uqFA-A8-4*, *qC182_N9*, *FAD7*, *FAD3A*, and *FAD3C* genes) using RILs, DHs, and inbred lines (**Table 1**).

QTLs for Protein and Amino Acid Content

Pulses are important sources of proteins, important nutrients, and calories, particularly in diets of peoples living in developing countries. Paying less attention to the genetic enhancement of pulses is likely to have a noticeable impact on global food and nutritional security (117). Advances in sequencing technologies have made significant improvements in the breeding lines of important pulses such as pigeonpea, chickpea, and groundnut; however, the pace in improvement does not match that of cereals. Importantly, the consumption of pulses has been shown to have several positive health effects that are inherently tied to their nutritional qualities (118). The development of improved cultivars of pulses can be accelerated through the identification and the deployment of the gene(s)/QTLs of nutritional quality traits. In chickpea germplasm lines (Desi and Kabuli), two significant QTLs for protein content linked with SSR markers TR26.205 and CaM1068.195, present on LG3 and LG5, were reported (63). In soybean, major QTLs *(qProt_Gm20*, *qPRO_13_1*, *qPRO001*, and *qPC)* for protein content flanked by SNP and SSR markers were identified in RILs (66, 73, 79, 83). In *B. carinata*, QTLs for protein content were identified (*qPRO10* and PRO-WH13) through DArTseq marker using DH and inbred lines (102) (**Table 1**). Among cereals, rice and wheat are also sources of protein in daily diets, though in limited quantity. The major effect of QTLs for grain protein content has been mapped in rice (28, 33, 39, 43) and Wheat (15, 17, 22) (**Table 1**).

Protein is composed of amino acids; these amino acids are the "building blocks" of the body, and improving essential amino acids is one of the better prospects envisaged. Some QTLs corresponding to essential amino acids have been identified in important food crops. Amino acid contents such as those of lysine, threonine, methionine, and cysteine were improved in soybean through the identification of QTLs in different RIL populations (70, 71, 73, 80). In *B. napus*, QTLs were identified for lysine, threonine, and methionine content (qLysC-16-3, qThrC-12-5, and qMetC-9-5) in DH lines on chromosomes C6, C2, and A9 (93). In *B. napus*, arginine, histidine, glutamic acid, glycine, alanine, and aspartic acid content was improved through QTLs such as *qArgC-8-5*, *qHisC-4-3*, *qGluC-1-1*, *qGlyC-4-1*, *qProC-1-1*, *qAlaC-7-3*, and *qAspC-9-4* in BC lines on chromosomes A1, A4,

TABLE 1 | List of important QTLs pertaining to nutritional quality in major food crops.

Crop	QTL /Loci	MP	Cross(s)	Markers	Chromosome	Nutritional parameters in the grain	References
Wheat	*QGpc-1B-2* *QGpc-4B-1.4*	RILs	Tainong 18/Linmai 6	D-1190331S-3222160 (SNP) D-1380792D-1094306(SNP)	1B 4B	Grain protein content	(15)
	QGZn.co-5A *QGZn.co-7A* *QGFe.co-3B.1* *QGFe.co-5A.2*	RILs	Roelfs F2007/Hong Hua Mai/./Blouk #1	1244217-1272027|F|0 (DArT) 5356706-5325178|F|0 (DArT) 1089107-1127875|F|0 (DArT) 1102433-988523 (DArT)	5A 7A 3B 5A	Zn Fe	(16)
	QGpc.uhw-4B *QGpc.uhw-5A.1* *QGpc.uhw-6B* *QGpc.uhw-7B.2*	RILs	Svevo/Y12-3	TG0010b(SNP) RAC875_rep_c106118-339(SNP) T.durum contig 9860-281(SNP) Kukri_c14766-484(SNP)	4B 5A 6B 7B	Grain protein content	(17)
	qFes1 *qfes2* *qZns1* *qZns2*	Inbreds	Synthetic hexaploid wheat	Xwmc399 Xgwm157 Xwmc357 Xcfd63	4D 2D 5D 1D	Fe Zn	(18)
	QZn 2A *QZn7B*	Inbreds	HPAM panel genotypes	RAC875_c34757_180 (SNP) wsnp_Ex_c5268_9320618 (SNP)	2A 7B	Zn Zn	(19)
	QGzn.ada_1B *QGzn.sar_1B* *QGFe.ada_2B*	RILs	Adana/70711 Saricanak 98/MM S4 Adana/70711	rpt-6561 (DArT) wPt- 6434-wPt1403 (DArT) wPt-9812 (DArT)	1B 1B 2B	Zn Zn Fe	(20)
	QZn.Across_4BS *QFe.Across_7DS*	RILs	Seri M 82/SHW Cw176364	TP91631-TP81797 (SNP) TP43715-TP37547 (SNP)	4BS 7DS	Zn Fe	(21)
	QZn.bhu_2B, QFe.bhu_2B *QGPC.bhu_1A*	DHs	Berkut/Krichaff	gwm120-wPt2430 (SSR-DArT) wPt 9592–GBM 1153 (DArT-SSR)	2B 1A	Zn and Fe Protein content	(22)
	QZn.bhu_2B *QZn.bhu_6A* *QFe.bhu_3B*	RILs	PI 348449 (*T. spelta*)/HUW 234	989092F0 (SNP) 998265F0 (DArT) 3022954F0 (SNP)	2B 6A 3B	Zn Zn Fe	(23)
	QGzncpk.cimmyt_2BL	RILs	PBW 343/Kenya Swara	wPt-6174 (DArT)	2BL	Zn	(24)
Barley	*QTL.Zn*	RILs	Clipper/Sahara	Xbcd175-Xpsr108 Vrs1-XksuF15	2HS 2HL	Zn	(25)
Rice	*QTL.Fe.9* *QTL.Zn.4* *QTL.pro.1*	Inbreds	Landraces	RM215 (SSR) RM551 (SSR) RM5 (SSR)	9 4 1	Fe Zn Protein content	(26)
	qFe3.3 *qFe7.3* *qZn2.2* *qZn8.3*	Inbreds	Germplasm lines	RM7 (SSR) RM1132 (SSR) RM300 (SSR) RM80 (SSR)	3 7 2 8	Fe Zn	(27)
	qAAC6.1 *qAAC7.1* *qPC1.2*	RILs	Milyang23/Tong88-7	3515361-3658340(SNP) 4856196-5206110(SNP) 39162572-39234399 (SNP)	6 7 1	Amino acid content Protein content	(28)

(Continued)

TABLE 1 | Continued

Crop	QTL /Loci	MP	Cross(s)	Markers	Chromosome	Nutritional parameters in the grain	References
	qFe1.1	BILs	RP-Bio226/Sampada	RM562-RM11943 (SSR)	1	Fe	(29)
	qFe1.2, qZn1.1			RM294A-RM12276 (SSR)	1	Fe and Zn	
	qFe6.1, qZn6.1			RM8226-RM400 (SSR)	6	Fe and Zn	
	qFe6.2, qZn6.2			RM400-RM162 (SSR)	6	Fe and Zn	
	qAA1	RILs	Zhenshan 97/Delong 208	RM493–RM562	1	Amino acid content	(30)
	qAA7			MRG186–MRG4499	7		
	qPC6.2	RILs	Yukihikari/Joiku 462	Indel 207-Indel 208	6	Protein content	(31)
	qB4.1	DH	R64/IR69428 (Pop1) and BR29/IR75862 (Pop2)	4333768-4349423(SNP)	4	Boron	(32)
	qCo7.1			7872824-7892971(SNP)	7	Cobalt	
	qCu8.1			89266819-8940497(SNP)	8	Copper	
	qK6.1			5883472-5895767(SNP)	6	Potassium	
	qMn3.1			3005168-2732340(SNP)	3	Manganese	
	qMg1.1			115078-117345(SNP)	1	Magnesium	
	qZn11.1			10907196-11001107(SNP)	11	Zinc	
	qP11.2			11391672-11407547(SNP)	11	Phosphorous	
	qGPC1.1	BILs	ARC10075/Naveen	93237905-93229368(SNP)	1	Grain protein content	(33)
	qSGPC2.1			93256429-93260438(SNP)	2		
	qSGPC7.1			93225742-93256949(SNP)	7		
	qFe7	DH	Goami2/Hwaseonchal	RM10-RM1973 (SSR)	7	Fe	(34)
	qZn7			RM10-RM1973 (SSR)		Zn	
	qFe1.2	BILs	Swarna/*O. nivara* (IRGC81832) P1 Swarna/*O. nivara* (IRGC81848) P2	RM243-RM81A (SSR)	1	Fe	(35)
	qFe11.1			RM332-RM287 (SSR)	11	Fe	
	qZn2.1			RM250-RM535 (SSR)	2	Zn	
	qZn3.2			RM55-RM520 (SSR)	3	Zn	
	qFe3.2			RM520-RM514 (SSR)	3	Fe	
	qFe4.1			RM241-RM348 (SSR)	4	Fe	
	qZn5.1			RM153-RM413 (SSR)	5	Zn	
	qZn12.1			RM415-RM19 (SSR)	12	Zn	
	qFe10.1	MAGIC	MAGIC Population	S10_13883426	10	Fe	(36)
	qZn6.2			S6_2939487	6	Zn	
	qZn7.1			S6_29350591	7	Zn	
	qFe 6 and	BILs	IR 7586/Ce 258 IR 7586/Zhongguangxiang1	RM3- RM340 (SSR)	6	Fe	(37)
	qZn 8			RM407- RM152 (SSR)	8	Zn	
	qFe 1.2 (gene OsYSL1)	RILs	Madhukar/Swarna	RM490 (SSR)	1	Fe	(38)
	qFe 5.1 (gene OsZIP6)			RM574 (SSR)	5		
	qFe 7.2 (gene OsZIP8)			RM8007 (SSR)	7		
	qPro-2	DHs	Cheongcheong/Nagdong	RM5619-RM1211 (SSR)	2	Protein content C	(39)
	qPro-10			RM24934-RM25128 (SSR)	10	Protein content LC	
	qLip-6			RM586-RM1163 (SSR)	6		
	qPro-2	DHs	Cheongcheong/Nagdong	RM12532-RM555	2	PC	(40)
	qFe 2, qZn 5, qCo 1, qCu 2, qMn 7, and *qMo 7*	BILs	Teqing/Lemont	RM452, RM421, RM490, RM6378, RM214, and RM11 (SSR)	2, 5, 1, 2, 7and 7	Fe, Zn, Co, Cu, Mn, Mo	(41)

(Continued)

TABLE 1 | Continued

Crop	QTL /Loci	MP	Cross(s)	Markers	Chromosome	Nutritional parameters in the grain	References
	OsZIP8a *OsNAC* *OsZIP4b*	RILs	IRRI 38/Jeerigesanna	RM8007 (SSR) - -	7 3 8	Zn	(42)
	qPC 1	RILs	Zhenshan97/Nanyangzhan	RM472- RM104	1	Protein content	(43)
Maize	*crtRB13'TE*	Inbred	1.V335/HP465-30 2.V345/HP465-35	crtRB1-3'TE-1 (SSR)	10	Pro-vitamin A	(12)
	crtRB1-5'TE-2 *crtRB1-3'TE-1*	Inbreds	Hp321-1	crtRB1-5'TE-2 (SSR) crtRB1-3'TE-1 (SSR)	10	Pro-vitamin A	(13)
	LcyE *crtRB1-3'TE*	F_2 $F_{3:4}$	L1-L5 H1-H15	LcyE-5'TE (SSR) LcyE-3'TE (SSR) crtRB1-3'TE (SSR)	10	Pro-vitamin A	(10)
	LCYE *PSY1* *crtRb1*	Inbreds	130 Tropical Inbred lines	LCYE-5'TE (285 Indel) LCYE-SNP (G-C SNP) LCYE-3'TE (8 bp Indel) PSY-SNP7 (A-C SNP) PSY1-1D1 (378 Indel) crtRB1-5'TE (397/206 bp Indel) crtRB1-InDel4 (12 bp Indel) crtRB1-3'TE (325/1250 bp Indel)		Pro-vitamin A	(11)
	crtRB1	RILs	1.B73/BY804 2.A619/SC55 3.K13/B77 4.K13/SC55	umc1506-bning1028 umc1506-crtRB1 crtRB1 crtRB1	10	β-carotene, βC/βCX, βC/Z and βC/ALL	(14)
	sQTL4.2 *sQTL4.1* *sQTl3.1*	RILs	Ye478/Wu312	umc1620-umc1994 (SSR) umc1346-bnlg2291 (SSR) mmc0132-umc1504 (SSR)	4 4 3	Zn Mn Mg	(44)
	qZn5,qMn1, qCo3, qCu8, qK4, qMo1, qNa5, qP4, qS1	RILs	B73/IL14H	PZA02411.3, PZA02135.2, PZA01925.1, PZA03698.1, PZA01751.1, PZA02269.3.4, PZA01327.1, PZA01751.2, PZA02698.3 (All SNP)	5, 1, 3, 5, 8, 4, 1, 5, 4, 1	Zn, Mn, Co, Cu, K, Mo, Na, P, S	(45)
Pearl millet	*qFe1/54 and qZn1/54*	RIL	ICMS 8511-S1-17-2-1-1-B-P03 _ AIMP 92901-S1-183-2-2-B-08	Pgpb10531-pgpb9130 (DArT)	LG1	Fe and Zn	(46)
	Fe and Zn	Inbreds	130 germplasm lines	Xpsmp2261, Xipes0180, Xipes0096		Fe and Zn	(47)
Common Bean	*MQTL_Fe,Zn 1.1* *MQTL_Fe,Zn 6.1* *MQTL_Fe,Zn 6.1*	RILs	7 Populations		1 6 6	*Fe and Zn*	(48)
	Seed coat Fe *Seed coat Zn*	BILs	Cerinza/G10022	Pv-at03, ATA16, ATA26 ATA247, ATA10	4, 2, 3 8, 6	Seed coat Fe Seed coat Zn	(49)
	Ca1, Ca7, Ca9 *Mg7* *Pt5, Pt7* *DF6, DF7*	RILs	Xana/Cornell 49242	McatEtc, P gene, McagEac P gene $OD12^{800}$, SAS8 DBD (AC), P gene	1, 7, 9 7 5, 7 6, 7	Ca Mg Protein content Dietary fiber	(50)
	Fe_cont8.1 *Zn_cont2.1, 5.1, 5.2 and 7.1*	BILs	Cerinza/G10022	BMc316 (SSR) PV109, BM155, BMd28 and PV35 (SSR)	8 3, 5, 5, 7	Fe Zn	(51)

(Continued)

TABLE 1 | Continued

Crop	QTL /Loci	MP	Cross(s)	Markers	Chromosome	Nutritional parameters in the grain	References
	Fe-AAS2a *Fe-AAS6c* *Zn-AAS2c* *Zn-AAS7c*	RILs	G21242/G21078	E0403A (SSR) N0401B (SSR) Pv11 (SSR) BM239 (SSR)	2 6 2 7	Fe concentration Zn concentration	(52)
	MnQTL1.1 *MnQTL3.2* *MnQTL3.3*	RILs	CDC Redberry/ILL7502	SNP	LG LG3 LG3	Mn content	(53)
Lentil	*FeQTL1.2, 2.2, 4.2, 5.2, 6.2, and 7.1*	RILs	ILL8006/CDC Milestone	15SNP, 81 SNP, 40 SNP, 239 SNP, 4 SNP and 12 SNP	1, 2, 4, 5, 6, 7	Fe concentration	(54)
	Zn content	Inbreds	143 diverse germplasm lines	LcC06739p564(SNP) LcC04105p1090 (SNP)	3 2	Zn content	(55)
	SeQTL2.1, 5.1, 5.2, 5.3,	RILs	PI320937/Eston	SNP	2,5,5,5	Selenium content	(56)
	qFe uptake	RILs	ILL8006-BM/CDC Milestone	–	–	Fe uptake	(57)
	121 QTLs for Mn and Zn concentartion	RILs	CDC-Redberry/ILL7602	–	–	*Mn and Zn concentartion*	(58)
Pea	*[B]-Ps5.1, [Ca]-Ps5.1, [Mg]-Ps5.1, [S]-Ps5.1* *[Fe]-Ps7.1, [Zn]-Ps7.1* *[K]-Ps5.1* *[Mn]-Ps5.1* *[Mo]-Ps5.1* *[P]-Ps3.1*	RILs	Aragorn/Kiflica	TP61763 (SSR) TP44143 (SSR) TP55189 (SSR) tip_SNP2_V (SNP) TP42330 (SSR) TP75231 (SSR)	5 7 5 5 5 3	B, Ca, Mg S Fe, Zn K Mn Mo P	(59)
	QTL.Fe1 *QTL.Fe2* *QTL.Fe3*	RILs	Carrera/CDC Striker	Sc1203_101100 and PsC17710p220 Sc9618_162688 and PsC4833p179 Sc2559_48386 and PsC908p622 (All SNP)	3 4 7	Fe concentration	(60)
	QTL FEBIO	RILs	1-2347-144/CDC Meadow	PA-P (SNP)	5	Fe bioavailability	(61)
Chickpea	*CaqFe1.1* *CaqZn2.1* *CaqFZ4.1* *CaqFZ5.1* *CaqFZ7.1*	Inbreds	92 Germplasm lines	SNP53-SNP55 SNP110 SNP300 SNP413 SNP471-SNP472	1 2 4 5 7	Fe Zn Fe and Zn Fe and Zn Fe and Zn	(62)
	QTL.PC	Inbreds	187 Germplasm lines	TR26.205 (SNP), CaM1068.195 (SNP)	3, 5	Protein content	(63)
	QTL.Zn *QTL.Fe*	Inbreds	94 Germplasm lines	Cav1Sc25.1p2052607 Cav1Sc19.1p1556596	4 1	Zn Fe	(64)
Peanut	*qOCB3* *qPAA8* *qOAA3* *qEAA5* *qBAB9* *qAAB9* *ASAA4*	F_2	Zhonghua10/ICG12625	AHGS1788-pPGSseq14C11 PM54-pPGPSseq2G3 TC4E10-ARS744 GM1577-TC6E1 AHGS1969 AHGS1969 GM2480	B3 A8 A3 A5 B9 B9 A4	Oil content Palmatic acid Oleic acid Eicosenic acid Behenic acid Arachidic acid Stearic acid	(65)

(Continued)

TABLE 1 | Continued

Crop	QTL /Loci	MP	Cross(s)	Markers	Chromosome	Nutritional parameters in the grain	References
	S_mqPAb09_4, *S_mqAA_b098* *S_mqGA_b09_1* *S_mqSAa07_1* *S_mqBA_a09* *S_meLA_b08_2*	RILs	Sunoleic97R/NC94022 (Population S)	FAD2B RN34A10 FAD2A Seq2A06	B9 A7 A9 B8	Palmatic acid, Arachidic acid, Gadoleic acid Stearic acid Behenic acid Lignoceric acid	(66)
	T_mqSA_b04_1, *T_mqAA)b04_1,* *T_mqLA_b04_1* *T_mqGA_b04_1, T_mqBA_b04_4* *T_mqPA_a09_4*	RILs	Tifrunner/GTC20 (Population T)	PM15 TC4H07 FAD2A	B9 B4 A9	Stearic acid Arachidic acid, Lignoceric acid Gadoleic acid, Behenic acid Palmatic acid	(66)
	mqQA181, *mqLA182, mqOLR* *mqOC*	RILs	Sunoleic97R/NC94022 (S population)	ahFAD2B GM1878	B9 A5	Oleic acid, Linoleic acid, Oleic/Linoleic acid Oil content	(67)
	mqQA181, *mqLA182, mqOLR* *mqOC*	RILs	Tifrunner/GT-C20 (T population)	ahFAD2A GM2690-1	B9 B8	Oleic acid, Linoleic acid, Oleic/Linoleic acid Oil content	(67)
Mung bean	*qFe4.1, qZn6.4,*	RILs	ML776/Sattya	PVBR82 (SSR), CEDG248 (SSR)	4,6	Fe, Zn	(68)
	SDPAP4.1 *SDIP10.1* *SDTP4.1*	F_2	V1725BG/Aus TRCF 321925	CEDG139-MB-SSR179 (SSR) VR226-CEDG097 (SSR) Bmd25-MB-SSR179 (SSR)	4 10 4	Low phytic acid Inorganic phosphate Total phosphorus	(69)
Soybean	*qCys-7-2* *qMet-8-1* *qSAA-15-1* *qPC-8-1*	RILs	Kefeng no. 1 and Nannong 1138-2	Bin 148 (SNP) Bin 34 (SNP) Bin 124 (SNP) Bin 37 (SNP)	7 8 15 8	Cysteine Methionine Sulfur containing amino acids Protein content	(70)
	qPRO001 qOIL001 qPAL002 qOLE003 qLIN001 qGLU001 qALA001 qCYS001 qVAL001 qHIS001	RILs	Hamilton/Spencer	ss249909538–ss249919445 ss246100375–ss245879277 ss245914593–ss245908292 ss249909538–ss249506152 ss245914593–ss245790648 ss245914593–ss245908292 ss246100375–ss245879277 ss245898080–ss245908292 ss249909538–ss249919445 ss245898080–ss245908292	18/G 6/C2 6/C2 18/G 6/C2 6/C2 6/C2 6/C2 18/G 6/C2	Protein Oil Palmitic Oleic Linolenic Glutamic Alanine Cysteine Valine Histidine	(71)
	qIF5-1	RILs	Huachun 2/Wayro	Bin 799-800 (SNP)	5	Isoflavone content	(72)
	qProt_Gm20 and *qLsy_Gm20* *qThr_Gm20, qMet_Gm20 and* *qMet+Cys_Gm20* *qCys_Gm10*	RILs	Benning/Danbaekkong	GSM0012 (SNP)-satt354 (SSR) GSM0012-BARC-020713 (SNP) satt592 (SSR)-BARC043247 (SNP)	20 20 10	Protein content and Lysine content Threonine, Methionine and Methionine + Cysteine content Cysteine content	(73)

(Continued)

TABLE 1 | Continued

Crop	QTL /Loci	MP	Cross(s)	Markers	Chromosome	Nutritional parameters in the grain	References
	suc1, suc3, suc2	F_3	MFS-553/PI243545	ss245668753, ss249186914, ss246796276 (SNP)	5, 9, 16	Sucrose content	(74)
	qPC	RILs	1.R05-1415/R05-638 2.V97-1346/R05-4256	satt451-satt614 (SSR) ss250447161-ss250327854 (SNP)	20 20	Protein content	(75)
	qOC	RILs	R05-1415/R05-638	satt451-satt614 (SSR)	20	Oil content	(76)
	qCa-8-1 qMg-13 qZn-11-1 qFe-3 qP-11	RILs	Kefeng1/Nannong1138-2	sat_162-AW132402 satt335-satt522 satt251 satt675-satt237 satt197	8 13 11 3 11	Calcium content Magnesium Zn Fe Phosphorous	(77)
	qPRO_B1 *qOIL_C1,J,O*	RILs	SD_{02}-4-59/A_{02}-381100	BARCSoYSSR-17-0621 (SSR) Sat_140, Sat_350, satt581 (SSR)	17 4,16,10	Protein content Oil content	(78)
	qPRO_13_1, *qPRO_13_1*	$F_{2:3}$	1. Jidou 12/ZYD 2738 (*G. soja*) 2. Jidou 9/ZYD 2738 (*G. soja*)	satt114 (SSR) satt114	13	Protein content	(79)
	qCys and *qCys-Met* *qMet*	RILs	1.Williams 82 /DSR-173 2. Williams 82/NKS19-90 3. Williams 82/Vinton81	BARC-038869-07364 - BARC-039753-07565 (SNP) BARC-018461-02916 – BARC-066103-17539 (SNP)	20	Cysteine and Cystein + Methionine content Methionine content	(80)
	qIF20-2	RILs	Luheidou2/Nanhuizao	M943408-M941848 (SLAF)	20	Isoflavone content	(81)
	Ca5	$F_{2:4}$ $F_{2:5}$	PI 408052B/PI 408052C PI408052B/KS 43035P	Sat_290- satt115	18	Calcium content	(82)
	qPRO001 *qOIL008* *qPAL002* *qSTEL001* *qOLE003* *qLINL003* *qLINN007*	RILs	MD96-5722/Spencer	ss248293401-ss248275088 (SNP) ss248308943-ss248309108 (SNP) ss249629157-ss249621644 (SNP) ss248977568-ss248979552 (SNP) ss249037210-ss249039670 (SNP) ss248981433-ss248993887 (SNP) ss249010538-ss249039670 (SNP)	14 14 18 16 16 16 16	Protein content Oil content Palmatic acid Stearic acid Oleic acid Linoleic acid Linolenic acid	(83)
	qPRO7-5 *qPRO20-1*	RILs	Charleston/Dongnong 594	satt358-Sat_001 (SSR) satt331-satt173 (SSR)	7 20	Protein content	(84)
Brassica napus	*qoil.1* *qpro-1* *qoil+pro-1*	DH	SGDH14/cv. Express	pP12638473-p12699181 (SNP) p18005556-p11609327 (SNP) p12699181-scaff_17119_1_p115218	A08 A07 A08	Oil content Protein content Oil and Protein content	(85)
	BnaA05g 23520 *BnaA05g 23930*	DHs	RIL324/RIL622	UQSNP0001565 UQSNP0001759		Oleic and Linoleic acid	(86)
	cqOC-A8-2 *uqFA-C3-3* *uqFA-A8-4*	DH	Tapidor × Ningyou 7		A8 C3 A8	Oil content C16:0/C18:0/C18:1/C18:2/C20:0/C20:1/C22:0/C22:1/C16:0/C18:0/C18:1/C18:2/C18:3/C20:0/C20:1/C22:0/C22:1/	(87)

(Continued)

TABLE 1 | Continued

Crop	QTL /Loci	MP	Cross(s)	Markers	Chromosome	Nutritional parameters in the grain	References
	Gene BrFAD5 *Gene BrFAD7*	DHs	YS 143 /Nai Bai Cai	Bra027203 Bra034863	A05 A05	Erucic acid Oleic and Linoleic acid	(88)
	qOIL-A10a qC16:0-C8b qC18:0-C3e, qC18:1-C3c, qC18:2-C3c, qC18:3-C3c, qC20:0-C3d	DHs	Polo 9/Topas	BnGMS288-311/CB10536-163 em1/bg9-434/bg23/pm59-285 pm88/pm45-177/odd3/pm3-399	A10 C8 C3	Oil content Palmatic acid Stearic, Oleic, Linoleic, Linolenic, Arachidic acid	(89)
	Oil content *Protein content* *Erucic acid* and *Stearic acid* *Linoleic acid* *Glucosinolate*	DHs	Tapidor/Ningyou7	Bn-scaff_23761_1-p249628 (SNP) Bn-scaff_17119_1-p349622 (SNP) Bn-scaff_15794_1-p347392 (SNP) Bn-A02-p10850012 (SNP) Bn-scaff_15794_1-p437864 (SNP)	C03 C03 C03 A02 C03	Oil content Protein content Erucic and Stearic acid Linoleic acid Glucosinolate	(90)
	OilC3-3	DHs	Tapidor/Ningyou7	Bn-scaff_23761_1-p249628 (SNP)	C03	Oil content	(91)
	PRT.C6.w.1 SUL.A2.w.1 OLA.A9.w.1 OIL.C3.w.1 LIA.C3.w.1 GSL.A2.w.1 ERA.C3.w.1	Inbreds	405 Inbred lines	Bn-ctg7180014756759-p1575 Bn-ctg7180014748062-p8451 Bn-Scaffold000110-p349432 Bn-ctg7180014717095-p1564 Bn-ctg7180014726380-p989 Bn-ctg7180014748062-p8451 Bn-ctg7180014717095-p1564 (SNP markers)	C6 A2 A9 C3 C3 A2 C3	Protein content Sulphur content Oleic acid Oil content Linolenic acid Glucosinolate Erucic acid	(92)
	qLysC-16-3 qThrC-12-5 *qMetC-9-5*	BC	BC_1F_1 (DHs/Tapidor) BC_2F_1 (DHs/Ningyou7)	HBR057/HBR047 EM18ME6-220/NA12C03HBR096/IGF5385F	C6 C2 A9	Lysine content Threonine content Methionine content	(93)
	qOC-2 *qPC-1* *qOAC-2* *qGLC-1*	RILs	827R/Darmor_Sin	CB10369-220 - me5em16-170 me1em1-400 - me4em7-400 Ol09-A06-400 - me2em10-240 me7em11-470 me8em20-230 (SRAP markers)	11 1 12 10	Oil content Protein content Oleic acid Glucosinolate	(94)
	qA8-5 *qC3-3*	DHs	Tapidor/Ningyou7	IGF1108c-sR7178 IGF0235b-BRMS-093	A8 C3	C16:0/C18:0/C18:1/ C18:2/C18:3/C20:0/ C20:1/C22:0/C22:1	(95)
	qC16:0, qC18:1, qC18:2, qC18:3, qC20:1	DHs	Tapidor/Ningyou7	HBr015 (A8) and JICB0633 (C3)	A8 and C3	C16:0, 18:1, C18:2, C18:3, C20:1	(96)
	FAD3A and*FAD3C* genes	Inbreds	21 Cultivars of winter and spring rape	LinAR-LinAF and LinCr-LinCF (dCAPS markers)		Low Linolenic acid	(97)
	BnFAD2-C5	Inbred	Xiangyou	YG-C5-FAD2-F/YG-C5-FAD2-R	C5	High oleic acid	(98)

(Continued)

TABLE 1 | Continued

Crop	QTL /Loci	MP	Cross(s)	Markers	Chromosome	Nutritional parameters in the grain	References
	qOil_N19	DHs	DH12075 /	-	N19	Oil content	(99)
	qC160180_N9		YN01-429	-	N9	C16:018:0	
	qC181_N9			-	N9	C18:1	
	qC182_N9			-	N9	C18:2	
	qC183_N6			-	N6	C18:3	
	qProtein_N16			-	N16	Protein content	
	qArgC-8-5	BC	BC1F1 1 (DHs/Tapidor)	*HAU348/B034P14-1-1*	*A8*	Arginine content	(100)
	qHisC-4-3		BC1F1 2 (DHs/'Ningyou7)	HS-k02-2/HBr094	A4	Histidine content	
	qGluC-1-1			znS13M26-100/CB10081	A1	Glutamic acid	
	qGlyC-4-1			HBr094/CNU256	A4	Glysine content	
	qProC-1-1			znS13M26-100/CB10081	A1	Protein content	
	qAlaC-7-3			znS13M26-340/JICB0571	A7	Alanine content	
	qAspC-9-4			HBr075/JICB0516	A9	Aspartic acid	
	BnaFAD2 gene	Inbred	Tapidor	BnaC.FAD2.a	A5	High Oleic and low PUFA	(101)
				BnaA.FAD2.b	A1		
Brassica carinata	*qPRO10*	DHs	Y-BcDH64/W-BcDH76	100059607 (DArTseq)	B7	Protein content	(102)
	qEru.C4-1			100036778 (DArTseq)	C4	Erucic acid	
	qLEN.C4-1			100065508 (DArTseq)	C4	Linolenic acid	
	qOLE.C4-1			100026342 (DArTseq)	C4	Oleic acid	
	qLEI.B8-4			100035893 (DArTseq)	B8	Linoleic acid	
	PRO-WH13	Inbreds	81 diverse accessions	5121285 (DArTseq)	C8	Protein content	(102)
	OLE-WH14			5859309S (DArTseq)	B2	Oleic acid	
	OC-WH13			5121285 (DArTseq)	C8	Oil content	
	LEN-WH14			5863483-1S (DArTseq)	B3	Linolenic acid	
	ERU-WH13			5121285 (DArTseq)	C8	Erucic acid	
Brassica juncea	TGLC-S	RILs	NUDH-YJ-04/RL-1359.	SB3739a -SB3739b (SSR)	J18	Glucosinolate	(103)
	GNA-S			NA14B05- NIA138 (SSR)	J1	Gluconapin	
	SIN-S			CNU566- NA12D08 (SSR)	J4	Sinigrin	
	IBE S			NIA010 -NIA046 (SSR)	J9	Glucoiberin	
	NEO-S			NIA045 -NIA043 (SSR)	J7	Gluconeobrasscin	
	NAS-S			NA12H09- NA12G08 (SSR)	J4	Gluconasturtin	
	TOC			CNU111 -CNU483 (SSR)	J6	Tocopherol	
	OIL			NIA044- CNU402b (SSR)	J9	Oil content	
	FAE1.1	Inbreds and	18 Inbred lines PM24/Pusa Vijay PM30/Pusa	CAPS591, CAPS1265	A8	Low erucic acid	(104)
	FAE1.2	BC	Bold	CAPS237	B7		
Brassica rapa	*BnFAE1 gene*	BC	Tori-7 / Kirariboshi	BnFAE1.1-dcapsF	A	Low erucic acid	(105)
				BnFAE1.1-3UTR-cR	C		
	Gene BrFAD5	DHs	YS 143 /Nai Bai Cai	Bra027203	A05	Erucic acid	(88)
	Gene BrFAD7			Bra034863	A05	Oleic and linoleic acid	
Brassica oleracea	BoFAD3-2 gene		alboglabra	BoFAD3-2FY1 - BoFAD3-2FY2 (gene specific markers)		Low α-linolenic acid	(106)

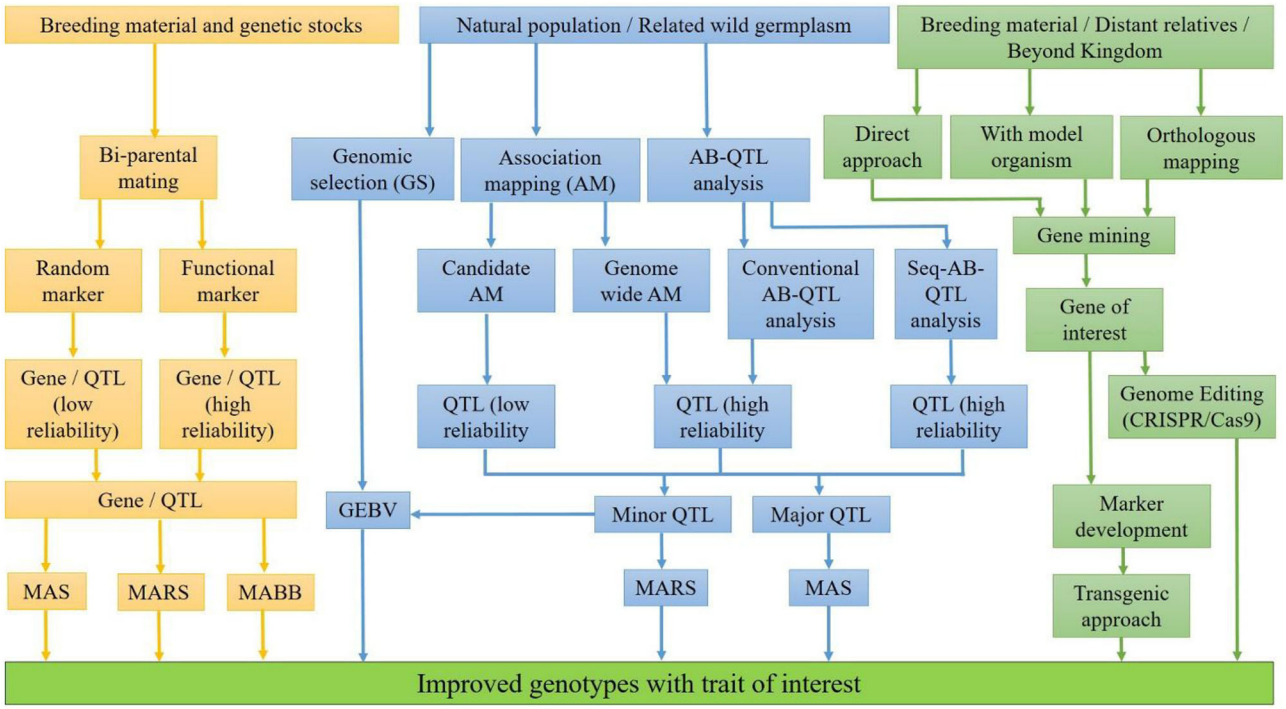

FIGURE 1 | Approaches for developing genotypes with enhanced nutritional traits.

A7, A8, and A9 (100). The details of the linked markers to these QTLs are given in **Table 1**.

QTLs for Minor Elements

Trace elements are required in a minimal amount, and these micronutrients are vital as they are often acting as catalysts in chemical reactions. In rice, QTLs, *viz*., *qCo 1*, *qCu 2*, *qMn 7*, and *qMo 7* in BILs for Co, Cu, Mn, and Mo on chromosomes 1, 2, 5, and 7, have been identified (41). Similarly, in maize, QTLs (*qMn1*, *qCo3*, *qCu8*, *qK4*, *qMo1*, *qNa5*, *qP4*, and *qS1*) for some trace elements (Mn, Co, Cu, K, Mo, Na, P, and S) have been identified on chromosomes 1, 3, 4, 5, and 8, respectively, by Baxter et al. (45). Furthermore, Ates et al. (56) identified four QTLs for selenium content using SNP markers in lentil (**Table 1**). QTLs [(K)—*Ps5.1*, (Mn)—*Ps5.1*, (Mo)—*Ps5.1*, and (P)—*Ps3.1*] for minor nutrients were identified through SSR markers in pea RILs (59). Sompong et al. (33) reported QTLs for low phytic acid (*SDPAP4.1*), inorganic phosphate (*SDIP10.1*), and total phosphorus (*SDTP4.1*) using SSR markers in mungbean. The details of the linked makers to these QTLs and QTLs linked to other minor elements are given in **Table 1**.

Improving the nutritional status of elite breeding lines requires functional information about the genetic network controlling important traits such as Fe, Zn, and protein content, oil content, pro-vitamin A, an array of amino acids, etc. Although impressive progress has been made in the identification of a large number of QTLs of nutritional traits in major food staples, an important challenge is their introgression into the agronomically superior background without compromising the yield and their popularization in malnourished areas.

GENOMIC APPROACHES FOR ENHANCING NUTRITIONAL QUALITY IN MAJOR FOOD CROPS

For the development of nutritionally rich cultivars, a suitable breeding strategy, available genetic diversity, and modern genomics information are needed. Sufficient genetic variability for nutritional quality traits has been explored in cultivars, direct progenitors, and wild relatives over the years. Moreover, a big treasurer of genetic diversity across the kingdoms can be explored and utilized through genetic engineering and genome editing. This diversity should be converted into nutritionally rich cultivars. Therefore, proper understanding of the genetic basis of nutritional quality traits and interaction with the environment is of utmost importance for an efficient breeding program. The integration of modern genomics, physiology coupled with precise phenotyping and advance breeding methodologies, has effectively revealed the genes and the metabolic pathways of quality traits. The identification and the incorporation of nutritional quality traits is being facilitated by several approaches (**Figure 1**), each having its own advantages and disadvantages.

The identification of loci for trait of interest can be facilitated with QTL mapping by the population developed from bi-parental mating or the natural population having large variability for the trait of interest. In bi-parental QTL mapping approach, two parents having a difference for the target trait are crossed

to generate a mapping population of the mortal (F2, F3, etc.) or immortal (RILs, NILs, etc.) in nature. QTLs can also be identified with natural population with the help of association mapping (AM), *viz*., genome-wide AM, candidate-gene based AM, and bulk segregant analysis (BSA). In genome-wide AM, polymorphism of the whole genome is considered at once for mapping of locus, while in candidate-gene based AM, the polymorphism of a particular genomic region for mapping of locus is considered. In BSA, target genomic regions linked with trait of interest are identified with simultaneous transfer from wild source into recipient during backcross program. The reliability of the identified QTLs depends on the closeness of the marker with the linked locus/QTLs. QTLs identified with functional markers/genic markers are more reliable because these markers are from the transcribed region and/or part of the identified locus. Therefore, SNP genotyping coupled with precise phenotyping, bioinformatics, and modern statistical algorithm specific to available data has enabled the identification of genomic regions controlling our traits of interest. Once genomic regions/QTLs associated with nutritional traits are identified, the next step is to utilize them for the development of nutritionally rich cultivars. The decision to transfer the target nutrional traits with molecular breeding depends on the relative eliteness and adaptation of the donor parent, the genetic complexity of the trait, the relative cost of phenotypic assays, the trait-linked marker assays, and genome profiling (119). If a nutritional trait is controlled by few loci, in elite genetic background, marker-assisted selection (MAS) is suggested. However, if the donor parents are landraces or distant relatives, marker-assisted backcross breeding (MABB) would be suitable for the transfer of trait. When the nutritional trait is controlled by a large number of genes with low genetic variance, genomic selection would be the choice to transfer the target trait (119). Genomic selection (GS) in crop breeding is gaining prominence due to the current availability of low-cost high-density genotyping with high-density DNA array chips and genotyping by sequencing (120, 121). In GS, all marker effects across the genome are estimated together by fitting the phenotypic and the marker data of the base training population with an appropriate statistical model (120). Then, genomic estimated breeding values (GEBVs) of any unknown genotypes having only marker information are generated with these marker effects. GEBVs predict the breeding values of unobserved genotypes for selection in the testing population (122, 123). With marker-assisted selection, nutritional traits like QPM and pro-vitamin A, controlled by major genes in maize, have been transferred in elite hybrids and released for commercial cultivation in India from ICAR-IARI, Delhi, India. These hybrids, namely, "Pusa Vivek QPM-9 Improved" having high QPM and pro-vitamin A in 2017, "Pusa Vivek Hybrid-27 Improved" having high pro-vitamin A, and "Pusa HQPM5 Improved" and "Pusa HQPM7 Improved," both having high QPM and pro-vitamin A in 2019, have been released for commercial cultivation in India so far (119, 124). In bread wheat, some success stories, but not of commercial success, of the introgression of major gene Gpc-B1 (high protein content) with MAS have been reported by several researchers (125–129). In durum wheat, one successful example of a cultivar having high protein content using MAS, namely, "Desert King-High Protein," had been developed at the University of California wheat breeding program (130). However, little success has been achieved in the development of nutritionally rich commercial cultivars with MAS in other food crops. The simple reason is that most of these traits are controlled by a large number of genes having a little effect individually in the expression of these traits. In bread wheat, the major QTLs for Zn content, namely, QZn2A and QZn7B, were each having only 11.9% genetic variance (19); therefore, introgression of these QTLs with MAS into an elite background is very difficult. In addition, genotype by environment interaction increases the complexity many folds in the transfer of target traits. Therefore, for complex quantitative traits governed by many minor QTLs, genomic selection (GS) would be a better choice than MAS (131). As a matter of fact that the genes or major QTLs for nutritional qualities are known, there are large numbers of minor QTL effects on the expression of trait. To support this fact, the betaine aldehyde dehydrogenase (BADH 2) gene solely is not able to explain the variation of aroma in rice because background minor QTLs contribute to the overall fragrance in rice (132, 133). Therefore, integration of minor effect QTLs for nutritional traits can be effectively accomplished by employing the recurrent selection under breeding cycles under GS.

Genome Editing Technologies for Enhancing Nutritional Quality in Major Food Crops

Although conventional breeding is currently the most preferred and widely used approach for developing nutrient-enriched genotypes/varieties, it is labor-intensive and resource consuming, and it usually takes several years to develop a final product right from the screening of large germplasm lines, making crosses, and selection of desired recombinants that can finally turn into the commercial variety. Genetically modified (GM) crops that have beneficial traits are developed by the transfer of transgene(s) of known function into elite crop varieties. The transgenic approach can be an alternative for the development of biofortified crops when the genetic variation for a particular nutrient is limited or absent in the germplasm. Unlimited genetic variation across the boundary of species can be explored through this approach. However, their use is largely affected by unsubstantiated health and environmental safety concerns raised by non-governmental organizations and many developed countries. Government regulatory frameworks in many developed countries that aim to safeguard human and environmental biosafety have led to significant cost barriers to the rapid widespread adoption of new GM traits (134). As a result, the advantages of GM traits have been restricted to a small number of cultivated crops such as maize (high lysine), soybean (high oleic acid), potato (high amylopectin), cassava (high content of iron, β-carotene, and protein), and canola (phytate degradation) (3). Genome editing technologies facilitate efficient, precise, and targeted modifications at the genomic loci (135). The first-generation genome editing technologies

that use ZFNs (136) and TALENs (137) have been around for two decades. They are labor-intensive, time-consuming, and involve a complex procedure to gain target specificity. However, second-generation genome editing techniques like CRISPR/Cas9 system (138) require less time and cost and provide simplicity and ease of targeted gene editing. All of these technologies use typical sequence-specific nucleases that can be induced to recognize specific DNA sequences and to generate double-stranded breaks (DSBs). The DSBs are repaired by plants' endogenous repair mechanisms, *viz*., non-homologous end joining, which can lead to the insertion or deletion of nucleotides, thereby causing gene knockouts, or by homologous recombination, which can cause gene replacements and insertions (139). The risks involved in altering genomes through the use of genome editing technology are significantly lower than those associated with GM crops because most edits alter only a few nucleotides, producing changes that are not unlike those found throughout naturally occurring populations (140). Once the genomic editing agents have segregated out, there is no way to distinguish between a "naturally occurring" mutation and a gene edit. Thus, the introduction of genome editing into modern breeding programs should facilitate rapid and precise crop improvement. Genome editing techniques often produce gene knockout mutants, gene replacement, and insertion mutants, thus becoming a potent tool in the improvement of traits of nutritional quality in major food crops (**Table 2**).

The first-generation genome editing technique TALENs has been used for modifying and enhancing the nutritional profile of major food crops. Soybean oil contains high levels of polyunsaturated linoleic and linolenic acid, which contribute to oxidative instability. This problem is often addressed through partial hydrogenation. However, partial hydrogenation increases the levels of *trans*-fatty acids, which have been associated with cardiovascular disease. Altering the composition of soybean oil by increasing the level of oleic acid and decreasing the levels of linoleic and linolenic acids may help reduce the need for hydrogenation. Soybean lines with high oleic acid and low linoleic acid contents were generated by introducing mutations in the two fatty acid desaturase 2 genes (FAD2-1A and FAD2-1B) (149) and fatty acid desaturase 3A (*FAD3A*) gene by directly delivering TALENs into *fad2-1a fad2-1b* soybean plants (150) and FAD2-2 gene using CRISPR/Cas9 (151), thus improving the shelf-life and heat stability of soybean oil. The same gene ahFAD2 was mutated for the isolation of high oleic acid lines in peanut using CRISPR/Cas9 (153).

Rice is a staple food crop feeding more than half of the world population. High amylose content and resistant starch improves human health and reduces the risk of serious diseases including hypertension, diabetes, and colon cancer (158). The CRISPR/Cas9 technology has been successfully used to create high-amylose rice by targeting two rice branching enzymes, *SBEI* and *SBEIIb* (143). Rice bran oil (RBO) contains many valuable healthy constituents, including oleic acid. Fatty acid desaturase 2 (FAD2) catalyzes the conversion of oleic acid to linoleic acid in plants. To produce high oleic/low linoleic RBO, Abe et al. (145) disrupted the OsFAD2-1 gene by CRISPR/Cas9-mediated targeted mutagenesis and developed rice lines with high oleic acid (twice that of the wild type) and low linoleic acid content. Enhancing the level of β-carotene is an important target of biofortification in major food crops because it is a precursor of pro-vitamin A. Endo et al. (146) were successful in accumulating β-carotene in rice callus (up to $2.86 \pm 1.41\,\mu g/g$ fresh weight) by identifying the putative ortholog of the cauliflower or gene in rice, *Osor*, and modifying the Osor gene *via* genome editing using CRISPR/Cas9. The majority of the rice varieties grown over the world are having a white pericarp. However, rich diversity also exists for brown, red, or purple/black pericarp. The red pericarp contains high levels of proanthocyanidins and anthocyanins which have been recognized as health-promoting nutrients (159). The red coloration in the grains of wild rice is controlled by two complementary genes, *Rc* and *Rd* (160, 161). Recently, Zhu et al. (147) successfully converted three elite white-pericarp rice varieties into red ones having high proanthocyanidin and anthocyanin content without compromising the yield potential through CRISPR/Cas9-mediated functional recovery of the recessive rc allele. Aromatic rice is popular worldwide for the characteristic fragrance of its grains. More than 100 volatile compounds were detected in the flavor of cooked fragrant rice. The presence of a defective *badh2* allele encoding BADH 2 results in the synthesis of 2-acetyl-1-pyrroline, which is a major fragrance compound in aromatic rice. Shan et al. (144) reported the creation of fragrant rice from a non-fragrant variety via the targeted knockout of *OsBADH2* using the TALEN method.

Using CRISPR/Cas9, DuPont Pioneer, in 2016, developed the first waxy corn hybrid by knocking out the maize waxy gene *Wx1*, making more than 97% amylopectin and essentially eliminating amylose from the kernel (148). The variety with high amylopectin starch content has higher digestibility and many industrial applications. The release of commercial hybrids with this trait is planned for 2020 (148). The researchers from Swedish Agricultural University developed high-amylopectin potatoes by knocking out the granule-bound starch synthase gene using CRISPR/Cas9 (154). The high-amylopectin potato starch has uses in both food and technical applications.

In potato tubers, the low temperature during cold storage stimulates the accumulation of reducing sugars that influence the quality of the product. Upon high-temperature processing, these reducing sugars react with free amino acids, resulting in brown, bitter-tasting products and elevated levels of acrylamide—a potential carcinogen. To minimize the accumulation of reducing sugars, vacuolar invertase (VInv) gene was knocked out by the TALENs technique (155).

The gluten protein of wheat which is responsible for the unique viscoelastic properties of wheat-derived foods also triggers gluten sensitivity in susceptible individuals commonly known as celiac disease. Because of the complexity of the Gli-2 locus and the high copy number of a-gliadin genes, conventional breeding and mutagenesis have failed to develop low-immunogenic wheat varieties for patients with celiac disease. Low-gluten, transgene-free wheat lines with much reduced immunoreactivity, using CRISPR/Cas9, have been developed

TABLE 2 | Nutritional quality traits improved by genome editing technologies in major food crops.

Crop species	Gene editing technology	Target gene	Trager trait	References
Wheat	CRISPR/Cas9	α-gliadin gene family	Low gluten content	(141)
Wheat	CRISPR/Cas9	α- and γ-gliadin gene family	Low gluten content	(142)
Rice	CRISPR/Cas9	*SEBIIb*	High amylose content	(143)
Rice	TALENs	*OsBadh2*	Enhanced aroma	(144)
Rice	CRISPR/Cas9	*OsFAD2-1*	Oleic acid content	(145)
Rice	CRISPR/Cas9	*Osor*	Enhanced β-carotene content	(146)
Rice	CRISPR/Cas9	*Rc*	Proanthocyanidinis and anthocunins	(147)
Maize	CRISPR/Cas9	*Wx1*	Waxy corn	(148)
Soybean	TALENs	*FAD2-1A*, *FAD2-1B*	High oleic acid content	(149)
Soybean	TALENs	*FAD2-1A*, *FAD2-1B*	High oleic, low linoleic acid content	(150)
Soybean	CRISPR/Cas9	*FAD2-2*	High oleic, low linoleic acid content	(151)
Soybean	TALENs	*GmPDS11* and *GmPDS18*	Carotenoid biosynthesis	(152)
Peanut	CRISPR/Cas9	*ahFAD2*	High oleic acid content	(153)
Potato	CRISPR/Cas9	*Wx1*	High amylopectin	(154)
Potato	TALENs	*VInv*	Low reducing sugars	(155)
Rapeseed	CRISPR/Cas9	*FAD2*	High oleic acid content	(156)
Rapeseed	CRISPR/Cas9	*BnaFAD2*	High oleic acid content	(157)

(141, 142) and will serve as donors for introgressing low gluten trait into elite wheat varieties.

Modification of fatty acid composition is one of the primary objectives in the improvement of oilseed brassicas. The fatty acid desaturase 2 gene, *FAD2*, is a key gene that affects oleic, linoleic, and linolenic acids. The CRISPR/Cas9-mediated genome editing system has been applied for developing lines with high oleic acid (more than 80%) of rapeseed (*Brassica napus*) (156, 157) and in the emerging oilseed plant *Camelina sativa* (162, 163) by modifying the *FAD2* gene. In yet another study in *Camelina*, the content of very long fatty acids was reduced by knocking out the FAE1 gene (164).

Genome editing technologies, because of their efficiency, high specificity, and amenability to multiplexing, have increasingly become popular genomic tools for enhancing the nutritional value of our major food crops. The mutant line generated through CRISPR/Cas9 will provide new genetic diversity for the traits of agronomic and nutritional importance for breeding in an unprecedented way.

Developing Nutrient-Enriched Varieties Through Conventional and Genomics-Assisted Breeding

Developing nutrient-enriched varieties of food crops through conventional plant breeding is the most preferred approach. The improvement of a particular trait requires the availability of sufficient and useful genetic diversity. This genetic diversity may be present in cultivated genotypes or can be introgressed from wild/related/progenitor species or can be created through directed mutagenesis. Work on the development of nutrient-rich varieties of major food crops is ongoing in various public and private sector organizations with the support of important programs like HarvestPlus, Global Alliance for Improved Nutrition (GAIN), Project ENABLE (Expanding Nutrition Access by Building capacity, Linking Initiative and Enhancing policy), HEALTHGRAIN Forum, etc. The purpose of all these programs is to improve the nutritional status of the food crops and make them available for all the people, especially the most vulnerable ones. Conventional breeding supported by the above-mentioned programs has yielded several nutrient-enriched varieties in important food crops (**Table 3**).

Wheat

Wheat is a staple cereal crop worldwide and one of the most preferred food crops for biofortification. The major target in wheat biofortification is the improvement of iron (Fe) and zinc (Zn) content in the grain. Wheat is blessed with a wider genetic variation for these two traits contributed by wild/progenitor species. This rich genetic diversity introgressed from *Triticum durum*-, *Triticum spelta*-, and *Triticum dicoccum*-based synthetic hexaploid wheat (SHW) into cultivated bread wheat led to the development of four high-Zn varieties in India and Pakistan. In India, the variety "Zinc Shakti" (Chitra), with 14 ppm Zn, was developed through participatory variety selection and further registered by private seed companies and growers. This variety has profitable yield potential and matures nearly 2 weeks earlier than common wheat. Two more varieties, *viz*., "WB 02" and "HPBW 01," were developed in 2017 by ICAR-India Institute of Wheat and Barley Research (ICAR-IIWBR) and Punjab Agricultural University (PAU), Ludhiana, respectively, for cultivation in the North-Western Plain Zone (NWPZ) of India (166) (**Table 3**). Wheat variety HD 3171, developed by ICAR-Indian Agricultural Research Institute (ICAR-IARI), New Delhi, is shown to have high iron content (47.2 ppm) (167). In Pakistan, the variety Zncol 2016 with +9 ppm Zn has been released for commercial cultivation (166) (**Table 3**). Apart from SHW, variation for high zinc content is present in wheat germplasm and varieties. In 2018, Bangladesh has witnessed the release of the first biofortified blast-resistant variety "Bari Gom 33″ with +7

TABLE 3 | Nutrient-enriched important biofortified crops developed through conventional and genomics-assisted breeding.

Crop	Biofortified nutrient	Name of variety	Country	Developing institute	References/source
Wheat	Zinc and iron	WB 02	India	ICAR-IIWBR, India, and CIMMYT	(165)
		HPBW 01	India	PAU, India, and CIMMYT	(166)
	Iron	HD 3171	India	ICAR-IARI, India	(167)
	Zinc	Zinc Shakti	India	CIMMYT	(166)
	Zinc	Zncol 2016	Pakistan	CIMMYT	(166)
	Zinc	BARI Gom 33	Bangladesh	Bangladesh Wheat and Maize Research Institute, and CIMMYT	(166)
	Zinc	BHU 1 BHU 3 (Akshay) BHU 5 BHU 6	India	Banaras Hindu University, CIMMYT, and HarvestPlus	(168)
	Protein content	HD 3226	India	ICAR-IARI, India	(169)
		PBW 757	India	PAU, India	(170)
	Yellow pigment (carotene)	HI 8627 HI 8759 HI 8777	India	ICAR-IARI, India	(170)
	Anthocyanin (colored wheat)	Scorpion	Austria	Crop Research Institute, Prague	(171)
		PS Karkulka	Slovak Republic	National Agricultural and Food Center	(172)
		NABIMG 9 NABIMG 10 NABIMG 11	India	National Agri-Food Biotechnology Institute, New Delhi	(173)
		Indigo	Austria	BOKU-University of Natural Resources and Life Sciences, Austria	(174)
		Black-grained wheat	China	Shanxi University, China	(175)
Rice	Zinc	Jalmagna	India	Landrace collection	(176)
	Iron	IR 72/Zawa Bonday (IR 68144-313-2-2-3)	India, Philippines	International Rice Research Institute, Philippines	(176)
	Zinc	BRRIdhan 62 BRRIdhan 64 BRRIdhan 72	Bangladesh	Bangladesh Rice Research Institute (BARI) and HarvestPlus	(177)
	Zinc	CR Dhan 45	India	National Rice Research Institute, India	(178)
	Protein	CR Dhan 310	India	National Rice Research Institute, India	(178)
Maize: Quality Protein Maize (QPM)	Lysine and tryptophan	Protina (composite) Ratan (composite)	India	GBPUAT, Pantnagar	(179)
		Shakti (composite) Shakti1 Shaktiman 1 Shaktiman 2 Shaktiman 3 Shaktiman 4	India	Indian Institute of Maize Research (IIMR), India	(179)

(Continued)

TABLE 3 | Continued

Crop	Biofortified nutrient	Name of variety	Country	Developing institute	References/source
		HQPM 1 HQPM 5 HQPM 7	India	CCS Haryana Agricultural University (CCSHAU), Hisar, India	(179)
		Vivek QPM 9	India	ICA-VPKAS Almora, India	(179)
		Pusa HM 4 Improved[a] Pusa HM 8 improved[a] Pusa HM 9 Improved [a]	India	ICAR-IARI, India	(124)
		QPHM 200 and QPHM 300	Pakistan	National Agricultural Research System and CIMMYT	(180)
		BHQPY 545 BHQP 542 Melkassa 1Q Melkassa 6Q MHQ 138	Ethiopia	Bako Agricultural Research Center, Ethiopia, and CIMMYT	(181–183)
		GH-132-2	Ghana	Agricultural Research Centers Ethiopia and CIMMYT	
		BR-451 (OPV) BR-473 (OPV) Assum Preto (OPV)	Brazil	University of Ghana and CIMMYT	
		HB-PROTICTA	Guatemala	Guatemala's Institute for Agricultural Science and Technology and CIMMYT	
		Obatampa GH Mambia CMS 473 CMS 475 K9101	Guinea	Crop Research Institute, Kumasi and International Institute of Tropical Agriculture (IITA)	
		Obatampa GH	Benin		
		Susuma (OPV)	Mozambique	CIMMYT, Mexico	
		Obatampa Espoir	Burkina Faso	CIMMYT, Mexico	
		Obatampa	Cameroon	CIMMYT, Mexico	
		Obangaina (OPV)	Uganda	CIMMYT, Mexico	
		HQ INTA-993 NB-Nutrinta (OPV)	Nicaragua	CIMMYT, Mexico	
		EV 99 QPM	Nigeria	CIMMYT, Mexico	
		Ev 99 QPM	Togo	CIMMYT, Mexico	
		Lishe-K1	Tanzania	CIMMYT, Mexico	
		EV 99 QPM DMRESR WQPM Susuma	Senegal	CIMMYT, Mexico	
		HQ-31	Honduras	CIMMYT, Mexico	
		HQ-61	El Salvador	CIMMYT, Mexico	
		ICA	Colombia	CIMMYT, Mexico	

(Continued)

TABLE 3 | Continued

Crop	Biofortified nutrient	Name of variety	Country	Developing institute	References/source
		KH 500Q KH 631Q WSQ 104	Kenya	CIMMYT, Mexico	
		Lishe-H1, Lishe-H2	Tanzania	CIMMYT, Mexico	
		ZS 261Q	Zimbabwe	CIMMYT, Mexico	
		Quian 2609 Zhongdan 9409 Zhongdan 3850 Yun Yao 19 Yun You 167 Lu Dan 167 Lu Dan 206 Lu Dan 207 Lu Dan 807 Hybrid 2075	China	Guizhou, China CAAS, China CAAS, China Yunnan, China Yunnan, China Shandong, China Shandong, China Shandong, China Shandong, China Sichuan, China	
		H551C H553C H519C H368EC H369EC	Mexico	CIMMYT, Mexico	
		QS-7705	South Africa	CIMMYT, Mexico	
		FONAIAP	Venezuela	CIMMYT, Mexico	
		INIA	Peru	CIMMYT, Mexico	
		HQ-2000	Vietnam	NMRI, Vietnam	
	QPM+ pro-vitamin A	Pusa Vivek QPM 9 Improved[a]	India	ICAR-IARI, India	(12, 178)
		Pusa HQPM5 Improved[a] Pusa HQPM7 Improved [a]	India	ICAR-IARI, India	(178, 184)
Maize: orange maize	Pro-vitamin A	Pusa Vivek Hybrid 27 Improved[a]	India	ICAR-IARI, India	(12, 178)
		Sam Vita 4-A, Sam Vita 4-B, Muibaki 3, Muibaki 2, Muibaki 1, GV662	DR Congo	CIMMYT, IITA, and HarvestPlus	(185)
		Ahoɔ*dzin*, *Dzifoo*, *Ahoɔfɛ*, CSIR-CRI Honampa, CSIR-CRI Odomfo, CSIR-CRI Owanwa	Ghana	CIMMYT, IITA, and HarvestPlus	(185)
		GV671A (HPH1301), GV673A (HPH1303), GV665A (HP1005), GV662A (HP1002), GV664A (HP1004)	Zambia	CIMMYT, IITA, and HarvestPlus	(186)
		Ife maizehyb-3 Ife maizehyb-4 Sammaz 38 (OPV) Sammaz 39 (OPV)	Nigeria	CIMMYT, IITA, and HarvestPlus	(185)
		Nafama, Abebe, Duba, Kodialan, Dakan	Mali	CIMMYT, IITA, and HarvestPlus	(185)
		HPH1317, HP1005	Tanzania	CIMMYT, IITA, and HarvestPlus	(185)

(Continued)

TABLE 3 | Continued

Crop	Biofortified nutrient	Name of variety	Country	Developing institute	References/source
		ZS242 (HP1005), ZS244 (HPH1301), ZS246 (HPH1302), ZS248 (HPH1303)	Zimbabwe	CIMMYT, IITA, and HarvestPlus	(185)
		MH39A, MH40A, MH42A, MH43A	Malawi	CIMMYT, IITA, and HarvestPlus	(185)
Maize: zinc maize	Zinc	BIO-MZn01	Colombia	CIMMYT and HarvestPlus	(187)
Pearl millet	Iron and zinc	Dhanshakti Hybrid ICMH 1201	India	ICRISAT and HarvestPlus	(188)
	Iron and zinc	HHB 299	India	CCSHAU, Hisar and ICRISAT	(178)
	Iron and zinc	HHB 311	India	CCSHAU, Hisar and ICRISAT	(188)
	Iron	AHB 1200	India	Vasantrao Naik Marathwada Krishi Vidyapeeth, Parbhani and ICRISAT	(178)
	Iron and zinc	RHB 233	India	SKN Agricultural University, Rajasthan and ICRISAT	(188)
Sorghum	Iron and zinc	ICSR 1401 (Parbhani Shakti)	India	VNMKV Agricultural University Parbhani and ICRISAT	(189)
	Ion	SAMSORG 45 SAMSORG 46	Nigeria	Nigerian National Agricultural Research System and ICRISAT	(190)
Common bean	High iron	NAROBEAN 1 NAROBEAN 2 NAROBEAN 3 NAROBEAN 4C NAROBEAN 5C	Uganda	HarvestPlus and International Center for Tropical Agriculture (CIAT)	(191)
	High iron	CAB 2, RWR 2145, RWR 2245, MAC 42, MAC 44, RWV 1129, RWV 2887, RWV 3006, RWV 3316, and RWV 3317	Rwanda	HarvestPlus and International Center for Tropical Agriculture (CIAT)	(192)
	High Iron	HM 21-7, RWR 2245, PAV 1438, Namulenga, Cod MLV 059, and Cuarentino	Democratic Republic of Congo	HarvestPlus and International Center for Tropical Agriculture (CIAT)	(193, 194)
Lentil	Iron and zinc	Pusa Ageti Masoor (L4717) L4704 Pusa Vaibhav	India	ICAR-IARI, New Delhi and ICARDA	(178, 195)
		Barimasur 4 Barimasur 5 Barimasur 6 Barimasur 7 Barimasur 8	Bangladesh	Bangladesh Agricultural Research Council (BARI) and ICARDA	(196, 197)
		Shekhar Khajuraho 1 Khajuraho 2 Sisir Shital Khajuraho Masuro 3 (RL 4)	Nepal	Nepal Agricultural Research Council and ICARDA	(198)

(Continued)

TABLE 3 | Continued

Crop	Biofortified nutrient	Name of variety	Country	Developing institute	References/source
		Alemaya	Ethiopia	ICARDA	(199)
		Idlib 2, Idlib 3, Idlib 4	Syria	ICARDA	
		Myveci-2001	Turkey	ICARDA	
		Beleza	Portugal	ICARDA	
Soybean	Kuntiz trypsin inhibitor (KTI)-free	NRC 127[a]	India	ICAR-Indian institute of Soybean Research, India	(200)
		UEL 175[a]	Brazil	Sate University of Londrina, Parana, Brazil	(201)
	KTI and lipoxygenase 2 free	NRC 142[a]	India	ICAR-Indian institute of Soybean Research, India	(202)

[a]*Varieties/hybrids developed through genomics-assisted breeding.*

ppm Zn advantage, developed with the support of International Maize and Wheat Improvement Center (CIMMYT). In 2012, the HarvestPlus program of CGIAR, in alliance with Banaras Hindu University, India, and CIMMYT, developed high-Zn genotypes designated as BHU 1, BHU 3 (Akshay), BHU 5, and BHU 6. Through public–private partnership, HarvestPlus has reached more than 50,000 wheat farmers in the Eastern Gangetic Plain of India (168) (**Table 3**). Improvement in protein content is also an important target in wheat breeding. In general, protein content is believed to be negatively correlated with grain yield. However, in 2018, ICAR-Indian Agricultural Research Institute (IARI) New Delhi and PAU Ludhiana become successful in developing high-protein (∼13%) and high-yielding bread varieties HD 3226 (169) and PBW 757, respectively, for NWPZ of India (170). Improvement of β-carotene content is a major target in durum wheat breeding program. The presence of yellow pigment in the grain generally measured as the Yellow Pigment Index (YPI), which is also a measure of β-carotene content, is an important quality trait and indicates an antioxidant in durum wheat. The bright yellow color of pasta due to the presence of yellow pigment in the flour is an indicator of good pasta quality. In India, durum wheat varieties such as WH 896, PDW 233, HI 8759, and HI 8777 with YPI >19 are developed and commercialized (170). In a wheat breeding program, one of the significant aims is to develop non-traditional genotypes having an added value which would impart a relevant health benefit. This biofortified colored wheat with high anthocyanin content is an extensive area of research. Anthocyanin can act as an antioxidant that removes harmful free radicals from the body and helps in the prevention of heart diseases, diabetes, obesity, and cancer. A winter wheat variety "Scorpion," with blue grain color, was registered in 2011 in Austria and in 2012 in Europe (171). Another winter wheat variety with purple grain "PS Karkulka" was developed by the National Agricultural and Food Center, Slovak Republic, in 2014 (172). In India, anthocyanin-rich wheat genotypes, *viz*., NABIMG 9, NABIMG 10, and NABIMG 11, with enhanced Zn content have been developed and registered (173) (**Table 3**).

Rice

The majority of the Asian population prefers rice as their staple food. This milled and polished rice is a poor source of some minerals and proteins; therefore, it becomes a preferred crop for biofortification. Gregorio et al. (176) screened improved cultivars, new plant type lines, local landraces, and lines from wild/related species for Fe and Zn content. It was reported that Fe (15.6 mg/kg) and Zn (378 mg/kg) content was higher in wild/related species. They also reported that an Indian floating rice local cultivar "Jalmagna" had 40% more Zn concentration than the mega variety IR 64. An improved line (IR 68144-313-2-2-3) with high grain Fe concentration (21 ppm in brown rice) was developed by the International Rice Research Institute. This line was derived from a cross of IR 72 and "Zawa Bonday," a traditional variety from India (176). In 2015, Bangladesh Rice Research Institute (BARI) has released their third high-Zn rice variety "BRRI dhan 72" with the support of HarvestPlus (177) (**Table 3**). Two earlier varieties, *viz*., BRRI dhan 62 (2013) and BRRI dhan 64 (2014), were also enriched with Zn. The Indian Council of Agricultural Research (ICAR) has one of the mandates to improve the nutritional quality in high-yielding varieties of cereals, pulses, oilseeds, vegetables, and fruit crops. In rice, India has released two biofortified varieties of rice, *viz*., "CR Dhan 310" having high protein content (10.3% in polished rice), which was developed by the National Rice Research Institute, Cuttack, and "DRR Dhan 45," having high Zn content (22.6 ppm) in polished grain, developed by the Indian Institute of Rice Research Hyderabad (178) (**Table 3**).

Maize

Maize is the third most important food grain, following wheat and rice. In recent years, the pro-vitamin A-enriched maize hybrids/varieties developed across the globe tell a success story of the biofortification program. In India, maize hybrid Pusa Vivek Hybrid Improved, enriched in pro-vitamin A, has been released for commercial cultivation (12). In Zimbabwe, orange maize rich in pro-vitamin A is bred especially for human consumption. The

CIMMYT with HarvestPlus has also released biofortified orange maize variety "ZS 422" for commercial cultivation in Zimbabwe. Similarly, in many African countries where malnutrition is a serious concern, biofortified orange maize varieties have been released (185, 186) (**Table 3**). Maize has a significant limitation; it lacks two essential amino acids, lysine and tryptophan, which are not synthesized in the human body. Maize breeders E. Villegas and SK Vasal at CIMMYT had developed quality protein maize (QPM) genotypes with high lysine and tryptophan by incorporating "opaque 2" gene along with genetic modifiers. In recent years, India has released a good number of QPM hybrids for commercial cultivation which includes Ratan, Protina, Shakti, Shakti 1, Shaktiman 1, Shaktiman 2, Shaktiman 3, Shaktiman 4, Shaktiman 5, HQPM 1, HQPM 5, HQPM 7, Vivel QPM 9, Vivek QPM 21, and Pratap QPM hybrid 1 with enhanced endospermic content of lysine and tryptophan (179) (**Table 3**). In addition to this, new QPM hybrids, *viz*., Pusa HM 4 Improved, Pusa HM 8 improved, and Pusa HM 9 Improved, with 0.68–1.06% lysine and 2.97–4.18% tryptophan were released by IARI, New Delhi, in 2017 (124) (**Table 3**). India released its first pro-vitamin A-enriched QPM hybrid "Pusa Vivek QPM 9 Improved" with high pro-vitamin A (8.5 ppm), lysine (2.76 ppm), and tryptophan (0.74%) (178). In the year 2020, IARI, New Delhi, again came up with three more hybrids, *viz*., Pusa HQPM 5 Improved and Pusa HQPM 7 Improved, having high pro-vitamin A in the QPM hybrid background, and one hybrid Pusa Vivek Hybrid 27 Improved, having high pro-vitamin A (200) (**Table 3**). The Global Maize Program of CIMMYT is the frontrunner in developing and releasing QPM hybrids for commercial cultivation in developing countries. The CIMMYT has also released QPM hybrids in Pakistan (QPHM 200 and QPHM 300), Ethiopia (BHQPY 545 and BHQP 542), Ghana (GH-132-28), Guatemala (HB-PROTICTA), Nicaragua (HQ INTA-993), Honduras (HQ-31), El Salvador (HQ-61), Colombia (ICA), Kenya (KH 500Q, KH 631Q), Tanzania (Lishe-H1, Lishe-H2), Ethiopia (Melkassa 1Q, Melkassa 6Q, MHQ 138), Zimbabwe (ZS 261Q), and China (Quian 2609, Zhong Dan 9409) (**Table 3**). Thus, QPM can be a better alternative to eliminate malnutrition in developing countries such as African nations where maize is a staple diet. In 2018, CIMMYT has released a new zinc-enriched maize variety, BIO-MZn01, in Colombia to help combat malnutrition in South America. This variety contains 36% more zinc on average than other maize varieties, with grain yield of 6–8 t/ha (187). Furthermore, QPM combined with high Fe and Zn, if developed, will be a favorable choice to tackle the problem of malnutrition.

Pearl Millet

Anemia is a significant health concern in millions of women and children in developing countries, especially in African and Asian countries. The reason for this is their staple diet which is deficient in iron (Fe). The pearl millet variety "Shakti" officially released in African country Niger and developed by International Crop Research Institute for Semi-Arid and Tropics (ICRISAT) has higher Fe (65 mg/kg) and Zn (icrisat.org). In India, high-Fe (75 mg/kg) and high-Zn (40 mg/kg) variety "Dhanashakti" and hybrid "ICMH 1201" have been released by ICRISAT during 2012–2014 (**Table 3**). Recently, ICRISAT, in collaboration with state agricultural universities in India, has developed and released four biofortified hybrids, *viz*., AHB 1200 (iron 73.0 ppm), HHB 299 (iron 73.0 ppm; zinc 41.0 ppm), and HHB 311 and RHB 233 with high Fe and Zn content (178) (**Table 3**). These biofortified varieties and hybrids have 7.5 to 8.0 mg of iron and 3.5 to 4.5 mg zinc per 100 gm of pearl millet. It is estimated that consumption of 200 g/day biofortified pearl millet-based food will provide 70% of dietary Fe and Zn requirement in men and women, and 130 g/day will give 100% requirement for children (188). With these improvements, pearl millet is becoming 'Smart Food Crop' on account of its high Fe and Zn content, improved levels of drought, heat, and salinity tolerance, high protein with balanced amino acids, high dietary fibers, and gluten-free protein.

Sorghum

Sorghum is the grain of the twenty first century for Africa. It is the only viable food grain for millions of food-insecure people. In Nigeria, the Nigerian National Agricultural Research System and ICRISAT have released two sorghum varieties, "SAMSORG 45″ and "SAMSORG 46," with 129 ppm Fe content. These biofortified varieties are higher yielding and resistant to Striga. In India, during 2018, ICRISAT has developed sorghum variety ICSR 14001 (later named as Parbhani Shakti), with 45 ppm Fe and 32 ppm Zn. Besides these, it has low phytate content (4.14 mg/100 g), which will increase the bioavailability of nutrients (**Table 3**).

Common Bean

Common bean is important for nutrient and poverty alleviation in the developing countries of Central America, Andean regions of South America, and Eastern and Southern Africa (203). Common bean has wide variation for Fe content, i.e., up to 110 ppm, which is much higher than the target of 40 ppm set by HarvestPlus. Therefore, developing Fe-rich varieties through breeding is quite easy. The HarvestPlus and International Center for Tropical Agriculture (CIAT) have released five varieties (NAROBEAN 1, NAROBEAN 2, NAROBEAN 3, NAROBEAN 4C, and NAROBEAN 5C) in Uganda, 10 varieties (CAB 2, RWR 2145, RWR 2245, MAC 42, MAC 44, RWV 1129, RWV 2887, RWV 3006, RWV 3316, and RWV 3317) in Rwanda, and six varieties (HM 21-7, RWR 2245, PAV 1438, Namulenga, Cod MLV 059, and Cuarentino) in the Democratic Republic of Congo (191–194) (**Table 3**).

Lentil

Lentil is a highly nutritious pulse crop, well-adapted to dry land, in poor-soil-fertility areas of Africa, Middle-East, Indian-Sub-Continent, Southern Europe, America, Australia, and West Asia (204). Lentil has been a choice in priority for biofortification of mainly Fe, Zn, and Se. The International Center for Agricultural Research in Dry Areas (ICARDA) is working with national programs in India, Bangladesh, and Nepal to breed Fe- and Zn-enriched lentil varieties. In recent years, ICARDA's research partnership with Bangladesh Agricultural Research Council (BARI) led to the release of five Fe- and Zn-enriched lentil varieties, namely, Barimasur 4, Barimasur 5, Barimasur 6,

Barimasur 7, and Barimasur 8 (196, 197) (**Table 3**). These biofortified lentil varieties have reached 8,20,000 farmers in Bangladesh. In Nepal, five varieties (Shekhar, Khajuraho 1, Khajuraho 2, Sisir, and Shital) with 81–98 ppm Fe and >54 ppm Zn are rapidly gaining acceptance and popularity among 4,00,000 farmers (198). In India, three biofortified Fe-rich lentil varieties "Pusa Vaibhav" (102 ppm), IPL 220 (73–114 ppm), and L 4704 are becoming popular among the farmers of North Eastern India (178, 195) (**Table 3**).

Soybean

Elimination of Kunitz trypsin inhibitor (KTI) from soybean seeds is one of the important breeding objectives. For its inactivation, preheating of soy flour is required before using it in food and feed products. The heat treatment not only increases the processing but also affects the seed protein quality and solubility (205). For the development of KTI-free soybean lines, MABB approach was adopted to introgress the kti allele (source PI542044) in the genetic background of popular varieties in India: DS9712, DS9814(204), and JS97-52 (206). At ICAR–Indian Institute of Soybean Research, India, a soybean variety NRC 142, free from KTI and lipoxygenase 2, has been developed through marker-assisted breeding (https://iisrindore.icar.gov.in/pdfdoc/NRC142News.pdf). In Brazil, a KTI-free variety UEL 175 has also been developed through marker-assisted breeding (201).

CONVENTIONAL AND GENOMIC-ASSISTED BREEDING: ADVANTAGES AND LIMITATIONS

Biofortification is targeted mainly to the rural poor in developing countries who are mainly dependent on staple foods for meeting their energy and nutrient demand and have very remote access to nutrient-rich fortified foods. The design of conventional plant breeding programs capitalizes on the available variability in the germplasm to breed nutrient-rich crops. The breeding for biofortification has been successful as several varieties have been released in different crops and are in cultivation in India, Pakistan, Zambia, etc. Both conventional and genomics-assisted breeding programs capitalize on the existing variability in the germplasm to breed nutrient-rich crops. The main advantage of both breeding strategies is that the varieties developed through these approaches have no yield penalty from an economic point of view and are also equally competitive to the earlier varieties in cultivation. The acceptability of varieties developed does not have an issue involved as in the case of genetically modified crops. Catering the nutrients through biofortification will uplift the livelihood as well as the nutritional status of the target population as this system is cost-effective and highly sustainable. Most of the quality traits are quantitative in nature and are highly influenced by environmental factors, because of which the progress of breeding varieties will be slow through conventional plant breeding. In many cases, the primary gene pool has limited variability; then, transferring the target trait from wild relatives poses serious constraints with respect to time duration and linked unfavorable genes (linkage drag). A large number of QTLs linked to key nutrient traits have been identified in different crops, which are amenable for marker-assisted selection, but in crops having a large genome size, like wheat, finding closely linked genes to the target traits was challenging, but with the advent of novel genomic tools, like genome-wide association mapping, genomic selection and also whole genome sequence data finding markers tightly linked to the trait of interest are possible. Marker-assisted selection can be applied to transfer the genes/QTLs once closely linked markers are available. Numerous studies have been reported in literature regarding mapping of QTLs linked to different biofortification traits. By using genomics-assisted breeding, the time duration for the development of biofortified varieties can be reduced to 5 to 6 years as compared to conventional breeding, and the problem of linkage drag is also addressed in genomics-assisted breeding while transferring the target trait from wild relatives.

However, both of these approaches have some limitations; for example, if the target micronutrient does not have variability in the crop germplasm, then genetic improvement is impossible. In some cases, it would be impossible to breed for a specific trait using conventional means, and the timescale and the effort involved may be quite unrealistic, e.g., improving Se concentration in wheat grains (207). Rice plants possess the whole pathway to synthesize β-carotene; this pathway is fully active in leaves but turned off in the grain. By adding only two genes, a plant phytoene synthase (psy) and a bacterial phytoene desaturase (crt I), the pathway is turned back on and β-carotene consequently accumulates in the grain. On the other hand, it is not possible either through conventional or genomics-assisted breeding to switch on the pathway. In marker-assisted breeding, the linkage between the trait and the marker determines the success of a biofortification program. In some cases, the genetic background can alter the expression of the transferred traits while transferring QTLs. In case of cross-pollinated crops, the biofortified hybrids may lose the trait in one or few generations of inbreeding, so the farmers have to buy a fresh hybrid seed to maintain the required trait. In most of the developing countries, there is no demarcation or segregated procurement with respect to normal or biofortified wheat crop after harvest. So, all wheats get equal treatment with respect to the procurement price. If there is a scope for the segregated procurement of biofortified crops to be linked with some monetary benefit, such will encourage the adoptability of biofortified crops. Given the widespread prevalence of micronutrient deficiencies, conventional breeding coupled with the integration of genomic tools will be effective in providing access to micronutrient-rich food to the target deficient population in developing countries as they capitalize on staple food crops for the delivery of micronutrients.

PERSPECTIVES OF ENHANCING NUTRIENTS WITH CONVENTIONAL AND GENOMICS-ASSISTED BREEDING

Increasing the content of nutritional elements/minerals through genetic biofortification has been demonstrated in major food

crops, and several varieties have been released (**Table 3**). The traits governing nutritional contents are mostly polygenic and involved GX E interaction; hence, improvement through conventional plant breeding is possible, provided that the time required for the release of the variety is generally long. This is a sustainable and cost-effective solution for the elimination of malnutrition. Advancements in molecular genetics have aided in precise mapping and dissecting the molecular basis of biofortification. Several major-effect and stable QTLs for biofortification traits, such as GPC, minerals, and pigments, have been identified in different crops (**Table 1**). These QTLs can be transferred into elite lines using marker-assisted selection, provided that the markers are tightly linked to the trait of interest, which is also sometimes challenging in crops having a large genome size. The advent of genomic selection has led to drastically reduced cost and increased throughput of genotyping assays, in combination with advances in high-throughput phenotyping. Through genomic selection, several QTLs pertaining to nutritional traits have been mapped and combined into elite lines in wheat (20). Increasing the micronutrients through genetic and genomic approaches is well-demonstrated where genetic variation exists in the germplasm. However, if genetic variation is not available in the germplasm, then novel techniques of genome editing (**Table 2**) and genetic engineering need to be employed, as demonstrated by establishing the pro-vitamin A pathway in golden rice (208). Golden rice has proved to elevate the pro-vitamin A level in rice, but due to strict GMO regulations and unexpected health risks, transgenic crops still have not found acceptance. Otherwise, golden rice has ample to offer to the world's malnutrition-prone population.

The other less explored nutritional elements, like low sugar and omega 3- fatty acids, are gaining importance with the upsurge of chronic illnesses like diabetes and cardiovascular diseases. The challenge is to alter the starch structure of rice to lower its digestibility so that it provides a slow and steady supply of energy but without sacrificing its cooking and eating quality. A low-glycemic-index (GI) rice variety with hard texture is not preferred in the Southeast Asian market, where soft and sticky types are the benchmarks of rice quality. Variations in rice germplasm for low GI along with hard and cohesive texture have been assessed using genome-wide association mapping and are currently being used as pre-breeding materials to develop low-GI lines targeted toward certain market segments, particularly in Asia (209). Omega-3 fatty acids have proven to be very essential for human health due to their multiple health benefits like skin and hair growth as well as for proper visual, neural, and reproductive functions of the body, but the concern is that they are not produced by the human body. The requirement of these fatty acids is generally met by deep sea fishes as variation in plants is not that much. The entire DHA biosynthetic pathway was reconstituted in oilseed crop *Brassica juncea* by stepwise metabolic engineering. Transgenic plants produced up to 25% arachidonic acid and 15% eicosapentaenoic acid (EPA), as well as up to 1.5% docosahexaenoic acid (DHA) in seeds (210). However, the accumulation of EPA and DHA was low in transgenic lines, and therefore it was difficult to commercialize. An emerging oilseed crop, *Camelina sativa* seed, contains >30% alpha-linolenic acid, which is the starting substrate required for the synthesis of EPA and DHA, making it a good platform for assembling the ω-3-LC-PUFA synthesis pathway. The co-expression of five genes (OtD6 desaturase from the eukaryotic microalgae *Ostreococcus tauri*, TcD5 desaturase from marine fungus *Thraustochytrium* sp, Piw-3 desaturase from *Phytophthora infestans*, PsD12 desaturase from *Phytophthora sojae*, and D6 fatty acid carbon chain elongation enzyme PSE1 from *Physcomitrella patens*) has demonstrated EPA accumulation of up to 31% in camelina oil (211). The contents of EPA and DHA in this transgenic camelina were comparable to the levels found in fish oil, representing the successful assembly of biosynthetic pathways in commercial oil crop seeds, but there are regulatory issues in the case of transgenics. Genome-wide analysis tools have revealed several genomic regions associated with fatty acid composition in seed oil that had never before been implicated in lipid metabolism (212). By mining sequence data, the candidate enzymes for target fatty acids were identified and subsequently used to manipulate *C. sativa* oil composition toward a superior biofuel and bio-based lubricant oil (213, 214).

In view of this, conventional and genomics-assisted breeding both provide a way of improving the micronutrient status of crops as they have no acceptance problem like the transgenics. An additional advantage that these approaches offer is that farmers can use and retain the seed for food as well as for the next crop plantation. Staple crop biofortification will offer the advantage of delivering the much-needed micronutrients through seeds directly to the malnuorished population in the developing world. Additionally, further studies are needed to improve the bioavailability of these elevated protein, mineral, and pigment concentrations, and the target micronutrients for each crop need to be set by a nutrition specialist in consultation with breeders. The most important issue is making the farmers and populations aware about the nutritional benefits of consuming biofortified foods for their acceptability among the masses. With the immense potential and scope of biofortification, it can be envisaged that this technology can breed nutrient-dense crops that will be crucial in addressing the problem of malnutrition.

CONCLUSION

Making our staple food crops nutrient-rich either by enriching our soils with nutrients or making nutrient-rich crops by modern breeding and biotechnological tools such as discovery and mapping of QTLs and their use in marker-assisted breeding will certainly help in eliminating malnutrition to a larger extent. The advanced QTL mapping tools like GWAS and the annotated whole-genome sequence of wheat and other cereals offer new opportunities to study the exact nature of allelic variation and explore the underlying genetic basis and putative gene(s) associated with quality traits. In conclusion, the compiled information on the QTLs on different nutrients flanked by linked marker systems, *viz.*, SSR, SNP, DArT, and CAPS, would help in their precise introgression into elite breeding lines/cultivated

varieties through MAS in biofortification breeding programs. Impressive work has been done in the identification of QTLs pertaining to nutritional quality and the release of nutritionally rich varieties through conventional breeding in major food crops. However, new avenues in genome editing research are making impressive findings in enriching our food crops to make them nutritionally rich. Genome editing technologies that can rapidly modify genomes in a precise way and will directly enrich the nutritional status of elite varieties could hold a bright future to address the challenge of malnutrition.

AUTHOR CONTRIBUTIONS

KG prepared the outline of the contents and coordinated the drafting of the manuscript. KG, SR, MK, VG, PB, and RY wrote the manuscript. NB assisted in final editing and addressing the comments. All authors contributed to the edits and the review.

CONSENT FOR PUBLICATION

The manuscript has been approved by all authors.

REFERENCES

1. Food and Agriculture Organization of the United Nations. *The State of Food Insecurity in the World 2014: Strengthening the Enabling Environment for Food. Security and Nutrition*. Food & Agriculture Org. (2015).
2. Field CB, Barros V, Stocker TF, Dahe Q. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press (2012).
3. Garg M, Sharma N, Sharma S, Kapoor P, Kumar A, Chunduri V, et al. Biofortified crops generated by breeding, agronomy, and transgenic approaches are improving lives of millions of people around the world. *Front Nutr.* (2018) 5:12. doi: 10.3389/fnut.2018.00012
4. White PJ, Broadley MR. Biofortifying crops with essential mineral elements. *Trends Plant Sci.* (2005) 10:586–93. doi: 10.1016/j.tplants.2005.10.001
5. Myers SS, Zanobetti A, Kloog I, Huybers P, Leakey AD, Bloom AJ, et al. Increasing CO 2 threatens human nutrition. *Nature.* (2014) 510:139–42. doi: 10.1038/nature13179
6. Loladze I. Hidden shift of the ionome of plants exposed to elevated CO2 depletes minerals at the base of human nutrition. *elife.* (2014) 3:e02245. doi: 10.7554/eLife.02245
7. Myers SS, Wessells KR, Kloog I, Zanobetti A, Schwartz J. Effect of increased concentrations of atmospheric carbon dioxide on the global threat of zinc deficiency: a modelling study. *Lancet Glob Health.* (2015) 3:e639–45. doi: 10.1016/S2214-109X(15)00093-5
8. Palmer AC, Healy K, Barffour MA, Siamusantu W, Chileshe J, Schulze KJ, et al. Provitamin A carotenoid–biofortified maize consumption increases pupillary responsiveness among Zambian children in a randomized controlled trial. *J Nutr.* (2016) 146:2551–8. doi: 10.3945/jn.116.239202
9. Talsma EF, Brouwer ID, Verhoef H, Mbera GN, Mwangi AM, Demir AY, et al. Biofortified yellow cassava and vitamin A status of Kenyan children: a randomized controlled trial. *Am J Clin Nutr.* (2016) 103:258–67. doi: 10.3945/ajcn.114.100164
10. Babu R, Rojas NP, Gao S, Yan J, Pixley K. Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. *Theor Appl Genet.* (2013) 126:389–99. doi: 10.1007/s00122-012-1987-3
11. Azmach G, Gedil M, Menkir A, Spillane C. Marker-trait association analysis of functional gene markers for provitamin A levels across diverse tropical yellow maize inbred lines. *BMC Plant Biol.* (2013) 13:227. doi: 10.1186/1471-2229-13-227
12. Muthusamy V, Hossain F, Thirunavukkarasu N, Choudhary M, Saha S, Bhat JS, et al. Development of β-carotene rich maize hybrids through marker-assisted introgression of β-carotene hydroxylase allele. *PLoS ONE.* (2014) 9:e113583. doi: 10.1371/journal.pone.0113583
13. Liu L, Jeffers D, Zhang Y, Ding M, Chen W, Kang MS, et al. Introgression of the crtRB1 gene into quality protein maize inbred lines using molecular markers. *Mol Breed.* (2015) 35:154. doi: 10.1007/s11032-015-0349-7
14. Yan J, Kandianis CB, Harjes CE, Bai L, Kim E-H, Yang X, et al. Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain. *Nat Genet.* (2010) 42:322–7. doi: 10.1038/ng.551
15. Guo Y, Zhang G, Guo B, Qu C, Zhang M, Kong F, et al. QTL mapping for quality traits using a high-density genetic map of wheat. *PLoS ONE.* (2020) 15:e0230601. doi: 10.1371/journal.pone.0230601
16. Liu J, Wu B, Singh RP, Velu G. QTL mapping for micronutrients concentration and yield component traits in a hexaploid wheat mapping population. *J Cereal Sci.* (2019) 88:57–64. doi: 10.1016/j.jcs.2019.05.008
17. Fatiukha A, Filler N, Lupo I, Lidzbarsky G, Klymiuk V, Korol AB, et al. Grain protein content and thousand kernel weight QTLs identified in a durum× wild emmer wheat mapping population tested in five environments. *Theor Appl Genet.* (2020) 133:119–31. doi: 10.1007/s00122-019-03444-8
18. Gorafi YS, Ishii T, Kim J-S, Elbashir AAE, Tsujimoto H. Genetic variation and association mapping of grain iron and zinc contents in synthetic hexaploid wheat germplasm. *Plant Genet Resour.* (2018) 16:9–17. doi: 10.1017/S1479262116000265
19. Velu G, Singh RP, Crespo-Herrera L, Juliana P, Dreisigacker S, Valluru R, et al. Genetic dissection of grain zinc concentration in spring wheat for mainstreaming biofortification in CIMMYT wheat breeding. *Sci Rep.* (2018) 8:13526. doi: 10.1038/s41598-018-31951-z
20. Velu G, Tutus Y, Gomez-Becerra HF, Hao Y, Demir L, Kara R, et al. QTL mapping for grain zinc and iron concentrations and zinc efficiency in a tetraploid and hexaploid wheat mapping populations. *Plant Soil.* (2017) 411:81–99. doi: 10.1007/s11104-016-3025-8
21. Crespo-Herrera L, Velu G, Singh R. Quantitative trait loci mapping reveals pleiotropic effect for grain iron and zinc concentrations in wheat. *Ann Appl Biol.* (2016) 169:27–35. doi: 10.1111/aab.12276
22. Tiwari C, Wallwork H, Arun B, Mishra VK, Velu G, Stangoulis J, et al. Molecular mapping of quantitative trait loci for zinc, iron and protein content in the grains of hexaploid wheat. *Euphytica.* (2016) 207:563–70. doi: 10.1007/s10681-015-1544-7
23. Srinivasa J, Arun B, Mishra VK, Singh GP, Velu G, Babu R, et al. Zinc and iron concentration QTL mapped in a Triticum spelta× *T. aestivum* cross. *Theor Appl Genet.* (2014) 127:1643–51. doi: 10.1007/s00122-014-2327-6
24. Hao Y, Velu G, Peña RJ, Singh S, Singh RP. Genetic loci associated with high grain zinc concentration and pleiotropic effect on kernel weight in wheat (*Triticumaestivum* L.). *Mol Breed.* (2014) 34:1893–902. doi: 10.1007/s11032-014-0147-7
25. Sadeghzadeh B, Rengel Z, Li C. Quantitative trait loci (QTL) of seed Zn accumulation in barley population clipper X Sahara. *J Plant Nutr.* (2015) 38:1672–84. doi: 10.1080/01904167.2014.991036
26. Islam M, Arifuzzaman M, Banik S, Hossain M, Ferdous J, Khalequzzaman M, et al. Mapping QTLs underpin nutrition components in aromatic rice germplasm. *PLoS ONE.* (2020) 15:e0234395. doi: 10.1371/journal.pone.0234395
27. Pradhan S, Pandit E, Pawar S, Naveenkumar R, Barik S, Mohanty S, et al. Linkage disequilibrium mapping for grain Fe and Zn enhancing QTLs useful for nutrient dense rice breeding. *BMC Plant Biol.* (2020) 20:57. doi: 10.1186/s12870-020-2262-4
28. Jang S, Han J-H, Lee YK, Shin N-H, Kang YJ, Kim C-K, et al. Mapping and validation of QTLs for the amino acid and total protein content in brown rice. *Front Genet.* (2020) 11:240. doi: 10.3389/fgene.2020.00240

29. Dixit S, Singh UM, Abbai R, Ram T, Singh VK, Paul A, et al. Identification of genomic region (s) responsible for high iron and zinc content in rice. *Sci Rep*. (2019) 9:8136. doi: 10.1038/s41598-019-43888-y
30. Zhong M, Wang LQ, Yuan DJ, Luo LJ, Xu CG, He YQ. Identification of QTL affecting protein and amino acid contents in rice. *Rice Sci.* (2011) 18:187–95. doi: 10.1016/S1672-6308(11)60026-7
31. Kinoshita N, Kato M, Koyasaki K, Kawashima T, Nishimura T, Hirayama Y, et al. Identification of quantitative trait loci for rice grain quality and yield-related traits in two closely related *Oryza sativa* L. subsp. *japonica* cultivars grown near the northernmost limit for rice paddy cultivation. *Breed Sci.* (2017) 67:191–206. doi: 10.1270/jsbbs.16155
32. Descalsota-Empleo GI, Amparado A, Inabangan-Asilo MA, Tesoro F, Stangoulis J, Reinke R, et al. Genetic mapping of QTL for agronomic traits and grain mineral elements in rice. *Crop J.* (2019) 7:560–72. doi: 10.1016/j.cj.2019.03.002
33. Chattopadhyay K, Behera L, Bagchi TB, Sardar SS, Moharana N, Patra NR, et al. Detection of stable QTLs for grain protein content in rice (*Oryza sativa* L.) employing high throughput phenotyping and genotyping platforms. *Sci Rep*. (2019) 9:3196. doi: 10.1038/s41598-019-39863-2
34. Jeong OY, Lee JH, Jeong EG, Chun A, Bombay M, Banzon Ancheta M, et al. Analysis of QTL responsible for grain iron and zinc content in doubled haploid lines of rice (*Oryza sativa*) derived from an intra-japonica cross. *Plant Breed.* (2020) 139:344–55. doi: 10.1111/pbr.12787
35. Swamy B, Kaladhar K, Anuradha K, Batchu AK, Longvah T, Sarla N. QTL analysis for grain iron and zinc concentrations in two *O. nivara* derived backcross populations. *Rice Sci.* (2018) 25:197–207. doi: 10.1016/j.rsci.2018.06.003
36. Descalsota GIL, Swamy B, Zaw H, Inabangan-Asilo MA, Amparado A, Mauleon R, et al. Genome-wide association mapping in a rice MAGIC Plus population detects QTLs and genes useful for biofortification. *Front Plant Sci.* (2018) 9:1347. doi: 10.3389/fpls.2018.01347
37. Xu Q, Zheng T-Q, Hu X, Cheng L-R, Xu J-L, Shi Y-M, et al. Examining two sets of introgression lines in rice (*Oryza sativa* L.) reveals favorable alleles that improve grain Zn and Fe concentrations. *PLoS ONE.* (2015) 10:e0131846. doi: 10.1371/journal.pone.0131846
38. Agarwal S, VGN TV, Kotla A, Mangrauthia SK, Neelamraju S. Expression patterns of QTL based and other candidate genes in Madhukar× Swarna RILs with contrasting levels of iron and zinc in unpolished rice grains. *Gene.* (2014) 546:430–6. doi: 10.1016/j.gene.2014.05.069
39. Yun BW, Kim MG, Handoyo T, Kim KM. Analysis of rice grain quality-associated quantitative trait loci by using genetic mapping. *Am J Plant Sci.* (2014) 5:1125–32. doi: 10.4236/ajps.2014.59125
40. Lee G-H, Yun B-W, Kim K-M. Analysis of QTLs associated with the rice quality related gene by double haploid populations. *Int J Genomics.* (2014) 2014:781832. doi: 10.1155/2014/781832
41. Zhang M, Pinson SR, Tarpley L, Huang X-Y, Lahner B, Yakubova E, et al. Mapping and validation of quantitative trait loci associated with concentrations of 16 elements in unmilled rice grain. *Theor Appl Genet.* (2014) 127:137–65. doi: 10.1007/s00122-013-2207-5
42. Gande NK, Kundur PJ, Soman R, Ambati R, Ashwathanarayana R, Bekele BD, et al. Identification of putative candidate gene markers for grain zinc content using recombinant inbred lines (RIL) population of IRRI38 X Jeerigesanna. *Afr J Biotechnol.* (2014) 13:657–63. doi: 10.5897/AJB2013.13334
43. Peng B, Wang L, Fan C, Jiang G, Luo L, Li Y, et al. Comparative mapping of chalkiness components in rice using five populations across two environments. *BMC Genet.* (2014) 15:49. doi: 10.1186/1471-2156-15-49
44. Gu R, Chen F, Liu B, Wang X, Liu J, Li P, et al. Comprehensive phenotypic analysis and quantitative trait locus identification for grain mineral concentration, content, and yield in maize (*Zea mays* L.). *Theor Appl Genet.* (2015) 128:1777–89. doi: 10.1007/s00122-015-2546-5
45. Baxter IR, Gustin JL, Settles AM, Hoekenga OA. Ionomic characterization of maize kernels in the intermated B73× Mo17 population. *Crop Sci.* (2013) 53:208–20. doi: 10.2135/cropsci2012.02.0135
46. Kumar S, Hash CT, Nepolean T, Mahendrakar MD, Satyavathi CT, Singh G, et al. Mapping grain iron and zinc content quantitative trait loci in an iniadi-derived immortal population of pearl millet. *Genes.* (2018) 9:248. doi: 10.3390/genes9050248
47. Anuradha N, Satyavathi CT, Bharadwaj C, Nepolean T, Sankar SM, Singh SP, et al. Deciphering genomic regions for high grain iron and zinc content using association mapping in pearl millet. *Front Plant Sci.* (2017) 8:412. doi: 10.3389/fpls.2017.00412
48. Izquierdo P, Astudillo C, Blair MW, Iqbal AM, Raatz B, Cichy KA. Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet.* (2018) 131:1645–58. doi: 10.1007/s00122-018-3104-8
49. Blair MW, Izquierdo P, Astudillo C, Grusak MA. A legume biofortification quandary: variability and genetic control of seed coat micronutrient accumulation in common beans. *Front Plant Sci.* (2013) 4:275. doi: 10.3389/fpls.2013.00275
50. Casanas F, Pérez-Vega E, Almirall A, Plans M, Sabaté J, Ferreira JJ. Mapping of QTL associated with seed chemical content in a RIL population of common bean (*Phaseolus vulgaris* L.). *Euphytica.* (2013) 192:279–88. doi: 10.1007/s10681-013-0880-8
51. Blair MW, Izquierdo P. Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. *Theor Appl Genet.* (2012) 125:1015–31. doi: 10.1007/s00122-012-1891-x
52. Blair MW, Astudillo C, Rengifo J, Beebe SE, Graham R. QTL analyses for seed iron and zinc concentrations in an intra-genepool population of Andean common beans (*Phaseolus vulgaris* L.). *Theor Appl Genet.* (2011) 122:511–21. doi: 10.1007/s00122-010-1465-8
53. Ates D, Aldemir S, Yagmur B, Kahraman A, Ozkan H, Vandenberg A, et al. QTL mapping of genome regions controlling manganese uptake in lentil seed. *G3.* (2018) 8:1409–16. doi: 10.1534/g3.118.200259
54. Aldemir S, Ateş D, Temel HY, Yagmur B, Alsaleh A, Kahriman A, et al. QTLs for iron concentration in seeds of the cultivated lentil (Lens culinaris Medic.) via genotyping by sequencing. *Turkish J Agric. Forest.* (2017) 41:243–55. doi: 10.3906/tar-1610-33
55. Kundu SS. *Zinc Biofortification of Lentil*. University of Saskatchewan (2016).
56. Ates D, Sever T, Aldemir S, Yagmur B, Temel HY, Kaya HB, et al. Identification QTLs controlling genes for Se uptake in lentil seeds. *PLoS ONE.* (2016) 11:e0149210. doi: 10.1371/journal.pone.0149210
57. Aldemir S, Sever T, Ates D, Yagmur B, Kaya H, Temel H, et al. QTL mapping of genes controlling Fe uptake in lentil (*Lens culinaris* L.) seed using recombinant inbred lines. In: *Plant and Animal Genome Conference* (San Diego, CA) (2014).
58. Ates D, Aldemir S, Sever T, Yagmur B, Kaya H, Temel H, et al. "Identification of QTLs controlling genes to Mn and Zn uptake in lentil seeds," in *Plant and Animal Genome XXII*, San Diego, CA (2014). p. 11–5.
59. Ma Y, Coyne CJ, Grusak MA, Mazourek M, Cheng P, Main D, et al. Genome-wide SNP identification, linkage map construction and QTL mapping for seed mineral concentrations and contents in pea (*Pisum sativum* L.). *BMC Plant Biol.* (2017) 17:43. doi: 10.1186/s12870-016-0956-4
60. Bangar P. *Effect of Iron and Carotenoid Concentration on Iron Bioavailability in Pea Pisum sativum L.* Saskatoon, SK: University of Saskatchewan (2016).
61. Shunmugam A, Liu X, Stonehouse R, Tar'An B, Bett K, Sharpe A, et al. Mapping seed phytic acid concentration and iron bioavailability in a pea recombinant inbred line population. *Crop Sci.* (2015) 55:828–36. doi: 10.2135/cropsci2014.08.0544
62. Upadhyaya HD, Bajaj D, Das S, Kumar V, Gowda C, Sharma S, et al. Genetic dissection of seed-iron and zinc concentrations in chickpea. *Sci Rep.* (2016) 6:24050. doi: 10.1038/srep24050
63. Jadhav A, Rayate S, Mhase L, Thudi M, Chitikineni A, Harer P, et al. Marker-trait association study for protein content in chickpea (*Cicer arietinum* L.). *J Genet.* (2015) 94:279–86. doi: 10.1007/s12041-015-0529-6
64. Diapari M, Sindhu A, Bett K, Deokar A, Warkentin TD, Tar'an B. Genetic diversity and association mapping of iron and zinc concentrations in chickpea (*Cicer arietinum* L.). *Genome.* (2014) 57:459–68. doi: 10.1139/gen-2014-0108
65. Huang L, He H, Chen W, Ren X, Chen Y, Zhou X, et al. Quantitative trait locus analysis of agronomic and quality-related traits in cultivated peanut (Arachis hypogaea L.). *Theor Appl Genet.* (2015) 128:1103–15. doi: 10.1007/s00122-015-2493-1

66. Wang ML, Khera P, Pandey MK, Wang H, Qiao L, Feng S, et al. Genetic mapping of QTLs controlling fatty acids provided insights into the genetic control of fatty acid synthesis pathway in peanut (*Arachis hypogaea* L.). *PLoS ONE.* (2015) 10:e0119454. doi: 10.1371/journal.pone.0119454
67. Pandey MK, Wang ML, Qiao L, Feng S, Khera P, Wang H, et al. Identification of QTLs associated with oil content and mapping FAD2 genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* L.). *BMC Genet.* (2014) 15:133. doi: 10.1186/s12863-014-0133-4
68. Vijayata S. *Genotypic Response and QTL Identification for Micronutrient (Iron and Zinc) Contents in Mungbean [Vigna radiata (L.) Wilczek]*. CCSHAU. (2013).
69. Sompong U, Somta P, Raboy V, Srinives P. Mapping of quantitative trait loci for phytic acid and phosphorus contents in seed and seedling of mungbean (*Vigna radiata* (L.) Wilczek). *Breeding Sci.* (2012) 62:87–92. doi: 10.1270/jsbbs.62.87
70. Ma Y, Ma W, Hu D, Zhang X, Yuan W, He X, et al. QTL mapping for protein and sulfur-containing amino acid contents using a high-density bin-map in soybean (Glycine max L. Merr.). *J Agric Food Chem.* (2019) 67:12313–21. doi: 10.1021/acs.jafc.9b04497
71. Akond M, Yuan J, Liu S, Kantartzi SK, Meksem K, Bellaloui N, et al. Detection of QTL underlying seed quality components in soybean [Glycine max (L.) Merr.]. *Can J Plant Sci.* (2018) 98:881–8. doi: 10.1139/cjps-2017-0204
72. Cai Z, Cheng Y, Ma Z, Liu X, Ma Q, Xia Q, et al. Fine-mapping of QTLs for individual and total isoflavone content in soybean (*Glycine max* L.) using a high-density genetic map. *Theor Appl Genet.* (2018) 131:555–68. doi: 10.1007/s00122-017-3018-x
73. Warrington C, Abdel-Haleem H, Hyten D, Cregan P, Orf J, Killam A, et al. QTL for seed protein and amino acids in the Benning× Danbaekkong soybean population. *Theor Appl Genet.* (2015) 128:839–50. doi: 10.1007/s00122-015-2474-4
74. Zeng A, Chen P, Shi A, Wang D, Zhang B, Orazaly M, et al. Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Sci.* (2014) 54:554–64. doi: 10.2135/cropsci2013.01.0036
75. Wang J, Chen P, Wang D, Shannon G, Zeng A, Orazaly M, et al. Identification and mapping of stable QTL for protein content in soybean seeds. *Mol Breed.* (2015) 35:92. doi: 10.1007/s11032-015-0285-6
76. Wang J, Chen P, Wang D, Shannon G, Shi A, Zeng A, et al. Identification of quantitative trait loci for oil content in soybean seed. *Crop Sci.* (2015) 55:23–34. doi: 10.2135/cropsci2014.04.0280
77. Ning L, Sun P, Wang Q, Ma D, Hu Z, Zhang D, et al. Genetic architecture of biofortification traits in soybean (Glycine max L. Merr.) revealed through association analysis and linkage mapping. *Euphytica.* (2015) 204:353–69. doi: 10.1007/s10681-014-1340-9
78. Wang X, Jiang G-L, Green M, Scott RA, Song Q, Hyten DL, et al. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. *Mol Genet Genomics.* (2014) 289:935–49. doi: 10.1007/s00438-014-0865-x
79. Yan L, Xing L-L, Yang C-Y, Chang R-Z, Zhang M-C, Qiu L-J. Identification of quantitative trait loci associated with soybean seed protein content using two populations derived from crosses between glycine max and glycine soja. *Plant Genet Resour.* (2014) 12:S104. doi: 10.1017/S1479262114000379
80. Ramamurthy RK, Jedlicka J, Graef GL, Waters BM. Identification of new QTLs for seed mineral, cysteine, and methionine concentrations in soybean [*Glycine max* (L.) Merr.]. *Mol Breed.* (2014) 34:431–45. doi: 10.1007/s11032-014-0045-z
81. Li B, Tian L, Zhang J, Huang L, Han F, Yan S, et al. Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in Glycine max. *BMC Genomics.* (2014) 15:1086. doi: 10.1186/1471-2164-15-1086
82. Orazaly M, Chen P, Zhang B, Zeng A. Quantitative trait loci mapping for seed calcium content of soybean. *Crop Sci.* (2014) 54:500–6. doi: 10.2135/cropsci2013.01.0026
83. Akond M, Liu S, Boney M, Kantarrtzi SK, Meksem K, Bellaloui N, et al. Identification of quantitative trait loci (QTL) underlying protein, oil, and five major fatty acids' contents in soybean. *Am J Plant Sci.* (2014) 5:158–67. doi: 10.4236/ajps.2014.51021
84. Qi Z, Hou M, Han X, Liu C, Jiang H, Xin D, et al. Identification of quantitative trait loci (QTL s) for seed protein concentration in soybean and analysis for additive effects and epistatic effects of QTL s under multiple environments. *Plant Breed.* (2014) 133:499–507. doi: 10.1111/pbr.12179
85. Behnke N, Suprianto E, Möllers C. A major QTL on chromosome C05 significantly reduces acid detergent lignin (ADL) content and increases seed oil and protein content in oilseed rape (*Brassica napus* L.). *Theor Appl Genet.* (2018) 131:2477–92. doi: 10.1007/s00122-018-3167-6
86. Gacek K, Bayer PE, Bartkowiak-Broda I, Szala L, Bocianowski J, Edwards D, et al. Genome-wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus. Front Plant Sci.* (2017) 7:2062. doi: 10.3389/fpls.2016.02062
87. Gu J, Chao H, Wang H, Li Y, Li D, Xiang J, et al. Identification of the relationship between oil body morphology and oil content by microstructure comparison combining with QTL analysis in *Brassica napus. Front Plant Sci.* (2017) 7:1989. doi: 10.3389/fpls.2016.01989
88. Basnet RK, Del Carpio DP, Xiao D, Bucher J, Jin M, Boyle K, et al. A systems genetics approach identifies gene regulatory networks associated with fatty acid composition in *Brassica rapa* seed. *Plant Physiol.* (2016) 170:568–85. doi: 10.1104/pp.15.00853
89. Javed N, Geng J, Tahir M, McVetty P, Li G, Duncan RW. Identification of QTL influencing seed oil content, fatty acid profile and days to flowering in *Brassica napus* L. *Euphytica.* (2016) 207:191–211. doi: 10.1007/s10681-015-1565-2
90. Zou J, Zhao Y, Liu P, Shi L, Wang X, Wang M, et al. Seed quality traits can be predicted with high accuracy in *Brassica napus* using genomic data. *PLoS ONE.* (2016) 11:e0166624. doi: 10.1371/journal.pone.0166624
91. Jiang C, Shi J, Li R, Long Y, Wang H, Li D, et al. Quantitative trait loci that control the oil content variation of rapeseed (*Brassica napus* L.). *Theor Appl Genet.* (2014) 127:957–68. doi: 10.1007/s00122-014-2271-5
92. Körber N, Bus A, Li J, Parkin IA, Wittkop B, Snowdon RJ, et al. Agronomic and seed quality traits dissected by genome-wide association mapping in *Brassica napus. Front Plant Sci.* (2016) 7:386. doi: 10.3389/fpls.2016.00386
93. Xu J, Long Y, Wu J, Xu H, Wen J, Meng J, et al. QTL mapping and analysis of the embryo and maternal plant for three limiting amino acids in rapeseed meal. *Eur Food Res Technol.* (2015) 240:147–58. doi: 10.1007/s00217-014-2316-7
94. Huang X-Q, Huang T, Hou G-Z, Li L, Hou Y, Lu Y-H. Identification of QTLs for seed quality traits in rapeseed (*Brassica napus* L.) using recombinant inbred lines (RILs). *Euphytica.* (2016) 210:1–16. doi: 10.1007/s10681-016-1675-5
95. Wang X, Long Y, Yin Y, Zhang C, Gan L, Liu L, et al. New insights into the genetic networks affecting seed fatty acid concentrations in *Brassica napus. BMC Plant Biol.* (2015) 15:91. doi: 10.1186/s12870-015-0475-8
96. Wen J, Xu J, Long Y, Xu H, Wu J, Meng J, et al. Mapping QTLs controlling beneficial fatty acids based on the embryo and maternal plant genomes in *Brassica napus* L. *J Am Oil Chem Soc.* (2015) 92:541–52. doi: 10.1007/s11746-015-2618-3
97. Lemesh V, Mozgova G, Grushetskaya Z, Sidorenko E, Piluk YE, Bakanovskaya A. The use of specific DNA markers for the identification of alleles of the FAD3 genes in rape (*Brassica napus* L.). *Russ J Genet.* (2015) 51:765–73. doi: 10.1134/S1022795415080049
98. Liu F, Wang G, Liu R, Guan C. The promoter of fatty acid desaturase on chromosome C5 in *Brassica napus* drives high-level expression in seeds. *Plant Biotechnol Rep.* (2016) 10:369–81. doi: 10.1007/s11816-016-0407-6
99. Yu B, Boyle K, Zhang W, Robinson SJ, Higgins E, Ehman L, et al. Multi-trait and multi-environment QTL analysis reveals the impact of seed colour on seed composition traits in *Brassica napus. Mol Breed.* (2016) 36:111. doi: 10.1007/s11032-016-0521-8
100. Wen J, Xu JF, Long Y, Wu JG, Xu HM, Meng JL, et al. QTL mapping based on the embryo and maternal genetic systems for non-essential amino acids in rapeseed (*Brassica napus* L.) meal. *J Sci Food Agric.* (2016) 96:465–73. doi: 10.1002/jsfa.7112
101. Wells R, Trick M, Soumpourou E, Clissold L, Morgan C, Werner P, et al. The control of seed oil polyunsaturate content in the polyploid crop species *Brassica napus*. *Mol Breed.* (2014) 33:349–62. doi: 10.1007/s11032-013-9954-5

102. Zhang W, Hu D, Raman R, Guo S, Wei Z, Shen X, et al. Investigation of the genetic diversity and quantitative trait loci accounting for important agronomic and seed quality traits in *Brassica carinata*. *Front Plant Sci.* (2017) 8:615. doi: 10.3389/fpls.2017.00615
103. Gupta S, Sangha M, Kaur G, Banga S, Gupta M, Kumar H, et al. QTL analysis for phytonutrient compounds and the antioxidant molecule in mustard (*Brassica juncea* L.). *Euphytica.* (2015) 201:345–56. doi: 10.1007/s10681-014-1204-3
104. Saini N, Singh N, Kumar A, Vihan N, Yadav S, Vasudev S. Development and validation of functional CAPS markers for the FAE genes in *Brassica juncea* and their use in marker-assisted selection. *Breed Sci.* (2016) 66:831–7. doi: 10.1270/jsbbs.16132
105. Karim MM, Tonu NN, Hossain MS, Funaki T, Meah MB, Hossain DM, et al. Marker-assisted selection of low erucic acid quantity in short duration *Brassica rapa*. *Euphytica.* (2016) 208:535–44. doi: 10.1007/s10681-015-1596-8
106. Singer SD, Weselake RJ, Rahman H. Development and characterization of low α-linolenic acid Brassica oleracea lines bearing a novel mutation in a 'class a'FATTY ACID DESATURASE 3gene. *BMC Genet.* (2014) 15:94. doi: 10.1186/s12863-014-0094-7
107. Cakmak I, Pfeiffer WH, McClafferty B. Biofortification of durum wheat with zinc and iron. *Cereal Chem.* (2010) 87:10–20. doi: 10.1094/CCHEM-87-1-0010
108. Singh U, Praharaj C, Singh SS, Singh NP. *Biofortification of Food Crops*. New Delhi: Springer (2016).
109. Khazaei H, Podder R, Caron CT, Kundu SS, Diapari M, Vandenberg A, et al. Marker–trait association analysis of iron and zinc concentration in lentil (Lens culinaris Medik.) seeds. *Plant Genome.* (2017) 10:1–8. doi: 10.3835/plantgenome2017.02.0007
110. Kumar J, Gupta DS, Kumar S, Gupta S, Singh NP. Current knowledge on genetic biofortification in lentil. *J Agric Food Chem.* (2016) 64:6383–96. doi: 10.1021/acs.jafc.6b02171
111. Singh A, Sharma V, Dikshit H, Singh D, Aski M, Prakash P, et al. Microsatellite marker-based genetic diversity analysis of elite lentil lines differing in grain iron and zinc concentration. *J Plant Biochem Biotechnol.* (2017) 26:199–207. doi: 10.1007/s13562-016-0382-6
112. Saini N, Koramutla MK, Singh N, Singh S, Singh R, Yadav S, et al. Promoter polymorphism in FAE1. 1 and FAE1. 2 genes associated with erucic acid content in *Brassica juncea*. *Mol Breed.* (2019) 39:75. doi: 10.1007/s11032-019-0971-x
113. Gunstone FD, Harwood JL, Dijkstra AJ. *The Lipid Handbook with CD-ROM*. Buca Raton, FL: CRC Press (2007).
114. Norden A, Gorbet D, Knauft D, Young C. Variability in oil quality among peanut genotypes in the Florida breeding program. *Peanut Sci.* (1987) 14:7–11. doi: 10.3146/i0095-3679-14-1-3
115. Moore K, Knauft D. The inheritance of high oleic acid in peanut. *J Hered.* (1989) 80:252–3. doi: 10.1093/oxfordjournals.jhered.a110845
116. Hu X, Zhang S, Miao H, Cui F, Shen Y, Yang W, et al. High-density genetic map construction and identification of QTLs controlling oleic and linoleic acid in peanut using SLAF-seq and SSRs. *Sci Rep.* (2018) 8:5479. doi: 10.1038/s41598-018-23873-7
117. Rehman HM, Cooper JW, Lam HM, Yang SH. Legume biofortification is an underexploited strategy for combatting hidden hunger. *Plant Cell Environ.* (2019) 42:52–70. doi: 10.1111/pce.13368
118. Nöthlings U, Murphy SP, Wilkens LR, Henderson BE, Kolonel LN. Dietary glycemic load, added sugars, and carbohydrates as risk factors for pancreatic cancer: the multiethnic cohort study. *Am J Clin Nutr.* (2007) 86:1495–501. doi: 10.1093/ajcn/86.5.1495
119. Prasanna BM, Palacios-Rojas N, Hossain F, Muthusamy V, Menkir A, Dhliwayo T, et al. Molecular breeding for nutritionally enriched maize: status and prospects. *Front Genet.* (2020) 10:1392. doi: 10.3389/fgene.2019.01392
120. Guzman C, Peña RJ, Singh R, Autrique E, Dreisigacker S, Crossa J, et al. Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Appl Transl Genomics.* (2016) 11:3–8. doi: 10.1016/j.atg.2016.10.004
121. Poland JA, Rife TW. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome.* (2012) 5:92–102. doi: 10.3835/plantgenome2012.05.0005
122. Heffner EL, Jannink JL, Sorrells ME. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome.* (2011) 4:65–75. doi: 10.3835/plantgenome2010.12.0029
123. Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, et al. Genomic selection in plant breeding: knowledge and prospects. *Adv Agron.* (2011) 110:77–123. doi: 10.1016/B978-0-12-385531-2.00002-5
124. Hossain F, Muthusamy V, Pandey N, Vishwakarma AK, Baveja A, Zunjare RU, et al. Marker-assisted introgression of opaque2 allele for rapid conversion of elite hybrids into quality protein maize. *J Genet.* (2018) 97:287–98. doi: 10.1007/s12041-018-0914-z
125. Kumar J, Jaiswal V, Kumar A, Kumar N, Mir R, Kumar S, et al. Introgression of a major gene for high grain protein content in some Indian bread wheat cultivars. *Field Crops Res.* (2011) 123:226–33. doi: 10.1016/j.fcr.2011.05.013
126. Tyagi S, Mir R, Kaur H, Chhuneja P, Ramesh B, Balyan H, et al. Marker-assisted pyramiding of eight QTLs/genes for seven different traits in common wheat (*Triticum aestivum* L.). *Mol Breed.* (2014) 34:167–75. doi: 10.1007/s11032-014-0027-1
127. Vishwakarma MK, Arun B, Mishra V, Yadav P, Kumar H, Joshi AK. Marker-assisted improvement of grain protein content and grain weight in Indian bread wheat. *Euphytica.* (2016) 208:313–21. doi: 10.1007/s10681-015-1598-6
128. Vishwakarma MK, Mishra V, Gupta P, Yadav P, Kumar H, Joshi AK. Introgression of the high grain protein gene Gpc-B1 in an elite wheat variety of Indo-Gangetic plains through marker assisted backcross breeding. *Curr Plant Biol.* (2014) 1:60–7. doi: 10.1016/j.cpb.2014.09.003
129. Randhawa HS, Asif M, Pozniak C, Clarke JM, Graf RJ, Fox SL, et al. Application of molecular markers to wheat breeding in Canada. *Plant Breed.* (2013) 132:458–71. doi: 10.1111/pbr.12057
130. *Wheat Cap Coordinated Agricultural Project.* University of California. Available online at: https://www.triticeaecap.org/t-cap-variety-and-germplasm-releases_2011_2013/ (accessed July 20, 2020)
131. Lau WCP, Latif MA. Current breeding approaches for developing rice with improved grain and nutritional qualities. In: A. M. Iqbal Qureshi, Z. A. Dar, and S. H. Wani, eds. *Quality Breeding in Field Crops.* Srinagar: Springer (2019). p. 199–216.
132. Amarawathi Y, Singh R, Singh AK, Singh VP, Mohapatra T, Sharma TR, et al. Mapping of quantitative trait loci for basmati quality traits in rice (*Oryza sativa* L.). *Mol Breed.* (2008) 21:49–65. doi: 10.1007/s11032-007-9108-8
133. Lau WCP, Rafii MY, Ismail MR, Puteh A, Latif MA, Asfaliza R, et al. Development of advanced fragrant rice lines from MR269× Basmati 370 through marker-assisted backcrossing. *Euphytica.* (2017) 213:11. doi: 10.1007/s10681-016-1794-z
134. Prado JR, Segers G, Voelker T, Carson D, Dobert R, Phillips J, et al. Genetically engineered crops: from idea to product. *Annu Rev Plant Biol.* (2014) 65:769–90. doi: 10.1146/annurev-arplant-050213-040039
135. Gao C. Genome editing in crops: from bench to field. *Natl Sci Rev.* (2015) 2:13–5. doi: 10.1093/nsr/nwu054
136. Kim Y-G, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA.* (1996) 93:1156–60. doi: 10.1073/pnas.93.3.1156
137. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics.* (2010) 186:757–61. doi: 10.1534/genetics.110.120717
138. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science.* (2012) 337:816–21. doi: 10.1126/science.1225829
139. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. *Annu Rev Genet.* (2011) 45:247–71. doi: 10.1146/annurev-genet-110410-132435
140. Voytas DF, Gao C. Precision genome engineering and agriculture: opportunities and regulatory challenges. *PLoS Biol.* (2014) 12:e1001877. doi: 10.1371/journal.pbio.1001877
141. Sánchez-León S, Gil-Humanes J, Ozuna CV, Giménez MJ, Sousa C, Voytas DF, et al. Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol J.* (2018) 16:902–10. doi: 10.1111/pbi.12837
142. Jouanin A, Schaart JG, Boyd LA, Cockram J, Leigh FJ, Bates R, et al. Outlook for coeliac disease patients: towards bread wheat with hypoimmunogenic

gluten by gene editing of α-and γ-gliadin gene families. *BMC Plant Biol.* (2019) 19:333. doi: 10.1186/s12870-019-1889-5
143. Sun Y, Jiao G, Liu Z, Zhang X, Li J, Guo X, et al. Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front Plant Sci.* (2017) 8:298. doi: 10.3389/fpls.2017.00298
144. Shan Q, Zhang Y, Chen K, Zhang K, Gao C. Creation of fragrant rice by targeted knockout of the Os BADH 2 gene using TALEN technology. *Plant Biotechnol J.* (2015) 13:791–800. doi: 10.1111/pbi.12312
145. Abe K, Araki E, Suzuki Y, Toki S, Saika H. Production of high oleic/low linoleic rice by genome editing. *Plant Physiol Biochem.* (2018) 131:58–62. doi: 10.1016/j.plaphy.2018.04.033
146. Endo A, Saika H, Takemura M, Misawa N, Toki S. A novel approach to carotenoid accumulation in rice callus by mimicking the cauliflower Orange mutation via genome editing. *Rice.* (2019) 12:81. doi: 10.1186/s12284-019-0345-3
147. Zhu Y, Lin Y, Chen S, Liu H, Chen Z, Fan M, et al. CRISPR/Cas9-mediated functional recovery of the recessive rc allele to develop red rice. *Plant Biotechnol J.* (2019) 17:2096–105. doi: 10.1111/pbi.13125
148. Pioneer. *DuPont Pioneer's Next Generation of Waxy Corn Shows the Green Side of CRISPR-Cas9.* Feburary 16, 2017. Available online at: https://biovox.eu/dupont-pioneer-rsquo-s-next-generation-of-waxy-corn-shows-the-gr/ (accessed July 16, 2020).
149. Haun W, Coffman A, Clasen BM, Demorest ZL, Lowy A, Ray E, et al. Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol J.* (2014) 12:934–40. doi: 10.1111/pbi.12201
150. Demorest ZL, Coffman A, Baltes NJ, Stoddard TJ, Clasen BM, Luo S, et al. Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biol.* (2016) 16:225. doi: 10.1186/s12870-016-0906-1
151. Al Amin N, Ahmad N, Wu N, Pu X, Ma T, Du Y, et al. CRISPR-Cas9 mediated targeted disruption of FAD2–2 microsomal omega-6 desaturase in soybean (*Glycine max.* L). *BMC Biotechnol.* (2019) 19:9. doi: 10.1186/s12896-019-0501-2
152. Du H, Zeng X, Zhao M, Cui X, Wang Q, Yang H, et al. Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *J Biotechnol.* (2016) 217:90–7. doi: 10.1016/j.jbiotec.2015.11.005
153. Yuan M, Zhu J, Gong L, He L, Lee C, Han S, et al. Mutagenesis of FAD2 genes in peanut with CRISPR/Cas9 based gene editing. *BMC Biotechnol.* (2019) 19:24. doi: 10.1186/s12896-019-0516-8
154. Andersson M, Turesson H, Nicolia A, Fält A-S, Samuelsson M, Hofvander P. Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Rep.* (2017) 36:117–28. doi: 10.1007/s00299-016-2062-3
155. Clasen BM, Stoddard TJ, Luo S, Demorest ZL, Li J, Cedrone F, et al. Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol J.* (2016) 14:169–76. doi: 10.1111/pbi.12370
156. Okuzaki A, Ogawa T, Koizuka C, Kaneko K, Inaba M, Imamura J, et al. CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in *Brassica napus. Plant Physiol Biochem.* (2018) 131:63–9. doi: 10.1016/j.plaphy.2018.04.025
157. Huang H, Cui T, Zhang L, Yang Q, Yang Y, Xie K, et al. Modifications of fatty acid profile through targeted mutation at BnaFAD2 gene with CRISPR/Cas9-mediated gene editing in *Brassica napus. Theor Appl Genet.* (2020) 133:2401–11. doi: 10.1007/s00122-020-03607-y
158. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives. *Nat Rev Endocrinol.* (2012) 8:228–36. doi: 10.1038/nrendo.2011.183
159. Qiu Y, Liu Q, Beta T. Antioxidant properties of commercial wild rice and analysis of soluble and insoluble phenolic acids. *Food Chem.* (2010) 121:140–7. doi: 10.1016/j.foodchem.2009.12.021
160. Furukawa T, Maekawa M, Oki T, Suda I, Iida S, Shimada H, et al. The Rc and Rd genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* (2007) 49:91–102. doi: 10.1111/j.1365-313X.2006.02958.x
161. Sweeney MT, Thomson MJ, Pfeil BE, McCouch S. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell.* (2006) 18:283–94. doi: 10.1105/tpc.105.038430
162. Jiang WZ, Henry IM, Lynagh PG, Comai L, Cahoon EB, Weeks DP. Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. *Plant Biotechnol J.* (2017) 15:648–57. doi: 10.1111/pbi.12663
163. Morineau C, Bellec Y, Tellier F, Gissot L, Kelemen Z, Nogué F, et al. Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa. Plant Biotechnol J.* (2017) 15:729–39. doi: 10.1111/pbi.12671
164. Ozseyhan ME, Kang J, Mu X, Lu C. Mutagenesis of the FAE1 genes significantly changes fatty acid composition in seeds of *Camelina sativa. Plant Physiol Biochem.* (2018) 123:1–7. doi: 10.1016/j.plaphy.2017.11.021
165. Chatrath R, V Tiwari, V Gupta, S Kumar, SK Singh, CN Mishra, et al. WB 2: a high yielding bread wheat variety for irrigated timely sown conditions of North Western plains zone of India. *Wheat Barley Res.* (2018) 10:40–4. doi: 10.25174/2249-4065/2018/81917
166. Singh R, Govindan V, Andersson MS. Zinc-biofortified wheat: harnessing genetic diversity for improved nutritional quality. Science Brief: Biofortification No. 1 (May 2017). CIMMYT, HarvestPlus, and the Global Crop Diversity Trust. Bonn, Germany. (2017).
167. Gupta A, Singh C, Kumar V, Tyagi B, Tiwari V, Chatrath R, et al. *Wheat Varieties Notified in India Since 1965.* ICAR-Indian Institute of Wheat & Barley Research, Karnal (2018).
168. Velu G, Singh R, Balasubramaniam A, Mishra VK, Chand R, Tiwari C, et al. Reaching out to farmers with high zinc wheat varieties through public-private partnerships: an experience from eastern-gangetic plains of India. *Adv Food Technol Nutr Sci.* (2015) 3:73–5. doi: 10.17140/AFTNSOJ-1-112
169. Yadav R, Gaikwad KB, Singh GP, Kumar M, Singh PK, Sai Prasad SV, et al. HD 3226 Notification of crop varieties and registration of germplasm. *Indian J Genet Plant Breed.* (2019) 79:632–3.
170. Ravish C, Tiwari V, Singh G, Tiwari R, Tyagi BS. ICAR-Indian Institute of Wheat and Barley Research, Karnal, Haryana, India. *ICAR-IIWBR Progress Report of AICRP on Wheat Barley 2017-18, Crop Improvement.* (2018). p. 206.
171. Martinek P, Škorpík M, Chrpova J, Schweiger JFP. Development of the new winter wheat variety Skorpion with blue grain. *Czech J Genet Plant Breed.* (2013) 49:90–4. doi: 10.17221/7/2013-CJGPB
172. Hanková A, Matúšková K, Rückschloss, L, Žofajová A. Winter wheat Ps Karkulka. *Agriculture.* (2014) 60:159–61.
173. Garg M, Chawla M, Chunduri V, Kumar R, Sharma S, Sharma NK, et al. Transfer of grain colors to elite wheat cultivars and their characterization. *J Cereal Sci.* (2016) 71:138–44. doi: 10.1016/j.jcs.2016.08.004
174. Eticha F, Grausgruber H, Siebenhandl-Ehn S, Berghofer E. Some agronomic and chemical traits of blue aleurone and purple pericarp wheat (Triticum L.). *J Agric Sci Technol B.* (2011) 1:48–58.
175. Bai Y, Li W, Sun S, Sun Y, Pei Z, Corke H. Evaluation on some quality characteristics of black-grained wheat 76. *J Chin Cereal Oil Assoc.* (2000) 15:6–9.
176. Gregorio GB, Senadhira D, Htut H, Graham RD. Breeding for trace mineral density in rice. *Food Nutr Bull.* (2000) 21:382–6. doi: 10.1177/156482650002100407
177. HarvestPlus. *Bangladesh Releases New, Improved Zinc Rice Variety.* (2015). Available online at: https://www.harvestplus.org/knowledge-market/in-the-news/bangladesh-releases-new-improved-zinc-rice-variety-0 (accessed October 07, 2020).
178. Yadava D, Choudhury P, Hossain F, Kumar D. *Biofortified Varieties: Sustainable way to Alleviate Malnutrition.* New Delhi: Indian Council of Agricultural Research (2017).
179. Gupta HS, Hossain F, Muthusamy V. Biofortification of maize: an Indian perspective. *Indian J Genet.* (2015) 75:1–22. doi: 10.5958/0975-6906.2015.00001.2
180. CIMMYT. *Pakistan Releases First Quality Protein Maize Varieties.* (2017). Available online at: https://www.cimmyt.org/news/pakistan-releases-first-quality-protein-maize-varieties/ (accessed July 10, 2020).
181. Vasal S. Quality protein maize development: an exciting experience. In: D. K. Friesen and A.F.E. Palmer eds. *Integrated Approaches to Higher Maize Productivity in the New Millennium.* Nairobi: CIMMYT and KARI (Kenya Agricultural Research Institute) (2002) 3–6.

182. Sofi P, Wani SA, Rather A, Wani SH. Quality protein maize (QPM): genetic manipulation for the nutritional fortification of maize. *J Plant Breed Crop Sci.* (2009) 1:244–53. doi: 10.5897/JPBCS.9000113
183. Chaudhary D, Kumar S, Yadav O. Nutritive value of maize: improvements, applications and constraints. In: *Maize: Nutrition Dynamics and Novel Uses.* New Delhi: Springer. (2014). p. 3–17.
184. Zunjare RU, Hossain F, Muthusamy V, Baveja A, Chauhan HS, Bhat JS, et al. Development of biofortified maize hybrids through marker-assisted stacking of β-carotene hydroxylase, lycopene-ε-cyclase and opaque2 genes. *Front Plant Sci.* (2018) 9:178. doi: 10.3389/fpls.2018.00178
185. Andersson MS, Saltzman A, Virk PS, Pfeiffer WH. Progress update: crop development of biofortified staple food crops under HarvestPlus. *Afr J Food Agric Nutr Dev.* (2017) 17:11905–35. doi: 10.18697/ajfand.78.HarvestPlus05
186. CIMMYT. *CIMMYT-SARO Steps Up Efforts in War Against Malnutrition.* (2015). Available online at: https://www.cimmyt.org/news/cimmyt-saro-steps-up-efforts-in-war-against-malnutrition/ (accessed July 10, 2020).
187. CIMMYT. Available online at: https://www.cimmyt.org/news/first-zinc-maize-variety-launched-to-reduce-malnutrition-in-colombia/ (accessed July 10, 2020).
188. ICRISAT. Retrieved from: http://www.icrisat.org/africas-first-biofortified-pearl-millet-variety-aims-to-combat-anaemia/ (accessed July 16, 2020).
189. ICRISAT. *India Gets Its First Biofortified Sorghum*. (2018). Retrieved from: https://www.icrisat.org/india-gets-its-first-biofortified-sorghum/ (accessed July 16, 2020).
190. ICRISAT. *New Sorghum Varieties Released in Nigeria With Higher Iron, Yields and Drought Resilience.* (2016). Available online at: https://www.icrisat.org/new-sorghum-varieties-will-fight-malnutrition-and-climate-change-in-sudan-and-sahel-region-of-nigeria/ (accessed July 10, 2020).
191. HarvestPlus. *Uganda Releases Biofortified Beans to Address Iron Deficiency Anemia.* (2016). Available online at: https://www.harvestplus.org/knowledge-market/in-the-news/uganda-releases-biofortified-beans-address-iron-deficiency-anemia (accessed July 10, 2020).
192. HarvestPlus. *RWANDA Country Brief*. Available online at: https://www.harvestplus.org/sites/default/files/HarvestPlus_CountryBrief_Rwanda.pdf (accessed July 10, 2020).
193. HarvestPlus. *Democratic Republic of Congo Country Brief*. Available online at: www.harvestplus.org
194. Mulambu J, Andersson M, Palenberg M, Pfeiffer W, Saltzman A, Birol E, et al. Iron beans in Rwanda: crop development and delivery experience. *Afr J Food Agric Nutr Dev.* (2017) 17:12026–50. doi: 10.18697/ajfand.78.HarvestPlus10
195. ICARDA. *Food From Fallows: Capitalizing on Idle Land for Better Food Security in South Asia*. Retrieved from: https://www.icarda.org/project/alleviating-malnutrition-biofortified-lentil (accessed July 16, 2020).
196. Sarker A, Erskine W, Abu Bakr M, Matiur Rahman M, Ali Afzal M, Saxena MC. Lentil improvement in Bangladesh. *A Success Story of Fruitful Partnership Between the Bangladesh Agricultural Research Institute and the International Center for Agricultural Research in the dry Areas*. APAARI Publication (2004). p. 1–38.
197. ICARDA. *'BARIMasur-8': A New Micronutrient-Rich Lentil Variety Released in Bangladesh*. Available online at: https://www.icarda.org/media/news/barimasur-8-new-micronutrient-rich-lentil-variety-released-bangladesh (accessed July 11, 2020).
198. ICARDA. *New Micronutrient Rich Lentil Varieties Released in Nepal and India*. Available online at: https://www.icarda.org/media/news/new-micronutrient-rich-lentil-varieties-released-nepal-and-india (accessed July 11, 2020).
199. ICARDA. *Pulses for Food Security, Nutrition and Environment: the Role of Science & Technology to Enhance Productivity and Production of Pulses.* (2015). Available online at: https://iyp2016.org/resources/documents/related-documents/72-iyp-launch-dr-m-solh-presentation/file (accessed July 11, 2020).
200. Yadav DK, Hossain Firoz, Choudhary PR, Kumar Dinesh, Singh AK, Mohapatra T. *Crop Cultivars Developed Through Molecular Breeding.* New Delhi: Indian Council of Agricultural Research (2020). p. 52.
201. Carpentieri-Pipolo V. UEL 175: a novel lipoxygenase-free soybean cultivar with kunitz trypsin inhibitor absence. *Crop Breed Appl Biotechnol.* (2015) 15:191–2. doi: 10.1590/1984-70332015v15n3c33
202. Available online at: https://iisrindore.icar.gov.in/pdfdoc/NRC142News.pdf (accessed July 27, 2020).
203. Singh SP. *Common Bean Improvement in the Twenty-First Century*. Springer Science & Business Media (2013).
204. Erskine W. Lessons for breeders from land races of lentil. *Euphytica.* (1997) 93:107–12. doi: 10.1023/A:1002939704321
205. Maranna S, Verma K, Talukdar A, Lal SK, Kumar A, Mukherjee K. Introgression of null allele of Kunitz trypsin inhibitor through marker-assisted backcross breeding in soybean (*Glycine max* L. Merr.). *BMC Genet.* (2016) 17:106. doi: 10.1186/s12863-016-0413-2
206. Kumar V, Rani A, Rawal R, Mourya V. Marker assisted accelerated introgression of null allele of kunitz trypsin inhibitor in soybean. *Breed Sci.* (2015) 65:447–52. doi: 10.1270/jsbbs.65.447
207. Lyons G, Ortiz-Monasterio I, Stangoulis J, Graham R. Selenium concentration in wheat grain: is there sufficient genotypic variation to use in breeding? *Plant Soil.* (2005) 269:369–80. doi: 10.1007/s11104-004-0909-9
208. Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, et al. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science.* (2000) 287:303–5. doi: 10.1126/science.287.5451.303
209. Anacleto R, Badoni S, Parween S, Butardo Jr VM, Misra G, Cuevas RP, et al. Integrating a genome-wide association study with a large-scale transcriptome analysis to predict genetic regions influencing the glycaemic index and texture in rice. *Plant Biotechnol J.* (2019) 17:1261–75. doi: 10.1111/pbi.13051
210. Wu G, Truksa M, Datla N, Vrinten P, Bauer J, Zank T, et al. Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat Biotechnol.* (2005) 23:1013–7. doi: 10.1038/nbt1107
211. Ruiz-Lopez N, Haslam RP, Napier JA, Sayanova O. Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. *Plant J.* (2014) 77:198–208. doi: 10.1111/tpj.12378
212. Branham SE, Wright SJ, Reba A, Linder CR. Genome-wide association study of Arabidopsis thaliana identifies determinants of natural variation in seed oil composition. *J Hered.* (2016) 107:248–56. doi: 10.1093/jhered/esv100
213. Kim HJ, Silva JE, Vu HS, Mockaitis K, Nam J-W, Cahoon EB. Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in Camelina seeds. *J Exp Bot.* (2015) 66:4251–65. doi: 10.1093/jxb/erv225
214. Nguyen HT, Park H, Koster KL, Cahoon RE, Nguyen HT, Shanklin J, et al. Redirection of metabolic flux for high levels of omega-7 monounsaturated fatty acid accumulation in camelina seeds. *Plant Biotechnol J.* (2015) 13:38–50. doi: 10.1111/pbi.12233

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Gaikwad, Rani, Kumar, Gupta, Babu, Bainsla and Yadav. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership