Genome-Wide Scans for Delineation of Candidate Genes Regulating Seed-Protein Content in Chickpea

Hari D. Upadhyaya[1†], Deepak Bajaj[2†], Laxmi Narnoliya[2], Shouvik Das[2], Vinod Kumar[3], C. L. L. Gowda[1], Shivali Sharma[1], Akhilesh K. Tyagi[2] and Swarup K. Parida[2]*

[1] International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India, [2] National Institute of Plant Genome Research, New Delhi, India, [3] National Research Centre on Plant Biotechnology, New Delhi, India

Identification of potential genes/alleles governing complex seed-protein content (SPC) is essential in marker-assisted breeding for quality trait improvement of chickpea. Henceforth, the present study utilized an integrated genomics-assisted breeding strategy encompassing trait association analysis, selective genotyping in traditional bi-parental mapping population and differential expression profiling for the first-time to understand the complex genetic architecture of quantitative SPC trait in chickpea. For GWAS (genome-wide association study), high-throughput genotyping information of 16376 genome-based SNPs (single nucleotide polymorphism) discovered from a structured population of 336 sequenced *desi* and *kabuli* accessions [with 150–200 kb LD (linkage disequilibrium) decay] was utilized. This led to identification of seven most effective genomic loci (genes) associated [10–20% with 41% combined PVE (phenotypic variation explained)] with SPC trait in chickpea. Regardless of the diverse *desi* and *kabuli* genetic backgrounds, a comparable level of association potential of the identified seven genomic loci with SPC trait was observed. Five SPC-associated genes were validated successfully in parental accessions and homozygous individuals of an intra-specific *desi* RIL (recombinant inbred line) mapping population (ICC 12299 × ICC 4958) by selective genotyping. The seed-specific expression, including differential up-regulation (>four fold) of six SPC-associated genes particularly in accessions, parents and homozygous individuals of the aforementioned mapping population with a high level of contrasting SPC (21–22%) was evident. Collectively, the integrated genomic approach delineated diverse naturally occurring novel functional SNP allelic variants in six potential candidate genes regulating SPC trait in chickpea. Of these, a non-synonymous SNP allele-carrying zinc finger transcription factor gene exhibiting strong association with SPC trait was found to be the most promising in chickpea. The informative functionally relevant molecular tags scaled-down essentially have potential to accelerate marker-assisted genetic improvement by developing nutritionally rich chickpea cultivars with enhanced SPC.

Keywords: chickpea, GBS, GWAS, QTL, seed-protein content, SNP

INTRODUCTION

Chickpea (*Cicer arietinum* L.) is the second most consumed important food legume crop globally and stands third in production among all legume crops world-wide (Kumar et al., 2011; FAOSTAT: Production-Crops, 2012; Varshney et al., 2013b). The chickpea seeds are considered highly nutritious and largely cultivated in arid and semi-arid regions of the world (Kumar et al., 2011; Varshney et al., 2013b). The seeds of chickpea serve as a staple source of protein in human diets. The seed-protein content (SPC) is thus an extremely vital component of nutrition for both humans and animals (Millerd, 1975; Monti and Grillo, 1983; Duranti and Gius, 1997; Kumar et al., 2011; Varshney et al., 2013b; Jadhav et al., 2015). Protein malnutrition, a commonly observed problem in infants/young children, especially of developing countries can be addressed by adequate intake of diet enriched with high protein content (Monti and Grillo, 1983; Haider and Haider, 1984; Duranti and Gius, 1997; World Health Organization [WHO], 2013). Besides, the chickpea seeds being significantly rich in almost all essential amino acids (except sulfur-containing types) can be consumed along with cereals as a daily human diet to mitigate the problem of protein-calorie malnutrition (Monti and Grillo, 1983; Haider and Haider, 1984; Duranti and Gius, 1997; Jukanti et al., 2012; World Health Organization [WHO], 2013). Considering the importance of SPC in global food as well as nutritional security, improving the protein content and quality in the most consumed part (seeds) of chickpea is now the most challenging task in molecular breeding and genomics research. The SPC is an important quality component agronomic trait in chickpea. The protein content in the chickpea seeds is considered the best among all legume proteins that significantly varies (17–22% of total dry seed mass) across core and mini-core germplasm lines, landraces, and cultivated *desi* and *kabuli* accessions (Upadhyaya and Ortiz, 2001; Upadhyaya et al., 2002, 2006; Jukanti et al., 2012; Jadhav et al., 2015). However, the proportionate storage and accumulation of protein content in chickpea seeds is a quantitative trait and regulated by many major genes/QTLs (quantitative trait loci) through a complex molecular genetic mechanism (Frimpong et al., 2009; Jadhav et al., 2015). Henceforth, it would be interesting to decipher the genetic factors and functional gene regulatory mechanism underlying SPC variation in chickpea.

The genomes of *desi* (∼520 of ∼740 Mb estimated genome size) and *kabuli* (sim544 Mb) chickpea cultivars with contrasting agro-morphological features representing diverse genepools, have been sequenced (Jain et al., 2013; Varshney et al., 2013a; Parween et al., 2015). These available genomic resources have assisted in NGS (next-generation sequencing)-based genome/transcriptome sequencing of a diverse array of *desi*, *kabuli*, and wild chickpea cultivars. These efforts further accelerated development of numerous informative genomic/genic microsatellite and single nucleotide polymorphism (SNP) markers and their high-throughput genotyping in numerous natural and mapping populations by

NGS-/array-based assay at a genome-wide scale in chickpea (Garg et al., 2011, 2014; Agarwal et al., 2012; Hiremath et al., 2012; Jhanwar et al., 2012; Singh et al., 2013, 2014; Kudapa et al., 2014; Pradhan et al., 2014; Parida et al., 2015). These inputs are known to expedite various high-throughput genetic analysis, including construction of high-resolution genetic linkage maps, molecular mapping of QTLs/genes and trait association analysis in chickpea (Nayak et al., 2010; Gujaria et al., 2011; Thudi et al., 2011, 2014; Gaur et al., 2012; Hiremath et al., 2012; Kujur et al., 2013, 2014, 2015a,b,c; Saxena et al., 2014a,b; Bajaj et al., 2015a,b,c; Das et al., 2015). Most of these genomics-assisted breeding efforts have uncovered informative markers tightly linked to the major QTLs/genes controlling abiotic (drought and salinity tolerance)/biotic (*Ascochyta* blight and *Fusarium* wilt resistance) stress tolerance and yield-contributing traits for marker-assisted genetic improvement of chickpea (Kujur et al., 2013, 2014, 2015a,c; Sabbavarapu et al., 2013; Saxena et al., 2014b; Thudi et al., 2014; Varshney et al., 2014a,b,c; Bajaj et al., 2015b,c; Das et al., 2015). However, no comprehensive efforts have been made so far to understand the genetic inheritance pattern as well as to identify the functionally relevant genetic loci governing SPC in chickpea. To decipher the complex genetic architecture of SPC, tremendous progress has been made in the area of genomics-assisted breeding by identification (fine mapping/map-based isolation and genetic association mapping) and introgression of useful genes/QTLs governing this vital quality trait in multiple crop plants, including cereals and legumes (Bourgeois et al., 2011; Krajewski et al., 2012; Lestari et al., 2013; Lu et al., 2013; Hwang et al., 2014; Peng et al., 2014; Jadhav et al., 2015; Wang et al., 2015; Warrington et al., 2015). However, very limited attention has been paid toward identification of QTLs/genes regulating SPC, which can be deployed in marker-assisted genetic improvement of chickpea. Only one such preliminary attempt has been made recently to identify SSR marker loci associated with SPC through 23 microsatellite marker-based trait association mapping strategy in chickpea (Jadhav et al., 2015). The added-advantage and broader applicability of integrated genomic approach combining marker-based genome-wide association study (GWAS), QTL mapping, and differential gene expression profiling in rapid quantitative dissection of complex yield component traits is well documented in chickpea (Kujur et al., 2013, 2014, 2015a,b,c; Saxena et al., 2014a; Bajaj et al., 2015b,c). Therefore, this integrated approach can be employed in natural/mapping populations for quantitative dissection of a quality component trait like SPC, with an ultimate aim of marker-assisted genetic enhancement to develop high protein-rich cultivars of chickpea.

In this context, our study utilized genome-wide SNPs discovered by genotyping-by-sequencing (GBS) of natural germplasm accessions in GWAS to identify potential genomic loci (genes) governing SPC in chickpea. Large-scale validation of these trait-associated genomic loci/genes in natural and mapping populations was performed by selective genotyping and differential expression profiling to delineate functionally relevant allelic variants in the candidate genes regulating SPC in chickpea.

MATERIALS AND METHODS

Genotyping of Genome-Wide SNPs

For discovery and high-throughput genotyping of genome-wide SNPs, 336, including 92 (39 *desi* and 43 *kabuli*) and 244 (167 *desi* and 77 *kabuli*) diverse accessions from the available chickpea germplasm collection were screened as per Kujur et al. (2014, 2015a,b). The genotyping as well as structural/functional annotation information of *kabuli* reference genome-based SNPs mined from the sequenced 92 diverse chickpea accessions using a GBS assay were acquired (Kujur et al., 2015b). Large-scale gene-based SNPs selected from afore-said GBS data were further genotyped in the genomic DNA of 244 accessions using Illumina GoldenGate SNP genotyping assay (following Bajaj et al., 2015c). The SNP genotyping information generated among accessions was utilized further for trait association mapping.

Phenotyping for Seed-Protein Content

A set of 336, including 92 and 244 chickpea accessions were grown in the field (randomized complete block design with two replications) during crop season for two consecutive years (2012 and 2013) at two diverse geographical locations (New Delhi: 28° 4′ N/77° 2′ E and Patancheru, Hyderabad: latitude 17° 3′ N/longitude 77° 2′ E) of India. The mature seeds of each accession (2–4 representative plants from each accession) were harvested and collected individually. To estimate the protein content, healthy (free from dust and metal particles contamination) mature seeds (20 g) of chickpea accessions were analyzed at Central Analytical Services Laboratory, ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), Patancheru, Hyderabad. Before grinding, the seed samples from each accession were washed with distilled water and oven-dried at 60°C for 48 h. The dried-seed samples (20 g) were powdered in a mill with Teflon chambers. The ground samples were further kept overnight in an oven at 60°C for drying. The digestion of the standards and samples was performed simultaneously with appropriate blanks in duplicates (two independent analyses). One g of ground sample was transferred to a digestion tube of 75 ml capacity containing 10 ml of tri-acid mixture of nitric, sulfuric and perchloric acid in the ratio of 10:0.5:2 (v/v). The contents were left overnight in a digestion chamber for cold digestion. To get clear and colorless digests, the samples were digested initially at 120°C for 1 h following digestion at 230°C for about 2 h. With the subsequent cooling of the digests, the contents were dissolved in distilled water and volume made up to 75 ml, and shaken well. Aliquots were retrieved from the digests and the protein content (% of total dry seed weight) of seed samples was determined in the digest using an Autoanalyzer (Singh and Jambunathan, 1980). The quality of SPC of accessions was estimated on a plot basis and their mean values were used for statistical analysis. The broad-sense heritability (H^2), coefficient of variation (CV), analysis of variance (ANOVA), and frequency distribution of SPC among accessions were estimated using SPSS v17.0 as per Bajaj et al. (2015c).

Association Mapping

The phylogenetic tree and population genetic structure information generated from 92 and 244 chickpea accessions were acquired from the earlier study of Kujur et al. (2014, 2015a,b), respectively. The principal component analysis (PCA) among accessions was performed using TASSEL v5.0[1] and GAPIT following Bajaj et al. (2015a). To determine genome-wide linkage disequilibrium (LD) decay, the genome-wide genotyping data of SNPs physically mapped on eight *kabuli* chromosomes were analyzed with PLINK and the full-matrix approach of TASSEL v5.0 as per Kujur et al. (2015b). Consequently, the genome-wide LD decay by plotting average r^2 (frequency correlation among pair of alleles across a pair of SNP loci) against 50 kb uniform physical intervals across chromosomes was determined. The genome-wide SNP genotyping and SPC phenotyping information along with population structure ancestry coefficient (Q), kinship matrix (K), and PCA (P) data of accessions were analyzed by mixed model (P+K, K, and Q+K)-based CMLM (compressed mixed linear model) approach of GAPIT (as per Kujur et al., 2015a and Kumar et al., 2015). The relative distribution of observed and expected −$\log_{10}(P)$-value in each trait-associated genomic loci was compared based on quantile-quantile plot. The corrections of these adjusted *P*-value threshold of significance for multiple comparisons were performed by false discovery rate (FDR cut-off ≤ 0.05) to determine the accuracy and robustness of SNP marker-trait association. The genomic (gene-derived) SNP loci exhibiting significant association with SPC at lowest FDR adjusted *P*-values (threshold $P < 1 \times 10^{-7}$) and highest R^2 (degree of SNP marker-trait association) were identified in chickpea.

Validation of SPC-Associated SNPs in a Bi-Parental Mapping Population

To ascertain the potential of detected genomic SNP loci for SPC trait association, these SNPs were targeted to validate in a traditional bi-parental intra-specific F_7 RIL (recombinant inbred line) mapping population [*C. arietinum desi* accession ICC 12299 (SPC: 16.5%) × *C. arietinum desi* accession ICC 4958 (SPC: 21.1%)] contrasting for SPC trait. For this, two parental accessions and five of each homozygous mapping individuals with contrasting level of low and high SPC (%) were selected for DNA isolation. The SPC-associated SNPs showing polymorphism between the parents of a mapping population were further genotyped in the above-mentioned homozygous mapping individuals using MALDI-TOF MS (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry) SNP genotyping assay as per Saxena et al. (2014a,b). The presence of SPC-associated SNP allelic variants in low and high SPC-containing homozygous mapping individuals and parental accessions was determined to validate the trait association potential of target genomic loci.

Expression Profiling

To assess the regulatory pattern of SNPs-carrying genes associated (validated by GWAS and selective genotyping in

---

[1]http://www.maizegenetics.net

mapping population) with SPC, the differential expression profiling of these genes was performed using quantitative RT-PCR assay. The leaves as well as early cell division [10–20 days after podding (DAP)] and late seed maturation (21–30 DAP) developmental stages (as defined by Bajaj et al., 2015b) of four chickpea accessions as well as parents and four homozygous RIL individuals of a mapping population (ICC 12299 × ICC 4958) with contrasting levels of low and high SPC (%) were used for RNA isolation. RNA isolated from three independent biological replicates of each sample and two technical replicates of each biological replicate with no template and primer as control were used in the quantitative RT-PCR assay. The quality of purified RNA was tested by denaturing agarose gel-based assay and NANODROP 2000 Spectrophotometer (Thermo Scientific, NanoDrop products, USA). One microgram of high quality total RNA was utilized for cDNA synthesis using first strand cDNA synthesis kit (Applied Biosystems, USA). The cDNA (1:100 dilution) along with 1X Fast SYBR Green Master Mix (Applied Biosystems) and 200 nM of forward and reverse gene-specific primers (Supplementary Table S1) in a total reaction volume of 10 μl were amplified in 7500 Fast Real-Time PCR system (Applied Biosystems). The differential expression profile of genes among tissues/stages of accessions, mapping parents and individuals was determined following Bajaj et al. (2015b). An internal control gene elongation factor 1-alpha ($EF1\alpha$) exhibiting consistent expression across diverse tissues and developmental stages of chickpea accessions (Garg et al., 2010) was utilized for normalization of expression value. Significant difference in gene expression at two seed developmental stages of low and high SPC-containing accessions as compared to leaf (considered as reference calibrator and assigned as 1) was determined by LSD-ANOVA significance test using SPSS 17.0[2]. The significant difference in gene expression was visualized by a heat map using MultiExperiment Viewer (MeV[3]).

RESULTS

GWAS of Seed-Protein Content in Chickpea

For GWAS, 2 years replicated phenotyping data of SPC (% of total dry seed weight) measured for 336 (including 92 and 244 *desi* and *kabuli* accessions) chickpea accessions (association panel) were analyzed with diverse statistical measures to infer the genetic inheritance pattern of traits under study (**Table 1**). This
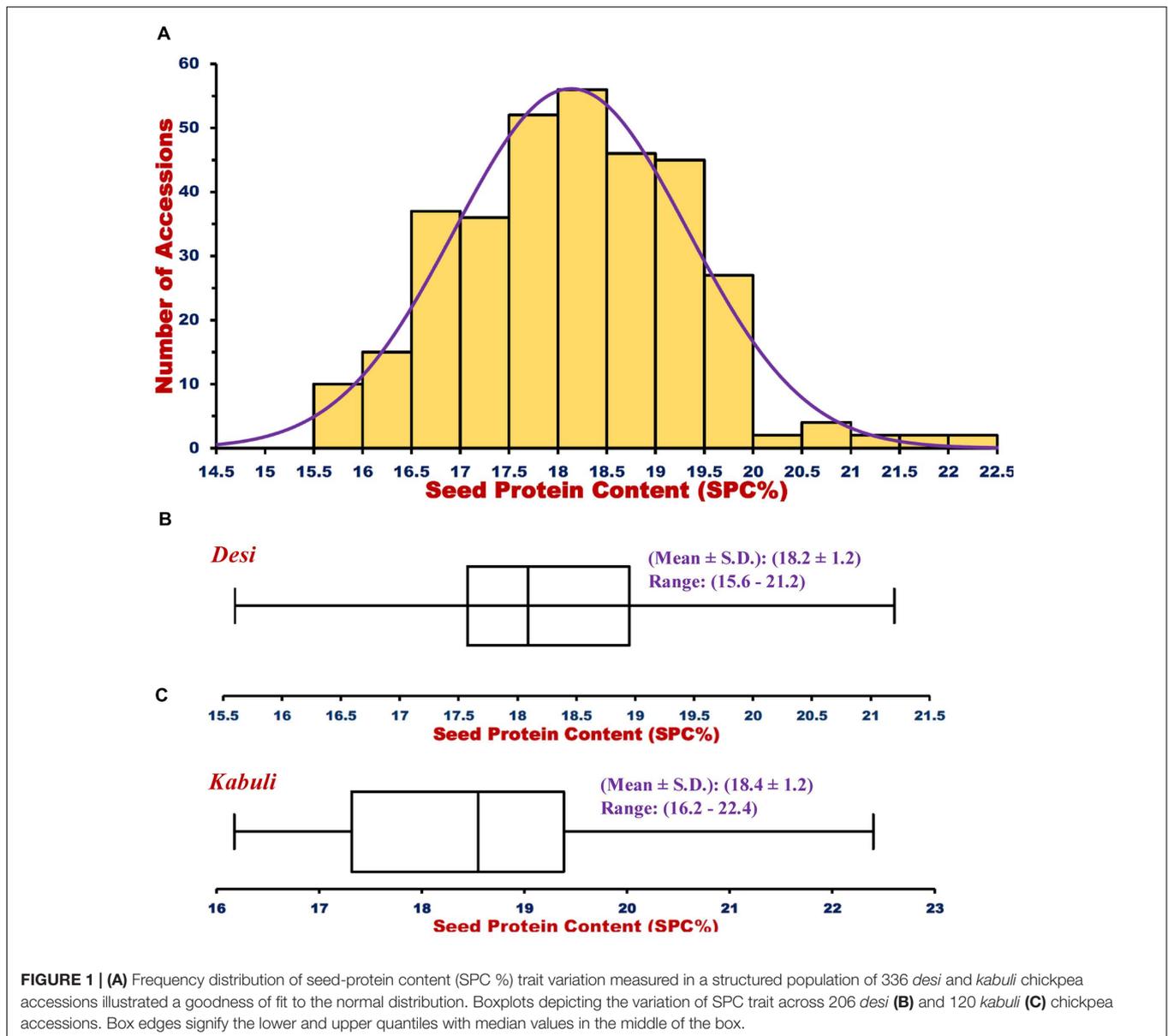
[2]http://www.spss.com/statistics

[3]http://www.tm4.org/mev.html

revealed a wider phenotypic variation [15.6–22.4% with a mean (17.6) $\pm$ SD (1.2) and 80% H^2] along with normal frequency distribution of SPC among chickpea accessions (**Figure 1**, **Table 1**). However, almost a similar level of phenotypic variation for SPC in 206 *desi* (15.6–21.2% with 18.2 $\pm$ 1.2) and 120 *kabuli* (16.2–22.4% with 18.4 $\pm$ 1.2) accessions was observed (**Figure 1**). To perform GWAS, the high-throughput genotyping data of 16376 reference genome-based GBS-SNPs generated from 92 sequenced and genotyped accessions (by GBS assay) were utilized. This included genotyping information of 3072 SNPs generated by their high-throughput genotyping in 244 accessions using Illumina GoldenGate assay (Supplementary Table S2). Of these, 14115 and 2261 SNPs were physically mapped on eight chromosomes and unanchored scaffolds of *kabuli* genome, respectively. The neighbor-joining phylogenetic tree, high-resolution population genetic structure, and PCA classified 336, including 92 and 244 *desi* and *kabuli* chickpea accessions into two distinct population groups (POP I and POP II; Kujur et al., 2014, 2015b) (**Figure 2A**). The determination of LD pattern (r^2) based on genotyping information of 14115 genome-wide SNPs (mapped on chromosomes) in 336 accessions exhibited 150–200 kb LD decay across chromosomes (**Figure 2B**). A non-linear regression curve exhibiting a decreasing trend of LD decay with an increase in the physical distance (kb) was observed. All these accessions sustained a significant level of LD ($r^2 \geq 0.1$) up to a physical distance of 1000 kb. However, r^2 decreased to half of its maximum value at approximately 150–200 kb physical distance in chickpea chromosomes.

The CMLM model-based GWAS in 336 accessions using 16376 genome-wide GBS-SNPs identified seven maximum effect genomic loci (gene-based SNPs) revealing significant association ($P \leq 10^{-8}$) with SPC of chickpea (**Figure 3A**, **Table 2**). The association potential of these seven genomic SNP loci with SPC trait was further ascertained by quantile–quantile plot analysis of their observed and expected $-\log_{10}(P)$-value at a FDR cut-off ≤ 0.05 (**Figure 3B**). These seven SPC-associated SNPs were physically mapped on five *kabuli* chromosomes (except chromosomes 3, 5, and 8). A maximum of two SPC-associated genomic loci were mapped on each of the chromosomes 1 and 7. Four and two of seven SPC-associated genomic loci was represented from diverse coding (two of each synonymous and non-synonymous SNPs) and non-coding (two intronic) sequence components of six genes, respectively (**Figure 3A**, **Table 2**). The remaining one genomic SNP was derived from the intergenic region of *kabuli* genome. The proportion of phenotypic variation for SPC explained (R^2) by maximum effect seven genomic SNP loci among 336 *desi* and *kabuli* chickpea accessions varied from 10 to 20% (P: 1.7×10^{-8} to 1.2×10^{-9}; **Table 2**). The phenotypic

TABLE 1 | Diverse statistical measures estimated for seed-protein content variation observed in a chickpea association panel.

Years	336 *desi* and *kabuli* accessions				
	(Mean ± *SD*)	Range (% of total dry seed weight)	Heritability (H[2]%)	Coefficient of variation (CV%)	ANOVA significance (*P*)
2012	18.1 ± 1.2	15.6–22.4	80	6.6	<0.0001
2013	17.6 ± 1.5	15.9–22.1	81	8.5	<0.0001

FIGURE 1 | (A) Frequency distribution of seed-protein content (SPC %) trait variation measured in a structured population of 336 *desi* and *kabuli* chickpea accessions illustrated a goodness of fit to the normal distribution. Boxplots depicting the variation of SPC trait across 206 *desi* **(B)** and 120 *kabuli* **(C)** chickpea accessions. Box edges signify the lower and upper quantiles with median values in the middle of the box.

variation for SPC explained (R^2) by all significant seven SNPs in combination was 41%. Notably, one non-synonymous SNP-containing zinc finger transcription factor gene exhibited strong association (R^2: 20% at P: 1.2×10^{-9}) with SPC in chickpea (**Table 2**). This was followed by another high SPC-associated (R^2: 17% at P: 2.5×10^{-9}) non-synonymous SNPs-containing candidate gene that encode cytidine (CMP) and deoxycytidylate (dCMP) deaminases. To determine the effect of diverse *desi* and *kabuli* genetic backgrounds in SPC trait association, GWAS was performed by selecting genotyping, phenotyping, population genetic structure and PCA data of 206 *desi* and 120 *kabuli* chickpea accessions independently. A similar level (R^2 and P-values) of association potential for seven genomic loci with SPC was observed irrespective of any genetic backgrounds in chickpea.

Validation of SPC-Associated Genomic Loci in an Intra-specific Mapping Population

To ascertain the association potential of GWAS-derived genomic loci for SPC trait, seven SPC-associated SNPs revealing parental polymorphism were genotyped in 10 of each low (15.1–16.3%) and high (21.0–22.2%) SPC-containing homozygous individuals derived from an intra-specific F_7 RIL mapping population (ICC 12299 × ICC 4958). Five SNPs in the candidate genes (encode ATP-dependent RNA helicase DEAD-box, cystathionine-beta synthase, CMP and dCMP deaminases, and G10 and zinc finger protein) associated with SPC, were also validated in a chickpea mapping population with contrasting level of SPC by selective genotyping. Interestingly, the presence of identical low and high SPC-associated alleles derived from
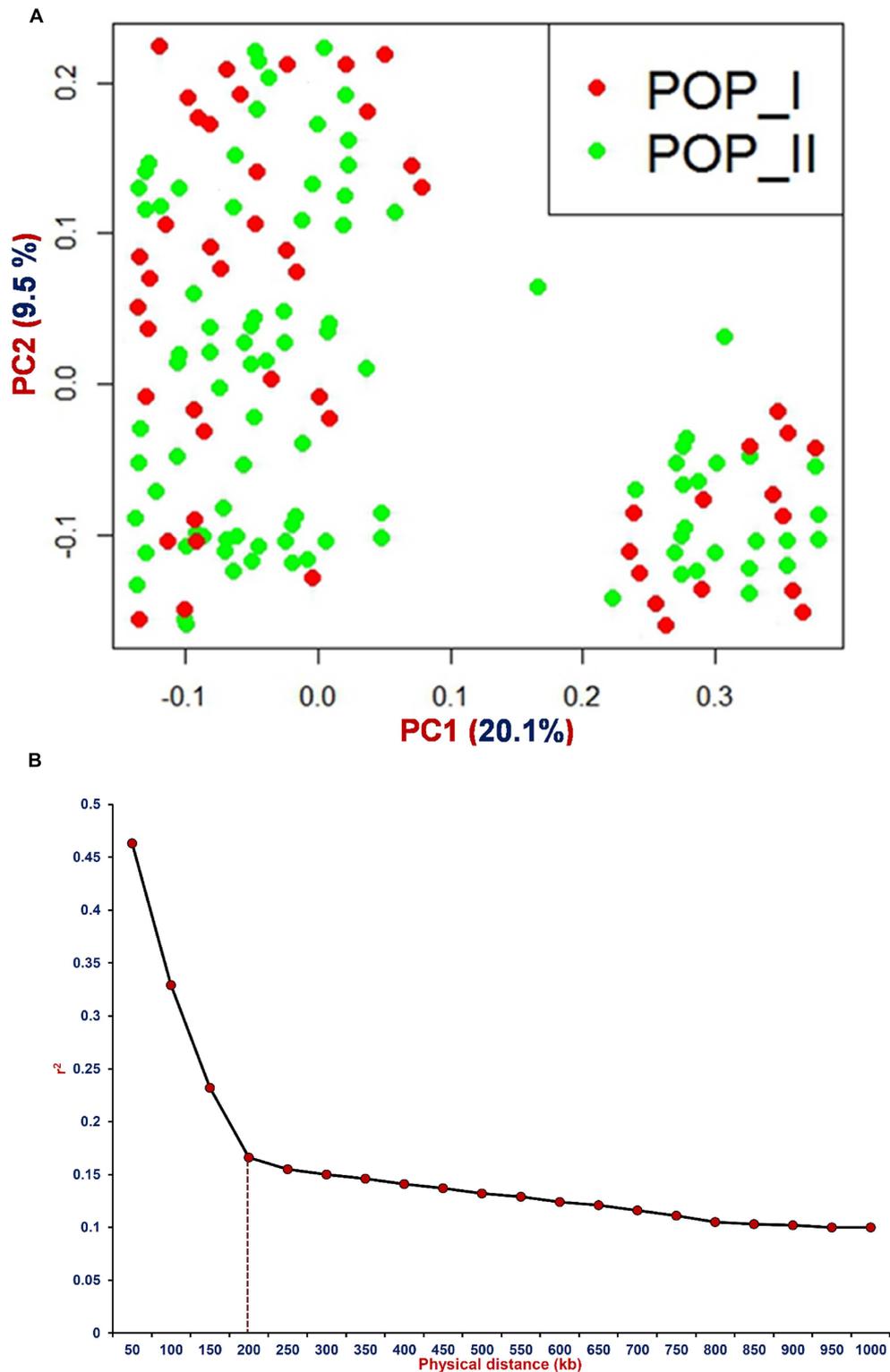
FIGURE 2 | The determination of (A) principal component analysis (PCA) employing 16376 genome-wide GBS-SNPs classified 336 *desi* and *kabuli* chickpea accessions into two major populations [POP I (red) and POP II (green)]. In PCA, the PC1 and PC2 explained 20.1 and 9.5% of the total variance, respectively. **(B)** LD decay (mean r^2) measured in a population of 336 chickpea accessions using 14115 chromosome-wise physically mapped SNPs. The plotted curved line denotes the average r^2 values among SNPs spaced with uniform 50 kb physical intervals from 0 to 1000 kb across *kabuli* chromosomes.
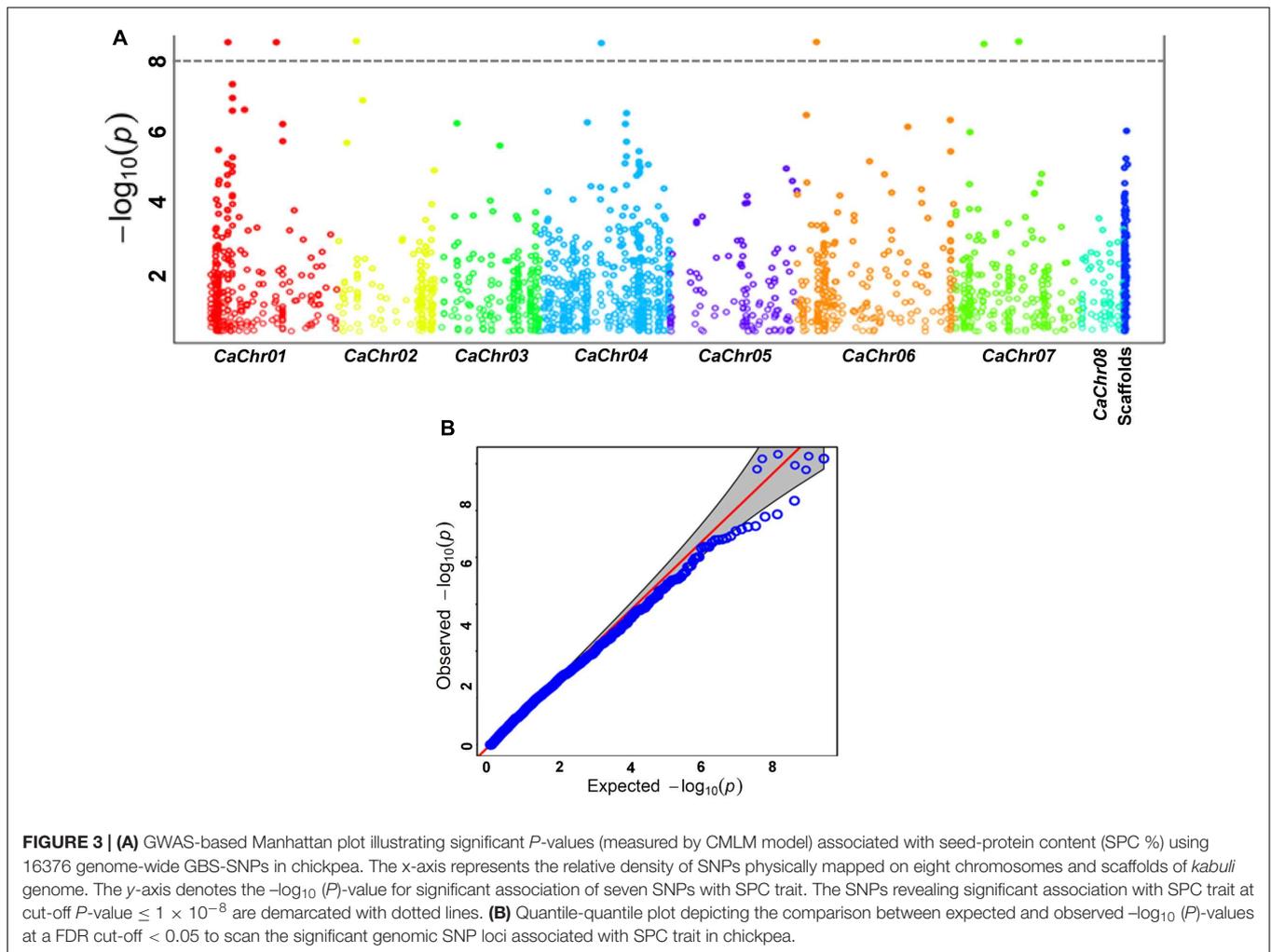
FIGURE 3 | (A) GWAS-based Manhattan plot illustrating significant *P*-values (measured by CMLM model) associated with seed-protein content (SPC %) using 16376 genome-wide GBS-SNPs in chickpea. The x-axis represents the relative density of SNPs physically mapped on eight chromosomes and scaffolds of *kabuli* genome. The *y*-axis denotes the –log$_{10}$ (*P*)-value for significant association of seven SNPs with SPC trait. The SNPs revealing significant association with SPC trait at cut-off *P*-value $\leq 1 \times 10^{-8}$ are demarcated with dotted lines. **(B)** Quantile-quantile plot depicting the comparison between expected and observed –log$_{10}$ (*P*)-values at a FDR cut-off < 0.05 to scan the significant genomic SNP loci associated with SPC trait in chickpea.

TABLE 2 | Seven SNPs-containing genes regulating seed protein content delineated by using an integrated genomic approach in chickpea.

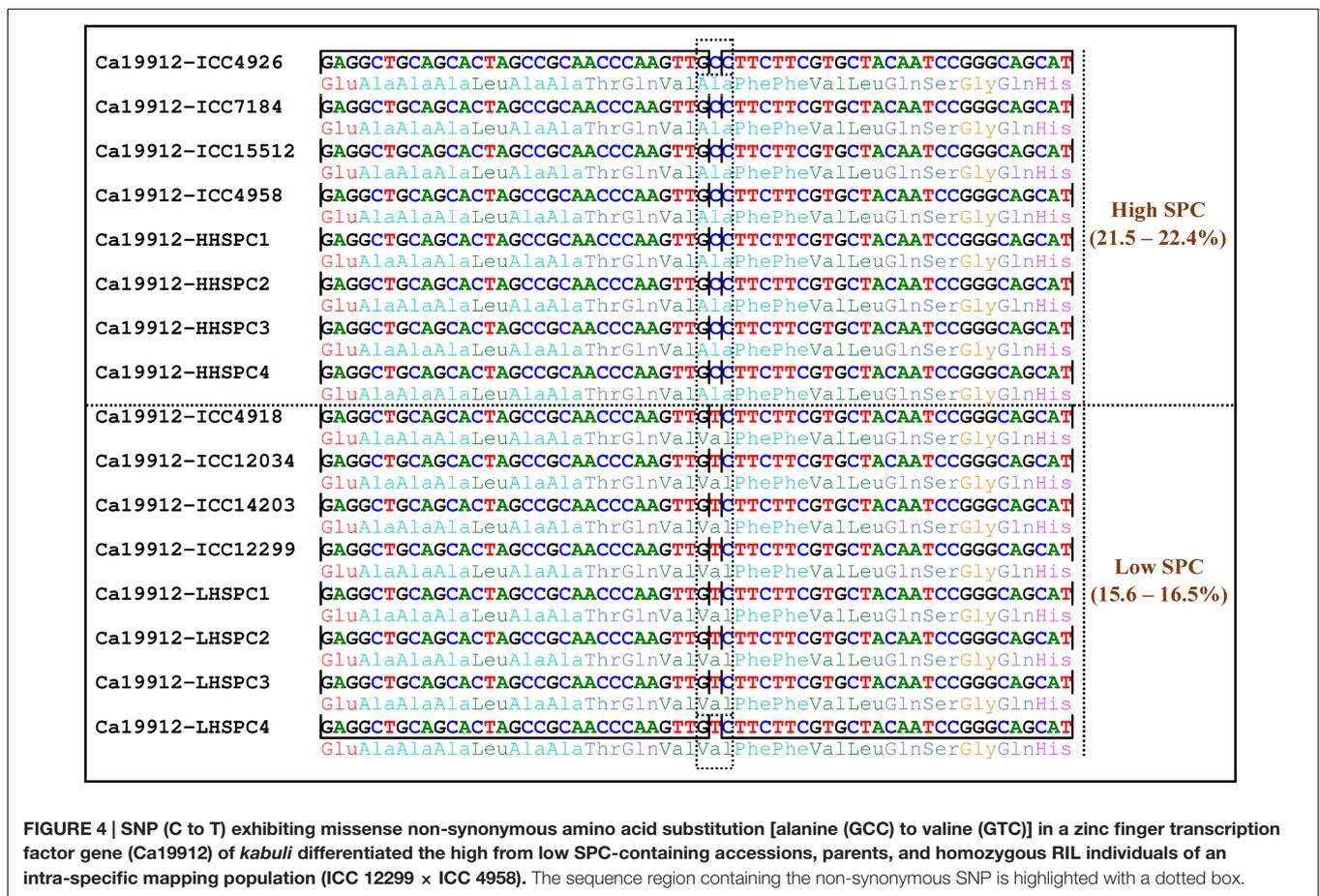| SNP IDs	*Kabuli* chromosomes	SNP physical positions (bp)	SNPs	Gene accession IDs	Structural annotation	Putative functions	Association analysis	
							P	R^2 (%)
*#CakSNP881	1	8112343	C/T	Ca08057	CDS (Syn)	ATP-dependent RNA helicase DEAD-box	3.7×10^{-9}	13
*#CakSNP1726	1	27017765	A/T	Ca20299	INTRON	Cystathionine-beta synthase	1.7×10^{-8}	10
CakSNP2438	2	9756955	G/A	–	INTERGENIC	–	2.1×10^{-8}	10
#CakSNP6395	4	17920616	C/G	Ca18632	CDS (Syn)	ABC transporter transmembrane domain-encoding protein	3.3×10^{-9}	14
*#CakSNP9644	6	3548121	C/A	Ca05955	CDS (NSyn)	Cytidine (CMP) and deoxycytidylate (dCMP) deaminases	2.5×10^{-9}	17
*#CakSNP12015	7	5854598	C/T	Ca06772	INTRON	G10 protein	1.5×10^{-8}	12
*#CakSNP12753	7	24181321	C/T	Ca19912	CDS (NSyn)	Zinc finger transcription factor	1.2×10^{-9}	20

*Validated in an intra-specific mapping population (ICC 12299 × ICC 4958) contrasting for SPC. #Validated by differential expression profiling. CDS: Coding sequence; Syn: Synonymous; NSyn: Non-synonymous.

these five gene-derived SNP loci in parental accessions (ICC 12299 and ICC 4958) and homozygous mapping individuals with contrasting low (15.6–16.5%) and high (21.5–22.4%) levels of SPC was observed. Like-wise, the five gene-derived SNP alleles were validated three of each low (*desi* cv. ICC 4918, *kabuli* cv. ICC 12034 and *kabuli* cv. ICC 14203) and high (*desi* cv. ICC 4926, *desi* cv. ICC 7184 and *kabuli* cv. ICC 15512) SPC-containing natural germplasm accessions (**Figure 4**, **Table 2**). Therefore, strong marker allele effect of these five genes with high and low SPC differentiation was apparent in chickpea. In contrast, low and high SPC-associated alleles of an intergenic SNP and a candidate gene (ABC transporter transmembrane domain-encoding protein)-derived SNP could not correspond to the phenotypes of low and high SPC-containing accessions/parents and homozygous mapping individuals. Henceforth, five SNPs-carrying genes validated by GWAS and selective genotyping in a mapping population were considered as target candidates for SPC trait regulation in chickpea (**Table 2**).

Validation of SPC-Associated Genes by Expression Profiling

The SPC-associated six SNPs-carrying genes (detected by GWAS), including five genes validated by selective genotyping

in a mapping population were used for differential expression analysis to infer the functional regulatory pattern of these genes controlling SPC in chickpea. The differential expression profiling of SPC-associated genes in the leaves as well as early cell division phase and late seed maturation developmental stages of eight chickpea accessions with contrasting levels of low (ICC 4918, IC 12034, ICC 14203, and ICC 12299 with 15.6–16.5%) and high (ICC 4926, ICC 7184, ICC 15512, and ICC 4958 with 21.5–22.4%) SPC content was performed. This analysis was further extended to low and high SPC-containing parental accessions and four homozygous RIL individuals of a mapping population (ICC 12299 × ICC 4958). The SPC-associated six SNPs-containing genes (including five genes validated by selective genotyping in a mapping population) exhibited seed-specific expression (>three fold, $P \leq 0.0001$) as compared to the leaves of low and high SPC-containing accessions/mapping individuals (Supplementary Figure S1, **Figure 5**). Interestingly, all six seed-specific genes revealed higher differential up-regulation (>four fold, $P \leq 0.0001$) in two seed developmental stages of high (21.5–22.4%) SPC-containing accessions/mapping individuals than that of low (15.6–16.5%) SPC-containing accessions/mapping individuals (Supplementary Figure S1, **Figure 5**). Remarkably, one non-synonymous SNP-containing zinc finger transcription factor gene showing strong association (R^2: 20%) with SPC based

FIGURE 4 | SNP (C to T) exhibiting missense non-synonymous amino acid substitution [alanine (GCC) to valine (GTC)] in a zinc finger transcription factor gene (Ca19912) of *kabuli* differentiated the high from low SPC-containing accessions, parents, and homozygous RIL individuals of an intra-specific mapping population (ICC 12299 × ICC 4958). The sequence region containing the non-synonymous SNP is highlighted with a dotted box.
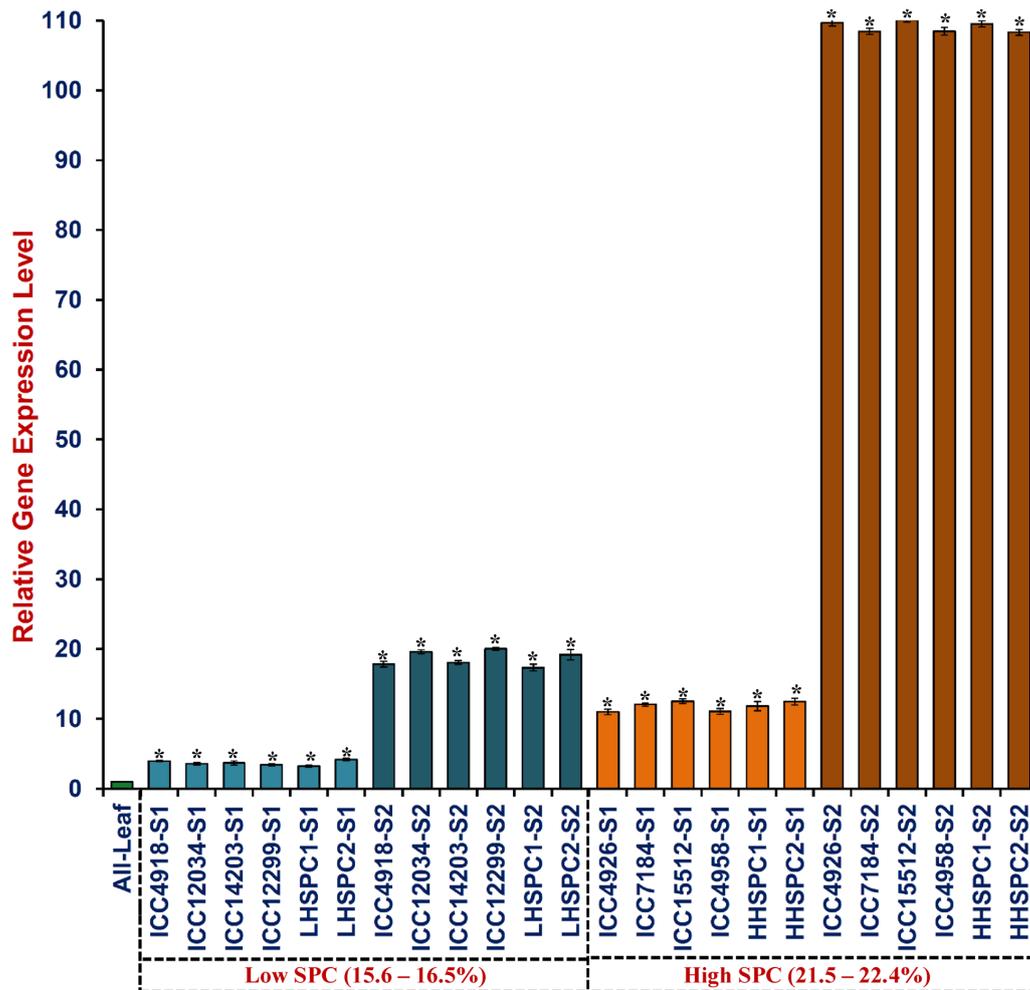
FIGURE 5 | Differential expression profile of a SPC-associated non-synonymous SNP-carrying zinc finger transcription factor gene in the leaves (green colored) as well as early cell division phase and late seed maturation developmental stages of low and high SPC-containing chickpea accessions, parents, and homozygous RIL individuals of an intra-specific mapping population (ICC 12299 × ICC 4958). The endogenous control *elongation factor-1 alpha* was used to normalize the expression value across different tissues/developmental stages of accessions, parents, and mapping individuals. The gene expression in leaf tissues (indicated as All-Leaf) of all the accessions/parents and mapping individuals was considered as reference calibrator (assigned 1). Each bars denote the mean (±SE) of three independent biological replicates with two technical replicates for each sample used in quantitative RT-PCR assay. LHSPC, Low homozygous seed protein containing mapping individual; HHSPC, High homozygous seed protein containing mapping individual; S1, seed development stage 1 (10–20 DAP); and S2, seed development stage 2 (21–30 DAP). The level of gene expression in S1 and S2 developmental stages of low SPC-containing accessions are depicted by light and dark blue-colored bars, respectively. The level of gene expression in S1 and S2 developmental stages of high SPC-containing accessions are illustrated by light and dark brown-colored bars, respectively. *Significant difference in expression of genes at two seed developmental stages of low and high SPC-containing accessions, parents and RIL mapping individuals as compared to leaf at $p < 0.01$ (LSD-ANOVA significance test).

on GWAS exhibited pronounced up-regulation (>five fold, $P \leq 0.00001$) in two seed developmental stages of high SPC-containing accessions/mapping individuals than that of low SPC-containing accessions/mapping individuals of chickpea (Supplementary Figure S1, **Figure 5**). A direct positive correlation between expression level of six genes with seed protein content both at early cell division and late seed maturation developmental stages of chickpea accessions, mapping parents and individuals was evident. Altogether, the GWAS, selective genotyping in bi-parental RIL mapping population and differential expression profiling in the present study delineated natural functional

SNP allelic variants in the six potential novel candidate genes regulating SPC in chickpea (**Table 2**).

DISCUSSION

The integrated genomics-assisted breeding strategy (combining association and QTL mapping with differential gene expression profiling) is found to be efficient for rapid delineation of potential candidate genes and alleles regulating complex yield component quantitative traits in chickpea (Kujur et al., 2013, 2014, 2015a,b,c;

Saxena et al., 2014a; Bajaj et al., 2015b,c). The clues obtained from these studies have essentially encouraged us to utilize this integrated genomic approach for quantitative dissection of complex seed protein content trait in chickpea. In present investigation, GWAS was integrated with selective genotyping in RIL mapping population and differential gene expression profiling to scale-down the natural allelic variants of candidate genes regulating SPC in chickpea. Primarily, we utilized a well-established GBS assay for simultaneous large-scale discovery and genotyping of genome-wide SNPs in chickpea (Deokar et al., 2014; Jaganathan et al., 2015; Kujur et al., 2015a,b,c). The high-throughput sequencing and genotyping of 92 *desi* and *kabuli* accessions (association panel) by GBS approach and subsequent genotyping of these mined genic SNPs in 244 *desi* and *kabuli* accessions by Illumina GoldenGate assay altogether discovered 16376 reference *kabuli* genome-based SNPs for performing GWAS of SPC trait in chickpea. These genome and gene-derived SNPs thus have potential to be utilized for various genomics-assisted breeding applications, including association and genetic mapping to identify potential genes/QTLs governing traits of agronomic importance in chickpea.

The observed wider phenotypic variation and normal frequency distribution of SPC among 336, including 92 and 244 *desi* and *kabuli* accessions reflect the complex quantitative genetic inheritance pattern of target trait in chickpea. Considering its efficacy for quantitative dissection of complex traits in diverse crop plants (Huang et al., 2010, 2012; Chen et al., 2014; Peng et al., 2014; Thudi et al., 2014; Yang et al., 2014; Kujur et al., 2015a; Kumar et al., 2015), the GWAS strategy can be deployed for understanding the complex genetic architecture of SPC trait in chickpea. The use of 16376 genome-wide GBS-SNP genotyping information scanned from a structured population of 336 accessions (with 150–200 kb chromosomal LD decay) in CMLM-based GWAS, identified seven genomic loci significantly associated with SPC trait in chickpea. The LD decay estimated for 336 accessions using 14115 chromosome-wise physically mapped SNPs in the present investigation is comparable with that documented by previous association mapping studies of complex yield component traits in chickpea (Bajaj et al., 2015a; Kujur et al., 2015a,b). This indicates that the SNP density required for effective GWAS is adequate to identify the valid potential genomic loci governing SPC trait in a larger chickpea genome with narrow genetic base. The added-advantage of computationally efficient CMLM model-based strategy adopted in our study for GWAS over other association model-based approaches as documented hitherto for rapid genome-wide scanning of non-spurious SNP marker-trait association (maximal statistical power and high prediction) is well-demonstrated in crop plants (Lipka et al., 2012; Stanton-Geddes et al., 2013; Saxena et al., 2014a; Thudi et al., 2014; Kumar et al., 2015). The SPC-associated seven SNP loci being derived from diverse non-synonymous coding sequence components of six candidate genes are assumed to be functionally relevant. The non-synonymous substitutions of SNP loci in the CDS (coding sequence) of genes encoding variable amino acid residues might create altered secondary structure of proteins that possibly affects the DNA binding and transcriptional activity of target genes associated

with multiple agronomic traits in crop plants. Such possible transcriptional mechanism of trait regulation due to non-synonymous SNP substitutions has already been demonstrated in one of the high seed weight-associated transcription factor gene in chickpea (Kujur et al., 2013) as well as grain size (Mao et al., 2010) and stress-responsive (Parida et al., 2012) genes in rice. Irrespective of any genetic backgrounds (*desi/kabuli*) in chickpea, we observed a similar level of association potential for seven genomic loci with SPC trait. This could be due to a comparable level of SPC phenotypic trait variation observed in *desi* and *kabuli* cultivar groups. Regardless of the genetic backgrounds, robust genomic loci (genes) especially the non-synonymous genic SNP allelic variants identified through GWAS in the present investigation can be effectively utilized for rapidly establishing marker-trait linkages and identification of genes/QTLs controlling SPC traits in chickpea. This in turn will expedite marker-assisted genetic improvement of chickpea for high protein content.

To ascertain the validity and robustness of identified seven SPC-associated candidate genes, the outcome of GWAS was further correlated with that of selective genotyping in mapping population and differential gene expression profiling. The seed-specific up-regulation and transcript accumulation of six SPC-associated genes specifically in early cell division and late maturation seed developmental stages of high SPC-containing accessions and mapping individuals was apparent. This included five genes that were validated by selective genotyping in a RIL mapping population. The seed protein content is a quantitative trait and being regulated by a complex genetic network involving a diverse array of genes in crop plants (Frimpong et al., 2009; Jadhav et al., 2015). In the present study, a combinatorial strategy involving GWAS, selective genotyping in mapping population and differential gene expression profiling helped us to narrow-down SNP allelic variants in six potential genes (encode ATP-dependent RNA helicase DEAD-box, cystathionine-beta synthase, ABC transporter transmembrane domain, CMP, and dCMP deaminases, and G10 and zinc finger protein) regulating SPC trait in chickpea.

Notably, one strong SPC-associated non-synonymous SNP-containing zinc finger transcription factor gene identified (integrating GWAS, selective genotyping in mapping population and differential expression profiling) in our study is known to be predominantly expressed in seed embryos and also involved in regulating seed storage protein content by accumulation of glutelins during seed development of rice (Chen et al., 2014). In maize, this zinc finger protein is known to bind prolamin box that are conserved across cereals and play a vital role in cereal storage protein gene expression and transcript accumulation (Vicente-Carbajosa et al., 1997). Such non-synonymous SNPs is known to be involved in amino acid substitutions and alteration of secondary structure of proteins and thus possibly affect the DNA binding and transcriptional activity of target genes regulating diverse complex yield component quantitative traits in chickpea (Kujur et al., 2013, 2014, 2015b). In this context, non-synonymous SNP identified in a zinc finger protein-coding gene regulating seed protein content by employing an integrated genomic approach is functionally relevant and

thus could be utilized for genetic dissection of this complex trait in chickpea. A SPC-associated ATP-dependent RNA helicase DEAD-box protein-coding gene identified in chickpea regulates ribosomal biogenesis as well as protein synthesis in multiple crop plants (Xu et al., 2015). Another SPC-associated ABC transporter gene detected in chickpea has a potential involvement in transport and storage of proteins inside the cell and thus plays a crucial role in plant nutrition (Kang et al., 2011). The genes code CMP/dCMP deaminases and G10 proteins associated with SPC trait in chickpea regulate nucleotide metabolism and translational recruitment, respectively, in crop plants (McGrew et al., 1989; Armenta-Medina et al., 2014). The gene code cystathionine-beta synthase is known to be involved in cysteine biosynthesis and methionine metabolism in plant species (Amir, 2008). The elevation of two sulfur-containing proteinogen amino acids (precursors of proteins), methionine and cysteine by altering the cystathionine-beta synthase is essential for improving the protein content, including nutritive value of plant seeds (Hesse et al., 2004; Nguyen et al., 2012). Collectively, the SPC-associated novel molecular tags (markers and alleles) of genes narrowed-down by integrated genomics-assisted breeding approach, essentially function by regulating overall signaling, transport, metabolism, and accumulation of proteins in crop plants. Therefore, these functionally relevant molecular tags can be utilized for marker-assisted genetic enhancement to develop superior cultivars with enriched seed protein content in chickpea.

AUTHOR CONTRIBUTIONS

HU and DB conducted experiments and drafted the manuscript. LN, SD, and VK involved in data analysis. HU, CG, and SS helped in constitution of association panel and performed phenotyping. SP and AT conceived and designed the study, guided data analysis and interpretation, participated in drafting and correcting the manuscript critically, and gave the final approval of the version to be published. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2016.00302

REFERENCES

Agarwal, G., Jhanwar, S., Priya, P., Singh, V. K., Saxena, M. S., Parida, S. K., et al. (2012). Comparative analysis of kabuli chickpea transcriptome with desi and wild chickpea provides a rich resource for development of functional markers. *PLoS ONE* 7:e52443. doi: 10.1371/journal.pone.0052443

Amir, R. (2008). Towards improving methionine content in plants for enhanced nutritional quality. *Funct. Plant Sci. Biotechnol.* 2, 36–46.

Armenta-Medina, D., Segovia, L., and Perez-Rueda, E. (2014). Comparative genomics of nucleotide metabolism: a tour to the past of the three cellular domains of life. *BMC Genomics* 15:800. doi: 10.1186/1471-2164-15-800

Bajaj, D., Das, S., Badoni, S., Kumar, V., Singh, M., Bansal, K. C., et al. (2015a). Genome-wide high-throughput SNP discovery and genotyping for understanding natural (functional) allelic diversity and domestication patterns in wild chickpea. *Sci. Rep.* 5:12468. doi: 10.1038/srep12468

Bajaj, D., Saxena, M. S., Kujur, A., Das, S., Badoni, S., Tripathi, S., et al. (2015b). Genome-wide conserved non-coding microsatellite (CNMS) marker-based integrative genetical genomics for quantitative dissection of seed weight in chickpea. *J. Exp. Bot.* 66, 1271–1290. doi: 10.1093/jxb/eru478

Bajaj, D., Upadhyaya, H. D., Khan, Y., Das, S., Badoni, S., Shree, T., et al. (2015c). A combinatorial approach of comprehensive QTL-based comparative genome mapping and transcript profiling identified a seed weight-regulating candidate gene in chickpea. *Sci. Rep.* 5:9264. doi: 10.1038/srep09264

Bourgeois, M., Jacquin, F., Cassecuelle, F., Savois, V., Belghazi, M., Aubert, G., et al. (2011). A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. *Proteomics* 11, 1581–1594. doi: 10.1002/pmic.201000687

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* 46, 714–721. doi: 10.1038/ng.3007

Das, S., Upadhyaya, H. D., Bajaj, D., Kujur, A., Badoni, S., Laxmi, et al. (2015). Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. *DNA Res.* 22, 193–203. doi: 10.1093/dnares/dsv004

Deokar, A. A., Ramsay, L., Sharpe, A. G., Diapari, M., Sindhu, A., Bett, K., et al. (2014). 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* 15:708. doi: 10.1186/1471-2164-15-708

Duranti, M., and Gius, C. (1997). Legume seeds: Protein content and nutritional value. *Field Crops Res.* 53, 31–45. doi: 10.1016/S0378-4290(97)00021-X

FAOSTAT: Production-Crops (2012). *The State of Food and Agriculture.* Rome: Food and Agriculture Organization of the United Nations.

Frimpong, A., Sinha, A., Tar'an, B., Warkentin, T. D., Gossen, B. D., and Chibbar, R. N. (2009). Genotype and growing environment influence chickpea (*Cicer arietinum* L.) seed composition. *J. Sci. Food Agric.* 89, 2052–2063. doi: 10.1002/jsfa.3690

Garg, R., Bhattacharjee, A., and Jain, M. (2014). Genome-scale transcriptomic insights into molecular aspects of abiotic stress responses in chickpea. *Plant Mol. Biol. Rep.* 33, 388–400. doi: 10.1007/s11105-014-0753-x

Garg, R., Patel, R. K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., et al. (2011). Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiol.* 156, 1661–1678. doi: 10.1104/pp.111.178616

Garg, R., Sahoo, A., Tyagi, A. K., and Jain, M. (2010). Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.). *Biochem. Biophys. Res. Commun.* 396, 283–288. doi: 10.1016/j.bbrc.2010.04.079

Gaur, R., Azam, S., Jeena, G., Khan, A. W., Choudhary, S., Jain, M., et al. (2012). High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). *DNA Res.* 19, 357–373. doi: 10.1093/dnares/dss018

Gujaria, N., Kumar, A., Dauthal, P., Dubey, A., Hiremath, P., Bhanu Prakash, A., et al. (2011). Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 122, 1577–1589. doi: 10.1007/s00122-011-1556-1

Haider, M., and Haider, S. (1984). Assessment of protein-calorie malnutrition. *Clin. Chem.* 30, 1286–1299.

Hesse, H., Kreft, O., Maimann, S., Zeh, M., and Hoefgen, R. (2004). Current understanding of the regulation of methionine biosynthesis in plants. *J. Exp. Bot.* 55, 1799–1808. doi: 10.1093/jxb/erh139

Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., et al. (2012). 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.* 10, 716–732. doi: 10.1111/j.1467-7652.2012.00710.x

Huang, X., Kurata, N., Wei, X., Wang, Z. X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695

Hwang, E. Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., et al. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics* 15:1. doi: 10.1186/1471-2164-15-1

Jadhav, A. A., Rayate, S. J., Mhase, L. B., Thudi, M., Chitikineni, A., Harer, P. N., et al. (2015). Marker-trait association study for protein content in chickpea (*Cicer arietinum* L.). *J. Genet.* 94, 279–286. doi: 10.1007/s12041-015-0529-6

Jaganathan, D., Thudi, M., Kale, S., Azam, S., Roorkiwal, M., Gaur, P. M., et al. (2015). Genotyping-by-sequencing based intra-specific genetic map refines a "QTL-hotspot" region for drought tolerance in chickpea. *Mol. Genet. Genomics* 290, 559–571. doi: 10.1007/s00438-014-0932-3

Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* 74, 715–729. doi: 10.1111/tpj.12173

Jhanwar, S., Priya, P., Garg, R., Parida, S. K., Tyagi, A. K., and Jain, M. (2012). Transcriptome sequencing of wild chickpea as a rich resource for marker development. *Plant Biotechnol. J.* 10, 690–702. doi: 10.1111/j.1467-7652.2012.00712.x

Jukanti, A., Guar, P., Gowda, C. L. L., and Chibbar, R. (2012). Nutritional quality and health benefits of chickpea. *Br. J. Nutr.* 108, S11–S26. doi: 10.1017/S0007114512000797

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., et al. (2011). Plant ABC transporters. *Arabidopsis Book* 9:e0153. doi: 10.1199/tab.0153

Krajewski, P., Bocianowski, J., Gawlowska, M., Kaczmarek, Z., Pniewski, T., Swiecicki, W., et al. (2012). QTL for field components and protein content: a multienvironmental study of two pea (*Pisum sativum* L.) populations. *Euphytica* 183, 323–336. doi: 10.1007/s10681-011-0472-4

Kudapa, H., Azam, S., Sharpe, A. G., Taran, B., Li, R., Deonovic, B., et al. (2014). Comprehensive transcriptome assembly of chickpea (*Cicer arietinum* L.) using Sanger and next generation sequencing platforms: development and applications. *PLoS ONE* 9:e86039. doi: 10.1371/journal.pone.0086039

Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2013). Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* 20, 355–373. doi: 10.1093/dnares/dst015

Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2014). An efficient and cost-effective approach for genic microsatellite marker-based large-scale trait association mapping: identification of candidate genes for seed weight in chickpea. *Mol. Breed.* 34, 241–265. doi: 10.1007/s11032-014-0033-3

Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015a). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Sci. Rep.* 5:11166. doi: 10.1038/srep11166

Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015b). Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. *Front. Plant Sci.* 6:162. doi: 10.3389/fpls.2015.00162

Kujur, A., Upadhyaya, H. D., Shree, T., Bajaj, D., Das, S., Saxena, M. S., et al. (2015c). Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. *Sci. Rep.* 5:9468. doi: 10.1038/srep09468

Kumar, A., Choudhary, A. K., Solanki, R. K., and Pratap, A. (2011). Towards marker-assisted selection in pulses: a review. *Plant Breed.* 130, 297–313. doi: 10.1111/j.1439-0523.2011.01851.x

Kumar, V., Singh, A., Amitha Mithra, S. V., Krishnamurthy, S. L., Parida, S. K., Jain, S., et al. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 22, 133–145. doi: 10.1093/dnares/dsu046

Lestari, P., Van, K., Lee, J., Kang, Y. J., and Lee, S. H. (2013). Gene divergence of homeologous regions associated with a major seed protein content QTL in soybean. *Front. Plant Sci.* 4:176. doi: 10.3389/fpls.2013.00176

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* 2, 2397–2399. doi: 10.1093/bioinformatics/bts444

Lu, W., Wen, Z., Li, H., Yuan, D., Li, J., Zhang, H., et al. (2013). Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. *Theor. Appl. Genet.* 126, 425–433. doi: 10.1007/s00122-012-1990-8

Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., et al. (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19579–19584. doi: 10.1073/pnas.1014419107

McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B., and Richter, J. D. (1989). Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.* 3, 803–815. doi: 10.1101/gad.3.6.803

Millerd, A. (1975). Biochemistry of legume seed proteins. *Ann. Rev. Plant Physiol.* 26, 53–72. doi: 10.1146/annurev.pp.26.060175.000413

Monti, L. M., and Grillo, S. (1983). Legume seed improvement for protein content and quality. *Plant Foods Hum. Nutr.* 32, 253–266. doi: 10.1007/BF01091190

Nayak, S. N., Zhu, H., Varghese, N., Datta, S., Choi, H. K., Horres, R., et al. (2010). Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. *Theor. Appl. Genet.* 120, 1415–1441. doi: 10.1007/s00122-010-1265-1

Nguyen, H. C., Hoefgen, R., and Hesse, H. (2012). Improving the nutritive value of rice seeds: elevation of cysteine and methionine contents in rice plants by ectopic expression of a bacterial serine acetyltransferase. *J. Exp. Bot.* 63, 5991–6001. doi: 10.1093/jxb/ers253

Parida, S. K., Mukerji, M., Singh, A. K., Singh, N. K., and Mohapatra, T. (2012). SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure. *BMC Genomics* 13:426. doi: 10.1186/1471-2164-13-426

Parida, S. K., Verma, M., Yadav, S. K., Ambawat, S., Das, S., Garg, R., et al. (2015). Development of genome-wide informative simple sequence repeat markers for large-scale genotyping applications in chickpea and development of web resource. *Front. Plant Sci.* 6:645. doi: 10.3389/fpls.2015.00645

Parween, S., Nawaz, K., Roy, R., Pole, A. K., Venkata Suresh, B., Misra, G., et al. (2015). An advanced draft genome assembly of a desi type chickpea (*Cicer arietinum* L.). *Sci. Rep.* 5:12806. doi: 10.1038/srep12806

Peng, B., Kong, H., Li, Y., Wang, L., Zhong, M., Sun, L., et al. (2014). OsAAP6 functions as an important regulator of grain protein content and nutritional quality in rice. *Nat. Commun.* 5:4847. doi: 10.1038/ncomms5847

Pradhan, S., Bandhiwal, N., Shah, N., Kant, C., Gaur, R., and Bhatia, S. (2014). Global transcriptome analysis of developing chickpea (*Cicer arietinum* L.) seeds. *Front. Plant Sci.* 5:698. doi: 10.3389/fpls.2014.00698

Sabbavarapu, M. M., Sharma, M., Chamarthi, S. K., Swapna, N., Rathore, A., Thudi, M., et al. (2013). Molecular mapping of QTLs for resistance to *Fusarium* wilt (race 1) and *Ascochyta* blight in chickpea (*Cicer arietinum* L.). *Euphytica* 193, 121–133. doi: 10.1007/s10681-013-0959-2

Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014a). An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* 21, 695–710. doi: 10.1093/dnares/dsu031

Saxena, M. S., Bajaj, D., Kujur, A., Das, S., Badoni, S., Kumar, V., et al. (2014b). Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea. *PLoS ONE* 9:e107484. doi: 10.1371/journal.pone.0107484

Singh, M., Bisht, I. S., Dutta, M., Kumar, K., Basandrai, A. K., Kaur, L., et al. (2014). Characterization and evaluation of wild annual *Cicer* species for agro-morphological traits and major biotic stresses under northwestern Indian conditions. *Crop Sci.* 54, 229–239. doi: 10.2135/cropsci2013.04.0225

Singh, U., and Jambunathan, R. (1980). Evaluation of rapid methods for the estimation of protein in chickpea (*Cicer arietinum* L.). *J. Sci. Food Agric.* 31, 247–254. doi: 10.1002/jsfa.2740310306

Singh, V. K., Garg, R., and Jain, M. (2013). A global view of transcriptome dynamics during flower development in chickpea by deep sequencing. *Plant Biotechnol. J.* 11, 691–701. doi: 10.1111/pbi.12059

Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., Yoder, J., Mudge, J., et al. (2013). Candidate genes and genetic architecture of symbiotic and agronomic

traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS ONE* 8:e65688. doi: 10.1371/journal.pone.0065688

Thudi, M., Bohra, A., Nayak, S. N., Varghese, N., Shah, T. M., Penmetsa, R. V., et al. (2011). Novel SSR markers from BAC-end sequences, DArT arrays and a comprehensive genetic map with 1,291 marker loci for chickpea (*Cicer arietinum* L.). *PLoS ONE* 6:e27275. doi: 10.1371/journal.pone.0027275

Thudi, M., Upadhyaya, H. D., Rathore, A., Gaur, P. M., Krishnamurthy, L., Roorkiwal, M., et al. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. *PLoS ONE* 9:e96758. doi: 10.1371/journal.pone.0096758

Upadhyaya, H. D., Furman, B. J., Dwivedi, S. L., Udupa, S. M., Gowda, C. L. L., Baum, M., et al. (2006). Development of a composite collection for mining germplasm possessing allelic variation for beneficial traits in chickpea. *Plant Genet. Resour.* 4, 13–19. doi: 10.1079/PGR2005101

Upadhyaya, H. D., and Ortiz, R. (2001). A mini-core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* 102, 1292–1298. doi: 10.1007/s00122-001-0556-y

Upadhyaya, H. D., Ortiz, R., Bramel, P. J., and Singh, S. (2002). Phenotypic diversity for morphological and agronomic characteristics in chickpea core collection. *Euphytica* 123, 333–342. doi: 10.1023/A:1015088417487

Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., et al. (2013a). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30, 83–89. doi: 10.1038/nbt.2022

Varshney, R. K., Gaur, P. M., Chamarthi, S. K., Krishnamurthy, L., Tripathi, S., Kashiwagi, J., et al. (2014a). Fast-track introgression of "QTL-hotspot" for root traits and other drought tolerance trait in JG 11, an elite and leading variety of chickpea (*Cicer arietinum* L.). *Plant Genome* 6, 1–26.

Varshney, R. K., Mohan, S. M., Gaur, P. M., Chamarthi, S. K., Singh, V. K., Srinivasan, S., et al. (2014b). Marker-assisted backcrossing to introgress resistance to *Fusarium* wilt race 1 and *Ascochyta* blight in C214, an elite cultivar of chickpea. *Plant Genome* 7, 1–11. doi: 10.3835/plantgenome2013.10.0035

Varshney, R. K., Mohan, S. M., Gaur, P. M., Gangarao, N. V., Pandey, M. K., Bohra, A., et al. (2013b). Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. *Biotechnol. Adv.* 31, 1120–1134. doi: 10.1016/j.biotechadv.2013.01.001

Varshney, R. K., Thudi, M., Nayak, S., Gaur, P., Kashiwagi, J., Krishnamurthy, L., et al. (2014c). Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 127, 445–462. doi: 10.1007/s00122-013-2230-6

Vicente-Carbajosa, J., Moose, S. P., Parsons, R. L., and Schmidt, R. J. (1997). A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7685–7690. doi: 10.1073/pnas.94.14.7685

Wang, J., Chen, P., Wang, D., Shannon, G., Zeng, A., Orazaly, M., et al. (2015). Identification and mapping of stable QTL for protein content in soybean seeds. *Mol. Breed.* 35:92. doi: 10.1007/s11032-015-0285-6

Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., et al. (2015). QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. *Theor. Appl. Genet.* 128, 839–850. doi: 10.1007/s00122-015-2474-4

World Health Organization [WHO] (2013). *Updates on the Management of Severe Acute Malnutrition in Infants, and Children.* Geneva: World Health Organization

Xu, J., Liu, C., Li, M., Hu, J., Zhu, L., Zeng, D., et al. (2015). A rice DEAD-box RNA helicase protein, *OsRH17*, suppresses 16S ribosomal RNA maturation in *Escherichia coli. Gene* 555, 318–328. doi: 10.1016/j.gene.2014.11.025

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi: 10.1038/ncomms6087