SOLANACEAE VII: BIOLOGY, GENETICS, AND EVOLUTION

EDITED BY: Peter Poczai, Nunzio D'Agostino, Rocio Deanna and Ezio Portis

PUBLISHED IN: Frontiers in Genetics and Frontiers in Plant Science

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-597-3
DOI 10.3389/978-2-88976-597-3

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

SOLANACEAE VII: BIOLOGY, GENETICS, AND EVOLUTION

Topic Editors:
Peter Poczai, University of Helsinki, Finland
Nunzio D'Agostino, University of Naples Federico II, Italy
Rocio Deanna, University of Colorado Boulder, United States
Ezio Portis, University of Turin, Italy

Citation: Poczai, P., D'Agostino, N., Deanna, R., Portis, E., eds. (2022). Solanaceae VII: Biology, Genetics, and Evolution. Lausanne: Frontiers Media. doi: 10.3389/978-2-88976-597-3

Table of Contents

frontiers | Frontiers in Genetics

EDITORIAL
published: 23 June 2022
doi: 10.3389/fgene.2022.932421

Editorial: Solanaceae VII: Biology, Genetics, and Evolution

Péter Poczai[1,2,3]*, *Nunzio D'Agostino*[4]*, *Rocio Deanna*[5,6]* and *Ezio Portis*[7]*

[1]Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland, [2]Museomics Research Group, Department of Biosciences, Viikki Plant Science Centre (ViPS), University of Helsinki, Helsinki, Finland, [3]Institute of Advanced Studies Kőszeg (iASK), Kőszeg, Hungary, [4]Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy, [5]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, United States, [6]Instituto Multidisciplinario de Biología Vegetal (IMBIV, CONICET-UNC), Córdoba, Argentina, [7]Department of Agricultural, Forest and Food Sciences (DISAFA), Plant Genetics, University of Turin, Grugliasco, Italy

Keywords: anthocyanin biosynthesis, biosphere reserves, biotic/abiotic stresses, conservation biogeography, endogenous hormones, fruit set, genome-wide association studies, karyotype

Editorial on the Research Topic

Solanaceae VII -biology, genetics and evolution

The Solanaceae family, also referred to as nightshades or the tomato family, includes a wide variety of morphologies and habits. It has a distinct evolutionary history driven by natural selection, among other processes, which has resulted in unique adaptations and rewarding biochemical pathways for pollinators. The Solanaceae family contains 99 genera and approximately 2,700 species, including economically important crops such as potatoes, tomatoes, peppers, and eggplants (**Figure 1**). It also includes several other underutilized nutrient-rich minor crops, such as turkey berry, African nightshades, or scarlet eggplant, which play a key role in the sustainability of local communities both in the Western and Eastern Hemisphere. Many other solanaceous plants are vital sources of medicinal drugs, folk remedies, and hallucinogens, while wild and cultivated species, such as tobacco, are commonly used as stimulants. Several groups, such as petunia seasides, angel's trumpets, butterfly flowers, and Jerusalem cherry have been grown as ornamentals for centuries due to the aesthetic appeal of their flowers and berries.

The declining resilience of ecosystems experiencing biodiversity loss has increased the urgency of promoting the sustainable conservation and use of globally important solanaceous plants. The toolbox available to scientists and breeders is increasingly diversifying to mitigate the emerging challenges imposed by climate change while honoring biodiversity. This diversification could lead to major transformations towards sustainability and how our global society functions and interacts with natural ecosystems. Recent developments in high-throughput sequencing technologies, data science, and internationally open biodiversity data infrastructures are effectively "linking biodiversity with omics" for a better understanding of this important family, as envisioned nearly two decades ago.

The list of Solanaceae with full or publicly available high quality and draft genomes is becoming richer, which allows the "sequence space" of most Solanaceae to be explored to: improve access to gene inventory catalogs; broaden our understanding of phylogeny, biogeography, and taxonomy; identify beneficial alleles for breeding; reveal the breeding potential of wild germplasm; study the mechanisms of gene expression; explore the architecture and organization of chromosomes, capturing the extent of genomic variation; use sequence-based information to perform large-scale genetic (bio)diversity studies; identify the genomic loci underlying key genes and selection sweeps; disentangle metabolic pathways; learn about plant-microbe interaction, (a)biotic stress responses, and defense mechanisms against pathogens and pests.

Edited and reviewed by:
Andrew H. Paterson, University of Georgia, United States

***Correspondence:**
Péter Poczai
peter.poczai@helsinki.fi
Nunzio D'Agostino
nunzio.dagostino@unina.it
Rocio Deanna
rocio.deanna@colorado.edu
Ezio Portis
ezio.portis@unito.it

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

Received: *29 April 2022*
Accepted: *12 May 2022*
Published: *23 June 2022*

Citation:
Poczai P, D'Agostino N, Deanna R and Portis E (2022) Editorial: Solanaceae VII: Biology, Genetics, and Evolution. Front. Genet. 13:932421. doi: 10.3389/fgene.2022.932421

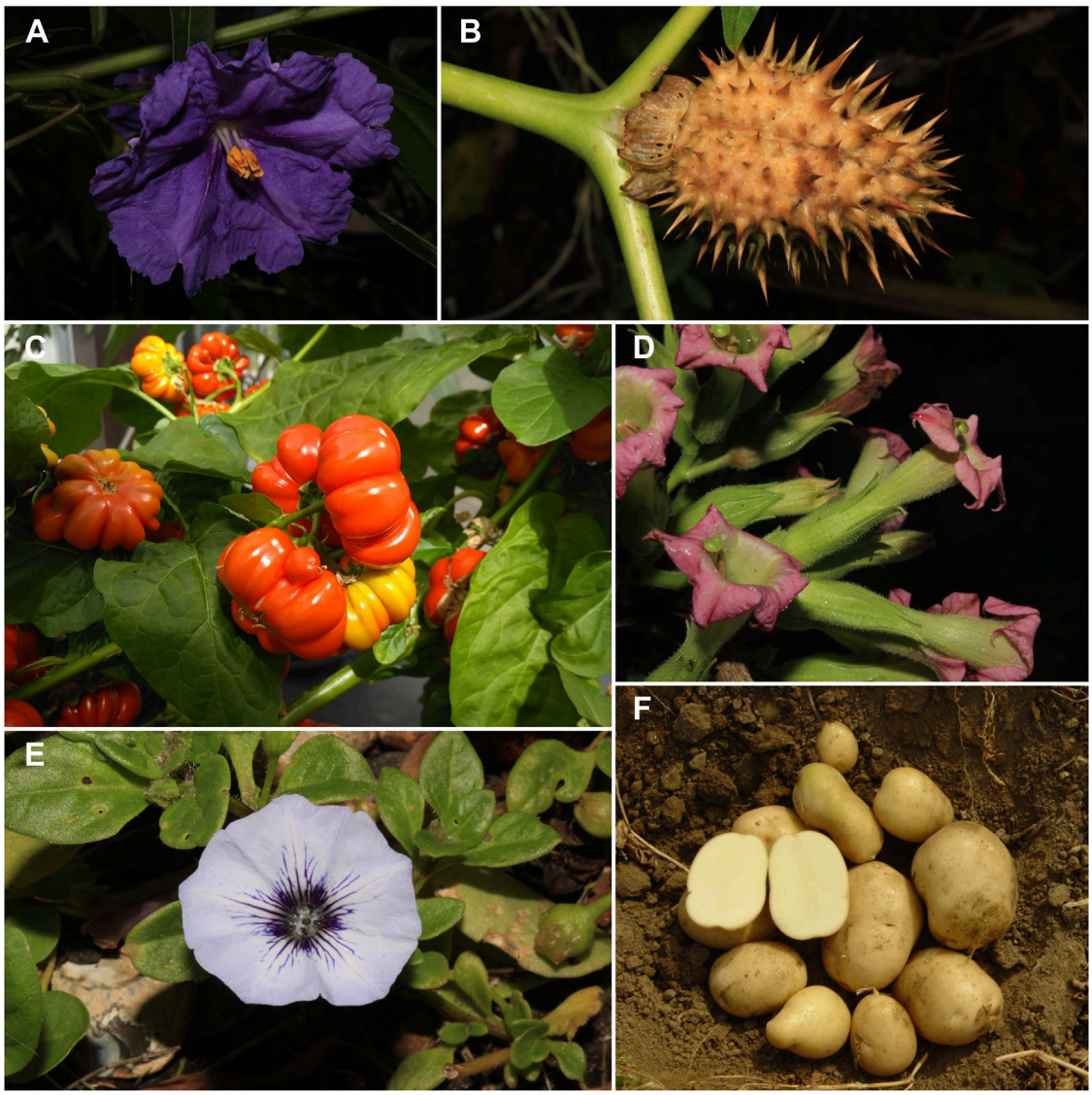

FIGURE 1 | Diversity of Solanaceae crops, poisonous weeds, and relatives. **(A)** *Solanum laciniatum* Aiton, **(B)** *Datura stramonium* L., **(C)** *Solanum lycopersicum* L., **(D)** *Nicotiana tabacum* L., **(E)** *Nolana humifusa* (Gouan) I.M.Johnst., **(F)** *Solanum tuberosum* L. A-B, D-E photographs by R. Deanna at Chelsea Physics Garden; C by P. Poczai; F by D. Carputo.

ADVANCES IN BIOLOGY AND EVOLUTION

For a solanaceous plant to survive, its growth, pollination, and development must be carefully coordinated by a variety of internal and external signals. Controlling gene expression is crucial for the regulation of biological processes including body planning, development, differentiation, and response to numerous environmental changes. Genome expression and modulation begin with transcription, which is the first stage in this process. To better understand how Solanaceae regulate gene expression, transcriptional regulators of the myeloblastosis (*MYB*) transcription factor family were studied in *Capsicum chinense* Jacq. by Islam et al. Their study revealed that several *MYB* genes are involved in determining fruit shape and size, creating opportunities for further trait improvement. Borràs et al. highlighted the different roles of two NAC homologs in *Capsicum* and discussed the complex role of NACs as transcriptional switches in the response to drought stress. The research reported by Wang et al. provided a genome-wide characterization and anthocyanin-related expression analysis of the B-BOX gene family in *C. annuum* L., providing a foundation for further investigation of the CaBBX genes involved in the anthocyanin synthesis mechanisms and development in pepper. The ubiquitous transcription factor nuclear factor Y (NF-Y) has been identified by Liu et al. as a key regulator in fruit development, abiotic stress tolerance, and anthocyanin biosynthesis in potato. In addition, Li et al. provided a characterization of members of the SRO gene family in

tomato, how widely expressed they are in different tissues, and their response to high temperature and salt stress, while mediating the hormone regulatory network.

Pollination is typically an uncertain process that depends on many factors that determine the reproductive success of Solanaceae. Self-incompatibility (SI), described by Charles Darwin (1809–1882) in "*The effects of cross and self-fertilization in the vegetable kingdom (1876),*" is one of these interesting factors often observed in solanaceous plants. Although it was demonstrated early on that self-fertility follows Mendelian inheritance, the underlying molecular basis of SI remained a black box until the mid-1980s. This mechanism was explored in wild tomatoes by Broz et al. who have highlighted that S-RNase expression is more dynamic than previously thought and that changes in expression can impact various reproductive barriers within or between natural populations.

When a solanaceous plant senses an external stimulus, hormones play a key role in its reaction. From blooming through fruit setting and ripening to phototropism and leaf fall, plant hormones affect all aspects of Solanaceae life. Every cell of a tomato plant, for example, may be able to produce these hormones, but this is not a certainty. Therefore, information about the molecular mechanisms underlying hormone synthesis, transport, and signaling is critical to understanding and later accurately engineering these pathways for crop improvement. Hormones have a critical effect on pollination and fruit development, which is pivotal for horticultural production. Endogenous hormones in the ovaries of parthenocarpic and non-parthenocarpic tomato lines were examined to investigate fruit development without pollination. The investigation by Zhang et al. highlighted that auxin plays a prominent role in controlling fruit set in a self-pollinating tomato, while other hormones are integrated in a synergistic or antagonistic way within this process.

Moreira-Munoz et al. described the diversity of Solanaceae in southern South America, assessing conservation gaps in relation to protected areas. They identified micro-hotspots of species richness and reiterated the need for conservation plans in Argentina and Chile. Ramírez-Ojeda et al. determined the most important edaphoclimatic descriptors of wild tomato species and their closely related species along with their natural geographic range in South America. Their work can be used to study the impact of climate change and anthropic activities across the range of these species. Deanna et al. summarized the current understanding of cytogenetics in Solanaceae, its applications, and prospects for making progress in fundamental systematic botany and plant evolution. The information compiled highlights the importance of basic chromosome features in understanding the evolution of the family, especially in delimiting clades.

GENETICS OF DISEASE RESISTANCE AND RESOURCES TO ACCELERATE BREEDING

Solanaceous plants have evolved sophisticated defense mechanisms to combat a variety of pests and diseases. Plant immune systems rely on their capacity to identify and transduce signaling molecules, as well as to respond defensively via pathways involving several genes and their products. Lai et al. performed GWAS on a collection of nearly 100 tobacco accessions to identify QTNs (quantitative trait nucleotides) associated with tobacco bacterial wilt resistance. Association tests returned 38 stable novel QTNs, which allowed the identification of superior alleles for the development of tobacco varieties highly resistant to *Ralstonia solanacearum*. Dubey et al. proceeded with the *in silico* identification and characterization of all members of the gene subfamily TNL (TIR-NBS-LRR) in the potato genome. Following quantitative RT-PCR, one of the genes, namely StTNLC7G2, showed induced expression after *Alternaria solani* attack and was thus functionally characterized. The results showed that StTNLC7G2 is responsible for the induction of reactive oxygen species and plays a key role in the hypersensitive response during plant defense.

Wild relatives of cultivated solanaceous crops are often sources of natural plant disease resistance. These plant genetic resources offer a wide variety of untapped traits that are beneficial for breeding programs. Mangino et al. described an 8-way MAGIC population developed using seven cultivated *Solanum melongena* L. accessions and a wild relative (*S. incanum* L.) as founders. The population, which includes 420 S3 individuals, has been genotyped and phenotyped for anthocyanin-related traits. Association tests returned strong signals for two MYB genes and a COP1 gene, all directly involved in the anthocyanin biosynthesis pathway. Gong et al. reviewed the germplasm resources, their collection strategies for breeding, and molecular-assisted breeding progress in *Lycium* (goji), a widely spread species of the arid to semiarid environments of Eurasia, Africa, North and South America, among which most species have high potential as healthy food. They also discussed recent strategies for QTL mapping and protocols for trait detection, by providing a number of traits (275 agronomic and 870 metabolic) in various organs based on reference studies on *Lycium*, tomato, and other Solanaceae species.

In summary, the current Research Topic (RT) provides further insight into understanding the important mechanisms of the biology of solanaceous plants in terms of biodiversity, plant growth, pollination, cytogenetics, fruit development, and resistance to pathogens.

AUTHOR CONTRIBUTIONS

All authors managed the peer review of manuscripts submitted to the Research Topic and contributed to manuscript writing and editing. All authors approved the final version of the editorial.

ACKNOWLEDGMENTS

The editors would like to thank all authors for their outstanding contributions and all reviewers for their valuable work, helpful

comments, and suggestions. We hope this collection of articles will be of interest to the whole Solanaceae community.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 02 August 2021
doi: 10.3389/fpls.2021.671713

Auxin Metabolism Is Involved in Fruit Set and Early Fruit Development in the Parthenocarpic Tomato "R35-P"

Shaoli Zhang[1,2], Xin Gu[3], Jingcheng Shao[2], Zhifeng Hu[2], Wencai Yang[3], Liping Wang[4], Hongyan Su[1] and Luying Zhu[1]*

[1] Key Laboratory of Molecular Module-Based Breeding of High Yield and Abiotic Resistant Plants in Universities of Shandong (Ludong University), College of Agriculture, Ludong University, Yantai, China, [2] Institute of Vegetable, Gansu Academy of Agricultural Science, Lanzhou, China, [3] College of Horticulture, China Agricultural University, Beijing, China, [4] Agricultural and Rural Bureau of Shouguang, Shouguang, China

OPEN ACCESS

Edited by:
Ezio Portis,
University of Turin, Italy

Reviewed by:
Andrea Mazzucato,
University of Tuscia, Italy
Nikolaos Nikoloudakis,
Cyprus University of Technology, Cyprus
Georgios Tsaniklidis,
Hellenic Agricultural Organization—ELGO, Greece

***Correspondence:**
Shaoli Zhang
shaolizh@163.com

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: 24 February 2021
Accepted: 09 July 2021
Published: 02 August 2021

Citation:
Zhang S, Gu X, Shao J, Hu Z, Yang W, Wang L, Su H and Zhu L (2021) Auxin Metabolism Is Involved in Fruit Set and Early Fruit Development in the Parthenocarpic Tomato "R35-P". Front. Plant Sci. 12:671713. doi: 10.3389/fpls.2021.671713

Parthenocarpic tomato can set fruit and develop without pollination and exogenous hormone treatments under unfavorable environmental conditions, which is beneficial to tomato production from late fall to early spring in greenhouses. In this study, the endogenous hormones in the ovaries of the parthenocarpic tomato line "R35-P" (stigma removed or self-pollination) and the non-parthenocarpic tomato line "R35-N" (self-pollination) at four stages between preanthesis and postanthesis investigated, using high-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS). A nearly twofold IAA (indoleacetic acid) content was found in "R35-P" rather than in "R35-N" at −2 and 0 days after anthesis (DAA). Except at −2 DAA, a lower ABA (abscisic acid) content was observed in Pe (stigma removed in "R35-P") compared to that in Ps (self-pollination in "R35-P") or CK (self-pollination in "R35-N"). After pollination, although the content of GA_1 (gibberellins acid 1) in CK increased, the levels of GAs (gibberellins acids) were notably low. At all four stages, a lower SA (salicylic acid) content was found in Ps and CK than in Pe, while the content and the change trend were similar in Ps and CK. The variation tendencies of JA (jasmonic acid) varied among Pe, Ps, and CK at the studied periods. Furthermore, KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analyses of transcriptomic data identified 175 differentially expressed genes (DEGs) related to plant hormone signal transduction, including 63 auxin-related genes, 27 abscisic acid-related genes, 22 ethylene-related genes, 16 cytokinin-related genes, 16 salicylic acid-related genes, 14 brassinosteroid-related genes, 13 jasmonic acid-related genes, and 4 gibberellin-related genes at −2 DAA and 0 DAA. Our results suggest that the fate of a fruit set or degeneration occurred before anthesis in tomato. Auxins, whose levels were independent of pollination and fertilization, play prominent roles in controlling a fruit set in "R35-P," and other hormones are integrated in a synergistic or antagonistic way.

Keywords: fruit set, endogenous hormones, transcriptome, parthenocarpy, tomato

INTRODUCTION

The off-season cultivation of tomato usually involves treatment with exogenous hormone regulators to stabilize a fruit set, which is a common practice in greenhouses with little wind and insects. However, the application of synthetic hormones results in the development of malformed fruits (Martinelli et al., 2009; Takisawa et al., 2019). The use of bumblebees can avoid these drawbacks but requires the maintenance of optimal temperatures to keep them active (Takisawa et al., 2017). All of these agricultural practices increase financial and labor costs for farmers.

Parthenocarpy can overcome these problems by converting the ovary into a developing fruit without pollination and exogenous hormone application, thereby producing seedless fruit. Currently, for highly desired agronomic traits, many vegetable crops, such as tomato (Gorguet et al., 2005, 2008; Takisawa et al., 2017, 2020), cucumber (Wu et al., 2016; Li et al., 2017), and eggplant (Du et al., 2016; Chen et al., 2017) can naturally produce parthenocarpic fruit. The use of parthenocarpic cultivars is considered to be the most cost-effective solution for a stable fruit set under suboptimal environmental conditions. The trait of parthenocarpy has increasingly attracted the interest of scholars and has become a popular topic for research.

Natural parthenocarpy in tomato has been widely studied for its potential use. To date, six parthenocarpic tomato resources and nine different parthenocarpic genes have been reported. "Soressi" and "Montfavet191" carry the *pat* gene (Soressi, 1975; Philouze and Pecaut, 1986), "Severianin" carries the *pat-2* gene (Philouze and Maisonneuve, 1978), "RP75/59" carries the *pat3/pat4* gene (Philouze, 1989), "IL5-1" carries the *pat4.1/pat5.1* gene (Finkers et al., 2007; Gorguet et al., 2008), "IVT-line1" carries the *pat4.2/pat9.1* gene (Zijlstra, 1985; Gorguet et al., 2008), and "MPK-1" carries the *pat-k* gene (Takisawa et al., 2017, 2020). The parthenocarpic genes *pat-2* and *pat-k* have been cloned. The *pat-2* gene is located on chromosome 4 and encodes a zinc-finger homeodomain (ZHD) protein, whereas the *pat-k* gene is located on chromosome 1 and encodes an E-class MADS-box gene, SlAGAMOUSLIKE6 (*SLAGL6*) (Takisawa et al., 2018, 2020). The *Pat* locus was mapped on the long arm of tomato chromosome 3, and the two SCAR markers SSR300 and SSR601 were located distally from the *Pat* locus at 7.5 and 3.7 cM, respectively (Beraldi et al., 2004). The *pat4.1* and *pat4.2* were likely allelic and located on chromosome 4, whose potential candidate gene was *ARF8* (Gorguet et al., 2008). The *pat5.1* and *pat9.1* were located on chromosomes 5 and 9, respectively (Gorguet et al., 2008). The *pat3/pat4* genes were located on chromosome 1 L site 152 and chromosome 6 L site 34 (Vardy et al., 1989).

It has been shown that parthenocarpic tomato is mainly controlled by endogenetic hormones. Hazra et al. (2010) showed that the content of auxin in "Oregon Pride" (*pat-2*) was 10.48 times higher than that of the control at 3 days before anthesis. Dai (1998) pointed out that the IAA content was significantly higher in parthenocarpic tomato than in non-parthenocarpic tomato from 2 days before anthesis to 4 days after anthesis. Furthermore, artificially increasing the endogenous auxin levels of the ovary by introducing *DefH9-iaaM* or *TIR1* or by inhibiting *AUX/IAA*, *ARFs,* or *Aucsia* could also stimulate parthenocarpy in tomato (Pandolfini et al., 2002; Rotino et al., 2005; Wang et al., 2005; Goetz et al., 2007; de Jong et al., 2009; Molesini et al., 2009; Mazzucato et al., 2015). Gibberellins have also been regarded as key regulators in the parthenocarpic fruit set and development of tomato (Gorguet et al., 2005; Serrani et al., 2008). Higher GA levels have been found in the ovaries of natural parthenocarpic tomato *pat*, *pat-2*, and *pat3/pat4* mutants, and endogenous GA levels increase in tomato ovaries, following pollination (Mariotti et al., 2011; and literature cited therein). DELLA proteins are plant nuclear factors that restrain growth and proliferation in response to hormonal signals, and they have been characterized as repressors of gibberellin signaling (Marti et al., 2007). Silencing *DELLA* induces fertilization-independent fruit growth in tomato (Marti et al., 2007). It is generally recognized that the parthenocarpy of these tomato resources is regulated by the coordination of endogenetic hormones, especially auxins and gibberellins (GAs) (Serrani et al., 2008; Pascual et al., 2009; de Jong et al., 2011; Mignolli et al., 2015). Although an increasing number of studies on hormones have been conducted, the mechanism governing parthenocarpy has not been fully elucidated to date.

The parthenocarpy of tomato "R35" is controlled by a recessive gene (Wang et al., 2008), but little is known about its physiological basis and molecular mechanism. In this study, to improve the parthenocarpic rate and growth potential of "R35," we developed a new line, "R35-P," by self-crossing and selection. During this period, we found a non-parthenocarpic plant, "R35-N," and we cultivated it by self-pollination and purification. Biochemical and molecular mechanisms of ovary development in "R35-P" were elucidated by combining the measurement of endogenous hormones with the analysis of DEGs related to plant hormones based on transcriptomic data. This study will not only be helpful to understand the molecular mechanisms by which hormones regulate the fruit set of "R35-P" but may also establish a theoretical foundation for the creation of genetic parthenocarpy in tomato by genetic engineering.

MATERIALS AND METHODS

Plant Materials and Treatment

Parthenocarpic and non-parthenocarpic tomato lines were used in this study. Both lines were isolated from the parthenocarpic tomato line "R35" (Wang et al., 2008) by selfing and selection.

The stigmas of "R35-P" buds were removed 2 days before anthesis to prevent self-pollination, and the flowers emasculated were denoted Pe; "R35-P" was self-pollinated at anthesis and denoted Ps; and "R35-N" was self-pollinated at anthesis and denoted CK. The ovaries of Pe, Ps, and CK were tagged and collected at −2, 0, 3, and 6 days after anthesis (DAA) (**Figure 1**), frozen in liquid N_2, and stored at −80°C until needed for hormone quantification or RNA-seq analysis.

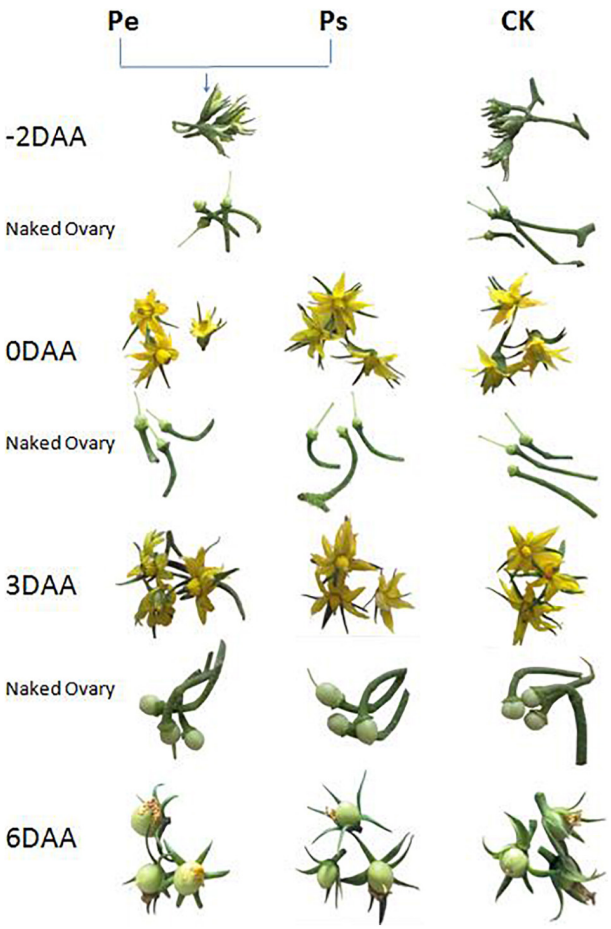

FIGURE 1 | Morphological changes of flowers and ovaries during the early development stages in Pe (stigma removed in "R35-P"), Ps (self-pollination in "R35-P"), and CK (self-pollination in "R35-N").

Measurement of Endogenous Hormone Levels

For endogenous plant hormone analyses, the naked ovaries, receptacles, and intact fruit petioles of Pe, Ps, and CK at −2, 0, 3, and 6 DAA were collected (**Figure 1**). About 1.5 g per bulked sample was used, which did not take from the fixed single plant but from the line consisting of more than 100 individual plants.

The levels of indole-3-acetic acid (IAA), gibberellin ($GA_{1,3,4,5,7}$), abscisic acid (ABA), jasmonic acid (JA), and salicylic (SA) were measured at Zoonbio Biotechnology Co., Ltd. (Nanjing, China). Approximately, 1.0-g ovaries were ground in a precooled mortar that contained a 10-ml extraction buffer composed of isopropanol/hydrochloric acid. The extract with 8-μl internal standard (1 μg/ml) (D-IAA, D-ABA, D-SA, D-GA, and 2HJA) added was shaken at 4°C for 30 min. Then, 20-ml dichloromethane was added, and the sample was shaken at 4°C for 30 min and centrifuged at 13,000 rpm for 5 min at 4°C. The lower, organic phase was extracted, which was dried under N_2 and dissolved in 400-μl methanol (0.1% methane acid) and filtered with a 0.22-μm filter membrane. The purified product was then subjected to high-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) analysis. HPLC analysis was performed, using a poroshell 120 SB-C18 (Agilent Technologies) column (2.1 × 150 mm; 2.7 μm). The mobile phase A solvents consisted of methanol/0.1% methanoic acid, and the mobile phase B solvents consisted of ultrapure water/0.1% methanoic acid. The injection volume was 2 μl. MS conditions were as follows: the spray voltage was 4,500 V; the pressure of the air curtain, nebulizer, and aux gas were 15, 65, and 70 psi, respectively, and the atomizing temperature was 400°C.

RNA Extraction and RNA-Seq Analysis

The RNA preparation, cDNA libraries construction, and Illumina sequencing were performed in Novogene (Beijing, China). Total RNA was extracted from freshly frozen naked ovaries (with receptacles and fruit pedicels), three biological replicates per sample, using the Trizol reagent (Invitrogen, Carlsbad, CA, United States) and purified using the RNeasy Plant Mini kit (Qiagen, Valencia, CA, United States). About 0.5 g ovaries were sampled per biological repeat sample for RNA-seq, and 0.2 mg of which was used for RNA extraction. RNA integrity and purity were assessed by agarose GEL electrophoresis and Agilent 2100 bioanalyzer. Then, mRNA was purified from total RNA, using poly-T oligo-attached magnetic beads and was interrupted into 250–300 bp short segments. The effective concentration of RNA library was higher than 3 nM. First strand cDNA was synthesized, using random oligonucleotide primers. Second strand cDNA synthesis was subsequently performed, using DNA Polymerase I and RNase H. After end repair, the adaptor was added to the cDNA with T_4 DNA ligase. The appropriate fragment ligated with adaptor was selected by AMPure XP beads as a template for PCR amplification. The cDNA library products were sequenced by the Illumina HiSeq 2000 platform.

The raw data obtained by sequencing were quality controlled, which mainly included removal of the reads with adapters, removal of the reads with a ratio of N more than 0.002 (N denotes unidentifiable base information) and removal of the paired reads when the number of low-mass bases in a single-ended read was more than 50% of the read length. Then, the clean reads got by examination of the sequencing error rate and GC content distribution. The clean reads were compared and analyzed with the Hisat2 software[1] and then were spliced. By comparing with the tomato reference genome (Heinz 1706), the genes expression of RNA-seq were obtained, which were expressed by FPKM (fragments per kilobase of transcript sequence per millions base pairs sequenced).

The transcription profiles of the unpollinated and pollinated ovaries at −2, 0, 3, and 6 DAA of Pe, Ps, and CK were compared, with three biological replicates per sample.

GO Enrichment and KEGG Pathway Analyses

The differentially expressed genes (DEGs) were identified through pairwise comparison between the ovaries of Pe, Ps, and CK libraries. The filtered DEGs were subjected to GO

[1]http://ccb.jhu.edu/software/hisat2

enrichment analysis with the GOseq R package, and the pathways of these DEGs were annotated with the BLASTALL program against the KEGG (the Kyoto Encyclopedia of Genes and Genomes) database, respectively. Categories with Padj < 0.05 were considered significantly enriched.

Quantitative Real-Time PCR Analysis

Ten genes were selected randomly for RNA-seq data validation, using qRT-PCR. Total RNA was extracted from ovaries samples of tomato as described above and was reverse-transcribed in a 20-μl reaction system with a cDNA synthesis kit (RR047A, TaKaRa, Dalian, China) according to the instructions of the manufacturers. Then, the products were subjected to 10-fold dilution and used in subsequent qPCR. qPCR was prepared in reactions containing 10-μl 2 × SYBR Green real-time PCR Master Mix (Takara, Dalian, China), 3-μl diluted first-strand cDNA,0.8 μl each PCR primer (10 μM), and distilled water to a 20-μl volume. The amplification procedure was set as follows: 95°C for 30 s, then 40 cycles of 95°C for 5 s, 60°C for 34 s, and 72°C for 45 s. The qRT-PCR was performed in three replicates per sample. Tomato *actin* gene (*Solyc03g078400*) was used as an internal control. The $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) was used for the calculation of relative expression levels of target genes. Primers used for qPCR assays are listed in **Supplementary Table 2**.

Statistical Analysis

All data were presented as the means ± standard errors of three biological replicate samples. The statistical significance of the pairwise comparison among the three groups was tested by the two-way ANOVA method, and $P < 0.05$ was indicated.

RESULTS

Performance of "R35-P" and "R35-N"

"R35-P" shows environmentally stable facultative parthenocarpy under different conditions, and displays accelerated ovary expansion before or at anthesis occasionally (**Figure 2**). It belongs to the infinite growth type, medium growth. The seeded and seedless fruits (parthenocarpic fruits with jelly) of "R35-P" were similar in size on the same plant (**Figure 2**), and single fruit weight is about 80 g, with 3–4 locules (the average ratio of polar and equatorial diameter is about 4/4.3). The fruits are pink and round. The percentage of a parthenocarpic fruit set is higher than 80%, if the stigmas of buds for the first three spikelet flowers in "R35-P" were removed 2 days before anthesis. And it can produce normal seeded fruits or seedless fruits (encounter unfavorable weather conditions) naturally when planted in an open field.

Unpollinated fruits of line "R35-N" will abort, and self-pollinated fruits are slightly larger than the seeded/seedless fruits of "R35-P"; and single fruit weight is about 100 g with 4–5 locules (the average ratio of polar and equatorial diameter is about 5/5.1) (**Figure 2**). "R35-N" has a genetic background similar to that of "R35-P."

Endogenous Hormone Levels in "R35-P" and "R35-N"

The levels of endogenous auxins (IAA), gibberellins ($GA_{1,3,4,5,7}$), abscisic acid (ABA), jasmonic acid (JA), and salicylic acid (SA) in naked ovaries of Pe, Ps, and CK were determined at four periods (−2, 0, 3, and 6 DAA) (**Figure 3**).

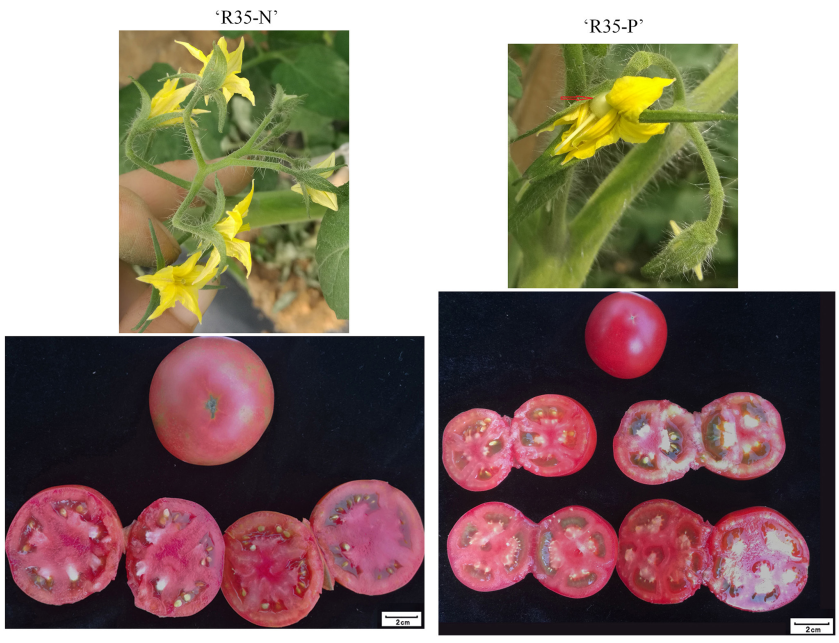

FIGURE 2 | The phenotypes of flowers and tomato fruits of "R35-N" and "R35-P." The seeded fruits of "R35-N" and "R35-P" obtained by natural pollination at an open field, while the seedless fruits of "R35-P" that obtained flowers emasculation carried 2 days before anthesis.

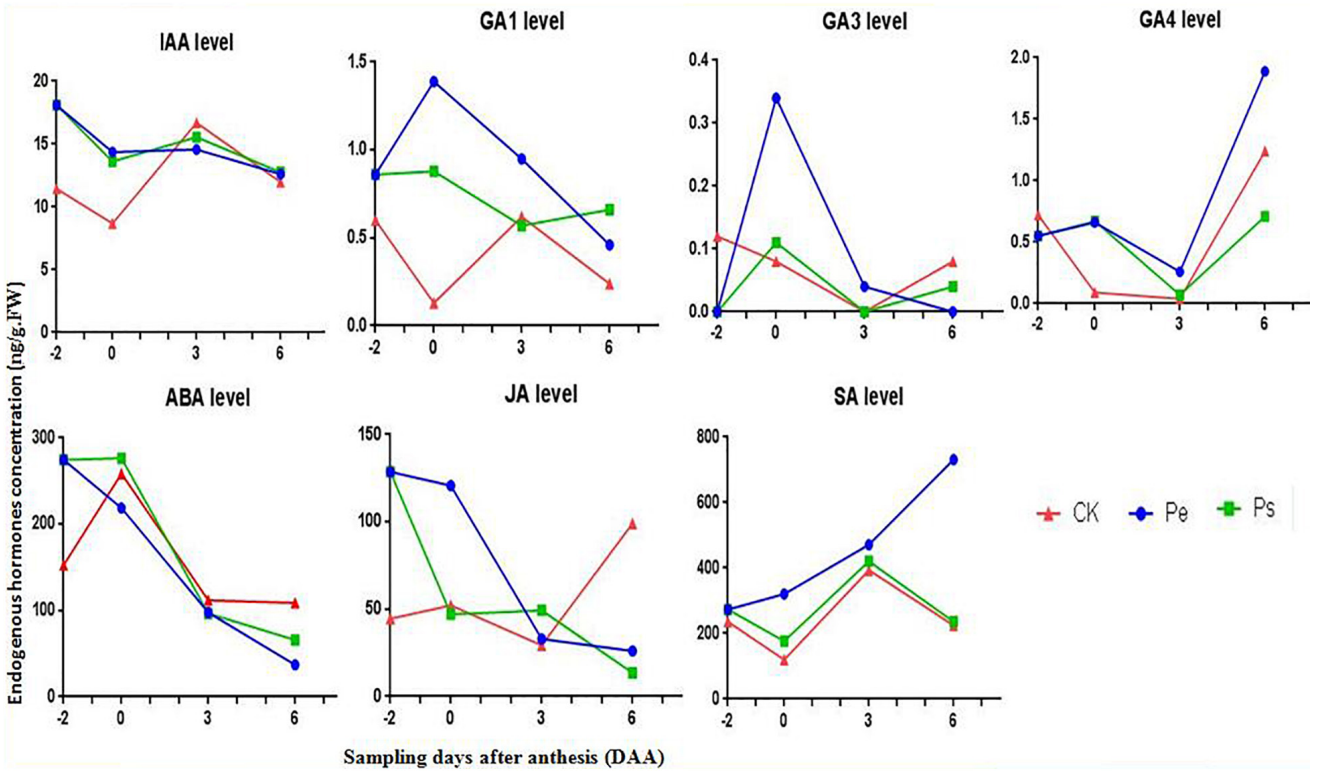

FIGURE 3 | Comparison of the phytohormone levels (IAA, ABA, $GA_{1,3,4}$, SA, and JA) between Pe (stigma removed in "R35-P"), Ps (self-pollination in "R35-P"), and CK (self-pollination in "R35-N") at −2, 0, 3, 6 days after anthesis (DAA).

A significantly higher IAA content (nearly twofold) was found in "R35-P" regardless of whether it was pollinated than in CK at −2 and 0 DAA, and a similar content was found at 3 and 6 DAA. The levels of IAA were almost identical in Pe and Ps, and the changes in the IAA level were similar in Pe, Ps, and CK at all four stages. This result indicated that IAA was associated with genotype (parthenocarpic or not) but not with fruit setting development direction (seeded or seedless in parthenocarpic tomato); IAA was independent of pollination.

The levels of GA_1, GA_3 and GA_4 were notably low, and GA_5 and GA_7 were under the detection limits at these stages. The variation tendencies of GA_1, GA_3, and GA_4 were similar in Pe and Ps, contrary to CK, from −2 DAA to 0 DAA. The contents of GA_1 and GA_3 were significantly higher in Pe than in Ps, while GA_4 was almost equal in Pe and Ps at 0 DAA. Additionally, only the content of GA_1 in CK was increased, following pollination (0–3 DAA), while that in the other treatments was reduced.

The content and the change trend of SA were similar in Ps and CK, but the content of SA was lower in Ps and CK than in Pe at all four stages.

The content of JA was considerably higher in Pe and Ps than in CK at −2 DAA, and a pronounced difference was observed between Pe and Ps or CK at 0 DAA. The variation tendencies of JA varied among Pe, Ps, and CK at the studied periods.

Except at −2 DAA, the ABA content of Pe was lower than that of Ps and CK. The highest content appeared at 0 DAA. Following pollination (0–3 DAA), the ABA contents in CK and Ps were reduced.

Analysis of RNA-Seq Data

Based on morphological observations and endogenous hormone level measurements, we found that the tomato ovaries were consistently triggered to set fruit before 3 DAA. Therefore, only the data of high-throughput RNA-seq of the samples at −2 and 0 DAA were used to analyze in this study. In total, 277 million reads, with approximately 5.5 million reads being obtained from each sample, were aligned with the reference genome (**Supplementary Table 1**).

Based on the FPKM method, the transcript abundance of each gene was analyzed. Padj < 0.05 was used as the threshold to judge the significance of the differences in gene expression. A total of 8,460 differentially expressed genes (DEGs) were detected among Pe, Ps, and CK at −2 DAA and 0 DAA; of which, 7,479 DEGs were detected between P (Pe/Ps) and CK at −2 DAA, while 2,378 DEGs were identified among Pe, Ps, and CK at 0 DAA (**Figure 4A**). Among the DEGs identified at −2 DAA, 1,323 DEGs were found between Ps and CK, 1,163 were found between Pe and CK, and 1,000 were found between Pe and Ps. Detailed information on these common and specific DEGs was screened by Venn analysis (**Figures 4B,C**). A total of 39 DEGs were common to Pe, Ps, and CK at 0 DAA (**Figure 4B**), and 31 were common at −2 DAA and 0 DAA (**Figure 4C**).

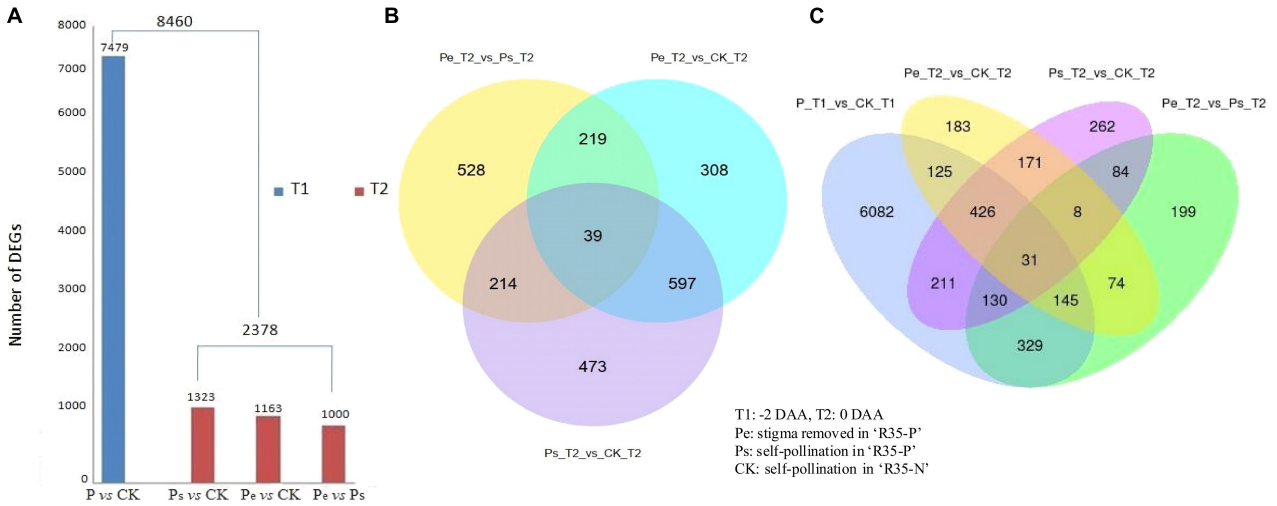

FIGURE 4 | DEGs (differentially expressed genes) of the RNA-seq data. **(A)** The numbers of DEGs distribution (Padj < 0.05). **(B)** The Venn diagrams of 2,378 DEGs at 0 days after anthesis (DAA). **(C)** The Venn diagrams of 8,460 DEGs at −2 DAA and 0 DAA.

To further validate the reliability of the RNA-seq results, 10 genes (**Supplementary Table 2**) were randomly selected for a qRT-PCR comparison of their expression levels in the aforementioned samples at −2 DAA and 0 DAA. The results showed that most of the genes had a similar expression pattern between qRT-PCR and RNA-seq analyses (**Figure 5**), except that the expression of two genes (*Solyc01g111310* and *Solyc02g037530*) had different patterns at 0 DAA, indicating that the RNA-seq data were reliable and suitable for further analysis.

Functional Enrichment Analysis of DEGs

The 7,479 DEGs (**Supplementary Table 4**) at −2 DAA and 2,378 DEGs (**Supplementary Table 5**) at 0 DAA among Pe, Ps, and CK identified by RNA-seq were subjected to GO and KEGG enrichment analyses.

Gene ontology (GO) enrichment analysis provided all the GO terms that were significantly enriched for DEGs. The DEGs were assigned to 30 GO terms at −2 DAA and 0 DAA, which were summarized into three major GO categories: biological process (BP), cellular components (CC), and molecular function (MF) (**Figures 6A,B**). In the BP category, at −2 DAA, the largest number of DEGs was clustered into 26 groups, of which "cellular processes," "metabolic processes," "organic substance metabolic process" and "single-organism process" were dominant. The term "response to hormone" was also found in BP. Only one term, "cell periphery," was found in the CC category at −2 DAA. Three terms were found in the MF category, with the most abundant DEGs being associated with "catalytic activity" at −2 DAA. Compared to the GO terms at −2 DAA, there were significant differences at 0 DAA. The number of groups in the BP category decreased from 26 to 11, and most DEGs were distributed in "metabolic processes," while "response to hormone" disappeared. The number of groups in MF rose from 3 to 15, but "catalytic activity" was still the most abundant. The CC category had the fewest DEGs. The DEGs related to "metabolic processes" and "catalytic activity" were dominant at the two periods, suggesting that the ovaries were highly vigorous. To further identify the biological pathways that were active in parthenocarpic tomato, the DEGs were mapped to the reference pathways in the KEGG database (**Figures 6C,D**). At −2 DAA, the most significant pathways consisted of "biosynthesis of secondary metabolites," "plant hormone signal transduction," and "plant-pathogen interaction." At 0 DAA, the DEGs were primarily distributed in "metabolic pathways," "biosynthesis of secondary metabolites," "starch and sucrose metabolism" pathways. As with GO enrichment, the "plant hormone signal transduction" pathway was only present at −2 DAA. Therefore, we inferred that hormone-related genes were actively expressed before anthesis to regulate a fruit set in tomato. Thus, we further analyzed plant hormone-related pathways and their gene expression patterns.

Differentially Expressed Genes in Response to Phytohormones

To investigate whether changes in hormone levels were associated with changes in the activity of hormone metabolism genes, transcript levels of hormone metabolism genes were analyzed. KEGG enrichment analysis identified 175 DEGs (**Supplementary Table 3**) in the plant hormone signaling pathway, including IAA, CTK (cytokinin), GAs, ABA, ETH (ethylene), BR (brassinosteroid), JA, and SA, which were found at −2 DAA and 0 DAA, especially at −2 DAA. The differential expression of these genes in the studied samples is displayed in heatmaps (**Figure 7**). DEGs in Pe, Ps, and CK demonstrated similar gene expression patterns at the same period, and the gene expression profiles of Pe and Ps were clustered together, except for GAs at 0 DAA. Furthermore, there was a darker color and stronger contrast in the heatmaps for the DEGs in samples at −2 DAA than those at 0 DAA (**Figure 7**), which indicated that hormone-related genes were actively expressed before anthesis to regulate a fruit set in

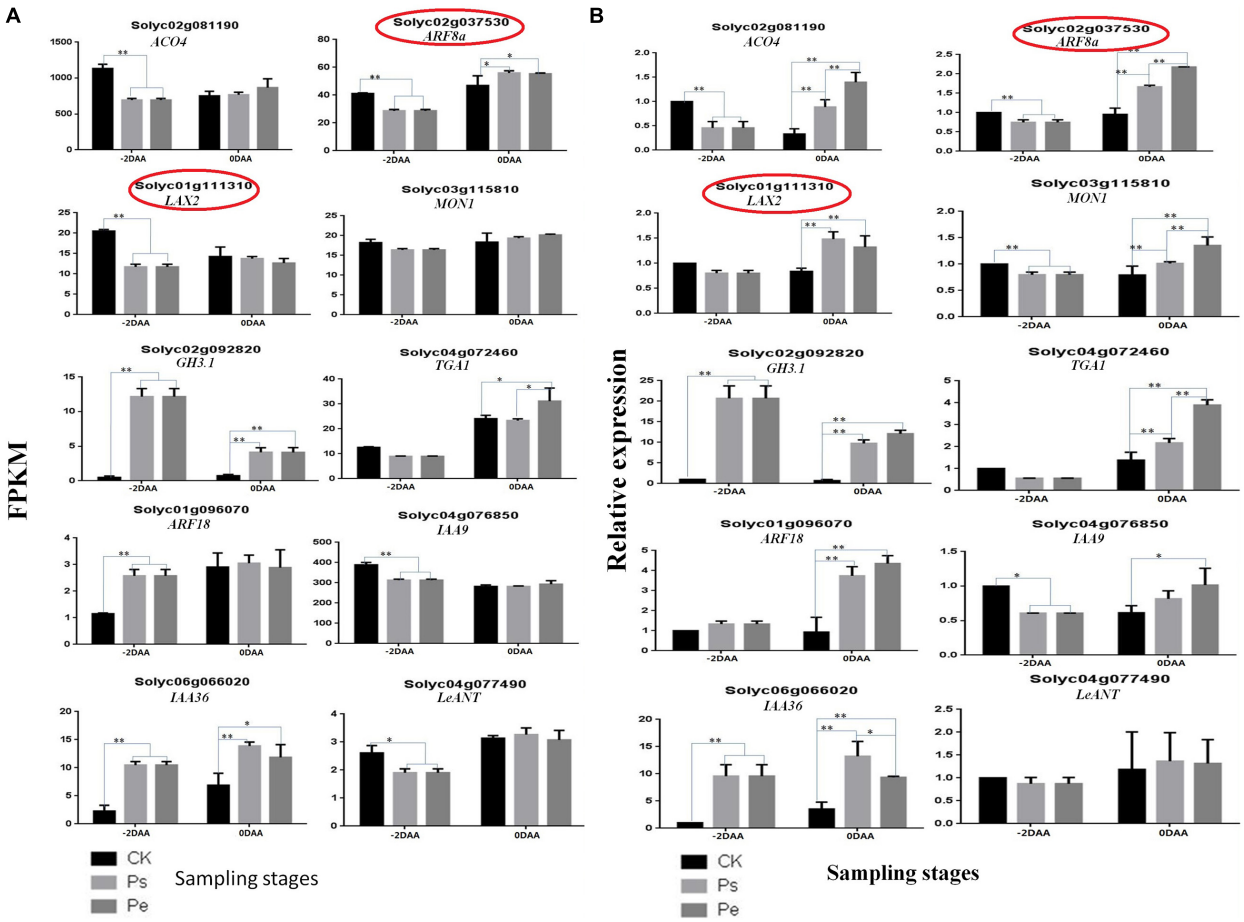

FIGURE 5 | Validation of RNA-seq data by RT-PCR. **(A)** Expression levels of 10 DEGs were detected by RT-PCR, normalized based on the expression of the internal control gene, at −2 DAA (days after anthesis) and 0 DAA. **(B)** Corresponding genes expression were detected by RNA-seq based on the FPKM method (Padj < 0.05). The genes denoted by○ had a different trend at 0 DAA between RT-PCR and RNA-seq. The bars represent standard errors of three replicates (n = 3). $^{*}P < 0.01$, $^{**}P < 0.001$ by two-way ANOVA.

tomato. It was consistent with previous findings regarding GO and KEGG enrichment analysis.

The number of DEGs involved in different hormone metabolism pathways showed that the greatest number involved IAA followed by ABA, ETR, CTK, SA, and BR, and the fewest involved JA and GAs. The Auxin-signaling pathway, which is regarded as a key regulator of a fruit set in parthenocarpy, included 63 auxins-related genes. These genes encode members of the SAUR, AUX/IAA, GH3, AUX1, TIR1, and ARF protein families. Most of these genes are involved in auxin response genes, such as *SAUR* (covered 28 genes), *AUX/IAA* (covered 16 genes) and *ARF* (covered 5 genes). The functions of SAUR proteins are capable to modulate auxin synthesis and transport (Zhu et al., 2014). In this study, most *SAURs*, such as *Solyc01g110647.1*, *Solyc01g110730.3*, *Solyc03g033590.1*, *Solyc10g018340.1* and so on, showed upregulated expression in "R35-P" and downregulated expression in "R35-N" at −2 DAA (**Figure 7**). In the auxin-signaling components *AUX/IAAs* and *ARFs* family, *IAA1* (*Solyc09g083280.3*), *IAA2* (*Solyc06g084070.3*), *IAA3* (*Solyc09g065850.3*), *IAA16* (*Solyc01g097290.3*), *IAA19* (*Solyc03g120380.3*), *IAA26* (*Solyc03g121060.3*), *IAA29* (*Solyc08g021820.3*), *IAA35* (*Solyc07g008020.3*), *IAA36* (*Solyc06g066020.3*), *ARF1* (*solyc01g103050.3*), and *ARF18* (*Solyc01g096070.3*) showed upregulated expression in "R35-P" and downregulated expression in "R35-N," while *IAA8* (*Solyc12g007230.2*), *IAA9* (*Solyc04g076850.3*), *IAA13* (*Solyc09g090910.2*), *IAA17* (*Solyc06g008590.3*), *IAA27* (*Solyc03g120500.3*), *ARF5* (*Solyc04g081235.1*), *ARF7* (*Solyc07g042260.3*), and *ARF9* (*Solyc08g082630.3*) were the opposite at −2 DAA (**Figure 7**). And the two libraries showed a slight difference among Pe, Ps, and CK, except the *IAA7* (*Solyc06g053830.3*) and *IAA14* (*Solyc09g083290.3*) showed different expression at 0 DAA (**Figure 7**). Validation by qRT-PCR showed that the expressions of auxin related genes *ARF18* (*Solyc01g096070.3*), *LAX2* (*Solyc01g111310.3*) and *IAA9* (*Solyc04g076850.3*) were extremely significant different among Pe, Ps and CK at −2 DAA, while which were similar expressed at 0DAA (**Figure 5**). The expressions of auxin-related genes *GH3.1* (*Solyc02g092820.3*) and *IAA36* (*Solyc06g066020.3*) were significantly or extremely significant

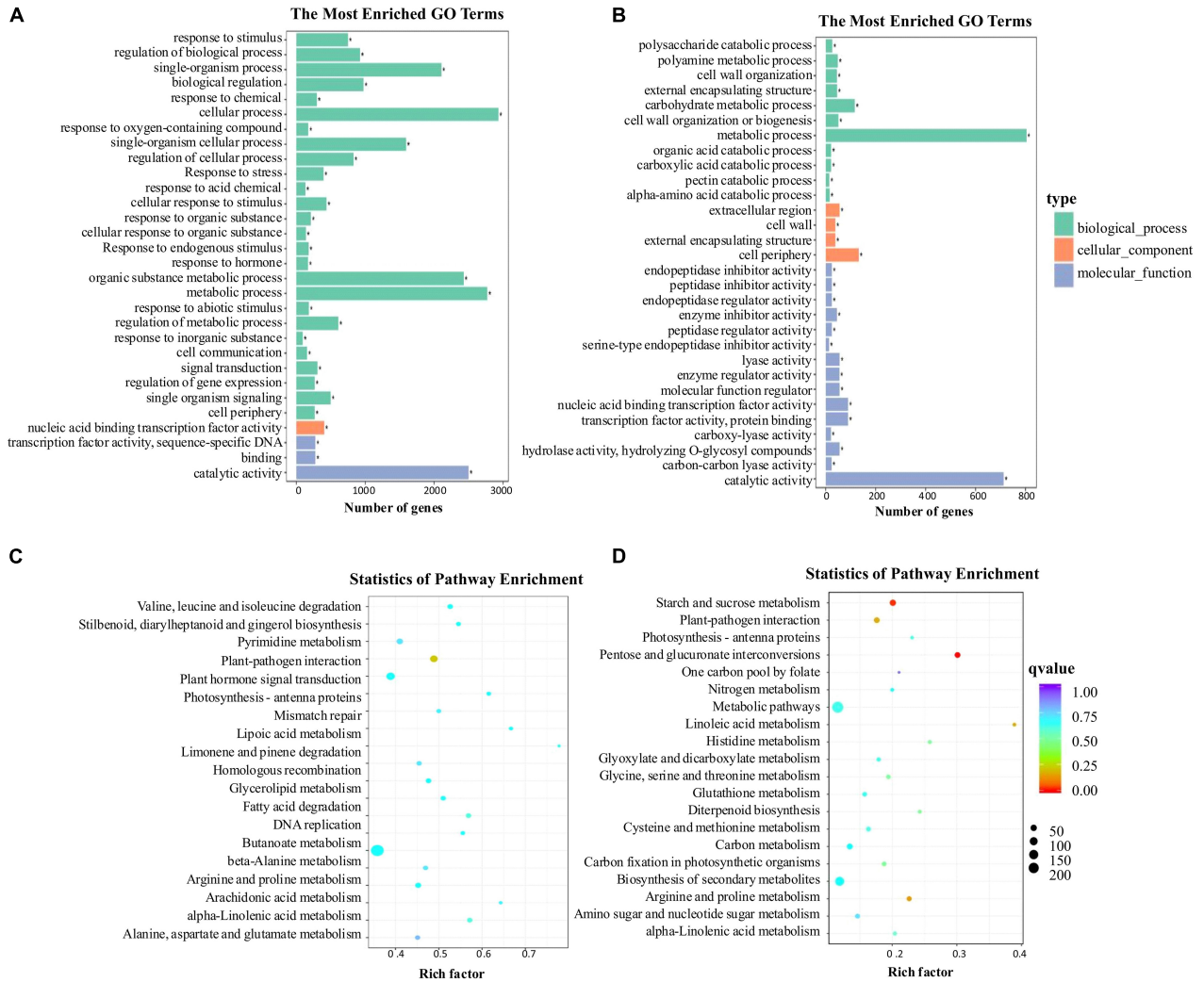

FIGURE 6 | Functional categorization with the DEGs (differentially expressed genes). **(A)** GO analysis was performed with the DEGs at −2 DAA (days after anthesis). **(B)** GO analysis was performed with the DEGs at 0 DAA. **(C)** KEGG analysis was performed with the DEGs at −2 DAA. **(D)** KEGG analysis was performed with the DEGs at 0 DAA.

different at −2 and 0 DAA among Pe, Ps, and CK (**Figure 5**). Twenty-seven related genes were found in the ABA-signaling pathway, which encodes members of the PYR/PYL, PP2C, SnPK2, and ABF protein families. The "ABA-PYR/PYL/RCAR-PP2C-SnRK2" signaling pathway of ABA may be suggested in tomato. Some *PP2C* genes (*Solyc03g121880.3*, *Solyc05g052980.3*, *Solyc06g076400.3*, *Solyc07g040990.3*) and *PYR/PYL* genes (*Solyc06g050500.2*, *Solyc08g076960.1*, *Solyc08g082180.3*, *Solyc09g015380.1*, *Solyc10g076410.1*, and *Solyc10g085310.1*) are greater expressed in "R35-P" than in "R35-N" at −2 DAA (**Figure 7**). In the ETH-signaling pathway, 22 related genes-encoding members of the ETR, CTR1, MPK6, EBF1/2, EIN3, ERF1/2, and EIN2 families were detected. Most of these genes mainly downregulated expressed, and more upregulated genes, such as *EIN3* (Solyc01g014480.3), *ERF1* (Solyc05g051200.1), *ETR* (Solyc05g055070.3), *MPK6* (Solyc06g005170.3), and *CTR1* (Solyc10g083610.2) were found in "R35-P" than in "R35-N" at −2 DAA (**Figure 7**). In the CTK-signaling pathway, 16 related genes-encoding members of the CRE1, AHP, B-ARR, and A-ARR families were detected. In the SA-signaling pathway, 16 related genes-encoding members of the NPR1, TGA, and PR-1 families were detected. In the BR-signaling pathway, 14 related genes-encoding members of the BSK, BZR1/2, CYCD3, BAK1, BRI1, and BIN2 families were detected. In the JA-signaling pathway, 13 related genes-encoding members of the JAR1, JAZ, MYC2, and COI1 families were detected. In the GA-signaling pathway, only four related genes-encoding members of the G1D1, DELLA, and TF families were detected.

DISCUSSION

The fruit set of plants largely depends on the biosynthesis and crosstalk of phytohormones. The activities of two plant growth

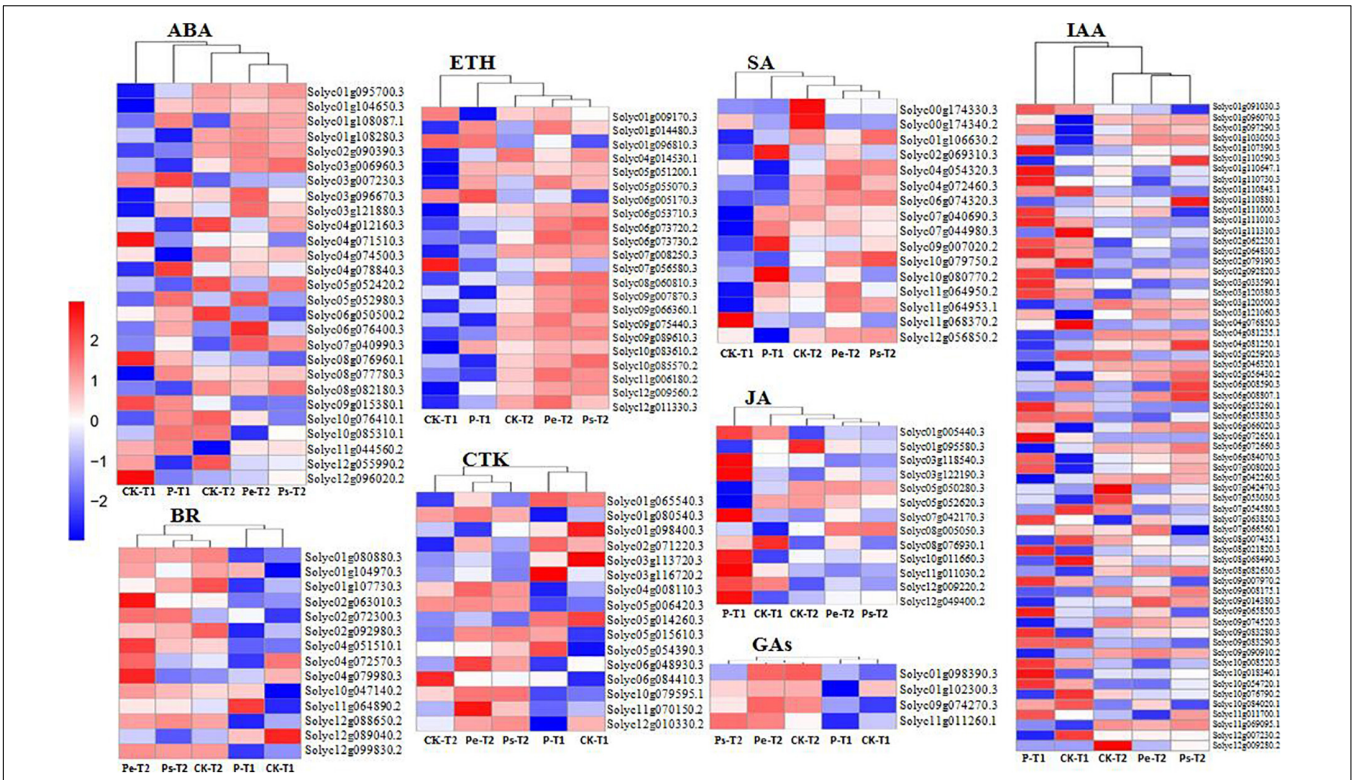

FIGURE 7 | The heatmaps about hormones biological functions DEGs enriched by KEGG analysis. The FPKMs were homogenized by Z-score, which was a log based on multiple and folded changes with pairwise comparisons. The red and the blue colors indicate high and low transcript abundance.

hormones, auxins, and gibberellins are considered to be the main stimulus in the induction of a fruit set (Fos et al., 2000; Olimpieri et al., 2007; Serrani et al., 2008; de Jong et al., 2011; Mariotti et al., 2011; Mignolli et al., 2015; Tang et al., 2015).

In our study, the IAA content in the ovaries of CK was significantly lower than that in "R35-P" at preanthesis and anthesis, while it increased rapidly to the level of "R35-P" (pollinated or not), following pollination at postanthesis. This result was consistent with the findings of previous research on the parthenocarpic fruits of the *pat* mutant (Mapelli et al., 1978). The IAA content of ovaries in Ps, which did not increase following pollination, was almost equal to that in Pe at all-time points. Therefore, we suggested that the threshold concentration of IAA in the ovaries of "R35-P" was sufficient to promote a fruit set and did not require pollination stimulation. Additionally, the largest number of IAA-related DEGs was shown in hormone metabolism pathways at the transcript level. ARF and Aux/IAA proteins physically interact to regulate auxin response, are the best-characterized pates, and crucial in a triggering fruit set (Wang et al., 2005; Dorcey et al., 2009; Ruiu et al., 2015). Consistent with them, our results showed that some Aux/IAAs and ARFs, which accounted for 21/63 of the IAA-related DEGs, displayed dramatic shifts in their expressions between "R35-P" and "R35-N," especially at −2 DAA. In our study, *IAA1*, *IAA2*, *IAA3*, *IAA16*, *IAA19*, *IAA26*, *IAA29*, *IAA35,* and *IAA36*, *ARF1*, *ARF9*, and *ARF18* upregulated expression in "R35-P" at −2 DAA, while *IAA8*, *IAA9*, *IAA13*, *IAA17*, *IAA27*, *ARF5*, and *ARF7* were reversed (**Figure 7** and **Supplementary Table 3**). In agreement with previous reports, *IAA2* and *ARF9* upregulation expression (*IAA2*, Pascual et al., 2009; *IAA2*, Vriezen et al., 2008; *ARF9*, Serrani et al., 2008; *ARF9,* Wang et al., 2009), and *IAA9* and *ARF7* downregulated expression (*IAA9*, Wang et al., 2009; *ARF7*, de Jong et al., 2009) were good for the parthenocarpic fruit set (**Figure 7** and **Supplementary Table 3**). Other *Aux/IAAs* and *ARFs* were inferred to be related to a parthenocarpic fruit set in tomato too. Validated by qRT-PCR, *LAX2* was significantly different at 0 DAA, and *GH3.1* was extremely significantly different between "R35-P" and "R35-N" at 0 and −2 DAA. *LAX2* related to IAA transport with a possible role in the regulation of auxin influx (Pattison and Catalá, 2012). As IAA-amido synthetase, GH3 can promote an IAA conjugation to amino acids to maintain auxin homeostasis (Molesini et al., 2020). In this study, *GH3.1* may be an important part in balancing the auxin synthesis and metabolism to ensure a fruit set in "R35-P" under any favorable or unprofitable conditions. Overall, these studies revealed that parthenocarpy can be achieved by manipulating auxin action at different levels, by acting on its biosynthesis, signaling cascade, and transport, corroborating the crucial role played by this hormone in the control of a fruit set (Molesini et al., 2020). Taken together, these results indicated that auxins played a pivotal role in controlling the fruit set of "R35-P."

In natural parthenocarpic tomato *pat*, *pat-2*, and *pat3/pat4* mutants, higher GA levels have been found in the ovaries, and endogenous GA levels increase in tomato ovaries, following

pollination (Mariotti et al., 2011; and the literature cited therein). These changes were not observed in our study. The levels of some active GAs, such as GA_1, GA_3, and GA_4, were quite low, and GA_5 and GA_7 were under the detection limits at all-time points, although the GA_1, GA_3, and GA_4 levels in Pe were higher than those in Ps and CK at most time points. Additionally, except for the content of GA_1 in CK, the contents of GA_3 and GA_4 in Ps and CK were reduced, following pollination (after 0 DAA). It was evident that the expression of *GA20ox* genes, which leads to increased synthesis of GA20, the precursor of active GAs, is a key regulatory step in the GA biosynthetic pathway in *pat* mutants (Olimpieri et al., 2007) and in *pat-2* mutants (Fos et al., 2000). However, *GA20ox* genes were not found among the DEGs related to GAs in our transcriptomic study. This discrepancy between our results and previous investigations might be due to the difference in parthenocarpic mutant genes in "R35-P." Fos et al. (2000) reported that there were significant differences in the levels and variation tendencies of GAs (except GA_{20}) in ovaries in different tomato lines with different genetic backgrounds. Differences in the alteration of GA metabolism appeared in *pat-2* and *pat3/pat4* (Srivastava and Handa, 2005). Therefore, we inferred that the effects of GAs on the induction of a fruit set vary with the different parthenocarpic tomato sources. The application of exogenous auxins leads to parthenocarpic development with filled locules, while treatment with GAs leads to fruits with almost empty locular cavities (Serrani et al., 2007). The "R35-P" set parthenocarpic fruits with completely filled locules. Based on the findings described above, the parthenocarpy in "R35-P" might be independent of GA action.

In this study, a sustained gradual decrease in ABA content was found in Pe at all the studied periods, while ABA levels were the highest in the ovaries at anthesis (0 DAA) and dramatically declined after pollination/fertilization in Ps and CK. These results are consistent with previous findings (Srivastava and Handa, 2005; Mariotti et al., 2011). The ovary of tomato ceases to undergo cell division shortly (1–2 days) before anthesis and enters an "ovary arrest" state (Klap et al., 2017). ABA and ETH are hormones that inhibit plant growth. Large numbers of ABA and ETH genes were found in DEGs in hormone metabolism pathways at the transcript level, which were in agreement with the previous results (Vriezen et al., 2008). ABA and ETH concentrated highly before a fruit set, and then the expression of ETH and ABA was attenuated with pollination, suggesting that they are involved as antagonists of IAA and GAs to keep the pre-anthesis unpollinated ovary in a temporally protected and dormant state (Vriezen et al., 2008; Pascual et al., 2009). We also inferred that ABA and ETH may be involved as antagonists of IAA (IAA levels were the lowest at 0 DAA) to keep the pre-anthesis unpollinated ovary in a temporally protected and dormant state at anthesis. The ovary subsequently changes to an active state upon pollination and a fruit set in Ps and CK, while, in the parthenocarpic line, there is no state of "ovary arrest" at anthesis; therefore, the ABA content sustainably decreases in Pe. The differences accompanying this state led to the differential expression of CTK genes, which play a critical role in the stimulation of cell division. These results suggested that ABA and ETH may also play an important role in the regulation of a fruit set.

JA and SA are two major defense hormones found in response to wounding and defense against insect herbivores. Their roles in a tomato fruit set are not clearly understood. The higher level of JA in Pe may be caused by the wounding after removal of stigmas before anthesis. In addition, the variation tendency of JA differed among Pe, Ps, and CK. Therefore, JA may not be involved in the parthenocarpy of tomato. Unlike JA, the variation tendency of SA was similar in Ps and CK and lower than that in Pe at all four stages. We inferred that SA could respond to pollination/fertilization, and a higher content of SA is favorable to parthenocarpy in tomato.

Overall, high levels of IAA, GA_1, and SA before anthesis and low levels of ABA and ETH were determined to be more conducive to the development of parthenocarpic tomato in this study. Auxins play a major role in a tomato fruit set (Vriezen et al., 2008; Pascual et al., 2009; Mariotti et al., 2011). Mariotti and others (2011) also suggested that elevated auxin levels in the ovaries initiate a cascade of events that finally promote fruit growth. These events include crosstalk with other hormones, especially the activation of GA metabolism, the increase in CTK, and the decrease in ABA levels (Mariotti et al., 2011). Consistent with these findings, we inferred that auxins govern the fate of fruit initiation events and play pivotal roles in integrating other endogenous hormone participation to promote a fruit set. Different parthenocarpic systems in tomato, such as normal parthenocarpic genotypes (Fos et al., 2000; Olimpieri et al., 2007; Pascual et al., 2009; Hazra et al., 2010; Takisawa et al., 2018, 2019; this study) or pollination-independent driven by hormonal treatment (Serrani et al., 2007, 2008; Vriezen et al., 2008; de Jong et al., 2009; Mariotti et al., 2011; Tang et al., 2015) or genetic engineering of phytohormone-related genes (Goetz et al., 2007; Marti et al., 2007; Martinelli et al., 2009; Molesini et al., 2009; Mazzucato et al., 2015; Mignolli et al., 2015) identified different pivotal phytohormones, and hormones-related genes showed differential expression in the fruit set and developing ovaries. In addition, the seedless fruits of these partenocarpic systems showed difference in fruit size or morphology or locular cavities and so on between the corresponding seeded fruits except with or without seeds. Collectively, this revealed that the pathway of the parthenocarpic fruit set and development might be varied in different parthenocarpic systems.

The earliest phase (up to anthesis) of fruit development involves the decision to abort or to continue with growth, and the molecular events observed in *pat-2* ovaries before anthesis, which are associated with parthenocarpic fruit growth, may modify the hormone content of the ovary before pollination (Fos et al., 2000). Consistent with this result, the difference in hormone contents mainly appeared at 0 DAA, and the DEGs present at the transcriptional level mainly appeared at −2 DAA in our study.

In this study, transcriptomic analyses showed that there were 7,479 DEGs at −2 DAA (**Supplementary Table 4**) and 2,378 DEGs (**Supplementary Table 5**) at 0 DAA among Pe, Ps, and CK, and the genes involved in hormone metabolism pathways composed only a fraction of those DEGs. Not only the complex hormonal regulatory network but also the interactions of many

other genes controlled various aspects of the parthenocarpic fruit set. Our continuing studies on the mapping of major parthenocarpic QTLs to identify major parthenocarpic candidate genes and deeply analyze the transcriptome profile data may enable us to gain a better understanding of the flower-to-fruit transition mechanism of parthenocarpy in "R35-P."

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: NCBI repository, Accession number: PRJNA710539 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA710539).

AUTHOR CONTRIBUTIONS

SZ, JS, and WY conceived and designed the research. SZ, XG, ZH, LW, and LZ performed the experiments, conducted the fieldwork, and analyzed the data. SZ, XG, WY, HS, and LZ wrote the original manuscript. All the authors contributed to the article and critically revised the article.

FUNDING

This work was supported by the National Natural Science Foundation of China (NSFC 31660572), the Modern Biological Breeding Program of Gansu Academy of Agricultural Sciences (2019 GAAS06), and the Natural Science Foundation of Shandong Province (ZR2020MC153 and ZR2020MC135).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.671713/full#supplementary-material

REFERENCES

Beraldi, D., Picarella, M. E., Siressi, G. P., Mazzucato, A. (2004). Fine mapping of the parthenocarpic fruit (pat) mutation in tomato. *Theor. Appl. Genet.* 108, 209–216. doi: 10.1007/s00122-003-1442-6

Chen, X., Zhang, M., Tan, J., Huang, S., Wang, C., Zhang, H., et al. (2017). Comparative transcriptome analysis provides insights into molecular mechanisms for parthenocarpic fruit development in eggplant (Solanum melongena L). *PLoS One* 12:e179491.

Dai, H. (1998). Relationship between parthenocarpy and endogenous hormone of tomato. *J. Changjiang Vegetab.* 7, 23–25.

de Jong, M., Wolters-Arts, M., Feron, R., Mariani, C., Vriezen, W.H. (2009). The Solanum lycopersicum auxin response factor 7 (*SlARF7*) regulates auxin signaling during tomato fruit set and development. *Plant J.* 57, 160–170. doi: 10.1111/j.1365-313X.2008.03671.X

de Jong, M., Wolters-Arts, M., García-Martínez, J. L., Mariani, C., Vriezen, W. H. (2011). The Solanum lycopersicum AUXIN RESPONSE FACTOR 7(*SlARF7*) mediates cross-talk between auxin and gibberellins signalling during tomato fruit set and development. *J. Exp. Botan* 62, 617–626. doi: 10.1093/jxb/erq293

Dorcey, E., Urbez, C., Blazquez, M., Carbonell, J., Perez, M. A. (2009). Fertilization-dependent auxin response in ovules triggers fruit development through the modulation of gibberellin metabolism in Arabidopsis. *Plant J.* 58, 318–332. doi: 10.1111/j.1365-313X.2008.03781.x

Du, L., Bao, C. L., Hu, T. H., Zhu, Q. M., Hu, H. J., He, Q. Y., et al. (2016). SmARF8, a transcription factor involved in parthenocarpy in eggplant. *Mol. Genet. Genomics* 291, 93–105. doi: 10.1007/s004387-015-1088-5

Finkers, R., van Heusden, A. W., Meijer Dekens, F., van Kan, J. A. L., Maris, P., Lindhout, P. (2007). The construction of a Solanum habrochaites LYC4 introgression line population and the identification of QTLs for resistance to Botrytis cinerea. *Theor. Appl. Genet*. 114, 1071–1080.

Fos, M., Nuez, F., García-Martínez, J. L. (2000). The gene *pat-2*, which induces natural parthenocarpy, alters the gibberellin content in unpollinated tomato ovaries. *Plant Physiol.* 122, 471–480.

Goetz, M., Hooper, L. C., Johnson, S. D., Rodrigues, J. C, Vivian-Smith, A., Koltunow, A. M. (2007). Expression of aberrant forms of AUXIN RESPONSE FACTOR8 stimulates parthenocarpy in Arabidopsis and tomato. *Plant Physiol.* 145, 351–366. doi: 10.1104/pp.107.104174

Gorguet, B., Eggink, P. M., Ocana, J., Tiwari, A., Schipper, D., Finkers, R., et al. (2008). Mapping and characterization of novel parthenocarpy QTLs in tomato. *Theor. Appl. Genet*. 116, 755–767.doi: 10.1007/s00122-007-0708-9

Gorguet, B., van Heusden, A. W and Lindhout, P. (2005). Parthenocarpic fruit development in tomato. *Plant Biol.* 7, 131–139.doi: 10.1055/s-2005-837494

Hazra, P., Dutta, A. K., Chatterjee, P. (2010). Altered gibberellin and auxin levels in the ovaries in the manifestation of genetic parthenocarpy in tomato (*Solanum lycopersicum*). *Curr. Sci.* 99, 1439–1443.

Klap, C., Yeshayahou, E., Bolger, A. M., Arazi, T., Gupta, S. K., Shabtai, S., et al. (2017). Tomato facultative parthenocarpy results from SlAGAMOUS-LIKE 6 loss of function. *Plant Biotechnol. J.* 15, 634–647. doi: 10.1111/pbi.12662

Li, J., Xu, J., Guo, Q. W., Wu, Z., Zhang, T., Zhang, K.J., et al. (2017). Proteomic insight into fruit set of cucumber (Cucumis sativus L.) suggests the cues of hormone-independent parthenocarpy. *BMC Genomics* 18:896. doi: 10.1186/s12864-017-4290-5

Livak, K. J., Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 25, 402-408. doi: 10.1006/meth.2001.1262

Mapelli, S., Frova, C., Torti, G., Soressi, G. P. (1978). Relationship between set, development and activities of growth regulators in tomato fruits. *Plant Cell Physiol.* 19, 1281–1288. doi: 10.1016/S0031-9422(00)89308-5

Mariotti, L., Picciarelli, P., Lombardi, L., Ceccarelli, N. (2011). Fruit-set and early fruit growth in tomato are associated with increases in indoleacetic acid, cytokinin, and bioactive gibberellin contents. *J. Plant Growth Regul.* 30, 405–415. doi: 10.1007/s00344-011-9204-1

Marti, C., Orzaez, D., Ellul, P., Moreno, V., Carbonell, J., Granell, A. (2007). Silencing of *DELLA* induces facultative parthenocarpy in tomato fruits. *Plant J.* 52, 865–876. doi: 10.1111/j.1365-313X.2007.03282.x

Martinelli, F., Uratsu, S. L., Reagan, R. L., Chen, Y., Tricoli, D., Fiehn, O., et al. (2009). Gene regulation in parthenocarpic tomato fruit. *J. Exp. Bot*. 60, 3873–3890. doi: 10.1093/jxb/erp227

Mazzucato, A., Cellini, F., Bouzayen, M., Zouine, M., Mila, I., Minoia, S., et al. (2015). A TILLING allele of the tomato *Aux/IAA9* gene offers new insights into fruit set mechanisms and perspectives for breeding seedless tomatoes. *Mol. Breeding* 35:22. doi: 10.1007/s11032-015-0222-8

Mignolli, F., Vidoz, M. L., Mariotti, L., Lombardi, L., Picciarelli, P. (2015). Induction of gibberellin 20-oxidases and repression of gibberellins 2β-oxidases in unfertilized ovaries of entire tomato mutant, leads to accumulation of active gibberellins and parthenocarpic fruit formation. *Plant Growth Regul.* 75, 415–425. doi: 10.1007/s10725-014-0002-1

Molesini, B., Pandolfini, T., Rotino, G.L., Dani, V., Spena, A. (2009). *Aucsia* gene silencing causes parthenocarpic fruit development in tomato. *Plant Physiol.* 149, 534–548. doi: 10.1104/pp.108.131367

Molesini, B., Dusi, V., Pennisi, F., Pandolfini, T. (2020). How hormones and MADS-Box transcription factors are involved in controlling fruit set and parthenocarpy in tomato. *Genes* 11:1411. doi: 10.3390/genes11121441

Olimpieri, I., Siligato, F., Caccia, R., Mariotti, L., Ceccarelli, N., Soressi, G., et al. (2007). Tomato fruit-set driven by pollination or by the parthenocarpic

fruit allele are mediated by transcriptionally regulated gibberellin biosynthesis. *Planta* 226, 877–888. doi: 10.1007/s00425-007-0533-z

Pandolfini, T., Rotino, G. L., Camerini, S., Defez, R., Spena, A. (2002). Optimisation of transgene action at the post-transcriptional level: high quality parthenocarpic fruits in industrial tomatoes. *BMC Biotechnol.* 2:1. doi: 10.1186/1472-6750-2-1

Pascual, L., Blanca, J. M., Canizares, J., Nuez, F. (2009). Transcriptomic analysis of tomato carpel development reveals alterations in ethylene and gibberellin synthesis during pat3/pat4 parthenocarpic fruit set. *BMC Plant Biol.* 9:67. doi: 10.1186/1471-2229/9/67

Pattison, R. J., Catalá, C. (2012). Evaluating auxin distribution in tomato (Solanum lycopersicum) through an analysis of the PIN and AUX/LAX gene families. *Plant J.* 70:585–598. doi: 10.111/j.1365-313x.2011.04895.x

Philouze, J., and Maisonneuve, B. (1978). Heredity of the natural ability to set parthenocarpic fruits in the Soviet variety Severianin. *Tomato Genet. Coop.* 28, 12–13.

Philouze, J. and Pecaut, P. (1986). Natural parthenocarpy in tomato.III. Study of the parthenocarpy due to the gene *pat* (parthenocarpic fruit) in the line Montfavet 191. *Agronomie* 6, 243–248.

Philouze, J. (1989). Natural parthenocarpy in tomato. IV. A study of the polygenic control of parthenocarpy in line 75/59. *Agronomie* 9, 63–75.

Rotino, G. L., Acciarri, N., Sabatini, E., Mennella, G., Scalzo, R. L., Maestrelli, A., et al. (2005). Open field trial of genetically modified parthenocarpic tomato: seedlessness and fruit quality. *BMC Biotechnol.* 5:32. doi: 10.1186/1472-6750-5-32

Ruiu, F., Picarella, M. E., Imanishi, S., Mazzucato, A. (2015). A transcriptomic approach to identify regulatory genes involved in fruit set of wild-type and parthenocarpic tomato genotypes. *Plant Mol. Biol.* 89, 263–278. doi: 10.1007/s11103-015-0367-1

Serrani, J. C., Fos, M., Atares, A., Garcia-Martinez, J. (2007). Effect of gibberellins and auxin on parthenocarpic fruit growth induction in the cv *micro-tom* of tomato. *J. Plant Growth Regul.* 26, 211–221. doi: 10.1007/s00344-007-9014-7

Serrani, J. C., Ruiz-Rivero, O., Fos, M., Garcia-Martinez, J. (2008). Auxin-induced fruit-set in tomato is mediated in part by gibberellins. *Plant J.* 56, 922–934. doi: 10.1111/j.1365-313X.2008.03654.x

Soressi, G. P. (1975). New spontaneous or chemically-induced fruit ripening tomato mutants. *Rep. Tomato Genet. Coop.* 25, 21–22

Srivastava, A and Handa, A. K. (2005). Hormonal regulation of tomato fruit development: a molecular perspective. *J. Plant Growth Regul.* 24, 67–82. doi: 10.1007/s00344-005-0015-0

Takisawa, R., Maruyama, T., Nakazaki, T., Kataoka, K., Saito, H., Koeda, S., et al. (2017). Parthenocarpy in the tomato (*Solanum lycopersicum* L.) Cultivar 'MPK-1' is controlled by a novel parthenocarpic gene. *Horticulture J.* 86, 487–492. doi: 10.2503/hortj.OKD-042

Takisawa, R., Nakazaki, T., Nunome, T., Fukuoka, H., Kataoka, K., Saito, H., et al. (2018). The parthenocarpic gene Pat-k is generated by a natural mutation of *SlAGL6* affecting fruit development in tomato (*Solanum lycopersicum* L.). *BMC Plant Biol.* 18: 72. doi: 10.1186/s12870-018-1285-6

Takisawa, R., Kusaka, H., Nishino, Y., Miyashita, M., Miyagawa, H., Nakazaki, T., et al. (2019). Involvement of indole-3-acetic acid metabolism in the early fruit development of the parthenocarpic tomato cultivar, MPK-1. *J. Plant Growth Regul.* 38, 189–198. doi: 10.1007/s00344-018-9826-7

Takisawa, R., Maai, E., Nakano, R. and Nakazaki, T. (2020). Effect of parthenocarpic genes *pat-2* and *pat-k* on vegetative and fruit traits in tomato (*Solanum lycopersicum* 'Micro-Tom'). *Horticulture J.* 89, 261–267. doi: 10.2503/hortj.UTD-127

Tang, N., Deng, W., Hu, G., Hu, N., Li, Z. (2015). Transcriptome profiling reveals the regulatorymechanism underlying pollination dependent and parthenocarpic fruit set mainly mediated by auxin and gibberellin. *PLoS One* 10:e0125355. doi: 10.1371/journal.pone.0125355

Vardy, E., Lapushner, D., Genizi, A and Hewitt, J. (1989). Genetics of parthenocarpy in tomato under a low temperature regime: I. Line RP 75/79. *Euphytica* 41, 1–8

Vriezen, W. H., Feron, R., Maretto, F., Keijman, J., Mariani, C. (2008). Changes in tomato ovary transcriptome demonstrate complex hormonal regulation of fruit set. *New Phytol.* 17, 60–76. doi: 10.1111/j.1469-8137.2007.02254.x

Wang, H., Jones, B., Li, Z., Frasse, P., Delalande, C., Regad, F., et al. (2005). The tomato Aux/IAA transcription factor IAA9 is involved in fruit development and leaf morphogenesis. *Plant Cell.* 17, 2676–2692. doi: 10.1105/tpc.105.033415

Wang, L. Y., Shi, Y., Liu, W. M., Yu, H. L. (2008). Genetic effect of parthenocarpy and new lines selection in tomato. *Tianjin Agric. Sci.* 14, 38–40.

Wang, H., Schauer, N., Usadel, B., Frasse, P., Zouine, M., Hernould, M. et al. (2009). Regulatory features underlying pollination-dependent and-independent tomato fruit set revealed by transcript and primary metabolite profiling. *Plant Cell Online* 21, 1428–1452.

Wu, Z., Zhang, T., Li, L., Xu, J., Qin, X., Zhang, T., et al. (2016). Identification of a stable major-effect QTL (*Parth 2.1*) controlling parthenocarpy in cucumber and associated candidate gene analysis via whole genome re-sequencing. *BMC Plant Biol.* 16:182. doi: 10.1186/s12870-016-0873-6

Zhu, Y. B., Kong, Y. Y and Wang, J. H. (2014). Researcg advances in auxin-responsive *SAUR* genes. *Chin. Bul. Life Sci.* 26, 407–413. doi: 10.13376/j.cbls/2014059

Zijlstra S. (1985). Parthenocarpie in tomaat: twee nieuwe lijnen uit soortkruising. *Zaadbelangen* 4, 92–94.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fgene.2021.739989

Genome-Wide Identification and Analysis of the NF-Y Gene Family in Potato (*Solanum tuberosum* L.)

*Zhen Liu[1], Yuanming Li[2], Jinyong Zhu[3], Wenjing Ma[3], Zhitao Li[3], Zhenzhen Bi[3], Chao Sun[3], Jiangping Bai[3], Junlian Zhang[2] and Yuhui Liu[1]**

[1] State Key Laboratory of Aridland Crop Science, Gansu Agricultural University, Lanzhou, China, [2] College of Horticulture, Gansu Agricultural University, Lanzhou, China, [3] College of Agronomy, Gansu Agricultural University, Lanzhou, China

Nuclear factor Y (NF-Y) is a ubiquitous transcription factor in eukaryotes, which is composed of three subunits (NF-YA, NF-YB, and NF-YC). NF-Y has been identified as a key regulator of multiple pathways in plants. Although the NF-Y gene family has been identified in many plants, it has not been reported in potato (*Solanum tuberosum*). In the present study, a total of 41 NF-Y proteins in potato (StNF-Ys) were identified, including 10 StNF-YA, 22 StNF-YB, and nine StNF-YC subunits, and their distribution on chromosomes, gene structure, and conserved motif was analyzed. A synteny analysis indicated that 14 and 38 pairs of StNF-Y genes were orthologous to *Arabidopsis* and tomato (*Solanum lycopersicum*), respectively, and these gene pairs evolved under strong purifying selection. In addition, we analyzed the expression profiles of NF-Y genes in different tissues of double haploid (DM) potato, as well as under abiotic stresses and hormone treatments by RNA-seq downloaded from the Potato Genome Sequencing Consortium (PGSC) database. Furthermore, we performed RNA-seq on white, red, and purple tuber skin and flesh of three potato cultivars at the tuber maturation stage to identify genes that might be involved in anthocyanin biosynthesis. These results provide valuable information for improved understanding of StNF-Y gene family and further functional analysis of StNF-Y genes in fruit development, abiotic stress tolerance, and anthocyanin biosynthesis in potato.

Keywords: potato (*Solanum tuberosum*), nuclear factor Y transcription factor, expression profiles, abiotic stress, anthocyanin biosynthesis

OPEN ACCESS

Edited by:
Rocio Deanna, University of Colorado Boulder, United States

Reviewed by:
Jingwei Yu, Southern University of Science and Technology, China
Tao Xu, Jiangsu Normal University, China
Fuyou Fu, Agriculture and Agri-Food Canada (AAFC), Canada

****Correspondence:***
Yuhui Liu
yuhuiliu2021@163.com

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

Received: *12 July 2021*
Accepted: *20 August 2021*
Published: *17 September 2021*

Citation:
Liu Z, Li Y, Zhu J, Ma W, Li Z, Bi Z, Sun C, Bai J, Zhang J and Liu Y (2021) Genome-Wide Identification and Analysis of the NF-Y Gene Family in Potato (Solanum tuberosum L.). Front. Genet. 12:739989. doi: 10.3389/fgene.2021.739989

INTRODUCTION

Nuclear factor Y (NF-Y) transcription factor, also known as CCAAT-box binding factor (CBF) or heme-associated protein (HAP) (Mantovani, 1999), is a ubiquitous transcription factor in eukaryotes (Zhao et al., 2017). NF-Y is a trimeric transcription factor composed of three distinct subunits, namely, NF-YA, NF-YB, and NF-YC (Nardini et al., 2013). To form the NF-Y complex, the histone folding motifs of NF-YB and NF-YC interact in the cytoplasm to form heterodimer,

Abbreviations: StNF-Y, nuclear factor Y transcription factor in potato; DM, double monoploid potato; HMM, hidden Markov model; Ka, non-synonymous substitution rate; Ks, synonymous substitution rate; PG, PGSC0003DMG40; qPCR, quantitative real-time PCR; CC, correlation coefficient.

which is transferred to the nucleus. Upon arrival in the nucleus, the NF-YA subunit is recruited and combines with NF-YB/NF-YC heterodimer to produce mature NF-Y complex (Kahle et al., 2005; Laloum et al., 2013). In mature NF-Y complex, the NF-YA subunit provides a distinctive sequence that specifically binds to the CCAAT *cis*-element and regulates the expression of target genes (Hackenberg et al., 2012; Yan et al., 2013; Cao et al., 2014; Xu et al., 2016). In yeast and mammals, each NF-Y subunit is encoded by a single gene; however, three NF-Y subunits are encoded by multiple genes in plants (Li et al., 1992; Petroni et al., 2012).

In recent years, many NF-Ys were identified in *Arabidopsis* (Riechmann et al., 2000), rice (*Oryza sativa*) (Yang W. et al., 2017), tomato (*Solanum lycopersicum*) (Li et al., 2016), banana (*Musa acuminata*) (Yan et al., 2019), sorghum (*Sorghum bicolor*) (Malviya et al., 2016), watermelon (*Citrullus lanatus*) (Yang J. et al., 2017), and canola (*Brassica napus*) (Liang et al., 2014). And, the functions of many NF-Y genes have been characterized to play multiple roles in plants, including embryogenesis (Kwong et al., 2003), seed germination (Liu et al., 2016), root growth (Sorin et al., 2014), flowering (Cao et al., 2014), fruit ripening (Li et al., 2016), as well as flavonoid biosynthesis (Wang et al., 2020) and response to various abiotic stresses (Li et al., 2008; Ni et al., 2013; Sato et al., 2014; Wang et al., 2018). For example, *AtLEC1* (*AtNF-YB9*) is a key regulator of late embryogenesis and seed development in *Arabidopsis*, and the ectopic expression of *AtLEC1* caused abnormality in tobacco transgenic seedling (Guo et al., 2013). The overexpression of *AtNF-YA1* and *AtNF-YA9* significantly affects male gametophyte development, embryo development, seed morphology, and seed germination (Mu et al., 2013). The overexpression of *AtNF-YA5*, which is strongly induced under drought, increases drought tolerance of *Arabidopsis* plants (Li et al., 2008). The constitutive expression of *TaNF-YA10* significantly increases the sensitivity of *Arabidopsis* plants to salinity (Ma et al., 2015). A recent report suggested that NF-Y complexes composed of NF-YB8, NF-YC1/9, and NF-YA1/9 play an important role in tomato flavonoid biosynthesis, and these complexes are involved in transcriptional regulation and H3K27me3 marking of the *chalcone synthase* (*CHS1*) locus (Wang et al., 2020). Although the NF-Y gene family has been extensively studied in plants, little report is yet available for potato.

Potato, which originated from the Andean regions of Peru and Bolivia (Spooner et al., 2005), is an important food crop worldwide. As a new kind of natural pigment and antioxidant resource, the pigmented potato has attracted more and more attention (Fossen and Andersen, 2000). Flavonoids not only attract insects for pollination but also protect plants against UV-induced damage (He and Giusti, 2010), as well as play a key role in cold and drought stress tolerance (Castellarin et al., 2007; Kim et al., 2017). Furthermore, many flavonoids have strong antioxidant activity and free radical-scavenging ability, which are beneficial to human health, such as preventing cardiovascular disease, controlling obesity, alleviating diabetes, anti-cancer, and so on (Hollman and Katan, 1997; Castañeda-Ovando et al., 2009). However, abiotic stresses can severely restrict the growth and productivity of potato, such as high-temperature stress, cold stress, salinity stress, and drought stress. Given the important role that NF-Y genes play in plants, it is essential to identify and study the NF-Y gene family in the potato genome.

In the present study, we identified a NF-Y gene family with 41 members in potato (*StNFYs*), and comprehensive analyses of the phylogenetic relationships, chromosome distribution, gene structure, sequence features, and gene duplications were further performed. In addition, we analyzed the expression profiles of NF-Y genes in different tissues of double haploid (DM) potato, as well as under abiotic stresses and hormone treatments by RNA-seq downloaded from the Potato Genome Sequencing Consortium (PGSC) database. Furthermore, we performed RNA-seq on white, red, and purple tuber skin and flesh of three potato cultivars at the tuber maturation stage to identify genes that might be involved in anthocyanin biosynthesis. The results could provide a basis for further study on the functional characterization of the NF-Y gene family.

MATERIALS AND METHODS

Identification of StNF-Ys in Potato

The amino acid and nucleotide sequences of potato were downloaded from the PGSC[1]. To identify the potato NF-Y members, two methods were used individually and then combined: (1) the 36 reported *Arabidopsis* NF-Y (AtNF-Y) protein sequences (Riechmann et al., 2000) were downloaded from the *Arabidopsis* information resource (TAIR;[2]). The BLASTP (Altschul et al., 1997) was used to search for StNF-Y members based on 36 AtNF-Y amino acid sequences with an *E*-value $\leq$ 1e-5. (2) The hidden Markov model (HMM) profiles of the NF-Y domain (PF00808 and PF02045) were downloaded from Pfam[3]. The StNF-Y proteins were identified by HMMER3.1 software[4]. After removing all redundant sequences, these candidate members were submitted to SMART[5] and NCBI Conserved Domain Data (CDD) to manually screen StNF-Y members for further analysis.

Sequence Analysis and Structural Characterization

The number of amino acids, theoretical isoelectric point (pl), and molecular weight (MW) were calculated using the ExPasy site[6] (Gasteiger et al., 2005). The MEME program[7] was used to identify motifs in StNF-Y sequence with the following parameters: 20 motifs with an optimal motif width of 6–50 amino acid residues and any number of repeats (Bailey et al., 2009). The gene structures of the StNF-Ys were drawn using the Gene Structure Display Server (GSDS 2.0, [8]) (Guo et al., 2007).

[1]http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml
[2]http://www.Arabidopsis.org
[3]http://pfam.xfam.org
[4]http://hmmer.org/download.html
[5]http://smart.embl-heidelberg.de
[6]http://web.expasy.org/protparam
[7]http://alternate.meme-suite.org/tools/meme
[8]http://gsds.cbi.pku.edu.cn

Chromosomal Localization and Gene Duplication

All *StNF-Y* genes were mapped to 12 potato chromosomes based on physical location information from PGSC using MapChart software (Voorrips, 2002). MCScanX (Wang et al., 2012) was used to analyze the duplication events of the *StNF-Y* genes and identified the synteny of *NF-Y* genes between potato and *Arabidopsis*. Ka (non-synonymous) and Ks (synonymous) of each pair of duplicated *NF-Y* genes were calculated using KaKs Calculator 2.0 (Wang et al., 2010).

Multiple Alignments, Phylogenetic Analysis, and Classification of NF-Ys

The multiple alignments were performed using ClustalW on the 41 StNF-Ys amino acid under default parameters. The full-length amino acid sequences of 41 StNF-Ys, 54 SlNF-Ys (tomato NF-Y), and 36 AtNF-Ys were used for phylogenetic analysis. The Molecular Evolutionary Genetics Analysis (MEGA) 7.0 software (Kumar et al., 2016) was used to construct an unrooted phylogenetic tree with the following parameters: Poisson model and 1,000 bootstrap replications.

Plant Materials and Treatments

Three potato cultivars, including "Xindaping" (XD—white skin and white flesh), "Lingtianhongmei" (LT—red skin and red flesh), and "Heimeiren" (HM—purple skin and purple flesh), were grown in a greenhouse at Gansu Agricultural University in Lanzhou, Gansu Province, China (**Figure 1**). Six fresh tubers (diameter = 4–5 cm) were collected from each cultivar for harvesting the skin tissue and flesh tissue. Skin tissues were carefully obtained from the cortical tissue using a scalpel. Flesh tissues were isolated at least 5-mm distance from the skin. The samples were immediately frozen in liquid nitrogen and stored in −80°C refrigerator for further use.

RNA-Seq Data Analysis

Total RNA of the aforementioned samples was chosen for further RNA-seq library construction. Each sample had three biological replications. Next-generation Illumina sequencing was performed by Sagene Biotech Corporation (Guangzhou, China). The raw RNA-seq data were uploaded to the NCBI (Project ID PRJNA541919). After RNA sequencing, the clean reads were obtained by trimming the raw reads and filtering out contaminants, adapters, Phred scores less than 20, and uncertain bases. The cleaned data were aligned to PGSC_DM_v3.4 gene models downloaded from Solanaceae Genomics Resource at Michigan State University[1] by Bowtie2 (v2.2.9). Only reads with a perfect match or one mismatch were further analyzed.

Differential Expression Genes Analysis

Differential genes were identified using fragments per kilobase of exon per million fragments mapped (FPKM). Genes with the absolute value of $\log_2$ fold change (FC) ≥ 1 and a false discovery rate (FDR) < 0.05 were considered significant differentially expressed genes (DEGs).

The DEGs were annotated against non-redundant database (Nr), SwissProt/UniProt Plant Proteins, Kyoto Encyclopedia of Genes and Genomes (KEGG), Cluster of Orthologous Groups of proteins (COG/KOG), and the potato protein database[9] by BLASTX, with a cut-off *E*-value of 1e-5. Then, DEGs were subjected to enrichment analysis of GO functions and KEGG pathways.

RNA Isolation and Quantitative Real-Time PCR

Total RNA extractions were isolated with the RNA extraction kit (Tiangen DP419, Beijing, China). The integrity of RNA was monitored by agarose gel electrophoresis, and the concentration was detected using a NanoDrop ND-2000 spectrophotometer (NanoDrop Technologies, United States). The cDNAs were synthesized using a FastKing RT kit with gDNase (Tiangen KR116, Beijing, China). Quantitative real-time PCR (qPCR) was performed with SuperReal PreMix Plus kit (SYBR Green FP205, Tiangen, Beijing, China) in Bio-Rad CFX96 (Bio-Rad, Hercules, CA, United States). The PCR program was as follows: 30 s at 95°C, followed by 40 cycles of 5 s at 95°C and 30 s at 60°C, followed by 65–95°C melting curve detection. Three independent

[9]ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Stuberosum

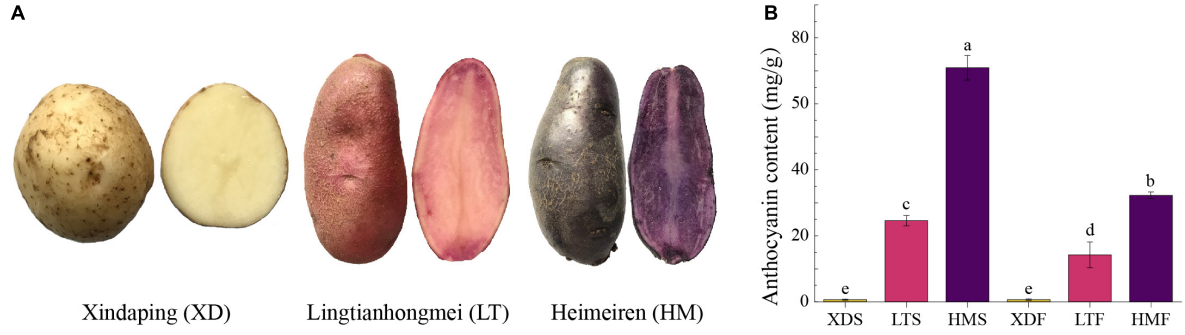

FIGURE 1 | Phenotypes of three potato cultivars in **(A)** "Xindaping" (XD—white skin and white flesh), "Lingtianhongmei" (LT—red skin and red flesh), and "Heimeiren" (HM—purple skin and purple flesh). **(B)** The content of anthocyanin in three potato cultivars. Data represent the mean of three biological replicates ± standard error. Standard errors are shown as bars above the columns. Different letters above bars denote significant difference at $P < 0.05$.

biological replications were performed. We used the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) to determine the relative expression levels of genes. The *StEF-1α* (AB061263) was taken as the internal control (Tang et al., 2017). The primers are listed in **Supplementary Table 1**.

Expression Pattern Analysis of StNF-Ys in Potato

The expressions of *StNF-Y* genes in 13 tissues (leaves, stamens, shoots, stolons, roots, tubers, carpels, petals, petioles, sepals, flowers, immature fruit, and mature fruit) in DM potato and the whole plant *in vitro* that was treated for abiotic stress (salt treatment: 150 mM NaCl, 24 h; mannitol-induced drought stress: 260 μM mannitol, 24 h; heat treatment: 35°C, 24 h) and hormone treatments [benzylaminopurine (BAP): 10 μM, 24 h; abscisic acid (ABA): 50 μM, 24 h; indole acetic acid (IAA): 10 μM, 24 h; gibberellic acid (GA3): 50 μM, 24 h] were analyzed based on the Illumina RNA-seq data that was downloaded from the PGSC (The Potato Genome Sequencing Consortium et al., 2011). TBtools software was used to draw the heat map (Chen et al., 2018).

The expression patterns of *StNF-Y* genes in three pigmented potato cultivars (XD, LT, and HM) were analyzed based on RNA-seq. The Illumina sequencing was performed by Sagene Biotech Corporation (Guangzhou, China). The raw data were uploaded on NCBI (Project ID PRJNA541919).

The Interaction Network of NF-Y Proteins

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a database of known and predicted protein–protein interactions. Based on the RNA-seq dataset of pigmented potato cultivars, we selected StNF-Ys with FPKM > 5 to predict the protein interaction network. The protein interactions were constructed using online STRING v11.0 software[10] with a combined score > 400 (medium confidence) (Szklarczyk et al., 2016). As active interaction sources, text mining, experiments, databases, co-expression, neighborhood, gene fusion, and co-occurrence were selected. The interaction network was visualized by Cytoscape v3.7.1 (Shannon et al., 2003).

Statistical Analysis

For qPCR analyses, data were presented as means (±SE) of three biological replicates. Statistical significance was determined by one-way ANOVA followed by the least significant difference (LSD) computed at $P < 0.05$.

RESULTS

Identification and Chromosomal Distribution of StNF-Ys

To obtain the NF-Y family members in potato, the HMM and BLASTP algorithm search were combined to analyze the potato genome. A total of 41 *StNF-Y* genes were identified and confirmed in the potato genome, including 10 *StNF-YA*, 22 *StNF-YB*, and 9 *StNF-YC* (**Supplementary Table 2**). The transcript IDs were converted to corresponding gene IDs and the "SC0003DMG40" in the gene ID was omitted for brevity. The bioinformatics data of StNF-Y members were analyzed, including the number of amino acid residues, theoretical molecular weight (MM), and theoretical isoelectric point (pI). The StNF-Y proteins varied in length and physicochemical properties, the number of amino acid residues of StNF-Y proteins ranged from 120 (PG0013302) to 311 (PG0021365), the molecular weights were between 13.82 kDa (PG0013302) and 33.97 kDa (PG0021365), and the pI values ranged from 4.64 (PG0023065) to 9.65 (PG0002484). The details are listed in **Supplementary Table 2**.

The identified 41 *StNF-Y* genes were distributed unevenly on 12 chromosomes (**Figure 2A**). Seven *StNF-Y* genes were located on chromosomes 1 and 5, which had the largest number of *StNF-Y* genes. In contrast, only one *STNF-Y* was distributed on chromosome 8, which had the least number of *StNF-Y* genes (**Figure 2B**). In addition, the high densities of *StNF-Y* genes were distributed at the proximal end of chromosome 1 and the distal end of chromosome 5.

Multiple Alignments, Phylogenetic Analysis, and Classification of StNF-Ys

Multiple sequence alignments showed that StNF-Y family proteins had conserved region domains as shown in **Figure 3**. The results showed that each member of StNF-Y contains a heterodimerization domain and a DNA-binding domain that recognizes the CCAAT site. The StNF-YA subunit contained two conserved domains (**Figure 3A**); one is the NF-YB and NF-YC interaction domain and the other is the DNA-binding domain. In this DNA-binding domain, three histidine (H) and three arginine (R) residues are absolutely conserved and essential (Xing et al., 1993). The conserved domain of StNF-YB subunit was about 93 AAs in length, which contained a NF-YA interaction region, a NF-YC interaction region, and a DNA-binding domain (**Figure 3B**). Previous studies have shown that the NF-YB subunits could be divided into LECl type and non-LECl type according to sequence similarity; the LECl type is composed of LEC1 and LEC1-like (L1L) (Lee et al., 2003; Siefers et al., 2009).

As shown in **Figure 3B**, the key amino acids that distinguish Lec1 from non-LEC1 were represented in the red box, and aspartate (D) at this site was considered to be specific for LEC1 type. The conserved domains of StNF-YC subunits contained three heterodimerization regions and one DNA-binding domain, which was about 79 AAs in length (**Figure 3C**).

To further investigate the phylogenetic relationships of NF-Ys among potato, tomato, and *Arabidopsis*, an unrooted phylogenetic tree of 41 StNF-Ys, 54 SlNF-Y, and 36 AtNF-Y protein sequences was constructed by MEGA 7.0 software, as shown in **Figure 4**. The 131 NF-Ys were divided into three distinct subunits (NF-YA, NF-YB, and NF-YC). Ten StNF-Ys, 10 SlNF-Ys, and 10 AtNF-Ys belonged to the NF-YA subunit. Twenty-two StNF-Ys, 27 SlNF-Ys, and 13 AtNF-Ys were assigned to the NF-YB subunit. Nine StNF-Ys, 17

[10]http://string-db.org/

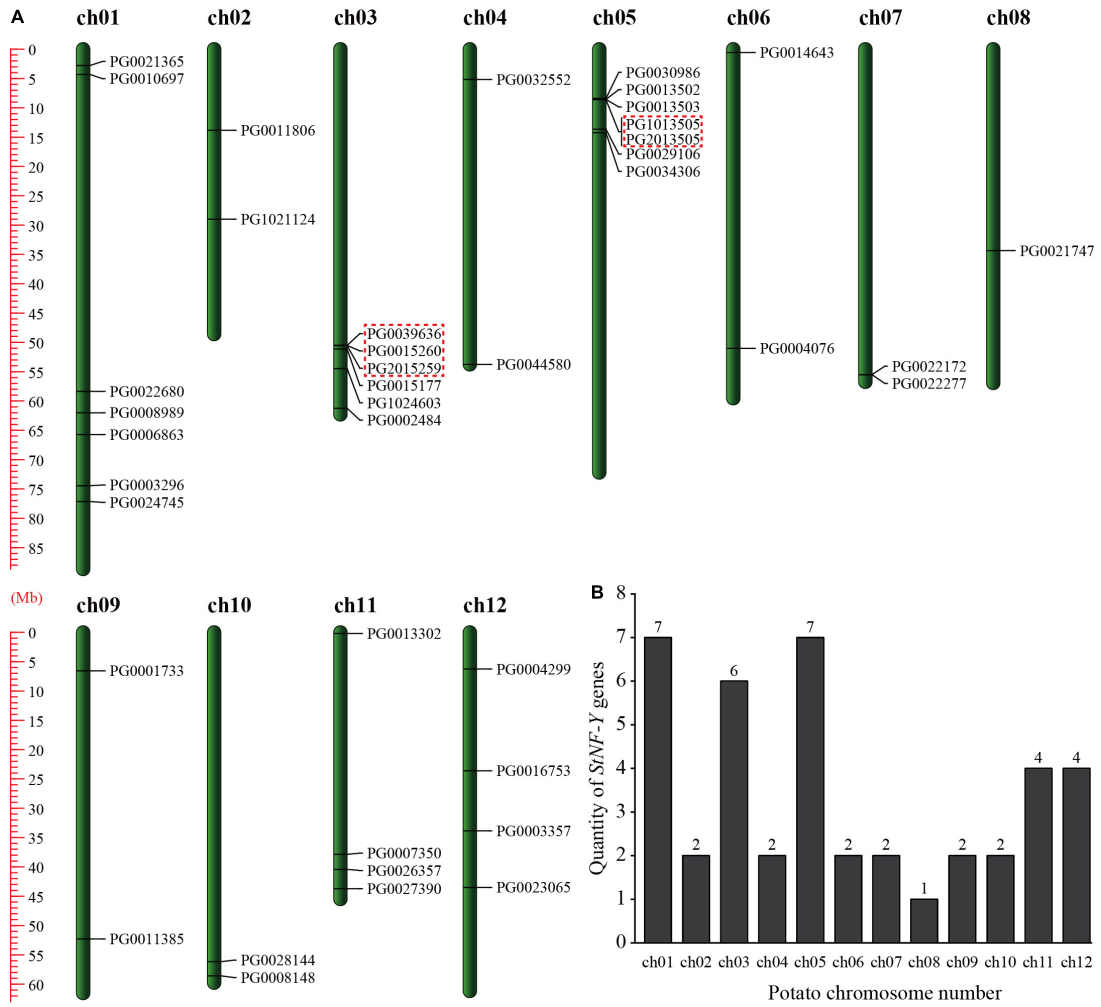

FIGURE 2 | Genomic distributions of 41 *StNF-Y* genes on 12 potato chromosomes. **(A)** "*StNF-Y* gene distribution map" on 12 potato chromosomes. Tandemly duplicated genes were marked by red boxes. **(B)** Numbers of *StNF-Y* genes on each chromosome in potato.

SlNF-Ys, and 13 AtNF-Ys belonged to the NF-YC subunit. Notably, a distinct cluster was identified as LEC1-type group in NF-YB, including AtNF-YB6 (LEC1-like), AtNF-YB9 (LEC1), 13 SlNF-Ys, and eight StNF-YBs (PG1013505, PG2013505, PG0013503, PG0013302, PF0013502, PG0030986, PG0022277, and PG0034306).

Gene Structure and Conserved Motifs of StNF-Ys

The full-length amino acid sequences of 41 StNF-Ys were used to construct an unrooted phylogenetic tree by MEGA 7.0 software. Based on phylogenetic analysis of potato, tomato, and Arabidopsis, 41 StNF-Ys were divided into three subunits (**Figure 5A**). A gene structure analysis can provide insight into the evolution of gene family. Therefore, we analyzed the organization of exons and introns (**Figure 5B**). The results showed that all of *StNF-YA* genes were separated by introns, of which eight *StNF-YAs* contained four introns, one *StNF-YA* contained three introns, and one *StNF-YA* contained two introns. Most of *StNF-YBs* and *StNF-YCs* did not have introns, including14 *StNF-YBs* and six *StNF-YCs*; the rest of *StNF-YBs* and *StNF-YCs* contained 1–4 introns. In general, the gene structure of *StNF-YA* members was more conserved than that of *StNF-YBs* and *StNF-YCs*.

In order to study the diversification of StNF-Ys, we searched for conserved motifs of these proteins using the online MEME program. The details of the 20 motifs are in **Supplementary Table 3**. The results showed that the three StNF-Y subunits have a unique motif distribution (**Figure 5C**). Motifs 12, 17, and 18 were specifically present in the StNF-YA members. Motifs 1, 8, 10, 11, 16, and 19 were unique to the StNF-YB members. Motifs 4, 9, 13, 15, and 20 were present only in StNF-YC members. Motif 2 was distributed in StNF-YA and StNF-YB subunits. Motif 3 was present in StNF-YB and StNF-YC subunits. Motif 6 was widely distributed in three StNF-Y subunits. Overall, each subunit had a conserved motif composition.

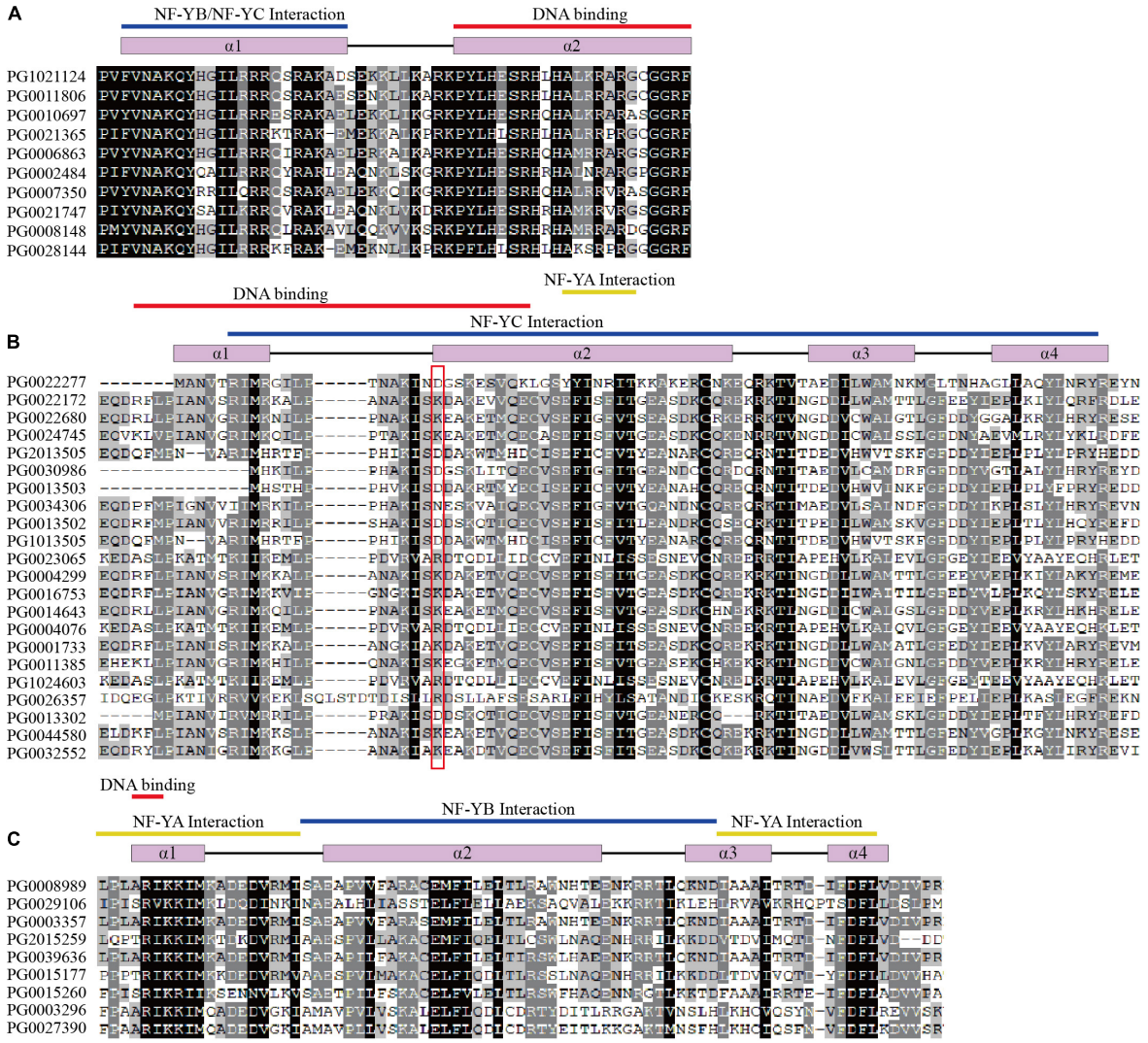

FIGURE 3 | Multiple sequence alignments of StNF-Y family members. **(A)** Multiple alignments of the StNF-YA conserved domains. **(B)** Multiple alignments of the StNF-YB conserved domains. **(C)** Multiple alignments of the StNF-YC conserved domains. The secondary structures of StNF-Ys are indicated at the top of each alignment. The alpha-helices are represented in purple rectangles, and the coils are represented in black lines on the top of the alignment. DNA-binding, yellow lines indicate NF-YA interaction regions, and blue lines indicate NF-YB/YC interaction domains, respectively. The key amino acids that distinguish LEC1 from non-LEC1 are represented in the red box.

Gene Duplication and Genome Synteny

Tandem and segmental duplications are crucial for the extension of gene family members and the realization of new functions (Cannon et al., 2004). In the present study, three pairs of *StNF-Y* genes (5/41, 12.20%) were identified as tandemly duplicated genes, of which chromosome 3 had two pairs and chromosome 5 had one pair (**Figure 2A** and **Supplementary Table 4**). In addition to tandem duplication, six pairs (10/41, 24.39%) of segmental duplication events were identified (**Figure 6** and **Supplementary Table 4**).

In order to further investigate the potential evolution processes of the NF-Y gene family, we analyzed the synteny relationship of *NF-Ys* among potato, *Arabidopsis*, and tomato. A total of 14 pairs of orthologs were identified between potato and *Arabidopsis*, and 38 pairs of orthologs between potato and tomato were identified (**Figure 6** and **Supplementary Tables 5, 6**).

The substitution rate of non-synonymous (Ka) and synonymous (Ks) is the basis for evaluating the positive selection pressure of duplication events. Ka/Ks = 1 indicates neutral selection, Ka/Ks < 1 denotes purification selection, and Ka/Ks > 1 signifies positive selection. The Ka/Ks of duplication *NF-Y* genes was calculated using KaKs Calculator 2.0. The results showed that the Ka/Ks of tandem and segmental duplications ranged from 0.0787 to 0.9482, with a mean value of 0.3569 (**Supplementary Table 4**). The Ka/Ks of the orthologous relationship between potato and *Arabidopsis* ranged from 0.0305 to 0.3507, with a mean value of 0.1736 (**Supplementary Table 5**). And, the Ka/Ks of the orthologous relationship between potato

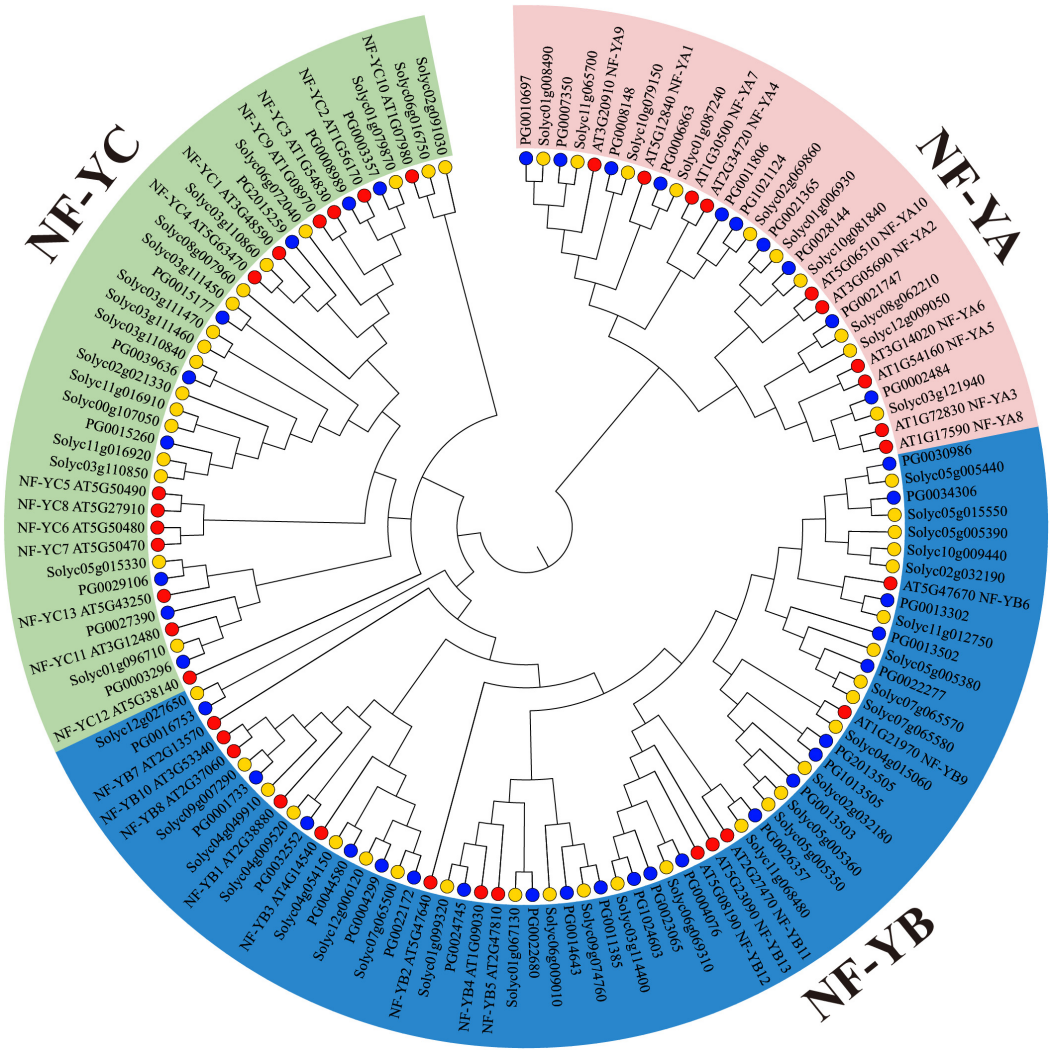

FIGURE 4 | Phylogenetic classification of NF-Ys from potato, tomato, and *Arabidopsis*. The three NF-Y subunits were marked with different colors. The blue circles represent StNF-Ys, the yellow circles represent SlNF-Ys, and the red circles represent AtNF-Ys.

and tomato ranged from 0.001 to 0.9260, with a mean value of 0.2863 (**Supplementary Table 6**). The Ka/Ks values were < 1, suggesting that these genes had evolved under the effect of purifying selection.

Expression Profiles of StNF-Y Genes in Different Tissues

Based on transcriptome data downloaded from PGSC, the expression patterns of *StNF-Y* genes in 13 tissues (leaves, roots, shoots, tubers, sepals, stamens, stolons, mature flowers, petioles, petals, carpels, immature fruit, and mature fruit) of DM potato were analyzed (**Figure 7** and **Supplementary Table 7**). The results showed that 14 *StNF-Y* genes (*PG0001733*, *PG0003296*, *PG0003357*, *PG0004076*, *PG0004299*, *PG0006863*, *PG0007350*, *PG0008989*, *PG0010697*, *PG2015259*, *PG1021124*, *PG1024603*, *PG0026357*, and *PG0027390*) were highly expressed in all tissues with FPKM > 5. Four *StNF-Y* genes (*PG0011385*, *PG1013505*, *PG0039636*, and *PG0024745*) were not expressed in all tissues (FPKM = 0). In addition, some StNF-Y genes showed tissue-specific expression patterns, e.g., *PG0021365* was highly expressed in stolon and *PG0044580* was highly expressed in the tuber. It is noteworthy that six *StNF-YB* genes (*PG0013502*, *PG0013503*, *PG0015177*, *PG0022277*, *PG0030986*, and *PG0034306*) were specifically expressed in immature fruit with FPKM > 5 and $\log_2$FC > 1, which were in the same cluster as *AtNF-YB9* (*LEC1*) and *AtNF-YB6* (*LEC1-like*).

Expression Profiles of StNF-Y Genes Under Abiotic Stresses and Hormone Treatments

The RNA-seq data was downloaded from the PGSC database to analyze the expression profiles of *StNF-Y* genes in DM potato under abiotic stresses. The results showed that six, nine, and 11 *StNF-Y* genes were differentially expressed

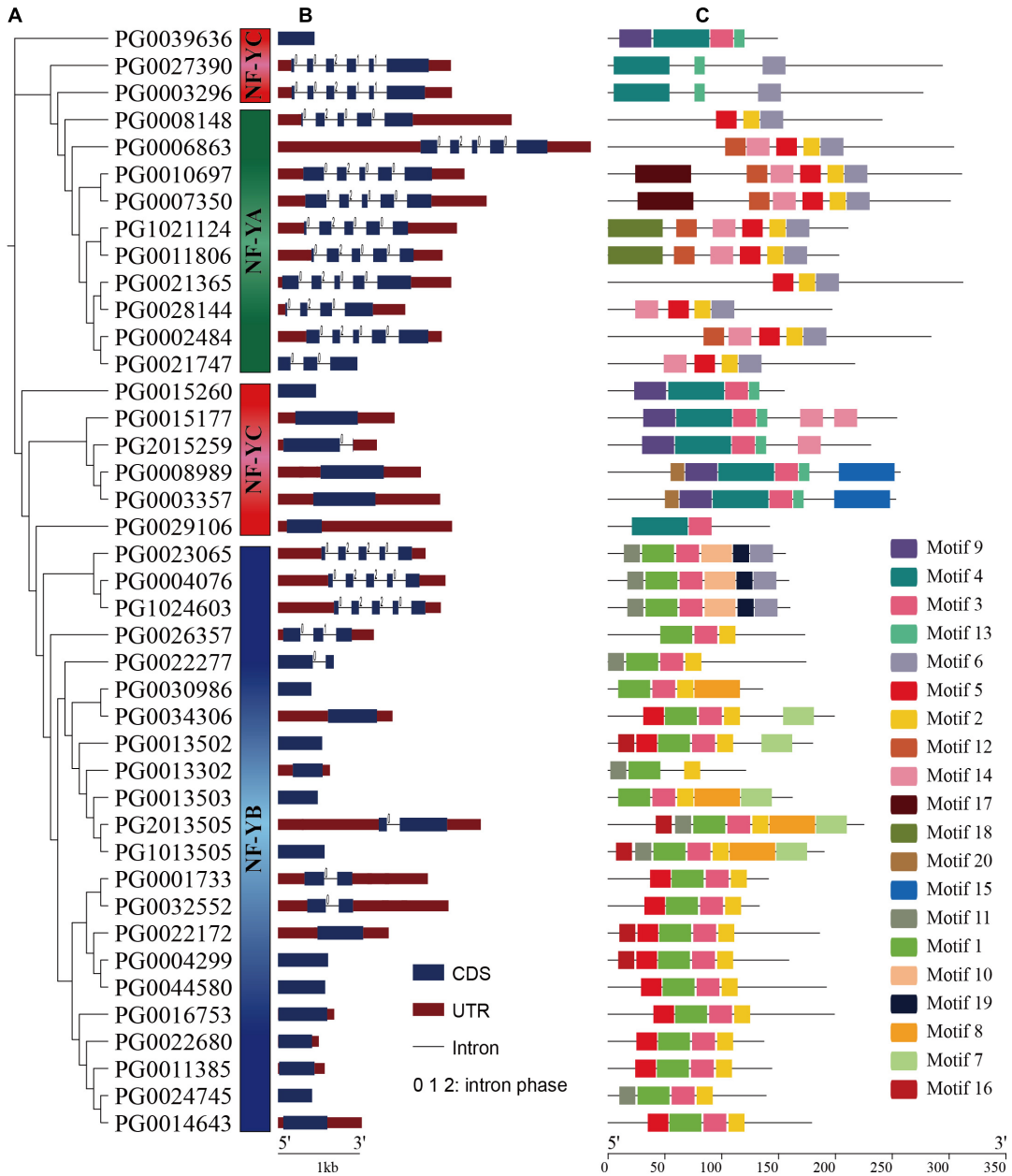

FIGURE 5 | Phylogenetic relationships, gene structure, and conserved motifs analysis of StNF-Ys. **(A)** Phylogenetic tree of StNF-Ys. **(B)** Exon/intron organization of the *StNF-Y* genes. The blue box indicates the exon, and the black line of the same length indicates the intron. The upstream/downstream area is indicated by a red box. The numbers 0, 1, and 2 represent the splicing phase of the intron. **(C)** Distribution of conserved motifs in StNF-Ys. The putative motifs were predicted by MEME program. The 20 boxes of different colors represent 20 different putative motifs. The details of the 20 putative motifs are in **Supplementary Table 3**.

(FPKM > 1 and $|\log_2 FC| > 1$) under salt, mannitol, and heat treatments, respectively. Among them, four, seven, and nine *StNF-Y* genes were upregulated under salt, mannitol, and heat treatments, respectively. In addition, two *StNF-Y* genes (*PG2015259* and *PG0021365*) were differentially expressed under three stresses, five StNF-Y genes (*PG0002484*, *PG0004299*, *PG0022172*, *PG0028144*, and *PG0044580*) were differentially expressed under two stresses, and 10 genes (*PG0004076*, *PG0008148*, *PG0008989*, *PG0011385*, *PG0014643*, *PG1021124*, *PG0021747*, *PG0022680*, *PG0023065*, and *PG0032552*) were only differentially expressed under single abiotic stress (**Figure 8A** and **Supplementary Table 8**).

To further explore the expression changes in the *StNF-Y* genes under various hormone treatments (BAP, ABA, IAA, and GA3), the expression of 41 *StNF-Y* genes was analyzed using the RNA-seq data downloaded from the PGSC database. The results indicated that 22 and 10 *StNF-Y* genes responded to BAP and ABA treatments (FPKM > 1 and $|\log_2 FC| > 1$),

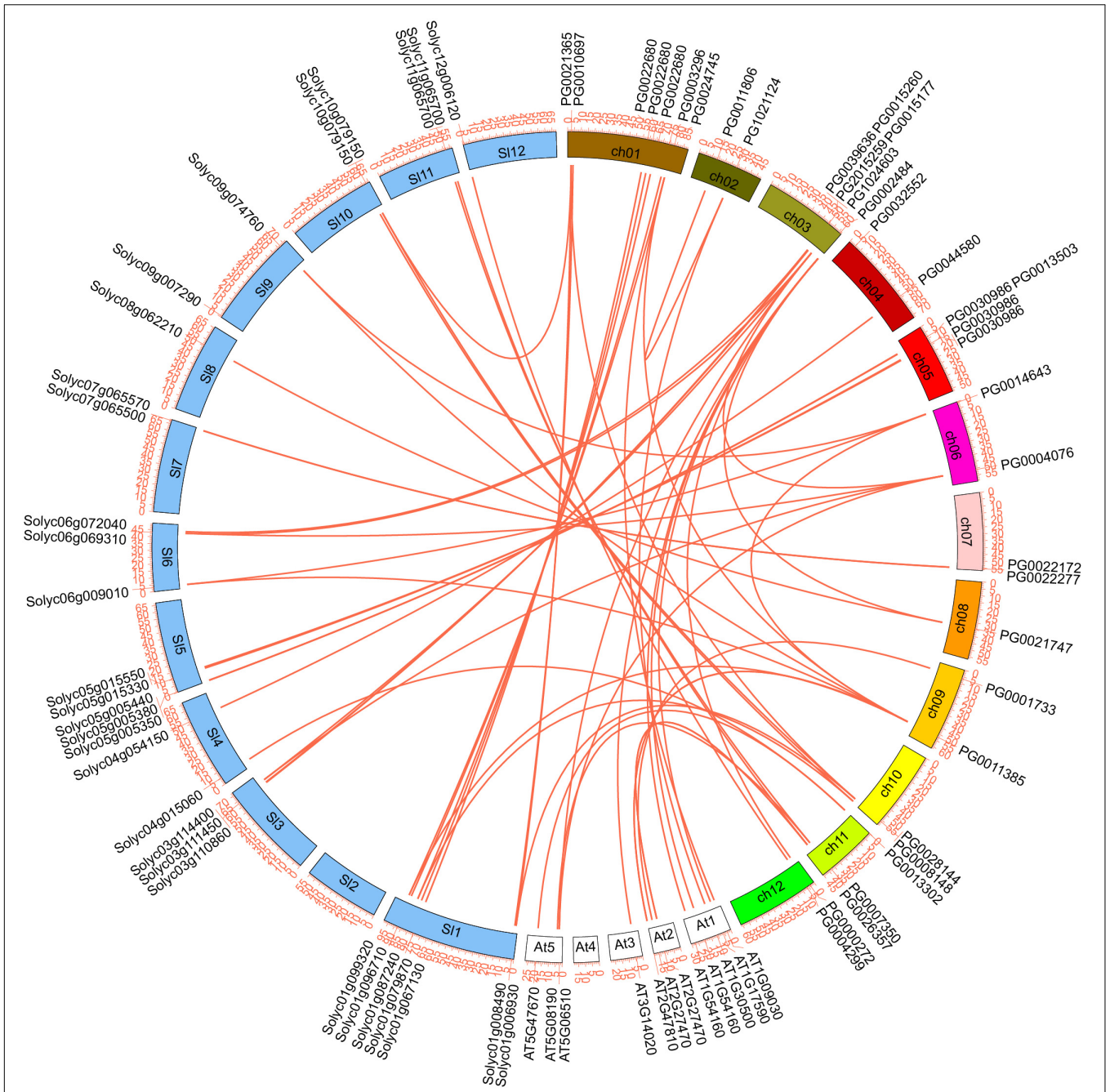

FIGURE 6 | The segmental replication events of *StNF-Y* genes in potato and the orthologous relationships of *StNF-Y* genes with tomato and *Arabidopsis*. Red lines indicate segmental duplication of *StNF-Y* genes and the orthologous relationships of *StNF-Y* genes with tomato and *Arabidopsis*. The chromosome number is indicated at the middle of each chromosome. The white boxes represent *Arabidopsis* chromosome, the blue boxes represent tomato chromosome, and the colored boxes represent potato chromosome, respectively.

respectively. All of 22 *StNF-Y* genes were downregulated under BAP treatments, and all of 10 *StNF-Y* genes were upregulated under ABA treatments (FPKM > 1 and $|\log_2 FC| > 1$). However, all of *StNF-Y* genes were not differentially expressed under IAA and GA3 treatments, implying that *StNF-Y* genes did not respond to IAA and GA3 treatments (**Figure 8A** and **Supplementary Table 8**).

Combined with RNA-seq data analysis under abiotic stress and hormone treatment, four *StNF-Y* genes (*PG0002484*, *PG0008989*, *PG0023065*, and *PG0028144*) were differentially expressed (FPKM > 1 and $|\log_2 FC| > 1$) under mannitol and ABA treatments in DM potato. The results indicate that these four *StNF-Y* genes may be involved in drought stress response through an ABA-dependent pathway.

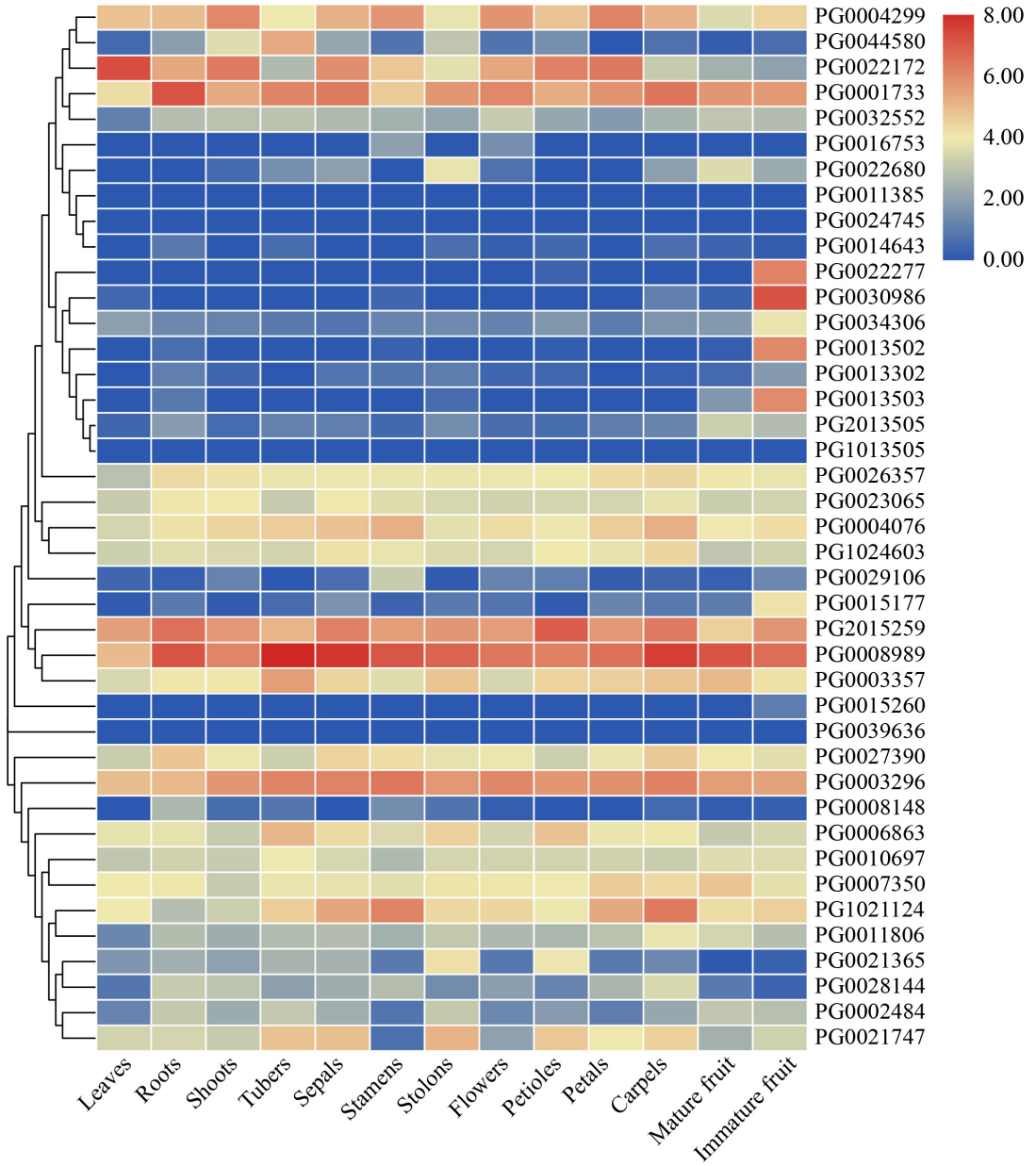

FIGURE 7 | An expression profile analysis of *StNF-Y* genes in different tissues (leaves, roots, shoots, tubers, sepals, stamens, stolons, mature flowers, petioles, petals, carpels, immature fruit, and mature fruit) in double haploid (DM) potato based on the transcriptome data. The color in the heat map represents the fragments per kilobase of exon per million fragments mapped (FPKM) value using logarithm with a base of 2.

Expression Profiles of StNF-Y Genes in Pigmented Potato Cultivars

To identify the potential functions of *StNF-Y* genes in flavonoid biosynthesis of potato, the expression patterns of the *StNF-Y* genes in tuber tissues (skin and flesh) of tetraploid pigmented potato cultivars were analyzed according to RNA-seq dataset. The results showed that eight *StNF-Y* genes (*PG0011385*, *PG0013502*, *PG0013503*, *PG0013302*, *PG0015260*, *PG0024745*, *PG0039636*, and *PG0022277*) were not expressed (FPKM = 0) in the skin, and 12 *StNF-Y* genes were differentially expressed in pigmented skin compared with the white skin of XD (FPKM $>$ 1 and $|\log_2 FC| > 1$). Among them, six genes (*PG0002484*, *PG0008148*, *PG0008989*, *PG0010697*, *PG0028144*, and *PG0044580*) were upregulated and four genes (*PG0022680*, *PG0034306*, *PG1013505*, and *PG2013505*) were downregulated in LTS and one gene (*PG0002484*) was upregulated and five genes (*PG0003357*, *PG0026357*, *PG0034306*, *PG1013505*, and *PG2013505*) were downregulated in HMS. In pigmented skin, one gene (*PG0002484*) was upregulated both in the skin of HM and LT,

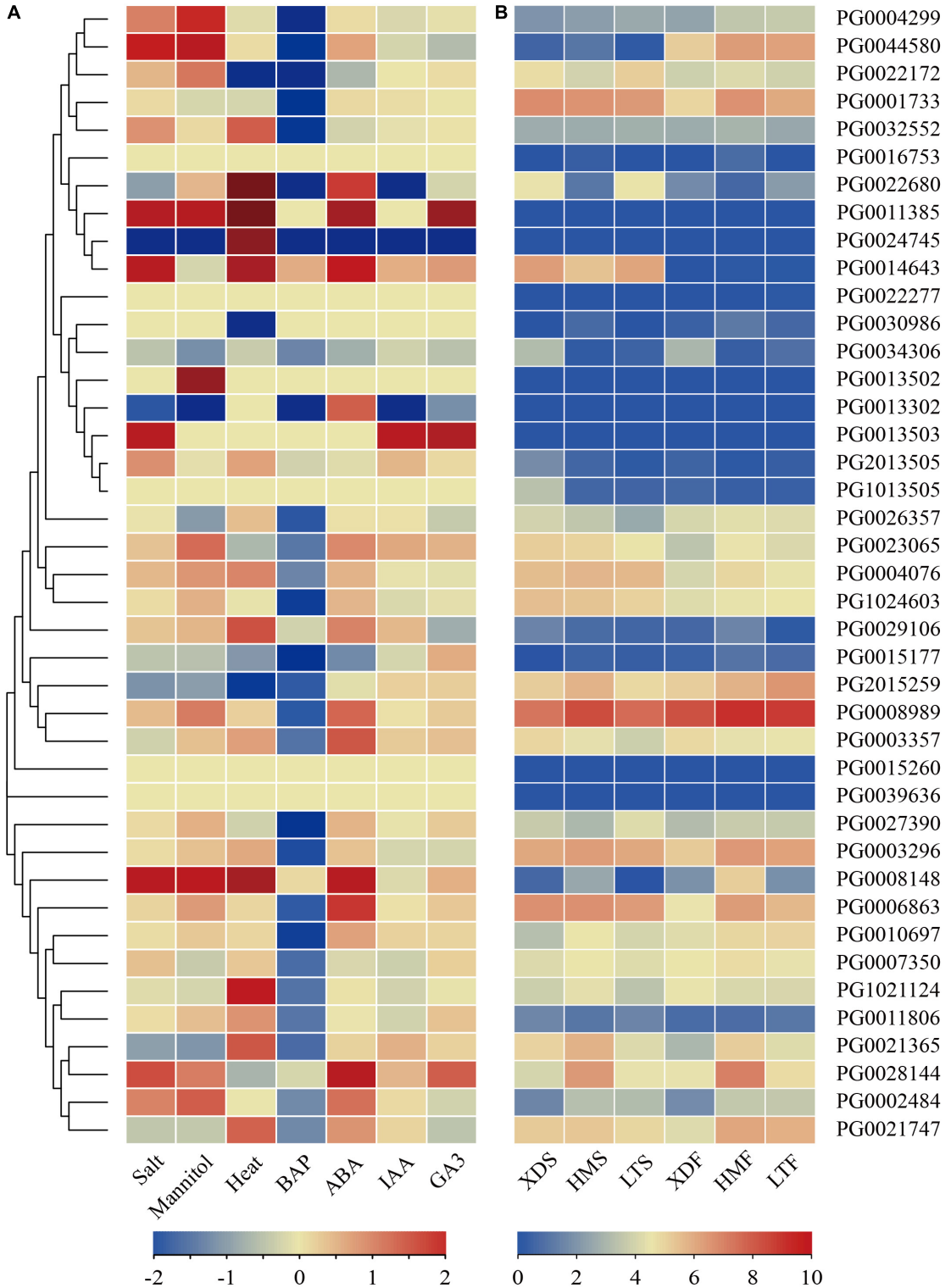

FIGURE 8 | The expression profiles of *StNF-Y* genes. **(A)** The expression profiles of *StNF-Y* genes under abiotic stresses (salt, mannitol, and heat stress) and hormone treatments (BAP, ABA, IAA, and GA3) in DM potato. The color scale was plotted using the $\log_2$FC of each gene. **(B)** The expression profiles of *StNF-Y* genes in white and pigmented potato tuber skin and flesh. XDS, LTS, HMS represent the white skin of white potato cultivar (Xindaping, XD), the red skin of red potato cultivar (Lingtianhongmei, LT), and the purple skin of purple potato cultivar (Heimeiren, HM), respectively. XDF, LTF, and HMF represent the white flesh of XD, the red flesh of LT, and the purple flesh of HM, respectively. The color in the heat map represents the FPKM value using logarithm with a base of 2. The heatmap was constructed by TBtools software. BAP, benzylaminopurine; ABA, abscisic acid; IAA, indole acetic acid; GA3, gibberellic acid.

and three genes (*PG0034306*, *PG1013505*, and *PG2013505*) were downregulated both in the skin HM and LT (**Figure 8B** and **Supplementary Table 9**).

In tuber flesh, eight *StNF-Y* genes were not expressed (FPKM = 0), and these eight genes also were not expressed in the skin. Compared with the white flesh of XD, a total of 17 *StNF-Y* genes were differentially expressed (FPKM > 1 and $|\log_2 FC| > 1$) in pigmented flesh, of which eight genes (*PG0001733*, *PG0002484*, *PG0003296*, *PG0004299*, *PG0006863*, *PG0021365*, *PG0021747*, and *PG0044580*) were upregulated in both LTF and HMF and one gene (*PG0034306*) was downregulated in pigmented flesh (**Figure 8B** and **Supplementary Table 9**). It is noteworthy that one gene (*PG0002484*) was upregulated, and one gene (*PG0034306)* was downregulated in pigmented tissues (skin and flesh).

Eight *StNF-Y* genes with relatively high expression levels in pigmented tissues were selected as targets, and the reliability of RNA-seq dataset was verified by qPCR. The results confirmed that the qPCR expression patterns were in agreement with the RNA-seq dataset (**Figure 9A**), showing a high correlation ($R^2 = 0.8393$) between the RNA-seq dataset and the simple linear regression equation of qPCR ($y = 0.6430x + 0.0924$), indicating the consistency of the two analysis methods (**Figure 9B**).

The Interaction Network of StNF-Y Proteins

Previous studies have shown that NF-Y subunits are involved in flavonoid synthesis, and they play a role through interaction with each other (Hackenberg et al., 2012; Mu et al., 2013; Wang et al., 2020). We further investigated the interaction between StNF-Y proteins that might be involved in flavonoid biosynthesis. Based on the RNA-seq dataset of pigmented potato cultivars, we selected highly expressed StNF-Y (FPKM > 5) to predict the protein interaction network with median confidence (combination score > 400). The results showed that a total of 89 pairs of interacting proteins were predicted. Twenty-one StNF-Y proteins (eight StNF-YAs, 10 StNF-YBs, and three StNF-YCs) were included in the protein interaction network, of which, eight StNF-Y genes (*PG0003296*, *PG0004299*, *PG0021365*, *PG0002484*, *PG0006863*, *PG0021747*, *PG0044580*, and *PG0034306*) were differentially expressed in pigmented tissues compared to that in white tissues. These eight StNF-Ys contained four StNF-YA subunits, three StNF-YBs, and one StNF-YC. In addition, we found that StNF-YA subunits did not interact with each other at high confidence (combined score > 400), whereas they showed a direct interaction with StNF-YBs or StNF-YCs (**Figure 10** and **Supplementary Tables 10, 11**).

DISCUSSION

NF-Ys are important transcription factors in plants, which play critical roles in plant growth, development, and stress responses (Yan et al., 2013; Liu et al., 2016; Xu et al., 2016; Yu et al., 2020). However, little is known about NF-Ys in potato. In the present study, a total of 41 *StNF-Y* genes were identified, including 10 *StNF-YAs*, 22 *StNF-YBs*, and nine *StNF-YCs*. The bioinformatics analysis included a study of the phylogenetic relationships between potato and *Arabidopsis* NF-Y proteins and an analysis of chromosomal location, gene structures, encoded protein conserved motifs, and gene duplication events of *StNF-Y* genes. In addition, the expression pattern of *StNF-Y* genes in different tissues and their response to various abiotic stresses and hormone treatments were analyzed. This study provided comprehensive information for further investigation of their biological functions and the evolution of StNF-Y gene family.

Expansions of StNF-Y Gene Family in Potato

Gene duplication plays a major role in the expansion of gene families and may be an important driver of evolutionary diversification (Lawton-Rauh, 2003). In the present study, nine pairs of duplication genes were identified in potato, of which the number of *StNF-Y* genes arranged in tandem duplications contributed to 12.20% (5/41) and segmental duplication events contributed to 24.39% (10/41). These results suggested that tandem and segmental duplication events were the main mechanisms of StNF-Y gene family evolution and that some of the *StNF-Y* genes were possibly generated by gene duplication events.

Duplication events often lead to increased functional diversity between duplicated gene pairs, and some new members may acquire new functions or lose their original functions (Dias et al., 2003; Singh et al., 2012). In our study, three *StNF-Y* genes (*PG0039636*, *PG0015260*, and *PG2015259*) were tandem duplication, of which *PG2015259* was downregulated under various abiotic stresses (salt, mannitol, and heat) and highly expressed in 13 tissues (FPKM > 5), while the *PG0039636* and *PG0015260* were not expressed (FPKM = 0), indicating that *PG0039636* and *PG0015260* may become pseudogenes during duplication events and evolution. In addition, some duplication gene pairs were expressed consistently, but some were expressed inconsistently, such as *PG0021747*/*PG0002484* and *PG0021365*/*PG0028144* that were segmental duplication gene pairs. Among them, *PG0021747* and *PG0002484* were upregulated in pigmented flesh; however, *PG0021365* was downregulated and *PG28144* was upregulated under salt and mannitol stresses. In conclusion, gene duplication provided a large number of functional or non-functional genes for the potato genome.

Phylogenetic Analysis and Evolution of StNF-Ys

To investigate the relationship of NF-Y proteins among potato, tomato, and *Arabidopsis*, an unroot phylogenetic tree was constructed using complete NF-Y protein sequences. The results showed that there were 10 NF-YA in all three plants (potato, tomato, and *Arabidopsis*). However, the amounts of NF-YB and NF-YC varied widely; there are 13 YBs and 13 YCs in *Arabidopsis*, 22 YBs and 9 YCs in potato, and 27 YBs and 17 YCs in tomato, respectively, suggesting that NF-YB and NF-YC play an important role in the expansion of NF-Y gene family.

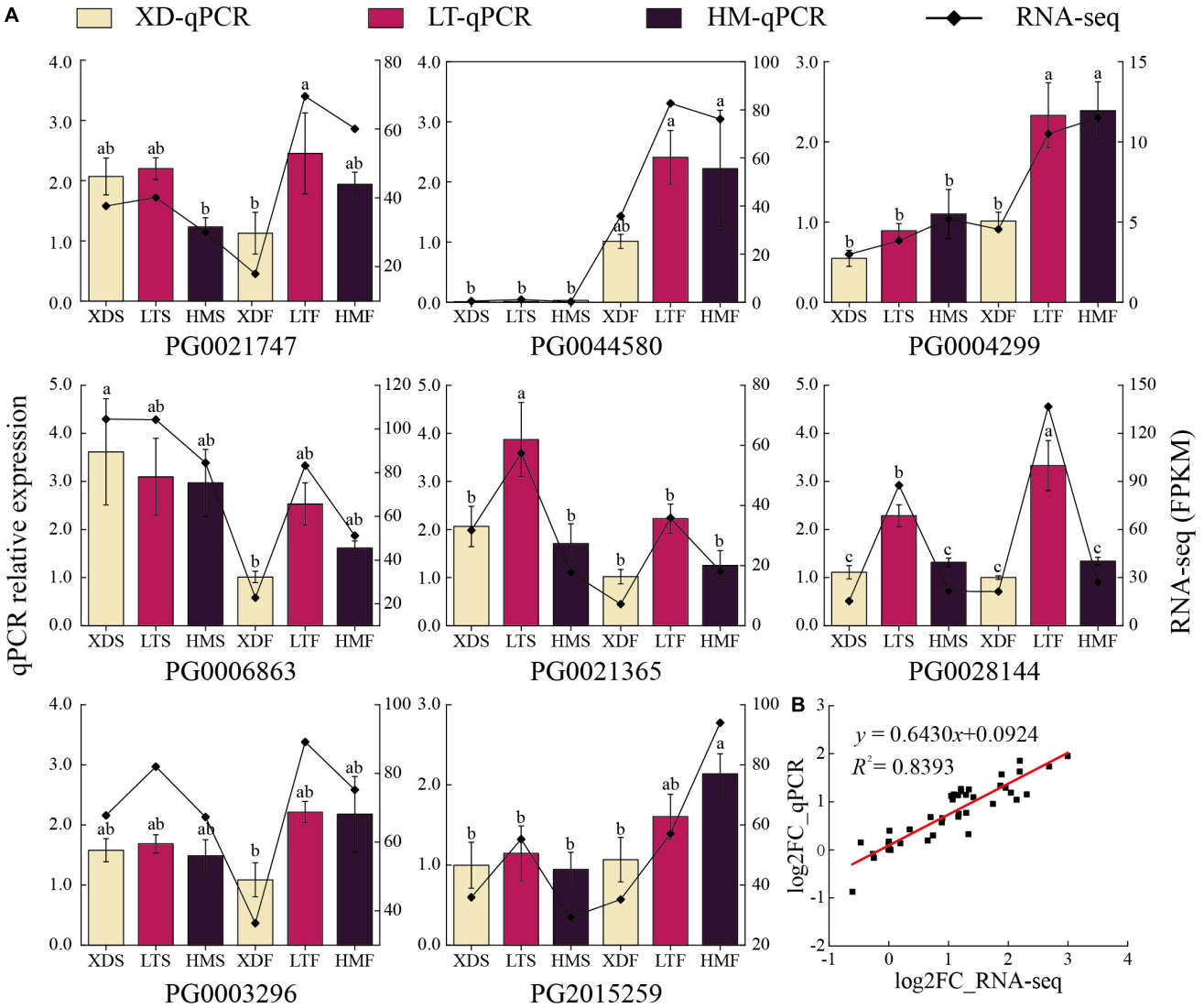

FIGURE 9 | The qPCR expression analyses of eight *StNF-Y* genes in **(A)** the skin/flesh (S/F) of white (XD), purple (HM), and red (LT) cultivars. **(B)** The correlation between qPCR and RNA-seq analysis. Error bars represent the means ± SE of three biological replicates. Different letters above bars denote significant difference at $p < 0.05$. Column graphs represent the relative expression levels of genes determined by qPCR; the values are shown on the left *y*-axis, and line graphs represent the FPKMs obtained by RNA-seq; the values are shown on the right *y*-axis.

Studies have shown that AtNF-YC1/2/3/4/9 controls flowering time (Siefers et al., 2009; Kumimoto et al., 2010; Hackenberg et al., 2012). Three StNF-YCs (PG0003357, PG0008989, and PG2015259) were clustered with AtNF-YC1/2/3/4/9, indicating that these StNF-Ys may have a similar function as AtNF-YC1/2/3/4/9. Previous studies have reported that the *AtLEC1* and *AtLEC1*-like play an important role in embryo development and fruit ripening (Kwong et al., 2003; Li et al., 2016; Gago et al., 2017). It is noteworthy that AtLEC1 (AtNF-YB9) and AtLEC1-like (AtNF-YB6) formed a distinct cluster in NF-YB (**Figure 4**). Our study found that eight *StNF-YB* genes were clustered with *AtLEC1* and *AtLEC1*-like in one cluster, of which six *StNF-YB* genes were specifically expressed in immature fruits. It is suggesting that *LEC1* and *LEC1*-like genes were extensively extended during potato evolution and that these genes may be associated with embryo development and fruit ripening.

Expression Profile Analysis of StNF-Y Genes Under Abiotic Stress and Hormone Treatment

Previous studies have shown that the NF-Y family is involved in various abiotic stresses (Li L. et al., 2013; Li Y. J. et al., 2013; Alam et al., 2015; Feng et al., 2015). In *Arabidopsis*, *AtNF-YA1* (*At5G12840*) is significantly induced by NaCl, mannitol, PEG, and ABA, and it is involved in the regulation of post-germination growth arrest under salt stress (Li Y. J. et al., 2013). *PG0006863* was homologous to *AtNF-YA1*, and it was highly

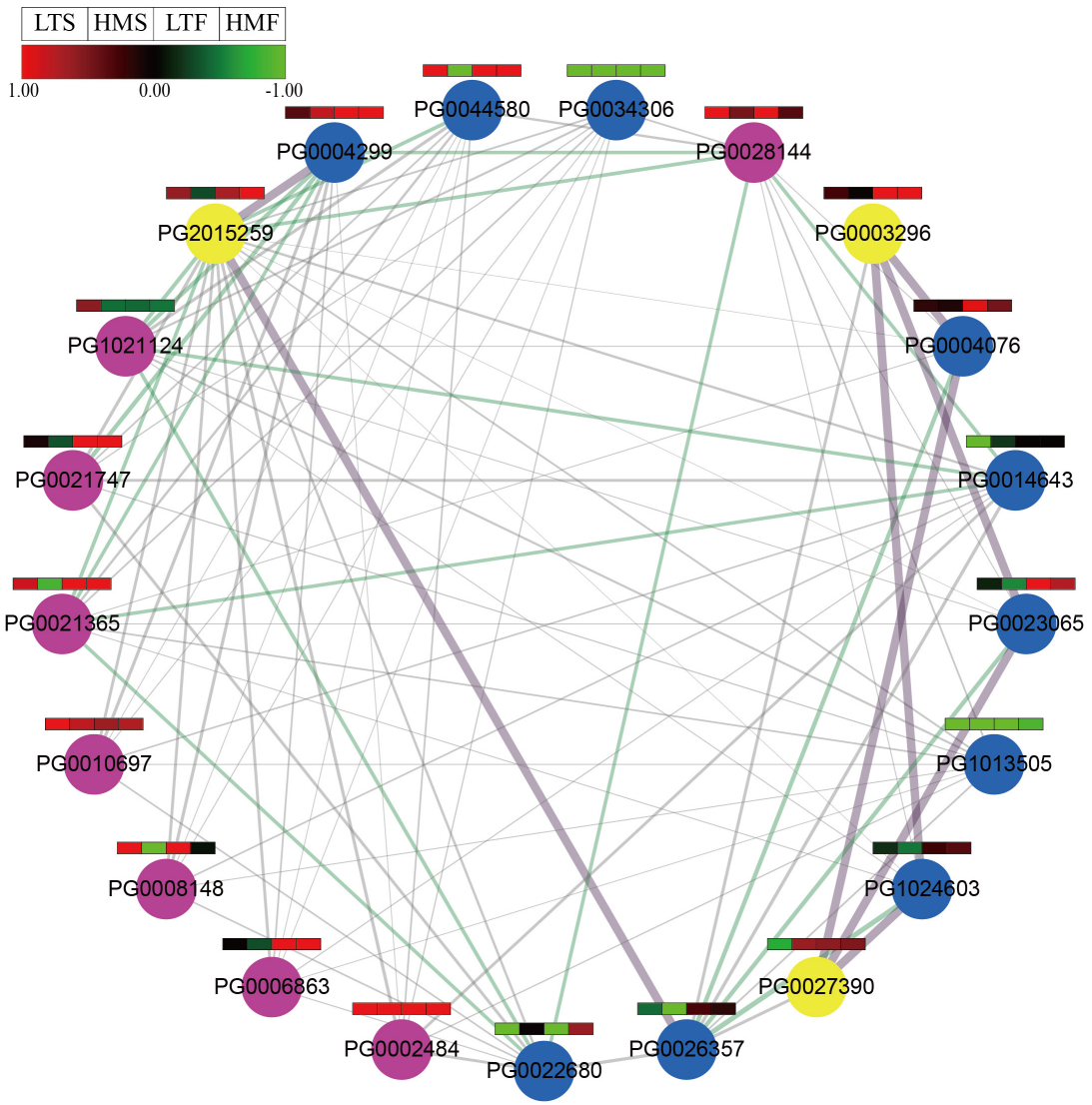

FIGURE 10 | Protein interaction network predicted for StNF-Y (combined score of 400, medium confidence). The purple lines represent the highest confidence (combined score > 900), the green lines indicate high confidence (combined score > 700), and the gray lines indicate medium confidence (combined score > 400). NF-YA subunits are highlighted in purple, NF-YB subunits are highlighted in blue, and NF-YC subunits are highlighted in yellow, respectively. The color scale was plotted using the $\log_2$FC of each gene.

expressed under mannitol treatment and upregulated under ABA treatment in DM potato, indicating that *PG0006863* may function in a similar way as *AtNF-YA1* under mannitol treatment. Li et al. (2008) found that overexpression of *AtNF-YA5* (*AT1G54160*) in *Arabidopsis* can increase the tolerance of plants to drought stress through the ABA-dependent pathway. *PG0002484* was homologous to *AT1G54160*, which was upregulated under salt, mannitol, and ABA treatments in DM potato, suggesting that *PG0002484* may have similar functions to *AtNF-YA5*. Overexpression of *AtNF-YA7* (*AT1G30500*) enhanced tolerance to various abiotic stresses in *Arabidopsis* (Leyva-González et al., 2012). *PG1021124*, homologous to *AtNF-YA7*, was upregulated under heat stress and differentially expressed under various hormone treatments in DM potato, suggesting that *PG1021124* may have similar functions to *AtNF-YA7* under abiotic stresses. Overexpression of *NF-YA10* can improve tolerance to drought and salt stress in *Arabidopsis* (Ma et al., 2015; Yu et al., 2020). *PG0028144* and *PG0021365* were a pair of segmental duplication genes, which were clustered with *AtNF-YA10* (*AT6G06510*). *PG0021365* was differentially expressed under mannitol and heat treatments, while *PG0028144* was upregulated under salt, mannitol, and ABA treatments. The results indicated that *PG0028144* may be induced by drought stress in an ABA-dependent manner, while *PG0021365* was not. Sato et al. (2019) found that overexpression of *AtNF-YB2* (*AT5G47640*) and *AtNF-YB3* (*AT4G14540*) specifically enhances drought and heat stress tolerance, respectively. *PG0004299*, *PG0044580*, and *PG0022172* were clustered with *AtNF-YB2* and *AtNF-YB3* in

one cluster. Among them, *PG0004299* and *PG0044580* had the same expression pattern, which were upregulated under salt and mannitol treatments, indicating that *PG0004299* and *PG0044580* may have similar functions to *AtNF-YB2*. The expression of *PG0022172* was upregulated under mannitol treatment, and downregulated under heat treatment, suggesting that *PG0022172* may have functions of both *AtNF-YB2* and *AtNF-YB3*. In addition, *PG0008989* and *PG0023065* were upregulated under mannitol and ABA treatments, while *PG2015259* was downregulated under three abiotic stresses (salt, mannitol, and heat), suggesting that these *StNF-Y* genes might be putative regulators of response to abiotic stress.

Expression of StNF-Y Genes in Pigmented Tuber Tissue

The expression profiles of *StNF-Y* genes in different tuber tissues (skin and flesh) of tetraploid pigmented potato cultivars were analyzed based on the RNA-seq dataset. Studies have shown that the NF-Y genes play crucial roles in flavonoid biosynthesis, e.g., NF-YB8 interacts with NF-YC1 and then recruits NF-YA1/9 to form NF-Y complexes that bind the CCAAT element in the CHS1 promoter to regulate the flavonoid biosynthesis in tomato (Wang et al., 2020). *PG0001733* was homologous to *AtNF-YB8* (*AT2G37060*) and clustered with SlNF-YB8c in one cluster, which was upregulated in pigmented flesh (LTF and HMF). *PG2015259* was clustered with AtNF-YC1 (AT3G48590) and SlNF-YC1a (Solyc03g110860) in one cluster and was highly expressed in LTF and upregulated in HMF. *PG0006863* and *PG0008148* were homologous to *SlNF-YA9* (*Solyc01g087240*), of which *PG0006863* was upregulated in LTF and HMF and *PG0008148* was upregulated in LTS and LTF. In addition, *PG0010697* was homologous to *SlNF-YA1a* (*Solyc01g008490*) and was highly expressed in pigmented tissues and upregulated in LTS. These results indicated that these five genes may have similar functions to *SlNF-YA1/9*, *SlNF-YB8c*, and *SlNF-YC1*a in flavonoid biosynthesis. Furthermore, six genes were upregulated in pigmented tissues (LTF and HMF), of which three genes (*PG0002484*, *PG0021365*, and *PG0021747*) belonged to StNF-YA subunit, two genes (*PG0004299* and *PG0044580*) were assigned to StNF-YB subunit, and one gene (*PG0003296*) was grouped into StNF-YC subunit.

In addition, we analyzed the correlation coefficient (CC) between differentially expressed *StNF-Ys* of potato tuber skin and flesh in RNA-seq data and anthocyanin content. The results showed that the expression of some *StNF-Y* genes was highly correlated to anthocyanin content in tuber flesh, such as *PG0044580* and *PG0021747* with CC > 0.7; *PG0004299*, *PG0002484*, and *PG0010697* with CC > 0.8; and *PG2015259* (CC = 0.8) (**Supplementary Table 12**), suggesting that these *StNF-Y* genes may be involved in the anthocyanin biosynthesis in potato tuber flesh, and their functions require further investigation.

Through the prediction of StNF-Y protein interaction network, we found that there were interaction relationships (high confidence with combined score > 700) between these differentially expressed genes (**Figure 10** and **Supplementary Table 10**), i.e., two StNF-YAs (PG0021747 and PG0021365) directly interacted with one StNF-YB (PG0004299), and one StNF-YB (PG0004299) and one StNF-YA (PG0021365) interacted with one StNF-YC (PG2021259). These StNF-Ys may be involved in the biosynthesis of anthocyanins by interacting with each other to form the completed NF-Y complex, and their functions need to be further studied.

CONCLUSION

In this study, we identified a total of 41 *StNF-Y* genes from the potato genome, including 10 *StNF-YA*, 22 *StNF-YB*, and nine *StNF-YC*, and their distribution on chromosomes, multiple alignments, gene structure, conserved motifs, and phylogenetic relationships were analyzed. Synteny analysis showed that nine pairs of duplication genes were identified in the potato genome, which played an important role in the expansion of the StNF-Y gene family. Synteny analysis indicated that 14 and 38 pairs of StNF-Y genes were orthologous to *Arabidopsis* and tomato, respectively. Six *StNF-YB* genes were specifically expressed in immature fruit, which may be involved in fruit ripening. In addition, the *StNF-Y* genes displayed different expression patterns under the salt, mannitol, heat, and hormone treatments in DM potato. Furthermore, we analyzed the expression of *StNF-Y* genes in tetraploid pigmented potato cultivars and the protein interaction networks that were highly expressed in anthocyanin biosynthesis. The results of this study provide valuable information for further studies on the functions of the StNF-Y gene family.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YHL, JB, and JLZ conceptualized the study. ZL, YML, JYZ, WM, ZB, and ZTL curated the data. ZL and CS conducted the formal analysis. ZL, YHL, and JLZ acquired funding. ZL contributed the software and wrote the original draft. YHL wrote, reviewed, and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Scientific Research Start-up Funds for Openly-recruited Doctors of Gansu Agricultural University (GAU-KYQD-2020-11), the Program for Innovation Ability Improvement of the Higher Education Institutions of Gansu Province (2020A-056), the Program for Longyuan Youth Innovation Team of Gansu Province (LRYCT-2020-1), the National Natural Science Foundation of China (31601356 and 31860398), the Gansu Science Foundation for Distinguished

Young Scholars (17JR5RA138), the Fuxi Talent Project of Gansu Agricultural University (Ganfx-02Y04), and the China Agriculture Potato Research System of MOF and MARA (CARS-09-P14).

ACKNOWLEDGMENTS

We would like to thank Yuanyuan Pu of Gansu Agricultural University for her help in the improvement of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.739989/full#supplementary-material

Supplementary Table 1 | Primers used for qPCR.

Supplementary Table 2 | Characteristic features of 41 StNF-Y genes.

Supplementary Table 3 | Conserved motifs identified by MEME among StNF-Ys.

Supplementary Table 4 | Ka/Ks ratios of tandemly and segmentally duplicated *StNF-Y* genes.

Supplementary Table 5 | Ka/Ks ratios of the syntenic relationships for NF-Y genes between potato and *Arabidopsis*.

Supplementary Table 6 | Ka/Ks ratios of the syntenic relationships for NF-Y genes between potato and tomato.

Supplementary Table 7 | RNA-seq data of StNF-Y genes in different tissues.

Supplementary Table 8 | RNA-seq data of StNF-Y genes under abiotic stress and hormone treatment.

Supplementary Table 9 | The FPKM value of StNF-Y genes in white and pigmented tissues.

Supplementary Table 10 | Protein interaction network of StNF-Ys.

Supplementary Table 11 | The FPKM value of genes in protein interaction networks.

Supplementary Table 12 | The correlation between the expression of StNF-Ys and anthocyanin content in potato tuber (skin and flesh).

REFERENCES

Alam, M. M., Tanaka, T., Nakamura, H., Ichikawa, H., Kobayashi, K., Yaeno, T., et al. (2015). Overexpression of a rice heme activator protein gene (OsHAP2E) confers resistance to pathogens, salinity and drought, and increases photosynthesis and tiller number. *Plant Biotechnol. J.* 13, 85–96. doi: 10.1111/pbi.12239

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4:10. doi: 10.1186/1471-2229-4-10

Cao, S., Kumimoto, R. W., Gnesutta, N., Calogero, A. M., Mantovani, R., and Holt, B. F. III (2014). A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the FLOWERING LOCUS T promoter and regulates the timing of flowering in *Arabidopsis*. *Plant Cell* 26, 1009–1017. doi: 10.1105/tpc.113.120352

Castañeda-Ovando, A., Pacheco-Hernández, M. D. L., Páez-Hernández, M. E., Rodríguez, J. A., and Galán-Vidal, C. A. (2009). Chemical studies of anthocyanins: a review. *Food Chem.* 113, 859–871. doi: 10.1016/j.foodchem.2008.09.001

Castellarin, S. D., Pfeiffer, A., Sivilotti, M. D., Peterlunger, E., and Di Gaspero, G. (2007). Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. *Plant Cell Environ.* 30, 1381–1399. doi: 10.1111/j.1365-3040.2007.01716.x

Chen, C., Xia, R., Chen, H., and He, Y. (2018). TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *BioRxiv* [Preprint] doi: 10.1101/289660v2

Dias, A. P., Braun, E. L., McMullen, M. D., and Grotewold, E. (2003). Recently duplicated maize *R2R3 Myb* genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiol.* 131:610. doi: 10.1104/pp.012047

Feng, Z.-J., He, G.-H., Zheng, W.-J., Lu, P. -, Chen, M., Gong, Y.-M., et al. (2015). Foxtail millet NF-Y families: genome-wide survey and evolution analyses identified two functional genes important in abiotic stresses. *Front. Plant Sci.* 6:1142. doi: 10.3389/fpls.2015.01142

Fossen, T., and Andersen, ØM. (2000). Anthocyanins from tubers and shoots of the purple potato, *Solanum tuberosum*. *J. Hortic. Sci. Biotechnol.* 75, 360–363. doi: 10.1080/14620316.2000.11511251

Gago, C. D., Paschalidis, K., Guerreiro, A., Miguel, G., Antunes, D., and Hilioti, Z. (2017). Targeted gene disruption coupled with metabolic screen approach to uncover the LEAFY COTYLEDON1-LIKE4 (L1L4) function in tomato fruit metabolism. *Plant Cell Rep.* 36, 1065–1082. doi: 10.1007/s00299-017-2137-9

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., et al. (2005). "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, ed. J. M. Walker (Totowa, NJ: Humana Press), 571–607. doi: 10.1385/1-59259-890-0:571

Guo, A. Y., Zhu, Q. H., and Xin, C. (2007). GSDS:a gene structure display server. *Hereditas* 29, 1023–1026. doi: 10.1360/yc-007-1023

Guo, F., Liu, C., Xia, H., Bi, Y., Zhao, C., Zhao, S., et al. (2013). Induced expression of AtLEC1 and AtLEC2 differentially promotes somatic embryogenesis in transgenic tobacco plants. *PLoS One* 8:e71714. doi: 10.1371/journal.pone.0071714

Hackenberg, D., Wu, Y., Voigt, A., Adams, R., Schramm, P., and Grimm, B. (2012). Studies on differential nuclear translocation mechanism and assembly of the three subunits of the *Arabidopsis thaliana* transcription Factor NF-Y. *Mol. Plant* 5, 876–888. doi: 10.1093/mp/ssr107

He, J., and Giusti, M. M. (2010). Anthocyanins: natural colorants with health-promoting properties. *Annu. Rev. Food Sci. Technol.* 1, 163–187. doi: 10.1146/annurev.food.080708.100754

Hollman, P., and Katan, M. B. (1997). Absorption, metabolism and health effects of dietary flavonoids in man. *Biomed. Pharmacother.* 51, 305–310. doi: 10.1016/s0753-3322(97)88045-6

Kahle, J., Baake, M., Doenecke, D., and Albig, W. (2005). Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin beta and importin 13. *Mol. Cell Biol.* 25, 5339–5354. doi: 10.1128/mcb.25.13.5339-5354.2005

Kim, J., Lee, W. J., Vu, T. T., Jeong, C. Y., Hong, S. W., and Lee, H. (2017). High accumulation of anthocyanins via the ectopic expression of AtDFR confers significant salt stress tolerance in *Brassica napus* L. *Plant Cell Rep.* 36, 1215–1224. doi: 10.1007/s00299-017-2147-7

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874.

Kumimoto, R. W., Zhang, Y., Siefers, N., and Holt, B. F. III (2010). NF–YC3, NF–YC4 and NF–YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*. *Plant J.* 63, 379–391. doi: 10.1111/j.1365-313x.2010.04247.x

Kwong, R. W., Bui, A. Q., Lee, H., Kwong, L. W., Fischer, R. L., Goldberg, R. B., et al. (2003). LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* 15, 5–18. doi: 10.1105/tpc.006973

Laloum, T., De Mita, S., Gamas, M. B., and Niebel, A. (2013). CCAAT-box binding transcription factors in plants: y so many? *Trends Plant Sci.* 18, 157–166. doi: 10.1016/j.tplants.2012.07.004

Lawton-Rauh, A. (2003). Evolutionary dynamics of duplicated genes in plants. *Mol. Phylogenet. Evol.* 29, 396–409. doi: 10.1016/j.ympev.2003.07.004

Lee, H., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (2003). *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2152–2156. doi: 10.1073/pnas.0437909100

Leyva-González, M. A., Ibarra-Laclette, E., Cruz-Ramírez, A., and Herrera-Estrella, L. (2012). Functional and transcriptome analysis reveals an acclimatization strategy for abiotic stress tolerance mediated by *Arabidopsis* NF-YA family members. *PLoS One* 7:e48138. doi: 10.1371/journal.pone.0048138

Li, L., Yu, Y., Wei, J., Huang, G., Zhang, D., Liu, Y., et al. (2013). Homologous HAP5 subunit from Picea wilsonii improved tolerance to salt and decreased sensitivity to ABA in transformed *Arabidopsis*. *Planta* 238, 345–356. doi: 10.1007/s00425-013-1894-0

Li, S., Li, K., Ju, Z., Cao, D., Fu, D., Zhu, H., et al. (2016). Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening. *BMC Genom.* 17:36. doi: 10.1186/s12864-015-2334-2

Li, W. X., Oono, Y., Zhu, J., He, X. J., Wu, J. M., Iida, K., et al. (2008). The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *Plant Cell* 20, 2238–2251. doi: 10.1105/tpc.108.059444

Li, X. Y., Hooft van Huijsduijnen, R., Mantovani, R., Benoist, C., and Mathis, D. (1992). Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain. *J. Biol. Chem.* 267, 8984–8990. doi: 10.1016/s0021-9258(19)50377-5

Li, Y. J., Fang, Y., Fu, Y. R., Huang, J. G., Wu, C. A., and Zheng, C. C. (2013). NFYA1 is involved in regulation of postgermination growth arrest under salt stress in *Arabidopsis*. *PLoS One* 8:e61289. doi: 10.1371/journal.pone.0061289

Liang, M., Yin, X., Lin, Z., Zheng, Q., Liu, G., and Zhao, G. (2014). Identification and characterization of NF-Y transcription factor families in canola (*Brassica napus* L.). *Planta* 239, 107–126. doi: 10.1007/s00425-013-1964-3

Liu, X., Hu, M. Huang, Tang, Y., Li, Y., Li, L., and Hou, X. (2016). The NF-YC-RGL2 module integrates GA and ABA signalling to regulate seed germination in *Arabidopsis*. *Nat. Commun.* 7:12768.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta\ CT}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Ma, X., Zhu, X., Li, C., Song, Y., Zhang, W., Xia, G., et al. (2015). Overexpression of wheat NF-YA10 gene regulates the salinity stress response in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 86, 34–43. doi: 10.1016/j.plaphy.2014.11.011

Malviya, N., Jaiswal, P., and Yadav, D. (2016). Genome- wide characterization of Nuclear Factor Y (NF-Y) gene family of sorghum [Sorghum bicolor (L.) Moench]: a bioinformatics approach. *Physiol. Mol. Biol. Plants* 22, 33–49. doi: 10.1007/s12298-016-0349-z

Mantovani, R. (1999). The molecular biology of the CCAAT-binding factor NF-Y. *Gene* 239, 15–27. doi: 10.1016/s0378-1119(99)00368-6

Mu, J., Tan, H., Hong, S., Liang, Y., and Zuo, J. (2013). *Arabidopsis* transcription factor genes NF-YA1, 5, 6, and 9 play redundant roles in male gametogenesis, embryogenesis, and seed development. *Mol. Plant* 6, 188–201. doi: 10.1093/mp/sss061

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., et al. (2013). Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell* 152, 132–143.

Ni, Z., Hu, Z., Jiang, Q., and Zhang, H. (2013). GmNFYA3, a target gene of miR169, is a positive regulator of plant tolerance to drought stress. *Plant Mol. Biol.* 82, 113–129. doi: 10.1007/s11103-013-0040-5

Petroni, K., Kumimoto, R. W., Gnesutta Calvenzani, N., Fornari, M., Tonelli, C., Holt, B. F., et al. (2012). The promiscuous life of plant nuclear factor Y transcription factors. *Plant Cell* 24:4777. doi: 10.1105/tpc.112.105734

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110. doi: 10.1126/science.290.5499.2105

Sato, H., Mizoi, J., Tanaka, H., Maruyama, K., Qin, F., Osakabe, Y., et al. (2014). *Arabidopsis* DPB3-1, a DREB2A interactor, specifically enhances heat stress-induced gene expression by forming a heat stress-specific transcriptional complex with NF-Y subunits. *Plant Cell* 26, 4954–4973. doi: 10.1105/tpc.114.132928

Sato, H., Suzuki, T., Takahashi, F., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2019). NF-YB2 and NF-YB3 have functionally diverged and differentially induce drought and heat stress-specific genes. *Plant Physiol.* 180, 1677–1690. doi: 10.1104/pp.19.00391

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.

Siefers, N., Dang, K. K., Kumimoto, R. W., Bynum, W. E. T., Tayrose, G., and Holt, B. F. III (2009). Tissue-specific expression patterns of *Arabidopsis* NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiol.* 149, 625–641. doi: 10.1104/pp.108.130591

Singh, A., Pandey Baranwal, A., Kapoor, S., and Pandey, G. K. (2012). Comprehensive expression analysis of rice phospholipase D gene family during abiotic stresses and development. *Plant Signal. Behav.* 7, 847–855. doi: 10.4161/psb.20385

Sorin, C., Declerck, M., Christ, A., Blein, T., Ma, L., Lelandais-Brière, C., et al. (2014). A miR169 isoform regulates specific NF-YA targets and root architecture in *Arabidopsis*. *New Phytol.* 202, 1197–1211. doi: 10.1111/nph.12735

Spooner, D. M., McLean, K., Ramsay, G., Waugh, R., and Bryan, G. J. (2005). A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14694–14699. doi: 10.1073/pnas.0507400102

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2016). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368.

Tang, X., Zhang, N., Si, H., and Calderón-Urrea, A. (2017). Selection and validation of reference genes for RT-qPCR analysis in potato under abiotic stress. *Plant Methods* 13:85.

The Potato Genome Sequencing Consortium, Xu X., Pan, S., Cheng, S., Zhang, B., Mu, D., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475:189.

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Heredity* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* 8, 77–80. doi: 10.1016/s1672-0229(10)60008-3

Wang, J., Li, G., Li, C., Zhang, C., Cui, L., Ai, G., et al. (2020). NF-Y plays essential roles in flavonoid biosynthesis by modulating histone modifications in tomato. *New Phytol.* 229, 3237–3252. doi: 10.1111/nph.17112

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Wang, Y., Xu, W., Chen, Z., Han, B., Haque, M. E., and Liu, A. (2018). Gene structure, expression pattern and interaction of Nuclear Factor-Y family in castor bean (*Ricinus communis*). *Planta* 247, 559–572. doi: 10.1007/s00425-017-2809-2

Xing, Y., Fikes, J. D., and Guarente, L. (1993). Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain. *EMBO J.* 12, 4647–4655. doi: 10.1002/j.1460-2075.1993.tb06153.x

Xu, J. J., Zhang, X. F., and Xue, H. W. (2016). Rice aleurone layer specific OsNF-YB1 regulates grain filling and endosperm development by interacting with an ERF transcription factor. *J. Exp. Bot.* 67, 6399–6411. doi: 10.1093/jxb/erw409

Yan, D. H., Xia, X., and Yin, W. (2013). NF-YB family genes identified in a poplar genome-wide analysis and expressed in populus euphratica are responsive to drought stress. *Plant Mol. Biol. Rep.* 31, 363–370. doi: 10.1007/s11105-012-0508-5

Yan, H., Wu, F., Jiang, G., Xiao, L., Li, Z., Duan, X., et al. (2019). Genome-wide identification, characterization and expression analysis of NF-Y gene family in relation to fruit ripening in banana. *Postharvest Biol. Technol.* 151, 98–110. doi: 10.1016/j.postharvbio.2019.02.002

Yang, J., Zhu, J., and Yang, Y. (2017). Genome-wide identification and expression analysis of NF-Y transcription factor families in watermelon (*Citrullus lanatus*). *J. Plant Growth Regul.* 36, 590–607. doi: 10.1007/s00344-017-9670-1

Yang, W., Lu, Z., Xiong, Y., and Yao, J. (2017). Genome-wide identification and co-expression network analysis of the OsNF-Y gene family in rice. *Crop J.* 5, 21–31. doi: 10.1016/j.cj.2016.06.014

Yu, Y., Bai, Y., Wang, Y. Wang, Wan, H., Liu, C., and Ni, Z. (2020). Soybean nuclear factor YA10 positively regulates drought resistance in transgenic *Arabidopsis thaliana*. *Environ. Exp. Bot.* 180:104249. doi: 10.1016/j.envexpbot.2020.104249

Zhao, H., Wu, D., Kong, F., Lin, K., Zhang, H., and Li, G. (2017). The *Arabidopsis thaliana* nuclear factor Y transcription factors. *Front. Plant Sci.* 7:2045. doi: 10.3389/fpls.2016.02045

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 September 2021
doi: 10.3389/fgene.2021.753638

Genome-Wide Identification and Evolutionary Analysis of the SRO Gene Family in Tomato

Ning Li[1,2,3†], *Ruiqiang Xu*[1,2,3†], *Baike Wang*[2,3], *Juan Wang*[2,3], *Shaoyong Huang*[1,2,3], *Qinghui Yu*[2,3*] *and Jie Gao*[1*]

[1]College of Forestry and Horticulture, Xinjiang Agricultural University, Urumqi, China, [2]Institute of Horticultural Crops, Xinjiang Academy of Agricultural Sciences, Urumqi, China, [3]Key Laboratory of Horticulture Crop Genomics and Genetic Improvement in Xinjiang, Urumqi, China

SRO (SIMILAR TO RCD ONE) is a family of plant-specific small molecule proteins that play an important role in plant growth and development and environmental responses. However, SROs still lack systematic characterization in tomato. Based on bioinformatics methods, *SRO* family genes were identified and characterized from cultivated tomatoes and several wild tomatoes. qRT-PCR was used to study the expression of *SRO* gene in cultivated tomatoes. Phylogenetic and evolutionary analyses showed that *SRO* genes in angiosperms share a common ancestor and that the number of *SRO* family members changed as plants diverged and evolved. Cultivated tomato had six *SRO* members, five of which still shared some degree of identity with the ancestral *SRO* genes. Genetic structure and physicochemical properties showed that tomato *SRO* genes were highly conserved with chromosomal distribution. They could be divided into three groups based on exon-intron structure, and cultivated tomato contained only two of these subclades. A number of hormonal, light and abiotic stress-responsive *cis*-regulatory elements were identified from the promoter of the tomato *SRO* gene, and they also interacted with a variety of stress-responsive proteins and microRNAs. RNA-seq analysis showed that *SRO* genes were widely expressed in different tissues and developmental stages of tomato, with significant tissue-specific features. Expression analysis also showed that *SRO* genes respond significantly to high temperature and salt stress and mediate the tomato hormone regulatory network. These results provide a theoretical basis for further investigation of the functional expression of tomato *SRO* genes and provide potential genetic resources for tomato resistance breeding.

Keywords: SRO gene family, tomato, biotic/abiotic stresses, bioinformatics, phylogenetic

OPEN ACCESS

Edited by:
Rocio Deanna,
University of Colorado Boulder, United States

Reviewed by:
Ali Raza,
Fujian Agriculture and Forestry University, China
Swarup Roy Choudhury,
Indian Institute of Science Education and Research, India

***Correspondence:**
Qinghui Yu
yuqinghui@xaas.ac.cn
Jie Gao
ofc111@163.com

†These authors share first authorship

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

Received: *05 August 2021*
Accepted: *06 September 2021*
Published: *21 September 2021*

Citation:
Li N, Xu R, Wang B, Wang J, Huang S, Yu Q and Gao J (2021) Genome-Wide Identification and Evolutionary Analysis of the SRO Gene Family in Tomato. Front. Genet. 12:753638. doi: 10.3389/fgene.2021.753638

INTRODUCTION

Plant growth and development are dynamic processes that interact with the surrounding environment. Environmental stress has always been one of the major factors limiting plant growth. The long evolutionary process has endowed plants with many means of coping with biotic and abiotic stresses. Transcription factors, as one of the main ways in which plants regulate their life activities, often play an important role in the plant stress response system (Nevo, 2001; Song et al., 2016). Many key stress response transcription factors have been identified in plants, such as MYB (Du et al., 2009), bHLH (Sun et al., 2018), and WRKY (Li et al., 2020). SRO (SIMILAR TO RCD

ONE) is a family of small plant-specific proteins commonly thought to be involved in plant growth and development dynamics and resistance to abiotic stresses (Jaspers et al., 2010a). They are characterized by a C-terminus containing a PARP structural domain involved in a wide range of life activities and an RST structural domain involved in protein-protein interactions, and some SRO members also contain a conserved WWE structural domain associated with the formation of protein globular structures (Jaspers et al., 2010b). *RCD1* was the first member of the SRO family to be discovered and was identified in a yeast 2-hybrid screen using turnip crinkle virus movement protein as bait. *RCD1* is considered to be related to overcoming the oxidative stress-sensitive phenotype of yeast cells (Ahlfors et al., 2004).

Arabidopsis contains 6 SRO family members (*AtSRO1-AtSRO6*). *AtSRO1* is a homologous protein with the same domain as *RCD1* and is involved in the plant oxidative stress response and a variety of hormone-induced gene expression systems (Jaspers et al., 2010a). *AtRCD1* loss-of-function mutants are more sensitive to salt stress and osmotic stress and exhibit the characteristics of early flowering and senescence (Overmyer et al., 2000). There is a functional redundancy between *AtSRO1* and *AtRCD1*, whose double mutants have been observed to be severely defective in Arabidopsis embryonic growth and development and have exhibited a pleiotropic phenotype with dwarf plants, short roots and reduced apical dominance (Jaspers et al., 2009; Teotia and Lamb, 2009; Teotia and Lamb, 2011). Overexpression of *AtSRO5* could mediate proline metabolism in Arabidopsis mitochondria, thereby improving plant salt stress and antioxidant capacity (Borsani et al., 2005). *AtSRO2*, *AtSRO3* and *AtSRO5* have shown changes in transcript levels in response to light stress, salt treatment and exposure to O_3 (Jaspers et al., 2010b; Li et al., 2013), but *AtSRO4* has not yet been reported.

The SRO family has also been characterized in some other species in addition to Arabidopsis. *OsSRO1c* in rice (*Oryza sativa*) is involved in a variety of abiotic stress response processes and interacts with a large number of transcription factors (You et al., 2014). In apple (*Malus domestica*), *MdRCD1* plays a crucial role in the regulation of ROS homeostasis. Its ectopic expression significantly enhances the resistance of transgenic lines to salt and oxidative stress (Li et al., 2017). All *ZmSROs* in maize (*Zea mays*) are specifically expressed in the roots and respond to high salt and drought stress to varying degrees (Jiang et al., 2018). The 30 *TaSRO* members in wheat (*Triticum aestivum*) are divided into two different groups. Most *TaSROs* are highly expressed in one or more tissues, participate in the wheat hormone regulation network and are induced by the wheat stress response (Jiang et al., 2020). Banana (*Musa nana*) contains 6 *MaSROs*, which actively respond to biotic/abiotic stresses by mediating a hormone regulatory network. *MaSRO4* could interact with *MaNAC6* and *MaMYB4* through the PARP domain to regulate downstream signalling pathways (Zhang et al., 2019). The above studies have shown that the SRO family participates in a variety of plant stress responses and regulates the processes of plant growth and development.

Tomato is the largest vegetable cash crop widely planted in the world and is favoured by consumers worldwide. However, tomato cultivation still has not eliminated the effects of biotic and abiotic stress. Every year, billions of tomato yield are lost due to adverse stress (Krishna et al., 2019). Tomato is rich in genetic diversity. Wild tomato usually has strong stress resistance and extremely rich variation. It has advantages over cultivated tomatoes in resisting biotic and abiotic stresses (Lin et al., 2014; Szymański et al., 2020). Studying the response dynamics of wild tomato to adverse environments can provide an important theoretical basis and genetic resources for research on the stress tolerance of cultivated tomato. Although there is evidence that the *JWS-26* gene, which is similar to the *AtSRO5* sequence, is significantly upregulated in tomato roots under salt stress (Babajani et al., 2009), systematic studies on the *SRO* gene family of tomato have not yet been reported. In this study, we used bioinformatics methods to comprehensively identify the *SRO* gene families in cultivated tomato (*S. lycopersicum*, *S. lycopersicum* var. *cerasiforme*) and wild relatives (*S. pennellii*, *S. pimpinellifolium*, *S. chilense*, and *S. lycopersicoides*). The physical and chemical properties, gene structure, evolutionary characteristics and functional expression of the SRO family were analysed, and the unregulated mechanism of the SRO family in tomato in response to different stresses was discussed. This study provides a basis for clarifying the function of the SRO protein and provides a theoretical reference for stress gene mining and breeding of cultivated tomato.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The plant materials used in this study were tomato cultivars (*Solanum lycopersicum*, M82) from our laboratory. Tomatoes were grown in a 24 ± 2°C common greenhouse under a 16 h light/8 h dark photoperiod, and the relative humidity was 60–70%. Four-week-old seedlings were used for stress and hormone treatments. Salt stress was applied to seedlings treated with 150 mM sodium chloride (NaCl), and seedlings were transferred to a growth chamber at 40°C to simulate heat shock stress. Leaves were collected after 0, 2, 4 and 8 h for the stress treatments. Seedlings were sprayed with 100 μM IAA, 100 μM MeJA or 100 μM ABA, and tomato leaves were collected after 0, 6, 12 and 24 h. The isolated tissues were frozen in liquid nitrogen and then transferred to −80°C. Three different biological sample sources were collected for subsequent experiments in each process.

Identification of *SRO* Genes in Multiple Species

Complete genome sequences of grape and coffee were downloaded from the Ensemble Plants database (https://plants.ensembl.org/index.html). The reported amino acid sequences of Arabidopsis atrcd1 and *AtSRO*1-5 (Jaspers et al., 2010a) were downloaded from The Arabidopsis Information Resource (TAIR: https://www.arabidopsis.org/) (Rhee et al., 2003). The genomes of the major Solanaceae plants were downloaded from the Solanaceae genome database (https://solgenomics.net/), and *AtSROs* were used as query sequences for the whole genome

sequence BLASTP search in the Phytozome database (https://phytozome.jgi.doe.gov) to extract *SRO* members from various plants (Goodstein et al., 2012; Fernandez-Pozo et al., 2015). Similarly, BLASTP was used to search the local *Solanaceae* plant protein database (E-value: $1e^{-5}$) for *AtSROs* PFAM database (http://pfam.xfam.org/) was used to download Hidden Markov Models for RST (PF12174), PARP (PF00644) and WWE (PF02825) domains (Bateman et al., 2004; Sonnhammer et al., 1997). The canonical domains were used to Hmmsearch (Finn et al., 2011) from the local *Solanaceae* protein database with HMMER 3.0 (E-value: 1e-5). All candidate gene domains were analysed in smart (http://smart.embl.de/), CDD search (HTTPS://www.ncbi.NLM.NIH.Gov/CDD/) and Pfam (http://pfam.xfam.org/) databases (Sonnhammer et al., 1997; Schultz et al., 2000; Marchler-Bauer et al., 2007). The *SRO* genes in *Solanaceae* were obtained by deleting the genes without any typical SRO family domains and retaining a representative transcript of each gene (**Supplementary Data S1**).

The ExPASy online database ProtParam tool (http://www.expasy.org/protparam/) was used to predict and analyse the amino acid number (Artimo et al., 2012), isoelectric point, fat index and other physical and chemical properties of the tomato SRO protein. Protein subcellular localization was predicted by WoLF PSORT Online software (https://wolfpsort.hgc.jp/) (Horton et al., 2007).

Construction of Conserved Motifs, *Cis*-regulatory Elements and Phylogenetic Tree of *SRO* Genes in Tomato

Meme software (v4.12.0) was used to search tomato *SRO* motifs (Grundy et al., 1997); the number of searches was 20, the maximum and minimum widths were set to 6 and 50, respectively. Tbtools was used to draw conservative motifs and gene structure maps (Chen et al., 2020). According to the position information of the *SRO* gene on the chromosome, the karyotype map of tomato was drawn using mapchart. MEGA 7.0 software was used for multiple sequence alignment, and the maximum likelihood (ML) and neighbour joining methods were used to construct the phylogenetic tree with Poisson correction (Kumar et al., 2016). The bootstrap value was set to 2000. The Itools online website (https://itol.embl.de/) was used to display the midpoint rooted base tree. The promoter sequence of the *SRO*genes in tomato (2000 bp upstream of the translation start point) was extracted, and the *cis*-regulatory element (CRE) of the *SRO* genes was predicted through the Search for CARE tool in the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Rombauts et al., 1999) GSDS (http://gsds.gao-lab.org/) Online software was used to draw a distribution map of CREs (Hu et al., 2015).

Tomato *SRO* Family Homologous Genes, Interaction Network Andexpression Analysis

Perl scripts were used to extract the *SRO* gene position on the chromosome, and McscanX was used to extract the collinearity relationship between *SRO* genes (Wang et al., 2012). The substitution rate of paralogous genes was calculated by KaKs_Calculator2.0 (Wang et al., 2010), and Tbtools was used to draw the collinearity analysis map of orthologous genes of each species. The protein-protein interaction relationship was predicted by the STRING online website (https://string-db.org/), and the microRNA targeting relationship was predicted by psRNATarget (http://plantgrn.noble.org/psRNATarget/) with default parameters (Dai et al., 2018; Szklarczyk et al., 2019). The interaction network was displayed by Cytoscape software (Su et al., 2014).

The expression data of *SRO* genes in different tissues and developmental stages, inculding leaves, roots, flower buds, fully opened flowers, 1 cm fruits, 2 cm fruits, 3 cm fruits, mature green fruits, breaker fruits and breaker + 10 days fruits, were retrieved from the Tomato Functional Genomics database (TFGD, http://ted.bti.cornell.edu/) (Fei et al., 2011). Seedings of M82 (salt-sensitive) and *S. pennellii* (elite salt-resistant) were exposed to salt stress (200 mM NaCl, Irrigation) after 6 weeks of normal growth, 0 and 12 h tomato roots were used for RNA-seq in illumina Hiseq 2500 platform. The expression level were normalized by Transcripts Per Million (TPM). The R package DESeq2 (Love et al., 2014) was then used to calculate the Fold Change (FC). All *SRO* genes expression profiles were analyzed and performed using software Tbtools. The raw data were deposited in the Genome Sequence Archive (GSA) of the China National Center for Bioinformation under accession number: PRJCA005251 (unpublished).

Ribo Nucleic Acid Extraction and Reverse Transcription Polymerase Chain Reaction Analysis

Total RNA was extracted using TRIzol reagent (Aidlab Biotechnologies, Beijing, China). First-strand cDNA was synthesized using a HiScript II 1st Strand cDNA Synthesis Kit (+gDNA wiper) (Vazyme, China). Gene-specific primers were designed using Primer Premier 5.0 (**Supplementary Table S1**), and the primers for these genes were synthesized by Sangon Biotech Co., Ltd. (Shanghai, China). Then, quantitative PCR (qPCR) was performed using Maxima SYBR Green/ROX qPCR Master Mix. The EF-1α gene was used as an internal reference. Each treatment contained three independent biological replicates, and each replicate contained three technical replicates. Gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

RESULT

Identification of the *SRO* Genes in Tomato

In this study, we used *Arabidopsis thaliana* amino acid sequences (*AtSROs*) for BLASTP and HMM searches (RST, PARP and WWE) to screen *SRO* members with at least one conserved domain in the genomes of multiple tomatoes and named them according to their positions on chromosomes. The cultivated tomatoes contained 6 *SRO* genes. The number of *SRO* family

TABLE 1 | Basic information of *SRO* genes identified in tomato.

Species	Gene id	Gene name	Chr	Length	MW (Da)	pI	Aliphatic index	GRAVY	Subcellular localization
S. lycopersicum	Solyc03g114360	*SolySRO1*	Soly-3	375	41317.55	8.61	69.15	−0.40	nucl
	Solyc05g005280	*SolySRO2*	Soly-5	304	34189.18	7.65	84.31	−0.28	cyto
	Solyc05g005290	*SolySRO3*	Soly-5	233	26286.90	6.71	62.93	−0.25	cyto
	Solyc06g066330	*SolySRO4*	Soly-6	595	67309.50	7.67	85.61	−0.43	nucl
	Solyc08g005270	*SolySRO5*	Soly-8	600	67858.03	6.39	73.70	−0.45	nucl
	Solyc08g076420	*SolySRO6*	Soly-8	598	67307.56	7.20	81.09	−0.44	nucl
S. lycopersicum var. *cerasiforme*	SLYcer01g04782	*SlycSRO1*	Slyc-1	300	34191.15	7.70	83.43	−0.39	nucl
	SLYcer01g04783	*SlycSRO2*	Slyc-1	442	50504.70	6.44	82.42	−0.37	cyto
	SLYcer03g04627	*SlycSRO3*	Slyc-3	376	41417.66	8.60	68.96	−0.41	nucl
	SLYcer04g05316	*SlycSRO4*	Slyc-4	680	77592.90	5.82	85.94	−0.36	nucl
	SLYcer04g05317	*SlycSRO5*	Slyc-4	483	55418.48	6.36	84.72	−0.43	cyto
	SLYcer04g05318	*SlycSRO6*	Slyc-4	443	49038.35	6.54	84.83	−0.18	chlo
	SLYcer05g00116	*SlycSRO7*	Slyc-5	315	35392.43	6.60	83.52	−0.31	cyto
	SLYcer05g00117	*SlycSRO8*	Slyc-5	320	36137.59	8.91	85.59	−0.29	cyto
	SLYcer06g04350	*SlycSRO9*	Slyc-6	594	67352.60	7.67	85.27	−0.44	nucl
	SLYcer08g00172	*SlycSRO10*	Slyc-8	600	67888.60	6.81	73.70	−0.45	nucl
	SLYcer08g05857	*SlycSRO11*	Slyc-8	507	57279.60	6.70	80.89	−0.48	nucl
S. chilense	SOLCI001453300	*SolcSRO1*	—	233	26254.13	7.08	89.14	−0.23	cyto
	SOLCI001453400	*SolcSRO2*	—	315	35419.46	6.60	83.84	−0.32	cyto
	SOLCI001464200	*SolcSRO3*	—	594	67595.43	7.54	86.50	−0.44	nucl
	SOLCI003930500	*SolcSRO4*	—	600	67793.00	6.90	76.63	−0.41	nucl
	SOLCI004134700	*SolcSRO5*	—	597	67250.52	6.69	81.41	−0.43	nucl
	SOLCI005404200	*SolcSRO6*	—	310	34869.80	6.21	85.45	−0.30	nucl
	SOLCI005589700	*SolcSRO7*	—	375	41302.52	8.63	69.50	−0.41	nucl
S. *pimpinellifolium*	SPI01g04931	*SpiSRO1*	Spi-1	442	50525.01	6.33	84.19	−0.37	cyto
	SPI03g04680	*SpiSRO2*	Spi-3	376	41417.00	8.70	68.96	−0.41	nucl
	SPI04g04808	*SpiSRO3*	Spi-4	679	77360.60	5.82	86.49	−0.34	nucl
	SPI04g04809	*SpiSRO4*	Spi-4	483	55194.90	6.83	83.91	−0.41	cyto
	SPI04g04810	*SpiSRO5*	Spi-4	443	49197.44	6.20	85.28	−0.20	chlo
	SPI05g00126	*SpiSRO6*	Spi-5	315	35392.39	6.41	83.84	−0.31	cyto
	SPI05g00127	*SpiSRO7*	Spi-5	320	36246.71	8.81	85.59	−0.31	cyto
	SPI06g04210	*SpiSRO8*	Spi-6	594	67352.60	7.67	85.70	−0.44	nucl
	SPI08g00091	*SpiSRO9*	Spi-8	600	67856.50	6.81	74.80	−0.44	nucl
	SPI08g05744	*SpiSRO10*	Spi-8	507	57285.60	6.51	81.60	−0.47	nucl
S. *pennellii*	Sopen03g033460	*SpenSRO1*	Spen-3	375	41250.55	8.73	70.96	−0.39	nucl
	Sopen04g030720	*SpenSRO2*	Spen-4	595	66837.97	5.74	81.98	−0.39	nucl
	Sopen04g030730	*SpenSRO3*	Spen-4	455	50100.49	5.94	88.79	−0.12	chlo
	Sopen05g001280	*SpenSRO4*	Spen-5	315	35408.45	6.51	83.84	−0.28	cyto
	Sopen05g001300	*SpenSRO5*	Spen-5	217	24569.40	9.58	92.53	−0.19	cyto
	Sopen06g021690	*SpenSRO6*	Spen-6	594	67113.79	7.56	85.44	−0.42	nucl
	Sopen08g001290	*SpenSRO7*	Spen-8	595	67457.00	7.09	75.60	−0.44	nucl
	Sopen08g025000	*SpenSRO8*	Spen-8	597	67299.00	7.62	80.75	−0.44	nucl
S. *lycopersicoides*	Solyd03g075660	*SlydSRO1*	Slyd-3	352	38871.74	8.79	67.87	−0.42	nucl
	Solyd05g050320	*SlydSRO2*	Slyd-5	376	41703.19	5.57	75.48	−0.30	nucl
	Solyd06g065810	*SlydSRO3*	Slyd-6	594	67162.05	8.50	85.62	−0.40	nucl
	Solyd08g050330	*SlydSRO4*	Slyd-8	539	60715.62	5.72	73.75	−0.42	nucl
	Solyd08g050340	*SlydSRO5*	Slyd-8	600	67979.41	6.50	74.70	−0.42	nucl
	Solyd08g068000	*SlydSRO6*	Slyd-8	597	67197.55	6.50	82.06	−0.39	nucl

genes in the wild tomatoes was 7–11. Analysis of protein physicochemical properties showed that the length of the *SRO* family amino acids in all tomatoes ranged from 217 (*SpenSRO5*) to 680 (*SlycSRO4*), the molecular weight ranged from 24569.40 (*SpenSRO5*) to 77592.90 (*SlycSRO4*), the pI ranged from 5.57 (*SlydSRO2*) to 9.58 (*SpenSRO5*), the aliphatic index of the SRO protein ranged from 62.93 (*SolySRO3*) to 92.53 (*SpenSRO5*), and the GRAVY value ranged from −0.12 to −0.48 (**Table 1**). Chromosome localization (**Supplementary Figure S1**) showed that the *SRO* family in tomato is distributed in 7 regions on 6 chromosomes. The *SRO* genes on Chr1 and Chr4 in cultivated tomato were lost. Subcellular localization showed that the *SRO* genes on Chr1, Chr4, and Chr5 were distributed in the cytoplasm and chloroplast, and the rest of the *SRO* was located in the nucleus (**Table 1**).

Phylogenetic Analysis of *SRO* Genes in Various Plants

The Arabidopsis Information Resource (TAIR), PlantGDB, Phytozome, and National Center for Biotechnology Information (NCBI) databases were used to retrieve reliable

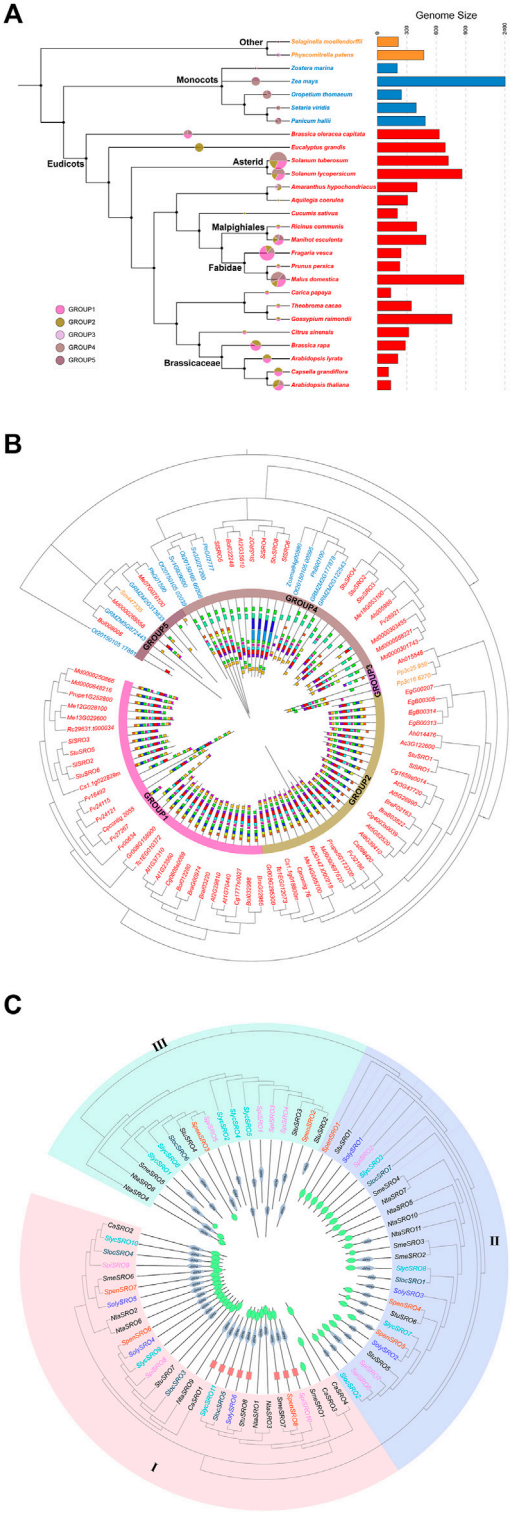

FIGURE 1 | Comparison of the SRO family in plants. **(A)** Evolutionary relationships and number of SRO families in multiple species, with species colours representing their Taxonomic characteristics and the size and colour of the sectors representing the number of SRO family in the species and

(*Continued*)

SRO sequences (Li et al., 2019), and 93 *SRO* potential homologous genes were retrieved from 27 plants (20 Eudicots, 5 Monocots, 1 Bryophyta, and 1 Tracheophyta). The *SRO* gene family of plants evolved continuously with the evolution of the complexity of life (**Figure 1A**). There were obvious taxonomic differences among Bryophytes, Tracheophytes, Monocots and Eudicots, but the expansion of the *SRO* family was relatively conservative, although the number of *SRO* genes in Asterid, Fabidae and Brassicaceae was significantly higher than that in *Physcomitrella patens* and *Selaginella moellendorffii*. However, the *SRO* families in some higher plants seemed to be under more selection pressure, and the number of genes was reduced. The amino acid sequences of 93 *SRO* homologous genes were used to construct the evolutionary tree by the neighbour joining method (**Figure 1B**). All SRO families were divided into 5 groups, among which GROUP1 and GROUP2 contained only Eudicots, GROUP3 contained only *P. patens* and *A. hypochondriacus*, and GROUP4 and GROUP5 contained both Eudicot and Monocot plants. Conserved motif analysis showed that each group exhibited higher similarities. Eudicots accumulated more subfamily types than Monocots. These results indicated that the expansion of SRO family members coincided with whole-genome duplication (WGD) during plant evolution.

To further analyse the lineage-specific amplification of the *SRO* family, we identified the *SRO* family in *Solanaceae* (*Capsicum annuum*, *Solanum melongena*, *Solanum tuberosum*, *Nicotiana tabacum*, *Solanum lycopersicum*, and *Lycopersicon*) and constructed a phylogenetic tree (**Figure 1C**). All *SRO* genes were divided into three classes. There were large differences in coding sequence (CDS) length and domain among them. The genes in class I showed the longest gene length and contained both PARP and RST domains, with the exception of *SmeSRO1*, *CaSRO3* and *CaSRO4*. Some class I genes also contained the WWE domain. The length of genes in class II was the shortest, and some of them only contained the RST domain. Among the class III genes, *NitaSRO4*, *NitaSRO8* and *SmeSRO5* contained a small RST domain at the N-terminus, and the other genes only contained the PARP domain. *S. lycopersicum* in particular was lost in class III. In fact, these *SRO* genes were mainly located on Chr1 and Chr4 of their respective species, which was consistent with the results of the chromosome localization map. We noticed that the four *SRO* genes (*CaSRO1* ~ *4*) identified in *C. annuum* were all classified in region I of the phylogenetic tree and were lost in particular in class II and class III. The domains of *CaSRO3* and *CaSRO4* were different from the others in class I.

FIGURE 1 | the subgroups to which they belong. **(B)** Phylogenetic trees were constructed for 93 *SRO* genes using the NJ method. Different colors represent species with different taxonomic characteristics. Gene structure and conserved motif were peformed inside the phylogenetic tree. **(C)** Phylogenetic tree of the *SRO* family in Solanaceae. The phylogenetic tree was constructed using the NJ method. The different coloured *SRO* genes were derived from different tomato species, and the conserved structural domains of the corresponding *SRO* genes are shown inside the evolutionary tree, with the WWE structural domain in red, the RST structural domain in green, and the **PARP** structural domain in grey.

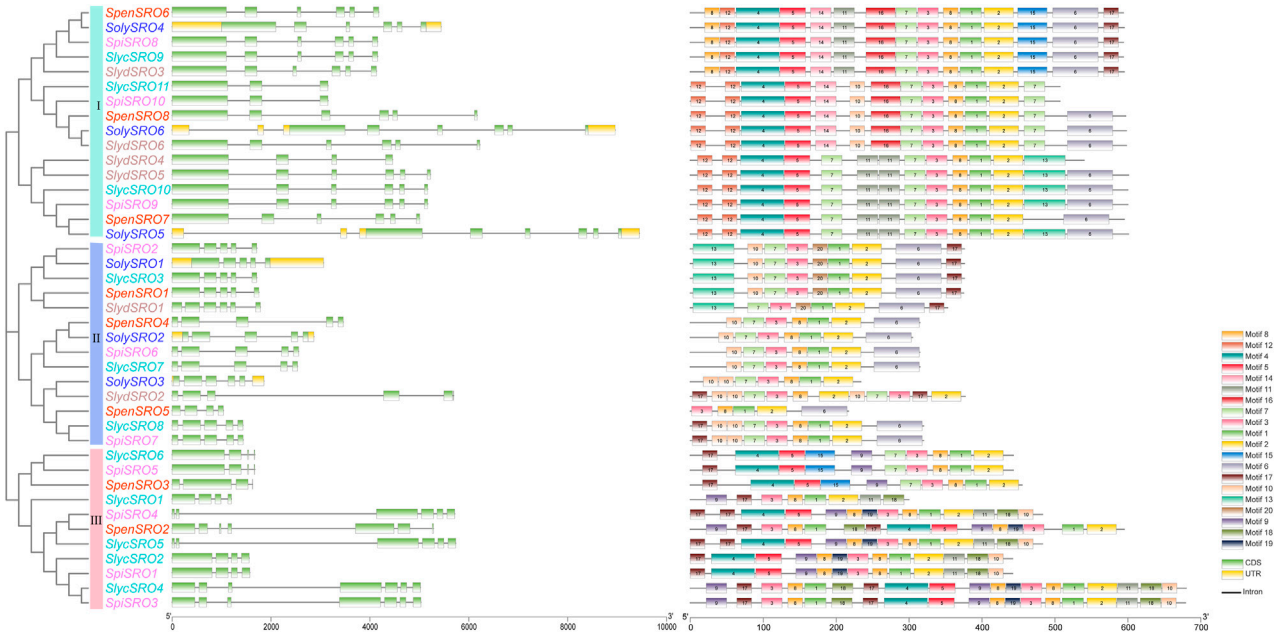

FIGURE 2 | Phylogenetic relationships, Structure and conserved motifs of *SRO* genes in tomato species. The different coloured *SRO* genes were derived from different tomato species. Green boxes indicate exons, yellow boxes indicate UTR and black lines indicate introns. The numbers 1–20 and the different colored boxes indicate motifs.

Structure and Conserved Motif Analysis of *SRO* Genes in Tomato

Exon-intron structural differences are important sources of gene family variation and plant biodiversity. Different structures determine the differential function and expression of genes (Xu et al., 2012). Except for *S. chilense,* whose *SRO* genes were not assembled on the chromosome, we extracted all *SRO* gene annotations from the whole genomes of cultivated tomato and multiple wild tomatoes. The comparison results of the positions and quantity of exons were visualized with TBtools (**Figure 2**). The results of the phylogenetic tree showed that all *SRO* genes were divided into three groups, among which, in group I, *SlycSRO11* and *SpiSRO10* contained three exons, *SlydSRO4* contained 4 exons, and the rest of the *SRO* genes contained 6 exons. In group II, *SlydSRO1* contained 6 exons, *SpenSRO5* contained 4 exons, and all other *SRO* genes contained 5 exons. Obviously, the length and structure of the *SRO* genes in groups I and II were relatively consistent. We noticed that even with the same number of exons, *SRO* genes of cultivated tomato in groups I and II still exhibited more introns and longer gene lengths than those of wild tomatoes. Group III, which contained only wild tomato, showed more structural diversity. The first *SRO* genes (*SpenSRO2*, *SlycSRO4*, and *SpiSRO3*) on Chr4 of all wild tomatoes showed higher structural similarity; they had 7 exons and almost the same gene length. The *SRO* genes (*SlycSRO1* and *SpiSRO1*) on Chr1 and the third *SRO* genes (*SlycSRO2*, *SlycSRO6*, *SpiSRO5*, and *SpenSRO3*) on Chr4 showed similar regularity. They had the same length and four exons. Only *S. lycopersicum* var. *cerasiforme* and *S. pimpinellifolium* contained the second *SRO* genes (*SlycSRO5* and *SpiSRO4*) on Chr4, and they had seven exons with the same distribution.

The conserved motifs of all *SRO* genes were predicted based on MEME software, and a total of 20 conserved motifs were identified (**Supplementary Figure S2**). Motif6 and motif8 are the RST and PARP domains, respectively, and they were distributed in all *SRO* genes. Similar to the exon-intron structure, the conserved motifs were also divided into three groups based on genetic relationships. The motif composition of the *SRO* gene in the same groups was similar. Group I contained the largest number of motifs, with a total of 16 motifs. Motif12, motif14 and motif16 only appeared in this group. Group II contained 11 motifs, including motif20, and group III contained 14 motifs, including motif9, motif18 and motif19. The *SRO* genes of cultivated tomato also only appeared in groups I and II, and each *SolySROs* was always genetically close to one or more *SROs* in wild tomatoes. The *SRO* gene motifs on the same branch cluster were highly similar in both cultivated and wild tomatoes, indicating that there were no significant differences in the sequence and function of *SRO* genes in tomato species, with the exception of group III.

Promoter Analysis of *SRO* Genes in Tomato

CRE control gene expression by combining with specific transcription factors, and the distribution of CREs in the promoter region is closely related to gene function (Biłas et al., 2016). We predicted the CRE in the 2000 bp sequence upstream of all *SRO* genes through the Plantcare online website (**Figure 3**, **Supplementary Table S2**). In addition to the core promoter and

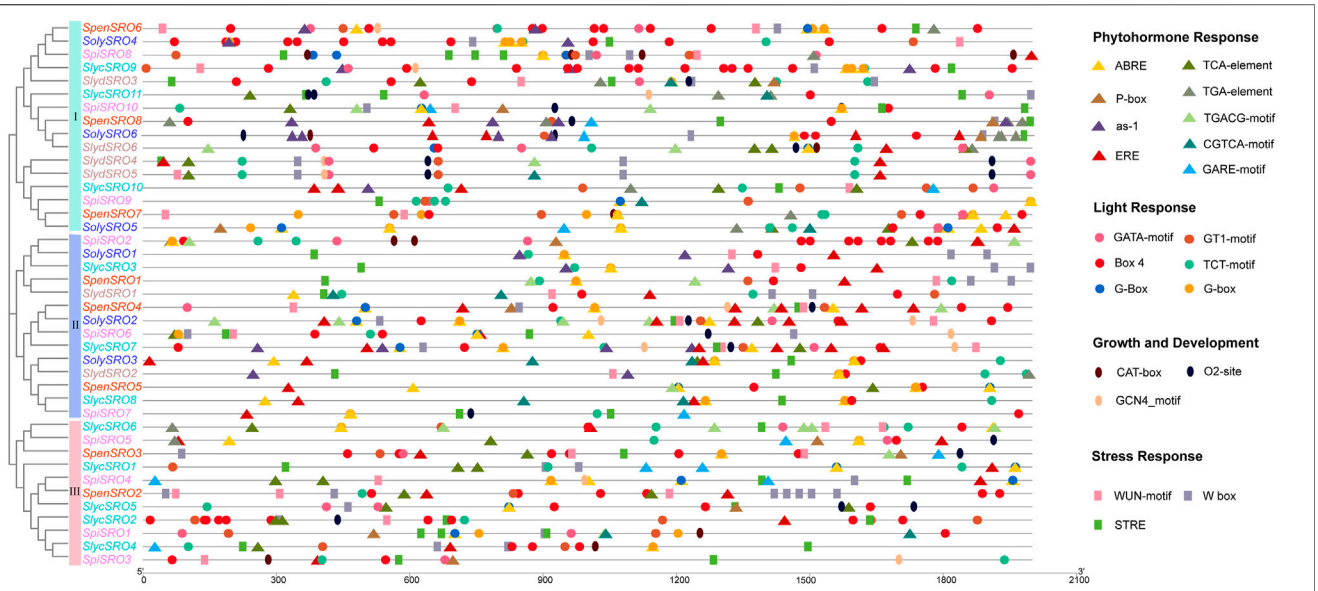

FIGURE 3 | Distribution of CREs of *SRO* genes in tomato. The different coloured *SRO* genes were derived from different tomato species. Different CREs were indicated by different shapes, triangles indicate hormone response elements, circles indicate light response elements, ellipses indicate growth and development related elements, boxes indicate stress response related elements, and different elements were indicateed by different colors.

enhancer, the promoter region of *SRO* genes in tomato contained a large number of plant hormone response elements. A total of 543 plant hormone response elements were identified and divided into 20 species, including 190 abscisic acid response elements of 5 types, 112 salicylic acid response elements of 4 types, 34 gibberellin response elements of 3 types, 33 auxin response elements of 3 types, 118 methyl jasmonate response elements of 2 types, 56 ethylene response elements of one type. Two auxin response elements (AuxRR-core, E2Fb) and one salicylic acid response element (SARE) were only specifically recognized in wild tomato. There were the mosttypes of light-responsive elements. Among all *SRO* genes, 23 types of light-responsive elements were identified, a total of 452, mainly including 102 conserved DNA modules involved in the Box4 light response, 73 light-induced stem- and leaf-specific expression promoter G-boxes, 48 photosynthetic element TCT motifs induced by sunlight time, and 45 photosyntheticelement GT1 motifs. Of these elements, 7 types of light-responsive elements (AAAC motif, AT1 motif, ATCT motif, chs-CMA2a, gap box, LAMP element and Sp1) were only specifically identified in wild tomatoes. Nine types of biotic/abiotic stress response elements were identified, for a total of 285, including 84 anaerobic inducing elements (AREs), 45 drought response elements (W-boxes), 48 high temperature response elements (STREs), and 37 wound inducing elements (WUN motifs). Sixty-seven growth and development response elements were also identified in all *SRO* genes, divided into 8 types, including 12 CAT boxes related to meristem expression, 13 GCN4 motifs related to endosperm expression, and 23 O2 sites participating in zein metabolism regulation. Among them, four growth and development response elements (AACA motif, CCGTCC box, HD-Zip 1, and MSA-like) were lost in cultivated tomato. At the same time, *SRO* genes in tomatoes also contained a large number of other regulatory elements. These results indicated that *SRO* genes were widely involved in various life activities, such as plant growth and development and stress responses.

Duplication Gene and Ka/Ks Analysis of *SRO* Genes in Tomato

Gene replication is an effective way for organisms to obtain new genes and maintain gene vitality (Zhang, 2003). Local blast and mcscanx software were used to extract the repeat sequences of the *SRO* gene in all tomato genomes, and the replacement rate of *SRO* homologous gene pairs was calculated using KaKs Calculator 2.0 (**Table 2**). The results showed two paralogous gene pairs in the *SRO* family of cultivated tomato, namely, *SolySRO2*/*SolySRO3* and *SolySRO4*/*SolySRO6*, and all were derived from segmental replication. *S. pennellii*, *S. Chilense* and *S. lycopersicoides* also contained two pairs of paralogous genes, while *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* contained 6 and 11 pairs of *SRO* paralogous gene pairs, respectively, which mainly came from multiple repeat pairs of the *SRO* gene on Chr1 (Slyc1, Slyc2, and Spi1) and Chr 4 (Slyc4, Slyc5, Slyc6, Spi3, and Spi4). In all wild tomatoes, *SlycSRO1*/*SlycSRO2*, *SlycSRO7*/*SlycSRO8*, *SpiSRO3*/*SpiSRO4* and *SpiSRO6*/*SpiSRO7* paralogue gene pairs were derived from chromosome tandem replication, and the rest of the repeat gene pairs were derived from segmental replication. The *Ka*/*Ks* of the two homologous gene pairs in cultivated tomato were both <1, indicating that the two pairs of paralogous genes had received strong environmental pressure, and the gene evolution and protein function had stabilized. There were still 9 pairs of paralogous genes *Ka*/*Ks* greater than 1 in wild tomatoes. These *SRO* genes were subjected to positive environmental

TABLE 2 | The *Ka/Ks* ratios and date of duplication for duplicate *SRO* genes in tomato.

Species	Chr	Duplicated gene pairs	Ka	Ks	Ka/Ks	Selective pressure	Type	Time (Mya[a])
S. lycopersicum	Soly5/Soly5	*SolySRO2/SolySRO3*	0.13	0.39	0.32	Purify selection	segmental	12.99
	Soly6/Soly8	*SolySRO4/SolySRO6*	0.97	1.16	0.83	Purify selection	segmental	38.68
S. pennellii	Spen4/Spen4	*SpenSRO2/SpenSRO3*	1.06	0.79	1.33	Purify selection	segmental	26.48
	Spen5/Spen5	*SpenSRO4/SpenSRO5*	0.94	1.36	0.69	Purify selection	segmental	45.33
S. chilense	—	*SolcSRO1/SolcSRO2*	0.93	1.36	0.69	Purify selection	segmental	45.21
	—	*SolcSRO5/SolcSRO6*	0.99	1.05	0.94	Purify selection	segmental	35.05
S. lycopersicum var. *cerasiforme*	Slyc1/Slyc1	*SlycSRO1/SlycSRO2*	1.01	0.96	1.05	Positive selection	tandem	32.05
	Slyc1/Slyc4	*SlycSRO1/SlycSRO4*	0.98	1.09	0.90	Purify selection	segmental	36.40
	Slyc1/Slyc4	*SlycSRO1/SlycSRO5*	0.98	1.08	0.90	Purify selection	segmental	36.15
	Slyc1/Slyc4	*SlycSRO1/SlycSRO6*	1.05	0.83	1.27	Positive selection	segmental	27.55
	Slyc1/Slyc4	*SlycSRO2/SlycSRO4*	1.01	0.96	1.05	Positive selection	segmental	32.12
	Slyc1/Slyc4	*SlycSRO2/SlycSRO5*	1.00	1.01	0.99	Purify selection	segmental	33.72
	Slyc1/Slyc4	*SlycSRO2/SlycSRO6*	1.02	0.90	1.14	Positive selection	segmental	30.02
	Slyc4/Slyc4	*SlycSRO4/SlycSRO6*	1.01	0.97	1.05	Positive selection	segmental	32.19
	Slyc4/Slyc4	*SlycSRO5/SlycSRO6*	0.99	1.03	0.96	Purify selection	segmental	34.49
	Slyc4/Slyc6	*SlycSRO5/SlycSRO9*	0.96	1.14	0.84	Purify selection	segmental	37.99
	Slyc5/Slyc6	*SlycSRO7/SlycSRO8*	0.12	0.45	0.27	Purify selection	tandem	14.86
S. pimpinellifolium	Spi1/Spi4	*SpiSRO1/SpiSRO3*	1.02	0.91	1.12	Positive selection	segmental	30.39
	Spi1/Spi4	*SpiSRO1/SpiSRO4*	1.01	0.97	1.03	Positive selection	segmental	32.48
	Spi4/Spi4	*SpiSRO3/SpiSRO4*	0.96	1.16	0.83	Purify selection	tandem	38.52
	Spi4/Spi4	*SpiSRO3/SpiSRO5*	0.95	1.19	0.80	Purify selection	segmental	39.68
	Spi5/Spi5	*SpiSRO6/SpiSRO7*	0.15	0.17	0.87	Purify selection	tandem	5.62
	Spi6/Spi8	*SpiSRO8/SpiSRO10*	0.99	1.06	0.93	Purify selection	segmental	35.30
S. lycopersicoides	Slyd5/Slyd8	solydSRO2/solydSRO5	1.00	0.99	1.01	Positive selection	segmental	33.05
	Slyd6/Slyd8	solydSRO3/solydSRO4	0.98	1.09	0.90	Purify selection	segmental	36.21

[a]*Millions years ago.*

selection and were still in the rapid evolutionary stage. According to the differentiation rate R (1.5×10^{-8}) of Solanaceae (Blanc and Wolfe, 2004), the differentiation time of all gene pairs was estimated. The duplication time of the *SRO* paralogous gene pairs in tomato was more dispersed, ranging from 5.62 to 45.33 Mya. Duplication of the *SolySRO2*/*SolySRO3* fragment on cultivated tomato chromosome 5 occurred at approximately 12.99 Mya. However, the homologous gene pairs of *SpenSRO4*/*SpenSRO5* and *SpiSRO6*/*SpiSRO7*, which were also distributed on chr5, replicated at 45.33 and 5.62 Mya in segmental and tandem manners, respectively. Duplication of the *SolySRO4*/*SolySRO6* homologous gene pair occurred at approximately 38.68 Mya, and duplication of homologous genes in the same region in *S. pimpinellifolium* and *S. lycopersicoides* occurred at 35.30 and 36.21 Mya, respectively. These three homologous gene pairs were relatively close in duplication time, while their *Ka*/*Ks* values converged to 1, which could mean that the *SRO* genes of tomatoes on Chr6 and Chr8 occurred early after whole genome duplication in Solanaceae, and these genes belong to the conserved members of the *SRO* family.

Evolutionary and Collinearity Analysis of *SRO* Genes in Tomato

To trace the evolutionary origin and orthologous relationship of the *SRO* genes in tomatoes, we used grape (*Vitis vinifera*. L) and coffee (*Coffea canephora*), which did not undergo a new specific genome-wide doubling event after a "gamma" whole-genome triplication event that was common to most ancient ancestors of eudicot plants (Wang et al., 2016). At the same time, according to the time of *Solanaceae* differentiation, the *SRO* genes were analysed for interspecies collinearity (**Figure 4A**, **Supplementary Table S3**). The SRO family expanded with the whole genome replication of angiosperms. Grape and coffee, which represent the ancient ancestors, each had three *SRO* genes, which were highly conserved in the evolutionary process and homologous with several *SRO* genes in *Solanaceae*. This indicated that the SRO family in plants may be copied from one *SRO* gene in the ancestral species after the γ event. Starting from tobacco, the number of homologous members of the SRO family increased to five, and the evolutionary speed was accelerated. The SRO family in tomato was divided into subfamilies I–III, in which subfamilies I and II each contained only one *SRO* gene on Chr3 and Chr4, respectively. Subfamily III contained three *SRO* genes on Chr6 and Chr8, and *SRO* genes in all subfamilies were highly homologous in both cultivated and wild tomatoes.

To further discover the origin of the SRO family in tomatoes, we extracted the collinearity of the SRO family in *Solanaceae* (*C. annuum*, *S. tuberosum*, and *S. melongena*), V. *vinifera*. L and various tomatoes (*S. lycopersicoides*, *S. pennellii*, and *S. lycopersicum*) (**Figures 4B,C**). We realized that the *SRO* genes in tomatoes actually showed five orthologous patterns based on the position of its chromosome. *SolySRO1*, located on Chr3, had

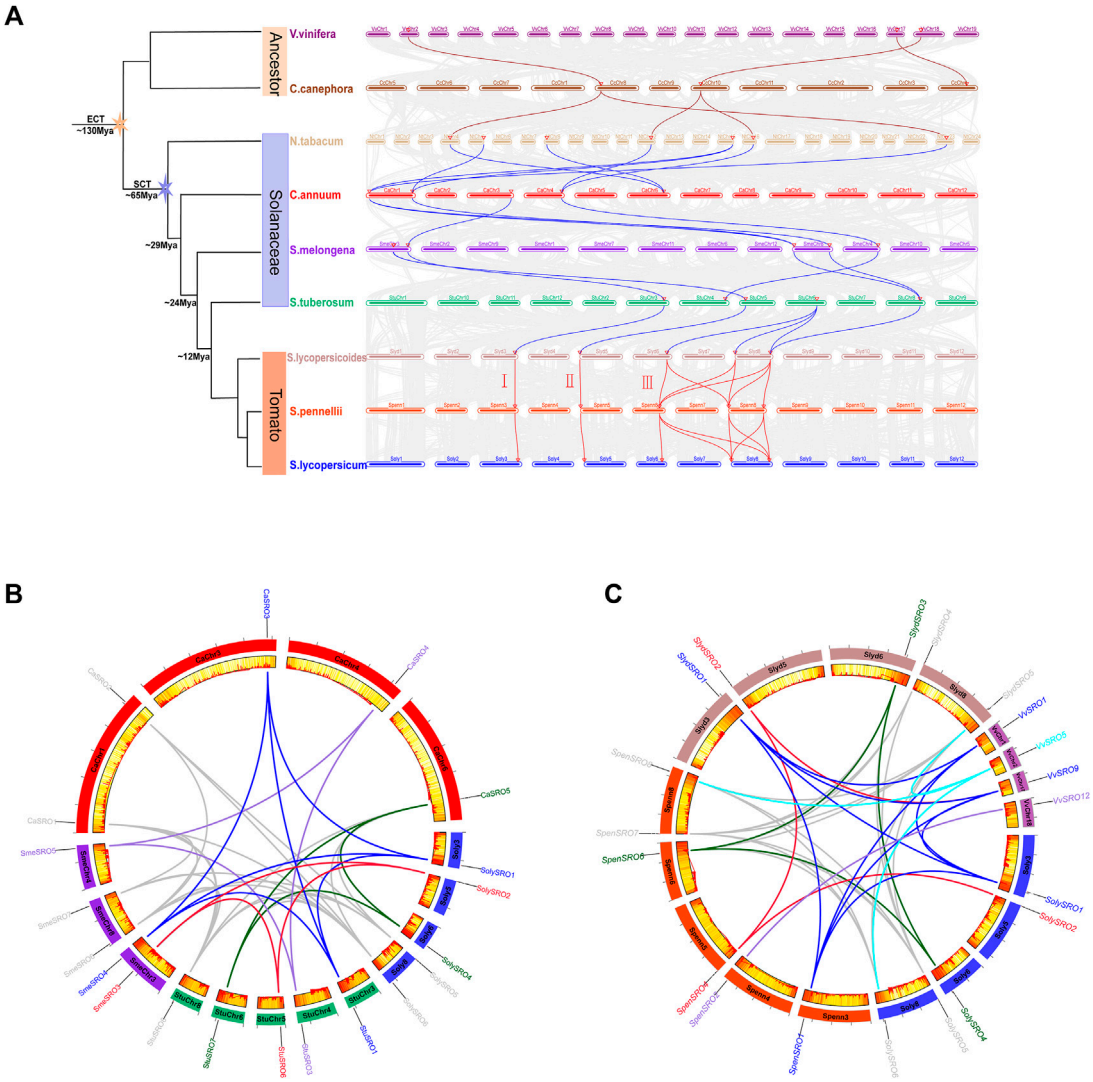

FIGURE 4 | Homologous genes and evolutionary analysis of the SRO family. **(A)** Co-lineage map for multiple species, with species genomes arranged in evolutionary order and coloured lines representing *SRO* genes with direct homologous relationships within each species. **(B)** Co-lineage map of *SRO* genes within *Solanaceae* (*C. annuum*, *S. melongena*, *S. tuberosum*, and *S. lycopersicum*), with the outer circle showing the chromosomes of each *Solanaceae*, the inner circle showing gene density, the ends of the lines representing direct homologous *SRO* genes, and the different coloured lines representing different evolutionary patterns. **(C)** Covariance of *SRO* genes within *V. vinifera*. L, *S. lycopersicoides*, *S. pennellii*, and *S. lycopersicum*, with the outer circle showing the chromosomes of each species, the inner circle showing the gene density, the two ends of the lines representing the direct homologous *SRO* genes, and different coloured lines representing different evolutionary patterns.

one orthologous gene in all *Solanaceae* and two orthologous genes (*VvSRO1* and *VvSRO9*) in grape. *SolySRO2*, located on Chr5, had one homologous gene with all other species and only lacked a homologous relationship in *C. annuum*. This gene was also derived from *VvSRO9* and maintained a certain degree of conservation during evolution. *SolySRO4*, located on Chr6, has one homologous gene in all *Solanaceae*, with the exception of *S. melongena*, and there is no homologous *SRO* member in grape. The *SolySRO5* and *SolySRO6* gene pairs located on Chr8 had highly homologous *SRO* genes in all species. Interestingly, there was only one *VvSRO5* homologous gene in grape. We also noticed that *SpenSRO2* on Chr4 in *S. pennellii* is highly homologous to *VvSRO12*. Chr4 in several *Solanaceae* also contained homologous *SRO* genes, which were lost in cultivated tomato.

Interaction Between Protein and microRNA of *SRO* Genes in Tomato

To better understand the function of *SRO* genes in tomatoes, we predicted the interactions between all SolySROs proteins based on the STRING online database. SolySRO4 and SolySRO6 had no predicted interactions with any protein. There was no direct interaction between SolySRO1, SolySRO2, SolySRO3, and SolySRO5, but they cooperated with other proteins to regulate

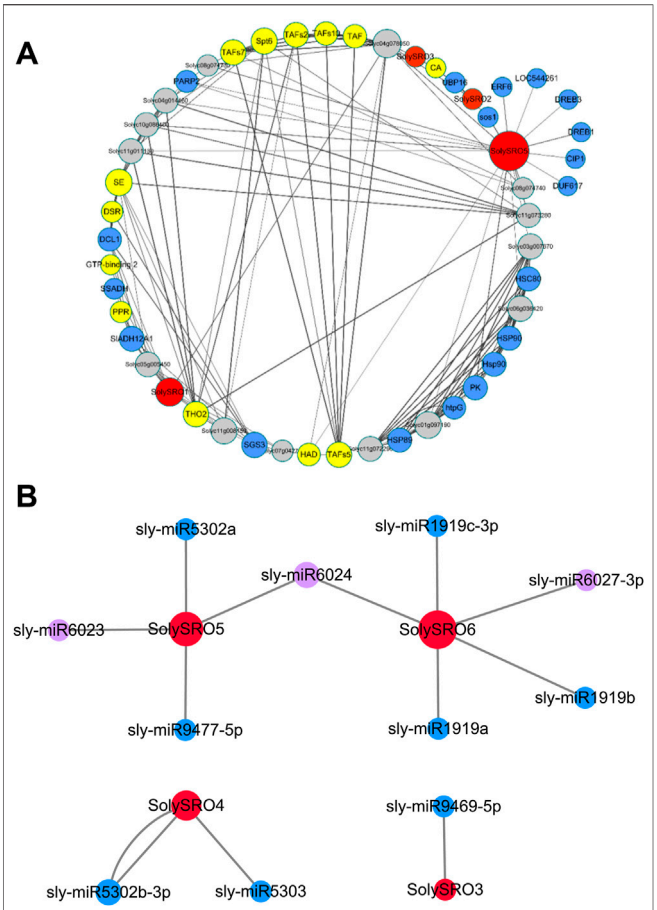

FIGURE 5 | Inter-crossing network diagram of *SRO* genes in tomatoes. **(A)** Interaction network of the SRO family with other proteins. Each node is a protein, each edge represents the presence of interactions, the size of the node represents the number of interactions, the thickness of the edge represents the value of the combined score, red nodes represent SRO proteins, blue nodes represent stress-related proteins, yellow nodes are transcription factors, and grey nodes represent proteins lacking annotation. **(B)** *SRO* genes and micoRNA targeting interactions. The red nodes are *SRO* genes, the size of the node represents the number of interactions, the blue nodes represent abiotic stress-related microRNAs, and the purple nodes are biotic stress-related microRNAs.

similar physiological functions and produced a total of 218 branches (**Supplementary Table S4**). SolySRO5 interacted with the most proteins with 31, and SolySRO2 and SolySRO3 each interacted with only three proteins. We excluded some proteins with lost annotations and low degree values and drew an interaction network diagram (**Figure 5A**). The results showed that proteins interacting with the SolySRO family could be divided into three categories. The number of proteins related to environmental stress response was the largest, including the protein families SLADH, SSADH, and LOC that regulate the balance of ROS products, the HSP, SOS, UBP, etc., which promote plant adaptation to low temperature and participate in plant salt and drought stress tolerance, the protein families that enhance plant biotic stress resistance, SGS and DCL, and the DREB, ERF, and AP2, etc., which are regulated and responded to by hormones. SolySROs also interacted with a large number of transcription factor protein families, including SPT, TAF, and DSR that regulate the transcription process, which may be related to their expression patterns under special circumstances. There were also some proteins with missing annotations in the interaction network diagram. They had a clear direct or indirect synergy with SolySRO proteins, but their functions were still unclear.

MicroRNAs have target regulatory relationships with *SolySROs* were predicted in the psRNATarget database (**Figure 5B**, **Supplementary Table S5**). Only four genes, *SolySRO3*, *SolySRO4*, *SolySRO5* and *SolySRO6*, were predicted to have a targeted regulatory relationship. *SolySRO6* was targeted by five microRNAs, with the greatest regulation. *SolySRO5* and *SolySRO4* were targeted by four and two microRNAs, respectively, and *SolySRO3* was only regulated by microRNA9469. Almost all microRNAs targeted a single *SolySRO* gene, with only micoRNA6024 targeting and regulating the *SolySRO5* and *SolySRO6* genes at the same time, and micoRNA5302 bound two specific target sites of *SolySRO4*. The above results of the protein interaction network and microRNA targeting regulation provided more possibilities for functional research on *SolySROs*.

Expression Profile Analysis of *SRO* Genes in Tomato

The published RNA-seq data were used to study the expression pattern of *SolySROs*. The results of *the SRO* genes expression profile in different tomato tissues showed that all *SolySROs* members exhibited strong tissue-specific expression, and they were obviously divided into two groups by expression level (**Figure 6A**, **Supplementary Table S6**). *SolySRO5* and *SolySRO6* had higher expression levels in all tomato tissues. The expression level of *SolySRO5* was the highest in fruit (3 cm), and this value of *SolySRO6* appeared in mature fruits, which indicated that these two *SRO* genes were the core genes of the *SRO* family and were highly expressed in fruit development and ripening. *SolySRO1*, *SolySRO2*, *SolySRO3*, and *SolySRO4* were all expressed at low levels in different tissues and were only highly expressed at secific periods. The expression level of *SolySRO2* was highest in flowers. The expression of *SolySRO3* in roots was higher than that in other tissues, while the maximum expression of *SolySRO4* and *SolySRO1* appeared in mature fruits. In addition, based on the RNA-seq data, the expression patterns of *SolySROs* in cultivated tomatoes (M82) and wild tomato (*S. pennellii*) under salt stress were studied (**Figure 6B**, **Supplementary Table S6**). The SRO family of cultivated and wild tomatoes exhibited the same expression patterns under a high salt environment. Compared to the control, *SolySRO1* was significantly down-regulated in both M82 ($\log_2$ FC = 1.39) and *S. pennellii* ($\log_2$ FC = 1.08). *SolySRO4* was significantly up-regulated in both M82 ($\log_2$ FC = 2.38) and *S. pennellii* ($\log_2$ FC = 1.68), which means that *SolySRO4* was the main salt stress response factors in the SRO family. In particular, the expression of *SolySRO2* significantly increased in M82 ($\log_2$ FC = 6.34) under a salt environment but did not change in *S. pennellii* Although the expression of

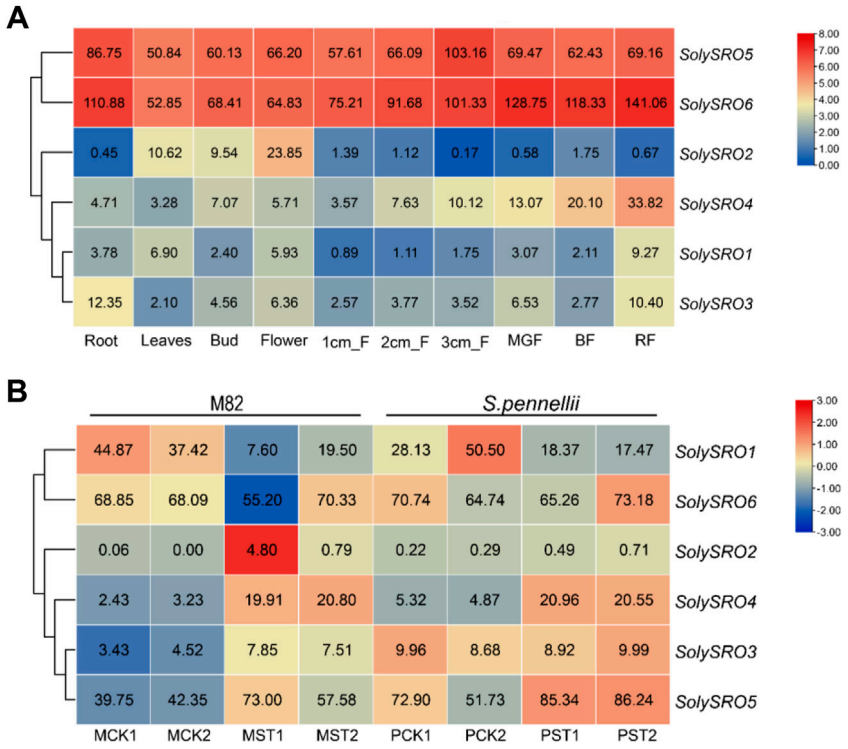

FIGURE 6 | Quantitative heat map of *SRO* gene expression. The color bar represents the $\log^2$ expression values, With red representing high expression levels and blue representing low expression levels. The gene name is shown on the right side. **(A)** Heat map of tissue-specific expression of *SRO* genes in tomato. Heinz roots (Root), Heinz leaves (Leaves), Heinz unopened flower buds (Bud), Heinz fully opened flowers (Flower), Heinz 1 cm fruits (1 cm_F), Heinz 2 cm fruits (2 cm_F), Heinz 3 cm fruits (3 cm_F), Heinz mature green fruits (MGF), Heinz breaker fruits (BF), Heinz breaker + 10 fruits (RF) **(B)** Heat map of *SRO* gene expression in M82 and *S. pennellii* under salt stress. Normal growth of M82 (MCK), Normal growth of *S. pennellii* (PCK), Salt-stressed M82 (MST), Salt-stressed *S. pennellii* (PST), Each treatment has two replicates.

SolySRO3 and *SolySRO5* increased, they did not reach the significant level. The expression level of *SolySRO6* remained basically unchanged.

Expression Profiles of *SolySROs* Under Abiotic Stress and Hormone Treatment

To investigate the expression pattern of the *SRO* genes in tomato, qRT-PCR experiments were performed to analyse six *SolySRO* genes under two abiotic stresses and three hormone treatments (**Figure 7**). Compared to the control, high-temperature stress caused a decrease in the expression of *SolySRO1* (81.60%) and *SolySRO3* (64.99%) in 2 h. The expressions of both *SolySRO2* (32.62%) and *SolySRO4* (953.09%) first increased in 2 h and then decreased by 8 h *SolySRO5* expression continued to increase, and *SolySRO6* expression remained unchanged throughout. The expression pattern of *SRO* under salt stress simulated by NaCl was different. The expression of *SolySRO1* decreased compared to the control, while the expressions of *SolySRO2* (1381.39%) and *SolySRO3* (720.26%) increased and reached a maximum at 4 h. The expressions of *SolySRO4* (1037.57%) and *SolySRO5* (563.12%) also increased, but their maximum expression occurred at 2 h. The expression of *SolySRO6* decreased first and then returned to normal at 8 h. The response of the tomato *SRO* genes was explored with auxin, methyl jasmonate and abscisic acid. The expressions of *SolySRO1*, *SolySRO2*, *SolySRO4*, and *SolySRO6* all increased under the IAA treatment, reaching maximum expression at 12 and 24 h, respectively, and the expressions of *SolySRO3* and *SolySRO5* did not change significantly. The expressions of *SolySRO5* and *SolySRO6* also did not change significantly under the MeJA treatment, while the expressions of *SolySRO1* and *SolySRO4* increased significantly and reached a maximum at 12 h, *SolySRO2* decreased significantly at 12 h and *SolySRO3* was the least expressed at 24 h. Under ABA stress, the expressions of *SolySRO1* and *SolySRO2* decreased throughout and did not recover, while the expressions of *SolySRO3*, *SolySRO4* and *SolySRO6* increased significantly and reached a maximum at 24, 12 and 6 h, respectively, while the expression of *SolySRO5* did not change significantly throughout the stress.

DISCUSSION

Next-generation sequencing (NGS) technology improves the resolution and accuracy of genomics research, focusing on repeated prediction and verification of a few genes, avoiding the annotation errors of individual gene sequences caused by

genome-wide sequencing, and enabling genetic improvement and directional breeding of plants (Rothan et al., 2019). As a small protein family unique to plants, *SRO* has been suggested to participate in a variety of abiotic stress and oxidative stress responses in plant growth, thereby enhancing plant stress tolerance. *SRO* has been isolated and identified in a variety of plants (You et al., 2014; Li et al., 2017; Jiang et al., 2018; Zhang et al., 2019; Jiang et al., 2020). In this study, we systematically identified SRO family members in a variety of tomatoes and studied their physical and chemical properties, structural characteristics, evolutionary classification and functional expression. Like most higher plants, cultivated tomato also contains 6 members of the SRO family. This number is the same as that of Arabidopsis and bananas but less than that of wheat (Ahlfors et al., 2004; Zhang et al., 2019; Jiang et al., 2020). In cultivated tomato, the SRO family is distributed on chromosomes 3, 5, 6 and 8, which was highly consistent with *S. lycopersicoides*but different from the wild tomatoes. The additional *SRO* members in wild tomatoes were mainly distributed on Chr1 and Chr4. The *SRO* genes in tomatoes show a certain degree of conservation and separation along with their distribution on the chromosome. *SRO* genes at the same or similar positions on different tomatoes chromosomes were highly consistent in their physical and chemical properties such as amino acid length, molecular weight, and isoelectric point. Similarly, *SRO* genes distributed on different chromosomes were quite different in both cultivated tomato and wild tomatoes. Based on the conservative characteristics of *SRO* genes in chromosome distribution, we can predict that *SRO* genes in *S. chilense* were also distributed on Chr1 (*SolcSRO6*), Chr3 (*SolcSRO7*), Chr5 (*SolcSRO1* and *SolcSRO2*), Chr6 and Chr8 (*SolcSRO3* and *SolcSRO4*, *SolcSRO5*), even if they were not mounted on chromosomes.

The differential functional expression of genes is closely related to their structures. Similar to the physical and chemical characteristics, whether in cultivated tomato or wild tomatoes, *SRO* genes distributed in the same or similar positions on chromosomes also had similar structures and conserved motifs. The *SRO* genes in cultivated tomato were divided into group I and group II, and group III was added by wild tomatoes. The SRO family in tomato is undoubtedly conserved. The *SRO* genes in the same group showed similar numbers of exons and conserved structures in a variety of tomatoes, especially the *SRO* genes in group I, which had a highly consistent exon distribution and the largest number of motifs and were likely the core gene cluster in the tomato SRO family. However, compared with wild tomatoes, *SRO* genes in cultivated tomato often have longer gene structures and more introns than other genes in the same group, which means that *SolySROs* can achieve transcriptional diversification through alternative splicing and other processes, thus regulating more complex and extensive functions (Liu et al., 2021). This was obviously not available in the *SRO* gene in wild tomatoes. We speculated that artificial domestication may cause the loss of *SolySROs* genes on Chr1 and Chr4. Mutations may also increase the complexity of the *SolySROs* gene structure, thereby maintaining the functional expression of the SRO family and reducing gene redundancy. Unfortunately, we have not found similar reports in SRO family studies of other species.

Predicting the promoter sequence of *SRO* genes in tomato, we found 87 CREs, it indicated that the SRO family was widely involved in mediating multiple life activities of tomato. The distribution of hormone response elements was the most widespread. Both cultivated and wild tomatoes *SRO* genes contained a large number of response elements, including gibberellin, ethylene, abscisic acid, jasmonic acid, and salicylic acid. *SRO* genes may affect tomato life activities by widely participating in hormone regulation networks, which is consistent with studies in other species (YongChun et al., 2019; Qiao et al., 2020). There were also a large number of light-responsive CREs in the *SRO* promoter region, mainly BOX-4 and G-BOX, and most light-responsive elements were significantly enriched in *S. lycopersicum* var. *cerasiforme* and *S. pimpinellifolium*, which was consistent with their light-loving and heat-resistant growth characteristics (Kumar et al., 2015). Stress-related response elements showed that many *SRO* genes were induced by an anaerobic response. The stress-related elements of SRO family members in cultivated tomato were far less abundant than those in wild tomatoes, which may lead to damage to their stress tolerance.

With the different evolutionary statuses of the plants, there were obvious differences in the *SRO* genes. The phylogenetic tree showed that genome evolution of the SRO family followed the differentiation of species, Bryophytes, Tracheophytes, Monocots and Eudicots were distributed in different branches. it was consistent with some previous studies (Zhang et al., 2019; Jiang et al., 2020). *P. patens* and *S. moellendorffii*, which have relatively simple life structures, naturally contained only a few *SRO* genes. With the occurrence of genome-wide replication events (WGD), the number of *SRO* genes gradually increases in some monocotyledons and dicotyledons, indicated that *SRO* genes did undergo lineage-specific amplification and evolution with plant differentiation. According to the phylogenetic tree of *Solanaceae*, the SRO family is more accurately divided into three subgroups. The *SRO* genes structure and typical domains in group I were relatively complete, while the *SRO* genes in groups II and III were either short in length or contained only one of the conserved RST or PARP domains. The *SRO* genes are relatively conserved in *Solanaceae*, and the genetic relationship could not be strictly divided. *SRO* genes in different *Solanaceae* may perform similar functions. Group III contained *C. annuum*, *S. tuberosum*, *S. melongena* and wild tomatoes, but cultivated tomatoes were lost from this group. Long-term artificial domestication caused the SRO family in tomato to shrink.

The proportion of nonsynonymous substitutions (*KAs*) and synonymous substitutions (*KSs*) reflects the selection pressure of gene evolution to a certain extent, generally believed that *Ka*/*Ks* > 1 represents positive selection of accelerated evolution and *Ka*/*Ks* < 1 exhibits gene duplication suffers purifying selection (Wang et al., 2010). The *Ka*/*Ks* ratio of all duplicated wheat SRO gene pairs were <1 (Jiang et al., 2020), The *Ka*/*Ks* ratio of both homologous gene pairs in cultivated tomatoes was also less than 1, these duplicated gene pairs were subject to greater selective pressure and did not produce significant functional differences during evolution. Interestingly, although The *Ka*/*Ks* ratio of most duplicated gene pairs were <1 in wild tomato, there

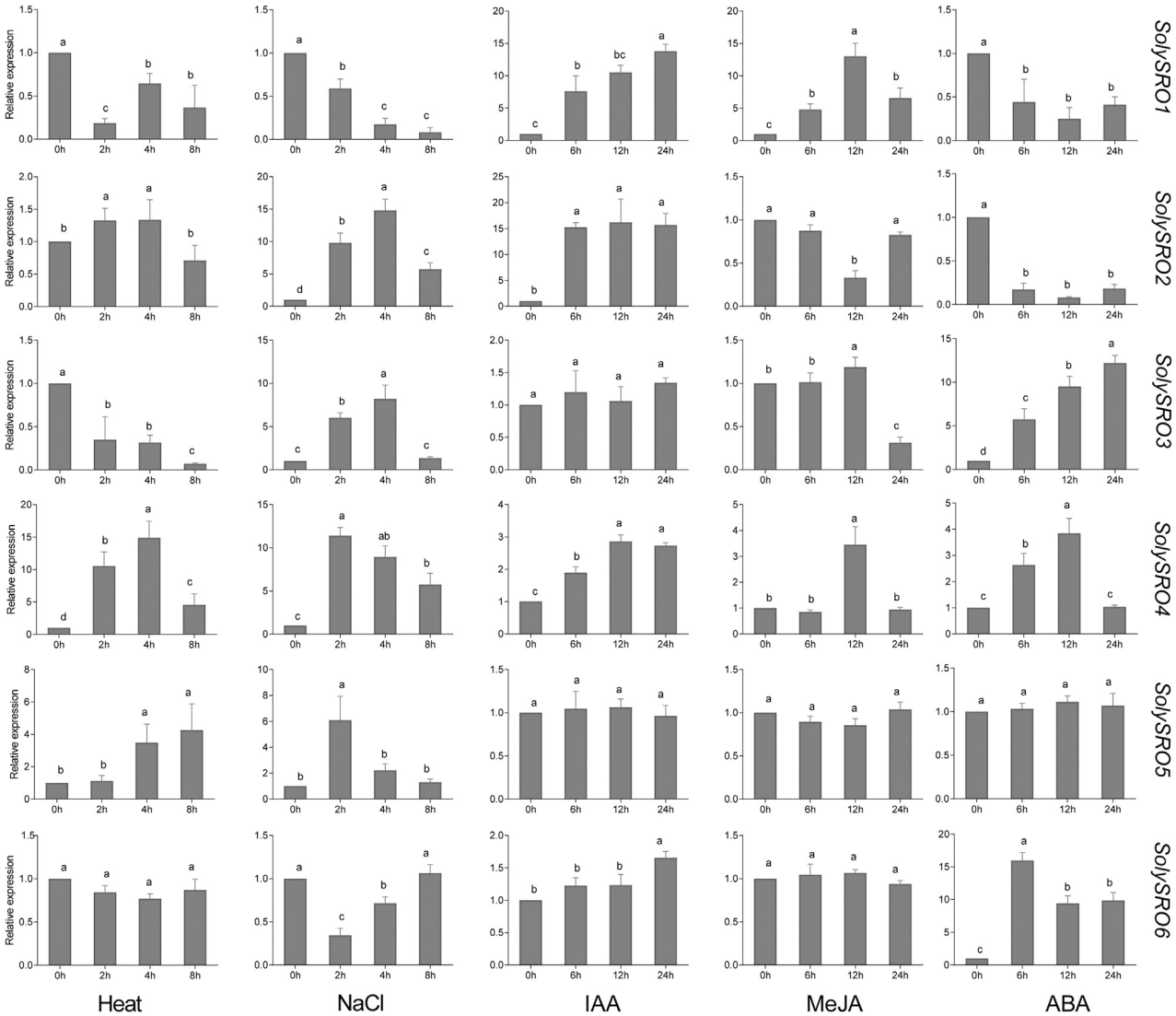

FIGURE 7 | Real-time quantitative PCR validation of *SRO* genes under abiotic stress and hormone induction, The standard deviations are shown with error bars.

were still a considerable number of duplicated genes $Ka/Ks > 1$, and some of them were from tandem repeats, implying that they were subject to environmental positive selection and still in a rapid evolutionary stage. We speculate that the more complex survival environment has forced wild tomatoes to retain the viability of some adaptive genes (Pailles et al., 2017; Gibson and Moyle, 2020) The evolution of genes in the same family often reflects certain key events in the process of species differentiation and maps the source of conservation and differential functions of its family members. Multispecies orthologous genes showed the complete evolutionary trajectory of the SRO family in tomato. The ancient ancestors of angiosperms contained only one *SRO* gene, and duplicated with the occurrence of WGT-γ. Approximately 65 Mya, The occurrence of *Solanaceae* exclusive polyploidization event drived massive expansion of SRO genes, the number of family members gradually increased, and the evolution speed accelerated. Approximately 12 Mya, with potato and tomato began to separate, the evolution of the SRO family slowed. *SolySRO1* is the most conserved member in *Solanaceae*. derived from the loss or degeneracy of two ancestral *SRO* genes after triploidization, *SolySRO2* also maintained a certain degree of similarity between the ancestral species. It formed *SolySRO3* through segmental duplication. *SolySRO4* only had orthologous genes in *Solanaceae* and no homologous relationship with grape. This meant that the *SRO* genes of Chr6 may only exists exclusively in *Solanaceae*. *SolySRO5* and *SolySRO6* have two highly homologous colinearity gene pairs in all *Solanaceae*, while only *VvSRO5* had homology with *SolySRO6* in grape. We suggested that an *SRO* genes that was triploidized in the ancestral species replicated in the genome-wide doubling event peculiar to the differentiation stage of *Solanaceae* and preserved in the evolutionary process, formed two members, *SolySRO5* and *SolySRO6*, and then tomato Chr6 and Chr8

underwent gene exchange to form *SolySRO4*. The *SRO* genes deleted on Chr4 in cultivated tomato had orthologous genes in both *Solanaceae* and grape, which further proved that the diversity of the SRO family in cultivated tomato was reduced by domestication.

The prediction results for SRO protein interactions in tomato showed that the SolySRO protein is widely involved in a variety of stress-related pathways. Among them, SLADH and SSADH respond to O_3 stress and encode aldehyde dehydrogenase to catalyse the conversion of ROS products (Sunkar et al., 2003; Timpson et al., 2012), LOC belongs to the glutathione peroxidase family, which catalyse the reduction of H_2O_2 or other organic hydroperoxides in to water or the corresponding alcohols (Islam et al., 2017), The heat shock protein family could significantly promote the ability of tomato to adapt to temperature (Hossain and Nakamoto, 2002), Overexpression of the *SOS* gene significantly improved the salt tolerance of *Arabidopsis thaliana* (Yang et al., 2009), Ubp 16 could interact with specific proteins to improve the tolerance of plants to the heavy metal cadmium (Zhao et al., 2013), The synergistic expression of SolySROs with these proteins undoubtedly improves the ability of tomato to withstand adverse environmental stresses. In Arabidopsis, *AtSRO5* mediates the formation of 24-nt-siRNA by biogenesis pathways such as DCL2 and SGS3 to accumulate proline and improve salt tolerance, while *AtSRO5* similarly reduces ROS products (Mourrain et al., 2000; Borsani et al., 2005; Deleris et al., 2006). Amazing, SolySRO1 is predicted to interact with DCL1 and SGS3 proteins, which may suggest that SolySRO1 mediates tomato proline metabolic synthesis and ROS homeostatic balance through a similar regulatory mode as Arabidopsis AtSRO5-siRNA. Six TaSROs proteins in wheat were predicted to interact with 14 transcription factors (Jiang et al., 2020). SolySROs also interacted with a large number of TFs and the RST domain always acts as the binding sites. This domain may be required for the interaction and co-expression of *SRO* genes with TFs to participate in plant stress resistance in tomato.

Poly (ADP-ribose) polymerase (PARP) widely mediates plant DNA repair, epigenetics and transcription by modifying (poly (ADP-ribosyl) ates) itself and other nuclear proteins (Vainonen et al., 2016; Briggs et al., 2017). Pharmacological inhibition assays suggest that PARP protein is involved in the natural immunity of plants against microorganisms (Adams- Phillips et al., 2009). However, the parp triple mutant whcih knocked out all three *Arabidopsis thaliana PARP* genes did not differ from wild type. Previous research hypothesized that the PARP-like structural domain of the *SRO* gene could serve as an alternative pathway when PARP activity is genetically reduced, even though the domain in the *SRO* gene did not possess any enzymatic activity and its protein sequence was similarly less similar to PARP proteins (Lamb et al., 2012; Rissel et al., 2017). Our study showed that SolySRO5 did have a direct interaction with PARP2 protein, which is the core member of the PARP family in plants (Song et al., 2015), It supported the possibility that *SRO* genes regulated active PARP proteins under specific conditions. Meanwhile *SolySRO5* and *SolysSRO6* were predicted to interact with sly-miR6023, sly-miR6024 and sly-miR6027-3p, these miRNAs regarded to be involved in plant-pathogen interactions and could regulate R gene expression in tomato (Prigigallo et al., 2019), and all four miRNA targeting sites were located in the PARP-like domain of *SRO* genes, suggesting the complexity of the active PARP protein being regulated by SRO genes. We know little about the involvement of the SRO family in plant biological stress, Four *MaSROs* showed significant dysregulation of expression in banana roots inoculated with *Fusarium oxysporum* f. sp. *Cubense* (Zhang et al., 2019), Transcriptomic data revealed that *TaSRO1b.3-4A* and *TaSRO2b.3-4B* genes in wheat were responsive to multiple fungal diseases (Jiang et al., 2020). MicroRNA family predicted in our study was conserved in Solanaceae and highly expressed in tomato leaves infected by potato virus (Li et al., 2012; Miozzi et al., 2014), these results likewise provided new insights into the involvement of *SRO* genes in biotic stresses. considering the conservation of SolySRO5 and SolySRO6 in the evolutionary process, we believed that the pattern of miRNA-SRO involvement in plant biotic stress response was at least conserved in *Solanaceae*.

Tissue-specific expression showed that the expression pattern of the SRO family members in tomato was significantly different from that in other plants. *SolySRO1* maintained low expression throughout the reproductive period. *SolySRO2* was highly expressed in seeds and flowers. *SolySRO3* had its highest expression level in roots. This gene may be related to tomato perception and response to stimuli. *SolySRO4* was highly expressed in mature tomato fruits and may be involved in the transformation of green tomato fruit to red fruit by regulating hormones such as ethylene. Compared with other *SolySRO* genes, *SolySRO5* and *SolySRO6* maintained absolute high expression throughout the growth period of tomato. These two genes were widely involved in the dynamics of tomato growth and development and reached maximum expression in the fruit. Salt stress caused an imbalance in SRO family expression, and the expressions of *SolySRO4* increased significantly to cope with the high-salt environment. In this study, the expressions of *SolySROs* under different stress environments were also verified using qRT-PCR. The expressions of *SolySRO2*, *SolySRO4*, and *SolySRO5* significantly increased under both high temperature and salt stress, and these three genes were likely to be more sensitive to the stress response and expressed rapidly in tomato in response to adverse conditions. The expression of *SolySRO4* was significantly increased at 6 h under the IAA, MeJA and ABA treatments after exogenous application of hormones, whereas the expression of *SolySRO5* did not change significantly under the three hormone environments; they all had many hormone-responsive elements distributed in their promoter regions, but the hormone response mechanisms were different. *SolySRO4*, *SolySRO5* and *SolySRO6* were evolutionarily homologous and highly similar in gene structure and conserved motifs, but their expression patterns were not identical. *SolySRO1*, *SolySRO2* and *SolySRO3*, which were distributed in the same subclade, were also highly divergent. The results that *SolySROs* expression patterns did not substantially vary in a simple linear fashion with time, and indeeded in other species (Ahlfors et al., 2004; Zhang et al., 2019; Jiang et al., 2020), provided evidence for the complex expression patterns of *SRO*.

CONCLUSION

In this study, we systematically identified the SRO family from the tomato genome and its wild relatives. We used bioinformatics method to describe the physical and chemical properties, gene structure, protein interactions, promoter elements and targeted microRNA regulation of different *SRO* genes. The evolutionary origin of the *SRO* genes in tomato was also discussed. Transcriptome analysis showed that only two genes, *SolySRO5* and *SolySRO6*, were highly expressed in different tissues of tomato and affected and regulated the dynamic changes of tomato development. Four *SolySROs* genes responded significantly to salt stress, of which *SolySRO4* and *SolySRO5* were the core genes. At the same time, the *SRO* genes were verified by qRT-PCR. These genes were involved in hormone-mediated pathways and played an important role in tomato resistance to abiotic stress. These results laid a foundation for further study of the function of the SRO family in tomato and had value for applications in tomato resistance breeding.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JG and QY guided the design of the experiment. NL and RX directed the data analysis. NL and BW conducted data analysis and manuscript writing. JW and SH finished plant material handling, QY and NL supervised the experiment and confirmed the manuscript. All authors contributed to the article and approved the submitted version. Thank all the above staff for the help in this study.

FUNDING

This work was financially supported by the Special Incubation Project of Science and Technology Renovation of Xinjiang Academy of Agricultural Sciences (xjkcpy-2021001) and the China Agriculture Research System of MOF and MARA (CARS-23-G25). The funders have no role in the study design, data analysis, and manuscript writing, but just provide the financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.753638/full#supplementary-material

REFERENCES

Adams-Phillips, L., Briggs, A. G., and Bent, A. F. (2009). Disruption of Poly(ADP-Ribosyl)ation Mechanisms Alters Responses of Arabidopsis to Biotic Stress. *Plant Physiol.* 152, 267–280. doi:10.1104/pp.109.148049

Ahlfors, R., Lång, S., Overmyer, K., Jaspers, P., Brosché, M., Tauriainen, A., et al. (2004). Arabidopsis RADICAL-INDUCED CELL DEATH1 Belongs to the WWE Protein-Protein Interaction Domain Protein Family and Modulates Abscisic Acid, Ethylene, and Methyl Jasmonate Responses. *Plant Cell* 16, 1925–1937. doi:10.1105/tpc.021832

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB Bioinformatics Resource Portal. *Nucleic Acids Res.* 40, W597–W603. doi:10.1093/nar/gks400

Babajani, G., Effendy, J., and Plant, A. L. (2009). Sl-SROl1 Increases Salt Tolerance and Is a Member of the Radical-Induced Cell Death 1-similar to RCD1 Gene Family of Tomato. *Plant Sci.* 176, 214–222. doi:10.1016/j.plantsci.2008.10.012

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., et al. (2004). The Pfam Protein Families Database. *Nucleic Acids Res.* 32, 138D–141D. doi:10.1093/nar/gkh121

Biłas, R., Szafran, K., Hnatuszko-Konka, K., and Kononowicz, A. K. (2016). *Cis*-regulatory Elements Used to Control Gene Expression in Plants. *Plant Cel Tiss Organ. Cult* 127, 269–287. doi:10.1007/s11240-016-1057-7

Blanc, G., and Wolfe, K. H. (2004). Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes[W]. *Plant Cell* 16, 1667–1678. doi:10.1105/tpc.021345

Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R., and Zhu, J.-K. (2005). Endogenous siRNAs Derived from a Pair of Natural *Cis*-Antisense Transcripts Regulate Salt Tolerance in Arabidopsis. *Cell* 123, 1279–1291. doi:10.1016/j.cell.2005.11.035

Briggs, A. G., Adams-Phillips, L. C., Keppler, B. D., Zebell, S. G., Arend, K. C., Apfelbaum, A. A., et al. (2017). A Transcriptomics Approach Uncovers Novel Roles for Poly(ADP-Ribosyl)ation in the Basal Defense Response in *Arabidopsis thaliana*. *Plos One* 12, e0190268. doi:10.1371/journal.pone.0190268

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* 13, 1194–1202. doi:10.1016/j.molp.2020.06.009

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a Plant Small RNA Target Analysis Server (2017 Release). *Nucleic Acids Res.* 46, W49–W54. doi:10.1093/nar/gky316

Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K. D., Carrington, J. C., and Voinnet, O. (2006). Hierarchical Action and Inhibition of Plant Dicer-like Proteins in Antiviral Defense. *Science* 313, 68–71. doi:10.1126/science.1128214

Du, H., Zhang, L., Liu, L., Tang, X.-F., Yang, W.-J., Wu, Y.-M., et al. (2009). Biochemical and Molecular Characterization of Plant MYB Transcription Factor Family. *Biochem. Mosc.* 74, 1–11. doi:10.1134/S0006297909010015

Fei, Z., Joung, J.-G., Tang, X., Zheng, Y., Huang, M., Lee, J. M., et al. (2011). Tomato Functional Genomics Database: a Comprehensive Resource and Analysis Package for Tomato Functional Genomics. *Nucleic Acids Res.* 39, D1156–D1163. doi:10.1093/nar/gkq991

Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., et al. (2015). The Sol Genomics Network (SGN)-from Genotype to Phenotype to Breeding. *Nucleic Acids Res.* 43, D1036–D1041. doi:10.1093/nar/gku1195

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* 39, W29–W37. doi:10.1093/nar/gkr367

Gibson, M. J. S., and Moyle, L. C. (2020). Regional Differences in the Abiotic Environment Contribute to Genomic Divergence within a Wild Tomato Species. *Mol. Ecol.* 29, 2204–2217. doi:10.1111/mec.15477

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a Comparative Platform for green Plant Genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi:10.1093/nar/gkr944

Grundy, W. N., Bailey, T. L., Elkan, C. P., and Baker, M. E. (1997). Meta-MEME: Motif-Based Hidden Markov Models of Protein Families. *Bioinformatics* 13, 397–406. doi:10.1093/bioinformatics/13.4.397

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: Protein Localization Predictor. *Nucleic Acids Res.* 35, W585–W587. doi:10.1093/nar/gkm259

Hossain, M. M., and Nakamoto, H. (2002). HtpG Plays a Role in Cold Acclimation in Cyanobacteria. *Curr. Microbiol.* 44, 291–296. doi:10.1007/s00284-001-0005-9

Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an Upgraded Gene Feature Visualization Server. *Bioinformatics* 31, 1296–1297. doi:10.1093/bioinformatics/btu817

Islam, S., Rahman, I. A., Islam, T., and Ghosh, A. (2017). Genome-wide Identification and Expression Analysis of Glutathione S-Transferase Gene Family in Tomato: Gaining an Insight to Their Physiological and Stress-specific Roles. *PLoS One* 12, e0187504. doi:10.1371/journal.pone.0187504

Jaspers, P., Blomster, T., Brosché, M., Salojärvi, J., Ahlfors, R., Vainonen, J. P., et al. (2009). Unequally Redundant *RCD1* and *SRO1* Mediate Stress and Developmental Responses and Interact with Transcription Factors. *Plant J.* 60, 268–279. doi:10.1111/j.1365-313X.2009.03951.x

Jaspers, P., Brosché, M., Overmyer, K., and Kangasjär, J. (2010a). The Transcription Factor Interacting Protein *RCD1* Contains a Novel Conserved Domain. *Plant Signaling Behav.* 5, 78–80. doi:10.4161/psb.5.1.10293

Jaspers, P., Overmyer, K., Wrzaczek, M., Vainonen, J. P., Blomster, T., Salojärvi, J., et al. (2010b). The RST and PARP-like Domain Containing SRO Protein Family: Analysis of Protein Structure, Function and Conservation in Land Plants. *BMC Genomics* 11, 170. doi:10.1186/1471-2164-11-170

Jiang, H., Xiao, Y., and Zhu, S. (2018). Genome-wide Identification, Systematic Analysis and Characterization of *SRO* Family Genes in maize (Zea mays L.) *Acta Physiol. Plant* 40, 176. doi:10.1007/s11738-018-2738-0

Jiang, W., Geng, Y., Liu, Y., Chen, S., Cao, S., Li, W., et al. (2020). Genome-wide Identification and Characterization of *SRO* Gene Family in Wheat: Molecular Evolution and Expression Profiles during Different Stresses. *Plant Physiol. Biochem.* 154, 590–611. doi:10.1016/j.plaphy.2020.07.006

Krishna, R., Karkute, S. G., Ansari, W. A., Jaiswal, D. K., Verma, J. P., and Singh, M. (2019). Transgenic Tomatoes for Abiotic Stress Tolerance: Status and Way Ahead. *3 Biotech.* 9, 143. doi:10.1007/s13205-019-1665-0

Kumar, P. P., Longjam, M., and Sikder, S. (2015). Morphological Characterisation of Tomato Wild Relatives. *Jrnl. Func. Env. Bot.* 5, 141. doi:10.5958/2231-1750.2015.00020.7

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054

Lamb, R. S., Citarelli, M., and Teotia, S. (2012). Functions of the poly(ADP-Ribose) Polymerase Superfamily in Plants. *Cell. Mol. Life Sci.* 69, 175–189. doi:10.1007/s00018-011-0793-4

Li, F., Pignatta, D., Bendix, C., Brunkard, J. O., Cohn, M. M., Tung, J., et al. (2012). MicroRNA Regulation of Plant Innate Immune Receptors. *Proc. Natl. Acad. Sci.* 109, 1790–1795. doi:10.1073/pnas.1118282109

Li, B.-Z., Zhao, X., Zhao, X.-L., and Peng, L. (2013). Structure and Function Analysis of *Arabidopsis thaliana* SRO Protein Family. *Hereditas (Beijing)* 35, 1189–1197. doi:10.3724/SP.J.1005.2013.01189

Li, H., Li, R., Qu, F., Yao, J., Hao, Y., Wang, X., et al. (2017). Identification of the SRO Gene Family in Apples (Malus×domestica) with a Functional Characterization of MdRCD1. *Tree Genet. Genomes* 13, 94. doi:10.1007/s11295-017-1175-3

Li, Z., Shen, J., and Liang, J. (2019). Genome-Wide Identification, Expression Profile, and Alternative Splicing Analysis of the Brassinosteroid-Signaling Kinase (BSK) Family Genes in Arabidopsis. *Ijms* 20, 1138. doi:10.3390/ijms20051138

Li, W., Pang, S., Lu, Z., and Jin, B. (2020). Function and Mechanism of WRKY Transcription Factors in Abiotic Stress Responses of Plants. *Plants* 9, 1515. doi:10.3390/plants9111515

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., et al. (2014). Genomic Analyses Provide Insights into the History of Tomato Breeding. *Nat. Genet.* 46, 1220–1226. doi:10.1038/ng.3117

Liu, H., Lyu, H. M., Zhu, K., Van de Peer, Y., Cheng, Z. M., and Max (2021). The Emergence and Evolution of Intron-poor and Intronless Genes in Intron-rich Plant Gene Families. *Plant J.* 105, 1072–1082. doi:10.1111/tpj.15088

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K., DeWeese-Scott, C., Gonzales, N. R., Gwadz, M., et al. (2007). CDD: A Conserved Domain Database for Interactive Domain Family Analysis. *Nucleic Acids Res.* 35, D237–D240. doi:10.1093/nar/gkl951

Miozzi, L., Napoli, C., Sardo, L., and Accotto, G. P. (2014). Transcriptomics of the Interaction between the Monopartite Phloem-Limited Geminivirus Tomato Yellow Leaf Curl Sardinia Virus and *Solanum lycopersicum* Highlights a Role for Plant Hormones, Autophagy and Plant Immune System Fine Tuning during Infection. *PLoS One* 9, e89951. doi:10.1371/journal.pone.0089951

Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.-B., et al. (2000). Arabidopsis *SGS2* and *SGS3* Genes Are Required for Posttranscriptional Gene Silencing and Natural Virus Resistance. *Cell* 101, 533–542. doi:10.1016/S0092-8674(00)80863-6

Nevo, E. (2001). Evolution of Genome-Phenome Diversity under Environmental Stress. *Proc. Natl. Acad. Sci.* 98, 6233–6240. doi:10.1073/pnas.101109298

Overmyer, K., Tuominen, H., Kettunen, R., Betz, C., Langebartels, C., Sandermann, H., Jr., et al. (2000). Ozone-Sensitive Arabidopsis *Rcd1* Mutant Reveals Opposite Roles for Ethylene and Jasmonate Signaling Pathways in Regulating Superoxide-dependent Cell Death. *Plant Cell* 12, 1849–1862. doi:10.1105/tpc.12.10.1849

Pailles, Y., Ho, S., Pires, I. S., Tester, M., Negrão, S., and Schmöckel, S. M. (2017). Genetic Diversity and Population Structure of Two Tomato Species from the Galapagos Islands. *Front. Plant Sci.* 8, 138. doi:10.3389/fpls.2017.00138

Prigigallo, M. I., Križnik, M., De Paola, D., Catalano, D., Gruden, K., Finetti-Sialer, M. M., et al. (2019). Potato Virus Y Infection Alters Small RNA Metabolism and Immune Response in Tomato. *Viruses* 11, 1100. doi:10.3390/v11121100

Qiao, Y., Gao, X., Liu, Z., Wu, Y., Hu, L., and Yu, J. (2020). Genome-Wide Identification and Analysis of SRO Gene Family in Chinese Cabbage (Brassica Rapa L). *Plants* 9, 1235. doi:10.3390/plants9091235

Rhee, S. Y., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A., et al. (2003). The Arabidopsis Information Resource (TAIR): A Model Organism Database Providing a Centralized, Curated Gateway to Arabidopsis Biology, Research Materials and Community. *Nucleic Acids Res.* 31, 224–228. doi:10.1093/nar/gkg076

Rissel, D., Heym, P. P., Thor, K., Brandt, W., Wessjohann, L. A., and Peiter, E. (2017). No Silver Bullet - Canonical Poly(ADP-Ribose) Polymerases (PARPs) Are No Universal Factors of Abiotic and Biotic Stress Resistance of *Arabidopsis thaliana*. *Front. Plant Sci.* 08, 59. doi:10.3389/fpls.2017.00059

Rombauts, S., Déhais, P., Van Montagu, M., and Rouzé, P. (1999). PlantCARE, a Plant *Cis*-Acting Regulatory Element Database. *Nucleic Acids Res.* 27, 295–296. doi:10.1093/nar/27.1.295

Rothan, C., Diouf, I., and Causse, M. (2019). Trait Discovery and Editing in Tomato. *Plant J.* 97, 73–90. doi:10.1111/tpj.14152

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000). SMART: A Web-Based Tool for the Study of Genetically Mobile Domains. *Nucleic Acids Res.* 28, 231–234. doi:10.1093/nar/28.1.231

Song, J., Keppler, B. D., Wise, R. R., and Bent, A. F. (2015). PARP2 Is the Predominant Poly(ADP-Ribose) Polymerase in Arabidopsis DNA Damage and Immune Responses. *Plos Genet.* 11, e1005200. doi:10.1371/journal.pgen.1005200

Song, L., Huang, S.-s. C., Wise, A., Castanon, R., Nery, J. R., Chen, H., et al. (2016). A Transcription Factor Hierarchy Defines an Environmental Stress Response Network. *Science* 354, aag1550. doi:10.1126/science.aag1550

Sonnhammer, E. L. L., Eddy, S. R., and Durbin, R. (1997). Pfam: A Comprehensive Database of Protein Domain Families Based on Seed Alignments. *Proteins* 28, 405–420. doi:10.1002/(sici)1097-0134(199707)28:3<405:aid-prot10>3.0.co;2-l

Su, G., Morris, J. H., Demchak, B., and Bader, G. D. (2014). Biological Network Exploration with Cytoscape 3. *Curr. Protoc. Bioinf.* 47, 8.13.1–8.13.24. doi:10.1002/0471250953.bi0813s47

Sun, X., Wang, Y., and Sui, N. (2018). Transcriptional Regulation of *bHLH* During Plant Response to Stress. *Biochem. Biophys. Res. Commun.* 503, 397–401. doi:10.1016/j.bbrc.2018.07.123

Sunkar, R., Bartels, D., and Kirch, H.-H. (2003). Overexpression of a Stress-Inducible Aldehyde Dehydrogenase Gene from *Arabidopsis thaliana* in Transgenic Plants Improves Stress Tolerance. *Plant J.* 35, 452–464. doi:10.1046/j.1365-313X.2003.01819.x

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* 47, D607–D613. doi:10.1093/nar/gky1131

Szymański, J., Bocobza, S., Panda, S., Sonawane, P., Cárdenas, P. D., Lashbrooke, J., et al. (2020). Analysis of Wild Tomato Introgression Lines Elucidates the Genetic Basis of Transcriptome and Metabolome Variation Underlying Fruit Traits and Pathogen Response. *Nat. Genet.* 52, 1111–1121. doi:10.1038/s41588-020-0690-6

Teotia, S., and Lamb, R. S. (2009). The Paralogous Genes RADICAL-INDUCED CELL DEATH1 and SIMILAR TO RCD ONE1 Have Partially Redundant Functions during Arabidopsis Development. *Plant Physiol.* 151, 180–198. doi:10.1104/pp.109.142786

Teotia, S., and Lamb, R. S. (2011). *RCD1* and *SRO1* Are Necessary to Maintain Meristematic Fate in *Arabidopsis thaliana*. *J. Exp. Bot.* 62, 1271–1284. doi:10.1093/jxb/erq363

Timpson, L. M., Alsafadi, D., Mac Donnchadha, C., Liddell, S., Sharkey, M. A., and Paradisi, F. (2012). Characterization of Alcohol Dehydrogenase (*ADH12*) from Haloarcula Marismortui, an Extreme Halophile from the Dead Sea. *Extremophiles* 16, 57–66. doi:10.1007/s00792-011-0405-0

Vainonen, J. P., Shapiguzov, A., Vaattovaara, A., and Kangasjärvi, J. (2016). Plant PARPs, PARGs and PARP-like Proteins. *Curr. Protein Pept. Sci.* 17, 713–723. doi:10.2174/1389203717666160419144721

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genomics, Proteomics Bioinf.* 8, 77–80. doi:10.1016/S1672-0229(10)60008-3

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40, e49. doi:10.1093/nar/gkr1293

Wang, X., Guo, H., Wang, J., Lei, T., Liu, T., Wang, Z., et al. (2016). Comparative Genomic De-convolution of the Cotton Genome Revealed a Decaploid Ancestor and Widespread Chromosomal Fractionation. *New Phytol.* 209, 1252–1263. doi:10.1111/nph.13689

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of Duplicate Genes in Exon-Intron Structure. *Proc. Natl. Acad. Sci.* 109, 1187–1192. doi:10.1073/pnas.1109047109

Yang, Q., Chen, Z.-Z., Zhou, X.-F., Yin, H.-B., Li, X., Xin, X.-F., et al. (2009). Overexpression of *SOS* (Salt Overly Sensitive) Genes Increases Salt Tolerance in Transgenic Arabidopsis. *Mol. Plant* 2, 22–31. doi:10.1093/mp/ssn058

YongChun, G., PengJie, W., Di, C., YuCheng, Z., XueJin, C., and NaiXing, Y. (2019). Genome-wide Identification and Expression Analysis of *SRO* Gene Family in Camellia Sinensis. *J. Tea Sci.* 39, 392–402. doi:10.3969/j.issn.1000-369X.2019.04.004

You, J., Zong, W., Du, H., Hu, H., and Xiong, L. (2014). A Special Member of the rice SRO Family, *OsSRO1c*, Mediates Responses to Multiple Abiotic Stresses through Interaction with Various Transcription Factors. *Plant Mol. Biol.* 84, 693–705. doi:10.1007/s11103-013-0163-8

Zhang, L., Zhou, D., Hu, H., Li, W., Hu, Y., Xie, J., et al. (2019). Genome-wide Characterization of a SRO Gene Family Involved in Response to Biotic and Abiotic Stresses in Banana (*Musa Spp*). *BMC Plant Biol.* 19, 211. doi:10.1186/s12870-019-1807-x

Zhang, J. (2003). Evolution by Gene Duplication: An Update. *Trends Ecol. Evol.* 18, 292–298. doi:10.1016/S0169-5347(03)00033-8

Zhao, J., Zhou, H., and Li, X. (2013). UBIQUITIN-SPECIFIC PROTEASE16 Interacts with a HEAVY METAL ASSOCIATED ISOPRENYLATED PLANT PROTEIN27 and Modulates Cadmium Tolerance. *Plant Signaling Behav.* 8, e25680. doi:10.4161/psb.25680

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 13 October 2021
doi: 10.3389/fpls.2021.721265

Capsicum chinense MYB Transcription Factor Genes: Identification, Expression Analysis, and Their Conservation and Diversification With Other Solanaceae Genomes

Khushbu Islam[1†], Abdul Rawoof[1†], Ilyas Ahmad[1], Meenakshi Dubey[2], John Momo[1] and Nirala Ramchiary[1]*

[1] School of Life Sciences, Jawaharlal Nehru University, New Delhi, India, [2] Department of Biotechnology, Delhi Technological University, New Delhi, India

OPEN ACCESS

Edited by:
Nunzio D'Agostino,
University of Naples Federico II, Italy

Reviewed by:
Zhangsheng Zhu,
South China Agricultural University, China
Magda Arce,
University of California, Davis, United States

***Correspondence:**
Nirala Ramchiary
nrudsc@gmail.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 08 June 2021
Accepted: 08 September 2021
Published: 13 October 2021

Citation:
Islam K, Rawoof A, Ahmad I, Dubey M, Momo J and Ramchiary N (2021) Capsicum chinense MYB Transcription Factor Genes: Identification, Expression Analysis, and Their Conservation and Diversification With Other Solanaceae Genomes. Front. Plant Sci. 12:721265. doi: 10.3389/fpls.2021.721265

Myeloblastosis (*MYB*) genes are important transcriptional regulators of plant growth, development, and secondary metabolic biosynthesis pathways, such as capsaicinoid biosynthesis in *Capsicum*. Although *MYB* genes have been identified in *Capsicum annuum*, no comprehensive study has been conducted on other *Capsicum* species. We identified a total of 251 and 240 *MYB* encoding genes in *Capsicum chinense MYBs* (*CcMYBs*) and *Capsicum baccatum MYBs* (*CbMYBs*). The observation of twenty tandem and 41 segmental duplication events indicated expansion of the *MYB* gene family in the *C. chinense* genome. Five *CcMYB* genes, i.e., *CcMYB101*, *CcMYB46*, *CcMYB6*, *CcPHR8*, and *CcRVE5*, and two *CaMYBs*, i.e., *CaMYB3* and *CaHHO1,* were found within the previously reported capsaicinoid biosynthesis quantitative trait loci. Based on phylogenetic analysis with tomato *MYB* proteins, the *Capsicum* MYBs were classified into 24 subgroups supported by conserved amino acid motifs and gene structures. Also, a total of 241 *CcMYBs* were homologous with 225 *C. annuum*, 213 *C. baccatum*, 125 potato, 79 tomato, and 23 Arabidopsis *MYBs*. Synteny analysis showed that all 251 *CcMYBs* were collinear with *C. annuum*, *C. baccatum*, tomato, potato, and Arabidopsis *MYBs* spanning over 717 conserved syntenic segments. Using transcriptome data from three fruit developmental stages, a total of 54 *CcMYBs* and 81 *CaMYBs* showed significant differential expression patterns. Furthermore, the expression of 24 *CcMYBs* from the transcriptome data was validated by quantitative real-time (qRT) PCR analysis. Eight out of the 24 *CcMYBs* validated by the qRT-PCR were highly expressed in fiery hot *C. chinense* than in the lowly pungent *C. annuum.* Furthermore, the co-expression analysis revealed several *MYB* genes clustered with genes from the capsaicinoid, anthocyanin, phenylpropanoid, carotenoid, and flavonoids biosynthesis pathways, and related to determining fruit shape and size. The homology modeling of 126 R2R3 CcMYBs showed high similarity with that of the Arabidopsis R2R3 *MYB* domain template, suggesting their potential functional similarity at the proteome level. Furthermore, we have identified simple

sequence repeat (SSR) motifs in the *CcMYB* genes, which could be used in *Capsicum* breeding programs. The functional roles of the identified *CcMYBs* could be studied further so that they can be manipulated for *Capsicum* trait improvement.

Keywords: *Capsicum chinense*, *baccatum*, *MYB*, Solanaceae, fruit development, transcription factors

INTRODUCTION

The myeloblastosis (*MYB*) gene family is one of the largest transcription factor (TF) families in plants (Romero et al., 1998; Riechmann et al., 2000). MYB TFs have one or more imperfect repeats of the characteristic DNA-binding domain (DBD) in the basic region of a protein (Klempnauer and Sippel, 1987). Each repeat comprises about 53 amino acids with three regularly placed tryptophan residues forming a helix-turn-helix structure (Ogata et al., 1994; König et al., 1996). *MYB* TFs with only one repeat are called "MYB1R" or MYB-related, while those with two, three, or four repeats are called "R2R3-MYB," "MYB3R," or "MYB4R," respectively. In plants, most MYBs belong to the R2R3 class, unlike in animals, and exhibit plant-specific responses (Martin and Paz-Ares, 1997). MYB TFs play key roles in the regulation of phenylpropanoid and flavonoid metabolism in plants; for instance, in Arabidopsis, *MYB123* partly determines the accumulation of proanthocyanidin (PA) in the coat of seeds (Nesi et al., 2001), *MYB11*, *MYB12*, and *MYB111* are involved in the transcriptional regulation of the *chalcone synthase* and *flavonol synthase* genes (Mehrtens et al., 2005), while *MYB14* and *MYB15*, along with WRKY53, are reported to regulate stilbene synthesis in Chinese wild grapes (Wang et al., 2020b). *MYB88* and *MYB124* were reported to have diverse roles (Lei et al., 2015), such as in the regulation of mitotic divisions of the stomatal guard mother cell (Lai et al., 2005; Lee et al., 2013) and direct transcriptional regulation of auxin transporter PIN-FORMED proteins in roots of *Arabidopsis thaliana* (Xie et al., 2010; Wang et al., 2015; Geng et al., 2018). Reports also suggest their role in female reproductive development (Makkena et al., 2012) and conditional repression of non-stomatal epidermal cells in Arabidopsis cotyledons (Yang, 2016). *MYB75* and *MYB90* are known to activate phenylpropanoid biosynthetic genes and the accumulation of purple anthocyanins in Arabidopsis (Kranz et al., 1998; Borevitz et al., 2000). Furthermore, several *MYBs* have been reported to induce anthocyanin production in different organs, including fruits in tomatoes (Kiferle et al., 2015; Jian et al., 2019; Yan et al., 2020), potato (Li et al., 2021), and in other plants (Quattrocchio et al., 2006; Takos et al., 2006; Cutanda-Perez et al., 2009; Czemmel et al., 2009; Kortstee et al., 2011; Wang et al., 2017; Yan et al., 2019). In *Capsicum*, MYB_A and *CaAN2* control anthocyanin pigmentation in flower and fruit tissues (Aguilar-Barragán and Ochoa-Alejo, 2014; Jung et al., 2019). However, MYB TFs have been scarcely studied for their protein structures; for example, a MYB-related motif in Arabidopsis recognizes the major groove of target DNA *via* the amino acid residues present in three alpha helices while binding to the minor groove using an N-terminal arm (Hosoda et al., 2002). Another report characterized the crystal structure of the MYB domain from an *Antirrhinum majus* single MYB repeat RADIALIS (RAD) TF, which functions in the development of floral asymmetry (Stevenson et al., 2006).

The genus *Capsicum* comprises several species grown worldwide mostly for vegetables and spices, which are of high economic and nutritional value. The *Capsicum* fruit is known for its unique attribute of pungency owing to alkaloids, known as capsaicinoids complex, mainly capsaicin and dihydrocapsaicin (Antonious and Jarret, 2006), which have pharmaceutical applications (Fattori et al., 2016). There exists a wide variation in capsaicin content in *Capsicum* genotypes, with the highest being reported in Bhut jolokia (*C. chinense*; Sarpras et al., 2016; Chhapekar et al., 2020). *R2R3-MYB31* was reported as a transcriptional regulator of capsaicinoid biosynthetic genes (CBGs; Arce-Rodríguez and Ochoa-Alejo, 2017). The same *MYB* was reported to be encoded by a pungency-controlling locus *Pun3* (Han et al., 2019), and its gene promoter showed natural variations between high and low pungent *C. annuum* species (Zhu et al., 2019). *CaMYB108* confers a pungent flavor to *Capsicum* genotypes and controls stamen development, and it is found to be induced by methyl jasmonate (Sun et al., 2019). Recently, *CaMYB48* was discovered to directly control the expression of *CBGs acyl transferase* (*AT3a*) and *ketoacyl-ACP synthase* (*KasIa*) and the accumulation of capsaicinoids in *C. annuum* (Sun et al., 2020).

Myeloblastosis genes have been identified only in *C. annuum* (Wang et al., 2020c; Arce-Rodríguez et al., 2021), and no comprehensive study has been reported in other *Capsicum* species, such as *C. chinense*, *C. baccatum*, and *C. frutescens*. In this study, we identified *MYB* genes in the *C. chinense*, *C. baccatum*, and *C. annuum* genomes, and the analysis of their expression was performed using transcriptome data and validated by quantitative real-time (qRT) PCR in the early green (EG), mature green (MG), and breaker (Br) fruit developmental stages of highly pungent *C. chinense* and lowly pungent *C. annuum*. Seven *MYB* genes were found within the previously reported capsaicinoids quantitative trait loci (QTLs) (Han et al.,

Abbreviations: CcMYB, *Capsicum chinense* MYB; CbMYB; *Capsicum baccatum* MYB; TF, transcription factor; DBD, DNA-binding domain; CSSs, conserved syntenic segments; qRT-PCR, quantitative real-time PCR; SSRs, simple sequence repeats; HMM, Hidden Markov Model; BLASTs, basic local alignment search tool sequence; NCBI, National Centre for Biotechnology information; CDD, Conserved Domains Database; pI, isoelectric point; GRAVY, grand average of hydropathicity; SMART, simple modular architecture research tool; CDS, coding sequences; QTLs, quantitative trait loci; GSDS, gene structure display server; MEME, multiple em for motif elicitation; NJ, neighbor-joining; ML, maximum likelihood; DPA, day post anthesis; EG, early green; MG, mature green; Br, breaker; RIN, RNA integrity number; RNAseq, RNA sequencing; FC, fold change; DEG, differentially expressed gene; Kb, kilobase; TSS, transcription start site; Chr, chromosome; FPKM, fragments per kilobase of transcript per million mapped reads; DE, differentially expressed; Mbp, million base pair.

2018; Park et al., 2019). The co-expression analysis revealed several *MYB* genes that clustered with capsaicinoid, anthocyanin, phenylpropanoid, flavonoid, fruit shape and size, carotenoid, and vitamin C biosynthesis pathway genes. Furthermore, we analyzed duplications of *MYB* genes in *C. chinense*, and comparative analysis with *C. baccatum, C. annuum*, tomato, potato, eggplant, and Arabidopsis showed conserved syntenic segments (CSSs) and collinear *MYB* genes, and diversification among them. In addition, we performed homology modeling of R2R3 CcMYB proteins and developed simple sequence repeat (SSR) markers in genic and promoter regions of *Capsicum MYBs*, which can be used to manipulate pungency levels in future *Capsicum* breeding programs. The identified *MYB* genes in this study could be used in the future to understand their regulatory roles in diverse biological functions including the capsaicinoid biosynthetic pathway.

MATERIALS AND METHODS

Identification of *MYB* Genes in *Capsicum* Spp

For the identification of *MYB* genes, a local blastp analysis (e-value < 1e-05) was performed against 34,974, 35,853, and 31,353 full-length protein sequences of *C. chinense* (GCA_002271895.2; Kim et al., 2017), *C. annuum* (GCA_000710875.1; Qin et al., 2014), and *C. baccatum* (GCA_002271885.2; Kim et al., 2017), respectively, using the R2R3-MYB family protein sequences of Arabidopsis (Stracke et al., 2001), tomato (Li et al., 2016), potato (Li et al., 2021), and *C. annuum* (Arce-Rodríguez et al., 2021) as a query. For the Hidden Markov Model (HMM) analysis, the sequences of the *Capsicum* proteins were queried against the protein sequences of MYB DNA-binding domains (PF00249 and PF13921) using HMMER (v3.2.1; http://hmmer.janelia.org/). The duplicate sequences were removed manually, and the remaining sequences were cross-checked for the presence of the MYB domain using the Conserved Domains Database (CDD) of the National Centre for Biotechnology Information (NCBI) (Lu et al., 2020) and the Simple Modular Architecture Research Tool (SMART); http://smart.embl-heidelberg.de/) database. The resultant *C. chinense* and *C. baccatum MYB* encoding genes were given acronyms according to their amino acid sequence homology with *C. annuum MYB* genes (Arce-Rodríguez et al., 2021). Physicochemical features, such as molecular weight, theoretical isoelectric point (pI), instability index, aliphatic index, and grand average of hydrophobicity (GRAVY) of the MYB protein sequences were estimated using ProtParam (https://web.expasy.org/protparam/; Gasteiger et al., 2005). The subcellular localization of *Capsicum* MYBs was determined using Cello (v.2.5 http://cello.life.nctu.edu.tw/).

Chromosomal Distribution, Gene Duplication, and Co-localization With Capsaicinoid QTLs

The chromosomal positions of *Capsicum MYB* genes were obtained from the gene feature file (gff) of their genomes. The physical locations of *Capsicum MYB* genes, along with the capsaicin and dihydrocapsaicin QTLs as reported earlier (Han et al., 2018; Park et al., 2019), were represented across 12 *Capsicum* chromosomes using TBtools (v1.068; Chen et al., 2020). The duplication of *MYB* genes within the *C. chinense* genome was identified based on filter criteria of >75% identity and query coverage of above 75% of the gene length. Gene pairs with a <100-kb (kilobase) distance on the same chromosome were considered as tandem duplicates, while those with >100 kb were considered as segmental duplicates. The rate of non-synonymous (*Ka*) and synonymous substitutions (*Ks*) and their ratio ($\omega = Ka/Ks$) for all duplicated gene pairs were estimated using KaKs_Calculator 2.0 (Wang et al., 2010). The value of $\omega \sim 0$ indicates neutral selection, $\omega < 1$ indicates purifying selection, and $\omega > 1$ indicates positive selection. The date of duplication (diversion time) was calculated using the formula $T = Ks/2\lambda$, assuming a clock-like rate (λ) of 6.96 synonymous substitutions per 10^{-9} years (Moniz de Sá and Drouin, 1996).

Gene Structure, Motifs, Cis-Elements Analysis, and Homology Modeling

The structure of the *MYBs* genes (exons and introns) was represented using Gene Structure Display Server (GSDS2.0; http://gsds.gao-lab.org/). The conserved motifs in the MYB protein sequences were identified using the Multiple Em for Motif Elicitation (MEME) suite (http://meme-suite.org/tools/meme). The maximum number of motif 40, the minimum width of each motif 6, and the maximum width of 120 were used as parameters. The identified conserved motifs were then confirmed with previously characterized Arabidopsis motifs (Stracke et al., 2001). The *cis*-regulatory elements and motifs in the 1,500 bp (base-pairs) upstream promoter region of *Capsicum MYB* genes were speculated using the PlantCARE website (Lescot et al., 2002). The 126 R2R3 CcMYBs were analyzed for protein tertiary (or 3D; three-dimensional) structure-based homology models using the Phyre2 server (Kelley et al., 2015). The models were predicted based on the alignment coverage, percent identity, and percent confidence score for the individual CcMYB protein sequences.

Phylogenetic Analysis

The multiple protein sequence alignment of *Capsicum* MYBs was performed using Clustal Omega (Madeira et al., 2019) with default parameters. The phylotree was constructed in MEGAX (v.10.1.8; Kumar et al., 2018) using maximum likelihood (ML) methods with a phylogeny test of 1,000 bootstrap replications and a Jones–Taylor–Thornton (JTT) model with uniform rates among sites applied to infer evolutionary history. The Nearest-Neighbor-Interchange (NNI), as an ML heuristic method, was used for phylogeny tree inference. The combined phylogenetic tree of *C. chinense* and *C. baccatum* R2R3 *MYBs* with already known Arabidopsis *MYBs* (*AtMYBs*) was generated using the above-mentioned parameters. The phylogenetic tree data with bootstrap values were visualized using the Interactive Tree of Life (iTOL) server (https://itol.embl.de/).

Plant Materials and Growth Conditions

Seeds of *Capsicum* genotypes belonging to highly pungent *C. chinense* (Bhut Jolokia; *Acc-Cc74*; 925084.8 Scoville Heat Unit; SHU) and lowly pungent *C. annuum* (*Acc-Ca18*; 7034.4 SHU) were sown in agro peat and vermiculite (in the proportion 3:1; Sarpras et al., 2019). The seedlings were grown in a glasshouse at 24–26°C temperature with 16-h light and 8-h dark photoperiod and 70% humidity. The 1-month-old plants were transferred into the soil and grown until fruit maturity in the glasshouse at Jawaharlal Nehru University, New Delhi. Fruit tissues of the early green (EG; 5–10 days postanthesis; DPA), mature green (MG; 20–25 DPA), and breaker (Br; 30–45 DPA) stages were harvested and immediately frozen in liquid nitrogen and stored in a deep freezer at −80°C until RNA extraction.

RNA Sequencing and Differential Gene Expression Analysis

Total RNA from the EG, MG, and Br fruit stages of *C. chinense* (*Acc-Cc74*) and *C. annuum* (*Acc-Ca18*) was extracted using an MN Nucleospin RNA Plant kit (Takara, Mountain View, CA, United States). The integrity of the RNA samples was checked using a bioanalyzer (Agilent Technologies, Santa Clara, CA, United States). The RNA samples from three biological replicates of each fruit stage, with RNA integrity number (RIN) > 8, were used for library preparation using TruSeq RNA Sample Prep Kits (Illumina, San Diego, CA, United States) and sequenced using a HiSeq XTen (Illumina, San Diego, CA, United States) paired-end platform with an average read length of 150 bp. The quality of raw reads was evaluated with FastQC (v0.11.5), and adapter sequences along with low quality reads (phred score < 20) were removed using TrimGalore (v0.4.4) as descrbied earlier (Chhapekar et al., 2020; Rawoof et al., 2020). The filtered good quality reads from *C. chinense* and *C. annuum* were mapped to their respective genomes (Qin et al., 2014; Kim et al., 2017) using the Hisat2 (Kim et al., 2019) program. The expression of all genes was estimated using StringTie (v2.0.6; Kim et al., 2019). Read counts of transcripts were quantified using the feature Counts (v1.5.1; Liao et al., 2014). The normalization of raw read counts was performed using TMM methods (Robinson and Oshlack, 2010), and differentially expressed genes (DEGs) between two tissues were identified using the glmQLFit and glmQLFTest functions in the edgeR package (Robinson et al., 2010). Genes with adjusted $p < 0.01$ and fold change (FC) >1.5 were considered as significantly expressed between the two tissues. The normalized expressions and differential expression pattern of *Capsicum MYBs* among fruit tissues were illustrated in the form of a heatmap using gplots (Warnes et al., 2020).

Expression Analysis of MYB Genes by Quantitative Real-Time (qRT) PCR

A total of 24 *CcMYB* genes showing DE among the fruit developmental stages of *C. chinense* and *C. annuum,* along with *CBG (AT3*, *KasI*, *AMT*, *ACS1*, *BCAT*, and *COMT*) genes, were validated by qRT-PCR. Gene-specific primers were designed from exonic sequences using standard criteria (Dieffenbach et al., 1993) (**Supplementary Table 1**). The total RNA was extracted as described above. The quality of RNA was checked on 1% agarose gel, and the quantity was measured using NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, United States). The total RNA (1 μg) was then converted into complementary DNA (cDNA) using PrimeScript IV 1st strand cDNA Synthesis Mix (Takara, United States) following the protocol of the manufacturer. The real-time PCR reaction was set up using SYBR Green Mix (Clontech, Mountain View, CA, United States) and run on the CFX96 Real-Time System (Bio-Rad Laboratories, Hercules, CA, United States). The thermal profile included the initial denaturation step (95°C for 2 min) and followed by a 40-cycle amplification step (95°C for 15 s and 60°C for 1 min). For qRT-PCR, three biological replicates of each fruit stage with three technical replicates were used. The relative expression of each gene was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The *actin* gene was taken as an internal control. The student's *t*-test was performed for calculating significant differences in the expression of *MYB* genes ($p < 0.05$).

Co-expression Analysis of *MYB* Genes

The co-expression of *Capsicum MYBs* with genes involved in capsaicinoid biosynthesis (*AT3*: *acyltransferase 3*; *Kas*: *ketoacyl-ACP synthase; pAMT*: *putative aminotransferase; BCKDH: branched-chain a-ketoacid dehydrogenase a-ketoacid decarboxylase; ACL: acyl carrier protein; FatA: acyl-ACP thioesterase*), phenylpropanoid biosynthesis (*PAL: phenylalanine ammonia-lyase; COMT*: *caffeic acid 3-O-methyltransferase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumaroyl-CoA ligase; BCAT*: *branched-chain amino acid aminotransferase*), carotenoid biosynthesis (*CCS: capsanthin-capsorubin synthase; BCH: b-carotene hydroxylase; PSY: phytoene synthase*), anthocyanin biosynthesis genes (*DFR: dihydroflavonol 4-reductase; F3′5′H: flavonoid 3′, 5′-hydroxylase; CHS: chalcone synthase*), vitamin C biosynthesis genes (*GLDH: L-galactono-1,4-lactone dehydrogenase; GalDH: L-galactose-1-dehydrogenase; GME: GDP-D-mannose-3′,5′-epimerase*), and fruit shape and size genes (*WD-40; CLAVATA1; auxin receptor; EAR1; SEC8: Arabidopsis Exocyst Complex; WUSCHEL; TTL3: TPR repeat-containing thioredoxin TTL3; OVATE; SUN*) was represented using a heatmap. The expression of genes was shown as $\log_2$FPKM, and clustering was performed based on Pearson correlation analysis.

Synteny and Homologous *MYB* Gene Pairs

Genome-wide CSSs and the collinearity of *C. chinense MYB* genes with *C. annuum*, *C. baccatum*, tomato, potato, eggplant, and Arabidopsis genomes were identified using the MCScanX toolkit (Wang et al., 2012). Blastp search (e-value <1e-10) results among the total protein sequences of each genome against *C. chinense* total protein sequences were used as input. The whole genome sequences of *C. chinense* (GCA_002271895.2), *C. annuum* (GCA_000710875.1), and *C. baccatum* (GCA_002271885.2) were downloaded from NCBI, and the eggplant genome was downloaded from Solanaceae Genomics Network (https://solgenomics.net/; Barchi et al., 2019). The Arabidopsis, tomato, and potato genome sequences were retrieved from Phytozome v12 (https://phytozome.jgi.doe.gov/pz/portal.html). The CcMYB homologous proteins from the *C. annuum*, *C. baccatum*, tomato,

potato, and Arabidopsis genomes were identified by blastp search with cut-off parameters: e-value <1e-03, percent identity > 75%, and coverage with > 80% of query length. The CSSs were represented using TBtools (v1.068; Chen et al., 2020), and collinear and homologous *MYBs*, along with physical positions on different chromosomes, were presented in a chord diagram using the circlize R package (Gu et al., 2014).

Simple Sequence Repeat (SSR) Prediction in *Capsicum MYBs*

Full length gene sequences and 1.5-kb upstream of promoter sequences from the Transcription Start Site (TSS) of 251 *C. chinense MYBs* were used to identify simple sequence repeats (SSRs) using the online WebSat tool (Martins et al., 2009) as described previously (Dubey et al., 2019; Jaiswal et al., 2020). The maximum size of an SSR motif was kept at 6 nucleotides, while the minimum number of repeats of the motif was kept at 6. Mononucleotide repeats were excluded from the analysis.

RESULTS

Identification of *MYB* Genes in *Capsicum* Spp

After the duplicate sequences were removed from the blastp search results and Hidden Markov Model (HMM) analysis, a total of 301, 433, and 292 potential *MYB*-encoding genes were predicted in the *C. chinense*, *C. annuum*, and *C. baccatum* genomes, respectively. The remaining sequences were screened using CDD and the SMART database to ascertain the presence of the MYB domain. Ultimately, a total of 251, 245, and 240 *MYB*-encoding genes were identified in *C. chinense, C. annuum,* and *C. baccatum* genomes, respectively. In the *C. chinense* genome, out of the total *CcMYBs*, 126 were R2R3 type, while the remaining 99 were MYB1R type, 25 belonged to MYB with other domains type, and 1 was atypical MYB. A total of 128 *C. annuum MYB* (*CaMYB*) and 123 *C. baccatum MYB* (*CbMYB*) were of R2R3-type. The molecular weight of CcMYBs varied from 11.5 to 114.6 kDa (kilodalton), and that of MYB related from 8.8 to 183.5 kDa. Most of the *CcMYBs* were localized in the nucleus, seven in mitochondria, two in the chloroplast, two in the cytoplasm, and one in extracellular (**Supplementary Table 2**). The coding (CDS) and protein sequences of *CcMYB* genes are given in **Supplementary File 1**.

Genome-Wide Distribution, Duplication, and Co-mapping of *Capsicum MYBs* With Capsaicinoid QTLs

Of all the *MYB* genes identified, 232 (92.43%) were physically mapped on 12 chromosomes (chrs), and the remaining were mapped on scaffolds of the *C. chinense* genome (**Figure 1**). In *C. annuum*, 222 (90.61%) and in *C. baccatum* 221 (92.08%) *MYBs* were mapped on their respective chrs (**Supplementary Figures 1A,B**). Uneven distribution of *MYB* genes on the 12 *Capsicum* chrs (average 11–12/ chr) was observed. In *C. chinense*, the upper end of the arm on chr1 and the lower end of the arm on chrs 2 and 6, respectively, have a greater density of *MYBs*. A similar distribution of *MYBs* was observed for chrs 1 and 2 in both *C. annuum* and *C. baccatum* and for chr 6 in *C. baccatum*.

In the *C. chinense* genome, nine clusters of 20 tandem and 41 segmental duplicated *CcMYB* gene pairs were observed with *Ka/Ks* ratios ranging from 0.001 to 1.0269 (**Table 1**). Among the duplicate pairs, *CcMYB60-1* and *CcMYB60-2* had the highest *Ka/Ks* ratio (1.0269) followed by *CcPHR10* and *CcPHR3-2* (0.8988), *CcDIV6-2* and *CcDIV6-1* (0.8396), and *CcARR9* and *CcARR8* (0.7781). *Ka/Ks* values of >1 indicate a positive selection, while a *Ka/Ks* ratio of <1 indicates purifying selection for these *MYB* gene pairs in the *C. chinense* genome. The minimum diversion time was 0.2414 MYA for the gene pair duplicated in tandem, *CcGLK1-1* and *CcGLK1-2*, while it was highest between the *CcMYB106* and *CcMYB104* segmental duplicated pair, i.e., 309.78 MYA.

Our analysis showed at least seven *MYB* genes (five *CcMYBs* and two *CaMYBs*) within the previously reported capsaicinoid QTLs (Han et al., 2018; Park et al., 2019) (**Figure 1**). At chr 6 of *C. chinense*, *CcMYB101* and *CcPHR8* are located within the dihydrocapsaicin QTL *SJ-dhc6.2*; *CcMYB46* and *CcMYB6* within the capsaicinoid QTL *HJ-tcp6.4;* and *CcRVE5* within capsaicinoid QTLs *HJ-tcp6.1* and *HJ-tcp6.2*. At chr 2 of *C. annuum*, *CaMYB3* and *CaHHO1* are located within the two QTLs, i.e., a dihydrocapsaicin QTL *PD-dicap2.2* and total capsaicinoid QTL *PD-total2* (**Supplementary Figure 1A**).

Gene Structure, Motif, and Cis-Element Analysis

Most of the *MYB* genes, i.e., 60.95% in *C. chinense*, 61.22% in *C. annuum,* and 60.41% in *C. baccatum*, have two to three exons (**Figure 2** and **Supplementary Figure 2**), and very few showed 11 or more exons. The motif analysis of CcMYB proteins revealed that in the R2 MYB domain, 18 (out of 52) amino acid (aa) positions were conserved in 80% of the R2R3-MYB protein sequences (**Figure 3**). However, 20 (out of 52) in the R3 domain were conserved, suggesting that R3 is relatively more conserved. The number and placement of tryptophan residues were found to be highly conserved; three tryptophan residues placed 20–21 residues apart in the R2 domain, and two tryptophan residues placed 19 residues apart in the R3 domain. The phenylalanine residue, which replaces the first tryptophan residue in the R3 repeat, is also found to be highly conserved. Apart from the MYB repeats, conserved motifs were observed on the C terminal of the MYB protein sequences (**Supplementary Table 3**). Previously identified motifs in *AtMYBs* were also identified in *CcMYBs*; for instance, motif-23 ([TY][SV]AN[LA][SR]HMA[QE]WES ARLEAEARL[VS]R[EK]S[KQ]), which has previously been defined in *Antirrhinum majus* MIXTA MYB, was observed in CcMYB89, CcMYB90, CcMYB91, and CcMYB87 (Kranz et al., 1998) protein sequences. The putative *cis*-elements were also identified in *C. chinense, C. annuum,* and *C. baccatum MYBs* (**Supplementary Table 4**). Both elements binding to basic transcriptional machineries like the TATA and CAAT boxes (Forde et al., 1985) and cis-elements like hormone-responsive ABRE (ABA-Responsive Element; Yamaguchi-Shinozaki and

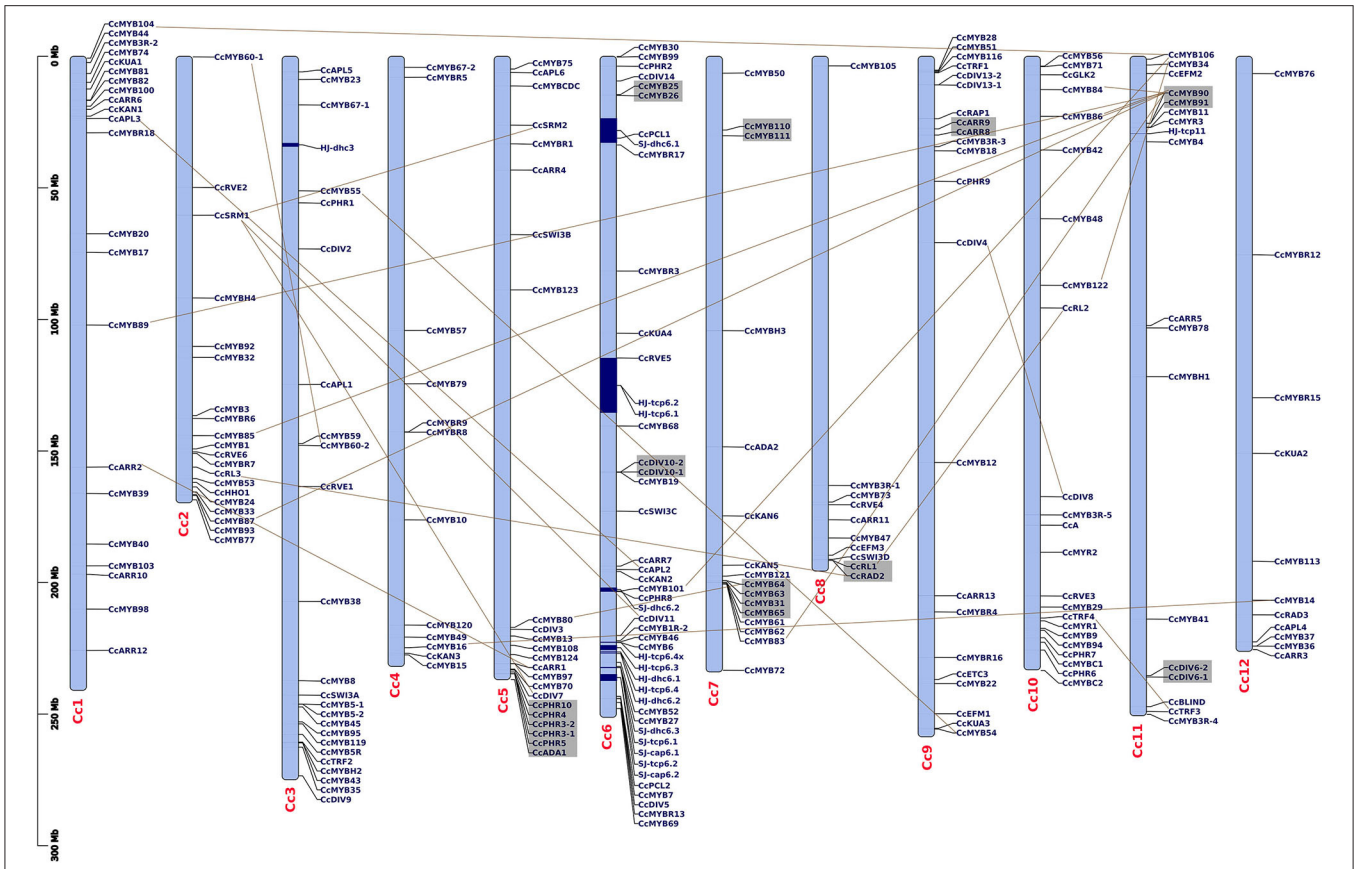

FIGURE 1 | Distribution of *CcMYB* genes on 12 *Capsicum* chromosomes along with the previously reported capsaicinoid QTLs (Han et al., 2018; Park et al., 2019). The QTLs are shown in blue blocks. The tandem duplicates are shown *via* gray blocks, while the segmental duplicates are shown *via* connecting lines. The scale is mentioned on the left of the chromosomes in million base pairs (Mbp).

Shinozaki, 1994); seed-specific like RY-element (Fujiwara and Beachy, 1994), AE-Box (Sevilla-Lecoq et al., 2003), AACA motif (Yoshihara et al., 1996), GCN4 motif (Müller and Knudsen, 1993) and Box II (Kim and Wu, 1990); light-responsive elements like MRE (Safrany et al., 2008), G-box (Schindler et al., 1992), GATA-motif (Reyes et al., 2004); low temperature-responsive LTRE (Dunn et al., 1994); and drought- and stress-responsive TC-rich repeats, were identified.

Phylogenetic Analysis of *Capsicum MYBs*

In the phylogenetic analysis, 126 *CcMYBs*, 123 *CbMYBs*, and 147 *AtMYBs* clustered into 24 subgroups (**Figure 4**) with >75% confidence in most of the branches, for instance, nearly 100% confidence between clades of *CcMYB97* and *CbMYB97* in subgroup III, *CcMYB7* and *CbMYB7* in subgroup IV, and *AtMYB121*, *CcMYB24,* and *CbMYB24* in subgroup XI were observed. Similarly, subgroup VIII MYBs like *CcMYB16*, *CcMYB13*, and *CcMYB14* share 99–100% confidence in their clades with *CbMYB16*, *CbMYB13*, and *CbMYB14*, respectively. However, *CbMYB18-1* in subgroup IX shares a sister clade with *AtMYB91* with only 51.6% confidence. MYB3Rs (subgroup V) formed a separate subgroup but share a distant common ancestor with *R2R3MYBs,* suggesting their common origin. Interestingly, *Capsicum MYBs*, such as *MYB31*, *MYB59*, *MYB60*, *MYB62*, *MYB63*, *MYB64*, and *MYB65*, were segregated separately in subgroup XIV. The phylogenetic tree constructed from a total of 734 MYB protein sequences from three *Capsicum* species resolved into 22 subgroups (**Supplementary Figure 3**). Several *MYB* genes, such as *CcMYBR14*, *CcMYBR17*, and *CbLHY,* clustered alone with no recent sister clade in the other two *Capsicum* species.

Expression Analysis of *MYB* Genes

The transcriptome data of the EG, MG, and Br fruit development stages of *C. chinense* and *C. annuum* were used to determine the expression values of all the *MYB* genes based on the normalized FPKM (fragments per kilobase of transcript per million mapped reads). A total of ∼458 million raw paired-end reads were generated from the fruit tissues of *C. chinense* and *C. annuum*. Around 196.8 and 200.9 million reads out of ∼237.4 and 220.8 million clean reads were aligned with an average alignment rate of 82.9 and 91.02% against the *C. chinense* and *C. annuum* genomes (Qin et al., 2014; Kim et al., 2017), respectively (**Supplementary Table 5**). We observed a variation in the expression of *MYB* genes among the different fruit developmental stages in *C. annuum* and *C. chinense*.

TABLE 1 | *MYB* gene duplicates in the *Capsicum chinense* genome.

C. chinense MYB name	Duplicated *MYB* pair	Duplication type	Ka	Ks	Ka/Ks	Diversion time (T = Ks/2λ) MYA
CcAPL3	*CcAPL2*	Segmental	0.131	0.526	0.248	37.77
CcARR1	*CcARR2*	Segmental	0.167	0.834	0.200	59.92
CcARR9	*CcARR8*	Tandem	0.092	0.118	0.778	8.50
CcDIV10-1	*CcDIV10-2*	Tandem	0.078	0.117	0.668	8.43
CcDIV11	*CcSRM1*	Segmental	0.427	3.501	0.122	251.48
CcDIV4	*CcDIV8*	Segmental	0.164	1.008	0.163	72.43
CcDIV6-2	*CcDIV6-1*	Tandem	0.091	0.108	0.840	7.77
CcDIV7	*CcSRM1*	Segmental	0.456	2.903	0.157	208.54
CcGLK1-1	*CcGLK1-2*	Segmental	0.000	0.003	0.001	0.24
	CcGLK1-3	Segmental	0.000	0.003	0.001	0.24
	CcGLK1-4	Segmental	0.000	0.003	0.001	0.24
CcGLK1-2	*CcGLK1-3*	Segmental	NA	NA	NA	–
	CcGLK1-4	Segmental	NA	NA	NA	–
CcGLK1-4	*CcGLK1-3*	Segmental	NA	NA	NA	–
CcMYB102	*CcMYB122*	Segmental	0.235	1.185	0.198	85.10
CcMYB106	*CcMYB102*	Segmental	0.254	4.157	0.061	298.60
	CcMYB101	Segmental	0.346	3.527	0.098	253.35
	CcMYB122	Segmental	0.334	4.018	0.083	288.64
	CcMYB104	Segmental	0.336	4.312	0.078	309.78
CcMYB110	*CcMYB111*	Tandem	0.014	0.083	0.164	5.94
CcMYB115	*CcA*	Segmental	0.369	0.993	0.372	71.32
CcMYB14	*CcMYB16*	Segmental	0.396	2.037	0.195	146.36
CcMYB25	*CcMYB26*	Tandem	0.035	0.136	0.261	9.74
CcMYB31	*CcMYB63*	Tandem	0.137	0.456	0.300	32.76
	CcMYB65	Tandem	0.184	0.529	0.349	37.99
CcMYB32	*CcMYB33*	Segmental	0.134	0.783	0.172	56.24
CcMYB55	*CcMYB54*	Segmental	0.233	2.843	0.082	204.22
CcMYB60-1	*CcMYB60-2*	Segmental	0.061	0.060	1.027	4.29
CcMYB64	*CcMYB31*	Tandem	0.090	0.334	0.269	23.98
	CcMYB63	Tandem	0.115	0.562	0.204	40.35
	CcMYB80	Segmental	0.281	3.748	0.075	269.24
CcMYB90	*CcMYB91*	Tandem	0.052	0.263	0.198	18.93
	CcMYB87	Segmental	0.172	3.742	0.046	268.83
	CcMYB89	Segmental	0.156	0.985	0.159	70.74
	CcMYB86	Segmental	0.347	2.739	0.127	196.74
	CcMYB85	Segmental	0.329	3.449	0.096	247.77
	CcMYB83	Segmental	0.359	3.568	0.101	256.29
	CcMYB84	Segmental	0.381	2.282	0.167	163.91
CcMYBC2	*CcMYBC1*	Segmental	0.158	1.221	0.130	87.75
CcMYBH1	*CcMYBH5*	Segmental	0.181	0.534	0.339	38.39
CcMYBR10	*CcMYBR11*	Tandem	0.046	0.070	0.655	4.99
CcMYR1	*CcMYR2*	Segmental	0.167	0.520	0.322	37.35
CcPHR10	*CcPHR3-2*	Tandem	0.196	0.218	0.899	15.69
	CcPHR4	Tandem	0.172	0.264	0.651	18.96
CcPHR10	*CcPHR3-1*	Tandem	0.240	0.333	0.721	23.94
CcPHR3-1	*CcPHR3-2*	Tandem	0.052	0.098	0.528	7.03
CcPHR4	*CcPHR3-2*	Tandem	NA	NA	NA	–
	CcPHR3-1	Tandem	0.044	0.107	0.415	7.69
	CcPHR5	Tandem	0.085	0.138	0.619	9.89
CcRAD3	*CcRAD6*	Segmental	0.144	1.636	0.088	117.50
	CcRAD4	Segmental	0.288	3.758	0.077	269.94
	CcRAD5	Segmental	0.213	2.091	0.102	150.25

(Continued)

TABLE 1 | Continued

C. chinense MYB name	Duplicated *MYB* pair	Duplication type	Ka	Ks	Ka/Ks	Diversion time (T = Ks/2λ) MYA
CcRAD4	*CcRAD6*	Segmental	0.188	3.621	0.052	260.12
	CcRAD3	Segmental	0.288	3.758	0.077	269.94
CcRAD5	*CcRAD6*	Tandem	0.078	0.143	0.549	10.25
	CcRAD3	Segmental	0.213	2.091	0.102	150.25
CcRL1	*CcRAD2*	Tandem	0.418	2.159	0.194	155.10
CcRL2	*CcRL1*	Segmental	0.210	1.744	0.120	125.30
CcRL3	*CcRAD2*	Segmental	0.225	4.043	0.056	290.42
CcSRM1	*CcSRM2*	Segmental	0.464	1.053	0.440	75.63
CcTRF3	*CcTRF4*	Segmental	0.066	0.107	0.617	7.70
CcDIV7	*CcSRM1*	Segmental	0.456	2.903	0.157	208.54
CcGLK1-1	*CcGLK1-2*	Segmental	0.000	0.003	0.001	0.24
	CcGLK1-3	Segmental	0.000	0.003	0.001	0.24
	CcGLK1-4	Segmental	0.000	0.003	0.001	0.24
CcGLK1-2	*CcGLK1-3*	Segmental	NA	NA	NA	–
	CcGLK1-4	Segmental	NA	NA	NA	–
CcGLK1-4	*CcGLK1-3*	Segmental	NA	NA	NA	–
CcMYB102	*CcMYB122*	Segmental	0.235	1.185	0.198	85.10
CcMYB106	*CcMYB102*	Segmental	0.254	4.157	0.061	298.60
	CcMYB101	Segmental	0.346	3.527	0.098	253.35
	CcMYB122	Segmental	0.334	4.018	0.083	288.64
	CcMYB104	Segmental	0.336	4.312	0.078	309.78
CcMYB110	*CcMYB111*	Tandem	0.014	0.083	0.164	5.94
CcMYB115	*CcA*	Segmental	0.369	0.993	0.372	71.32
CcMYB14	*CcMYB16*	Segmental	0.396	2.037	0.195	146.36
CcMYB25	*CcMYB26*	Tandem	0.035	0.136	0.261	9.74
CcMYB31	*CcMYB63*	Tandem	0.137	0.456	0.300	32.76
	CcMYB65	Tandem	0.184	0.529	0.349	37.99
CcMYB32	*CcMYB33*	Segmental	0.134	0.783	0.172	56.24
CcMYB55	*CcMYB54*	Segmental	0.233	2.843	0.082	204.22
CcMYB60-1	*CcMYB60-2*	Segmental	0.061	0.060	1.027	4.29
CcMYB64	*CcMYB31*	Tandem	0.090	0.334	0.269	23.98
	CcMYB63	Tandem	0.115	0.562	0.204	40.35
	CcMYB80	Segmental	0.281	3.748	0.075	269.24
CcMYB90	*CcMYB91*	Tandem	0.052	0.263	0.198	18.93
	CcMYB87	Segmental	0.172	3.742	0.046	268.83
	CcMYB89	Segmental	0.156	0.985	0.159	70.74
	CcMYB86	Segmental	0.347	2.739	0.127	196.74
	CcMYB85	Segmental	0.329	3.449	0.096	247.77
	CcMYB83	Segmental	0.359	3.568	0.101	256.29
	CcMYB84	Segmental	0.381	2.282	0.167	163.91
CcMYBC2	*CcMYBC1*	Segmental	0.158	1.221	0.130	87.75
CcMYBH1	*CcMYBH5*	Segmental	0.181	0.534	0.339	38.39
CcMYBR10	*CcMYBR11*	Tandem	0.046	0.070	0.655	4.99
CcMYR1	*CcMYR2*	Segmental	0.167	0.520	0.322	37.35
CcPHR10	*CcPHR3-2*	Tandem	0.196	0.218	0.899	15.69
	CcPHR4	Tandem	0.172	0.264	0.651	18.96
CcPHR10	*CcPHR3-1*	Tandem	0.240	0.333	0.721	23.94
CcPHR3-1	*CcPHR3-2*	Tandem	0.052	0.098	0.528	7.03
CcPHR4	*CcPHR3-2*	Tandem	NA	NA	NA	–
	CcPHR3-1	Tandem	0.044	0.107	0.415	7.69
	CcPHR5	Tandem	0.085	0.138	0.619	9.89

(Continued)

TABLE 1 | Continued

C. chinense MYB name	Duplicated *MYB* pair	Duplication type	Ka	Ks	Ka/Ks	Diversion time (T = Ks/2λ) MYA
CcRAD3	*CcRAD6*	Segmental	0.144	1.636	0.088	117.50
	CcRAD4	Segmental	0.288	3.758	0.077	269.94
	CcRAD5	Segmental	0.213	2.091	0.102	150.25
CcRAD4	*CcRAD6*	Segmental	0.188	3.621	0.052	260.12
	CcRAD3	Segmental	0.288	3.758	0.077	269.94
CcRAD5	*CcRAD6*	Tandem	0.078	0.143	0.549	10.25
	CcRAD3	Segmental	0.213	2.091	0.102	150.25
CcRL1	*CcRAD2*	Tandem	0.418	2.159	0.194	155.10
CcRL2	*CcRL1*	Segmental	0.210	1.744	0.120	125.30
CcRL3	*CcRAD2*	Segmental	0.225	4.043	0.056	290.42
CcSRM1	*CcSRM2*	Segmental	0.464	1.053	0.440	75.63
CcTRF3	*CcTRF4*	Segmental	0.066	0.107	0.617	7.70
	CcMYB104	Segmental	0.336	4.312	0.078	309.78
CcMYB110	*CcMYB111*	Tandem	0.014	0.083	0.164	5.94
CcMYB115	*CcA*	Segmental	0.369	0.993	0.372	71.32
CcMYB14	*CcMYB16*	Segmental	0.396	2.037	0.195	146.36
CcMYB25	*CcMYB26*	Tandem	0.035	0.136	0.261	9.74
CcMYB31	*CcMYB63*	Tandem	0.137	0.456	0.300	32.76
	CcMYB65	Tandem	0.184	0.529	0.349	37.99
CcMYB32	*CcMYB33*	Segmental	0.134	0.783	0.172	56.24
CcMYB55	*CcMYB54*	Segmental	0.233	2.843	0.082	204.22
CcMYB60-1	*CcMYB60-2*	Segmental	0.061	0.060	1.027	4.29
CcMYB64	*CcMYB31*	Tandem	0.090	0.334	0.269	23.98
	CcMYB63	Tandem	0.115	0.562	0.204	40.35
	CcMYB80	Segmental	0.281	3.748	0.075	269.24
CcMYB90	*CcMYB91*	Tandem	0.052	0.263	0.198	18.93
	CcMYB87	Segmental	0.172	3.742	0.046	268.83
	CcMYB89	Segmental	0.156	0.985	0.159	70.74
	CcMYB86	Segmental	0.347	2.739	0.127	196.74
	CcMYB85	Segmental	0.329	3.449	0.096	247.77
	CcMYB83	Segmental	0.359	3.568	0.101	256.29
	CcMYB84	Segmental	0.381	2.282	0.167	163.91
CcMYBC2	*CcMYBC1*	Segmental	0.158	1.221	0.130	87.75
CcMYBH1	*CcMYBH5*	Segmental	0.181	0.534	0.339	38.39
CcMYBR10	*CcMYBR11*	Tandem	0.046	0.070	0.655	4.99
CcMYR1	*CcMYR2*	Segmental	0.167	0.520	0.322	37.35
CcPHR10	*CcPHR3-2*	Tandem	0.196	0.218	0.899	15.69
	CcPHR4	Tandem	0.172	0.264	0.651	18.96
CcPHR10	*CcPHR3-1*	Tandem	0.240	0.333	0.721	23.94
CcPHR3-1	*CcPHR3-2*	Tandem	0.052	0.098	0.528	7.03
CcPHR4	*CcPHR3-2*	Tandem	NA	NA	NA	–
	CcPHR3-1	Tandem	0.044	0.107	0.415	7.69
	CcPHR5	Tandem	0.085	0.138	0.619	9.89
CcRAD3	*CcRAD6*	Segmental	0.144	1.636	0.088	117.50
	CcRAD4	Segmental	0.288	3.758	0.077	269.94
	CcRAD5	Segmental	0.213	2.091	0.102	150.25
CcRAD4	*CcRAD6*	Segmental	0.188	3.621	0.052	260.12
	CcRAD3	Segmental	0.288	3.758	0.077	269.94
CcRAD5	*CcRAD6*	Tandem	0.078	0.143	0.549	10.25
	CcRAD3	Segmental	0.213	2.091	0.102	150.25
CcRL1	*CcRAD2*	Tandem	0.418	2.159	0.194	155.10
CcRL2	*CcRL1*	Segmental	0.210	1.744	0.120	125.30

(Continued)

TABLE 1 | Continued

C. chinense MYB name	Duplicated *MYB* pair	Duplication type	Ka	Ks	Ka/Ks	Diversion time (T = Ks/2λ) MYA
CcRL3	*CcRAD2*	Segmental	0.225	4.043	0.056	290.42
CcSRM1	*CcSRM2*	Segmental	0.464	1.053	0.440	75.63
CcTRF3	*CcTRF4*	Segmental	0.066	0.107	0.617	7.70
	CcPHR3-1	Tandem	0.044	0.107	0.415	7.69
	CcPHR5	Tandem	0.085	0.138	0.619	9.89
CcRAD3	*CcRAD6*	Segmental	0.144	1.636	0.088	117.50
	CcRAD4	Segmental	0.288	3.758	0.077	269.94
	CcRAD5	Segmental	0.213	2.091	0.102	150.25
CcRAD4	*CcRAD6*	Segmental	0.188	3.621	0.052	260.12
	CcRAD3	Segmental	0.288	3.758	0.077	269.94
CcRAD5	*CcRAD6*	Tandem	0.078	0.143	0.549	10.25
	CcRAD3	Segmental	0.213	2.091	0.102	150.25
CcRL1	*CcRAD2*	Tandem	0.418	2.159	0.194	155.10
CcRL2	*CcRL1*	Segmental	0.210	1.744	0.120	125.30
CcRL3	*CcRAD2*	Segmental	0.225	4.043	0.056	290.42
CcSRM1	*CcSRM2*	Segmental	0.464	1.053	0.440	75.63
CcTRF3	*CcTRF4*	Segmental	0.066	0.107	0.617	7.70

The type of duplication, Ka (rate of non-synonymous substitution per non-synonymous site), Ks (rate of synonymous substitution per synonymous site), Ka/Ks ratio, and the duplication time in million years ago (MYA) are mentioned (T = Ks/2λ; where λ = 6.96 × 10^{-9} years).

A total of 236 and 238 *MYB* genes were expressed in at least one tissue in *C. chinense* and *C. annuum*, respectively. Based on the expression patterns of these genes in the fruit developmental stages, the co-expression analysis identified 15 and 20 different clusters of *MYB* genes in *C. chinense* and *C. annuum*, respectively (**Figure 5**). Also, several *MYB* genes, such as *CcMYB16*, *CcMYB28*, *CcMYB100*, *CcA, CcDIV4, CcMYB46,* and *CcMYB74,* were co-expressed with *DFR* and *CHS* from the anthocyanin/flavonoid pathway, while *CcDIV1*, *CcMYB4*, *CcMYB31*, *CcMYB52*, *CcMYB86*, *CcMYB108*, *CcMYBR6*, and *CcARR11* were co-expressed with *Kas*, *FatA*, and *BCKDH* from the capsaicinoid biosynthesis pathway. Moreover, the *CcMYB10*, *CcMYB82*, *CcMYB102*, *CcMYB1R1*, and *CcRVE4* genes showed similar expression patterns with genes related to fruit shape and size (**Figure 5**). Further analysis revealed a total of 54 DE *CcMYBs* (adjusted $p < 0.01$) in *C. chinense*, 36 in MG compared with EG (12 upregulated and 24 downregulated), 66 in Br compared with EG (20 upregulated and 46 downregulated.), and 50 in MG compared with Br (34 upregulated and 16 downregulated; **Figure 6**). While a total of 81 *CaMYBs* were DEGs (adjusted $p < 0.01$) in *C. annuum*, 39 and 42 were DEs in MG and Br compared with the EG fruit stage and 13 were DEGs in MG with respect to the Br fruit stage (**Figure 6**). Furthermore, we analyzed the expression patterns of *MYB* genes in *C. chinense* and *C. annuum* using the transcriptome data and compared them using representative heatmaps (**Figure 7A**). Most of the MYB genes showed similar expression patterns, but several MYB genes showed contrasting expression levels in the two *Capsicum* spp. For instance, *CcMYBR12* shows a higher expression in *C. chinense* and a moderate expression in *C. annuum* in all three fruit stages. *CcPHR8* is highly expressed throughout the fruit developmental stages in *C. chinense,* while its expression decreases from high to low during fruit development in *C. annuum*. *CcMYB31* is exclusively highly expressed in the MG stage in *C. chinense*, moderately expressed in EG, and lowly expressed in the Br fruit stage, while its homolog in *C. annuum* showed a low expression in the EG stage and a negligible expression in the rest of the two stages. *MYB48* showed a higher expression throughout the fruit developmental stages in *C. annuum* but only slightly higher in MG in *C. chinense* than the rest of the fruit stages.

The expression of 24 MYB genes showing DE in transcriptome data was validated by qRT-PCR analysis (**Figure 7B**). The *MYB* genes showed 74–83% similarity in their qRT-PCR expression profile with the RNA-seq data. Eight MYB genes-*CcMYB100* (BC332_00785), *CcMYB16 (*BC332_11900), *CcETC3* (BC332_24253), *CcMYBR12* (BC332_30379), *CcMYB106* (BC332_27082), *CcMYB3R-1* (BC332_21354), *CcMYB3* (BC332_04434), and *CcMYB31* (BC332_19185) showed a significantly higher expression in one or more of the fruit stages of highly pungent *C. chinense* compared with lowly pungent *C. annuum*. The expression of *CcMYBR12* increased from EG to the Br fruit stage in both the *Capsicum spp.,* but the level of expression remains higher in *C. chinense* throughout. *CcMYBR12* expression shows a high similarity between RNA seq and qRT-PCR data, except that it was highest in the Br stage in qRT-PCR and the MG stage in RNA-seq data. *CcMYB16* and *CcETC3* showed two to four times higher expression in *C. chinense*. The transcriptome data for *CcMYB16* suggest the same for the MG and Br stages; however, it was the opposite for the EG stage. *CcETC3* showed similar expression patterns in the transcriptome data. *CcMYB106* and *CcMYB100* showed expression only in the EG stage with a slightly higher level of expression in *C. chinense* as compared with that of *C. annuum*.

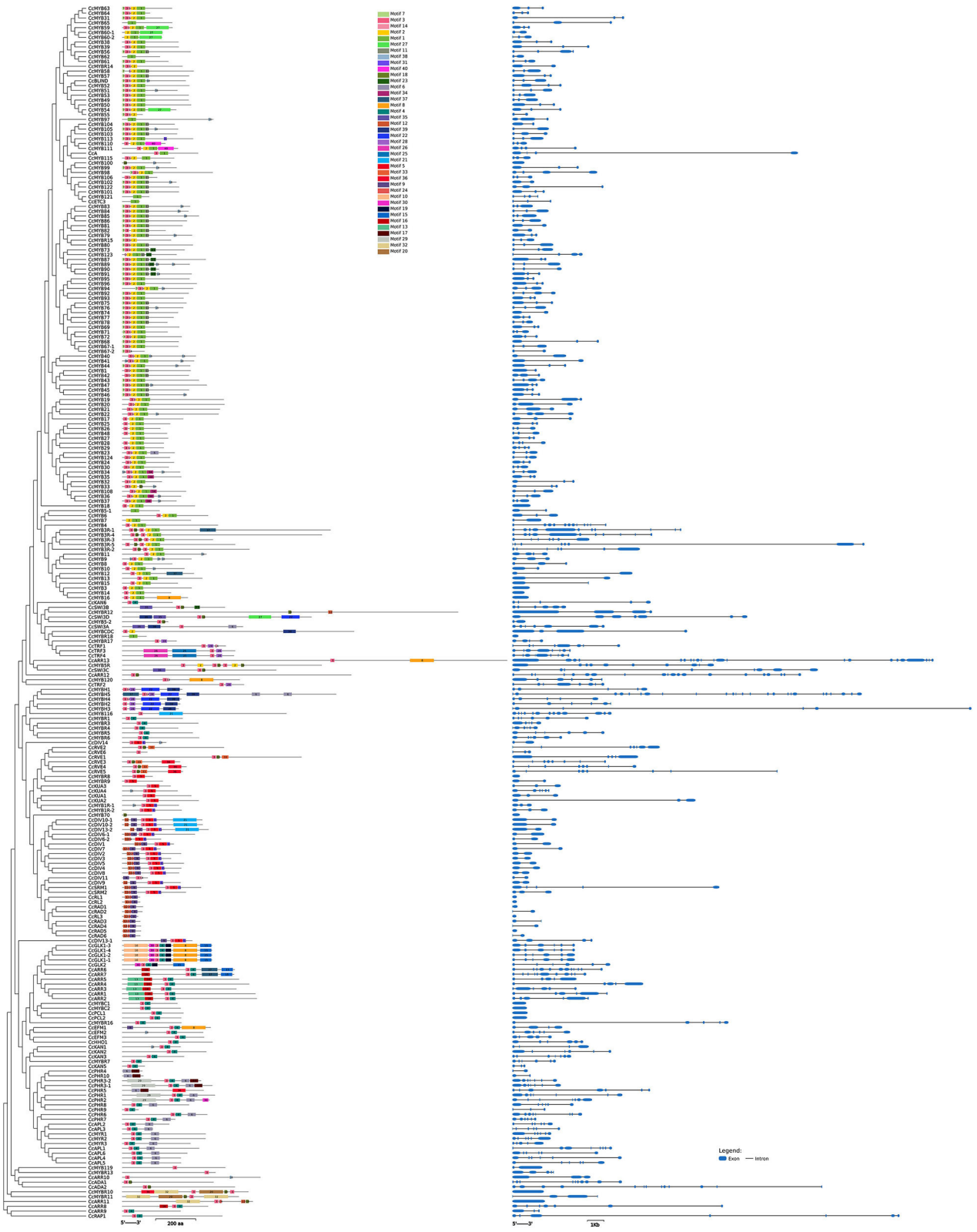

FIGURE 2 | Conserved amino acid motifs and structure of *CcMYB* genes. The motifs are differently colored with a specific number. The exons are shown in blue blocks, while the connecting lines in between represent intronic regions in kb (kilobase).

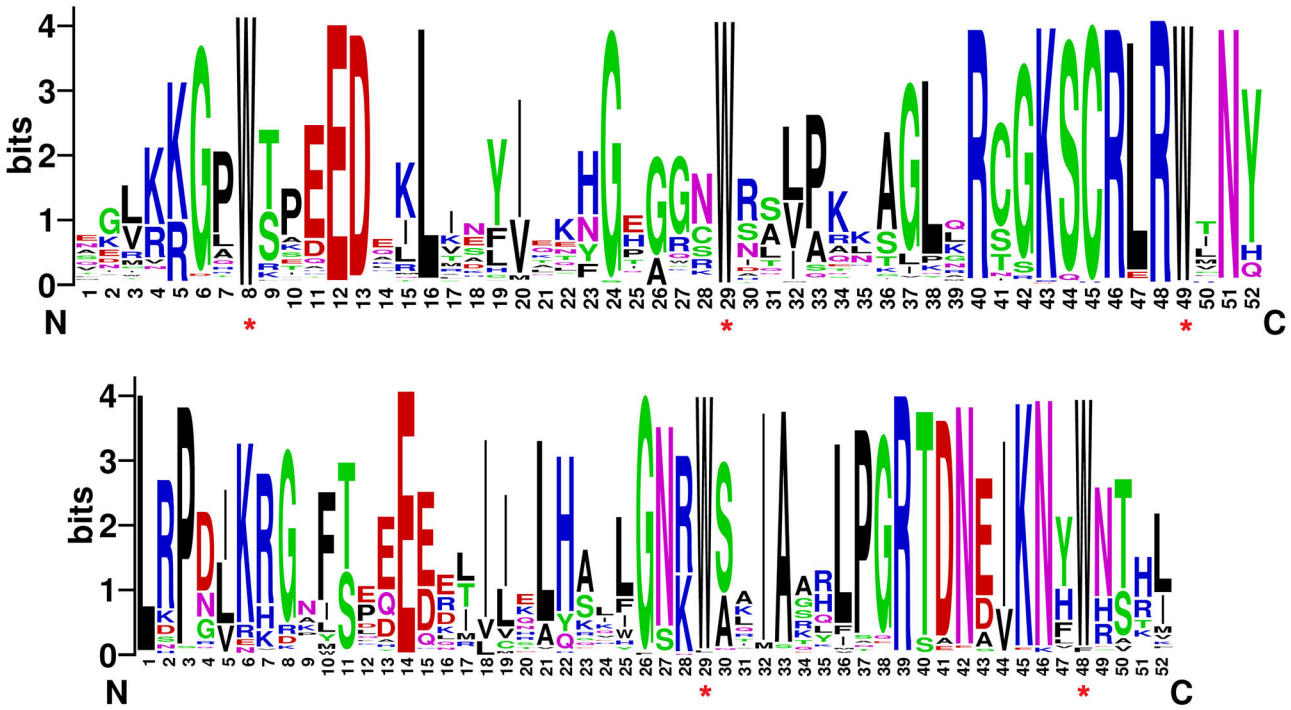

FIGURE 3 | Sequence logo of **(A)** R2 and **(B)** R3 MYB repeats in *C. chinense*.

CcMYB100 showed three times higher expression in the EG fruit of *C. chinense* as compared with that of *C. annuum,* and *CcMYB31* showed approximately two times higher expression in the MG fruit stage in *C. chinense* compared with that of *C. annuum*. *CcMYB3R-1* was 2-3 times highly expressed in all the three fruit stages of *C. chinense*, while *CcMYB3* showed a lower expression in EG and MG, and a higher expression in the Br stage of *C. chinense* compared with those of *C. annuum*.

Protein Structure Prediction of *C. chinense* R2R3 *MYB* Genes

The 126 CcMYBs having the R2R3 MYB domain were analyzed for their secondary and tertiary structures using the Phyre2 server (**Supplementary Figure 4**). The best models for CcMYB proteins showed 20–68% identity and 98.6–100% query coverage with their template sequences. The majority of CcMYB (116) genes were observed to have high coverage and similarity with the c6kksA protein model/template of Arabidopsis R2R3 type MYB2 TF (*WEREWOLF*, *WER*). For instance, both CcMYB106 and CcMYB113 showed 68% similarity, while CcMYB67-2 was just 21% similar to this template sequence. The remaining seven CcMYBs were modeled with different protein templates, out of which two, namely, CcMYB3R-4 and CcMYB5R, were modeled with a template c1h88C of ternary protein-DNA complex1 of MYB TF, while CcMYB70, CcMYB5-1, and CcMYB100 were modeled with a template named d1mbja of c-MYB DNA-binding domain repeat 3. Two MYBs, i.e., CcMYB121 and CcMYB5-2, were modeled with templates d1gv2a2 and d1h8ac1, respectively, and were related to the Myb/SANT DNA-binding domain family and three MYBs, CcMYB116, CcMYB119, and CcMYB120, with c2yqkA, d2crga1, and c5ylzJ, respectively (**Supplementary Table 6**). The CcMYBs contained 21–61% α-helix in their secondary structure, while 1–6% β-strands were predicted in only 20 CcMYBs (**Supplementary Table 6**). Overall, the modeled 3D structures suggested the helix-turn-helix structure similarity of CcMYB proteins with already known Arabidopsis models and were highly reliable.

Synteny and Gene Duplication Analysis

We analyzed the synteny and collinearity of five Solanaceae and Arabidopsis genomes with *C. chinense* (**Figure 8**) and identified a total of 717 conserved syntenic segments (CSSs) in all the species analyzed, ranging from 0.02 to 33.56 Mbp (million base pair) in size, which have at least one *CcMYB* gene in them along with other protein coding genes (**Supplementary Table 7**, **Table 2**). The highest number, i.e., 171 and 176 CSSs, was shared with *C. annuum* and *C. baccatum*, respectively. In these CSSs, 203 and 168 unique *CcMYBs* were homologous with *MYBs* of *C. annuum* and *C. baccatum*. Among the *Capsicum* species, most of the CSSs were present on the same chromosomes and had an order of genes similar to that of *C. chinense,* but few were found to be diverged (**Table 2**). For instance, 47 CSSs were present on different chromosomes, and 75 had a reversed gene order in *C. annuum* with respect to *C. chinense*. In particular, a CSS, harboring *MYB44* along with other genes on chr 1, showed a reverse order of genes in *C. annuum*. The size of this CSS is 1.607 Mbp in *C. chinense* and 1.662 Mbp in *C. annuum*. Another CSS of 1.9 Mbp in *C. chinense* and of 3.17 Mbp in *C. annuum*

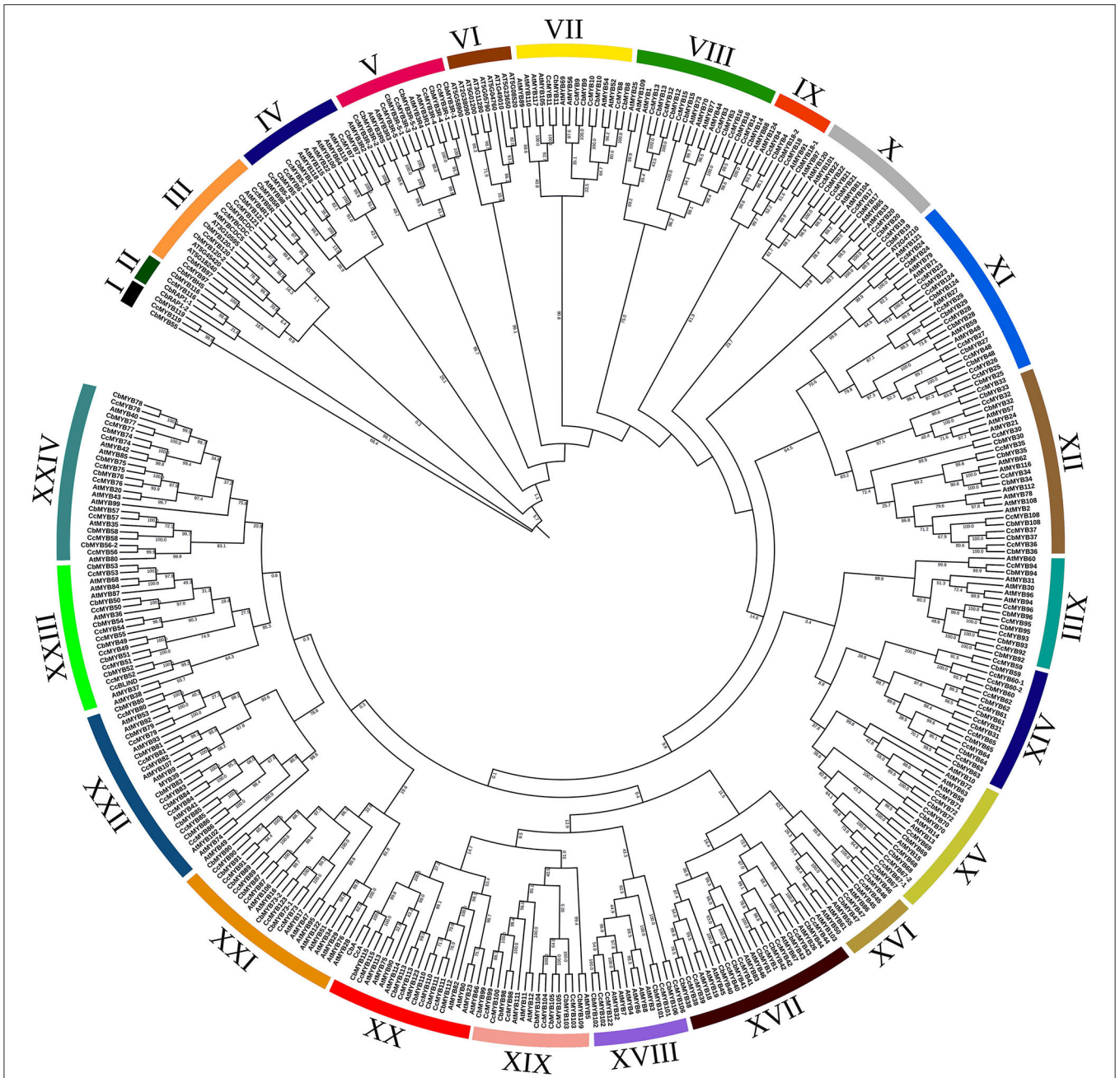

FIGURE 4 | Phylogenetic tree constructed using R2R3 MYBs from *C. chinense*, *C. baccatum,* and Arabidopsis. The tree was divided into 24 subgroups, and bootstrap values are shown on the branches of the tree.

were present on different chromosomes, chr 7 in *C. chinense* and on chr 9 in *C. annuum* and had a reverse order of genes. In *C. baccatum*, a CSS with *CcMYB105* and *CbMYB104* and other genes was 2.29 Mbp in size on chr 1, while it was 2.47 Mbp in *C. chinense* on chr 8 (**Figure 9A**, **Tables 2**, **3**).

As expected, the genomes of potato (154), tomato (108), and Arabidopsis (93) shared a lesser number of CSSs with the C. *chinense* genome. A number of CSSs were spread on different chromosomes compared with *C. chinense*, i.e., 55 in tomato, 74 in potato, and 86 in Arabidopsis, which was expected as they are more diverged compared with different species of *Capsicum*.

To study the effects of duplication events on the expansion of the *MYB* gene family, *Capsicum* MYB homolog proteins were identified in *C. annuum*, *C. baccatum*, tomato, potato, and Arabidopsis (**Figure 9B**). We found 758 pairs of MYB duplicates across the genomes of five species (**Supplementary Table 8**). Of these, our analysis showed 435 MYB duplicate pairs to be under purifying selection, 23 pairs under neutral selection, and 47

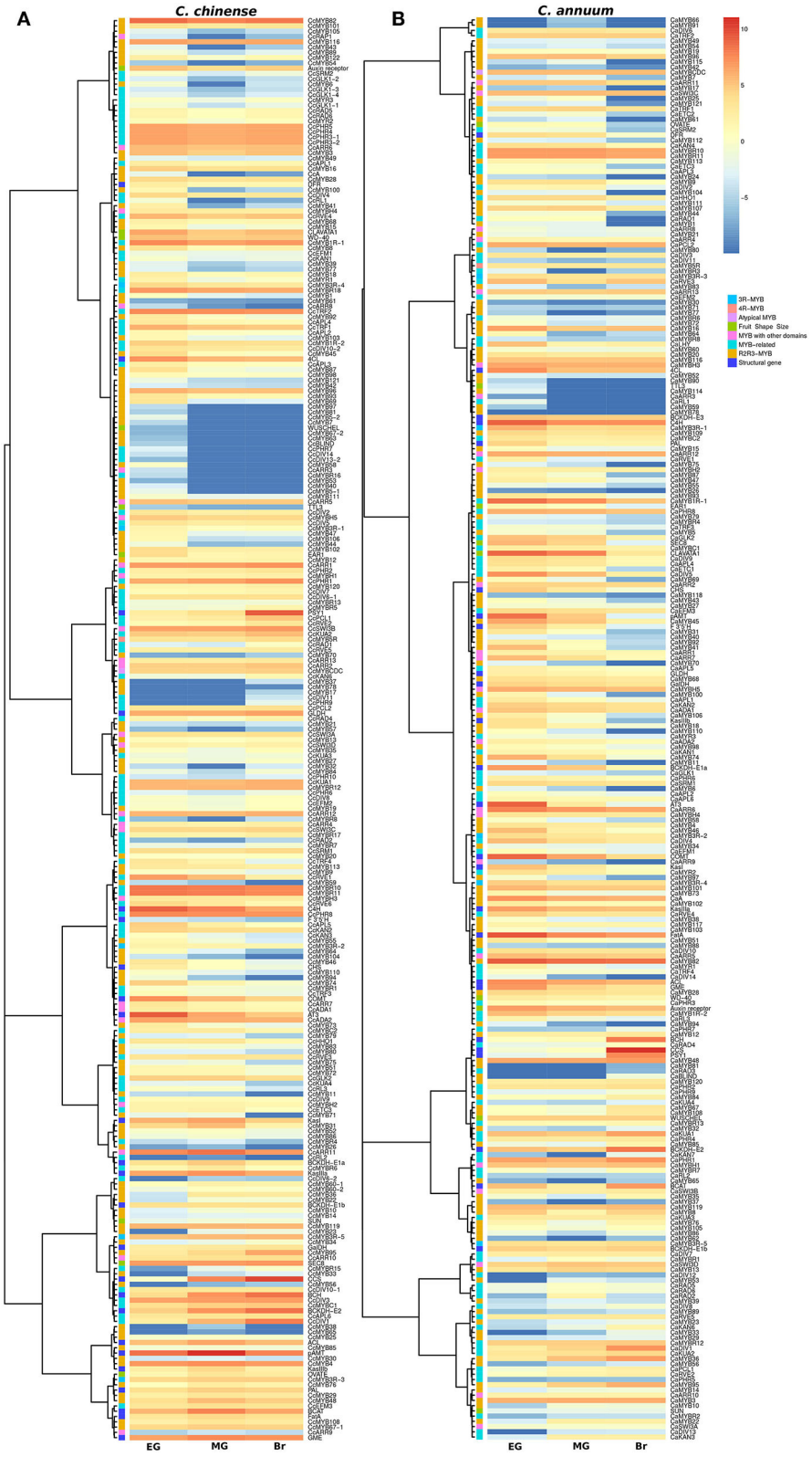

FIGURE 5 | Expression profiles of *MYB* genes in **(A)** *C. chinense*, **(B)** *C. annuum* at the three different fruit developmental stages. The heatmap color scale represents the log2 fragments per kilobase of transcript per Million mapped reads (FPKM) expression of *MYB* genes.

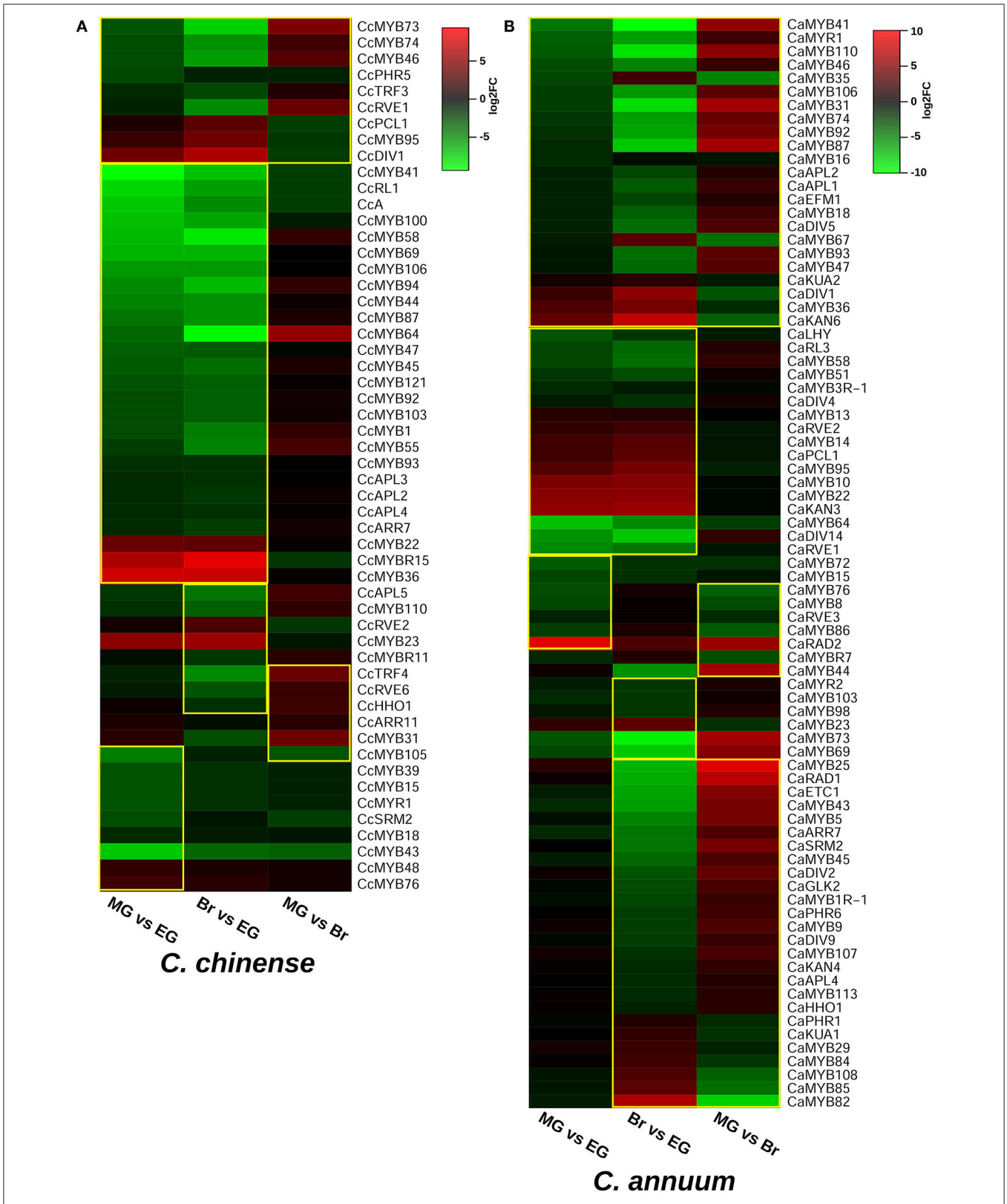

FIGURE 6 | Differentially expressed *MYB* genes in **(A)** *C. chinense* and **(B)** *C. annuum*. Significant DEG *MYBs* are enclosed within the yellow box.

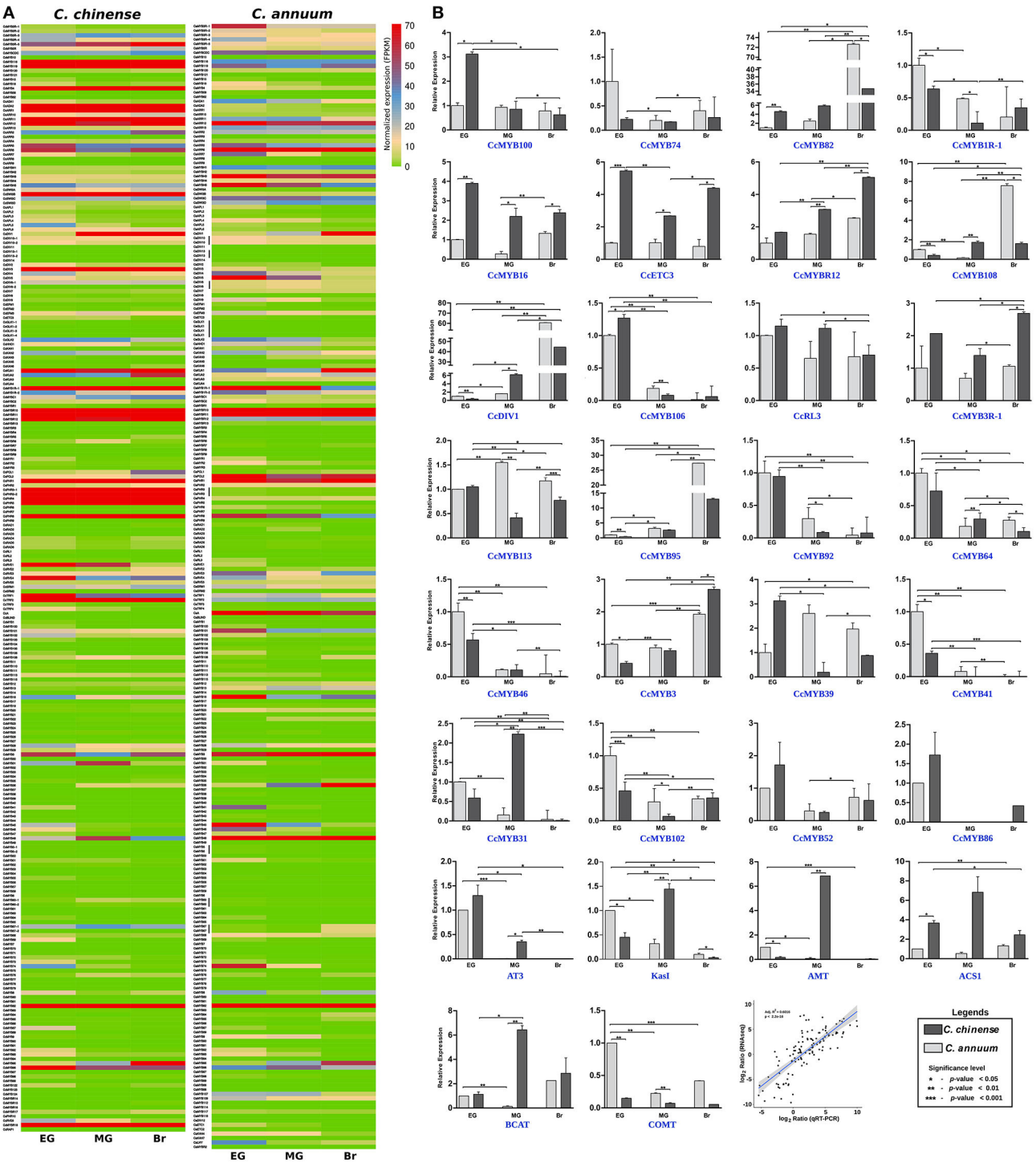

FIGURE 7 | Ribonucleic acid sequencing (RNAseq) and quantitative real-time (qRT)-PCR expression data of *Capsicum MYBs* at the early green (EG), mature green (MG), and breaker (Br) fruit development stages. **(A)** Normalized FPKM expression of *C. chinense MYBs* and their *C. annuum* homologs by RNAseq, and **(B)** relative expression of 24 *CcMYB* genes and capsaicinoid biosynthesis pathway genes (*AT3*, *KasI*, *pAMT*, *ACS1*, *BCAT,* and *COMT*) between *C. chinense* (Bhut Jolokia; Acc-Cc74) and *C. annuum* (Acc-Ca18) by qRT-PCR in the three fruit developmental stages and correlation between RNAseq and qRT-PCR expression data. A Student *t*-test was performed to calculate the significant difference of expression. The significance level was represented as $^{***}p < 0.001$; $^{**}p < 0.01$, and $^{*}p < 0.05$).

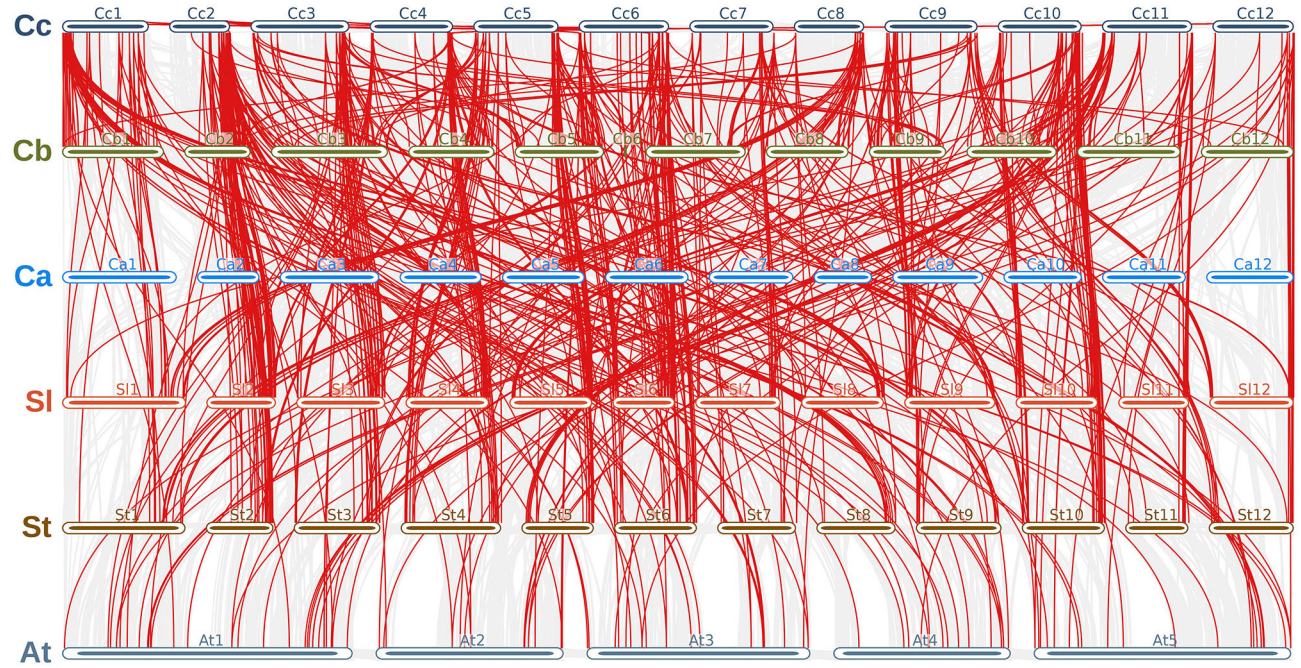

FIGURE 8 | Synteny and collinearity of the *C. chinense* (*CcMYB*) genes with the *Capsicum annuum (Ca)*, *C. baccatum (Cb)*, *Solanum lycoperisucum (Sl)*, *Solanum tuberosum (St)*, *Solanum melongena (Sm)*, and *Arabidopsis thaliana (At)* genomes. Gray lines in the background show collinear segments, while red lines represent syntenic *MYB* gene pairs with respect to the *C. chinense* genome.

TABLE 2 | Number of conserved syntenic segments (CSSs), with at least one *MYB* gene in them, their size in million base pairs (Mbp), and their diversification in different Solanaceae genomes and *Arabidopsis thaliana*.

Species	*C. annuum*	*C. baccatum*	*S. lycopersicum*	*S. tuberosum*	*A. thaliana*	*C. chinense*
No. of CSSs	171	176	108	154	93	15
Size (Mbp)	0.06–28.79	0.07–33.56	0.05–2.75	0.06–3.07	0.02–0.89	0.29–10.95
CSSs on diff chr	47	71	55	74	86	12
CSSs with reverse gene order	75	76	42	65	37	5
MYB genes in CSSs	203	168	90	131	77	27

pairs under positive selection. Among these, 225, 213, 125, 79, and 23 CcMYB proteins were orthologous to *C. annuum, C. baccatum,* potato, tomato, and Arabidopsis MYBs, respectively. The value of Ka/Ks ranged from 0.001 to 50 in the homolog pairs. Homologous pairs CcMYB25 and CaMYB25, and CcMYB90 and CaMYB90 were under positive selection. Among others, CcMYB31 and CaMYB31, and CcMYB3 and SlMYB73 were under purifying selection. The average divergence time for CcMYB orthologs with tomato was 37.1 MYA, while that with potato was 33.56 MYA (**Supplementary Table 8**).

Identification of Simple Sequence Repeat (SSR) Motifs

We identified SSRs in the *C. chinense MYB* genes and their 1.5-kb upstream promoter, which can be used as molecular markers in future *Capsicum* breeding programs. A total of 169 SSRs were identified in the *C. chinense MYBs*. Out of these, 114 were gene-based from 77 *C. chinense MYB* genes. The remaining 55 SSRs were in the 1.5-Kb upstream regions from the TSS of 49 *C. chinense MYB* genes. Among all the SSRs, the dinucleotide repeats were the most common, i.e., 70.4% in *C. chinense*, followed by tri-, tetra-, penta-, and hexanucleotide repeats. The most common dinucleotide repeat was "AT," with a frequency of 42. The maximum size of a SSR motif in *C. chinense* was seven repeats of hexanucleotide [ATTTTA] in *CcMYBR11*. The primer sequences and the expected amplicon length of all the SSR repeat motifs in the *Capsicum* species are given in **Supplementary Table 9**.

DISCUSSION

MYB genes are important TFs involved in the regulation of several biological, developmental, and metabolic processes in plant species such as *Capsicum* (Liu et al., 2015; Xu et al.,

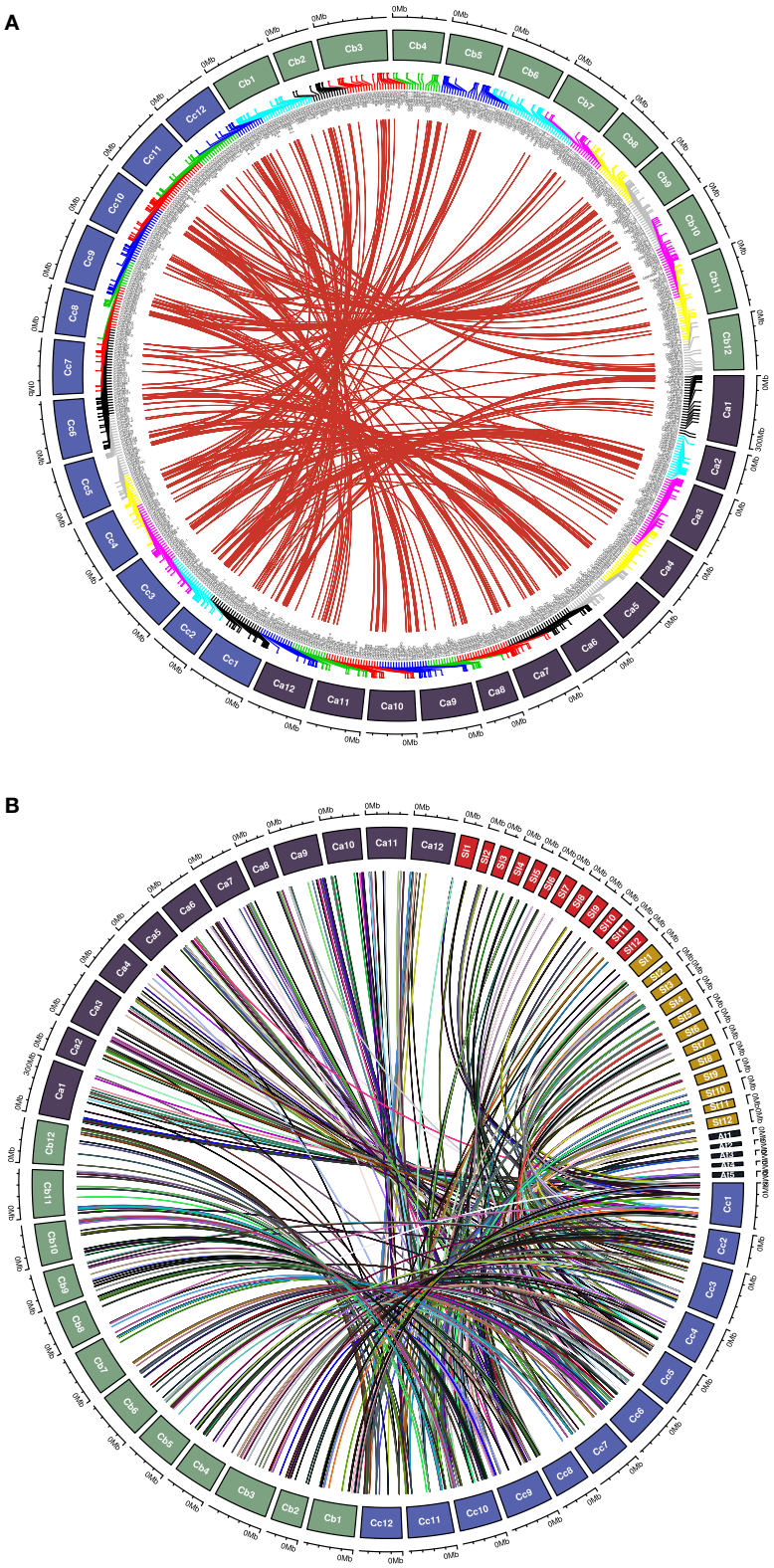

FIGURE 9 | Chord diagram representing **(A)** syntenic *CcMYB* gene pairs with *C. annuum* and *C. baccatum MYBs* and **(B)** *CcMYB* homologs in *C. annuum*, tomato (Sl), potato (St), and Arabidopsis (At) across their respective chromosomes.

TABLE 3 | Chromosome-wise distribution and the number of collinear *CcMYB* genes on the same chromosomes as *C. chinense* in Solanaceae members.

chr No.	*C. chinense*	*C. annuum*		*C. baccatum*		*S. lycopersicum*		*S. tuberosum*	
	MYB count	*MYB* Count	Collinear	*MYB* count	*Collinear*	*MYB count*	Collinear	*MYB* count	Collinear
chr01	16	16	16	18	18	8	4	8	2
chr02	18	15	15	12	12	10	10	14	14
chr03	18	15	15	15	6	8	6	14	9
chr04	10	12	9	7	6	9	5	13	7
chr05	20	17	17	17	10	11	5	16	7
chr06	28	20	19	2	2	11	10	17	16
chr07	12	12	8	8	7	5	3	7	6
chr08	9	8	6	8	7	5	2	7	2
chr09	18	11	11	14	12	6	4	7	4
chr10	21	14	14	12	12	8	7	14	14
chr11	15	6	6	11	11	5	3	6	2
chr12	10	9	6	9	8	3	1	8	3
Scaffolds	9	9	–	32	3	0	–	0	–
Total	**204**	**164**	**142**	**165**	**114**	**89**	**60**	**131**	**86**

2015b; Sun et al., 2019; Cao et al., 2020). Although previously reported in *C. annuum*; in *C. chinense*, however, being one of the most important species, including the naturally occurring highest pungency containing *Capsicum* genotype, Bhut jolokia/ghost chili, no report of *MYB* gene identification was reported. Additionally, in this study, we have identified *MYB* genes in *C. baccatum* (240). Previously, Arce-Rodríguez et al. (2021) reported 235 *MYBs* in *C. annuum,* but this study identified additional 10 new *MYB* genes (6 atypical R2R3 *MYBs* and 4 *MYBs* with other domains). The greater number of *MYB* genes in *Capsicum* might be due to genome expansion. Alternatively, it may also mean the deletion or loss of genes from other lineages. Despite a large number of *MYB* genes in *Capsicum*, there are few chances that their functions are redundant but more likely to overlap in their functionalities (Jin and Martin, 1999) or mask the functions of each other. The large size of the *MYB* gene family in plants can be attributed to the high rates of duplication and retention of duplicate copies compared with other TF families (Shiu et al., 2005). The retention of such large numbers of *MYB* genes, during evolution, in the members of Solanaceae indicates their positive selection and acquisition of new functions. R2R3-MYBs were the most common type of *MYBs* in other *Capsicum* like other plant species (Katiyar et al., 2012; Du et al., 2015; Li et al., 2019). The greater number of R2R3-MYBs in plants suggests their selective amplification and expansion after the loss of R1 repeat in ancestral three repeat MYBs (Lipsick, 1996) and can be involved in the plant speciation process. Whether an increase in genome size or the number of genes explains the huge number of *MYB* genes in phylogenetically related genomes or not remains to be explored further with broad sampling strategies.

CcMYBs Within Capsaicinoid QTLs Were Differentially Expressed

The chromosomal distribution of *MYB* genes was found to be random in the three *Capsicum* species. Seven *Capsicum* chromosomes had around 60% of the *MYB* genes (**Figure 1** and **Supplementary Figure 1**). Five *CcMYBs*, *CcMYB101*, *CcMYB46*, *CcMYB6*, *CcPHR8*, and *CcRVE5*, on chr 6 of *C. chinense* and two *CaMYBs*, namely, *CaMYB3* and *CaHHO1*, on chr 2 of *C. annuum* were found inside the previously reported capsaicinoid QTLs (Han et al., 2018; Park et al., 2019). Among these, two MYB genes, *MYB3* and *MYB46*, also showed a significant differential expression in the fruit tissues of lowly pungent *C. annuum* and highly pungent *C. chinense* in the qRT-PCR analysis (**Figure 7B**). *MYB46* was also found to be co-expressed with *CHS* in the co-expression analysis, suggesting its possible involvement in the transcriptional regulation of anthocyanin biosynthesis. Other two *MYB* genes in the QTLs, *CcPHR8* and *CaHHO1*, were co-expressed with phenylpropanoid gene *C4H* and anthocyanin biosynthesis gene *DFR* (**Figure 5**), suggesting their potential roles in the respective pathways. The motif identification analysis revealed motifs that are important to impart functional significance to *MYB* proteins. For example, recently, motif-23 was reported to mediate the interaction between *Cucumis sativus* MYB6 and a MYB-related protein CsTRY (Yang et al., 2018). Although functions for most of these motifs are still unknown, they may be involved in protein-protein interactions or other biological roles and are subject to further exploration (Millard et al., 2019). *MYB* genes also respond to environmental and hormonal changes to regulate gene expression during abiotic stress responses in plants (Urao et al., 1996; Abe et al., 1997; Li et al., 2015). Here, the *cis*-element analysis also suggests environmental and hormonal regulatory mechanisms for the *Capsicum MYB* genes. MYB binding motifs (MBSI) and MYB recognition elements (MREs) observed in the promoter regions of *Capsicum MYB* genes may indicate their regulatory roles in flavonol biosynthesis (Solano et al., 1997; Mehrtens et al., 2005). Previously, *MYB14* has been shown to contain MRE alongside AT-rich element, GATA-motif, ARE, Box 4, and circadian in its promoter region, and to be activated by UV-C light in order

to activate stilbene synthesis in *Vitis labrusca* (Bai et al., 2019). Likewise, another *MYB15* that controls basal immunity in *V. quinquangularis* was found to harbor cis-elements like GCN4-motif, MBS, and TCA-element in its promoter region (Luo et al., 2019). The identification of cis-elements may aid in deciphering regulatory networks and may further lead to the isolation and characterization of corresponding TFs.

Phylogenetic Analysis Showed Functional Importance of *Capsicum MYBs*

The phylogenetic relationships revealed in the MYB protein sequences can be utilized to relate their function. For instance, *CcMYB3, CcMYB12, CcMYB14, CcMYB16,* etc., along with their sister clades in *C. baccatum* and Arabidopsis, form subgroup VIII. *AtMYB1* in the same subgroup is involved in pollen development, and *AtMYB44*, *AtMYB73*, and *AtMYB77* are involved in lateral root growth and salinity response (Reňák, 2012; Kim et al., 2013; Zhao et al., 2014). Similar functions may be stipulated for *Capsicum MYBs* of the same subgroup. Subgroup XIII *MYBs* such as *AtMYB30* positively regulate fatty acid biosynthesis as well as hypersensitive cell death response, while subgroup XXIV *MYBs,* such as *AtMYB80,* is, again, important for pollen development and tapetal growth (Raffaele et al., 2008; Phan et al., 2011). Subgroup XXII includes *AtMYB49* and *AtMYB74* that provide tolerance to abiotic stresses like salinity, and *AtMYB102* provides resistance to biotic stresses like insect herbivores (De Vos et al., 2006; Xu et al., 2015a; Zhang et al., 2020). Subgroup X contains *AtMYB33* and *AtMYB101*, which are targets of the miR159 family upon ABA accumulation during seed germination, and are shown to be involved in drought stress (Reyes and Chua, 2007). *CcMYBs* such as *MYB31*, *MYB59*, *MYB60*, *MYB61*, *MYB63*, *MYB64*, and *MYB65* in subgroup XIV did not cluster with any *AtMYBs* (**Figure 4**). This was similar to the results of previous analysis on *C. annuum*, where clades 24 and 25 harbored *Capsicum*-specific *MYBs* (Arce-Rodríguez et al., 2021). Also, the same *Capsicum MYBs* clustered with the tomato and potato *MYB* homologs in a distinct subgroup, suggesting these *MYBs* to be Solanaceae-specific (**Figure 4**, **Supplementary Figure 5**). However, contrary to the findings of Arce-Rodríguez et al. (2021), other *MYBs,* such as *MYB116* and *MYB119*, which were not reported earlier, clustered separately from At*MYBs* and were present in subgroups I and II, respectively (**Figure 4**). *CcMYB98, CcMYB99*, and *CcMYB100* lie in subgroup XIX with *AtMYB11*, *AtMYB12*, and *AtMYB111* that control flavonol glycoside accumulation (Stracke et al., 2010). *CcMYB115* and *CcA* in subgroup XX share a sister clade with *AtMYB75* and *AtMYB90*. *AtMYB75* has been extensively studied for its role in anthocyanin biosynthesis and *AtMYB90* for phenylpropanoid biosynthesis (Borevitz et al., 2000; Teng et al., 2005). *CcMYB46* lies along with *AtMYB61* in subgroup XVI, which functions as a transcriptional control for the development of root, seed, and vascular tissues (Romano et al., 2012). *MYB3Rs* have been shown to regulate the cell cycle like c-MYB in animals by regulating cyclin genes *via* MYB recognition elements (MREs) in cyclin gene promoters *via* the R3 MYB repeat (Ito et al., 2001). One major challenge in the functional characterization of *MYB* genes in *Capsicum* could be the functional redundancy of duplicated *MYBs* (Dubos et al., 2010). The second phylogenetic tree among the *MYBs* in the three *Capsicum* species revealed *C. chinense*- and *C. baccatum*-specific lineages, leading to species-specific phenotypes. Similarly, when we analyzed the evolutionary relationship among *MYBs* between the *Capsicum* species and other Solanaceae family members along with Arabidopsis, *C. chinense*-specific clades having *CcMYBR14* were observed (**Supplementary Figures 3**, **5**). However, the phylogenetic tree has a low coverage sampling and may have omitted the evolutionary processes leading to the emergence of these clades and is less representative of the process of speciation in *Capsicum* species.

Expression Analysis Revealed Eight *MYB* Genes Were Highly Expressed in *C. chinense*

The expression analysis revealed the spatio-temporal expression pattern of *CcMYB* genes in the EG, MG, and Br fruit development stages of two *Capsicum* genotypes (**Figures 6**, **7**). Of the 24 *MYB* genes showing DE in RNA-seq data and validated by qRT-PCR, eight showed a high expression in one or more of the fruit stages in *C. chinense* as compared with those of *C. annuum*. For instance, *CcMYB106* and *CcMYB100* showed a significantly higher expression in the EG fruit in *C. chinense* compared with that of *C. annuum* in both the transcriptome data and qRT analysis. *CcMYB100* is a homolog of *SlMYB12,* which has been reported to regulate the flavonol biosynthesis pathway in tomatoes (Ballester et al., 2010). Moreover, our co-expression analysis also showed that *CcMYB100* clustered with *DFR*, an anthocyanin pathway gene (**Figure 5**). *CcMYB16* and *CcETC3* show a maximum expression in the EG stage, which then decreases in the MG stage and then again increases in the Br stage in both the species. *CcMYB16* is also clustered with *DFR* in the co-expression analysis, suggesting its role in anthocyanin biosynthesis. In the phylogenetic analysis, *CcMYB16* had clustered with *AtMYB44*, *AtMYB70*, *AtMYB77,* etc., in subgroup VIII, which may suggest its role in plant growth and development, and abiotic and biotic stress responses (**Figure 4**; Jung et al., 2008; Shim et al., 2013; Zhao et al., 2014). *CcETC3,* on the other hand, was segregated with *AtETC2*, which plays a crucial role in trichome development and patterning (**Supplementary Figure 3**; Kirik et al., 2004; Hilscher et al., 2009). Here, the overall expression pattern of both the *MYB* genes indicates their possible functions in fruit ripening and development in *Capsicum*. *CcMYB3R-1* showed a similar pattern of expression in both *Capsicum* species; however, the level of expression was higher in *C. chinense*. *AtMYB3R-1*, which lies in the same subgroup V as *CcMYB3R-1*, is known to regulate cell cycle and abiotic stress-responsive genes, suggesting similar functions for *CcMYB3R-1* (**Figure 4**; Dai et al., 2007; Ma et al., 2009; Haga et al., 2011). *CcMYB3* showed a similar pattern of increased expression from EG to Br in both the *Capsicum* species. *CcMYB3* was found within the capsaicinoid QTL and clustered with *AtMYB73* in

subgroup VIII, suggesting its role in capsaicinoid biosynthesis and abiotic stress response (**Figure 4**, **Supplementary Figure 1A**; Kim et al., 2013). Another *MYB* gene, *CcMYB46*, however, present in capsaicinoid QTL, did not show a higher expression in *C. chinense* as well as *C. annuum*. In the co-expression analysis, *CcMYB*46 was also clustered with the anthocyanin pathway gene *CHS* (**Figure 5**). *CcMYB31* showed a significantly higher expression in the MG fruit in *C. chinense* compared with that of *C. annuum* (**Figure 7B**). The qRT expression trends of *CcMYB31* and its homolog *CaMYB31* during fruit development were similar to those of their respective RNA-seq expression data (**Figures 6**, **7A**). *CcMYB31* and *CcMYB48* were coexpressed with CBGs- *KasI* and *PAL*, respectively, in the co-expression analysis. Previously, the expression levels of their homologs, *CaMYB31* and *CaMYB48,* have been reported to correlate with capsaicinoid levels in *C. annuum* (Arce-Rodríguez and Ochoa-Alejo, 2017; Han et al., 2019; Sun et al., 2020). Furthermore, other *CcMYBs*, such as *CcDIV1*, *CcMYB4*, *CcMYB52*, *CcMYB86*, *CcMYB108*, *CcMYBR6*, and *CcARR11* also co-expressed with *Kas*, *FatA*, and *BCKDH* from the capsaicinoid biosynthesis pathway, suggesting their potential role in capsaicinoid biosynthesis regulation. However, the expression of *CcMYB52* and *CcMYB86* did not show a significant difference between *C. chinense* and *C. annuum.* Although *pAMT*, *KasIIIb*, *ACL*, *BCAT*, *PAL,* and *FatA* were grouped into the same major cluster in the co-expression analysis, other capsaicinoid pathway genes, such as *AT3*, clustered along with *COMT* and *BCKDH* with *KasIIIa* in their respective clusters (**Figure 5**). However, previously in *C. annuum*, *AT3*, *pAMT*, *Kas*, and *BCKDH* have been reported to be present in the same co-expression cluster (Arce-Rodríguez et al., 2021), which may be due to the selection of different species/genotypes and distinct fruit developmental stages. Sarpras et al. (2016) have previously reported that capsaicinoid pathway genes are highly expressed in *C. chinense* as compared with *C. annuum* and showed correlations with pungency levels of *C. chinense*, EG (315936 SHU), MG (763411.2 SHU), and Br (925084.8 SHU), and of *C. annuum*, EG (3478.4 SHU), MG (6656 SHU), and Br (7034.4 SHU) (Sarpras et al., 2016, 2019). In our study, several *CcMYB* genes also showed a significantly high expression in highly pungent *C. chinense* compared with lowly pungent *C. annuum,* while some of them co-expressed with capsaicinoid/phenylpropanoid biosynthesis pathway genes, which can be selected for further validation in correlation to pungency regulation (**Figures 5**–**7**). Additionally, we have also analyzed the expression of eight *MYB* genes in four other *Capsicum* accessions (two each from *C. chinense* and *C. annuum*) to further understand their expression pattern in high and low pungent *Capsicum* during fruit development (**Supplementary Figure 6**). Overall, the qRT-PCR expression and co-expression analysis suggested that *CcMYBs* potentially have a diverse role in the regulation of capsaicinoid, phenylpropanoid, and anthocyanin biosynthesis (**Supplementary Figure 6**). Furthermore, *CcMYB10*, *CcMYB82*, *CcMYB1R-1*, *CcRVE4*, and *CcMYB102* were co-expressed with fruit shape and size genes like *Auxin receptor*, *WD-40*, *SUN,* and *EAR1,* suggesting their potential roles in fruit development (**Figure 5**).

Homology Modeling Suggested a High Structural Similarity of R2R3 CcMYBs With Arabidopsis R2R3 MYB Domain

With only sparse 3D structures characterized compared with increasingly known protein sequences, there is a massive need for the prediction of protein structures in order to bridge this ever-widening gap. In the absence of experimentally determined protein structures, computational tools for protein structure predictions provide a reliable prerequisite. For example, the 3D structure for MYB108-like involved in responses to drought and salt stresses in cotton was predicted using a Swiss model to better understand its mechanism of action (Ullah et al., 2020). We predicted the 3D models for the identified CcMYB protein sequences by sequence based homology modeling (**Supplementary Figure 4**). Most of the CcMYBs were modeled with a single chain model, c6kksA of MYB DNA-binding domain repeat with a unique helix-loop-helix structure in Arabidopsis R2R3 MYB2 TF WER (**Supplementary Figure 4**, **Supplementary Table 6**). The crystal structure of the WER complex with its target DNA was determined recently by X-Ray diffraction and showed that third recognition helices of both R2 and R3 MYB repeats bind to the major groove of DNA in a sequence-specific manner (Wang et al., 2020a). WER is a MYB-related protein that transcriptionally regulates the expression of *GLABRA2* to control epidermal cell patterning in a position-dependent manner in Arabidopsis roots (Lee and Schiefelbein, 1999). The CcMYB R2R3 domains shared a moderate degree of sequence similarity (<68%) to the identified template sequence of the MYB domain, which may indicate similarity in the mechanism of binding to its target DNA. Provided there is still no method to determine protein structures solely based on sequence information without known reference structures, the computational based analysis of CcMYB protein sequences can be a stepping stone toward structural determination.

MYB Containing Conserved Syntenic Segments Showed Diversification Among Solanaceae Members

With an increasing number of sequenced plant genomes, little has been understood about the genomic divergence, chromosome evolution, and evolutionary relationships among them. Synteny and collinearity are one way to detect complex evolutionary relationships among plant genomes, especially in reference to multigene families. Synteny analysis among Solanaceae genomes displays the diversification and conservation of chromosomal segments containing *MYB* genes (**Table 2**). The *Capsicum* genomes shared more numbers of CSSs, most of them on similar chromosomes and with similar order of genes, as compared with Solanum genomes, which is expected. However, several CSSs were diversified in *Capsicum* spp. despite belonging to the same genus and being closely related. The least conservation in CSSs was observed with the Arabidopsis genome with only

seven CSSs present on similar chromosomes as *C. chinense*. Only 142 (56.57%) and 114 (45.41%) *C. chinense MYB* genes were collinear with the *C. annuum* and *C. baccatum* genomes, respectively (**Table 3**). The lower level of collinearity suggests that *Capsicum* genomes have undergone large-scale chromosomal rearrangements during their evolution. Therefore, it may indicate that *Capsicum* genomes, and *MYB* genes, have diverged, and that there is a need to study specific genes and genomes. In the genomes of species belonging to the *Solanum* genus, even a lower level of conservation and collinearity of *MYB* genes was observed with respect to *C. chinense* as expected (**Table 2**).

Apart from the genome-wide duplications, tandem and segmental duplications within *C. chinense* highlight the duplication frequency of *MYB* genes within the *Capsicum* genome (**Table 1**). Genome-wide and tandem duplications have been implicated in the expansion of the *R2R3-MYB* gene family and are an important measure for the same (Du et al., 2015). In our study, we reported the average duplication time of *Capsicum MYB* homolog pairs as 28.56 MYA. The *Ka/Ks* ratios of 435 *MYB* duplicate pairs were <1, which indicates purifying selection. Forty-seven *MYB* pairs that had *Ka/Ks* >1 showed positive selection. We found 241 *Capsicum MYBs* with orthologs in tomato, potato, and Arabidopsis that were operating under purifying selection (**Supplementary Table 8**). The strong purifying selection of the *Capsicum MYB* gene family is similar to the tandem expansion and positive selection observed in the GRAS TF family and *R*-genes in Arabidopsis (Chen et al., 2010; Wu et al., 2014).

MYB-Specific SSR Repeats Can Serve as Potential Molecular Markers

In an earlier study, gene-specific SSRs related to *Capsicum* fruit ripening showed high polymorphism among *Capsicum spp* (Dubey et al., 2019). In this study, the SSRs predicted in the genic and non-genic regions of *Capsicum MYBs* have rendered them useful in *Capsicum* breeding and improvement programs. Compared with the 1.5-kb upstream regions, more SSRs were present in the genic regions, 67.45% (**Supplementary Table 9**). However, previous studies have suggested higher conservation and less variation in the genic regions of different species (Kim et al., 2008; Zhang et al., 2014; Chhapekar et al., 2020). This may be due to the less coverage of the upstream region of *MYB* genes for SSR prediction, which is 1.5 Kb from the TSS in our study. The di- and tri-nucleotide repeats have been found to be varying in number from species to species but are the most common SSRs in plants (Saha et al., 2017). We also found di-nucleotide repeats to be abundant among all the SSRs in *Capsicum* species. Both gene-based and non-genic SSRs in *Capsicum MYBs* can be used as potential markers for the selection of associated genes in fruit breeding programs.

CONCLUSION

A total of 251, 240, and 245 *MYB* genes were identified in the *C. chinense*, *C. baccatum*, and *C. annuum* genomes. Twenty tandem and 41 segmental duplication events may have led to the expansion of the *MYB* gene family in the *C. chinense* genome. Also, 225, 213, 125, 79, and 23 CcMYB proteins were orthologous to *C. annuum*, *C. baccatum,* potato, tomato, and Arabidopsis MYBs, respectively. The transcriptome analysis revealed that 54 *CcMYB* and 81 *CaMYB* genes were differentially expressed during fruit development in *C. chinense* and *C. annuum*, respectively. Eight *CcMYB* genes were highly expressed in highly pungent *C. chinense* compared with lowly pungent *C. annuum* in the qRT-PCR analysis. Additionally, our finding also suggests the *CcMYBs*, such as *CcMYB16*, *CcMYB28*, *CcMYB100*, *CcDIV4*, *CcMYB46*, and *CcMYB74*, as potential anthocyanin biosynthesis regulators in *Capsicum*. While along with already characterized *MYB31* and *MYB108* (Arce-Rodríguez and Ochoa-Alejo, 2017) other *MYBs* such as *CcMYB4, CcDIV1*, *CcMYBR6,* and *CcARR11* may be used as potential targets for the regulation of capsaicinoid biosynthesis. On the other hand *CcMYBs*, such as *CcMYB10*, *CcMYB82*, *CcMYB1R-1*, *CcDIV1*, *CcRVE4*, and *CcMYB102*, may be investigated for their role in fruit development/shape-size regulation in fruits of *Capsicum* species. The *MYB* genes identified could be studied for their functional roles, so that they can be manipulated for *Capsicum* trait improvement.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: The RNA sequencing data related to this study were submitted on NCBI under BioProject (PRJNA679780). Sequence Read Archive (SRA) accessions for *C. chinense* samples includes SRR12963502, SRR12963513, and SRR12963514 for early green (EG), SRR12963488, SRR12963489, and SRR12963490 for mature green (MG), and SRR12963491, SRR12963492, and SRR12963493 for breaker (Br) fruit samples. SRA accessions for *C. annuum* samples are SRR12963501, SRR12963503 and SRR12963504 for EG, SRR12963495, SRR12963496, and SRR12963497 for MG and SRR12963498, SRR12963499, and SRR12963500 for Breaker fruit samples.

AUTHOR CONTRIBUTIONS

NR conceived and designed the research. KI, AR, IA, MD, and JM conducted field and lab experiments. KI and AR performed *in silico* analysis and analyzed the data. KI and AR drafted and NR corrected and finalized the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Department of Biotechnology (DBT), Ministry of Science and Technology, Government of India in the form of Ramalingaswami Re-entry Fellowship cum Research Grant (BT/RLF/Re-entry/46/2011), and Science and Engineering Research Board (SERB) (SB/EMEQ-086/2014), Department of Science and Technology (DST), Government of India.

ACKNOWLEDGMENTS

KI acknowledges the University Grant Commission (UGC), Government of India, for the fellowship received in the form of a Senior Research Fellowship. AR, IA, and JM acknowledge the fellowships received from the Council of Scientific and Industrial Research (CSIR), Government of India, New Delhi in the form of Senior Research fellowships. The authors also acknowledge the DST, Govt. of India for DST-FIST-II given to School of Life Sciences, Jawaharlal Nehru University, New Delhi.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.721265/full#supplementary-material

Supplementary Figure 1 | Chromosomal distribution of *MYB* genes across 12 *Capsicum* chromosomes of **(A)** *C. annuum* and **(B)** *C. baccatum.* The capsaicinoid QTLs in *C. annuum* adapted from Han et al., 2018 are shown in shaded blocks.

Supplementary Figure 2 | Gene structure and phylogenetic tree of *MYB* genes in **(A)** *C. annuum* and **(B)** *C. baccatum*. The length of *MYB* genes are in a kb (kilobase) scale, and the exons are represented using purple and brown blocks, while the connecting lines represent the intronic regions.

Supplementary Figure 3 | Phylogenetic tree of MYB protein sequences from the three *Capsicum* species (*C. chinense*, *C. annuum,* and *C. baccatum*). The tree was constructed using the Maximum Likelihood method with 1,000 bootstrap replications.

Supplementary Figure 4 | Predicted 3D structural models of 126 *C. chinense* R2R3 MYB proteins. Models were shown in rainbow color from N to the C terminus.

Supplementary Figure 5 | The phylogenetic tree is drawn from MYB protein sequences of *C. chinense*, *C. annuum*, *C. baccatum*, tomato, potato, and Arabidopsis using the Maximum likelihood method with 1,000 bootstrap values.

Supplementary Figure 6 | Relative expression of eight *CcMYB* genes in six genotypes, three *C. annuum* (Kosom Moso, JH15 and JH7) and three *C. chinense* (Lota, CCGHY6, and BJS) by qRT-PCR in the three fruit developmental stages. A Student *t*-test was performed to calculate the significant difference of expression. The significance level was represented as $^{***}p < 0.001$; $^{**}p < 0.01$, and $^{*}p < 0.05$.

Supplementary Table 1 | Primer sequences used for the qRT-PCR expression analysis of *CcMYBs* and capsaicinoids biosynthesis pathway genes.

Supplementary Table 2 | The characteristic features of *Capsicum MYBs.* **(A)** *C. chinense (CcMYBs)* and **(B)** *C. annuum* (*CaMYBs*) genes with RNAseq expression data in the three fruit developmental stages. **(C)** *C. baccatum* (*CbMYBs*) along with their genomic coordinates.

Supplementary Table 3 | Conserved amino acid motifs identified in CcMYB protein sequences using the MEME suite.

Supplementary Table 4 | Cis-regulatory elements identified in the 1.5-kb upstream promoter region of *CcMYBs* using the PLANTCare database. **(A)** Cis-acting elements along with their sequences, function, and presence in the *MYB* genes and **(B)** their frequency in the upstream region of each *MYB* gene.

Supplementary Table 5 | RNA sequencing alignment summary. Raw reads and aligned read summary with Hisat2 in the **(A)** *C. chinense* and **(B)** *C. annuum* fruit samples.

Supplementary Table 6 | The 126 R2R3 CcMYBs and the probability and identity of their corresponding homologous protein model/template.

Supplementary Table 7 | *MYB* containing conserved syntenic segments (CSSs) from the *Capsicum chinense* genome compared with the *C. annuum, C. baccatum,* tomato, potato, and Arabidopsis genomes [Note: Cc, Ca, Cb, St, Sl, and At represent the chromosomes from the *C. chinense, C. annuum, C. baccatum,* potato, tomato, and Arabidopsis genomes].

Supplementary Table 8 | The *C. chinense* (*CcMYBs*) and their homologous gene pairs from *C. annuum*, *C. baccatum*, tomato, potato, and Arabidopsis along with the *Ka/Ks* ratios and duplication time (in MYA) for each homologous pair.

Supplementary Table 9 | SSRs identified in the *MYB* genes and 1.5-kb promoter regions in *C. chinense* with the primer sequences and the amplicon size of the SSR motifs.

Supplementary File 1 | CDS and protein sequences of *C. chinense MYB* family members.

REFERENCES

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., and Shinozaki, K. (1997). Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* 9, 1859–1868. doi: 10.1105/tpc.9.10.1859

Aguilar-Barragán, A., and Ochoa-Alejo, N. (2014). Virus-induced silencing of MYB and WD40 transcription factor genes affects the accumulation of anthocyanins in chilli pepper fruit. *Biol. Plant.* 58, 567–574. doi: 10.1007/s10535-014-0427-4

Antonious, G., and Jarret, R. (2006). Screening Capsicum accessions for capsaicinoids content. *J. Environ. Sci. Heal. Part B Pestic. Food Contam. Agric. Wastes* 41, 717–729. doi: 10.1080/03601230600701908

Arce-Rodríguez, M. L., Martínez, O., and Ochoa-Alejo, N. (2021). Genome-wide identification and analysis of the MYB transcription factor gene family in chili pepper (*Capsicum* spp.). *Int. J. Mol. Sci.* 22:2229. doi: 10.3390/ijms22052229

Arce-Rodríguez, M. L., and Ochoa-Alejo, N. (2017). An R2R3-MYB transcription factor regulates capsaicinoid biosynthesis. *Plant Physiol.* 174, 1359–1370. doi: 10.1104/pp.17.00506

Bai, R., Luo, Y., Wang, L., Li, J., Wu, K., Zhao, G., et al. (2019). A specific allele of MYB14 in grapevine correlates with high stilbene inducibility triggered by Al 3+ and UV-C radiation. *Plant Cell Rep.* 38, 37–49. doi: 10.1007/s00299-018-2347-9

Ballester, A.-R., Molthoff, J., de Vos, R., Hekkert, B., te, L., Orzaez, D., et al. (2010). Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. *Plant Physiol.* 152, 71–84. doi: 10.1104/pp.109.147322

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., et al. (2019). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Sci. Rep.* 9:11769. doi: 10.1038/s41598-019-47985-w

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, Y., Li, K., Li, Y., Zhao, X., and Wang, L. (2020). MYB transcription factors as regulators of secondary metabolism in plants. *Biology* 9:61. doi: 10.3390/biology9030061

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, Q., Han, Z., Jiang, H., Tian, D., and Yang, S. (2010). Strong positive selection drives rapid diversification of R-Genes in arabidopsis relatives. *J. Mol. Evol.* 70, 137–148. doi: 10.1007/s00239-009-9316-4

Chhapekar, S. S., Brahma, V., Rawoof, A., Kumar, N., Gaur, R., Jaiswal, V., et al. (2020). Transcriptome profiling, simple sequence repeat markers development

and genetic diversity analysis of potential industrial crops *Capsicum chinense* and *C. frutescens* of Northeast India. *Ind Crops Prod*. 154:112687.

Cutanda-Perez, M.-C., Ageorges, A., Gomez, C., Vialet, S., Terrier, N., Romieu, C., et al. (2009). Ectopic expression of VlmybA1 in grapevine activates a narrow set of genes involved in anthocyanin synthesis and transport. *Plant Mol. Biol.* 69, 633–648. doi: 10.1007/s11103-008-9446-x

Czemmel, S., Stracke, R., Weisshaar, B., Cordon, N., Harris, N. N., Walker, A. R., et al. (2009). The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. *Plant Physiol.* 151, 1513–1530. doi: 10.1104/pp.109.142059

Dai, X., Xu, Y., Ma, Q., Xu, W., Wang, T., Xue, Y., et al. (2007). Overexpression of an R1R2R3 MYB Gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic Arabidopsis. *Plant Physiol.* 143, 1739–1751. doi: 10.1104/pp.106.094532

De Vos, M., Denekamp, M., Dicke, M., Vuylsteke, M., Van Loon, L., Smeekens, S. C., et al. (2006). The *Arabidopsis thaliana* transcription factor AtMYB102 functions in defense against the insect herbivore *Pieris rapae*. *Plant Signal. Behav.* 1, 305–311. doi: 10.4161/psb.1.6.3512

Dieffenbach, C. W., Lowe, T. M. J., and Dveksler, G. S. (1993). General concepts for PCR primer design. *Genome Res.* 3, S30–S37. doi: 10.1101/gr.3.3.S30

Du, H., Liang, Z., Zhao, S., Nan, M. G., Tran, L. S. P., Lu, K., et al. (2015). The evolutionary history of R2R3-myb proteins across 50 eukaryotes: new insights into subfamily classification and expansion. *Sci. Rep*. 5:11037. doi: 10.1038/srep11037

Dubey, M., Jaiswal, V., Rawoof, A., Kumar, A., Nitin, M., Chhapekar, S. S., et al. (2019). Identification of genes involved in fruit development/ripening in Capsicum and development of functional markers. *Genomics* 111, 1913–1922. doi: 10.1016/j.ygeno.2019.01.002

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. *Trends Plant Sci.* 15, 573–581. doi: 10.1016/j.tplants.2010.06.005

Dunn, M. A., Goddard, N. J., Zhang, L., Pearce, R. S., and Hughes, M. A. (1994). Low-temperature-responsive barley genes have different control mechanisms. *Plant Mol. Biol.* 24, 879–888. doi: 10.1007/BF00014442

Fattori, V., Hohmann, M. S. N., Rossaneis, A. C., Pinho-Ribeiro, F. A., and Verri, W. A. (2016). Capsaicin: current understanding of its mechanisms and therapy of pain and other pre-clinical and clinical uses. *Molecules* 21:844. doi: 10.3390/molecules21070844

Forde, B. G., Heyworth, A., Pywell, J., and Kreis, M. (1985). Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. *Nucleic Acids Res.* 13, 7327–7339. doi: 10.1093/nar/13.20.7327

Fujiwara, T., and Beachy, R. N. (1994). Tissue-specific and temporal regulation of a β-conglycinin gene: roles of the RY repeat and other cis-acting elements. *Plant Mol. Biol.* 24, 261–272. doi: 10.1007/BF00020166

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, ed J. M. Walker (Hatfield: Humana Press), 571–607.

Geng, D., Chen, P., Shen, X., Zhang, Y., Li, X., Jiang, L., et al. (2018). MdMYB88 and MdMYB124 enhance drought tolerance by modulating root vessels and cell walls in Apple. *Plant Physiol.* 178, 1296–1309. doi: 10.1104/pp.18.00502

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi: 10.1093/bioinformatics/btu393

Haga, N., Kobayashi, K., Suzuki, T., Maeo, K., Kubo, M., Ohtani, M., et al. (2011). Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in Arabidopsis. *Plant Physiol.* 157, 706–717. doi: 10.1104/pp.111.180836

Han, K., Jang, S., Lee, J. H., Lee, D. G., Kwon, J. K., and Kang, B. C. (2019). A MYB transcription factor is a candidate to control pungency in *Capsicum annuum*. *Theor. Appl. Genet.* 132, 1235–1246. doi: 10.1007/s00122-018-03275-z

Han, K., Lee, H. Y., Ro, N. Y., Hur, O. S., Lee, J. H., Kwon, J. K., et al. (2018). QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in Capsicum. *Plant Biotechnol. J.* 16, 1546–1558. doi: 10.1111/pbi.12894

Hilscher, J., Schlötterer, C., and Hauser, M.-T. (2009). A single amino acid replacement in ETC2 shapes trichome patterning in natural Arabidopsis populations. *Curr. Biol.* 19, 1747–1751. doi: 10.1016/j.cub.2009.08.057

Hosoda, K., Imamura, A., Katoh, E., Hatta, T., Tachiki, M., Yamada, H., et al. (2002). Molecular structure of the GARP family of plant myb-related DNA binding motifs of the Arabidopsis response regulators. *Plant Cell* 14, 2015–2029. doi: 10.1105/tpc.002733

Ito, M., Araki, S., Matsunaga, S., Itoh, T., Nishihama, R., Machida, Y., et al. (2001). G2/M-phase-specific transcription during the plant cell cycle is mediated by c-Myb-like transcription factors. *Plant Cell* 13, 1891–1905. doi: 10.1105/TPC.010102

Jaiswal, V., Rawoof, A., Dubey, M., Chhapekar, S. S., Sharma, V., and Ramchiary, N. (2020). Development and characterization of non-coding RNA based simple sequence repeat markers in Capsicum species. *Genomics* 112, 1554–1564. doi: 10.1016/j.ygeno.2019.09.005

Jian, W., Cao, H., Yuan, S., Liu, Y., Lu, J., Lu, W., et al. (2019). SlMYB75, an MYB-type transcription factor, promotes anthocyanin accumulation and enhances volatile aroma production in tomato fruits. *Hortic. Res.* 6, 1–15. doi: 10.1038/s41438-018-0098-y

Jin, H., and Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol. Biol.* 41, 577–585. doi: 10.1023/A:1006319732410

Jung, C., Seo, J. S., Han, S. W., Koo, Y. J., Kim, C. H., Song, S. I., et al. (2008). Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic Arabidopsis. *Plant Physiol.* 146, 623–635. doi: 10.1104/pp.107.110981

Jung, S., Venkatesh, J., Kang, M. Y., Kwon, J. K., and Kang, B. C. (2019). A non-LTR retrotransposon activates anthocyanin biosynthesis by regulating a MYB transcription factor in Capsicum annuum. *Plant Sci.* 287:110181. doi: 10.1016/j.plantsci.2019.110181

Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., and Bansal, K. C. (2012). Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics* 13:544. doi: 10.1186/1471-2164-13-544

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Kiferle, C., Fantini, E., Bassolino, L., Povero, G., Spelt, C., Buti, S., et al. (2015). Tomato R2R3-MYB proteins SlANT1 and SlAN2: same protein activity, different roles. *PLoS ONE* 10:e0136365. doi: 10.1371/journal.pone.0136365

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Kim, J. H., Nguyen, N. H., Jeong, C. Y., Nguyen, N. T., Hong, S.-W., and Lee, H. (2013). Loss of the R2R3 MYB, AtMyb73, causes hyper-induction of the SOS1 and SOS3 genes in response to high salinity in Arabidopsis. *J. Plant Physiol.* 170, 1461–1465. doi: 10.1016/j.jplph.2013.05.011

Kim, K. S., Ratcliffe, S. T., French, B. W., Liu, L., and Sappington, T. W. (2008). Utility of EST-derived SSRs as population genetics markers in a Beetle. *J. Hered.* 99, 112–124. doi: 10.1093/jhered/esm104

Kim, S., Park, J., Yeom, S. I., Kim, Y. M., Seo, E., Kim, K. T., et al. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* 18, 1–11. doi: 10.1186/s13059-017-1341-9

Kim, S. Y., and Wu, R. (1990). Multiple protein factors bind to a rice glutelin promoter region. *Nucleic Acids Res.* 18, 6845–6852. doi: 10.1093/nar/18.23.6845

Kirik, V., Simon, M., Wester, K., Schiefelbein, J., and Hulskamp, M. (2004). ENHANCER of TRYand CPC 2(ETC2) reveals redundancy in the region-specific control of trichome development of Arabidopsis. *Plant Mol. Biol.* 55, 389–398. doi: 10.1007/s11103-004-0893-8

Klempnauer, K. H., and Sippel, A. E. (1987). The highly conserved amino-terminal region of the protein encoded by the v-myb oncogene functions as a DNA-binding domain. *EMBO J.* 6, 2719–2725. doi: 10.1002/j.1460-2075.1987.tb02565.x

König, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996). The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell* 85, 125–136. doi: 10.1016/S0092-8674(00)81088-0

Kortstee, A. J., Khan, S. A., Helderman, C., Trindade, L. M., Wu, Y., Visser, R. G. F., et al. (2011). Anthocyanin production as a potential visual

selection marker during plant transformation. *Transgenic Res.* 20, 1253–1264. doi: 10.1007/s11248-011-9490-1

Kranz, H. D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R. C., et al. (1998). Towards functional characterisation of the members of the R2R3-MYB gene family from Arabidopsis thaliana. *Plant J.* 16, 263–276. doi: 10.1046/j.1365-313x.1998.00278.x

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lai, L. B., Nadeau, J. A., Lucas, J., Lee, E.-K., Nakagawa, T., Zhao, L., et al. (2005). The Arabidopsis R2R3 MYB proteins FOUR LIPS and MYB88 restrict divisions late in the stomatal cell lineage. *Plant Cell* 17, 2754–2767. doi: 10.1105/tpc.105.034116

Lee, E., Liu, X., Eglit, Y., and Sack, F. (2013). FOUR LIPS and MYB88 conditionally restrict the G1/S transition during stomatal formation. *J. Exp. Bot.* 64, 5207–5219. doi: 10.1093/jxb/ert313

Lee, M. M., and Schiefelbein, J. (1999). WEREWOLF, a MYB-related protein in Arabidopsis, is a position-dependent regulator of epidermal cell patterning. *Cell* 99, 473–483. doi: 10.1016/S0092-8674(00)81536-6

Lei, Q., Lee, E., Keerthisinghe, S., Lai, L., Li, M., Lucas, J. R., et al. (2015). The FOUR LIPS and MYB88 transcription factor genes are widely expressed in Arabidopsis thaliana during development. *Am. J. Bot.* 102, 1521–1528. doi: 10.3732/ajb.1500056

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Li, C., Ng, C. K. Y., and Fan, L. M. (2015). MYB transcription factors; active players in abiotic stress signaling. *Environ. Exp. Bot.* 114, 80–91. doi: 10.1016/j.envexpbot.2014.06.014

Li, W., Liu, Y., Zhao, J., Zhen, X., Guo, C., and Shu, Y. (2019). Genome-wide identification and characterization of r2r3-myb genes in medicago truncatula. *Genet. Mol. Biol.* 42, 611–623. doi: 10.1590/1678-4685-gmb-2018-0235

Li, Y., Liang, J., Zeng, X., Guo, H., Luo, Y., Kear, P., et al. (2021). Genome-wide Analysis of MYB gene family in potato provides insights into tissue-specific regulation of anthocyanin biosynthesis. *Hortic. Plant J.* 7, 129–141. doi: 10.1016/j.hpj.2020.12.001

Li, Z., Peng, R., Tian, Y., Han, H., Xu, J., and Yao, Q. (2016). Genome-wide identification and analysis of the MYB transcription factor superfamily in Solanum lycopersicum. *Plant cell Physiol. Physiol.* 57, 1657–1677. doi: 10.1093/pcp/pcw091

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Lipsick, J. S. (1996). One billion years of myb. *Oncogene* 13, 223–235.

Liu, J., Osbourn, A., and Ma, P. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* 8, 689–708. doi: 10.1016/j.molp.2015.03.012

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using Real-Time Quantitative PCR and the 2−ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991

Luo, Y., Bai, R., Li, J., Yang, W., Li, R., Wang, Q., et al. (2019). The transcription factor MYB15 is essential for basal immunity (PTI) in Chinese wild grape. *Planta* 249, 1889–1902. doi: 10.1007/s00425-019-03130-5

Ma, Q., Dai, X., Xu, Y., Guo, J., Liu, Y., Chen, N., et al. (2009). Enhanced tolerance to chilling stress in OsMYB3R-2 transgenic rice is mediated by alteration in cell cycle and ectopic expression of stress genes. *Plant Physiol.* 150, 244–256. doi: 10.1104/pp.108.133454

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi: 10.1093/nar/gkz268

Makkena, S., Lee, E., Sack, F. D., and Lamb, R. S. (2012). The R2R3 MYB transcription factors FOUR LIPS and MYB88 regulate female reproductive development. *J. Exp. Bot.* 63, 5545–5558. doi: 10.1093/jxb/ers209

Martin, C., and Paz-Ares, J. (1997). MYB transcription factors in plants. *Trends Genet.* 13, 67–73. doi: 10.1016/S0168-9525(96)10049-4

Martins, W. S., César, D., Lucas, S., Fabricio, K., Neves, D. S., and John, D. (2009). Bioinformation WebSat - a web software for microsatellite marker development Bioinformation. *Biomed. Informatics Publ. Gr.* 3, 282–283. doi: 10.6026/97320630003282

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol.* 138, 1083–1096. doi: 10.1104/pp.104.058032

Millard, P. S., Kragelund, B. B., and Burow, M. (2019). R2R3 MYB transcription factors – functions outside the DNA-binding domain. *Trends Plant Sci.* 24, 934–946. doi: 10.1016/j.tplants.2019.07.003

Moniz de Sá, M., and Drouin, G. (1996). Phylogeny and substitution rates of angiosperm actin genes. *Mol. Biol. Evol.* 13, 1198–1212. doi: 10.1093/oxfordjournals.molbev.a025685

Müller, M., and Knudsen, S. (1993). The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box. *Plant J.* 4, 343–355. doi: 10.1046/j.1365-313X.1993.04020343.x

Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* 13, 2099–2114. doi: 10.1105/TPC.010098

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., et al. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* 79, 639–648. doi: 10.1016/0092-8674(94)90549-5

Park, M., Lee, J. H., Han, K., Jang, S., Han, J., Lim, J. H., et al. (2019). A major QTL and candidate genes for capsaicinoid biosynthesis in the pericarp of *Capsicum chinense* revealed using QTL-seq and RNA-seq. *Theor. Appl. Genet.* 132, 515–529. doi: 10.1007/s00122-018-3238-8

Phan, H. A., Iacuone, S., Li, S. F., and Parish, R. W. (2011). The MYB80 transcription factor is required for pollen development and the regulation of tapetal programmed cell death in Arabidopsis thaliana. *Plant Cell* 23, 2209–2224. doi: 10.1105/tpc.110.082651

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proc. Natl. Acad. Sci.* 111, 5135–5140. doi: 10.1073/pnas.1400975111

Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J., and Koes, R. (2006). PH4 of petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* 18, 1274–1291. doi: 10.1105/tpc.105.034041

Raffaele, S., Vailleau, F., Léger, A., Joubès, J., Miersch, O., Huard, C., et al. (2008). A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in Arabidopsis. *Plant Cell* 20, 752–767. doi: 10.1105/tpc.107.054858

Rawoof, A., Chhapekar, S. S., Jaiswal, V., Brahma, V., Kumar, N., and Ramchiary, N. (2020). Single-base cytosine methylation analysis in fruits of three Capsicum species. *Genomics*, 112, 3342–3353.

Reňák, D., Dupl'áková, N., and Honys, D. (2012). Wide-scale screening of T-DNA lines for transcription factor genes affecting male gametophyte development in Arabidopsis. *Sex Plant Reprod.* 25, 39–60. doi: 10.1007/s00497-011-0178-8

Reyes, J. C., Muro-Pastor, M. I., and Florencio, F. J. (2004). The GATA family of transcription factors in arabidopsis and rice. *Plant Physiol.* 134, 1718–1732. doi: 10.1104/pp.103.037788

Reyes, J. L., and Chua, N.-H. (2007). ABA induction of miR159 controls transcript levels of two MYB factors during Arabidopsis seed germination. *Plant J.* 49, 592–606. doi: 10.1111/j.1365-313X.2006.02980.x

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., et al. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110. doi: 10.1126/science.290.5499.2105

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25. doi: 10.1186/gb-2010-11-3-r25

Romano, J. M., Dubos, C., Prouse, M. B., Wilkins, O., Hong, H., Poole, M., et al. (2012). AtMYB61, an R2R3-MYB transcription factor, functions as a pleiotropic regulator via a small gene network. *New Phytol.* 195, 774–786. doi: 10.1111/j.1469-8137.2012.04201.x

Romero, I., Fuertes, A., Benito, M. J., Malpica, J. M., Leyva, A., and Paz-Ares, J. (1998). More than 80R2R3-MYB regulatory genes in the genome of Arabidopsis thaliana. *Plant J.* 14, 273–284. doi: 10.1046/j.1365-313X.1998.00113.x

Safrany, J., Haasz, V., Mate, Z., Ciolfi, A., Feher, B., Oravecz, A., et al. (2008). Identification of a novel cis-regulatory element for UV-B-induced transcription in Arabidopsis. *Plant J.* 54, 402–414. doi: 10.1111/j.1365-313X.2008.03435.x

Saha, D., Rana, R. S., Chakraborty, S., Datta, S., Kumar, A. A., Chakraborty, A. K., et al. (2017). Development of a set of SSR markers for genetic polymorphism detection and interspecific hybrid jute breeding. *Crop J.* 5, 416–429. doi: 10.1016/j.cj.2017.02.006

Sarpras, M., Ahmad, I., Rawoof, A., and Ramchiary, N. (2019). Comparative analysis of developmental changes of fruit metabolites, antioxidant activities and mineral elements content in Bhut jolokia and other *Capsicum* species. *LWT* 105, 363–370. doi: 10.1016/j.lwt.2019.02.020

Sarpras, M., Gaur, R., Sharma, V., Chhapekar, S. S., Das, J., Kumar, A., et al. (2016). Comparative analysis of fruit metabolites and pungency candidate genes expression between Bhut jolokia and other *Capsicum* species. *PLoS ONE* 11:e0167791. doi: 10.1371/journal.pone.0167791

Schindler, U., Beckmann, H., and Cashmore, A. R. (1992). TGA1 and G-Box binding factors: two distinct classes of arabidopsis leucine zipper proteins compete for the G-box-like element TGACGTGG. *Plant Cell* 4, 1309–1319.

Sevilla-Lecoq, S., Deguerry, F., Matthys-Rochon, E., Perez, P., Dumas, C., and Rogowsky, P. M. (2003). Analysis of ZmAE3 upstream sequences in maize endosperm and androgenic embryos. *Sex. Plant Reprod.* 16, 1–8. doi: 10.1007/s00497-003-0176-6

Shim, J. S., Jung, C., Lee, S., Min, K., Lee, Y.-W., Choi, Y., et al. (2013). AtMYB44 regulates WRKY70 expression and modulates antagonistic interaction between salicylic acid and jasmonic acid signaling. *Plant J.* 73, 483–495. doi: 10.1111/tpj.12051

Shiu, S. H., Shih, M. C., and Li, W. H. (2005). Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol.* 139, 18–26. doi: 10.1104/pp.105.065110

Solano, R., Fuertes, A., Sánchez-Pulido, L., Valencia, A., and Paz-Ares, J. (1997). A single residue substitution causes a switch from the dual DNA binding specificity of plant transcription factor MYB.Ph3 to the animal c-MYB specificity. *J. Biol. Chem.* 272, 2889–2895. doi: 10.1074/jbc.272.5.2889

Stevenson, C. E. M., Burton, N., Costa, M. M. R., Nath, U., Dixon, R. A., Coen, E. S., et al. (2006). Crystal structure of the MYB domain of the RAD transcription factor from *Antirrhinum majus*. *Proteins Struct. Funct. Genet.* 65, 1041–1045. doi: 10.1002/prot.21136

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

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* 4, 447–456. doi: 10.1016/S1369-5266(00)00199-0

Sun, B., Zhou, X., Chen, C., Chen, C., Chen, K., Chen, M., et al. (2020). Coexpression network analysis reveals an MYB transcriptional activator involved in capsaicinoid biosynthesis in hot peppers. *Hortic. Res.* 7:162. doi: 10.1038/s41438-020-00381-2

Sun, B., Zhu, Z., Chen, C., Chen, G., Cao, B., Chen, C., et al. (2019). Jasmonate-inducible R2R3-MYB transcription factor regulates capsaicinoid biosynthesis and stamen development in Capsicum. *J. Agric. Food Chem.* 67, 10891–10903. doi: 10.1021/acs.jafc.9b04978

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, 1216–1232. doi: 10.1104/pp.106.088104

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

Ullah, A., Ul Qamar, M. T., Nisar, M., Hazrat, A., Rahim, G., Khan, A. H., et al. (2020). Characterization of a novel cotton MYB gene, GhMYB108-like responsive to abiotic stresses. *Mol. Biol. Rep.* 47, 1573–1581. doi: 10.1007/s11033-020-05244-6

Urao, T., Noji, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1996). A transcriptional activation domain of ATMYB2, a drought-inducible Arabidopsis Myb-related protein. *Plant J.* 10, 1145–1148. doi: 10.1046/j.1365-313X.1996.10061145.x

Wang, B., Luo, Q., Li, Y., Yin, L., Zhou, N., Li, X., et al. (2020a). Structural insights into target DNA recognition by R2R3-MYB transcription factors. *Nucleic Acids Res.* 48, 460–471. doi: 10.1093/nar/gkz1081

Wang, D., Jiang, C., Liu, W., and Wang, Y. (2020b). The WRKY53 transcription factor enhances stilbene synthesis and disease resistance by interacting with MYB14 and MYB15 in Chinese wild grape. *J. Exp. Bot.* 71, 3211–3226. doi: 10.1093/jxb/eraa097

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics. Proteomics Bioinformatics* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

Wang, H.-Z., Yang, K.-Z., Zou, J.-J., Zhu, L.-L., Xie, Z. D., Morita, M. T., et al. (2015). Transcriptional regulation of PIN genes by FOUR LIPS and MYB88 during Arabidopsis root gravitropism. *Nat. Commun.* 6:8822. doi: 10.1038/ncomms9822

Wang, J., Liu, Y., Tang, B., Dai, X., Xie, L., Liu, F., et al. (2020c). Genome-wide identification and capsaicinoid biosynthesis-related expression analysis of the R2R3-MYB gene family in *Capsicum annuum* L. *Front. Genet.* 11:598183. doi: 10.3389/fgene.2020.598183

Wang, M., Hao, J., Chen, X., and Zhang, X. (2020d). SlMYB102 expression enhances low-temperature stress resistance in tomato plants. *PeerJ* 8:e10059. doi: 10.7717/peerj.10059

Wang, N., Xu, H., Jiang, S., Zhang, Z., Lu, N., Qiu, H., et al. (2017). MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (Malus sieversii f. *niedzwetzkyana). Plant J.* 90, 276–292. doi: 10.1111/tpj.13487

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., and Lum, T. (2020). *gplots: Various R Programming Tools for Plotting Data*. R package version 3.1.1. Available online at: https://CRAN.R-project.org/package=gplots

Wu, N., Zhu, Y., Song, W., Li, Y., Yan, Y., and Hu, Y. (2014). Unusual tandem expansion and positive selection in subgroups of the plant GRAS transcription factor superfamily. *BMC Plant Biol.* 14:373. doi: 10.1186/s12870-014-0373-5

Xie, Z., Li, D., Wang, L., Sack, F. D., and Grotewold, E. (2010). Role of the stomatal development regulators FLP/MYB88 in abiotic stress responses. *Plant J.* 64, 731–739. doi: 10.1111/j.1365-313X.2010.04364.x

Xu, R., Wang, Y., Zheng, H., Lu, W., Wu, C., Huang, J., et al. (2015a). Salt-induced transcription factor MYB74 is regulated by the RNA-directed DNA methylation pathway in Arabidopsis. *J. Exp. Bot.* 66, 5997–6008. doi: 10.1093/jxb/erv312

Xu, W., Dubos, C., and Lepiniec, L. (2015b). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 20, 176–185. doi: 10.1016/j.tplants.2014.12.001

Yamaguchi-Shinozaki, K., and Shinozaki, K. (1994). A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell* 6, 251–264. doi: 10.1105/tpc.6.2.251

Yan, C., An, G., Zhu, T., Zhang, W., Zhang, L., Peng, L., et al. (2019). Independent activation of the BoMYB2 gene leading to purple traits in *Brassica oleracea*. *Theor. Appl. Genet.* 132, 895–906. doi: 10.1007/s00122-018-3245-9

Yan, S., Chen, N., Huang, Z., Li, D., Zhi, J., Yu, B., et al. (2020). Anthocyanin Fruit encodes an R2R3-MYB transcription factor, SlAN2-like, activating the transcription of SlMYBATV to fine-tune anthocyanin content in tomato fruit. *New Phytol.* 225, 2048–2063. doi: 10.1111/nph.16272

Yang, M. (2016). The FOUR LIPS (FLP) and MYB88 genes conditionally suppress the production of nonstomatal epidermal cells in Arabidopsis

cotyledons. *Am. J. Bot.* 103, 1559–1566. doi: 10.3732/ajb.1600238

Yang, S., Cai, Y., Liu, X., Dong, M., Zhang, Y., Chen, S., et al. (2018). A CsMYB6-CsTRY module regulates fruit trichome initiation in cucumber. *J. Exp. Bot.* 69, 1887–1902. doi: 10.1093/jxb/ery047

Yoshihara, T., Washida, H., and Takaiwa, F. (1996). A 45-bp proximal region containing AACA and GCN4 motif is sufficient to confer endosperm-specific expression of the rice storage protein glutelin gene, GluA-3. *FEBS Lett.* 383, 213–218. doi: 10.1016/0014-5793(96)00233-5

Zhang, M., Mao, W., Zhang, G., and Wu, F. (2014). Development and characterization of polymorphic ESTSSR and genomic SSR markers for tibetan annual wild barley. *PLoS ONE* 9:e94881. doi: 10.1371/journal.pone.0094881

Zhang, P., Wang, R., Yang, X., Ju, Q., Li, W., Lü, S., et al. (2020). The R2R3-MYB transcription factor AtMYB49 modulates salt tolerance in Arabidopsis by modulating the cuticle formation and antioxidant defence. *Plant. Cell Environ.* 43, 1925–1943. doi: 10.1111/pce.13784

Zhao, Y., Xing, L., Wang, X., Hou, Y.-J., Gao, J., Wang, P., et al. (2014). The ABA receptor PYL8 promotes lateral root growth by enhancing MYB77-dependent transcription of auxin-responsive genes. *Sci. Signal.* 7:ra53. doi: 10.1126/scisignal.2005051

Zhu, Z., Sun, B., Cai, W., Zhou, X., Mao, Y., Chen, C., et al. (2019). Natural variations in the MYB transcription factor MYB31 determine the evolution of extremely pungent peppers. *New Phytol.* 223, 922–938. doi: 10.1111/nph.15853

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 October 2021
doi: 10.3389/fpls.2021.744175

Identification of Novel Quantitative Trait Nucleotides and Candidate Genes for Bacterial Wilt Resistance in Tobacco (*Nicotiana tabacum* L.) Using Genotyping-by-Sequencing and Multi-Locus Genome-Wide Association Studies

*Ruiqiang Lai[1†], Muhammad Ikram[1†], Ronghua Li[1], Yanshi Xia[1], Qinghua Yuan[2], Weicai Zhao[3], Zhenchen Zhang[2], Kadambot H. M. Siddique[4] and Peiguo Guo[1]**

[1] International Crop Research Center for Stress Resistance, School of Life Sciences, Guangzhou University, Guangzhou, China, [2] Crop Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China, [3] Nanxiong Research Institute of Guangdong Tobacco Co., Ltd., Nanxiong, China, [4] The UWA Institute of Agriculture, UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA, Australia

Tobacco bacterial wilt (TBW) is a devastating soil-borne disease threatening the yield and quality of tobacco. However, its genetic foundations are not fully understood. In this study, we identified 126,602 high-quality single-nucleotide polymorphisms (SNPs) in 94 tobacco accessions using genotyping-by-sequencing (GBS) and a 94.56 KB linkage disequilibrium (LD) decay rate for candidate gene selection. The population structure analysis revealed two subpopulations with 37 and 57 tobacco accessions. Four multi-locus genome-wide association study (ML-GWAS) approaches identified 142 quantitative trait nucleotides (QTNs) in E1–E4 and the best linear unbiased prediction (BLUP), explaining 0.49–22.52% phenotypic variance. Of these, 38 novel stable QTNs were identified across at least two environments/methods, and their alleles showed significant TBW-DI differences. The number of superior alleles associated with TBW resistance for each accession ranged from 4 to 24; eight accessions had more than 18 superior alleles. Based on TBW-resistant alleles, the five best cross combinations were predicted, including MC133 × Ruyuan No. 1 and CO258 × ROX28. We identified 52 candidate genes around 38 QTNs related to TBW resistance based on homologous functional annotation and KEGG enrichment analysis, e.g., *CYCD3;2*, *BSK1*, *Nitab4.5_0000641g0050*, *Nitab4.5_0000929g0030*. To the best of our knowledge, this is the first comprehensive study to identify QTNs, superior alleles, and their candidate genes for breeding TBW-resistant tobacco varieties. The results provide further insight into the genetic architecture, marker-assisted selection, and functional genomics of TBW resistance, improving future breeding efforts to increase crop productivity.

Keywords: tobacco germplasm, bacterial wilt resistance, SNP, genome-wide association analysis, quantitative trait nucleotide, superior alleles

OPEN ACCESS

Edited by:
Ezio Portis,
University of Turin, Italy

Reviewed by:
Shengping Zhang,
Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences (CAAS), China
Wen Yao,
Henan Agricultural University, China

****Correspondence:***
Peiguo Guo
guopg@gzhu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *19 July 2021*
Accepted: *22 September 2021*
Published: *21 October 2021*

Citation:
Lai R, Ikram M, Li R, Xia Y, Yuan Q, Zhao W, Zhang Z, Siddique KHM and Guo P (2021) Identification of Novel Quantitative Trait Nucleotides and Candidate Genes for Bacterial Wilt Resistance in Tobacco (Nicotiana tabacum L.) Using Genotyping-by-Sequencing and Multi-Locus Genome-Wide Association Studies. Front. Plant Sci. 12:744175. doi: 10.3389/fpls.2021.744175

INTRODUCTION

Tobacco (*Nicotiana tabacum* L.; 2n = 48) is an important cash crop in many countries, including China, and a valuable model system in genetic engineering and molecular biology. Tobacco bacterial wilt (TBW) caused by *Ralstonia solanacearum* is a destructive soil-borne disease in many regions worldwide (Nishi et al., 2003; Lan et al., 2014; Drake-Stowe et al., 2017). Infected tobacco plants typically exhibit symptoms such as leaf wilt, root and stem necrosis, and growth retardation, followed by plant death, which reduce yield and quality (Qian et al., 2013). TBW is prevalent in tobacco-growing countries with moist tropical or warm-temperate climates (Denny, 2006). In China, the disease occurrence has been steadily rising, reaching 15–30% in some areas (Jiang et al., 2017), posing a serious threat to tobacco production in the four main tobacco-growing regions, including 14 provinces (Li et al., 2016). Several methods, including crop rotation and soil fumigation, can reduce some economic losses from the disease; however, none provide sufficient protection (Nishi et al., 2003; Lan et al., 2014).

Tobacco bacterial wilt resistance is a quantitative trait controlled by multiple genes and/or quantitative trait loci/nucleotides (QTLs/QTNs) (Smith and Clayton, 1948; Nishi et al., 2003; Gao et al., 2010; Ni et al., 2011). It is challenging to improve TBW resistance using traditional breeding methods (Yang et al., 2012). Marker-assisted selection (MAS) is an alternative tool for combining different resistance genes/alleles into a single plant, which has been used to improve different traits in crop breeding programs (Kuchel et al., 2005; Ribaut et al., 2010; Nakaya and Isobe, 2012). It is important to identify significant QTLs/QTNs to develop superior TBW-resistant tobacco cultivars. To date, only four QTL mapping studies for TBW resistance have been conducted in bi-parental and different genetic populations, using simple sequence repeat (SSR) and amplified fragment length polymorphism (AFLP) markers (Nishi et al., 2003; Qian et al., 2013; Lan et al., 2014; Drake-Stowe et al., 2017). Nishi et al. (2003) identified one QTL with 43.8% phenotypic variance using 117 AFLP markers in 125 doubled haploid populations. Similarly, Qian et al. (2013); Lan et al. (2014), and Drake-Stowe et al. (2017) identified four, eight, and two QTLs for TBW resistance, respectively. Thus, only 15 QTLs underlying TBW resistance have been identified, which is relatively small compared to other members of the Solanaceae family (Sharma et al., 2021). Unfortunately, the identified QTLs have large genomic regions unsuitable for detecting candidate genes, and markers linked to these QTLs have limited application in tobacco breeding programs (Li et al., 2017).

With the development of next-generation sequencing (NGS) technology, genotyping-by-sequencing (GBS) has been used widely as a high-throughput and low-cost genotyping platform for discovering genome-wide single-nucleotide polymorphisms (SNPs) in many crops, including tobacco (Elshire et al., 2011; Lee et al., 2017; Sakiroglu and Brummer, 2017). Genome-wide association studies (GWAS) can use these millions of SNPs as molecular markers to screen many accessions simultaneously without needing to construct segregating populations in advance (Buckler and Thornsberry, 2002; Flint-Garcia et al., 2003; Yu et al., 2006; Wang et al., 2016; Zhang et al., 2019). For instance, Thapa et al. (2021) identified 23 and 38 QTNs associated with shoot and blossom blight resistance, respectively, using GBS markers in 273 apple accessions, while Jing et al. (2021) identified 18 QTNs related to Sclerotinia stem rot resistance in soybean. Thus, GWAS is an efficient tool for QTN identification in natural populations with high-quality SNPs to overcome the shortcomings of bi-parental QTL mapping (Zhang et al., 2019) and has great potential for discovering interrelationships among complex traits conditioned by multiple genes/alleles (Buckler and Thornsberry, 2002; Yu et al., 2006; Hyun et al., 2021; Thapa et al., 2021). However, no GWAS studies have been undertaken to detect QTNs associated with TBW resistance in tobacco. Identifying QTNs/alleles/genes related to TBW resistance is an important step for improving tobacco production.

This study assembled a panel of 94 tobacco accessions from seven countries and used GBS sequencing to identify high-density SNPs. The study aimed to: (1) analyze the SNP distribution, linkage disequilibrium (LD), and population structure using GBS data; (2) detect QTNs related to TBW resistance using GWAS; (3) identify TBW-resistant superior alleles of stable QTNs for MAS and the best parental cross combination based on superior alleles; (4) predict potential candidate genes for TBW resistance in the region of stable QTNs. The results of this study will provide information for uncovering the genetic basis of TBW resistance and facilitating MAS in tobacco breeding.

MATERIALS AND METHODS

Plant Material and Phenotyping

Ninety-four tobacco accessions were obtained from the Nanxiong Scientific Research Institute of Guangdong Tobacco Company, China. These accessions came from the United States, Japan, Canada, Somalia, Australia, Zimbabwe, and China, including 90 flue-cured, two sun-cured, and two burley tobacco accessions (**Supplementary Table 1**).

The 94 accessions were planted at the Hukou experimental station in Nanxiong city in 2013, 2014, and 2015 (denoted E1, E2, and E4) and Xikou experimental station in Nanxiong city in 2014 (denoted E3) in a randomized complete block design with two replicates at each location. Each plot had 20 plants spaced 0.5 m within rows and 1.2 m between rows, with local management practices applied. We used the biochemical type III bacterial wilt pathogen race-1 strain of *R. solanacearum*. The inoculum was applied in early May (May 8, 2013, May 4, 2014, and May 5, 2015) using the stem puncture inoculation method. Disease ratings for each accession occurred on May 30, 2013, May 29, 2014, and May 29, 2015 using the 0–9 scale described in "China National Tobacco Pests Classification and Survey Methods (GB/T23222-2008)": 0 = no lesions; 1 = flecks on stem or leaf wilt $<1/2$ leaf; 3 = lesion on $<1/2$ stem or leaf wilt on 1/2 to 2/3 leaf; 5 = lesion on $>1/2$ but not entire stem or leaf wilt on $>2/3$ leaf; 7 = lesion on entire stem or wilt on entire leaf, and 9 = dead plant (**Figure 1**).

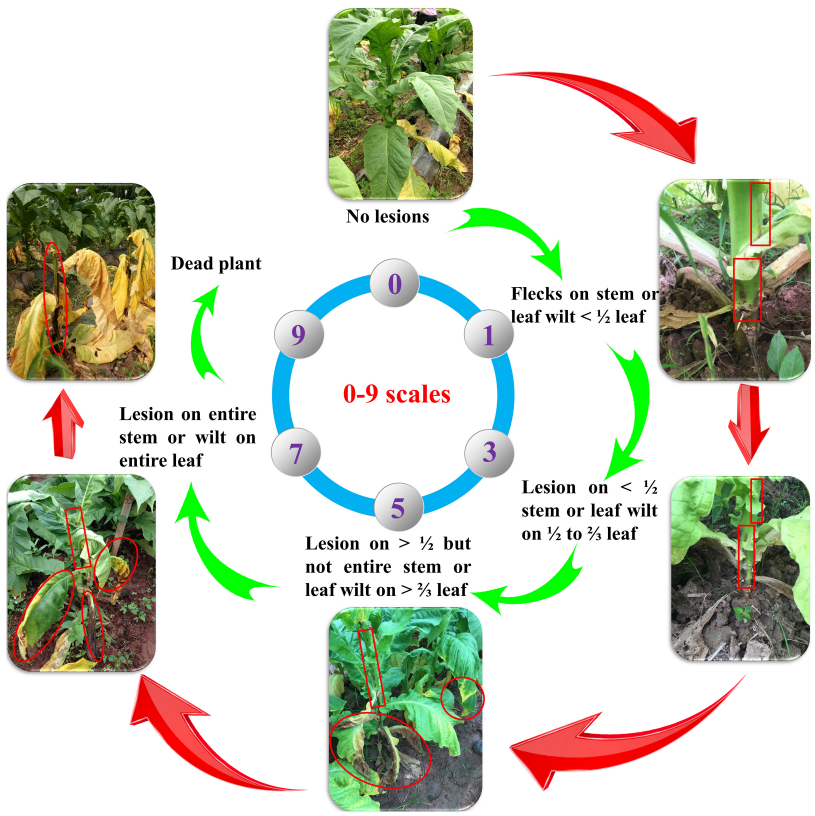

FIGURE 1 | Tobacco bacterial wilt disease rating (0–9 scale).

The disease index for TBW (TBW-DI) was calculated according to Lan et al. (2014):

$$\text{TBW}-\text{DI} = (\text{accession mean rating}/9) \times 100$$

The disease resistance of the 94 accessions was classified using a 0–100 scale following the standard method of "China National Tobacco Varieties Resistance Identification (YC/T 41-1996)": $0 < \text{TBW-DI} \leq 25$ as highly resistant, $25 < \text{TBW-DI} \leq 50$ as moderately resistant, $50 < \text{TBW-DI} \leq 75$ as moderately susceptible, and $75 < \text{TBW-DI} \leq 100$ as highly susceptible.

Statistical Analysis of Phenotypic Data

Mean, range, standard deviation (SD), coefficient of variation (CV%), skewness, kurtosis, and analysis of variance (ANOVA) were calculated for TBW-DI of the 94 tobacco accessions in each environment using R4.0.3[1] software. A best linear unbiased prediction (BLUP) value of TBW-DI for each tobacco accession was calculated using the *lme4* (Bates et al., 2014) statistical package of R. The mixed linear model (MLM) was applied to calculate polygenic and residual error variance components for heritability (Wang et al., 2016) as follows: $y = X\alpha + \varphi + \varepsilon$, where y = phenotypic vector, X = incident matrix for fixed effects, α = vector of fixed effects, $\varphi \sim \text{MVN}(0, K\sigma_g^2)$ = polygenic effect with a multivariate normal distribution with zero mean, and $\varepsilon \sim \text{MVN}(0, I\sigma_e^2)$ = vector of residues. Moreover, σ_g^2, σ_e^2, and K were used as polygenic variance, residual variance, and kinship matrix, respectively. The above two variance components were estimated from the restricted maximum likelihood (REML) method, whereas the kinship matrix was calculated from marker information (Xu, 2013). Broad-sense heritability was calculated as: $h_B^2 = \frac{\sigma_g^2}{\sigma_g^2+\sigma_e^2}$

DNA Extraction and Quantification

Total genomic DNA of the 94 tobacco accessions was isolated from 0.1 g fresh young leaves. DNA extraction was performed using a NuClean Plant Genomic DNA Kit (CWBIO, Beijing, China), according to the manufacturer's protocol. DNA quality and concentration were evaluated using 1% agarose gel electrophoresis and a NanoDrop spectrophotometer. The DNA concentration was normalized to 30 ng/uL for library construction.

Genotyping-by-Sequencing Library Construction, Sequencing, and Single-Nucleotide Polymorphism Calling

For GBS library preparation, genomic DNA was digested with the restriction enzyme *ApeK1*, and libraries with 250–500 bp were constructed in 96-plex using the protocols developed by

[1]http://www.R-project.org/

Elshire et al. (2011). The GBS libraries were sequenced on an Illumina HiSeq^{TM} 2000 instrument. Raw reads were de-multiplexed using a barcode sequence, and the adapter sequences were trimmed using the standard Illumina GA Pipeline v1.5. High-quality clean short reads were aligned to the tobacco reference genome, *N. tabacum* Nitab4.5 (Edwards et al., 2017), using Burrows-Wheeler Aligner (BWA, V0.7.12) software (Li and Durbin, 2009). The SNP variants were extracted using the Unified Genotyper module of GATK software (v3.4-46) in multiple samples (McKenna et al., 2010). The extracted variants were filtered using the following filter parameters: -Window 4, -filter "QD < 4.0 || FS > 60.0 || MQ < 40.0," -G_filter "GQ < 20." ANNOVAR software was used to annotate all filtered high-density SNPs (Wang et al., 2010). The GAPIT software package was used to create the kinship matrix between accessions, kinship matrix heatmap, and physical map of SNPs (Lipka et al., 2012).

Linkage Disequilibrium and Population Structure Analysis

PLINK v1.90 software (Purcell et al., 2007) was used to analyze the LD by calculating the squared correlation coefficients (R^2) of SNPs, using minor allele frequency (MAF) ≥0.05 and a missing rate <20%, and the LD plot was generated using R script. The LD decay rate was observed when the average R^2 decreased to half of its maximum value. The population structure for the 94 accessions was evaluated using STRUCTURE v2.3.4 (Pritchard et al., 2000). The hypothetical subgroup (K) values were set from 2 to 10, with 20,000 iterations for each run, followed by 200,000 Markov chain Monte Carlo (MCMC) replications after burn-in. According to Evanno et al. (2005), the best K was identified using STRUCTURE HARVESTER (Earl and vonHoldt, 2012). A neighbor-joining phylogenetic tree of the 94 accessions was constructed using the filtered SNPs by the Tassel 5.2 software (Bradbury et al., 2007).

Genome-Wide Association Mapping

Genome-wide association studies used SNPs with less than 20% missing data, MAF > 0.05, and sequencing depth ≥3. Four multi-locus (ML) GWAS approaches were used to identify significant QTNs, including mrMLM (Wang et al., 2016), pLARmEB (Zhang et al., 2017), ISIS EM-BLASSO (Tamba et al., 2017), and FASTmrMLM (Tamba and Zhang, 2018), while the Q and K matrix were incorporated into a MLM. These methods were implemented using the R package mrMLM (version 4.0.2).[2] All multi-locus genome-wide association study (ML-GWAS) models use a modified Bonferroni; the number of markers is replaced by the effective number of markers in the correction formulas (Wang et al., 2016; Zhang et al., 2017). Two-step algorithms are involved in all these methods. In the first step, the single (SL) GWAS method scans the entire genome, with putative QTNs identified according to a less stringent threshold level. In the second step, the effects of selected markers are estimated by empirical Bayesian, the significance of the effects from zero were obtained using the likelihood ratio test, and the threshold level LOD ≥ 3 (P = 0.0002) was used to determine significant trait-associated QTNs (Wang et al., 2016; Tamba et al., 2017).

Superior Allele Analysis for Tobacco Bacterial Wilt Resistance

For this purpose, we used stable QTNs identified in multiple environments and/or by multiple GWAS methods. The resistance allele of each stable QTN was determined using code 1 for genotype and QTN effect value. If the QTN effect value is negative, the genotype with code 1 is considered the TBW-resistant superior allele; if the QTN effect value is positive, the alternative genotype is considered the TBW-resistant superior allele (Wang et al., 2016; Zhang et al., 2019). Correlation coefficients between TBW-DI and the number of superior alleles were calculated using R4.0.3 (see text footnote 1) software. The TBW superior allele percentage was calculated for each accession as the number of superior alleles divided by the total number of stable QTNs. For each QTN, the TBW superior allele percentage in the GWAS population was calculated as the number of accessions with superior alleles divided by the total number of accessions. The best parental cross combinations for tobacco breeding programs were predicted using TBW-resistant superior alleles and stable QTN information.

Prediction of Potential Candidate Genes

The search for potential candidate genes based on the stable QTNs detected by multiple methods and/or in multiple environments/BLUP was performed using the *N. tabacum* Nitab4.5 reference genome[3], according to the genome-wide LD decay distance (Edwards et al., 2017). Next, homologous genes related to bacterial wilt in *Arabidopsis* were determined by BLAST analysis with 1E-30 critical *E*-value. These candidate genes were assigned to different biological processes related to bacterial wilt based on the function of their homologs in *Arabidopsis* in literature, such as WRKY transcription factors (TFs) (Cai et al., 2015; Hussain et al., 2018), ethylene-responsive factors (Gutterson and Reuber, 2004), pathogenesis-related proteins (PRs) (Kuhn et al., 2017), Cytochrome P450 family (Li et al., 2021), and brassinosteroids (Sun and Li, 2017). KEGG analysis was used to identify the functional categories (metabolic or signal transduction pathways) of predicted candidate genes, using the KOBAS v3.0 tool[4], with *P*-value < 5% as threshold criteria (Xie et al., 2011).

RESULTS

Phenotypic Variation of Tobacco Bacterial Wilt in a Natural Population

The mean values for TBW-DI across 94 accessions in the four environments (E1–E4) were 52.75, 51.23, 13.34, and 72.04, with SDs of 35.24, 29.42, 16.88, and 26.47, respectively (**Table 1**). The CVs were >50% in all environments except

[2]https://cran.r-project.org/web/packages/mrMLM/index.html

[3]https://solgenomics.net/organism/Nicotiana_tabacum/genome

[4]http://kobas.cbi.pku.edu.cn/kobas3

E4 (36.74%), indicating the highly dispersed distribution of TBW-DI among accessions. The frequency distribution for TBW-DI in the four environments and BLUP is in **Supplementary Figure 1**. Skewness and kurtosis values were <1 in E1, E2, and E4, indicating that TBW-DI followed a normal distribution; in E3, TBW-DI was skewed slightly to the left (**Table 1** and **Supplementary Figure 1**). The two-way ANOVA exhibited significant differences ($P < 0.001$) for genotype, environment, and genotype × environment interaction, suggesting that environmental factors also influence TBW-DI (**Table 1**). Moreover, the heritability estimates for TBW-DI in the four environments ranged from 61.37 to 81.36%, using residual and polygenic variances (**Table 1**), indicating that genetic effects play a significant role in TBW-DI variation.

Genotyping-by-Sequencing of the Test Population

We obtained 1412.73 million raw Illumina sequencing reads for the 94 accessions from the GBS library. After quality control and data filtering, 1370.27 million clean reads were generated, with an average 97.06% effective rate (**Supplementary Table 2**). On average, 98.63% of clean reads had a base error rate of <1% (Q20), and 95.90% of the reads had a base error rate of <0.1% (Q30), with an average GC distribution of 40.92%. Overall, 4.81–23.54 million high-quality sequencing reads were obtained per sample from clean reads, with an average of 9.76 million reads (**Supplementary Table 2**). Finally, 4.45–22.06 million reads were aligned to the reference genome, with an average mapping ratio of 93.01% (**Supplementary Table 2**).

After completing the sequencing, 938,799 SNP variants were called from GBS sequencing data using the GATK process. Among these, 573,312 SNPs were transitions, and 365,487 SNPs were transversions. The SNP data were filtered with MAF ≥ 5%, missing rate <20%, sequence depth ≥ 3 to obtain 126,602 high-quality SNPs, comprising 90,276 transitional and 36,326 transversional SNPs. SNP functional annotation revealed that most identified SNPs were located within intergenic regions (85.83%) of the genome followed by intronic regions (6.68%), coding variants (5.64%), upstream (0.95%), downstream (0.75%), and UTR regions (0.11%) (**Figure 2A**). Further classification of coding SNPs revealed that 54.66 and 44.56% are synonymous and non-synonymous, while stop-gain and stop-loss constituted <1% (**Figure 2B**). Moreover, the SNPs mentioned earlier were distributed on all 24 chromosomes of tobacco (**Figure 2C**), with an average of 5275.08 SNPs per chromosome (**Supplementary Table 3**). The maximal number of SNPs were identified on chromosome Nt17 (9583), while those with minimal numbers were on chromosome Nt02 (2966). The average marker density was approximately 24.46 kb/SNP at the genome level (**Supplementary Table 3**). The highest marker density (16.23 kb/SNP) was on chromosome Nt11, while the lowest marker density (37.94 kb/SNP) was on chromosome Nt15 (**Supplementary Table 3**). These results demonstrate the uneven distribution of markers throughout the tobacco genome.

Population Structure, Linkage Disequilibrium, and Kinship Analysis

A total of 126,602 high-density SNPs were used to define the subgroups/subpopulations within the panel of 94 tobacco accessions. Delta K (K = 1–10) analysis revealed two subpopulations (selected K = 2) comprising 37 (39.40%) and 57 (60.60%) tobacco accessions, respectively (**Figures 3A,B**). Each subpopulation comprised accessions from different ecological zones (**Figure 3B**), indicating that the division of two subpopulations was not related to their geographical origins. Furthermore, a neighbor-joining phylogenetic tree was conducted based on their genetic distances derived from the SNP differences in these accessions. The population could be divided into two subpopulations (**Figure 3C**), and the phylogenetic analysis agreed well with the clustering results in STRUCTURE. The squared correlation coefficient (r^2) values were calculated for all SNP pairs to determine LD decay. The r^2-values decreased rapidly with increasing physical distance between pairwise SNPs (**Figure 3D**). The overall LD decay distance for all chromosomes was estimated at ∼94.56 kb, where r^2 = 0.381 decreased to half its maximum value (**Figure 3D**). Moreover, the pairwise relative kinship coefficients showed a lower level of genetic relatedness among 94 tobacco accessions (**Figure 3E**).

Identification of Quantitative Trait Nucleotides by Multi-Locus Genome-Wide Association Study Methods

The four multi-locus methods identified 142 significant QTNs associated with TBW resistance based on LOD scores ≥3 in the four environments and BLUP (**Figure 4**

TABLE 1 | Statistical analysis of TBW-DI in 94 tobacco accessions in four environments.

Env.	Mean	Range	SD	CV (%)	Skew	Kur	F_G	F_E	$F_{G\times E}$	h^2B (%)
E1	52.75	0.00–100	35.24	66.80	−0.05	−1.38	3.42**	117.76**	14.86**	71.07
E2	51.23	0.00–100	29.42	57.39	−0.1	−1.22				77.78
E3	13.34	1.39–79.44	16.88	78.75	1.53	1.54				81.36
E4	72.04	2.78–100	26.47	36.74	−0.76	−0.32				61.37

*Env: environments, E1: Nanxiong (2013), E2: Nanxiong (2014), E3: Xikou (2014), E4: Nanxiong (2015), SD: standard deviation, CV: coefficient of variation, Skew: skewness, Kur: kurtosis, F_G, F_E, and $F_{G\times E}$: F-values for genotype, environment, and genotype × environment, respectively, h^2B: broad sense heritability. **Significance at the 0.01 level.*

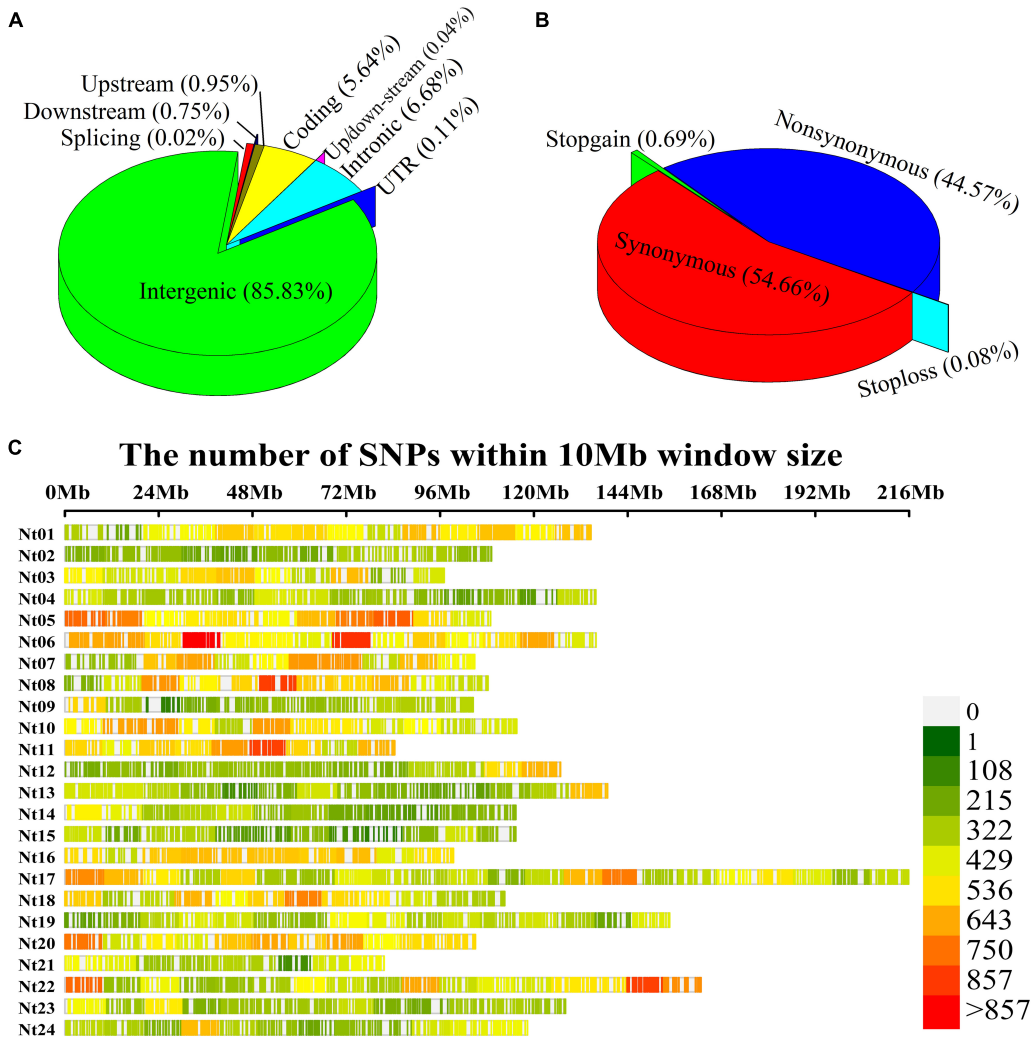

FIGURE 2 | Single-nucleotide polymorphism (SNP) annotation, density, and distribution on 24 chromosomes of tobacco. **(A)** SNP variant classification; **(B)** Coding variant classification; **(C)** Tobacco physical map developed using GBS-SNPs. *Y*-axis represents chromosomes from Nt01 to Nt24, and *X*-axis represents the physical positions of the SNPs on each chromosome.

and **Supplementary Figure 2**). Of these, 26, 34, 38, 26, and 28 QTNs were detected in E1, E2, E3, E4, and BLUP, respectively, explaining 8.18, 7.95, 7.03, 7.93, and 8.22% of the phenotypic variation (PVE) on average (range 0.49–22.52%) (**Table 2**). The corresponding LOD scores ranged from 3.20 to 12.41, 3.02 to 14.19, 3.01 to 15.20, 3.02 to 13.23, and 3.11 to 13.26 (**Supplementary Figure 2**). Among the 142 QTNs, 7–12, 5–13, 8–17, and 10–13 QTNs were identified using FASTmrMLM, ISIS EM-BLASSO, mrMLM, and pLARmEB, respectively, in E1–E4 and BLUP (**Table 2**). The corresponding PVE values ranged from 1.56 to 22.52, 1.14 to 19.16, 0.63 to 20.32, and 0.49 to 19.26, and LOD values ranged from 3.02 to 12.11, 3.10 to 9.56, 3.02 to 14.18, and 3.01 to 15.20 (**Table 2**). Significant QTNs were disseminated on 24 chromosomes, with more than eight QTNs located on chromosomes 1, 4, 7, 10, 17, 20, and 22 (**Supplementary Figure 2**). Additionally, the QTNs identified in this study were not located in or overlapped with the genomic region of previously reported QTLs for bacterial wilt resistance in tobacco.

Environment- and Method-Stable Quantitative Trait Nucleotides for Tobacco Bacterial Wilt Resistance

Two types of QTNs were defined as stable QTNs: those detected in at least two environments/BLUP (environment-stable) and/or by at least two ML-GWAS models (method-stable). In this study, 38 QTNs were identified as stable for TBW resistance (**Table 3** and **Figures 4**, **5A–C**), of which nine were environment-stable (**Table 3** in bold and **Figure 5B**), 37 were method-stable (**Table 3** and **Figure 5C**), and eight were co-detected as environment-stable and method-stable. For example, *qTBW-20-1* was detected in E2, E4,

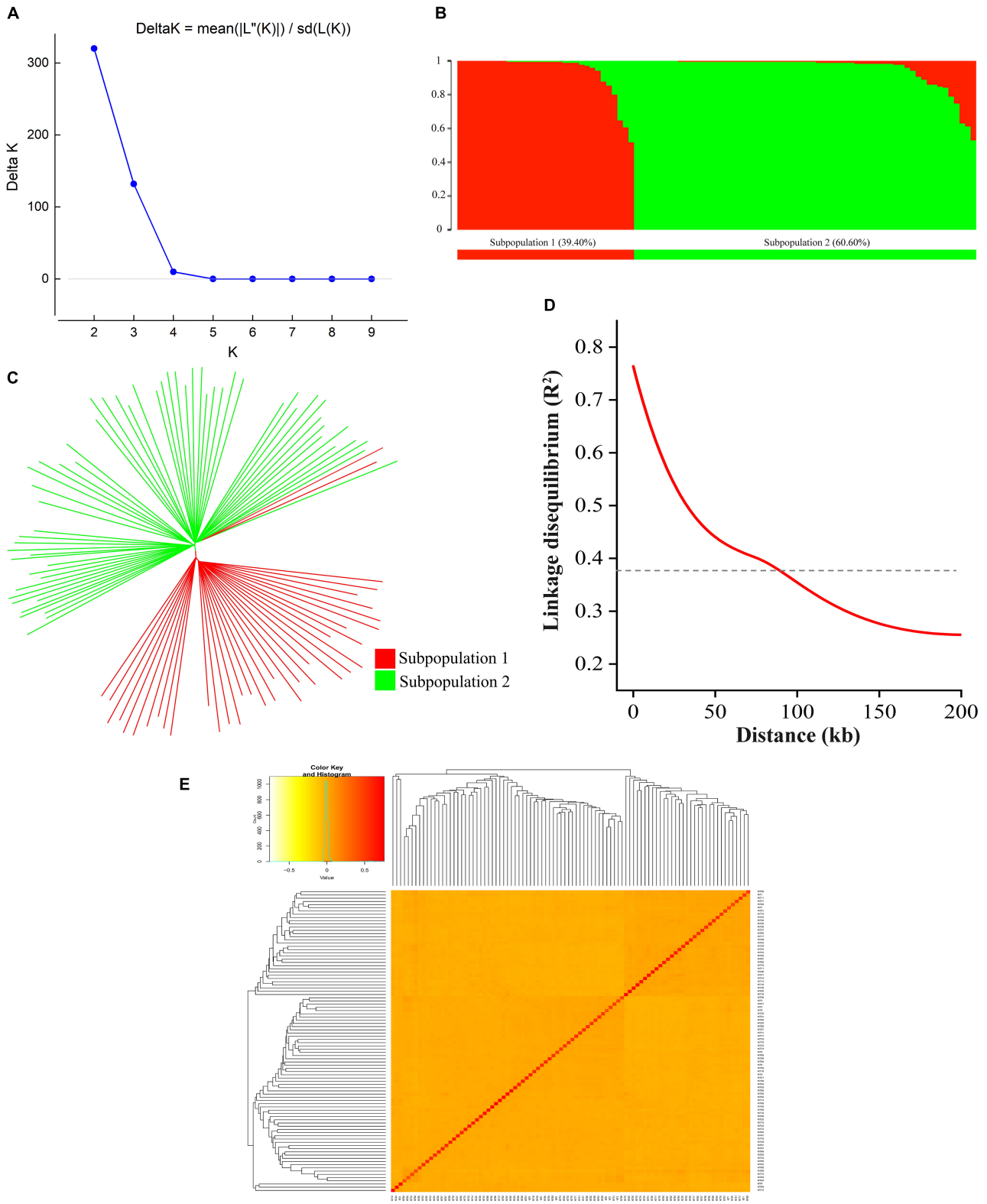

FIGURE 3 | Population structure, phylogenetic tree, LD decay, and kinship of 94 tobacco accessions. **(A)** Relationship between K and Delta K, and determination of subpopulations using Delta method of Evanno et al. (2005); **(B)** Distribution of accessions into subgroups: red and green bars indicate subgroup I and subgroup II, respectively; **(C)** Neighbor-joining phylogenetic tree based on Nei's genetic distances; **(D)** Entire genome LD decay of the population; **(E)** Heatmap of kinship matrix of 94 accessions.

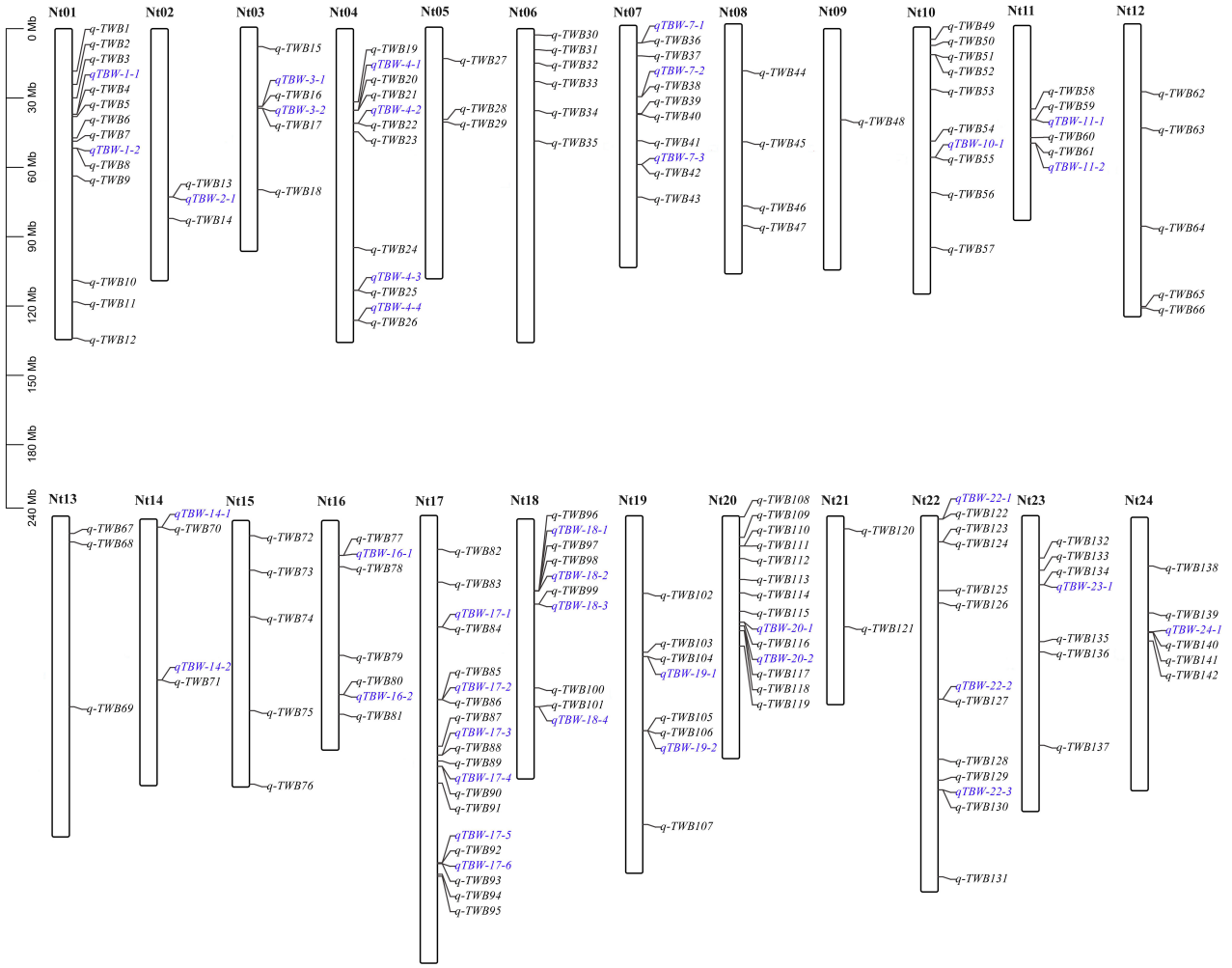

FIGURE 4 | Genetic map based on identified QTNs. The black font represents QTNs identified in E1–E4 and BLUP using four ML-GWAS approaches. The blue font represents QTNs repeatedly identified across at least two environments and/or two ML-GWAS methods.

TABLE 2 | Summary of QTNs detected in multiple environments and BLUP model using four ML-GWAS methods.

Method	Total	E1	E2	E3	E4	BLUP	QTN effect	LOD score	r^2 (%)
Total	142	26	34	38	26	28	−37.02 to 30.64	3.01–15.20	0.49–22.52
FASTmrMLM	45	12	8	7	11	9	−17.91 to 20.13	3.02–12.11	1.56–22.52
ISIS EM-BLASSO	44	10	13	10	5	9	−16.01 to 17.76	3.10–9.56	1.14–19.16
mrMLM	54	10	8	17	12	9	−31.02 to 20.12	3.02–14.18	0.63–20.32
pLARmEB	57	10	11	13	12	13	−17.29 to 30.64	3.01–15.20	0.49–19.26

E1: Nanxiong (2013), E2: Nanxiong (2014), E3: Xikou (2014), E4: Nanxiong (2015), r^2: phenotypic variance explained by each QTN.

and BLUP by three methods, with LOD and PVE values ranging from 3.10 to 9.43 and 8.05 to 19.98, respectively. Likewise, six stable QTNs, *qTBW-7-1*, *qTBW-7-2*, *qTBW-11-1*, *qTBW-18-3*, *qTBW-18-4*, and *qTBW-19-1* were identified in two environments using two, three, two, four, four, and two ML-GWAS methods; their respective LOD values were 4.55–7.55, 3.49–12.60, 3.37–3.78, 3.86–13.27, 3.10–5.40, and 4.78–14.18, and PVE values were 8.42–17.26, 5.79–16.87, 0.49–1.17, 4.77–14.83, 3.91–12.75, and 4.03–13.37 (**Table 3**). Three QTNs (*qTBW-2-1*, *qTBW-14-2*, and *qTBW-17-1*) were detected by all four multi-locus methods in one environment, and eight QTNs (*qTBW-3-1*, *qTBW-10-1*, *qTBW-14-1*, *qTBW-17-4*, *qTBW-19-2*, *qTBW-22-3*, *qTBW-23-1*, and *qTBW-24-1*) were detected by three methods in one environment. Interestingly, only one QTN (*qTBW-22-1*) was identified by a single method

TABLE 3 | Significant QTNs for TBW resistance detected by multiple ML-GWAS methods and/or in multiple environments/BLUP.

QTN name[a]	Chr.	Position (bp)	Effect	LOD score	r^2 (%)[b]	MAF[c]	Environment[d]	Method[e]
qTBW-1-1	Nt01	37016222	−16.95 to −16.82	5.62–9.31	12.86–14.41	0.29	E1	M1, M3
qTBW-1-2	**Nt01**	**51696366**	**4.87 to 5.91**	**3.54–7.79**	**2.50–5.07**	**0.48**	**E3, BLUP**	**M3, M4**
qTBW-2-1	Nt02	72820038	13.19 to 15.98	3.56–6.30	9.68–14.06	0.23	E1	M1, M2, M3, M4
qTBW-3-1	Nt03	34243911	4.64 to 6.30	3.42–4.16	4.17–6.69	0.44	BLUP	M1, M2, M4
qTBW-3-2	Nt03	35123140	11.94 to 15.35	4.58–6.90	6.71–12.29	0.32	E1	M2, M3
qTBW-4-1	Nt04	35325384	5.67 to 8.68	3.53–5.87	1.56–5.64	0.49	E4	M2, M3
qTBW-4-2	Nt04	40894957	15.10 to 16.36	6.70–8.02	12.60–13.35	0.21	E1	M2, M3
qTBW-4-3	Nt04	113176005	−8.05 to −7.19	3.93–4.60	5.06–6.22	0.28	E4	M2, M3
qTBW-4-4	Nt04	126224139	−6.35 to −5.23	3.22–5.91	7.05–10.77	0.38	BLUP	M1, M4
qTBW-7-1	**Nt07**	**7571216**	**5.21** to **14.57**	**4.55–7.55**	**8.42–17.26**	**0.20**	**E2, BLUP**	**M2, M3**
qTBW-7-2	**Nt07**	**30847693**	**7.17** to **11.31**	**3.49–12.60**	**5.79–16.87**	**0.43**	**E3, BLUP**	**M1, M3, M4**
qTBW-7-3	Nt07	60177410	−15.01 to −9.73	3.39–5.72	5.24–12.35	0.21	E1	M2, M4
qTBW-10-1	Nt10	56383019	17.11 to 30.64	4.88–15.20	9.21–19.26	0.48	E3	M1, M3, M4,
qTBW-11-1	**Nt11**	**40714656**	**3.50** to **10.41**	**3.37–3.78**	**0.49–1.17**	**0.45**	**E4, BLUP**	**M1, M3**
qTBW-11-2	Nt11	50906143	7.31 to 9.58	4.51–5.18	4.11–7.95	0.28	E4	M1, M2
qTBW-14-1	Nt14	3465200	−9.70 to −5.04	3.85–7.95	3.38–17.79	0.14	E3	M2, M3, M4,
qTBW-14-2	Nt14	69653438	−14.26 to −8.99	3.98–12.41	3.12–8.68	0.25	E1	M1, M2, M3, M4
qTBW-16-1	Nt16	15141918	13.11 to 15.35	5.87–7.20	11.04–19.16	0.14	E4	M2, M4
qTBW-16-2	Nt16	75282825	−12.77 to −10.82	3.31–5.21	5.22–9.98	0.42	E4	M1, M3
qTBW-17-1	Nt17	48243892	11.26 to 17.76	4.05–12.11	13.97–22.52	0.17	E4	M1, M2, M3, M4
qTBW-17-2	Nt17	79641667	10.73 to 11.80	3.25–6.85	5.08–6.78	0.35	E1	M1, M3
qTBW-17-3	Nt17	103649413	4.74 to 6.53	6.59–11.28	8.06–13.36	0.43	BLUP	M1, M2
qTBW-17-4	Nt17	108558750	−14.37 to −6.04	3.61–12.26	1.93–8.92	0.42	E1	M1, M2, M4,
qTBW-17-5	Nt17	172367035	5.42 to 5.66	3.99–6.90	8.31–9.35	0.33	BLUP	M1, M3
qTBW-17-6	Nt17	172718671	−10.87 to −10.37	3.59–3.75	3.23–3.26	0.45	E1	M2, M3
qTBW-18-1	Nt18	31032756	−10.05 to −7.92	4.59–5.29	5.29–12.68	0.07	E3	M2, M3
qTBW-18-2	Nt18	31188535	4.93 to 7.03	5.68–6.69	8.62–15.28	0.22	BLUP	M1, M2
qTBW-18-3	**Nt18**	**36818226**	**5.99 to 19.46**	**3.86–13.27**	**4.77–14.83**	**0.47**	**E4, BLUP**	**M1, M2, M3, M4**
qTBW-18-4	**Nt18**	**81285311**	**−11.26 to−3.64**	**3.10–5.40**	**3.91–12.75**	**0.36**	**E4, BLUP**	**M1, M2, M3, M4**
qTBW-19-1	**Nt19**	**60933888**	**3.60 to 14.06**	**4.78–14.1**	**4.03–13.37**	**0.48**	**E2, BLUP**	**M1, M3, M4**
qTBW-19-2	Nt19	92990637	−19.86 to −17.29	5.10–6.53	5.80–6.92	0.47	E1	M1, M2, M3,
qTBW-20-1	**Nt20**	**45986131**	**−14.90 to−5.04**	**3.10–9.43**	**8.05–19.98**	**0.21**	**E2, E4, BLUP**	**M2, M3, M4**
qTBW-20-2	Nt20	47568998	3.51 to 5.26	3.78–6.57	1.77–6.66	0.35	E3	M1, M3
qTBW-22-1	**Nt22**	**1437500**	**6.46 to 12.37**	**4.06–5.48**	**7.03–11.16**	**0.14**	**E1, BLUP**	**M4**
qTBW-22-2	Nt22	79361894	3.88 to 7.45	3.13–6.27	4.24–14.13	0.37	BLUP	M2, M4
qTBW-22-3	Nt22	118588843	7.63 to 9.65	3.59–4.10	0.63–1.14	0.33	E4	M1, M3, M4,
qTBW-23-1	Nt23	29888578	10.73 to 20.12	3.68–8.22	1.97–5.77	0.45	E2	M1, M2, M3,
qTBW-24-1	Nt24	49518556	15.12 to 20.13	4.21–5.81	7.32–12.39	0.48	E3	M1, M2, M3,

Normal font indicates stable QTNs detected by at least two methods. Bold font indicates stable QTNs identified in at least two environments/BLUP. [a]*qTBW- -: QTNTBW-chromosome-number.* [b]r^2 *(%): phenotypic variance explained by each QTN.* [c]*MAF: Minor allele frequency.* [d]*M1: mrMLM, M2: FASTmrMLM, M3: pLARmEB, M5: ISIS EM-BLASSO.* [e]*E1: Nanxiong (2013), E2: Nanxiong (2014), E3: Xikou (2014), E4: Nanxiong (2015).*

in two environments, with LOD 4.06–5.48 and PVE 7.03–11.16 (**Table 3**).

Identification of Superior Alleles

The 38 stable QTNs were used to identify superior alleles for TBW resistance using QTN effect values. Thirty-eight superior alleles were identified and significantly ($P < 0.05$) differed from the alternative alleles (**Supplementary Table 4**). The TBW-DI values for the accessions with superior alleles ranged from 31.73 to 46.91, while those for the alternative alleles ranged from 47.1 to 72.44 (**Supplementary Table 4**). For example, *qTBW-14-2* had CC as a superior allele and TT as an alternative allele, and TBW-DI values of 44.22 and 58.04, respectively. Similarly, three stable QTNs, *qTBW-20-2*, *qTBW-22-3*, and *qTBW-23-1*, had TT, CC, and TT superior alleles with TBW-DI values <45 (**Supplementary Table 4**). Moreover, all TBW-DI values for superior alleles of the 38 stable QTNs were <47; according to the disease index, these alleles are considered TBW-resistant superior alleles. A negative correlation was detected between the number of superior alleles and TBW-DI ($r = -0.83$, $p \leq 1 \times 10^{-5}$) (**Figure 6A**). A similar trend was found between the number of superior alleles and TBW-DI in E1 (−0.53,

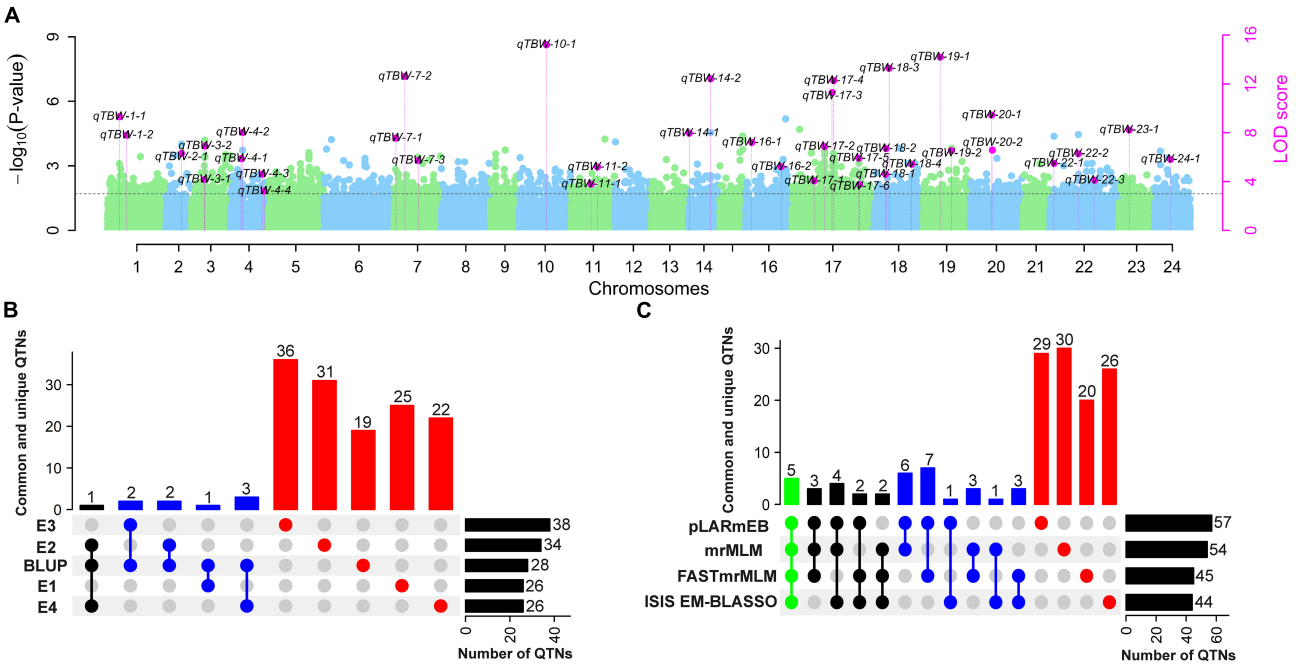

FIGURE 5 | Number of stable QTNs detected for TBW resistance using different ML-GWAS methods in four environments and their BLUP values. **(A)** Pink dots represent stable QTNs identified by more than two methods and/or two environments/BLUP. Light blue and green dots represent the –log10 (*P*-value) of 126,602 markers on 24 chromosomes in the first step of ML-GWAS methods. The threshold level LOD ≥ 3 ($P = 0.0002$) was used to determine significant QTNs and estimated by empirical Bayesian in the second step of ML-GWAS methods; **(B)** Common and unique QTNs identified in E1–E4 and BLUP; **(C)** Common and unique QTNs detected by mrMLM, FastmrMLM, pLARmEB, and ISIS EM-BLASSO. The connected circles below the histogram show overlapping QTNs; green, black, and blue represent four, three, and two overlapping QTNs, respectively, while red represents unique QTNs. Horizontal bars show the total number of QTN set sizes.

$p \leq 1 \times 10^{-5}$), E2 ($r = -0.72$, $p \leq 1 \times 10^{-4}$), E3 ($r = -0.61$, $p \leq 1 \times 10^{-6}$), E4 ($r = -0.63$, $p \leq 1 \times 10^{-4}$), and BLUP ($r = -0.81$, $p \leq 1 \times 10^{-5}$) (**Figures 6B–F**). Based on these results, the superior alleles can be used in MAS for TBW resistance in tobacco.

Distribution of Superior Alleles and Prediction of Best Cross Combination for Tobacco Bacterial Wilt Resistance

The number of TBW-resistant superior alleles for each stable QTN in the 94 accessions ranged from 1 (1.06%) to 91 (96.81%). Among the 38 stable QTNs, 11 had >50% superior alleles, and 27 had <50% superior alleles; only four had >80% superior alleles (**Supplementary Table 4**). The number of superior alleles for each accession ranged from 4 (10.52%) to 24 (63.15%); 15 accessions had >50% superior alleles, and 79 accessions had <50% superior alleles. In addition, K326, Ruyuan No. 1, MC133, C176, ROX28, CO258, H66B, and RG17 had 24, 21, 20, 20, 20, 20, 18, and 18 TBW-resistant superior alleles. These accessions can be used in tobacco breeding programs to increase the number of superior alleles for TBW resistance in one cultivar. For example, a cross between K326 (24 TBW-resistant superior alleles) and C176 (20 TBW-resistant superior alleles) could produce offspring with 30 superior alleles. Based on this information, we predicted the following best parental cross combinations: K326 × C176, MC133 × H66B, MC133 × Ruyuan No. 1, CO258 × ROX28, and CO258 × RG17.

Candidate Genes Underlying Stable Quantitative Trait Nucleotides for Tobacco Bacterial Wilt Resistance

The genomic regions (±95 kb around the associated QTNs) of QTN-linked candidate genes were adopted according to the genome-wide LD decay distance (about 94.5 kb) in this study. As a result, 642 genes were presented in the above regions of the 38 stable QTNs, according to *N. tabacum* reference genome Nitab4.5, of which 489 were homologous in *Arabidopsis*. These 489 genes belonged to different functional categories: stress and defense-related, unknown functional families, hormonal signaling, transcription, translation, transporter, and cell metabolism. Further, the Kyoto encyclopedia of genes and genomes (KEGG, see text footnote 4) analysis of the above 642 genes indicated that 74 genes were involved in 19 KEGG pathways (adjusted *P*-value ≤ 0.05), including flavonoid biosynthesis, glutathione metabolism, MAPK signaling pathway, phenylpropanoid biosynthesis, phosphatidylinositol signaling system, gingerol biosynthesis, and plant–pathogen interactions. Finally, 52 genes were considered potential candidate genes associated with disease resistance in plants (**Supplementary Table 5**) based on

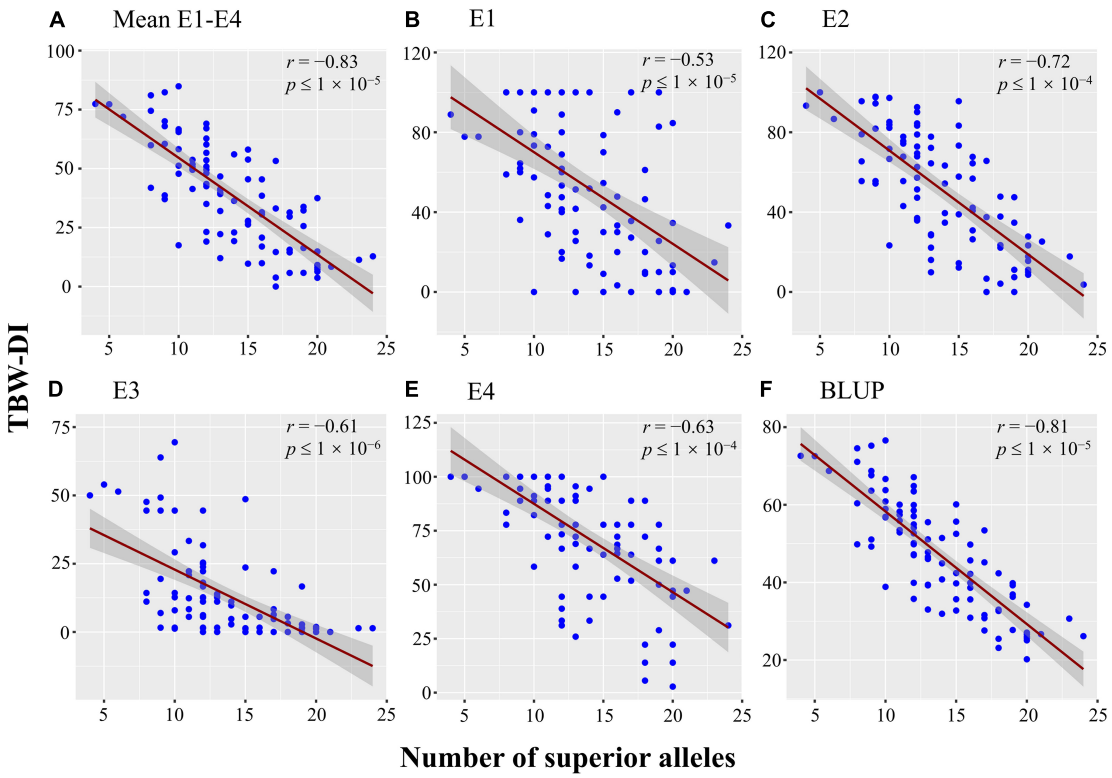

FIGURE 6 | Scatter plot with fitted regression lines and 95% confidence interval bands, showing a negative correlation between TBW-DI in **(A)** Mean E1–E4, **(B)** E1, **(C)** E2, **(D)** E3, **(E)** E4, **(F)** BLUP and number of alleles. The E1, E2, E3, and E4 denote the populations planted in Nanxiong (2013), Nanxiong (2014), Xikou (2014), and Nanxiong (2015), respectively. The *X*-axis represents the number of superior alleles, while the *Y*-axis corresponds to the TBW-DI.

functional annotation, homologous to known genes, and pathway enrichment analysis. For example, candidate gene *Nitab4.5_0002694g0030* underlying the stable QTN *qTBW-3-2* was homologous to *AT4G34050* (*CCoAOMT1*), which annotates caffeoyl-CoA O-methyltransferase and phenylpropanoid in *Arabidopsis*. Similarly, *Nitab4.5_0000274g0070* and *Nitab4.5_0000123g0350*, located near *qTBW-17-3* and *qTBW-24-1*, respectively, corresponded with *Arabidopsis thaliana* genes *AT5G15130* (*WRKY72*) and *AT2G40890* (*CYP98A3*) involved in diarylheptanoid, stilbenoid, gingerol biosynthesis, phenylpropanoid biosynthesis, and flavonoid biosynthesis (**Supplementary Table 5**). Therefore, these candidate genes may regulate tobacco growth to increase plant defense and disease resistance.

DISCUSSION

To study the mechanism of TBW resistance in tobacco plants, GWAS is a useful tool for dissecting the genetic basis and candidate genes for the natural variations in a targeted quantitative trait (Zhang et al., 2019). Here, four ML-GWAS methods were used to analyze TBW-DI and BLUP values using 126,602 high-density SNP markers. We identified 38 stable QTNs, superior alleles, and 52 candidate genes associated with TBW resistance. The markers associated with TBW resistance can be used to develop resistant varieties.

Population Selection for Association Mapping

To gain some insight, we evaluated 94 tobacco accessions for TBW in four different environments at two locations; the variance components indicated that TBW is affected by environmental conditions. The broad-sense heritability for TBW-DI was moderate (61–81%) and differed between environments (**Table 1**). These observations are similar to other studies in tobacco (Nishi et al., 2003; Qian et al., 2013; Lan et al., 2014), including some heritability problems (Gao et al., 2010).

Genetic diversity in modern tobacco cultivars is low (Wernsman, 1999; Leng et al., 2010); only a few genotypes are ancestors of most cultivars. Thus, it is challenging to assemble a natural population with rich genetic diversity. However, based on the 126,602 SNPs, we found 24.46 kb/SNP in the whole genome (**Supplementary Table 3** and **Figure 2C**) and higher coverage density than reported elsewhere (Leng et al., 2010; Wang et al., 2021). We also found an LD decay distance of ∼94.56 kb (**Figure 3D**), much smaller than that reported by Fricano et al. (2012). STRUCTURE analysis identified two subpopulations in the panel (**Figure 3B**), and clustering results showed that the genetic background of 94 accessions is diverse and complex,

consistent with most studies on tobacco (Fricano et al., 2012; Dadras et al., 2014; Wang et al., 2021). There are high genome-wide SNP variations in the panel used in this study that are suitable for association mapping (Mackay and Powell, 2007; Wang et al., 2009; Ikram et al., 2020).

Statistical Power of Multi-Locus Genome-Wide Association Methods and Significance of Stable Quantitative Trait Nucleotides

The four ML-GWAS methods—FASTmrMLM, ISIS EM-BLASSO, mrMLM, and pLARmEB—identified 45, 44, 54, and 57 significant QTNs for TBW-DI, with small to large effects (**Table 2**). While the statistical power of QTN detection has improved, after controlling the polygenic background, most small-effect QTNs of complex traits are not captured by SL-GWAS methods (Zhang et al., 2019). However, ML-GWAS studies have shown that these methods have high-resolution power; e.g., Hou et al. (2018) used SL- and ML-GWAS models to identify 20 QTNs related to the drought stress response using mrMLM, but only three by EMMAX, suggesting that the ML-GWAS methods are more powerful. Likewise, Su et al. (2018) reported that multi-locus methods are robust and more potent than MLM method.

In this study, 38 QTNs were identified in more than two environments and/or ML-GWAS models (**Table 3** and **Figures 4**, **5**); nine were considered environment-stable, 37 were considered method-stable, and eight were considered both. In previous studies, environmental-stable QTNs have gained more attention than method-stable QTNs (Zhang et al., 2019), but recent studies have shown that QTNs detected using multiple methods are also reliable (Cui et al., 2018; Li et al., 2018; Peng et al., 2018; Chaurasia et al., 2020; Ikram et al., 2020). For example, 42 QTNs related to salt stress in wheat were detected using multiple methods (Chaurasia et al., 2020). Hence, environment-stable and method-stable QTNs are more reliable for breeding programs, with similar results reported in other crop plants (Zhang et al., 2019), including soybean (Ikram et al., 2020), maize (Ma et al., 2018), wheat (Chaurasia et al., 2020; Danakumara et al., 2021), and rice (Verma et al., 2021). In addition, the 38 stable QTNs identified in our study are considered novel as they are not located in the genomic region of previously reported QTLs for TBW.

Application of Superior Alleles in Breeding Programs

Tobacco cultivars exhibit low genetic diversity, and their existing alleles may not improve TBW resistance (Qian et al., 2013; Wang et al., 2021). New alleles identified through germplasm screening will improve TBW resistance. Marker-assisted breeding has dramatically improved breeding efficiency. The alleles of stable QTNs significantly differed, with 4–24 superior alleles for TBW resistance found in the 94 accessions (**Supplementary Table 4**). Eight resistant genotypes were identified with superior alleles for TBW resistance that can be used to breed highly resistant varieties. The best cross combinations were identified based on these superior alleles for TBW resistance, similar to previous studies for complex traits in different crops (Wang et al., 2006; Zeng et al., 2017). The concept of molecular design breeding (Peleman and Van Der Voort, 2003; Zeng et al., 2017) was used by Tian and his co-workers to successfully selected the LYP9 rice variety with high yield and quality by transferring several alleles into the new cultivar (Tian et al., 2009; Zeng et al., 2017).

Candidate Genes for Tobacco Bacterial Wilt Resistance

The identification of candidate genes associated with quantitatively inherited traits is challenging in genetic research. The present study identified 52 candidate genes underlying the 38 stable QTNs, based on homology with *Arabidopsis* and KEGG pathways for plant defense and disease resistance (**Supplementary Table 5**). The gene *Nitab4.5_0002694g0030* encodes the caffeoyl-CoA O-methyltransferase-like protein that may prevent TBW by regulating the phenylpropanoid pathway and lignin production (Yang et al., 2017). *Nitab4.5_0001039g0060* is homologous to *CYCD3;2* in *Arabidopsis* (**Supplementary Table 5**), and *CYCD3* genes appear to be positive regulators of plant resistance because mutations in the target gene conferred increased disease susceptibility to plant pathogens (Hamdoun et al., 2016). Similarly, *Nitab4.5_0000337g0220* encodes BR-signaling kinase 1 (*BSK1*), and a bsk1-1 mutation displayed enhanced susceptibility to a range of pathogens, demonstrating that *BSK1* plays an important role in plant immunity (Shi et al., 2013). *BSK1* is a substrate of the brassinosteroid receptor BRI1 and plays a critical role in brassinosteroid signaling to regulate plant immunity (Tang et al., 2008; Sun and Li, 2017). Most of the candidate genes (*Nitab4.5_0000430g0170*, *Nitab4.5_0000016g0210*, *Nitab4.5_0002576g0050*, *Nitab4.5_0000553g0050*, and *Nitab4.5_0002890g0050*) were involved in signaling pathways, and their homologous genes (*MKK9*, *bZIP65*, *AT1G17345*, *EPF2*, and *EMB14*, respectively) in *A. thaliana* play a significant role in plant disease resistance (Feys and Parker, 2000; Eshraghi et al., 2014; Liu and Lam, 2019). *Nitab4.5_0000641g0050*—involved in the glutathione metabolism pathway and glutathione—is the most abundant antioxidant in cells and crucial for life processes. It protects DNA, biomolecules, and proteins against oxidative damage, which favored resistance against environmental stresses (Freeman et al., 2004; Li et al., 2021) and increased resistance in eggplant after infection with *R. solanacearum* (Avinash et al., 2017). TFs are essential regulatory genes in plants, and WRKY TFs play a significant role in the immune response of plants to various biological stresses (Chen et al., 2017). Two WRKY TF genes (*Nitab4.5_0000929g0030* and *Nitab4.5_0000274g0070*) were identified in this study (**Supplementary Table 5**). Numerous research findings have shown that *WRKY22* and *WRKY40* TF genes have positive regulatory effects on the resistance of Solanaceae crops to bacterial wilt

(Dang et al., 2014; Cai et al., 2015; Hussain et al., 2018). These results suggest that WRKY genes could be important positive regulators of tobacco plant resistance against bacterial wilt (Hussain et al., 2018). The relationship between these candidate genes and TBW resistance needs to be verified.

The fundamental task is to find excellent genes or QTNs related to the target trait to achieve the precise breeding, design, and breed aggregate of excellent genes/alleles. Most researchers have only used linkage analysis or association analysis to identify QTLs or SNPs/QTNs. Previous studies only contain basic theoretical results (Nishi et al., 2003; Lan et al., 2014), with few researchers using these results to screen material. However, pleiotropic genes regulate quantitative traits, their genetic laws are complex (Yang et al., 2012), and it is difficult for a single QTL to reflect the advantages of traits. Here, 38 stable QTNs and 52 candidate genes were detected by GWAS, filling a gap and laying a theoretical foundation for subsequent design and breeding. Using research results to evaluate material phenotypes will assist in selecting material containing multiple superior alleles to increase the probability of selecting material with desired traits, which has important implications for molecular marker-assisted screening. Therefore, this study screened tobacco varieties that carry the target QTN alleles and candidate genes that could be used as resistant parents for gene pyramiding to improve TBW resistance.

CONCLUSION

In this study, we used GBS technology for the first time to conduct GWAS for TBW resistance to identify QTNs, superior alleles, and candidate genes for breeding highly resistant tobacco varieties. We demonstrated that TBW resistance is genetically complex. We identified 38 novel stable QTNs with significantly different alleles in the association panel. We predicted the five best parental cross combinations based on superior allele information for developing tobacco varieties that are highly resistant to *R. solanacearum*. Moreover, 52 candidate genes were associated with TBW resistance. The results from this study serve as the basis for resistance gene cloning and further understanding of the molecular mechanisms of tobacco resistance to *R. solanacearum*.

DATA AVAILABILITY STATEMENT

The sequence read data from the genotyping-by-sequencing (GBS) for the 94 sequenced tobacco accession are available in NCBI Sequence Read Archive (SRA) under the accession number PRJNA759331 (https://www.ncbi.nlm.nih.gov/sra/PRJNA759331).

AUTHOR CONTRIBUTIONS

PG conceived and designed the experiments. RLa, RLi, YX, MI, and WZ performed the experiments and analyzed data. YX, MI, RLi, QY, ZZ, and WZ contributed to reagents, materials, and analysis tools. RLa, MI, RLi, YX, KS, and PG wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

The authors declare that this study received funding from the Scientific and Technological Projects of Guangdong Tobacco Corporation (201403 and 201702). The funder was not involved in the study design, data collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.744175/full#supplementary-material

REFERENCES

Avinash, P., Umesha, S., and Marahel, S. (2017). Role of hydrogen peroxide and ascorbate-glutathione pathway in host resistance to bacterial wilt of eggplant. *Russ. J. Plant Physiol.* 64, 375–385. doi: 10.1134/S1021443717030049

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

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

Buckler, E. S., and Thornsberry, J. M. (2002). Plant molecular diversity and applications to genomics. *Curr. Opin. Plant Biol.* 5, 107–111. doi: 10.1016/S1369-5266(02)00238-8

Cai, H., Yang, S., Yan, Y., Xiao, Z., Cheng, J., Wu, J., et al. (2015). CaWRKY6 transcriptionally activates CaWRKY40, regulates Ralstonia solanacearum resistance, and confers high temperature and high-humidity tolerance in pepper. *J. Exp. Bot.* 66, 3163–3174. doi: 10.1093/jxb/erv125

Chaurasia, S., Singh, A. K., Songachan, L. S., Sharma, A. D., Bhardwaj, R., and Singh, K. (2020). Multi-locus genome-wide association studies reveal novel genomic regions associated with vegetative stage salt tolerance in bread wheat (*Triticum aestivum* L.). *Genomics* 112, 4608–4621. doi: 10.1016/j.ygeno.2020.08.006

Chen, F., Hu, Y., Vannozzi, A., Wu, K., Cai, H., Qin, Y., et al. (2017). The WRKY transcription factor family in model plants and crops. *CRC. Crit. Rev. Plant Sci.* 36, 311–335. doi: 10.1080/07352689.2018.1441103

Cui, Y., Zhang, F., and Zhou, Y. (2018). The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. *Front. Plant Sci.* 9:1464. doi: 10.3389/fpls.2018.01464

Dadras, A. R., Sabouri, H., Nejad, G. M., Sabouri, A., and Shoai-Deylami, M. (2014). Association analysis, genetic diversity and structure analysis of tobacco based on AFLP markers. *Mol. Biol. Rep.* 41, 3317–3329. doi: 10.1007/s11033-014-3194-6

Danakumara, T., Kumari, J., Singh, A. K., Sinha, S. K., Pradhan, A. K., Sharma, S., et al. (2021). Genetic dissection of seedling root system architectural traits in a diverse panel of hexaploid wheat through multi-locus genome-wide association mapping for improving drought tolerance. *Int. J. Mol. Sci.* 22:7188. doi: 10.3390/ijms22137188

Dang, F., Wang, Y., She, J., Lei, Y., Liu, Z., Eulgem, T., et al. (2014). Overexpression of CaWRKY27, a subgroup IIe WRKY transcription factor of *Capsicum annuum*, positively regulates tobacco resistance to *Ralstonia solanacearum* infection. *Physiol. Plant.* 150, 397–411. doi: 10.1111/ppl.12093

Denny, T. (2006). "Plant pathogenic Ralstonia species," in *Plant-Associated Bacteria*, ed. S. S. Gnanamanickam (Dordrecht: Springer Netherlands), 573–644.

Drake-Stowe, K., Bakaher, N., Goepfert, S., Philippon, B., Mark, R., Peterson, P., et al. (2017). Multiple disease resistance loci affect soilborne disease resistance in tobacco (*Nicotiana tabacum*). *Phytopathology* 107, 1055–1061. doi: 10.1094/PHYTO-03-17-0118-R

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Edwards, K. D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A. D., Bombarely, A., et al. (2017). A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* 18:448. doi: 10.1186/s12864-017-3791-6

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

Eshraghi, L., Anderson, J. P., Aryamanesh, N., McComb, J. A., Shearer, B., and Giles, G. E. (2014). Defence signalling pathways involved in plant resistance and phosphite-mediated control of Phytophthora Cinnamomi. *Plant Mol. Biol. Rep.* 32, 342–356. doi: 10.1007/s11105-013-0645-5

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

Feys, B. J., and Parker, J. E. (2000). Interplay of signaling pathways in plant disease resistance. *Trends Genet.* 16, 449–455. doi: 10.1016/S0168-9525(00)02107-7

Flint-Garcia, S. A., Thornsberry, J. M., and Edward, S. B. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907

Freeman, J. L., Persans, M. W., Nieman, K., Albrecht, C., Peer, W., Pickering, I. J., et al. (2004). Increased glutathione biosynthesis plays a role in nickel tolerance in Thlaspi nickel hyperaccumulators W inside box sign. *Plant Cell* 16, 2176–2191. doi: 10.1105/tpc.104.023036

Fricano, A., Bakaher, N., Corvo, M. D., Piffanelli, P., Donini, P., Stella, A., et al. (2012). Molecular diversity, population structure, and linkage disequilibrium in a worldwide collection of tobacco (*Nicotiana tabacum* L.) germplasm. *BMC Genet.* 13:18. doi: 10.1186/1471-2156-13-18

Gao, J. M., Wang, Z. D., and Zhang, W. X. (2010). Genetic analysis on resistance to bacterial wilt in oriental tobacco. *Chin. Tob. Sci* 31, 1–4.

Gutterson, N., and Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. *Curr. Opin. Plant Biol.* 7, 465–471. doi: 10.1016/j.pbi.2004.04.007

Hamdoun, S., Zhang, C., Gill, M., Kumar, N., Churchman, M., Larkin, J. C., et al. (2016). Differential roles of two homologous cyclin-dependent kinase inhibitor genes in regulating cell cycle and innate immunity in arabidopsis1. *Plant Physiol.* 170, 515–527. doi: 10.1104/pp.15.01466

Hou, S., Zhu, G., Li, Y., Li, W., Fu, J., Niu, E., et al. (2018). Genome-wide association studies reveal genetic variation and candidate genes of drought stress related traits in cotton (*Gossypium hirsutum* l.). *Front. Plant Sci.* 9:1276. doi: 10.3389/fpls.2018.01276

Hussain, A., Li, X., Weng, Y., Liu, Z., Ashraf, M., Noman, A., et al. (2018). CaWRKY22 acts as a positive regulator in pepper response to *Ralstonia Solanacearum* by constituting networks with CaWRKY6, CaWRKY27, CaWRKY40, and CaWRKY58. *Int. J. Mol. Sci.* 19:1426. doi: 10.3390/ijms19051426

Hyun, D. Y., Sebastin, R., Lee, G., Lee, K. J., Kim, S., Yoo, E., et al. (2021). Genome-wide SNP markers for genotypic and phenotypic differentiation of melon (*Cucumis melo* L.) Varieties Using Genotyping-by-Sequencing. *Int. J. Mol. Sci.* 22:6722.

Ikram, M., Han, X., Zuo, J. F., Song, J., Han, C. Y., Zhang, Y. W., et al. (2020). Identification of QTNs and their candidate genes for 100-seed weight in soybean (*Glycine max* L.) using multi-locus genome-wide association studies. *Genes (Basel)* 11, 1–22. doi: 10.3390/genes11070714

Jiang, G., Wei, Z., Xu, J., Chen, H., Zhang, Y., She, X., et al. (2017). Bacterial wilt in China: History, current status, and future perspectives. *Front. Plant Sci.* 8:1549. doi: 10.3389/fpls.2017.01549

Jing, Y., Teng, W., Qiu, L., Zheng, H., Li, W., Han, Y., et al. (2021). Genetic dissection of soybean partial resistance to sclerotinia stem rot through genome wide association study and high throughput single nucleotide polymorphisms. *Genomics* 113, 1262–1271. doi: 10.1016/j.ygeno.2020.10.042

Kuchel, H., Ye, G., Fox, R., and Jefferies, S. (2005). Genetic and economic analysis of a targeted marker-assisted wheat breeding strategy. *Mol. Breed.* 16, 67–78. doi: 10.1007/s11032-005-4785-7

Kuhn, H., Lorek, J., Kwaaitaal, M., Consonni, C., Becker, K., Micali, C., et al. (2017). Key components of different plant defense pathways are dispensable for powdery mildew resistance of the arabidopsis mlo2 mlo6 mlo12 triple mutant. *Front. Plant Sci.* 8:1006. doi: 10.3389/fpls.2017.01006

Lan, T., Zheng, S., Yang, L., Wu, S., Wang, B., Zhang, S., et al. (2014). Mapping of quantitative trait loci conferring resistance to bacterial wilt in tobacco (*Nicotiana tabacum* L.). *Plant Breed.* 133, 672–677. doi: 10.1111/pbr.12202

Lee, S. J., Ban, S. H., Kim, G. H., Kwon, S. 2nd, Kim, J. H., and Choi, C. (2017). Identification of potential gene-associated major traits using GBS-GWAS for Korean apple germplasm collections. *Plant Breed.* 136, 977–986. doi: 10.1111/pbr.12544

Leng, X., Xiao, B., Wang, S., Gui, Y., Wang, Y., Lu, X., et al. (2010). Identification of NBS-type resistance gene homologs in tobacco genome. *Plant Mol. Biol. Rep.* 28, 152–161. doi: 10.1007/s11105-009-0134-z

Li, C., Fu, Y., Sun, R., Wang, Y., and Wang, Q. (2018). Single-locus and multi-locus genome-wide association studies in the genetic dissection of fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Front. Plant Sci.* 9:1083. doi: 10.3389/fpls.2018.01083

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, T., Ma, X., Li, N., Zhou, L., Liu, Z., Han, H., et al. (2017). Genome-wide association study discovered candidate genes of Verticillium wilt resistance in upland cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* 15, 1520–1532. doi: 10.1111/pbi.12734

Li, Y., Feng, J., Liu, H., Wang, L., Hsiang, T., Li, X., et al. (2016). Genetic diversity and pathogenicity of ralstonia solanacearum causing tobacco bacterial wilt in China. *Plant Dis.* 100, 1288–1296. doi: 10.1094/PDIS-04-15-0384-RE

Li, Y. Y., Wang, L., Sun, G. W., Li, X. H., Chen, Z. G., Feng, J., et al. (2021). Digital gene expression analysis of the response to *Ralstonia solanacearum* between resistant and susceptible tobacco varieties. *Sci. Rep.* 11:3887. doi: 10.1038/s41598-021-82576-8

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

Liu, J. Z., and Lam, H. M. (2019). Signal transduction pathways in plants for resistance against pathogens. *Int. J. Mol. Sci.* 20:2335. doi: 10.3390/ijms20092335

Ma, L., Liu, M., Yan, Y., Qing, C., Zhang, X., Zhang, Y., et al. (2018). Genetic dissection of maize embryonic callus regenerative capacity using multi-locus genome-wide association studies. *Front. Plant Sci.* 9:561. doi: 10.3389/fpls.2018.00561

Mackay, I., and Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* 12, 57–63. doi: 10.1016/j.tplants.2006.12.001

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Nakaya, A., and Isobe, S. N. (2012). Will genomic selection be a practical method for plant breeding? *Ann. Bot.* 110, 1303–1316. doi: 10.1093/aob/mcs109

Ni, C., Xu, X. H., and Zhang, X. W. (2011). Genetic analysis of easy curing potential in flue-cured tobacco with the mixed major-gene plus polygene inheritance model. *Chin. Tob. Sci* 32, 1–4.

Nishi, T., Tajima, T., Noguchi, S., Ajisaka, H., and Negishi, H. (2003). Identification of DNA markers of tobacco linked to bacterial wilt resistance. *Theor. Appl. Genet.* 106, 765–770. doi: 10.1007/s00122-002-1096-9

Peleman, J. D., and Van Der Voort, J. R. (2003). Breeding by design. *Trends Plant Sci.* 8, 330–334. doi: 10.1016/S1360-1385(03)00134-1

Peng, Y., Liu, H., Chen, J., Shi, T., Zhang, C., Sun, D., et al. (2018). Genome-wide association studies of free amino acid levels by six multi-locus models in bread wheat. *Front. Plant Sci.* 9:1196. doi: 10.3389/fpls.2018.01196

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multi-locus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

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., Wang, X., Wang, D., Zhang, L., Zu, C., Gao, Z., et al. (2013). The detection of QTLs controlling bacterial wilt resistance in tobacco (*N. tabacum* L.). *Euphytica* 192, 259–266. doi: 10.1007/s10681-012-0846-2

Ribaut, J. M., de Vicente, M. C., and Delannay, X. (2010). Molecular breeding in developing countries: challenges and perspectives. *Curr. Opin. Plant Biol.* 13, 213–218. doi: 10.1016/j.pbi.2009.12.011

Sakiroglu, M., and Brummer, E. C. (2017). Identification of loci controlling forage yield and nutritive value in diploid alfalfa using GBS-GWAS. *Theor. Appl. Genet.* 130, 261–268. doi: 10.1007/s00122-016-2782-3

Sharma, S., Katoch, V., and Banyal, D. K. (2021). Review on harnessing biotechnological tools for the development of stable bacterial wilt resistant solanaceous vegetable crops. *Sci. Hortic. (Amsterdam).* 285:110158. doi: 10.1016/j.scienta.2021.110158

Shi, H., Yan, H., Li, J., and Tang, D. (2013). BSK1, a receptor-like cytoplasmic kinase, involved in both BR signaling and innate immunity in Arabidopsis. *Plant Signal. Behav.* 8:e24996. doi: 10.4161/psb.24996

Smith, T. E., and Clayton, E. E. (1948). Inheritance of resistance to bacterial wilt in tobacco. *J. Agric. Res.* 76, 27–32.

Su, J., Ma, Q., Li, M., Hao, F., and Wang, C. (2018). Multi-locus genome-wide association studies of fiber-quality related traits in Chinese early-maturity upland cotton. *Front. Plant Sci.* 9:1169. doi: 10.3389/fpls.2018.01169

Sun, C., and Li, J. (2017). Biosynthesis, catabolism, and signal transduction of brassinosteroids. *Zhiwu Shengli Xuebao/Plant Physiol. J.* 53, 291–307. doi: 10.13592/j.cnki.ppj.2017.1002

Tamba, C. L., Ni, Y.-L., and Zhang, Y.-M. (2017). Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput. Biol.* 13:e1005357. doi: 10.1371/journal.pcbi.1005357

Tamba, C. L., and Zhang, Y.-M. (2018). A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv [preprint]* 341784. doi: 10.1101/341784

Tang, W., Kim, T. W., Oses-Prieto, J. A., Sun, Y., Deng, Z., Zhu, S., et al. (2008). BSKs mediate signal transduction from the receptor kinase BRI1 in Arabidopsis. *Science (80-.).* 321, 557–560. doi: 10.1126/science.1156973

Thapa, R., Singh, J., Gutierrez, B., Arro, J., and Khan, A. (2021). Genome-wide association mapping identifies novel loci underlying fire blight resistance in apple. *Plant Genome* 14:e20087. doi: 10.1002/tpg2.20087

Tian, Z., Qian, Q., Liu, Q., Yan, M., Liu, X., Yan, C., et al. (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21760–21765. doi: 10.1073/pnas.0912396106

Verma, R. K., Chetia, S. K., Dey, P. C., Rahman, A., Saikia, S., Sharma, V., et al. (2021). Genome-wide association studies for agronomical traits in winter rice accessions of Assam. *Genomics* 113, 1037–1047. doi: 10.1016/j.ygeno.2020.11.033

Wang, J., Wan, X., Crossa, J., Crouch, J., Weng, J., Zhai, H., et al. (2006). QTL mapping of grain length in rice (*Oryza sativa* L.) using chromosome segment substitution lines. *Genet. Res.* 88, 93–104. doi: 10.1017/S0016672306008408

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164–e164. doi: 10.1093/nar/gkq603

Wang, M. L., Zhu, C., Barkley, N. A., Chen, Z., Erpelding, J. E., Murray, S. C., et al. (2009). Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theor. Appl. Genet.* 120, 13–23. doi: 10.1007/s00122-009-1155-6

Wang, S.-B., Feng, J.-Y., Ren, W.-L., Huang, B., Zhou, L., Wen, Y.-J., et al. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* 6:19444. doi: 10.1038/srep19444

Wang, Y., Lv, H., Xiang, X., Yang, A., Feng, Q., Dai, P., et al. (2021). Construction of a SNP fingerprinting database and population genetic analysis of cigar tobacco germplasm resources in China. *Front. Plant Sci.* 12:618133. doi: 10.3389/fpls.2021.618133

Wernsman, E. A. (1999). An overview of tobacco breeding-past, present, and future. *TSRC Sept.* 25, 5–35.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322. doi: 10.1093/nar/gkr483

Xu, S. (2013). Mapping quantitative trait loci by controlling polygenic background effects. *Genetics* 195, 1209–1222. doi: 10.1534/genetics.113.157032

Yang, G., Li, Y., Wang, Q., Zhou, Y., Zhou, Q., Shen, B., et al. (2012). Detection and integration of quantitative trait loci for grain yield components and oil content in two connected recombinant inbred line populations of high-oil maize. *Mol. Breed.* 29, 313–333. doi: 10.1007/s11032-011-9548-z

Yang, Q., He, Y., Kabahuma, M., Chaya, T., Kelly, A., Borrego, E., et al. (2017). A gene encoding maize caffeoyl-CoA O-methyltransferase confers quantitative resistance to multiple pathogens. *Nat. Genet.* 49, 1364–1372. doi: 10.1038/ng.3919

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702

Zeng, D., Tian, Z., Rao, Y., Dong, G., Yang, Y., Huang, L., et al. (2017). Rational design of high-yield and superior-quality rice. *Nat. Plants* 3:17031. doi: 10.1038/nplants.2017.31

Zhang, J., Feng, J. Y., Ni, Y. L., Wen, Y. J., Niu, Y., Tamba, C. L., et al. (2017). PLARmEB: Integration of least angle regression with empirical Bayes for multi-locus genome-wide association studies. *Heredity* 118, 517–524. doi: 10.1038/hdy.2017.8

Zhang, Y. M., Jia, Z., and Dunwell, J. M. (2019). Editorial: the applications of new multi-locus gwas methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* 10:100. doi: 10.3389/fpls.2019.00100

Conflict of Interest: WZ is employed by Nanxiong Research Institute of Guangdong Tobacco Co., Ltd., Nanxiong, China.

The remaining 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 October 2021
doi: 10.3389/fgene.2021.743902

Transcriptome-Based Identification and Functional Characterization of NAC Transcription Factors Responsive to Drought Stress in *Capsicum annuum* L.

Dionis Borràs[1†], Lorenzo Barchi[1†], Karina Schulz[2†], Andrea Moglia[1†], Alberto Acquadro[1†], Iman Kamranfar[3†], Salma Balazadeh[2,4†] and Sergio Lanteri[1†]*

[1]Department of Agricultural, Forest and Food Sciences, Plant Genetics and Breeding, University of Torino, Turin, Italy, [2]Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany, [3]Department Molecular Biology, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany, [4]Plant Sciences and Natural Products, Institute of Biology Leiden (IBL), Leiden University, Leiden, Netherlands

Capsicum annuum L. is one of the most cultivated Solanaceae species, and in the open field, water limitation leading to drought stress affects its fruit quality, fruit setting, fruit size and ultimately yield. We identified stage-specific and a common core set of differentially expressed genes, following RNA-seq transcriptome analyses of a breeding line subjected to acute drought stress followed by recovery (rewatering), at three stages of plant development. Among them, two NAC transcription factor (TF) genes, i.e., CaNAC072 and CaNAC104, were always upregulated after drought stress and downregulated after recovery. The two TF proteins were observed to be localized in the nucleus following their transient expression in *Nicotiana benthamiana* leaves. The expression of the two NACs was also induced by NaCl, polyethylene glycol (PEG) and abscisic acid (ABA) treatments, suggesting that *CaNAC072* is an early, while *CaNAC104* is a late abiotic stress-responsive gene. Virus-induced gene silencing (VIGS) of *CaNAC104* did not affect the pepper plantlet's tolerance to drought stress, while VIGS of *CaNAC072* increased drought tolerance. Heterologous expression of *CaNAC072* in *Arabidopsis thaliana* as well as in plants mutated for its homolog *ANAC072* did not increase drought stress tolerance. This highlights a different role of the two NAC homologs in the two species. Here, we discuss the complex role of NACs as transcriptional switches in the response to drought stress in bell pepper.

Keywords: bell pepper, transcriptome, drought tolerance, NAC, VIGS, functional characterization

OPEN ACCESS

Edited by:
Nunzio D'Agostino,
University of Naples Federico II, Italy

Reviewed by:
Neftali Ochoa-Alejo,
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Mexico
Allen Van Deynze,
University of California, Davis, United States

****Correspondence:***
Sergio Lanteri
sergio.lanteri@unito.it

†ORCID:
Dionis Borràs
orcid.org/0000-0003-2793-7832
Lorenzo Barchi
orcid.org/0000-0001-6414-4239
Karina Schulz
orcid.org/0000-0002-6540-2209
Andrea Moglia
orcid.org/0000-0001-7431-844X
Alberto Acquadro
orcid.org/0000-0002-5322-9701
Iman Kamranfar
orcid.org/0000-0001-9632-3717
Salma Balazadeh
orcid.org/0000-0002-5789-4071
Sergio Lanteri
orcid.org/0000-0003-3012-8710

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

Received: *19 July 2021*
Accepted: *28 September 2021*
Published: *22 October 2021*

Citation:
Borràs D, Barchi L, Schulz K, Moglia A, Acquadro A, Kamranfar I, Balazadeh S and Lanteri S (2021) Transcriptome-Based Identification and Functional Characterization of NAC Transcription Factors Responsive to Drought Stress in Capsicum annuum *L. Front. Genet. 12:743902. doi: 10.3389/fgene.2021.743902*

INTRODUCTION

Drought stress is one of the key limiting factors affecting plant growth, development and survival, with a substantial impact on crop yield (Sinclair, 2011; Boyer et al., 2013). The lack of water, due to rising temperatures and changes in precipitation patterns, represents a serious problem in many parts of the world, and it is expected to become a cause of food shortage and malnutrition with the increase in population and food demand (Newton et al., 2011; Fitton et al., 2019). Plants have evolved

various molecular mechanisms to adjust their growth to limited water availability, and the elucidation of such mechanisms is essential for implementing breeding strategies aimed at developing crop varieties more resilient and capable to deal with water shortage while maintaining yield (Osakabe et al., 2014; Iovieno et al., 2016). Bell pepper (*Capsicum annuum* L.), also known as sweet pepper, is one of the most widely cultivated solanaceous vegetables (FAOSTAT, 2018), and particularly at flowering, water deficit causes flower dropping which reduces the setting and size of the fruits, and alters their biochemical composition (Borras et al., 2020).

Comparative transcriptomic analyses have proven to be a valuable approach for identifying key genes controlling the response to low water availability (Seki et al., 2002) in plants such as Arabidopsis (Zhang et al., 2019), rice (Chung et al., 2018), maize (Kakumanu et al., 2012), potato (Sprenger et al., 2018), and tomato (Lee et al., 2018). However, in C. *annuum*, studies on transcriptional changes occurring in response to water shortage are to date rather limited (Zinselmeier et al., 2002; Ji et al., 2010; Sharoni et al., 2012; Lee and Choi, 2013).

Previous studies in plants have highlighted that transcription factors (TFs) play a crucial role as molecular regulators to activate or inhibit the expression of stress-related genes and promote adaptation to stressful environments. Important abiotic stress-responsive TF families are AP2/ERF, DREB, bZIP, AREB/ABF, MYB, bHLH, WRKY and NAC (Nakashima et al., 2012; Singh and Laxmi, 2015; Gamboa-Tuz et al., 2018; Mittal et al., 2018; Mohanta et al., 2020). With respect to NACs, several members of this family have been identified and functionally characterized in crops, among which is the solanaceous species tomato (Yang et al., 2011; Han et al., 2012; Ma et al., 2013; Thirumalaikumar et al., 2018; Devkar et al., 2020). The typical NAC protein is characterized by a highly conserved DNA-binding NAC domain at its N-terminus and a divergent transcription regulatory region (TRR) at its C-terminus (Mao et al., 2007).

Whole-genome sequences of both peppers and closely related species have recently been made available (Kim et al., 2014; Qin et al., 2014; Kang et al., 2016; Ahn et al., 2018; Acquadro et al., 2020), and a comprehensive analysis of the NAC gene family has been performed. A whole of 104 *CaNAC* genes, of which 24 are exclusive to the Solanaceae family, were identified and found to be mostly located on chromosomes 1, 2, 3, and 6 (Diao et al., 2018). However, few NAC genes have been functionally characterized in C. *annuum* with respect to their function in stress responses. Recently, Zhang et al. (2020) isolated and functionally characterized CaNAC035, a positive regulator of abiotic stress tolerance acting through multiple signaling pathways. The involvement in the response to abiotic stress tolerance has also been demonstrated for CaNAC064 (Hou et al., 2020) and CaNAC2 (Guo et al., 2015).

In this study, we employed RNA-sequencing to identify differentially expressed genes (DEGs) in a *C. annuum* inbred line subjected to acute drought stress, and recovery from it, at different developmental stages. Common core sets and plant developmental stage-specific DEGs were identified. Among them, two members of the NAC family (namely *CaNAC072* and *CaNAC104*) were significantly up-regulated transcriptionally following acute drought stress, and down-regulated after recovery across all developmental stages, and were thus selected for further functional characterization. Transient expression in *Nicotiana benthamiana* leaves revealed a nuclear localization of CaNAC072 and CaNAC104; however, a different time of transcriptional activation of the two pepper NAC genes was detected following NaCl, PEG or ABA treatments.

Both NACs were also functionally characterized by Virus-Induced Gene Silencing (VIGS) in pepper plantlets, and CaNAC072 was further assessed in three *Arabidopsis thaliana* genetic backgrounds.

MATERIAL AND METHODS

Drought Stress Treatments, RNA Extraction and Sequencing

Seeds of a yellow-fruited breeding line of *Capsicum annuum* (Cuneo CCu07), selected at the DISAFA (University of Turin), were placed on filter paper soaked with sterilized tap water. Petri dishes were placed in a growth chamber at 350 µmol photons m^{-2} s^{-1}, 50% relative humidity, 26–24°C and a 16–8 h light-dark regime. After germination, seedlings were transferred to individual plastic pots filled with peat and grown for about 15 days. Before transfer to the greenhouse, the plantlets were transplanted in 5 L pots filled with a mixture of peat and perlite at a ratio of 1:1 by volume and grown following standard horticultural practices.

Groups of 15 plants were irrigated up to the soil water retention, and grown up to: 1) the production of five true leaves (stage 1); 2) the production of the third flower (stage 2); 3) the setting of the first fruit (stage 3). Ten plants at each developmental stage were subjected to drought stress by stopping irrigation, while irrigation was maintained on five other plants (Controls). The effect of drought stress was assessed by evaluating stomatal conductance (g_s) through a Leaf Porometer Model SC-1 (Meter Group) on five fully expanded and randomly chosen leaves of each plant. Each measurement was replicated three times in each plant. After drought stress, a group of five plants was recovered following re-watering up to the soil water retention, until they reached the same g_s as control plants.

At each stage of plant development, RNA was extracted using the NucleoSpin RNA kit (Macherey-Nagel) from three leaves of three control plants as well as three plants subjected to drought stress, and three plants subjected to drought stress followed by recovery, chosen randomly. The RNAs were treated with RNase-free DNase I at room temperature to remove contaminant DNA.

Prior to library construction, an Agilent 2,100 Bioanalyzer (Agilent, CA, United States) was used to confirm the quality and quantity of RNA such that the 28S/18 S rRNA ratio was >1.5 and the RNA integrity number >7. For each sample, mRNA was fragmented randomly by adding fragmentation buffer, then the cDNA was synthesized by using mRNA template and random hexamer primers, after which a custom second-strand synthesis buffer (Illumina), dNTPs, RNase H and DNA polymerase I were added to initiate the second-strand synthesis. After terminal-end repair and ligation of sequencing adaptors, the double-stranded cDNA library was completed through size selection and PCR

enrichment. Libraries were sequenced on an Illumina NovaSeq 6,000 platform at Novogene, and 150 bp paired-end reads were generated. Raw reads are publicly available in the NCBI Sequence Read Archive under Bioproject PRJNA668245.

Identification of Differentially Expressed Genes and Enrichment Analysis

Raw reads obtained from the Illumina NovaSeq 6,000 platform were processed for standard quality controls (QC) using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Then they were cleaned with Scythe (v0.994, https://github.com/vsbuffalo/scythe) for removing contaminant residual adapters, and Sickle (v1.33, https://github.com/najoshi/sickle) to remove reads with poor quality ends (Q < 30). The cleaned reads were mapped to the reference CM334 pepper genome (Kim et al., 2017) with Hisat2 (Kim et al., 2015). Data quality analysis was conducted by visualizing results of principal component analysis (PCA). The counts of mapped reads per gene per sample were obtained with Stringtie (Pertea et al., 2015). Then, the DESeq2 R package was applied to identify differentially expressed genes (DEGs) (Love et al., 2014) at a False Discovery Rate (FDR) threshold of 0.05. The genes with an absolute log2 fold change ≥1 were considered as DEGs.

Functional enrichment analyses of Gene Ontology (GO) annotation was performed to identify which DEGs were significantly enriched in GO terms. GO enrichment analysis was conducted using AGRIGOv2, using the available annotation of the previously published sequence of the inbred line CCu07 (Acquadro et al., 2020).

NAC Gene Prediction and Transcription Factor Identification

The Hidden Markov Model (HMM) profile of the NAC domain (PF02365), downloaded from the Protein families (Pfam) database (El-Gebali et al., 2019), was used for identifying NAC genes in the pepper CM334 genome (Kim et al., 2017) using HMMER (V 3.2.1). Positive matches were manually checked to confirm the presence of NAC domains, using InterProScan (Mitchell et al., 2019). Transcription factors in the reference genome CM334 were identified following a BlastP analysis against the Plant Transcription Factor Database (http://planttfdb.gao-lab.org/). To identify Arabidopsis orthologs of two NACs, which were always upregulated after drought stress and downregulated after recovery (i.e., CA.PGAv.1.6.scaffold213.1 and CA.PGAv.1.6.scaffold1156.46), a multiple alignment with the entire Arabidopsis NAC family was carried out by applying Clustal Omega (1.2.4) (Madeira et al., 2019). A further Pairwise Sequence Alignment with Emboss Needle (Madeira et al., 2019), for both pepper NACs with the closest Arabidopsis gene was used to retrieve identity and similarity percentage.

Subcellular Localization

Subcellular localization of CaNAC072 and CaNAC104 was predicted by identifying Nuclear Localization Signal (NLS) sequences using the Localizer software (http://localizer.csiro.au/).

The coding sequences (without stop codon) of CaNAC072 and CaNAC104 were amplified from CCu07 cDNA by using primers CaNAC072-F, CaNAC072-R, CaNAC104-F and CaNAC104-R (**Supplementary Table S1**). PCR was performed with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) and the amplicon sequences were validated (Eurofins MWG Operon). Amplicons were cloned into pDONR207 vector using BP clonase (Invitrogen). The sequence-confirmed entry vector was recombined into pK7FWG2, containing kanamycin as selection marker (Karimi et al., 2002), using LR reaction mix II (Invitrogen). The recombined plasmids (35S:CaNAC072-GFP and 35S:CaNAC104-GFP) were transformed into *Agrobacterium tumefaciens* (strain GV3101) and then used for infiltration of *Nicotiana benthamiana* leaves (Senthil-Kumar and Mysore, 2014). GFP signals were analyzed 2 days after leaf infiltration using a Leica DM6000B/SP5 confocal laser-scanning microscope (Leica Microsystems, Wetzlar, Germany). The laser excitation wavelength was 488 nm, and the detection was set at 507 nm.

Further confirmation of nuclear signal was assessed with nuclear-specific DAPI (4,6-diamidino-2-phenylindole) staining. Transformed leaves were incubated for 30 min in darkness with a solution containing 10 mg L^{-1} DAPI. Afterwards, nuclear-specific signal was assessed by confocal laser-scanning microscopy as described by Sampathkumar and Wightman (2015).

Stress and Hormonal Treatments

CCu07 seeds were germinated on full-strength Murashige-Skoog (MS) medium containing 2% (w/v) sucrose, and grown under a 24–26°C and a 16–8 h light-dark regime. Ten-day-old seedlings were transferred to MS liquid media containing 120 mM NaCl (Sigma-Aldrich), 10% (w/v) PEG 6000 (Sigma-Aldrich), or 100 µM ABA (Sigma-Aldrich), and treated for 2, 4, 6, 10 and 24 h. Stressors and hormones were omitted from control seedlings.

Virus-Induced Gene Silencing

For the virus-induced gene silencing (VIGS) experiment, CCu07 seeds were sown directly into soil containing a mixture of peat and vermiculite (2:1, v/v), and plants grown in a phytotron at 450 µmol photons $m^{-2}s^{-1}$, 50% relative humidity, 21°C day/night under a 14 h day/10 h night regime.

VIGS was performed with vectors pTRV1 (pYL192) (Liu et al., 2002) and pTRV2 (pYL170) following the protocol described by Senthil-Kumar and Mysore, 2014. Briefly, VIGS constructs were designed with the SGN VIGS Tool (Fernandez-Pozo et al., 2015). The selected gene fragments (**Supplementary Table S2**) were PCR-amplified using primers VIGS_CaNAC072-F, VIGS_CaNAC072-R, and VIGS_CaNAC104-F, VIGS_CaNAC104-R (**Supplementary Table S1**) from CCu07 cDNA and cloned into pDONR207 using BP clonase (Invitrogen). The sequence-confirmed entry vector was recombined into pK7FWG2 using LR reaction mix II (Life Technologies). The plasmids pTRV1, pTRV2:CaNAC072, pTRV2:CaNAC104 as well as pTRV2:AtPDS (as a positive control for efficient VIGS) and empty vector pTRV2:00 were used to transform A. tumefaciens (strain GV3101), and the latter used for infecting cotyledons of 10 days-old pepper seedlings. After 12 days of infection, drought stress was performed for 7 and 9 days. Recovery

treatments were performed by re-watering plants for 7 days. For assessing plant survival rate, plants were subjected to 9 days severe drought stress followed by recovery through re-watering. The percentage of surviving plants was determined after 7 days of recovery. The drought and recovery phenotypic analyses were assessed in 15 biological replicates per treatment.

Identification of Pepper Homologous Genes in Arabidopsis and Plant Transformation

Pepper orthologs of Arabidopsis drought marker genes [i.e., *ANAC019* (AT1G52890), *ANAC055* (AT3G15500), *DREB2A* (AT5G05410), *DELLA/GAI* (AT1G14920), and *SnRK2* (AT3G50500)], were identified using PLAZA 4.5 (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_5_dicots/) and annotated using the Sol Genomics Database (https://solgenomics.net/).

The 35S:CaNAC072-GFP and 35S:CaNAC104-GFP constructs were inserted into *A. tumefaciens* (strain GV3101) and transformed into *A. thaliana* genotypes Columbia-0 (Col-0), *anac072* mutant (*rd26-2*, SALK_083,756) and *anac019/anac055/anac072* triple mutant backgrounds using the floral dip method (Clough and Bent, 1998). NAC T-DNA insertion lines (*anac019*: SALK_096,295, *anac055*: SALK_014331, and *anac072*: SALK_083756) were obtained from NASC (http://arabidopsis.info). Homozygous knockout plants were identified by PCR-based genotyping, qRT-PCR, and end-point PCR. The triple NAC mutant was generated by crossing. Plants were grown at 22°C under long-day (LD) condition, i.e., 16 h light/8 h dark. Seeds of T2 plants were harvested.

From T1 transgenic plants showing a 3:1 segregation ratio when grown on MS plates supplemented with the same antibiotic. Ten-day-old seedlings were transferred to soil [potting and quartz sand; 2:1 (v/v)]. After 14 days, their drought tolerance was assessed by withholding water for 20 days.

Relative Water Content and Water Loss Assay

Relative water content (RWC) was measured as described by Thirumalaikumar et al., 2018. Briefly, second true leaves were harvested and their fresh weights (FW) were determined. Leaves were then immersed in distilled water and incubated overnight at 4°C to obtain the saturated weight (SW). Next, leaves were dried at 70°C for 48 h to measure the dry weight (DW). RWC was calculated as % by the formula (FW-DW)/(SW-DW) x 100. RWC was measured in 15 biological replicates per genotype per treatment.

For water loss assays, second true leaves were detached and immediately weighted (FW). Leaves were then placed on filter paper in a growth chamber at 22°C and the weight measured after 1, 2, 3, 4, 5 and 6 h. Water loss was measured in 15 biological replicates each.

Expression Profiling

Total RNA extraction, cDNA synthesis, and qRT-PCR were performed as described by Balazadeh et al., 2008. The qRT-PCR primers (**Supplementary Table S3**) were designed using QuantPrime (Arvidsson et al., 2008). ABI-PRISM 7900 HT sequence detection system (Applied Biosystems, Darmstadt, Germany) and SYBR Green (Applied Biosystems, Darmstadt, Germany) were used for qRT-PCR and detection of amplified product.

The transcription pattern of *CaNAC072* and *CaNAC104* was assessed in ten-day-old wild-type pepper seedlings subjected to salinity induced by 120 mM NaCl, drought induced by 10% (w/v) PEG 6000, as well as after treatments with 100 μM ABA over 24 h. Three biological replicates, each a pool of five seedlings, were used for gene expression profiling. The homologue of the Arabidopsis *RD29A* gene, known to be induced by various abiotic stressors (Msanne et al., 2011), i.e., *CaRD29A* (CA.PGAv.1.6.scaffold490.32), was used as a positive control. Data were quantified using the $2^{-\Delta\Delta Ct}$ method based on Ct values of *CaNAC072* and *CaNAC104*, and pepper *ACTIN*, (CA.PGAv.1.6.scaffold786.11) and *GADPH* (TC.CA12g18170) as housekeeping genes.

In the VIGS experiment, the expression pattern of *CaNAC072* (CA.PGAv.1.6.scaffold213.1), *CaNAC104* (CA.PGAv.1.6.scaffold1156.46), *CaNAC019* (CA.PGAv.1.6.scaffold643.12), *CaNAC055* (TC.CA07g18020), *CaDREB2A* (TC.CA05g15500), *CaDELLA* (TC.CA12g02780), and *CaSnRK2* (TC.CA05g02700) was assessed. RNA was extracted from pepper plants infiltrated with TRV2:CaNAC072, TRV2:CaNAC104 and TRV2:00 and subjected to 7 days of water deprivation. Three biological replicates were used for gene expression analyses.

RESULTS

Effects of Drought Stress and Recovery on Stomatal Conductance

We studied the effects of drought stress and recovery (re-watering) on a *C. annuum* inbred line at different plant developmental stages. To this end, plants at the status of five true leaves (stage 1), the third flower (stage 2), and at the setting of the first fruit (stage 3) were exposed to drought stress, by withholding water, and then allowed to recover by re-watering. In the control condition, plants were regularly irrigated. Drought stress was imposed by interruption of irrigation. For recovery experiments, drought-stressed plants were re-watered up to soil water retention (**Figure 1**). Leaves were sampled after each treatment and used for RNA-seq-based transcriptome profiling.

The effect of water deprivation was assessed through a repeated evaluation of stomatal conductance (g_s). Control plants showed values of 400 ± 60 mmol $m^{-2}s^{-1}$ on average, while drought-stressed plants displayed a progressive decrease of g_s until about 50 mmol $m^{-2}s^{-1}$ and an evident foliar withering. Afterwards, plants were re-watered, and after 24 h reached a g_s value analogous to control plants, and they recovered their leaf turgor.

RNA-Seq Analyses

RNA-sequencing profiles were obtained from pepper leaves of control, drought-stressed, and drought stressed and then

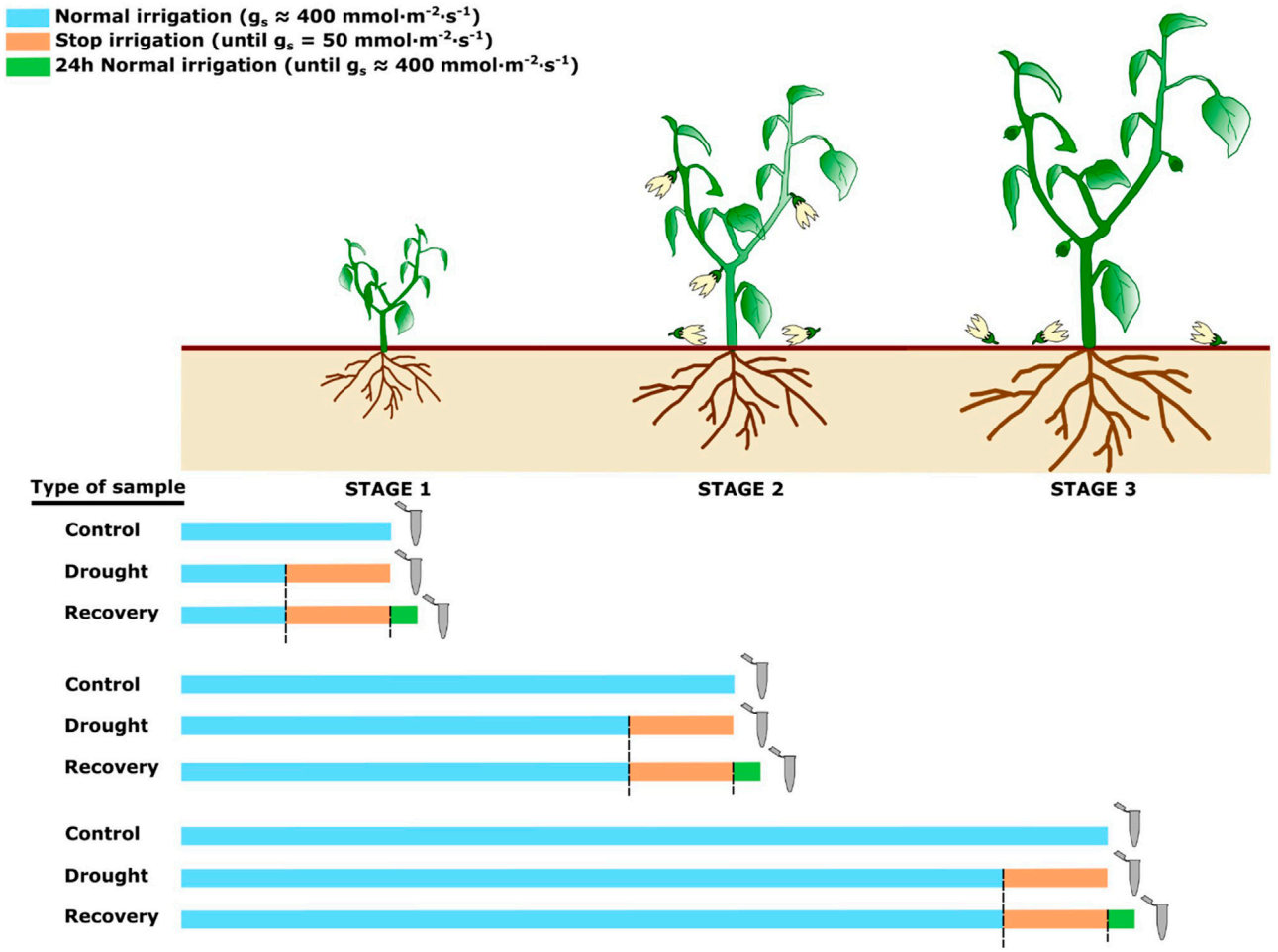

FIGURE 1 | Experimental design of the treatments for control, drought and recovery plants. Grey micro test tubes indicate when the samples were collected. The colour of the bars represents the irrigation status of the plants (light blue = normal irrigation; orange = no irrigation; green = irrigation, 24 h recovery).

recovered plants at the three developmental stages, as described above. About 33 million paired reads were obtained from each sample. After removing low-quality sequences, adapters and possible contaminations, around 30 million reads (PE and singlets) per sample were obtained (**Supplementary Table S4**). Clean reads were aligned against the CM334 pepper genome (Kim et al., 2017), resulting in approximately 92% mapping reads on average.

Following quality control by PCA analysis, the three replicates for each condition (control, drought and recovery) clustered together. Only one drought stress replicate at fruit stage was found to be separated from the other two replicates. In addition, a clear separation was found between drought stress replicates and control/recovery samples (**Supplementary Figure S1**).

Using a false discovery rate threshold of 0.05 and an absolute $\log_2$ fold change ≥1 (adjusted $p \leq 0.01$) as the significance cutoff, the expression of 3,680, 3,976 and 1,282 DEGs was found to be significantly regulated by comparing control to drought-stressed plants at stage 1, 2 and 3, respectively. By comparing drought-stressed and recovered plants a whole of 4,155, 2010, 1,328 DEGs were found to be significantly regulated at stage 1, 2 and 3, respectively. At last, when recovered plants were compared with control plants 1,285, 271 and 56 genes where differentially expressed at stages 1, 2 and 3, respectively (**Supplementary Table S5**).

Annotation and GO Enrichment of Core Sets of Commonly up- or Downregulated Genes

All DEG sets were compared in order to identify groups of commonly regulated genes. At all stages of plant development, a whole of 197 and 123 genes were found to be commonly up- and downregulated after drought, respectively **(Figures 2A,B)**. On the other hand, 90 and 187 genes were found to be commonly up- and downregulated following plant recovery, respectively (**Figures 2C,D**).

In order to compare the function enrichment of commonly up- or downregulated genes at all stages of plant development, a gene ontology (GO) enrichment analysis was performed. The common over-represented GO terms at all stages are summarized in **Figure 2E**. Nineteen significant GO terms were identified among all developmental stages. When comparing drought vs control treatments, three enriched GO terms were identified among

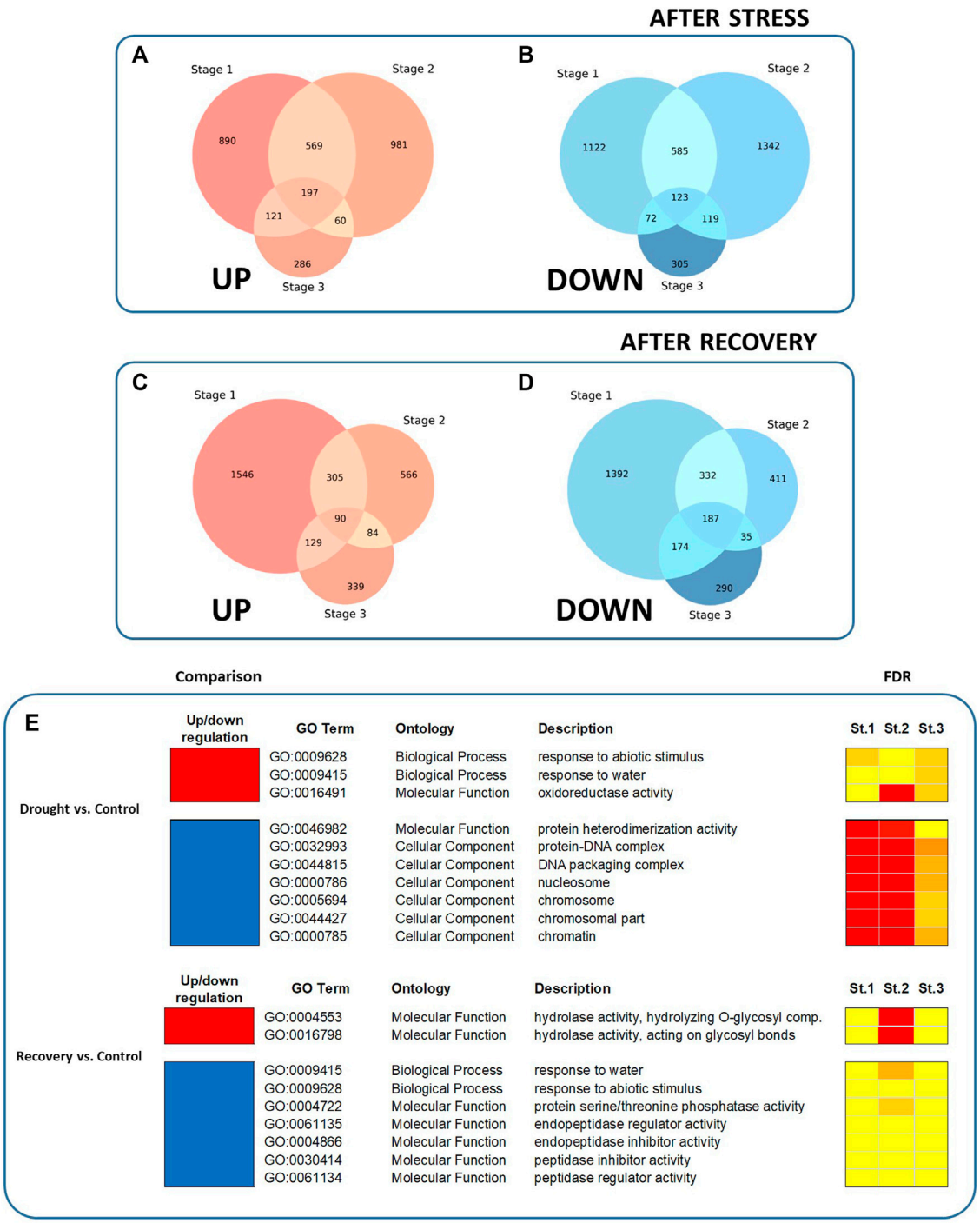

FIGURE 2 | Venn diagram showing intersections of up- or downregulated DEGs, and enriched GO terms, among the DEGs in response to drought and after recovery. **(A)** Genes upregulated after water stress, compared to control plants. **(B)** Genes downregulated after water stress. **(C)** Genes upregulated after recovery. **(D)** Genes downregulated after recovery. **(E)** Enriched GO terms are shown for the three stages of plant development (i.e., St.1 = five true leaves, St.2 = production of third flower, and St.3 = setting of first fruit). Red color represents upregulated GO terms whereas blue color represents downregulated GO terms. Colors on the FDR columns represent different significance levels (yellow: FDR <0.05; orange: FDR <0.01; red: FDR <0.001.

upregulated DEGs (i.e., GO:0009628, "response to abiotic stimulus"; GO:0009415, "response to water", and GO:0016491, "oxidoreductase activity"), while seven GO terms were identified among the downregulated DEGs (including e.g. GO:0046982, "protein heterodimerization activity"; GO:0032993, "protein-DNA complex"). On the other hand, when comparing recovery vs

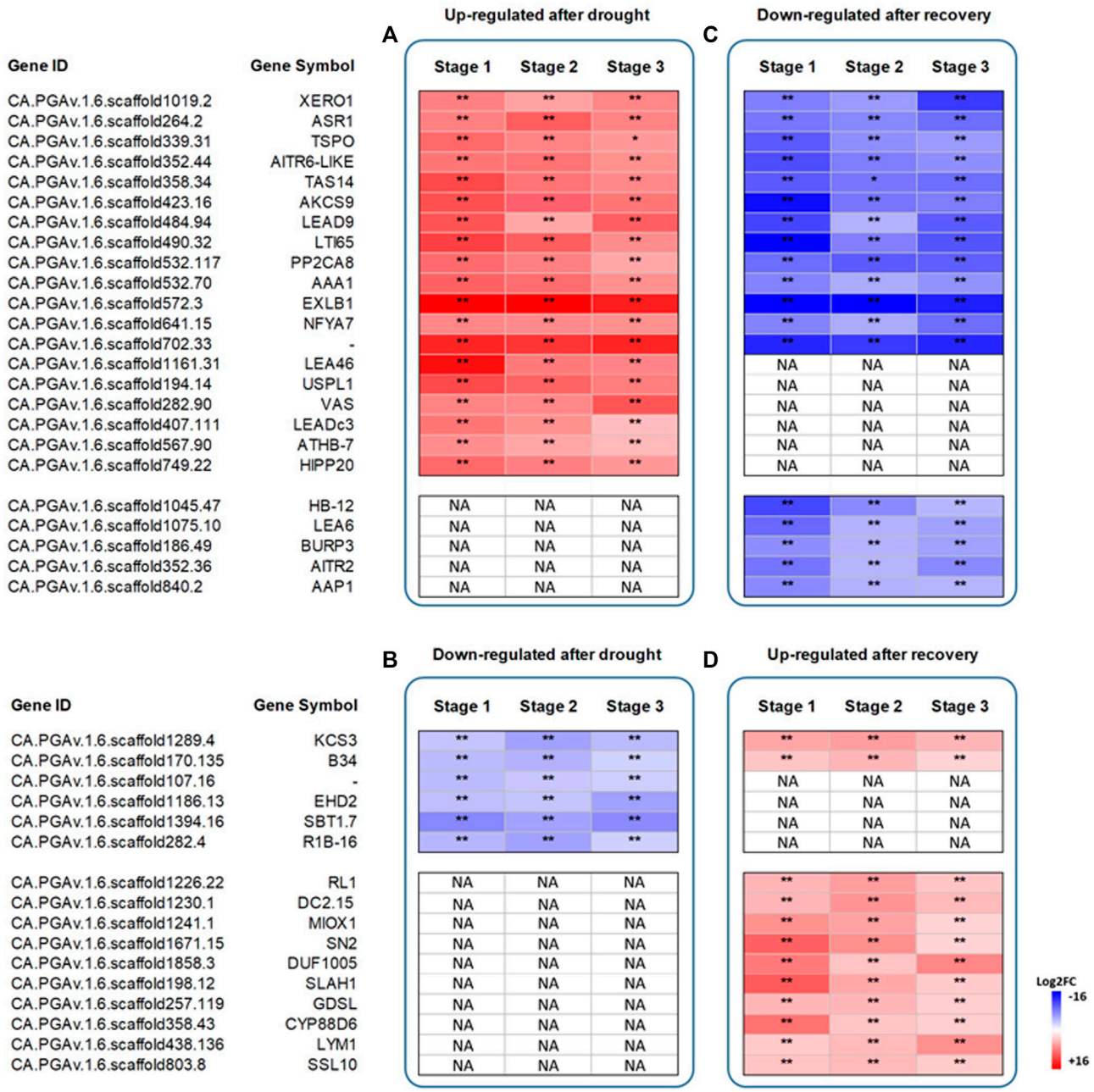

FIGURE 3 | Heatmap showing the log2 fold change differences in the expression profiles of the top 30 common DEGs at the three stages of plant development. Genes upregulated after drought stress or recovery (Panel **A** and **D**) and genes downregulated after drought stress or recovery (Panel **B** and **C**). Data represent the means of three biological replicates. Asterisks represent statistically significant differences (*p* values) between control plants and stressed (droughted) plants, and between stressed (droughted) plants and recovery plants. The *p* values were adjusted using the Benjamini and Hochberg method. Corrected *p* value of 0.005 and |log2 (Fold Change)| of 1 were set as the threshold for significantly different expression. White fields indicate the absence of significant differences between treatments. NA stands for "Not applicable".

drought treatments, two enriched GO terms were identified on upregulated DEGs, i.e., GO:0004553, "hydrolase activity, hydrolyzing O-glycosyl compounds", and GO:0016798, "hydrolase activity, acting on glycosyl bonds"), while seven were identified among the downregulated DEGs (including e.g., GO:0009415, "response to water"; GO:0009628, "response to abiotic stimulus"). No common enriched GO terms were found by comparing recovery vs. control. All GO terms, including stage-specific enriched GO terms, can be inspected in the **Supplementary Tables S6–S17.**

Among the 197 commonly up-regulated genes after drought stress, we focused our attention on the 30 most expressed genes at all stages of plant development, and 19 were found to be shared (**Figure 3A**).

Among them are two Late Embryogenesis Abundant protein genes, *LEA46* and *LEA-D29* (CA.PGAv.1.6.scaffold1161.31 and CA.PGAv.1.6.scaffold407.111), several stress-induced genes, such as ABA metabolism-related *ASR1*, *TAS14* and *ATHB-7* (CA.PGAv.1.6.scaffold264.2, CA.PGAv.1.6.scaffold358.34, and CA.PGAv.1.6.scaffold567.90) as well as some plant cell wall-related genes, i.e. *EXLB1, AAA1 KATANIN* (CA.PGAv.1.6.scaffold572.3 and CA.PGAv.1.6.scaffold532.70). Likewise, among the 123 commonly downregulated genes after drought stress, we focused our attention on the 30 most downregulated genes. Six of them were found to be shared and included one gene involved in fatty acid biosynthesis, i.e., *3-KETOACYL-COA SYNTHASE 3* (*KCS3*; CA.PGAv.1.6.scaffold1289.4), and a gene encoding a serine protease (i.e., SBT1.7; CA.PGAv.1.6.scaffold1394.16) (**Figure 3B**).

After recovery, 12 upregulated genes were found to be in common among the top 30 most upregulated ones. Among them, *MYO-INOSITOL OXYGENASE 1* (*MIOX1*; CA.PGAv.1.6.scaffold1241.1) is involved in providing nucleotide sugars for cell wall polymers, and *3-KETOACYL-COA SYNTHASE 3* (*KCS3*). Lastly, 18 genes were downregulated after recovery, 13 of which were observed to be upregulated after stress. Among them are BURP DOMAIN CONTAINING PROTEIN 3 and homeobox-leucine zipper protein gene ATHB-7 (CA.PGAv.1.6.scaffold186.49 and CA.PGAv.1.6.scaffold1045.47) (**Figures 3C,D**).

Differentially Expressed TFs After Drought Stress and Recovery

After drought, the RNA-seq expression profiling revealed 295, 370 and 108 differentially expressed (DE) TFs at stage 1, 2 and 3 of plant development, respectively. Among the ones at stage 1, 162 were up- and 133 downregulated, while at stage 2, 188 were up- and 183 downregulated, and at stage 3, 80 were up- and 28 downregulated (**Supplementary Table S18**). On the other hand, after the recovery phase a whole of 325, 176 and 106 DE TFs were identified at stage 1, 2 and 3 of plant development, respectively. Of those, at stage 1, 138 were up- and 187 downregulated; at stage 2, 84 were up- and 92 downregulated, while at stage 3, 33 were up- and 73 downregulated (**Supplementary Table S18**).

Overall, after drought and recovery significant changes occurred in the transcript levels of the five following TF families: MYB, bHLH, ERF, NAC and HSF. Stage-specific TF families were also identified (**Supplementary Figure S2**). A core set of 31 TF genes were found to be always up regulated after drought and downregulated after recovery at all stages of plant development, among which where MYB, bHLH, CO-like and NF-YA, as well as two NAC TFs (**Supplementary Table S19**).

Genome-wide Identification of NAC TFs and Their Expression After Drought and Recovery

We performed a comprehensive analysis to identify NAC genes in the pepper CM334 genome sequence (Kim et al., 2017) on the basis of the NAC domain (PF02365). A whole of 113 NAC TFs were identified (**Supplementary Table S20**), of which 13, 20 and 9 NACs were differentially expressed after drought stress at stage 1, 2 and 3, respectively (**Figures 4A,B**). On the other hand, 13, 10, and 8 NAC TFs were found differentially expressed after recovery (**Figures 4C,D**).

On the basis of our RNA-seq experiment, two members of the NAC TF family, i.e. CA.PGAv.1.6.scaffold213.1 (referred to as *CaNAC072* in the following) and CA.PGAv.1.6.scaffold1156.46 (*CaNAC104*) were found to be always upregulated after drought stress and downregulated after recovery, and thus were selected for functional characterization.

Subcellular Localization of *CaNAC072* and *CaNAC104*

In-silico prediction analyses of the encoding sequences identified one NLS in CaNAC072 (RKNGSSKLDEWVLCRIYKK) and one in CaNAC104 (KKRK), suggesting a nuclear localization of both proteins. This was confirmed by fusing in-frame the coding sequences of CaNAC072 and CaNAC104, without stop codon, to the green fluorescent protein (GFP) coding sequence under the control of the CaMV 35S promoter. The subcellular localization of each protein was analyzed by confocal laser scanning microscopy after transient expression in *N. benthamiana* leaves. The analysis showed that the GFP signal was localized in the nucleus of leaves transfected with both CaNAC072- and CaNAC104-GFP fusion constructs (**Figure 5**). This observation is consistent with the role of CaNAC072 and CaNAC104 being transcription factors.

Effects of Abiotic Stresses on *CaNAC072* and *CaNAC104* Expression

NAC TFs play key roles in plant stress responses. CaNAC072 is the closest homologue of Arabidopsis ANAC072, with an identity of 56.3% and similarity of 67.6% (E value: 5e-132). Lower similarity percentages were detected for the ANAC072 homologues: ANAC019 (identity 53.0%, similarity 64.4%), and ANAC055 (identity 52.8%, similarity 63.1%).

On the other hand, CaNAC104 was found the closest homolog of Arabidopsis ANAC104, with an identity of 40.5% and similarity of 54.8% (E value: 1e-43), the latter being induced by several abiotic stresses and playing a major role in root xylem vessels (Kilian et al., 2007; Tang et al., 2018). Lower similarity percentages were also detected for the ANAC0104 homologues: ANAC02 (identity 5.7%, similarity 10.1%), and ANAC05 (identity 3.8%, similarity 6.6%).

To validate the response of *CaNAC072* and *CaNAC104* to abiotic stresses, we assessed the effects of salinity induced by NaCl, and drought induced by PEG, over 24 h in CCu07 pepper seedlings. Since both salinity and drought stimulate the accumulation of ABA (Marusig and Tombesi, 2020), we also validated gene expression induction upon ABA treatment.

Our treatments were effective in inducing the expression of the stress-related gene *CaRD29A*. The NaCl treatment induced its highest level of expression after 2 h, which progressively decreased in the following hours. Differently, the PEG

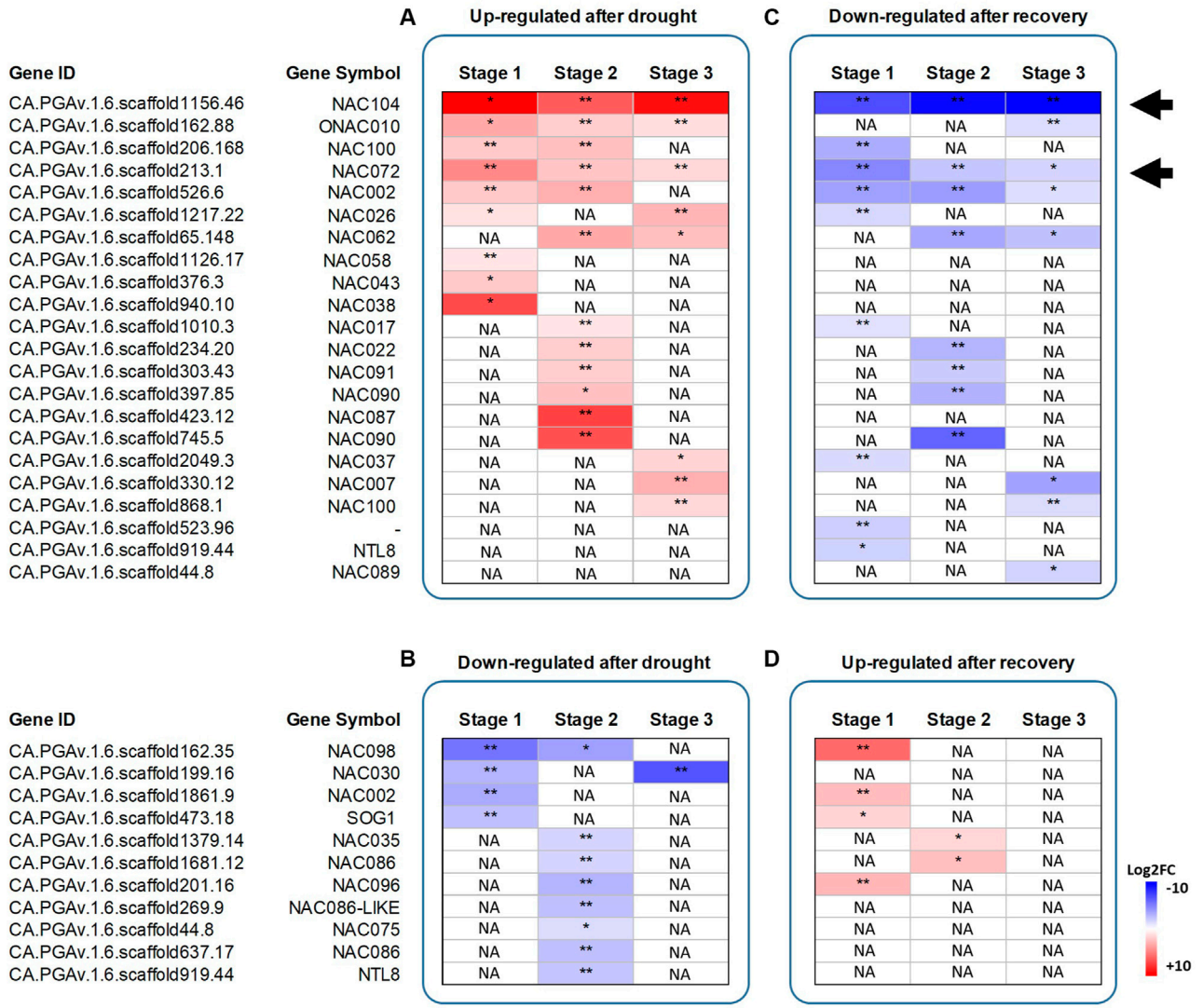

FIGURE 4 | Heatmap showing the log2 fold change differences in the expression of the differentially expressed NAC TFs identified. Genes upregulated after drought stress or recovery (Panel **A** and **D**) and genes downregulated after drought stress or recovery (Panel **B** and **C**). Black arrows highlight the two NACs selected for functional characterization, being always upregulated after drought, and downregulated after recovery. Data represent the means of three biological replicates. Asterisks represent statistically significant differences (*p* values < 0.05*, Pvalues < 0.01**) between control and stressed plants, and between drought-stressed and recovered plants. The *p* values were adjusted using the Benjamini and Hochberg method. Corrected *p* value of 0.005 and ||log2 (Fold Change)| of 1 were set as the threshold for significantly different expression. White fields indicate the absence of significant differences between treatments. NA stands for "Not applicable".

treatment induced an increase of *CaRD29A* expression after 4 h and a further increase after 6 h, which was followed by a progressive decrease. At last, treatment with ABA caused an increase of *CaRD29A* expression after 2 h onwards following the hormonal treatment (**Supplementary Figure S3**).

As shown in **Figure 6**, *CaNAC072* transcript levels significantly increased after 2 h of NaCl treatment and gradually decreased in the following hours. The PEG treatment induced the expression of *CaNAC072* only after 4 and 6 h of treatment. Treatment with ABA resulted in a rapid rise (already after 2 h) in the expression of *CaNAC072*. Interestingly, this upregulation was maintained for up to 24 h. In contrast, the NaCl treatment induced the upregulation of *CaNAC104* after 4 h and its transcript level progressively decreased in the following hours, while the PEG treatment did not induce its expression. On the other hand, ABA treatment induced a progressive increase of *CaNAC104* expression from 4 h onwards (**Figure 6**). These results confirm that *CaNAC072* and *CaNAC104* are differentially induced by salinity and drought and suggest that they are ABA-dependent, with *CaNAC072* being an early-responsive gene and *CaNAC104* acting as a late-responsive gene.

Effect of Drought Stress on *CaNAC072*- and *CaNAC104*-Silenced Pepper Plants

To investigate the role of *CaNAC072* and *CaNAC104* for the response to drought, we reduced their expression through Virus-

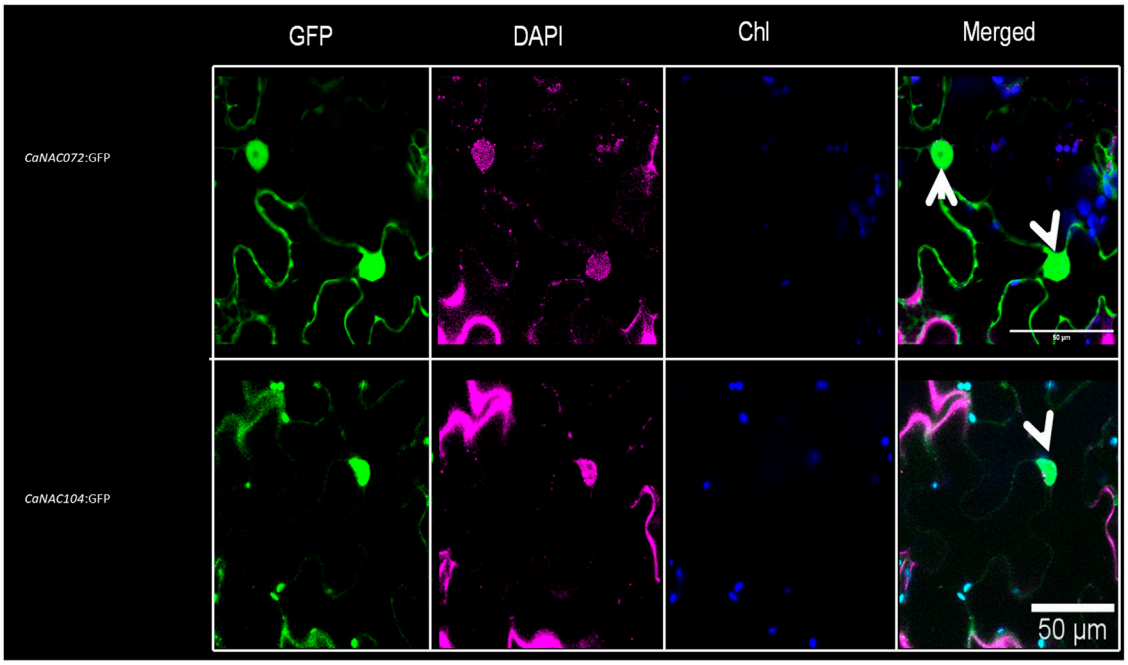

FIGURE 5 | Nuclear localization of CaNAC072 and CaNAC104. Confocal microscopy analysis showing the nuclear localization of CaNAC072:GFP and CaNAC104:GFP upon transient expression in *Nicotiana benthamiana* leaves. Nuclear-specific staining using DAPI confirmed nuclear localization of both TFs. "Merged" indicates the combination of GFP and chlorophyll (Chl) autofluorescence signals. White arrows indicate the nuclear-specific GFP signal. Scale bar, 50 µm.

Induced Gene Silencing (VIGS) in Cuneo CCu07 pepper plantlets. The efficiency of VIGS was confirmed by generating TRV2:AtPDS plants, in which the silencing of the PHYTOENE DESATURASE (PDS) is easily displayed by photo-bleaching of leaves. As expected, plants infected with the empty vector TRV2:00 did not show leaf bleaching (**Figure 7A**).

CaNAC072- and *CaNAC104*-silenced plantlets (TRV2:CaNAC072 and TRV2:CaNAC104, respectively), as well as the ones infected with the empty vector (TRV2:00), were subjected to drought stress (water withholding for 7 days), followed by recovery (rewatering for 7 days). No phenotype alterations were observed between non-infiltrated plants and TRV2:00-infected plants. Similarly, no phenotype alterations were detected between *CaNAC072*- and *CaNAC104*-silenced plants compared to TRV2:00 plants under well-watered conditions. After drought stress, a significant increase in the transcript levels of both *CaNAC072* and *CaNAC104* was detected in plantlets carrying the empty vector TRV2:00, but not in plantlets in which the two NAC TFs were silenced, further confirming the effectiveness of the VIGS (**Figure 7B**).

After 7 days of drought (water withholding), the TRV2:00 and TRV2:CaNAC104 plantlets showed severe leaf wilting and a marked decrease in their relative leaf water content (RWC), while TRV2:CaNAC072 plants did not, and their RWC was not significantly different from that of well-watered control plants (**Figures 7C,D**). In order to assess the rate of water loss in silenced plants, detached leaves of TRV2:00, TRV2:CaNAC072 and TRV2:CaNAC104 plants were analysed over 6 h and, as expected, *CaNAC072*-silenced plants showed a better water retention than TRV2:00 and *CaNAC104*-silenced plants (**Figure 7E**).

We also assessed the survival rate of TRV2:CaNAC072, TRV2:CaNAC104 and TRV2:00 plants after 9 days of water withholding, followed by 7 days of recovery (rewatering) (**Supplementary Figure S4**). Consistent with their higher RWC and not-withering phenotype, 15.2% of the *CaNAC072*-silenced plants survived, while none of the TRV2:00 and CaNAC104-silenced plants did. This result suggests that CaNAC072 acts as a negative regulator of the drought stress response.

In pepper infiltrated with TRV2:CaNAC072, TRV2:CaNAC104 and TRV2:00, we assessed the transcript profile of five pepper genes orthologous to drought stress-responsive genes in Arabidopsis, i.e. *CaNAC019*, *CaNAC055*, *CaDREB2A*, *CaDELLA* and *CaSnRK2* (Sakuma et al., 2002; Ye et al., 2017). The qRT-PCR analyses revealed that, after 7 days of water withholding, the expression of all of them significantly increased in TRV2:CaNAC072 leaves, while only a bland activation of *CaDELLA* was observed in TRV2:CaNAC104 (**Figure 7E**). This might indicate a more rapid perception and response to drought stress in *CaNAC072*-silenced plants, which ultimately might explain their improved drought tolerance.

CaNAC072 Does Not Recover the Function of Its Ortholog in Arabidopsis

ANAC072 (also known as *RD26*) and its homologs *ANAC019* and *ANAC055* are drought-responsive genes promoting drought stress tolerance in Arabidopsis (Tran et al., 2004). Previous studies showed that the overexpression of either *ANAC019*, *ANAC055* or *ANAC072*

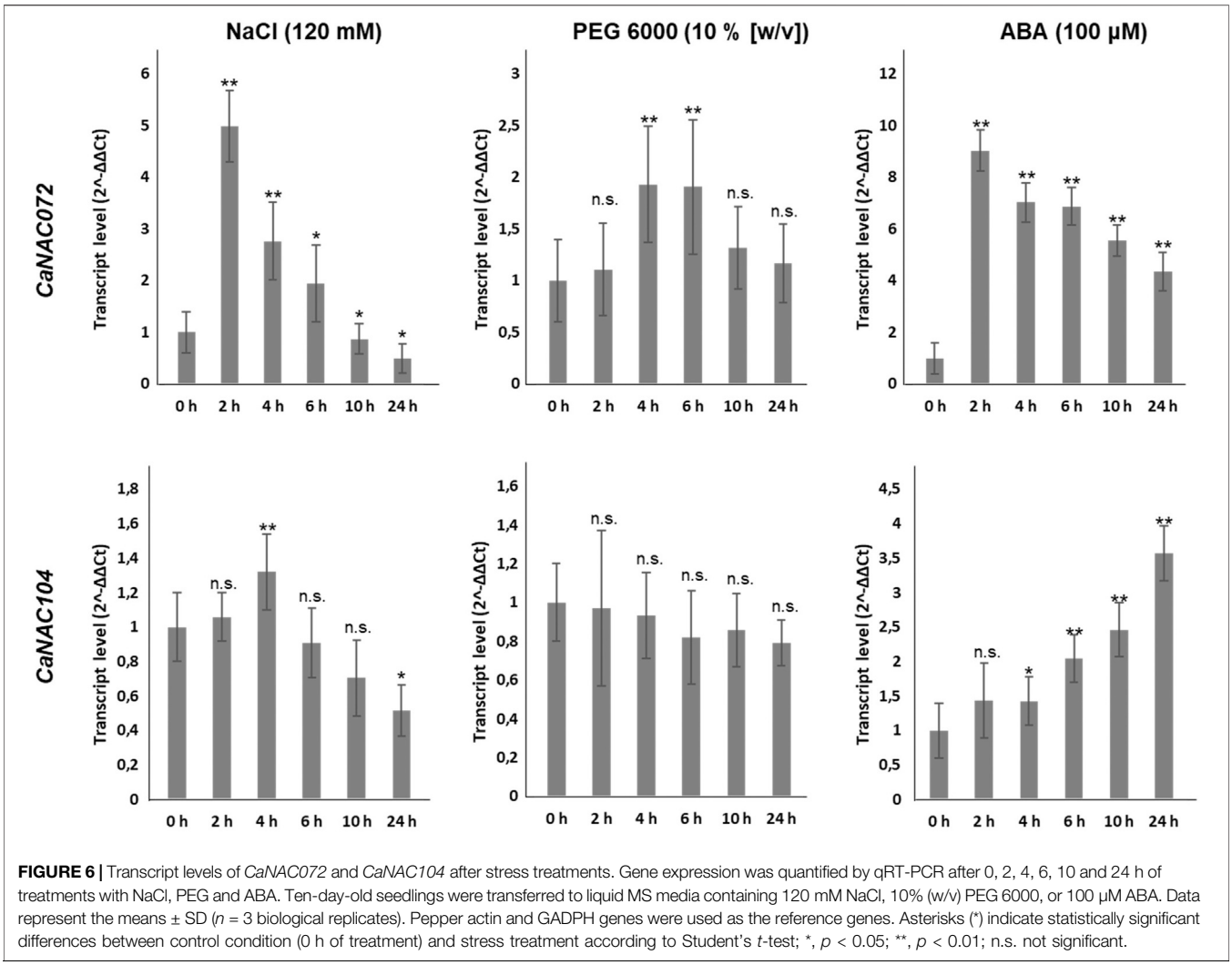

FIGURE 6 | Transcript levels of *CaNAC072* and *CaNAC104* after stress treatments. Gene expression was quantified by qRT-PCR after 0, 2, 4, 6, 10 and 24 h of treatments with NaCl, PEG and ABA. Ten-day-old seedlings were transferred to liquid MS media containing 120 mM NaCl, 10% (w/v) PEG 6000, or 100 µM ABA. Data represent the means ± SD (n = 3 biological replicates). Pepper actin and GADPH genes were used as the reference genes. Asterisks (*) indicate statistically significant differences between control condition (0 h of treatment) and stress treatment according to Student's t-test; *, $p < 0.05$; **, $p < 0.01$; n.s. not significant.

confers drought tolerance, while the *anac019/anac055/anac072* triple mutant shows enhanced drought sensitivity (Tran et al., 2004; Ye et al., 2017; Sukiran et al., 2019). In this study, we observed a drought tolerance increase in *CaNAC072*-silenced pepper plants, suggesting a different function in comparison to its homolog *ANAC072* in Arabidopsis. Furthermore, we generated Arabidopsis plants overexpressing *CaNAC072* fused to green fluorescent protein (GFP) (hereafter, CaNAC072:GFP-OX plants) in Col-0 as well as in *anac072* single and *anac019/anac055/anac072* triple mutants. Neither differences in the growth rate nor in the response to severe drought stress were detected between CaNAC072:GFP/Col-0, CaNAC072:GFP/rd26-2, and CaNAC072:GFP/triple mutant plants compared to Col-0, *rd26-2* and the triple mutant (**Supplementary Figure S5**). Our results thus suggest that CaNAC072 from bell pepper plays a different function than its homolog ANAC072 in Arabidopsis.

DISCUSSION

The response to drought in plants is a complex and coordinated process involving many biological mechanisms. Here, we subjected *Capsicum annuum* plants of a highly homozygous breeding line to acute water stress until leaf withering, followed by rewatering (recovery phase), at three developmental stages. Following sequencing of RNA extracted from leaves of drought-stressed, recovered and control plants, thousands of DEGs were identified, among which were core sets of commonly and plant development-specific up- and downregulated genes.

Abscisic acid (ABA) is known to act as an endogenous messenger in the regulation of the plant's water status, and its action can induce stomatal closure, but also systemically signal for adjustment towards severe water shortage (Tuteja, 2007; Todaka et al., 2019). We detected ABA metabolism-related genes downregulated (e.g., *ATHB-7*) or activated (e.g., *ASR1* and *TAS14*) upon drought stress (Godoy et al., 1994; Söderman et al., 1996; Dominguez and Carrari, 2015). Indeed, *TAS14* was found to increase drought and salinity tolerance when overexpressed in transgenic tomato plants, and it was suggested as a promising candidate for enhancing drought tolerance in pepper as well (Muñoz-Mayor et al., 2012).

In plants exposed to drought stress, hydrophilic proteins such as Late Embryogenesis Abundant (LEA) proteins accumulate and

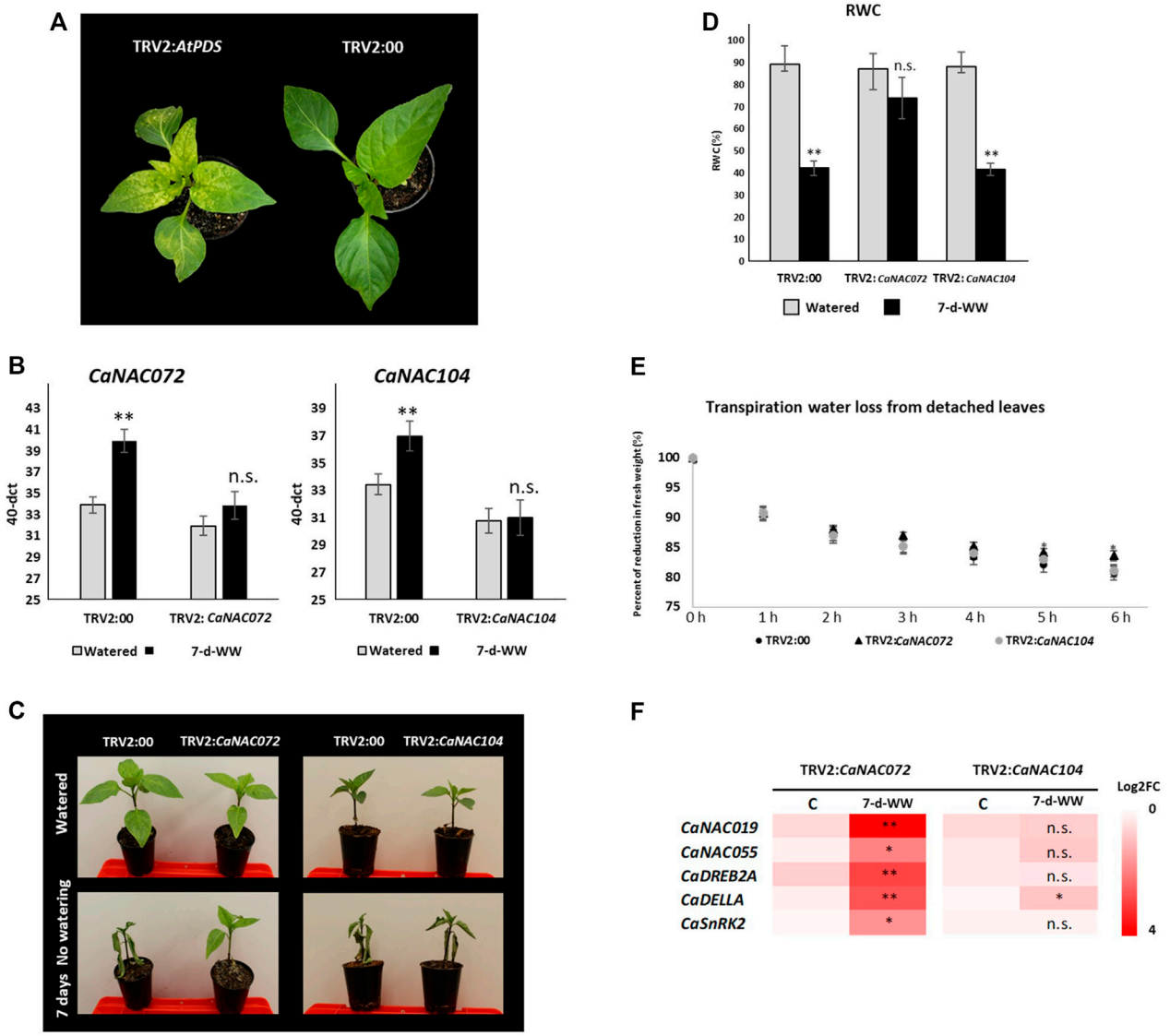

FIGURE 7 | Effect of drought stress on *CaNAC072*- and *CaNAC104*-silenced pepper plants. **(A)** Phenotype of a 25-day-old pepper plant after 12 days of agro-infiltration with TRV2:AtPDS **(left)** and with the empty vector TRV2:00 **(right)**. **(B).** Expression of *CaNAC072* and *CaNAC104* in well-watered control plants (grey bar) and drought-stressed plants (black bar) after 7 days of water withholding (7-days-WW), determined by qRT-PCR. Data represent the means ± SD ($n = 3$ biological replicates). Values are expressed as the difference between an arbitrary value of 40 and dCt, so that high 40-dCt values indicate high gene expression levels. Pepper actin and GADPH genes were used as the reference genes. **(C)** Phenotype of TRV2:00 vis-à-vis with TRV2:CaNAC072 and TRV2:CaNAC104 plants in watering conditions **(upper panel)**, and after 7 days of water withholding **(lower panel)**. **(D)** Relative Water Content (RWC) of TRV2:00, TRV2:CaNAC072 and TRV2:CaNAC104 plants watered (grey bars) and after 7 days of no watering (black bars). Data represent the means ± SD ($n = 15$ biological replicates). **(E)** Trend of leaf fresh weights over 6 h (index of water loss) in detached leaves of TRV2:00, TRV2:CaNAC072 and TRV2:CaNAC104 plants. Data represent means ± SD of 15 biological replicates. **(F)** Heatmap showing the log2 fold change (FC) differences in the expression of five pepper genes (i.e., *CaNAC019*, *CaNAC055*, *CaDREB2A*, *CaDELLA* and *CaSnRK2*) which are orthologues of Arabidopsis drought-responsive genes, in TRV2:CaNAC072, TRV2:CaNAC104, and TRV2:00 plants watered and after 7 days of water withholding. Transcript levels were evaluated by qRT-PCR. Data represent the means of three biological replicates per assay. Asterisks indicate statistically significant differences between *CaNAC072*-non silenced and *CaNAC072*-silenced plants, and between *CaNAC104*-non silenced and *CaNAC104*-silenced plants according to Student's *t*-test; $^{*}p < 0.05$, $^{**}p < 0.01$.

cell walls reorganize in order to preserve cell turgor (Shinozaki and Yamaguchi-Shinozaki, 2007; Gong et al., 2010). We detected upregulation of two LEA protein genes: LEA46 and LEA-D29, whose overexpression in Arabidopsis confers tolerance to severe drought (Olvera-Carrillo et al., 2010), as well as of some cell wall-related genes, such as *EXLB1* and *AAA1 KATANIN*, which were upregulated following drought stress and downregulated after recovery. Furthermore, genes such as *MIOX1*, involved in providing nucleotide sugars for cell wall polymers, and *3-KETOACYL-COA SYNTHASE 3 (KCS3)*, which is required for cuticular wax and root suberin biosynthesis, showed upregulation just after recovery. All in all, our transcriptomic data confirm that

cell wall reorganization is a complex cellular process actively involved in the response of C. annuum to water stress and deserves to be better explored for enhancing drought tolerance in the future.

As expected, we identified significant changes in the expression of members of TF families, which are known to be related to stress responses. Among them are bHLH genes involved in the ABA response, HSF genes acting as regulators that prevent the accumulation of damaged proteins, as well as MYBs, ERFs and NACs (Trujillo et al., 2008; Gong et al., 2010); NACs represent one of the largest plant-specific transcription factor gene families (Nakashima et al., 2012; Franco-Zorrilla et al., 2014; Thirumalaikumar et al., 2018).

Previously in *Capsicum annuum* a whole of 104 CaNAC genes were identified by using 47 NAC proteins as query (Diao et al., 2018); here, we identified 113 sequences putatively assigned to the NAC family, presumably due to the different bioinformatics approach adopted. However, to date, only three NACs have been functionally characterized in *C. annuum*: CaNAC035, which acts through multiple signaling pathways and regulates the tolerance to abiotic stress (Zhang et al., 2020); CaNAC064, which positively modulates cold tolerance (Hou et al., 2020); and CaNAC2, which is mainly expressed in seeds and roots, and whose silencing results in increased susceptibility to cold stress (Guo et al., 2015).

On the basis of our RNA sequencing data, only two NACs, i.e., *CaNAC072* and *CaNAC104*, were always upregulated after drought stress and downregulated after recovery, regardless of the stage of plant development; furthermore, following their transient expression in Nicotiana benthamiana leaves, both TFs were localized in the nucleus. Our results thus suggested a possible role of CaNAC072 and CaNAC104 as drought response switchers, and we, therefore, focused our attention on them.

The upregulation of the two NACs was induced at different times following treatments with NaCl, PEG and ABA, whereby *CaNAC072* appeared to be a more rapid responder to abiotic stress than *CaNAC104*. This appears to be confirmed by the not significant upregulation of the slower responsive gene *CaNAC104* after PEG treatment, likely due to a less marked drought stress induced by this treatment compared to the one we imposed (i.e., stop of irrigation up to leaf withering) before our transcriptome analyses.

VIGS represents a fast and powerful reverse genetics tool for analyzing the function of genes in many plant species. In the last decade, VIGS has been successfully applied in Capsicum for the functional characterization of the RING Type E3 ligase CaAIR1, involved in ABA signaling and drought stress response (Park et al., 2014), the ethylene-responsive TF CaAIEF1, involved in enhancing ABA sensitivity and drought tolerance (Hong et al., 2017) and, more recently, the CaNAC035 TF (Zhang et al., 2020). Here, a VIGS protocol, whose efficiency was validated by disabling phytoene desaturase (PDS) as a visible reporter/marker, was applied for transient gene silencing of *CaNAC072* and *CaNAC104*. No difference in the tolerance to drought stress was observed in *CaNAC104*-silenced pepper in comparison to control plants, and a similar relative water content (RWC) and water loss was detected in detached leaves of both plants. This result indicates that the stress-responsive TF CaNAC104 is a component of a complex regulatory network activated by different abiotic stresses, but seemingly it does not play a key role in influencing the response to drought stress. However, considering that CaNAC104 is the ortholog of the Arabidopsis *XND1* gene, which plays a major role in root xylem tracheids (vessels) (Tang et al., 2018), it cannot be excluded that its inhibition by VIGS was not effective and thus did not alter the response of the silenced plants in comparison to the control plants.

Unexpectedly, the silencing of *CaNAC072* improved the tolerance of bell pepper to drought stress, as silenced plants did not display a wilting phenotype after 7 days of no watering, and showed a higher RWC and lower water loss of detached leaves in comparison to control plants. Notably, in contrast to *CaNAC104*-silenced plants, the silencing of *CaNAC072* induced a marked upregulation of stress-responsive genes, including *CaNAC019* and *CaNAC055*. Furthermore, after severe drought (9 days of water withholding), about 15% of the *CaNAC072*-silenced plants survived, while none of the TRV2:00 and *CaNAC104*-silenced plants did.

In Arabidopsis, both ANAC019 and ANAC055 are closely related and functionally redundant homologs of ANAC072, and they act as positive regulators of the drought response (Tran et al., 2004; Ye et al., 2017). On the basis of our RNA-seq experiment, bell pepper plants exposed to acute water stress episodes activated the expression of *CaNAC072*, but not of *CaNAC019* and *CaNAC055* (**Supplementary Table S21**). However, since the latter two were upregulated in drought-stressed *CaNAC072*-silenced plants, it can be assumed that their upregulation might compensate the disabling of *CaNAC072*, because of a complex gene regulatory network involved in the response to abiotic stresses (Bouché and Bouchez, 2001; El-Brolosy and Stainier, 2017). Similarly, in the Arabidopsis *anac019* mutant, drought stress induced the upregulation of *ANAC055* and *ANAC072* (Sukiran et al., 2019).

CaNAC072 is the orthologue of the Arabidopsis gene *ANAC072*, a gene activated following drought stress and conferring drought stress tolerance when overexpressed (Tran et al., 2004). With the goal to shed light on its role, we generated transgenic Arabidopsis plants overexpressing *CaNAC072* fused to a green fluorescent protein (GFP) coding sequence (CaNAC072:GFP-OX plants) in Col-0 as well as in *anac072* single and *anac019/anac055/anac072* triple mutants. After 20 days of water withholding, no difference in water stress tolerance between the Col-0 wild type and *CaNAC072* overexpressors was observed. Furthermore, no recovery of function, i.e. increased drought tolerance, was observed when *CaNAC072* was overexpressed in the two Arabidopsis mutant backgrounds. This suggests that the interactions between a NAC protein and the cis-acting elements of target genes as well as the gene regulatory pathway and downstream signaling can vary between plant species. Indeed, previous studies highlighted that orthologous genes might not play the same function in different species, like in the case of Arabidopsis CBF homologs identified in tomato (i.e., LeCBF1, LeCBF2 or

LeCBF3) and known to be involved in cold acclimation, and of which only LeCBF1 increased freezing tolerance in transgenic Arabidopsis plants (Zhang et al., 2004).

CONCLUSION

Following RNA-sequencing, we identified two transcription factors, CaNAC072 and CaNAC104, that were always up-regulated by acute drought stress and down-regulated after recovery when water was again supplied, at three stages of bell pepper development. The transcriptional regulation and fine-tuning of NAC gene expression is determined by complex signaling pathways: a single NAC gene often responds to several stress factors and participates in the regulation of different processes, as a positive or negative regulator. In our study, the VIGS-mediated silencing of *CaNAC104* did not affect drought tolerance, while silencing of *CaNAC072* increased drought stress tolerance in pepper plantlets. An important result of our study is that *CaNAC072* did not recover the function of its homolog *ANAC072* in *Arabidopsis thaliana*. This observation suggests that the two TF bind to different cis-acting elements of target genes in the two plant species, or that interactions with other TFs or regulatory proteins that ANAC072 might undergo are not faithfully replicated by CaNAC072.

Future research may take advantage of CRISPR/Cas9-mediated genome editing to stably knockout the NAC genes in *C. annuum* for further characterization of their mode of action. This, however, requires the availability of a robust *in vitro* culture and regeneration protocol for this recalcitrant species, which we are currently establishing.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA668245.

AUTHOR CONTRIBUTIONS

DB, SB, and SL planned the experiments. SB, IK, and SL provided plant material. DB carried out the drought stress experiments, sampling, phenotype analyses, plant transformation, and statistical analyses. KS performed the cloning and primer design. LB performed the bioinformatics analyses. DB, AA, SB, AM, LB, and SL contributed to data interpretation and preparation of graphic material. DB and SL wrote the manuscript and DB curated the supplementary materials. DB, SB, AM, AA, LB, and SL discussed the results and revised the manuscript.

FUNDING

This research was funded by the "Cassa di Risparmio di Cuneo" Foundation, under the research project: Resequencing of sweet pepper ecotypes for enhancing their quality and for their traceability (RISEPP).

ACKNOWLEDGMENTS

The authors thank Anna Maria Milani, Danila Valentino, Dr. Shuchao Dong and Dr. Venkatesh Thirumalaikumar for technical support. This research was funded by the "Cassa di Risparmio di Cuneo" Foundation, under the research project: Resequencing of sweet pepper ecotypes for enhancing their quality and for their traceability (RISEPP). DB thanks the University of Turin for havingfunded his PhD fellowship. SB thanks the Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, for financial support. Funding of the ERA-CAPS project "AbioSen" by the Deutsche Forschungsgemeinschaft (MU 1199/16-1) is greatly acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.743902/full#supplementary-material

REFERENCES

Acquadro, A., Barchi, L., Portis, E., Nourdine, M., Carli, C., Monge, S., et al. (2020). Whole Genome Resequencing of Four Italian Sweet Pepper Landraces Provides Insights on Sequence Variation in Genes of Agronomic Value. *Sci. Rep.* 10, 1–16. doi:10.1038/s41598-020-66053-2

Ahn, Y.-K., Manivannan, A., Karna, S., Jun, T.-H., Yang, E.-Y., Choi, S., et al. (2018). Whole Genome Resequencing of Capsicum Baccatum and *Capsicum annuum* to Discover Single Nucleotide Polymorphism Related to Powdery Mildew Resistance. *Sci. Rep.* 8, 1–11. doi:10.1038/s41598-018-23279-5

Arvidsson, S., Kwasniewski, M., Riaño-Pachón, D. M., and Mueller-Roeber, B. (2008). QuantPrime - A Flexible Tool for Reliable High-Throughput Primer Design for Quantitative PCR. *BMC Bioinformatics* 9, 465. doi:10.1186/1471-2105-9-465

Balazadeh, S., Riaño-Pachón, D. M., and Mueller-Roeber, B. (2008). Transcription Factors Regulating Leaf Senescence in *Arabidopsis thaliana*. *Plant Biol.* 10, 63–75. doi:10.1111/j.1438-8677.2008.00088.x

Borràs, D., Plazas, M., Moglia, A., and Lanteri, S. (2021). The Influence of Acute Water Stresses on the Biochemical Composition of bell Pepper (*Capsicum annuum* L.) Berries. *J. Sci. Food Agric.* 101, 4724–4734. doi:10.1002/jsfa.11118

Bouché, N., and Bouchez, D. (2001). Arabidopsis Gene Knockout: Phenotypes Wanted. *Curr. Opin. Plant Biol.* 4, 111–117. doi:10.1016/S1369-5266(00)00145-X

Boyer, J. S., Byrne, P., Cassman, K. G., Cooper, M., Delmer, D., Greene, T., et al. (2013). The U.S. Drought of 2012 in Perspective: A Call to Action. *Glob. Food Security* 2, 139–143. doi:10.1016/j.gfs.2013.08.002

Chung, P. J., Jung, H., Choi, Y. D., and Kim, J.-K. (2018). Genome-wide Analyses of Direct Target Genes of Four rice NAC-Domain Transcription Factors Involved in Drought Tolerance. *BMC Genomics* 19, 40. doi:10.1186/s12864-017-4367-1

Clough, S. J., and Bent, A. F. (1998). Floral Dip: a Simplified Method for Agrobacterium-Mediated Transformation of Arabidopsis Thaliana. *Plant J.* 16, 735–743. doi:10.1046/j.1365-313x.1998.00343.x

Devkar, V., Thirumalaikumar, V. P., Xue, G. P., Vallarino, J. G., Turečková, V., Strnad, M., et al. (2020). Multifaceted Regulatory Function of Tomato SlTAF1 in the Response to Salinity Stress. *New Phytol.* 225, 1681–1698. doi:10.1111/nph.16247

Diao, W., Snyder, J., Wang, S., Liu, J., Pan, B., Guo, G., et al. (2018). Genome-wide Analyses of the NAC Transcription Factor Gene Family in Pepper (*Capsicum annuum* L.): Chromosome Location, Phylogeny, Structure, Expression Patterns, Cis-Elements in the Promoter, and Interaction Network. *Ijms* 19, 1028. doi:10.3390/ijms19041028

Dominguez, P. G., and Carrari, F. (2015). ASR1 Transcription Factor and its Role in Metabolism. *Plant Signaling Behav.* 10, e992751–5. doi:10.4161/15592324.2014.992751

El-Brolosy, M. A., and Stainier, D. Y. R. (2017). Genetic Compensation: A Phenomenon in Search of Mechanisms. *PLoS Genet.* 13, e1006780. doi:10.1371/journal.pgen.1006780

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam Protein Families Database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi:10.1093/nar/gky995

Fernandez-Pozo, N., Rosli, H. G., Martin, G. B., and Mueller, L. A. (2015). The SGN VIGS Tool: User-Friendly Software to Design Virus-Induced Gene Silencing (VIGS) Constructs for Functional Genomics. *Mol. Plant* 8, 486–488. doi:10.1016/j.molp.2014.11.024

Fitton, N., Alexander, P., Arnell, N., Bajzelj, B., Calvin, K., Doelman, J., et al. (2019). The Vulnerabilities of Agricultural Land and Food Production to Future Water Scarcity. *Glob. Environ. Change* 58, 101944. doi:10.1016/j.gloenvcha.2019.101944

FAOSTAT (2018). *Food Organ*. United Nations.Rome, Lazio, Italy

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding Specificities of Plant Transcription Factors and Their Potential to Define Target Genes. *Proc. Natl. Acad. Sci. USA* 111, 2367–2372. doi:10.1073/pnas.1316278111

Gamboa-Tuz, S. D., Pereira-Santana, A., Zamora-Briseño, J. A., Castano, E., Espadas-Gil, F., Ayala-Sumuano, J. T., et al. (2018). Transcriptomics and Co-expression Networks Reveal Tissue-specific Responses and Regulatory Hubs under Mild and Severe Drought in Papaya (Carica Papaya L.). *Sci. Rep.* 8, 1–16. doi:10.1038/s41598-018-32904-2

Godoy, J. A., Lunar, R., Torres-Schumann, S., Moreno, J., Rodrigo, R. M., and Pintor-Toro, J. A. (1994). Expression, Tissue Distribution and Subcellular Localization of Dehydrin TAS14 in Salt-Stressed Tomato Plants. *Plant Mol. Biol.* 26, 1921–1934. doi:10.1007/BF00019503

Gong, P., Zhang, J., Li, H., Yang, C., Zhang, C., Zhang, X., et al. (2010). Transcriptional Profiles of Drought-Responsive Genes in Modulating Transcription Signal Transduction, and Biochemical Pathways in Tomato. *J. Exp. Bot.* 61, 3563–3575. doi:10.1093/jxb/erq167

Guo, W.-L., Wang, S.-B., Chen, R.-G., Chen, B.-H., Du, X.-H., Yin, Y.-X., et al. (2015). Characterization and Expression Profile of *CaNAC2* Pepper Gene. *Front. Plant Sci.* 6, 755. doi:10.3389/fpls.2015.00755

Han, Q., Zhang, J., Li, H., Luo, Z., Ziaf, K., Ouyang, B., et al. (2012). Identification and Expression Pattern of One Stress-Responsive NAC Gene from *Solanum lycopersicum*. *Mol. Biol. Rep.* 39, 1713–1720. doi:10.1007/s11033-011-0911-2

Hong, E., Lim, C. W., Han, S.-W., and Lee, S. C. (2017). Functional Analysis of the Pepper Ethylene-Responsive Transcription Factor, CaAIEF1, in Enhanced ABA Sensitivity and Drought Tolerance. *Front. Plant Sci.* 8, 1407. doi:10.3389/fpls.2017.01407

Hou, X.-m., Zhang, H.-f., Liu, S.-y., Wang, X.-k., Zhang, Y.-m., Meng, Y.-c., et al. (2020). The NAC Transcription Factor CaNAC064 Is a Regulator of Cold Stress Tolerance in Peppers. *Plant Sci.* 291, 110346. doi:10.1016/j.plantsci.2019.110346

Iovieno, P., Punzo, P., Guida, G., Mistretta, C., Van Oosten, M. J., Nurcato, R., et al. (2016). Transcriptomic Changes Drive Physiological Responses to Progressive Drought Stress and Rehydration in Tomato. *Front. Plant Sci.* 7, 371. doi:10.3389/fpls.2016.00371

Ji, X., Shiran, B., Wan, J., Lewis, D. C., Jenkins, C. L. D., Condon, A. G., et al. (2010). Importance of Pre-anthesis Anther Sink Strength for Maintenance of Grain Number during Reproductive Stage Water Stress in Wheat. *Plant Cel Environ.* 33, 926–942. doi:10.1111/j.1365-3040.2010.02130.x

Kakumanu, A., Ambavaram, M. M. R., Klumas, C., Krishnan, A., Batlang, U., Myers, E., et al. (2012). Effects of Drought on Gene Expression in Maize Reproductive and Leaf Meristem Tissue Revealed by RNA-Seq. *Plant Physiol.* 160, 846–867. doi:10.1104/pp.112.200444

Kang, Y. J., Ahn, Y.-K., Kim, K.-T., and Jun, T.-H. (2016). Resequencing of Capsicum Annuum Parental Lines (YCM334 and Taean) for the Genetic Analysis of Bacterial Wilt Resistance. *BMC Plant Biol.* 16, 235. doi:10.1186/s12870-016-0931-0

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY Vectors for Agrobacterium-Mediated Plant Transformation. *Trends Plant Sci.* 7, 193–195. doi:10.1016/S1360-1385(02)02251-3

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., et al. (2007). The AtGenExpress Global Stress Expression Data Set: Protocols, Evaluation and Model Data Analysis of UV-B Light, Drought and Cold Stress Responses. *Plant J.* 50, 347–363. doi:10.1111/j.1365-313X.2007.03052.x

Kim, D., Langmead, B., and Salzberg, S. L. (2015). hisat2. *Nat. Methods* 12, 357–360. doi:10.1038/nmeth.3317

Kim, S., Park, J., Yeom, S.-I., Kim, Y.-M., Seo, E., Kim, K.-T., et al. (2017). New Reference Genome Sequences of Hot Pepper Reveal the Massive Evolution of Plant Disease-Resistance Genes by Retroduplication. *Genome Biol.* 18, 1–11. doi:10.1186/s13059-017-1341-9

Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J. M., Lee, H.-A., et al. (2014). Genome Sequence of the Hot Pepper Provides Insights into the Evolution of Pungency in Capsicum Species. *Nat. Genet.* 46, 270–278. doi:10.1038/ng.2877

Lee, J., Shim, D., Moon, S., Kim, H., Bae, W., Kim, fnm., et al. (2018). Genome-wide Transcriptomic Analysis of BR-Deficient Micro-tom Reveals Correlations between Drought Stress Tolerance and Brassinosteroid Signaling in Tomato. *Plant Physiol. Biochem.* 127, 553–560. doi:10.1016/j.plaphy.2018.04.031

Lee, S., and Choi, D. (2013). Comparative Transcriptome Analysis of Pepper (*Capsicum annuum*) Revealed Common Regulons in Multiple Stress Conditions and Hormone Treatments. *Plant Cel Rep.* 32, 1351–1359. doi:10.1007/s00299-013-1447-9

Liu, Y., Schiff, M., and Dinesh-Kumar, S. P. (2002). Virus-induced Gene Silencing in Tomato. *Plant J.* 31, 777–786. doi:10.1046/j.1365-313X.2002.01394.x

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Ma, N.-N., Zuo, Y.-Q., Liang, X.-Q., Yin, B., Wang, G.-D., and Meng, Q.-W. (2013). The Multiple Stress-Responsive Transcription Factor SlNAC1 improves the Chilling Tolerance of Tomato. *Physiol. Plantarum* 149, 474–486. doi:10.1111/ppl.12049

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* 47 (W1), W636–W641. doi:10.1093/nar/gkz268

Mao, C., Ding, W., Wu, Y., Yu, J., He, X., Shou, H., et al. (2007). Overexpression of a NAC-domain Protein Promotes Shoot Branching in Rice. *New Phytol.* 176, 288–298. doi:10.1111/j.1469-8137.2007.0217710.1111/j.1469-8137.2007.02177.x

Marusig, D., and Tombesi, S. (2020). Abscisic Acid Mediates Drought and Salt Stress Responses in Vitis Vinifera-A Review. *Ijms* 21, 8648. doi:10.3390/ijms21228648

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: Improving Coverage, Classification and Access to Protein Sequence Annotations. *Nucleic Acids Res.* 47, D351–D360. doi:10.1093/nar/gky1100

Mittal, S., Banduni, P., Mallikarjuna, M. G., Rao, A. R., Jain, P. A., Dash, P. K., et al. (2018). Structural, Functional, and Evolutionary Characterization of Major Drought Transcription Factors Families in Maize. *Front. Chem.* 6, 177. doi:10.3389/fchem.2018.00177

Mohanta, T. K., Yadav, D., Khan, A., Hashem, A., Tabassum, B., Khan, A. L., et al. (2020). Genomics, Molecular and Evolutionary Perspective of NAC Transcription Factors. *PLoS One* 15, e0231425. doi:10.1371/journal.pone.0231425

Msanne, J., Lin, J., Stone, J. M., and Awada, T. (2011). Characterization of Abiotic Stress-Responsive *Arabidopsis thaliana RD29A* and *RD29B* Genes and

Evaluation of Transgenes. *Planta* 234, 97–107. doi:10.1007/s00425-011-1387-y

Muñoz-Mayor, A., Pineda, B., Garcia-Abellán, J. O., Antón, T., Garcia-Sogo, B., Sanchez-Bel, P., et al. (2012). Overexpression of Dehydrin *Tas14* Gene Improves the Osmotic Stress Imposed by Drought and Salinity in Tomato. *J. Plant Physiol.* 169, 459–468. doi:10.1016/j.jplph.2011.11.018

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). NAC Transcription Factors in Plant Abiotic Stress Responses. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1819, 97–103. doi:10.1016/j.bbagrm.2011.10.005

Newton, A. C., Johnson, S. N., Gregory, P. J., Newton, C., Johnson, N., and Gregory, J. (2011). Implications of Climate Change for Diseases, Crop Yields and Food Security. *Euphytica* 179, 3–18. doi:10.1007/s10681-011-0359-4

Olvera-Carrillo, Y., Campos, F., Reyes, J. L., Garciarrubio, A., and Covarrubias, A. A. (2010). Functional Analysis of the Group 4 Late Embryogenesis Abundant Proteins Reveals Their Relevance in the Adaptive Response during Water Deficit in Arabidopsis. *Plant Physiol.* 154, 373–390. doi:10.1104/pp.110.158964

Osakabe, Y., Osakabe, K., Shinozaki, K., and Tran, L.-S. P. (2014). Response of Plants to Water Stress. *Front. Plant Sci.* 5, 86. doi:10.3389/fpls.2014.00086

Park, C., Lim, C. W., Baek, W., and Lee, S. C. (2015). RING Type E3 Ligase CaAIR1 in Pepper Acts in the Regulation of ABA Signaling and Drought Stress Response. *Plant Cel Physiol* 56, 1808–1819. doi:10.1093/pcp/pcv103

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome Sequencing of Cultivated and Wild Peppers Provides Insights into Capsicum Domestication and Specialization. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5135–5140. doi:10.1073/pnas.1400975111

Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002). DNA-binding Specificity of the ERF/AP2 Domain of Arabidopsis DREBs, Transcription Factors Involved in Dehydration- and Cold-Inducible Gene Expression. *Biochem. Biophysical Res. Commun.* 290, 998–1009. doi:10.1006/bbrc.2001.6299

Sampathkumar, A., and Wightman, R. (2015). Live Cell Imaging of the Cytoskeleton and Cell wall Enzymes in Plant Cells. *Methods Mol. Biol.* 1242, 133–141. doi:10.1007/978-1-4939-1902-4_12

Seki, M., Ishida, J., Narusaka, M., Fujita, M., Nanjo, T., Umezawa, T., et al. (2002). Monitoring the Expression Pattern of Around 7,000 Arabidopsis Genes under ABA Treatments Using a Full-Length cDNA Microarray. *Funct. Integr. Genomics* 2, 282–291. doi:10.1007/s10142-002-0070-6

Senthil-Kumar, M., and Mysore, K. S. (2014). Tobacco Rattle Virus-Based Virus-Induced Gene Silencing in *Nicotiana benthamiana*. *Nat. Protoc.* 9, 1549–1562. doi:10.1038/nprot.2014.092

Sharoni, A. M., Nuruzzaman, M., Satoh, K., Moumeni, A., Attia, K., Venuprasad, R., et al. (2012). Comparative Transcriptome Analysis of AP2/EREBP Gene Family under normal and Hormone Treatments, and under Two Drought Stresses in NILs Setup by Aday Selection and IR64. *Mol. Genet. Genomics* 287, 1–19. doi:10.1007/s00438-011-0659-3

Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Gene Networks Involved in Drought Stress Response and Tolerance. *J. Exp. Bot.* 58, 221–227. doi:10.1093/jxb/erl164

Sinclair, T. R. (2011). Challenges in Breeding for Yield Increase for Drought. *Trends Plant Sci.* 16, 289–293. doi:10.1016/j.tplants.2011.02.008

Singh, D., and Laxmi, A. (2015). Transcriptional Regulation of Drought Response: A Tortuous Network of Transcriptional Factors. *Front. Plant Sci.* 6, 895. doi:10.3389/fpls.2015.00895

Söderman, E., Mattsson, J., and Engström, P. (1996). The Arabidopsis Homeobox Gene *ATHB-7* Is Induced by Water Deficit and by Abscisic Acid. *Plant J.* 10, 375–381. doi:10.1046/j.1365-313X.1996.10020375.x

Sprenger, H., Erban, A., Seddig, S., Rudack, K., Thalhammer, A., Le, M. Q., et al. (2018). Metabolite and Transcript Markers for the Prediction of Potato Drought Tolerance. *Plant Biotechnol. J.* 16, 939–950. doi:10.1111/pbi.12840

Sukiran, N. L., Ma, J. C., Ma, H., and Su, Z. (2019). ANAC019 Is Required for Recovery of Reproductive Development under Drought Stress in Arabidopsis. *Plant Mol. Biol.* 99, 161–174. doi:10.1007/s11103-018-0810-1

Tang, N., Shahzad, Z., Lonjon, F., Loudet, O., Vailleau, F., and Maurel, C. (2018). Natural Variation at XND1 Impacts Root Hydraulics and Trade-Off for Stress Responses in Arabidopsis. *Nat. Commun.* 9, 1–12. doi:10.1038/s41467-018-06430-8

Thirumalaikumar, V. P., Devkar, V., Mehterov, N., Ali, S., Ozgur, R., Turkan, I., et al. (2018). NAC Transcription Factor JUNGBRUNNEN1 Enhances Drought Tolerance in Tomato. *Plant Biotechnol. J.* 16, 354–366. doi:10.1111/pbi.12776

Todaka, D., Takahashi, F., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2019). ABA-responsive Gene Expression in Response to Drought Stress: Cellular Regulation and Long-Distance Signaling. *Adv. Bot. Res.* 92, 83–113. doi:10.1016/bs.abr.2019.05.001

Tran, L.-S. P., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., et al. (2004). Isolation and Functional Analysis of Arabidopsis Stress-Inducible NAC Transcription Factors that Bind to a Drought-Responsive Cis-Element in the Early Responsive to Dehydration Stress 1 Promoter[W]. *Plant Cell* 16, 2481–2498. doi:10.1105/tpc.104.022699

Trujillo, L. E., Sotolongo, M., Menéndez, C., Ochogavía, M. E., Coll, Y., Hernández, I., et al. (2008). SodERF3, a Novel Sugarcane Ethylene Responsive Factor (ERF), Enhances Salt and Drought Tolerance when Overexpressed in Tobacco Plants. *Plant Cel Physiol* 49, 512–525. doi:10.1093/pcp/pcn025

Tuteja, N. (2007). Abscisic Acid and Abiotic Stress Signaling. *Plant Signaling Behav.* 2, 135–138. doi:10.4161/psb.2.3.4156

Yang, R., Deng, C., Ouyang, B., and Ye, Z. (2011). Molecular Analysis of Two Salt-Responsive NAC-Family Genes and Their Expression Analysis in Tomato. *Mol. Biol. Rep.* 38, 857–863. doi:10.1007/s11033-010-0177-0

Ye, H., Liu, S., Tang, B., Chen, J., Xie, Z., Nolan, T. M., et al. (2017). RD26 Mediates Crosstalk between Drought and Brassinosteroid Signalling Pathways. *Nat. Commun.* 8, 14573. doi:10.1038/ncomms14573

Zhang, H., Ma, F., Wang, X., Liu, S., Saeed, U. H., Hou, X., et al. (2020). Molecular and Functional Characterization of CaNAC035, an NAC Transcription Factor from Pepper (*Capsicum annuum* L.). *Front. Plant Sci.* 11, 14. doi:10.3389/fpls.2020.00014

Zhang, P., Fan, Y., Sun, X., Chen, L., Terzaghi, W., Bucher, E., et al. (2019). A Large-scale Circular RNA Profiling Reveals Universal Molecular Mechanisms Responsive to Drought Stress in maize and Arabidopsis. *Plant J.* 98, 697–713. doi:10.1111/tpj.14267

Zhang, X., Fowler, S. G., Cheng, H., Lou, Y., Rhee, S. Y., Stockinger, E. J., et al. (2004). Freezing-sensitive Tomato Has a Functional CBF Cold Response Pathway, but a CBF Regulon that Differs from that of Freezing-tolerantArabidopsis. *Plant J.* 39, 905–919. doi:10.1111/j.1365-313X.2004.02176.x

Zinselmeier, C., Sun, Y., Helentjaris, T., Beatty, M., Yang, S., Smith, H., et al. (2002). The Use of Gene Expression Profiling to Dissect the Stress Sensitivity of Reproductive Development in maize. *Field Crops Res.* 75, 111–121. doi:10.1016/S0378-4290(02)00021-7

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 November 2021
doi: 10.3389/fgene.2021.748979

Edaphoclimatic Descriptors of Wild Tomato Species (*Solanum* Sect. Lycopersicon) and Closely Related Species (*Solanum* Sect. Juglandifolia and Sect. Lycopersicoides) in South America

Gabriela Ramírez-Ojeda[1], *Iris Edith Peralta*[2,3], *Eduardo Rodríguez-Guzmán*[4], *Jaime Sahagún-Castellanos*[1], *José Luis Chávez-Servia*[5], *Tulio Cecilio Medina-Hinostroza*[6], *Jorge Rodrigo Rijalba-Vela*[7], *Leopoldo Pompeyo Vásquez-Núñez*[7] and *Juan Enrique Rodríguez-Pérez*[1]*

[1]Crop Science Department, Horticulture Institute, Chapingo Autonomous University (UACh), Chapingo, Mexico, [2]Agronomy Department, Agricultural Sciences Faculty, National University of Cuyo (UNCUYO), Mendoza, Argentina, [3]Scientific Technological Center CONICET, Argentine Institute for Arid Zones Research, Mendoza, Argentina, [4]Agronomy Department, University Center for Biological and Agricultural Sciences, University of Guadalajara (UdG), Zapopan, Mexico, [5]Interdisciplinary Research Center for Integral Regional Development Oaxaca Unit, National Polytechnic Institute (IPN), Oaxaca, Mexico, [6]Directorate of Genetic Resources and Biosafety, Ministry of Environment, Lima, Peru, [7]Biological Sciences Faculty, National University Pedro Ruiz Gallo (UNPRG), Lambayeque, Peru

Wild species related to cultivated tomato are essential genetic resources in breeding programs focused on food security to face future challenges. The ecogeographic analysis allows identifying the species adaptive ranges and most relevant environmental variables explaining their patterns of actual distribution. The objective of this research was to identify the diversity, ecological descriptors, and statistical relationship of 35 edaphoclimatic variables (20 climatic, 1 geographic and 14 edaphic variables) from 4,649 accessions of 12 wild tomato species and 4 closely related species classified in *Solanum* sect. Lycopersicon and clustered into four phylogenetic groups, namely "Lycopersicon group" (*S. pimpinellifolium*, *S. cheesmaniae*, and *S. galapagense*), "Arcanum group" (*S. arcanum*, *S. chmielewskii*, and *S. neorickii*), "Eriopersicon group" (*S. habrochaites*, *S. huaylasense*, *S. corneliomulleri*, *S. peruvianum*, and *S. chilense*), "Neolycopersicon group" (*S. pennellii*); and two phylogenetically related groups in *Solanum* sect. Juglandifolia (*S. juglandifolium* and *S. ochranthum*), and section Lycopersicoides (*S. lycopersicoides* and *S. sitiens*). The relationship between the climate and edaphic variables were determined by the canonical correlation analysis, reaching 89.2% of variation with the first three canonical correlations. The most significant climatic variables were related to humidity (annual evapotranspiration, annual precipitation, and precipitation of driest month) and physicochemical soil characteristics (bulk density, pH, and base saturation percentage). In all groups, ecological descriptors and diversity patterns were consistent with previous reports. Regarding edaphoclimatic diversity, 12 climate types and 17 soil units were identified

OPEN ACCESS

Edited by:
Nunzio D'Agostino, University of Naples Federico II, Italy

Reviewed by:
Carolina Carrizo García, Instituto Multidisciplinario de Biología Vegetal (IMBIV), Argentina
Muhammad Azhar Nadeem, Sivas University of Science and Technology, Turkey

***Correspondence:**
Juan Enrique Rodríguez-Pérez
erodriguezx@yahoo.com.mx

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

Received: *28 July 2021*
Accepted: *11 October 2021*
Published: *17 November 2021*

Citation:
*Ramírez-Ojeda G, Peralta IE, Rodríguez-Guzmán E, Sahagún-Castellanos J, Chávez-Servia JL, Medina-Hinostroza TC, Rijalba-Vela JR, Vásquez-Núñez LP and Rodríguez-Pérez JE (2021) Edaphoclimatic Descriptors of Wild Tomato Species (*Solanum *Sect. Lycopersicon) and Closely Related Species (*Solanum *Sect. Juglandifolia and Sect. Lycopersicoides) in South America. Front. Genet. 12:748979. doi: 10.3389/fgene.2021.748979*

among all species. This approach has promissory applications for biodiversity conservation and uses valuable genetic resources related to a leading crop.

Keywords: wild tomatoes, edaphoclimatic diversity, ecological descriptors, genetic resources, canonical correlation analysis

INTRODUCTION

Latin America and the Caribbean are regions rich in biodiversity, hosting nearly 60% of the world's biological diversity (UNEP-WCMC, 2016). Within this region, Mesoamerica is recognized as one of the main centers of origin, diversification, domestication, and biological plant diversity of various species of agricultural interest and animal consumption (Fortuny-Fernández et al., 2017). The complex evolutionary history, phylogenetics, geology, biogeography, and climatic variability are some factors that enhance the diversity in this area (UNEP-WCMC, 2016). This condition is essential to ensure food, socioeconomic, and cultural sovereignty for sustainable development and offers a large number of ecosystem services (FAO et al., 2019).

In this sense, tomato (*Solanum lycopersicum* L.) is one of the most cultivated vegetables due to its wide distribution and environmental adaptation in warm, subtropical, and tropical regions with nutritional and commercial importance worldwide (Peralta et al., 2008; Ramírez-Ojeda et al., 2021a). Regarding the place of origin and diversification of tomato, Peru is considered the center of origin with two transitions that involve tomato diversification process; the first one in South America, from wild species *S. pimpinellifolium* L. to a partially domesticated species *S. lycopersicum* L. var. *cerasiforme* (SLC); the second transition occurred in Mesoamerica from SLC to the completely domesticated species *S. lycopersicum* L. var. *lycopersicum*. However, new findings indicate that the origin of SLC may be prior to its domestication since many typical characteristics of tomatoes grown in South America come from this species; SLC is subsequently considered to have been lost or declined once the partially domesticated forms extended to the north (Razifard et al., 2020).

Wild species related to cultivated tomatoes are essential genetic resources in breeding programs focused on food security to face future challenges. Therefore, it is of strategic importance to study the climatic and edaphic factors that help to understand their current distribution patterns, as well as to establish the best indicators predicting possible effects of climate change and natural or anthropic environmental alterations. This is why it is necessary to undertake national and regional strategies for the conservation and use of cultivated and wild tomato genetic resources (Sandoval-Ceballos et al., 2021).

Based on an integrative taxonomy, which includes multiple evidences, the classification of wild tomatoes and their wild relatives was proposed: *Solanum* section Lycopersicon (Mill.) Wettst. comprises cultivated tomato (*S. lycopersicum* L.) and 12 wild tomato species: *S. arcanum* Peralta, *S. cheesmaniae* (L. Riley) Fosberg, *S. chilense* Dunal, *S. chmielewskii* (C. M. Rick, Kesicki, Fobes, and M. Holle), D. M. Spooner, G. J. Anderson, and R.K. Jansen, *S. corneliomulleri* J. F. Macbride, *S. galapagense* S. C. Darwin and Peralta, *S. habrochaites* S. Knapp and D. M. Spooner, *S. huaylasense* Peralta, *S. neorickii* D. M. Spooner, G. J. Anderson and R. K. Jansen, *S. pennellii* Correll, *S. peruvianum* L., and *S. pimpinellifolium* L. Four phylogenetically related *Solanum* species are also considered in the present study: *S. juglandifolium* Dunal, *S. ochranthum* Dunal (*Solanum* sect. Juglandifolia (Rydb.) A. Child), *S. lycopersicoides* Dunal, and *S. sitiens* I. M. Johnston (*Solanum* section *Lycopersicoides* (A. Child) Peralta) (Peralta et al., 2008; Causse et al., 2016; Tropicos.org, 2021).

Evidence of phylogenetic relationships of these species have been studied in detail by Peralta et al. (2008), who have proposed a six-group classification of wild tomatoes and phylogenetically closely related species: Section Lycopersicon: "Lycopersicon group" (*S. pimpinellifolium*, *S. cheesmaniae*, and *S. galapagense*), "Arcanum group" (*S. arcanum*, *S. chmielewskii*, and *S. neorickii*), "Eriopersicon group" (*S. habrochaites*, *S. huaylasense*, *S. corneliomulleri*, *S. peruvianum*, and *S. chilense*), "Neolycopersicon group" (*S. pennellii*); and two outgroups: Section Juglandifolia (*S. juglandifolium* and *S. ochranthum*) and Section Lycopersicoides (*S. lycopersicoides* and *S. sitiens*). This classification has been verified by molecular, genomic, and transcriptomic evidence of wild tomatoes (Rodríguez et al., 2009; Aflitos et al., 2014; Pease et al., 2016) and recently has been used for ecogeographic studies with satisfactory results (Ramírez-Ojeda et al., 2021a).

By considering plant genetic resources as the biological foundation for maintaining and improving crop productivity (Kantar et al., 2015), wild tomato species constitute an important gene pool due to the presence of genes with tolerance and resistance to biotic and abiotic factors (Arellano-Rodríguez et al., 2013; Cervantes-Moreno et al., 2014; Nosenko et al., 2016; Razali et al., 2018; Dinh et al., 2019) with potential use for breeding programs. Additionally, several questions arise about these gene pools, such as current distribution, population dynamics *in situ* or *ex situ*, and how are they used directly or as sources of genes to generate new varieties that respond to current and future basic problems of tomato cultivation (for example, climate change, diseases, pests), including the contribution of genes capable of conferring a greater nutritional–nutraceutical quality to new varieties (Chávez-Servia et al., 2011; Hernández-Bautista et al., 2014).

Identification of variables that derive in adaptation and speciation processes requires a large amount of field data of significant variables in natural populations. Recent developments and the use of remote sensing technologies, as well as a great availability of environmental information derived from Geographic information systems (GIS), have made it possible

to identify patterns of species environmental variations at different scales (Nakazato et al., 2010). These tools and the availability of databases, with passport information of specimens collected in natural areas, allow for verification of the presence of species in a geographic range, as well as possible ecological descriptors, that is, to describe in detail the environmental conditions associated with the distribution of natural populations (Nakazato et al., 2010; Sánchez-González et al., 2018; Vilchez et al., 2019; Ministerio del Ambiente, 2020; Ramírez-Ojeda et al., 2021a).

One way to identify the adaptive ranges and most relevant variables that determine species distribution of valuable genetic resources is through ecogeographic studies, focusing on collection, conservation, characterization, documentation, and use of these resources (Parra-Quinajo et al., 2012; Pease et al., 2016), with the purpose of describing and explaining spatial patterns and processes involved in biodiversity distribution through time and space (Martiny et al., 2006; Tofalo et al., 2013; Délices et al., 2019). Ecogeographic studies of plant genetic resources allow the identification of the adaptive ranges of the species and the most relevant environmental variables that define their distribution (Parra-Quinajo et al., 2012; Ramírez-Ojeda et al., 2021a). Through ecogeographic studies, it is also possible to predict the environmental characteristics of the accession sites (Steiner and Greene, 1996) from ecological descriptors obtained through GIS tools using the geographical location and environmental variables (Lobo-Burle et al., 2013; Sánchez-González et al., 2018; Ramírez-Ojeda et al., 2021a, 2021b).

Currently, several information sources about geographical distribution of tomato species can be found in public databases (GBIF, 2021; Solanaceae Source, 2021; TGRC, 2021), conservation programs and gene banks (Córdoba-Téllez and Molia-Moreno, 2006; Florido et al., 2009; Magallanes-López et al., 2020), and genetic resources baseline studies (Ministerio del Ambiente, 2020), as well as some studies on geographic distribution patterns and ecological and climatic descriptors of wild tomato species (Peralta et al., 2008; Chetelat et al., 2009; Nakazato et al., 2010; Grandillo et al., 2011; Gonzá lez et al., 2013; Vilchez et al., 2019; Ramírez-Ojeda et al., 2021a). However, information regarding edaphic conditions of the sites where these species are located is limited or unknown (Balaguera-López et al., 2009).

Soil, a finite and nonrenewable natural resource, is of great importance in a large number of environmental services such as food and biomass production, climate regulation, carbon fixation, water storage and filtration, biogeochemical cycles, biodiversity reserve, and human physical and cultural environment (Burbano-Orjuela, 2016). Therefore, when considering edaphic together with climatic characteristics, it allows having a better understanding of the ecological and distribution patterns of the species.

Due to the limited edaphic information available regarding optimal characteristics for development of wild tomato species, the aim of the present work was to study ecological descriptors associated with soil characteristics and their relationship and the statistical association with climatic variables. Likewise, it was also analyzed whether the classification of wild tomatoes is related to the edaphoclimatic descriptors and supports the proposed groups of species.

MATERIALS AND METHODS

Database

Initial database consisted of 12,131 accessions of 12 wild tomato species and 4 phylogenetically related species. Of these, 7,482 accessions were eliminated due to atypical data, repeated records, or accessions with little geographic precision and outside natural areas identified according to the altitude and ecological ranges reported (Peralta et al., 2008; Grandillo et al., 2011; Ministerio del Ambiente, 2020). The final 4,649 accessions database came from scientific reports, articles (Sotomayor et al., 2019; Razifard et al., 2020), international plant repositories (Tomato Genetic Resource Center, Global Biodiversity Information Facility, Solanaceae Source) (GBIF, 2021; Solanaceae Source, 2021; TGRC, 2021), and new accessions collected in 2018–2019 in Peru (Ministerio del Ambiente, 2020). The distribution of 16 species is shown in **Figure 1**. The species distribution is shown in Figure A1 in the **Supplementary Material**. It should be noted that *S. lycopersicum* was not included because its wide distribution would not reflect a natural but artificial distribution due to anthropic dispersal as a cultivated or ruderal species.

Environmental Information

For the statistical analysis and ecological descriptors, an environmental information system with 900 m spatial resolution was built with 35 variables (**Table 1**). Nineteen bioclimatic variables were obtained from WorldClim version 2.1 from period 1970 to 2000 (Fick and Hijmans, 2017). Annual evapotranspiration (ET) was calculated from the sum of monthly values reported by Trabucco and Zomer (2019). Altitude (Alt), a geographic variable, was obtained with an elevation model from WorldClim (Fick and Hijmans, 2017). Alt was analyzed together with climatic variables due to the strong influence on the definition of climates. Finally, 14 edaphic variables obtained from the Harmonized World Soil Database version 1.1 (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009) were used.

Edaphoclimatic diversity patterns were identified from climate types corresponding to world climatic classification proposed by Beck et al. (2018) with the Köppen–Geiger system and soil units from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009) (**Table 2**).

Canonical Correlation Analysis and Ecological Descriptors

A selection of climatic and edaphic variables was made in order to identify a strong linear dependence (collinearity) between more than two explanatory variables. For this purpose, Pearson's correlations were obtained, between variables, eliminating one

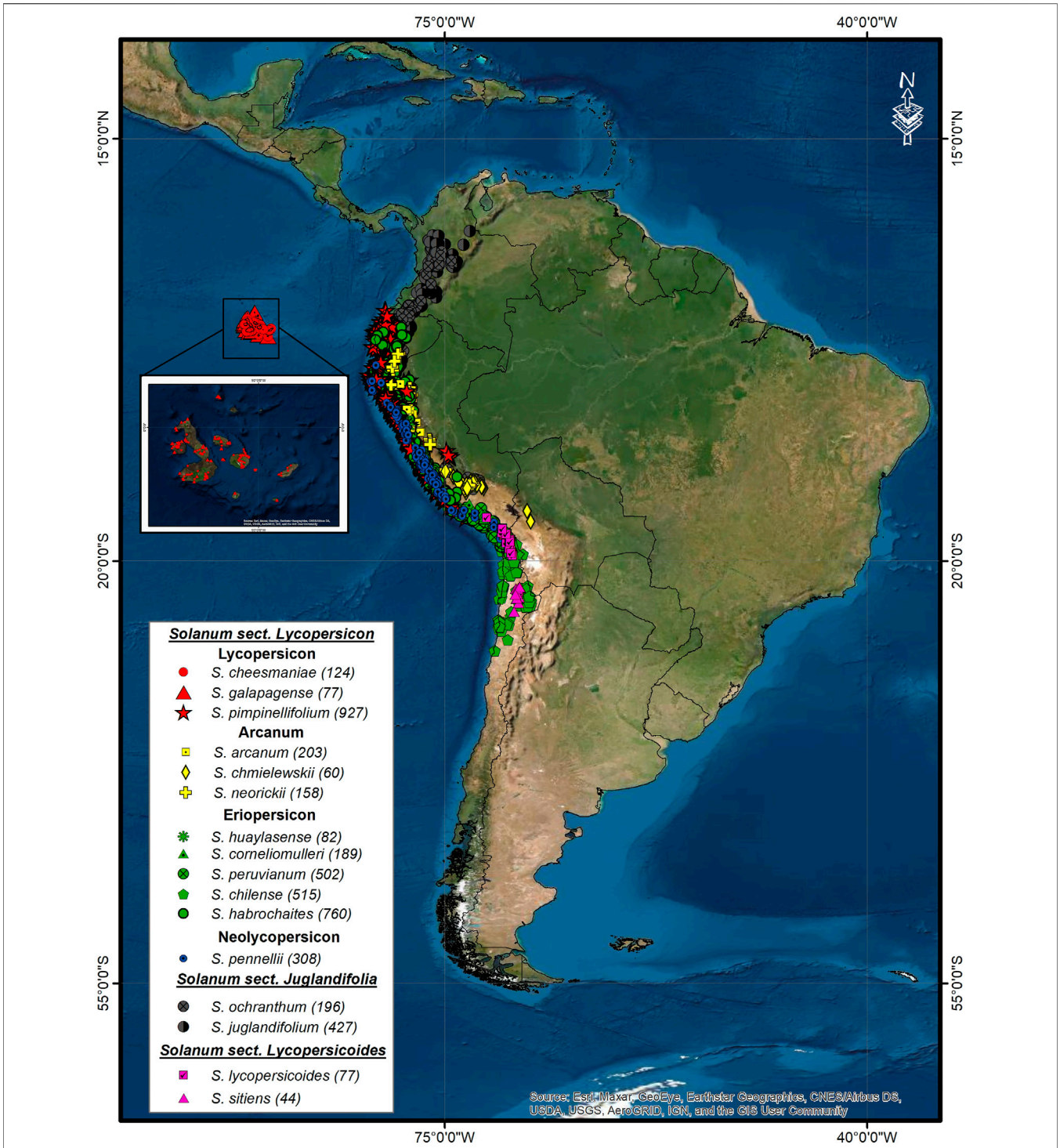

FIGURE 1 | Geographic distribution of 12 wild tomato (*Solanum* sect. Lycopersicon) and 4 closely related species (*Solanum* sect. Juglandifolia, and Lycopersicoides). In parenthesis, number of accessions after each species name.

of each pair whose absolute coefficient was greater than 0.90. The conserved variable was the one that showed the highest number of correlations with other variables, and therefore, the lowest number of non–linearly associated variables was maintained.

With the selected variables, a canonical correlation analysis was carried out to identify the relationship between the group of climatic variables and the group of edaphic variables. All statistical analyses were performed using SAS

TABLE 1 | Climatic, geographic, and edaphic variables used in the canonical correlation analysis and ecological descriptors.

Climatic variables
WorldClim variables (1970–2000): Annual mean temperature (**Bio1**, °C), mean diurnal range (**Bio2**, °C), isothermality (**Bio3**, Bio2/Bio7 × 100), temperature seasonality (**Bio4**, standard deviation × 100), maximum temperature of the warmest month (**Bio5**, °C), minimum temperature of coldest month (**Bio6**, °C), temperature annual range (**Bio7**, Bio5-Bio6), mean temperature of wettest quarter (**Bio8**, °C), mean temperature of driest quarter (**Bio9**, °C), mean temperature of warmest quarter (**Bio10**, °C), mean temperature mean of coldest quarter (**Bio11**, °C), annual precipitation (**Bio12**, mm), precipitation of wettest month (**Bio13**, mm), precipitation of driest month (**Bio14**, mm), precipitation seasonality (**Bio15**, coefficient of variation), precipitation of wettest quarter (**Bio16**, mm), precipitation of driest quarter (**Bio17**, mm), precipitation of warmest quarter (**Bio18**, mm), and precipitation of coldest quarter (**Bio19**, mm) (Fick and Hijmans, 2017). Annual evapotranspiration (**ET**, mm) (Trabucco and Zomer, 2019)
Geographic variables
Altitude (**Alt**, masl) (Fick and Hijmans, 2017)
Edaphic variables
Percentage of gravel (**GR**, %), sand (**SA**, %), silt (**SI**, %), and clay (**CL**, %), bulk density (**BD**, kg/dm^3), organic carbon (**CO**, %), **pH**, cation exchange capacity (**CEC** cmol/kg), base saturation (**BS**, %), calcium carbonate (**$CaCO_3$**, %), total exchangeable bases (**TEB**, cmol/kg), calcium sulfate (**$CaSO_4$**, %), salinity (**SAL**, dS/m), and sodium (**SOD**, %) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009)

TABLE 2 | Climate types and soil units used to determine edaphoclimatic diversity patterns.

Climate type
Af (tropical and rainforest), Am (tropical and monsoon), Aw (tropical and savannah), BWh (arid, desert, and hot), BWk (arid, desert, and cold), BSh (arid, steppe, and hot), BSk (arid, steppe, and cold), Csa (temperate, dry summer, and hot summer), Csb (temperate, dry summer, and warm summer), Csc (temperate and dry and cold summer), Cwa (temperate, dry winter, and hot summer), Cwb (temperate, dry winter, and warm summer), Cwc (temperate, dry winter, and cold summer), Cfa (temperate, no dry season, and hot summer), Cfb (temperate, no dry season, and warm summer), Cfc (temperate, no dry season, and cold summer), Dsa (cold, dry summer, and hot summer), Dsb (cold, dry summer, and warm summer), Dsc (cold, dry summer, and cold summer), Dsd (cold, dry summer, and very cold winter), Dwa (cold, dry winter, and hot summer), Dwb (cold, dry winter, and warm summer), Dwc (cold, dry winter, and cold summer), Dwd (cold, dry winter, and very cold winter), Dfa (cold, no dry season, and hot summer), Dfb (cold, no dry season, and warm summer), Dfc (cold, no dry season, and cold summer), Dfd (cold, no dry season, and very cold winter), ET (polar and tundra), and EF (polar and frost) (Beck et al., 2018)
Soil units
AC (Acrisol), **AL** (Alisol), **AN** (Andosol), **AR** (Arenosol), **AT** (Anthrosol), **CH** (Chernozem), **CL** (Calcisol), **CM** (Cambisol), **FL** (Fluvisol), **FR** (Ferralsol), **GL** (Gleysol), **GR** (Greysem), **GY** (Gypsisol), **HS** (Histosol), **KS** (Kastanozem), **LP** (Leptosol), **LV** (Luvisol), **LX** (Lixisol), **NT** (Nitisol), **PD** (Podzoluvisol), **PH** (Phaezem), **PL** (Planosol), **PT** (Plinthosol), **PZ** (Podzol), **RG** (Regosol), **SC** (Solonchak), **SN** (Solonetz), and **VR** (Vertisol) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009)

software (Statistical Analysis System) version 9.3 (SAS Institute, 2011).

Regarding ecological descriptors, these were calculated for each variable and each species (12 wild tomato and 4 phylogenetically related species) with the methodology proposed by Steiner and Greene (1996). Ecological descriptors were determined by vectors calculated with the geographic coordinates of each accession and the punctual value of each variable extracted with GIS.

Subsequently, the edaphic and climatic variables were identified as significant in the canonical correlation analysis; the extreme values (maximum and minimum), the median, and the coefficient of variation (CV = (Q/Med) × 100, where Q = (Q3 − Q1)/2 (interquartile range), and Med = median) were identified.

Finally, to identify the ecological distribution patterns of every group of species, altitude, annual mean temperature, precipitation, and annual evapotranspiration were considered as climatic variables and pH, cation exchange capacity (CEC), bulk density (BD), and base saturation (BS) as edaphic variables. These variables were chosen due to the importance and influence they have on the distribution and development of the species (Ramírez-Ojeda et al., 2021b), in addition to the importance and significance that they showed in the statistical analyses.

Edaphoclimatic Diversity

Edaphoclimatic diversity was identified using GIS tools with the vector of geographic coordinates of each accession and raster images of climate types and soil units (**Table 2**). **Figure 2** shows the distribution of climate types and soil units in South America. With the resulting information, a frequency table by climate type and soil unit was obtained for each species group (6) and for each individual species (16).

Hot spot Analysis

Critical points of species abundance and areas with the greatest diversity concentration were established using ArcGIS with the "Spatial Statistics Tools" module. Spatial density maps were constructed by adding all those accessions of each species with a distance between accessions of 1 km. A distance criterion was chosen based on previous diversity studies of potato species (*Solanum* Sect. *Petota*), the sister group of tomatoes (Hijmans et al., 2002; Spooner et al., 2010). Subsequently, hot spot spatial analysis was performed with Getis-Ord Gi* statistic (Getis and Ord, 1992) to quantify the specific areas of high clustering and spatial significance for species abundance and diversity.

The hot spot analysis determines the spatial grouping of points higher (hot spot) or lower (cold spot) than the expected by a

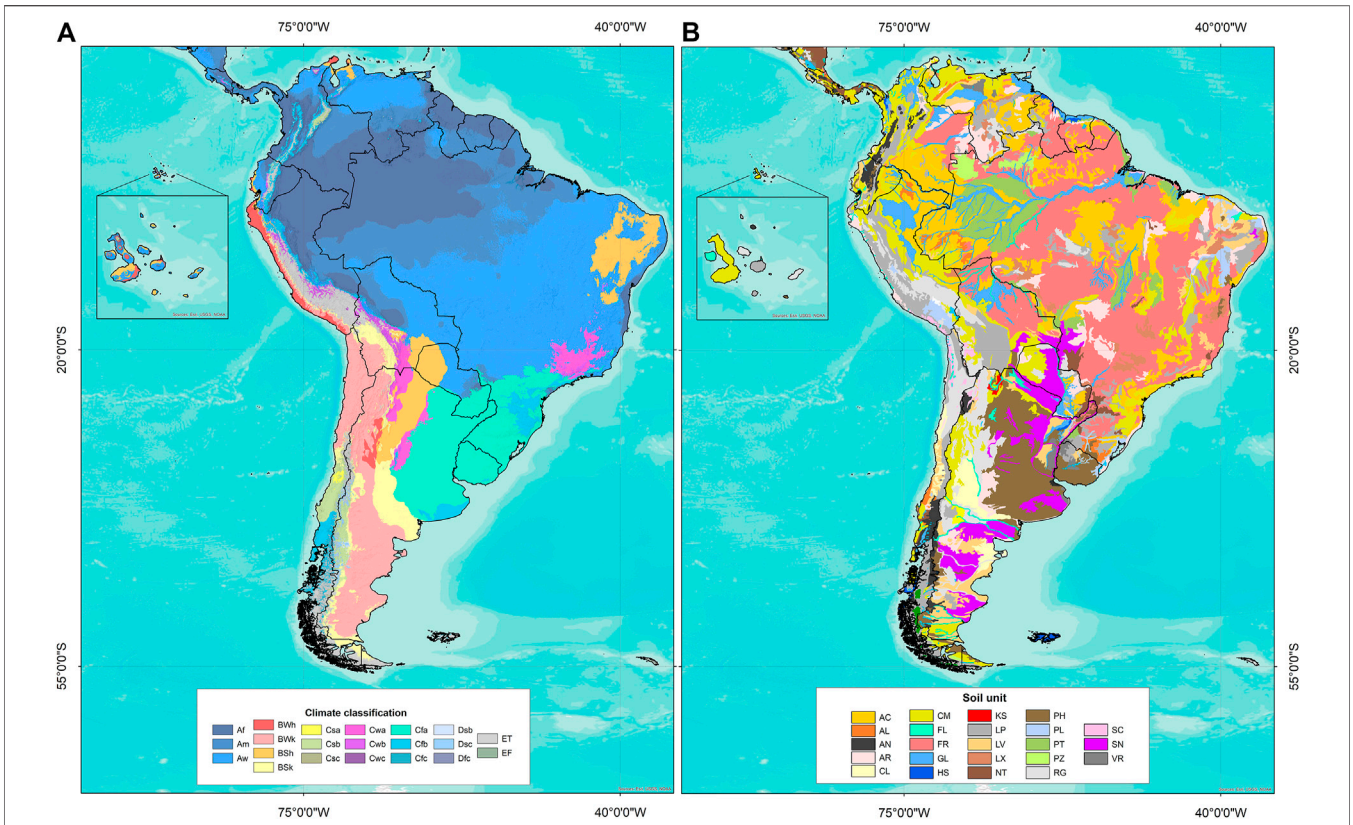

FIGURE 2 | (A) Climate classification according to Beck et al. (2018): Af (tropical and rainforest), Am (tropical and monsoon), Aw (tropical and savannah), BWh (arid, desert, and hot), BWk (arid, desert, and cold), BSh (arid, steppe, and hot), BSk (arid, steppe, and cold), Csb (temperate, dry summer, and warm summer), Cwb (temperate, dry winter, and warm summer), Cfb (temperate, no dry season, and warm summer), and ET (polar, frost). **(B)** Soil units according to FAO/IIASA/ISRIC/ISSCAS/JRC (2009): AC (acrisol), AL (alisol), AN (andosol), AR (arenosol), CL (calcisol), CM (cambisol), FL (fluvisol), FR (ferralsol), GL (gleysol), HS (histosol), KS (kastanozem), LP (leptosol), LV (luvisol), LX (lixisol), NT (nitisol), PH (phaezem), PL (planosol), PT (plinthosol), PZ (podzol), RG (regosol), SC (solonchak), SN (solonetz), VR (vertisol).

random distribution. Significance tests were calculated using z-values (Getis and Ord, 1992).

RESULTS

Canonical Correlation Analysis and Ecological Descriptors

According to Pearson's correlation coefficients, out of the 34 edaphoclimatic variables, 19 did not present collinearity. The variables selected for subsequent statistical analyses and ecological descriptors were annual evapotranspiration, altitude, precipitation of dries month, annual precipitation, temperature annual range, isothermality, mean diurnal range, annual mean temperature, percentage of sand, silt and clay, BD, pH, organic carbon, CEC, BS, calcium carbonate $CaCO_3$), sodicity, and salinity.

The canonical correlation analysis (CCA), performed with two groups of variables (climatic and edaphic), indicated that the first three canonical correlations had values of 0.800, 0.436, and 0.415, respectively, and percentages of explanation of data variation of 71.45, 9.38, and 8.36%, respectively, with a total of 89.20%. Likelihood ratio tests indicated that the three canonical correlations are different from zero ($p \leq 0.0001$).

Correlations between climate characteristics and their canonical variables indicated that CLIMATE1 is associated with annual evapotranspiration (ET, −0.907), annual precipitation (Bio12, −0.947), and precipitation of driest month (Bio14, −0.864). The CLIMATE2 vector is associated with altitude (Alt, −0.5164) and isothermality (Bio3, −0.5642), which roughly quantifies the "hot" (low values) and "cold" (high values) regions. The CLIMATE3 vector represents the mean annual temperature (Bio1, 0.7145) and mean diurnal range (Bio2, 0.793).

Regarding correlations of soil canonical variables, the SOIL1 vector represent BD (0.801), pH (0.658), and BS percentage (BS, 0.707). High values of the SOIL2 vector identify soils with high content of sand (0.478), pH (0.552), $CaCO_3$ concentration (0.546), and low clay content (−0.427). Finally, the SOIL3 vector does not show important correlations.

The correlations between the CLIMATE vectors with the original soil variables indicate that CLIMATE1 had correlations of importance with BD (0.641), pH (0.527), CEC (−0.452), BS (0.566), and $CaCO_3$ content (0.424). CLIMA2 and

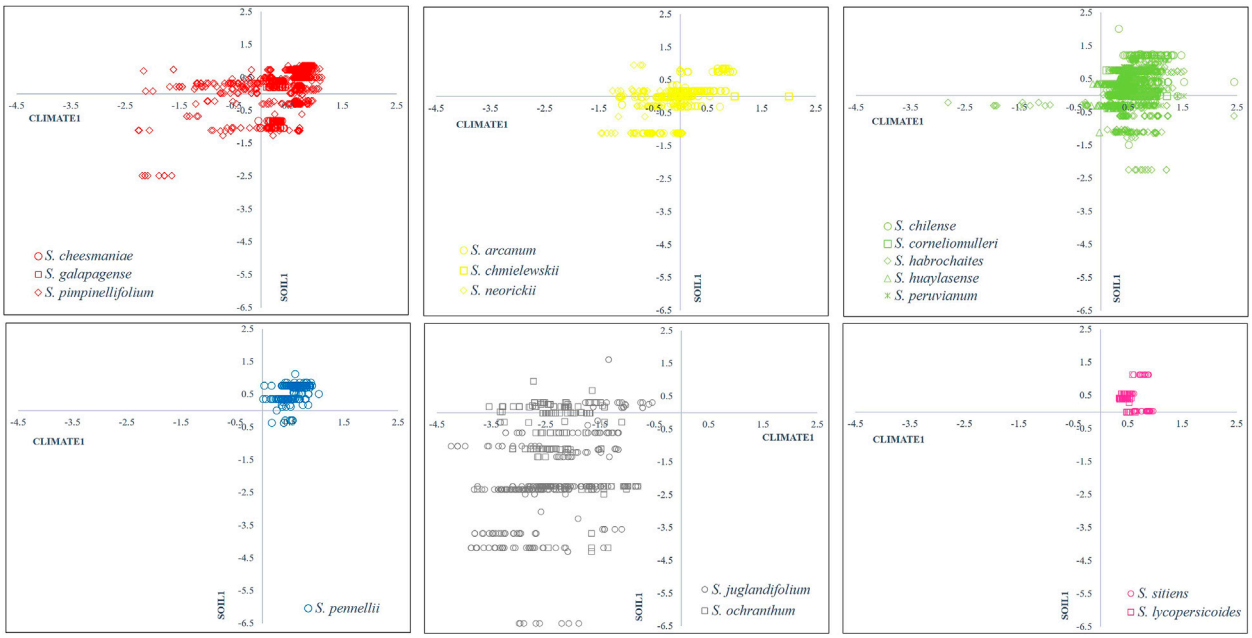

FIGURE 3 | Scatter plot of SOIL1 vs CLIMATE1 canonical vectors from CCA with 0.80 correlation and 74% of total variation. Group 1 (red): Section Lycopersicon: "Lycopersicon group", Group 2 (yellow): "Arcanum group", Group 3 (green): "Eriopersicon group", Group 4 (blue): "Neolycopersicon group" outgroup close related species in Section Juglandifolia Group 5 (grey): and Group 6 (pink): Section Lycopersicoides. In each scatter plot first-quadrant (I) represents environments with low annual evapotranspiration (ET), annual precipitation (Bio12), precipitation of driest month (Bio14) and high pH, BD, and base saturation (BS); second-quadrant (II) represents environments with high ET, Bio12, Bio14 and high pH, BD and BS; third-quadrant (III) represents environments with high ET, Bio12, Bio14 and low pH, BD and BS; fourth-quadrant (IV) represents environments with low ET, Bio12, Bio14, and low pH, BD, and BS.

CLIMA3 did not show correlation of importance. SOIL1 with climate variables showed associations with annual evapotranspiration (ET, −0.7267), annual precipitation (Bio12, −0.7585), precipitation of driest month (Bio14, −0.6924), isothermality (Bio3, 0.5404), and temperature annual range (Bio7, 0.5117). SOIL2 and SOIL3 did not show correlation of importance.

Figure 3 shows the relationship between canonical variables CLIMATE1 and SOIL1, representing 71.45% of the total data variability and a positive correlation of both canonical variables of 0.80. This figure shows the distribution and ecological adaptation of every species regarding canonical correlations.

Table 3 and **Table 4** show the ecological descriptors of edaphic and climatic variables identified as significant in the first and second canonical correlation. These results are mostly consistent with the environmental ranges previously reported in other studies. Table A1 in **Supplementary Material** shows the ecological descriptors of the rest of the variables.

Figure 4 shows the boxplots for four climatic variables for each of the six species groups, as well as the amplitude observed for each variable.

Among the main findings, it can be observed that groups 4 (*S. pennellii*) and 6 (*S. lycopersicoides* and *S. sitiens*) are ones that contain the species that distributes in environments with the lowest availability of precipitation and evapotranspiration. Considering altitude, group 1 (*S. pimpinellifolium*, *S. cheesmaniae*, and *S. galapagense*) has the lowest average altitude, while group 6 (*S. lycopersicoides* and *S. sitiens*) has the highest average altitude. Group 1 was located in environments with the highest mean annual temperature; by contrast, group 6 had the lowest average annual temperature. Groups 2, 3, and 5 remained in transition climatic conditions with the rest of the phylogenetic groups.

The analysis of four edaphic variables in **Figure 5** determines that group 5 (*S. juglandifolium* and *S. ochranthum*) has the lowest pH average. In all groups, BD was relatively constant, with similar values in all species. The mean BS in most of the groups was greater than 80%, except for group 5, with an average value around 40%. In general, soil characteristics in all groups of species were relatively similar, except for group 5 (*S. ochranthum* and *S. juglandifolium*) which presented an opposite trend.

Edaphoclimatic Diversity

The edaphoclimatic diversity found in 16 species is shown in **Figure 6** and **Figure 7**. Regarding, climate diversity, it was possible to identify 12 climate types of the 21 reported for Latin America by Beck et al. (2018).

Within the six phylogenetically related groups identified by Peralta et al. (2008) and used by Ramírez-Ojeda et al. (2021a), specific climate type patterns can be observed, with the same climate types occurring in different proportions within each group (**Figure 6**), confirming in most of the groups, the environmental distribution similarity between the species that make them up.

S. habrochaites has the greatest diversity (11 climate types), intermedium diversity (8 climate types) was found in *S. arcanum*,

TABLE 3 | Ecological descriptors of climatic and edaphic variables associated with the first canonical correlation (71.4%) for12 species of wild tomato and 4 closely related species. Bio12 = annual precipitation, Bio14 = precipitation of driest month, pH = hydrogen ion concentration. *Range (maximum–minimum value), **median, ***coefficient of variation.

Group/Section	Species	Bio12 (mm)	Bio14 (mm)	ET (mm)	BD (kg/dm^3)	pH	BS (%)
Lycopersicon	*S. cheesmaniae*	107–562*	0–15	187–1,125	1.1–1.4	4.3–8.5	31.0–100
		277** (21.3)***	3 (66.6)	45 (41.9)	1.3 (1.1)	7.1 (15.4)	100 (34.5)
	S. galapagense	135–546	0–15	160–1,052	1.1–1.4	4.3–8.5	31.0–100
		274 (16.0)	4 (50)	531 (30.6)	1.3 (0.4)	4.9 (22.4)	32.0 (107.8)
	S. pimpinellifolium	1–2,828	0–143	1–1,710	0.3–1.5	3.2–8.5	10.0–100
		68 (146.3)	0 (0)	43 (195.9)	1.4 (6.5)	7.6 (9.2)	100 (4.5)
Arcanum	*S. arcanum*	22–1,193	0–55	11–1,094	0.3–1.5	4.6–8.5	17.0–100
		487 (44.0)	4 (137.5)	390 (43.2)	1.3 (1.5)	6.4 (13.2)	87.0 (5.1)
	S. chmielewskii	504–1,318	4–19	429–874	1.2–1.5	5.2–8.1	19.0–100
		944 (18.2)	11 (22.7)	610 (16.9)	1.2 (2.5)	8.1 (4.3)	100 (0)
	S. neorickii	426–1,366	3–68	326–1,031	0.2–1.7	4.6–8.1	17.0–100
		817 (21.2)	18.5 (45.9)	672 (13.7)	1.2 (10.2)	5.6 (26.7)	81 (44.0)
Eriopersicon	*S. huaylasense*	128–507	0–3	73–424	1.2–1.4	5.2–7.9	21.0–100
		328 (21.5)	1 (50)	238 (23.0)	1.4 (7.9)	5.6 (12.5)	38.0 (81.5)
	S. corneliomulleri	19–434	0–2	12–354	1.1–1.5	4.2–8.1	19.0–100
		205 (59.2)	0 (0)	141 (45.7)	1.4 (6.7)	5.6 (16.9)	38.0 (86.8)
	S. peruvianum	0–534	0–3	0–427	1.2–1.5	3.2–8.6	10.0–100
		25 (97.4)	0 (0)	13 (134.6)	1.3 (4.1)	7.6 (6.3)	100 (5.0)
	S. chilense	0–355	0–1	3–275	1.0–1.5	4.2–8.6	28.0–100
		29 (68.9)	0 (0)	20 (72.5)	1.3 (2.3)	7.5 (10.6)	100 (6)
	S. habrochaites	11–2,358	0–143	8–1,682	0.3–1.7	4.3–8.5	14.0–100
		605 (42.0)	3 (266.6)	535 (43.9)	1.3 (7.6)	5.7 (15.7)	87 (35.6)
Neolycopersicon	*S. pennellii*	1–404	0–3	0–289	1.2–1.5	5.1–8.5	19.0–100
		49 (94.8)	0 (0)	33 (95.8)	1.3 (4.9)	7.9 (14.5)	100 (31.0)
Juglandifolia	*S. juglandifolium*	555–3,214	1–194	413–1,648	0.3–1.9	4.1–7.7	13.0–100
		1,895 (28.7)	60 (49.1)	1,177 (10.0)	0.9 (1.0)	5.2 (1.9)	23.0 (36.9)
	S. ochranthum	507–2,358	2–131	387–1,474	0.3–1.7	3.2–8.5	10.0–100
		1,010 (11.0)	36 (43)	814 (14.1)	1.2 (11.2)	5.6 (13.3)	45.0 (67.7)
Lycopersicoides	*S. lycopersicoides*	13–215	0–0	9–182	1.3–1.4	4.7–8.1	34.0–100
		104 (50.4)	0 (0)	82 (58.5)	1.3 (1.5)	7.5 (10.6)	100 (6.5)
	S. sitiens	8–31	0–0	9–26	1.2–1.4	6.4–7.9	88.0–100
		17 (24.2)	0 (0)	21 (16.6)	1.2 (8.5)	6.5 (10.7)	88.0 (6.8)

while *S. sitiens* has the greatest climatic restriction, located only in climates BWk (arid, desert, cold). The climate type identified in most of the accession sites was associated with the 16 species was BSk (arid, steppe, and cold), and only absent in species of Lycopersicon group (S. *cheesmaniae*, *S. galapagense*, and *S. pimpinellifolium*) and in *S. juglandifolium* and *S. sitiens*. The opposite case was presented with Cwc climate (temperate, dry winter, and cold summer) present only in some areas where *S. habrochaites* was collected. *S. juglandifolium* and *S. ochranthum* share similar climatic types but were most frequently found in Cfb (temperate, no dry season, warm summer).

Diversity of soil units among wild tomato species (**Figure 7**) found 17 different soil units out of the 23 reported for Latin America in Harmonized World Soil Database from FAO/IIASA/ISRIC/ISSCAS and JRC (2009). Most frequent soils types were leptosol (LP), present in all species, regosol (except *S. huaylasense*) and less frequently acrisol (AC), present only in some *S. pimpinellifolium* accessions.

The greatest edaphic diversity was found in *S. pimpinellifolium*, with accessions in 16 of the 17 reported soil units (except VR). The opposite case was identified for species of Lycopersicoides section *S. sitiens* and *S. lycopersicoides*, with two and four soil units, respectively. Likewise in the patterns of climatic diversity described, edaphic diversity is similar within species, integrating each of the six phylogenetically related groups.

Hot spot Analysis

Areas with a high number of species and accessions were determined by hot spot analysis. **Figure 8** shows the result of hot spot analysis applied with a distance of 1 km between accessions for 4,649 accessions of 12 wild tomato and 4 phylogenetically related species. The highest concentration of species is located in two areas of Peru, one near Trujillo and Chimbote, and the second area around Lima. Likewise, a small area with high diversity is located in southern Peru and northern limit of Chile. The zone in Trujillo–Chimbote is characterized by the presence of seven species (*S. pennellii*, *S. arcanum*, *S. neorickii*, *S. huaylasense*, *S. habrochaites*, *S. pimpinellifolium*, and *S. ochranthum*). The region of high diversity around *Lima* also

TABLE 4 | Ecological descriptors of climatic and edaphic variables associated with the second canonical correlation (9.3%) for 12 species of wild tomato and 4 closely related species. Bio3 = isothermality, Sand = sand percentage, Clay = clay percentage. *Range (maximum–minimum value), **median, ***coefficient of variation.

Group/Section	Species	Alt (m)	Bio3 (°C × 100)	Sand (%)<	$CaCO_3$ (%)	Clay (%)
Lycopersicon	*S. cheesmaniae*	5–1,478*	58.3–74.8	33–60	0–3.1	3–37
		87** (152.4)***	65 (4.5)	43 (11.6)	2 (67.3)	28 (16.0)
	S. galapagense	4–868	59.1–73.9	33–60	0–3.1	3–37
		45 (240.0)	68.3 (4.9)	34 (14.7)	0 (0)	36 (12.5)
	S. pimpinellifolium	1–1,774	48.2–89.4	0–94	0–4.3	0–56
		92 (101.9)	65.6 (8.0)	44 (51.1)	2 (65)	17 (47)
Arcanum	*S. arcanum*	132–3,292	65.5–90.1	0–83	0–4.3	0–45
		1,767 (33.1)	87 (2.7)	54 (10.1)	0 (0)	17 (8.8)
	S. chmielewskii	1,803–3,195	73.6–86.6	25–63	0–2.0	14–31
		2,445 (11.3)	82.6 (2.9)	63 (12.6)	2 (17.5)	14 (14.2)
	S. neorickii	1,202–3,262	76.4–89.1	0–76	0–2.4	0–56
		2,230 (12.2)	84.5 (1.9)	63 (30.1)	0 (0)	12 (20.8)
Eriopersicon	*S. huaylasense*	978–3,304	80.7–90.2	25–67	0–3.5	10–28
		2,301 (18.8)	87.8 (1.1)	67 (31.3)	0 (0)	10 (85)
	S. corneliomulleri	1,018–3,097	64.6–87.4	25–80	0–4.0	4–32
		2,310 (17.0)	75.8 (3.9)	57 (16.6)	0 (0)	16 (43.7)
	S. peruvianum	2–3,191	40.5–87.8	25–94	0–21.6	2–32
		532, (128.0)	62.5 (15.3)	54 (21.2)	2.9 (55.1)	16 (31.2)
	S. chilense	0–3,995	41.6–87.2	30–96	0–21.6	1–32
		1,910 (57.9)	68.1 (10.1)	54 (24.0)	3.1 (59.6)	19 (31.5)
	S. habrochaites	40–3,692	50.1–91.0	0–94	0–4.3	0–46
		2,137 (30.6)	84.5 (5.1)	51 (32.3)	0 (0)	17 (44.1)
Neolycopersicon	*S. pennellii*	5–2,921	48.0–87.4	25–94	0–7.2	2–28
		831 (52.4)	68.5 (8.2)	54 (12)	2.9 (60.3)	16 (18.7)
Juglandifolia	*S. juglandifolium*	1,005–3,153	76.5–94.5	9–94	0–2.0	2–56
		2,195 (14.3)	89.2 (1.9)	40 (25)	0 (0)	13 (46.1)
	S. ochranthum	1,195–4,008	72.4–93.8	0–83	0–2.4	0–46
		2,742 (10.5)	85.8 (3.3)	60 (16.6)	0 (0)	12 (26.0)
Lycopersicoides	*S. lycopersicoides*	1,290–3,775	65.8–85.5	30–57	0–7.2	18–32
		2,960 (13.2)	73.1 (4.6)	50 (11)	3.1 (53.2)	20 (20)
	S. sitiens	2,276–3,330	67.9–71.6	43–69	0.3–7.2	12–28
		2,740 (5.7)	69.1 (1.2)	69 (8.6)	0.3 (1,150)	12 (25)

features seven species: *S. pennellii*, *S. neorickii*, *S. corneliomulleri*, *S. peruvianum*, *S. chilense*, *S. habrochaites*, and *S. pimpinellifolium*.

Finally, the region of high diversity on the border between Chile and Peru is home to five species: *S. pennellii*, *S. peruvianum*, *S. chilense*, *S. pimpinellifolium*, and *S. lycopersicoides*.

Cold spots correspond to the geographical distribution of *S. ochranthum* and *S. juglandifolium* accessions in Colombia and Ecuador, and *S. sitiens* in the northern region of Chile. The rest of the areas of distribution are insignificant according to the statistical criteria, assuming a random distribution.

DISCUSSION

This research provides a relevant ecogeographic characterization to understand the distribution patterns of wild species that complement the phenotypic and genetic information. Characterization of genetic resources through environmental characteristics associated with accession areas and use of GIS tools allows the identification of adaptive ranges and most relevant environmental factors affecting species distribution and ecological adaptation (Parra-Quinajo et al., 2012).

Likewise, through GIS and georeferenced information of species locations, it is possible to quantify geographical distances and distribution patterns of germplasm accession sites. From this perspective, it is likely to determine specific environmental conditions in which wild species and local varieties of crops have acquired their adaptive characters (Hijmans and Spooner, 2001). Therefore, the results obtained in this research constitute a source of updated and valuable information on the edaphoclimatic characteristics in which wild tomatoes and phylogenetically related species are distributed along its natural geographic range.

In general, geographical distribution of 16 wild species related to the cultivated tomato is wide, from Colombia through Peru, comprising Pacific coastal region to Chile and the Andean mountains, with an altitudinal range from sea level to 3,300 m (Peralta et al., 2008; Bergougnoux, 2014. However, within this distribution, there are overlapping areas between several species

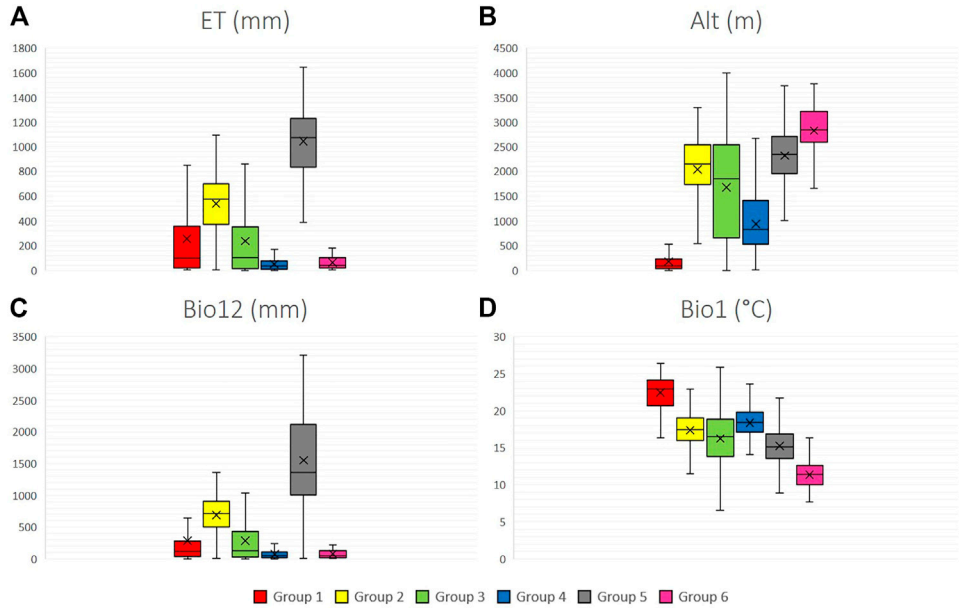

FIGURE 4 | Climate variables boxplots for six groups of 12 wild tomato species and 4 closely related species. **(A)** ET (annual evapotranspiration), **(B)** Alt (altitude), **(C)** Bio12 (annual precipitation), **(D)** Bio1 (mean annual temperature). Group 1: Section Lycopersicon: "Lycopersicon group" (*S. pimpinellifolium*, *S. cheesmaniae*, and *S. galapagense*), Group 2: "Arcanum group" (*S. arcanum*, *S. chmielewskii*, and *S. neorickii*), Group 3: "Eriopersicon group" (*S. habrochaites*, *S. huaylasense*, *S. corneliomulleri*, *S. peruvianum*, and *S. chilense*), Group 4: "Neolycopersicon group" (*S. pennellii*); outgroup close related species in section Juglandifolia Group five (*S. juglandifolium* and *S. ochranthum*) and Group six: Section Lycopersicoides (*S. lycopersicoides* and *S. sitiens*).

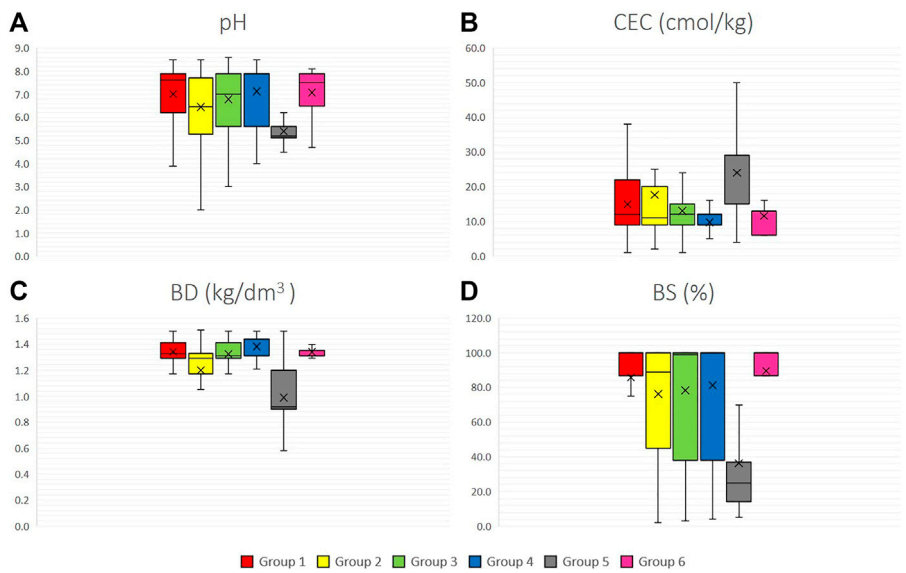

FIGURE 5 | Edaphic variables boxplots for six groups of the 12 wild tomato species and 4 related species **(A)** pH, **(B)** CEC (cation exchange capacity), **(C)** BD (bulk density), **(D)** BS (base saturation). Group 1: Section Lycopersicon: "Lycopersicon group" (*S. pimpinellifolium, S. cheesmaniae* and *S. galapagense*), Group 2: "Arcanum group" (*S. arcanum*, *S. chmielewskii*, and *S. neorickii*), Group 3: "Eriopersicon group" (*S. habrochaites*, *S. huaylasense*, *S. corneliomulleri*, *S. peruvianum*, and *S. chilense*), Group 4: "Neolycopersicon group" (*S. pennellii*); outgroup close related species in Section Juglandifolia Group 5 (*S. juglandifolium* and *S. ochranthum*) and Group 6: Section Lycopersicoides (*S. lycopersicoides* and *S. sitiens*).

or regions with specific distribution such as the endemic species of the Galapagos Islands (*S. cheesmaniae* and *S. galapagense*) or hyper arid regions of northern Chile with other rare endemic species, *S. sitiens*. Within these distribution patterns, it is also possible to identify differences and similarities between the species that conform each group, for example, the similarity

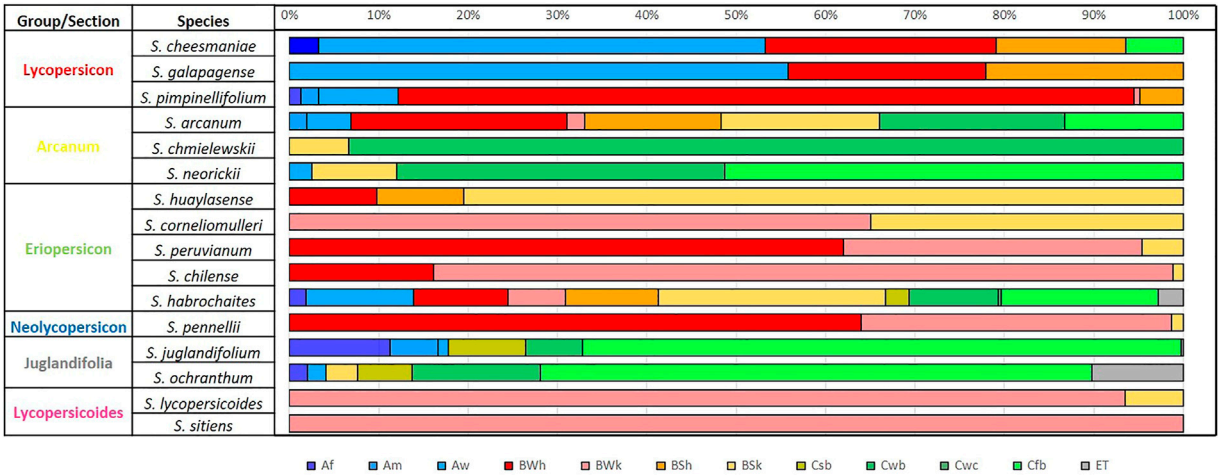

FIGURE 6 | Percentage of climate type by species according to Beck et al. (2018) of 12 wild tomato (*Solanum* Sect. Lycopersicon) and 4 closely related species (*Solanum* Sect. Juglandifolia and Sect. Lycopersicoides). Climate type: Af (tropical, rainforest), Am (tropical, monsoon), Aw (tropical, savannah), BWh (arid, desert, hot), BWk (arid, desert, cold), BSh (arid, steppe, hot), BSk (arid, steppe, cold), Csb (temperate, dry summer, and warm summer), Cwb (temperate, dry winter, and warm summer), Cfb (temperate, no dry season, and warm summer), ET (polar, frost).

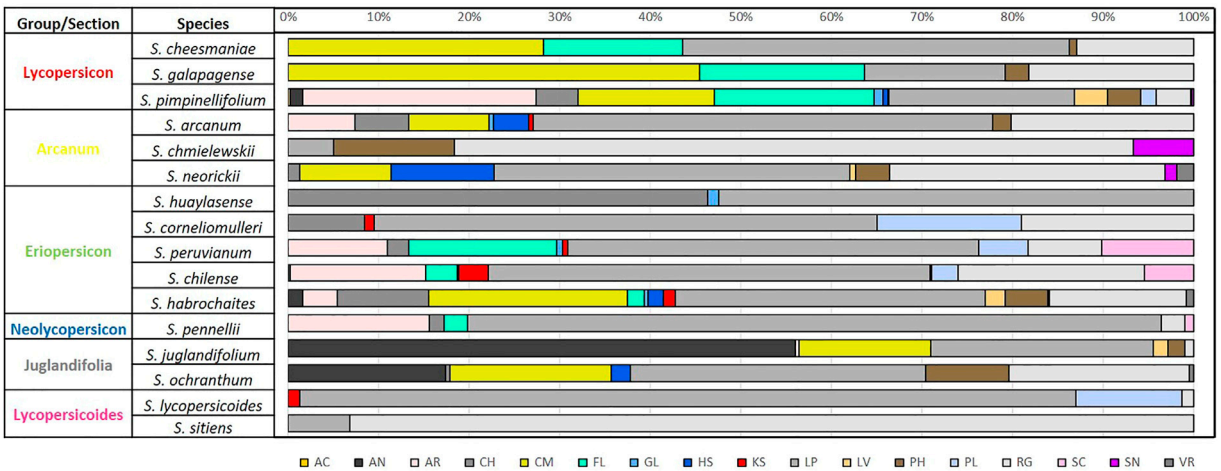

FIGURE 7 | Percentage of soil type by species according to FAO/IIASA/ISRIC/ISSCAS and JRC (2009), for 12 wild tomato (*Solanum* sect. Lycopersicon) and 4 closely related species (*Solanum* sect. Juglandifolia and sect. Lycopersicoides). Soil type: AC (acrisol), AN (andosol), AR (arenosol), CH (chernozem), CM (cambisol), FL (fluvisol), GL (gleysol), HS (histosol), KS (kastanozem), LP (leptosol), LV (luvisol), PH (phaezem), PL (planosol), RG (regosol), SC (solonchak), SN (solonetz), and VR (vertisol).

between *S. arcanum* and the species of Lycopersicon group (**Figure 6**), reflecting a wider distribution and adaptations to local sites of ecotypes (Peralta et al., 2008). These environmental characteristics reflect the ecological adaptation patterns and habitat preference of each species (Nakazato et al., 2010; Vilchez et al., 2019) (**Figures 6**, **7**, Table A1 and A3 in **Supplementary Material**). It is worth mentioning that these results also suggest a thorough revision of the proposed groups, incorporating the new passport data as well as genetic and molecular information to corroborate the belonging of each species to the phylogenetic assigned groups. The aforementioned are under the assumption that the species are closely and genetically related and in expecting that their adaptation areas are similar.

Regarding wild tomato species and phylogenetically related species, few studies have been carried out with an ecogeographic or climatic focus. A comprehensive treatment integrates main botanical, biological, and ecological characteristics of each wild tomato and related species (Peralta et al., 2008); other studies focused on distribution of species richness and diversity through the analysis with GIS (González, 2013) and established conservation priorities (Vilchez et al., 2019); further

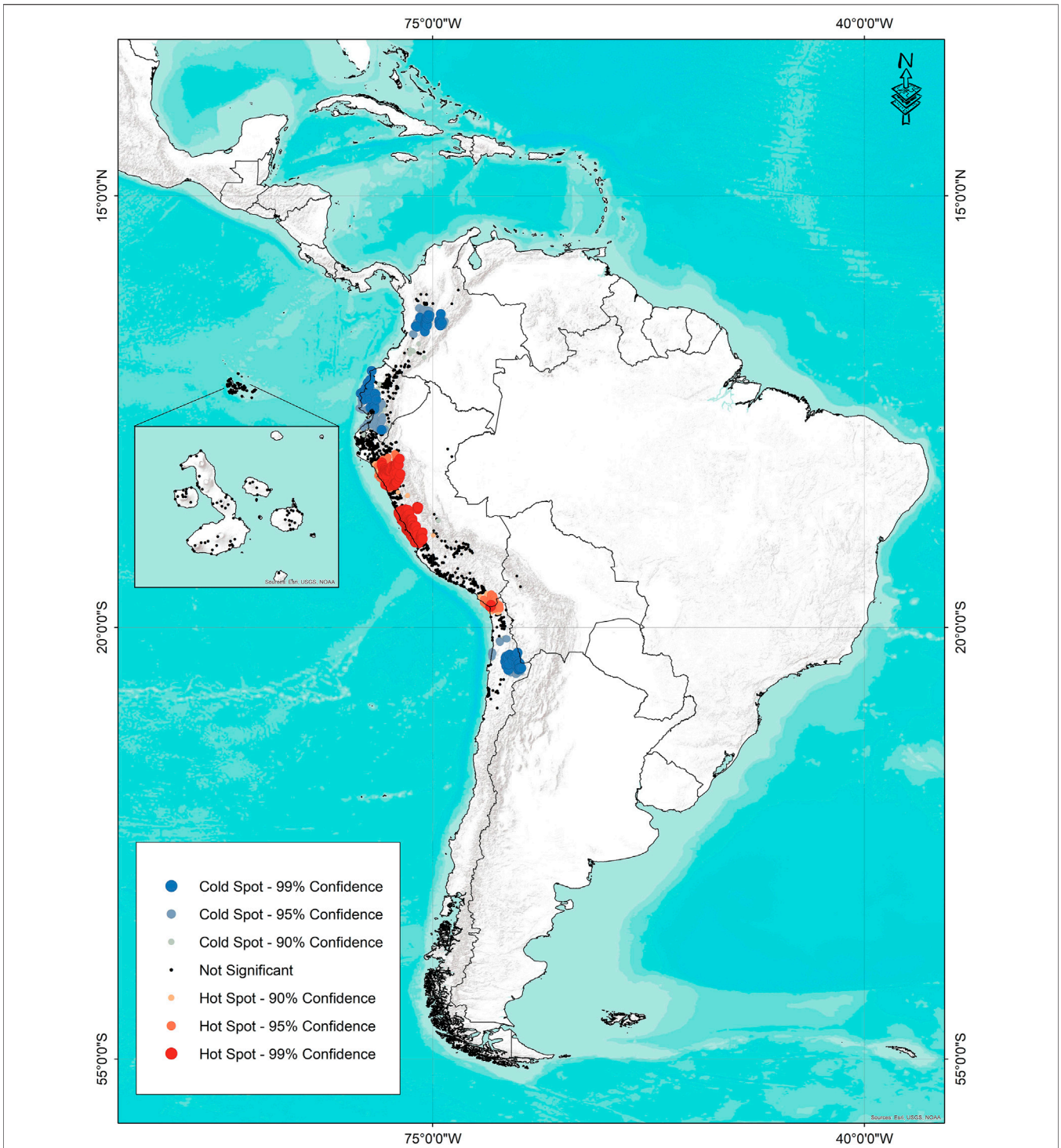

FIGURE 8 | Species diversity map hot spots (red) and cold spots (blue) for 12 wild tomato (*Solanum* Sect. Lycopersicon) and 4 closely related species (*Solanum* sect. Juglandifolia and sect. Lycopersicoides).

geographical and ecological characterization have been investigated in 10 tomato species determining soil and climate variables (Nakazato et al., 2010); studies have been conducted on tomato biogeography, *S. lycopersicum* var. *cerasiforme*, in its center of origin and domestication (Délices et al. (2019); and finally climatic effects on species distribution (Lin et al., 2020) and bioclimatic characterization, and identification of ecological descriptors and patterns of climatic diversity of 12 wild

tomato and 4 closely related species (Ramírez-Ojeda et al., 2021a) have been studied. In this sense, this study complements the information available, providing information on soil characteristics that had not been analyzed.

The canonical correlation analysis satisfactorily identifies climatic variables with greatest influence on edaphic variables and vice versa, with a correlation of 0.80 representing 74% of total variation in 4,649 accessions. One main conclusion is that variables related to water availability (ET, Bio12, Bio14) have a great influence on physical (BD) and chemical soil characteristics (BS, pH). This pattern is persistent in all six groups. This relationship can be better observed in group 5 (*S. juglandifolium* and *S. ochranthum*) accessions with greater availability of annual precipitation and evapotranspiration, which present lower pH, BD, and BS than the rest of species; that is, they are located in soils with the lowest agricultural quality (**Figures 4**, **5**). This methodological approach is promising to be applied at other scales, considering the analysis at population level of each species and climatic and edaphic factors limited to smaller areas of distribution. This basis of ecogeographic characterization could incorporate information from genetic and ecological studies. A better understanding of these variables would allow the generation of projection models in different climate change scenarios (Violle and Jiang, 2009;Luebert and Weigend, 2014; Godoy-Bürki, 2016; Lin et al., 2020).

Ecological descriptors obtained, despite the incorporation of new accessions, are very similar to the ranges reported by Peralta et al. (2008) and Ramírez-Ojeda et al. (2021a) and generally identify the groups of species proposed in the classification. It is important to mention that this methodology has been widely used in the study of other species (Ruiz-Corral et al., 2008; Cerda-Hurtado et al., 2018; Sánchez-González et al., 2018; Ramírez-Ojeda et al., 2021a; 2021b). With this information, it is also possible to identify those species with tolerance to extreme conditions, for example, low and high temperatures, humidity conditions, altitude, pH, BD, and all the possible conditions when associating a species with a climate type or soil unit (**Table 3** and **Table 4**, Tables A1, A2, and A3 in **Supplementary Material**).

Edaphic diversity (**Figure 7**) tends to be more constant between species groups and sections with respect to climate diversity. In general, considering climate and soil characteristics, specific adaptation patterns for each species group can be identified: Lycopersicon group (group 1) corresponds to species with lower altitude and higher mean annual temperature; species of Juglandifolia section (group 5) are those with the highest water availability, lowest pH, BD, and base saturation; species of Lycopersicoides section (group 6) are the ones with the highest altitude, the lowest mean annual temperature, and lowest water availability, groups 4 and 6 have the lowest water availability and soils with favorable agricultural characteristics, differing by altitude. The rest of the species (groups 2 and 3) are in the transition zones with the rest of the wild tomato species. One aspect to highlight is that when combining or considering climatic and edaphic information, it is possible to characterize in a better way the different groups, being able to better identify their differences and similarities.

Among possible uses of this approach is the identification of the germplasm with tolerance to adverse biotic and abiotic factors (Foolad and Lin., 2000; Mittova et al., 2004; Venema et al., 2005; Zhao et al., 2005; Ruiz-Corral et al., 2008; Chetelat et al., 2009; Arellano-Rodríguez et al., 2013; Ruiz-Corral et al., 2013; Cervantes-Moreno et al., 2014; Chen et al., 2015; Nosenko et al., 2016; Stam et al., 2017a; Stam et al., 2017b; Flores-Hernández et al., 2017; Razali et al., 2018; Dinh et al., 2019; Vilchez et al., 2019) with potential use for genetic breeding, identification of routes of germplasm accession, and areas of high and low diversity for use and conservation (Vilchez et al., 2019). In the information contained in **Table 3**, **Table 4**, and **Figures 4**, **5**, it is possible to identify species with extreme values that indicate tolerance or resistance to climatic and edaphic factors, with potential use as germplasm for genetic breeding.

Finally, the hot spot analysis could satisfactorily identify regions with the greatest diversity of species. These are priority areas for conservation, either due to high or low diversity. Regions identified as of great importance for conservation comprise endemism. Diversity contained in populations with few isolated individuals or with restricted distribution could be more affected by environmental and anthropic changes. This result could be explained by the quantity and geographic distance between the accessions of species studied. However, this first approximation is very useful and agrees with the diversity results obtained for wild potato species in Peru (Hijmans and Spooner, 2001).

This research determines the most important edaphoclimatic descriptors of wild tomato species and its closely related species along their natural geographic range in South America. Patterns of climatic diversity correlate with species groups and sections proposed in current classification. New edaphic characteristics analyzed in the same areas were also useful, although with less discrimination than the climatic variables. Interaction between climatic and edaphic factors allows for understanding species distribution and their adaptation patterns. Another feature to highlight is the incorporation of new data from recent collections of specimens being properly identified (Ministerio del Ambiente, 2020) that were not considered before in other studies, and thus expanding precision and reliability of these results. Most important areas for conservation of wild tomato species and related outgroups were detected. Under this premise, this contribution is promissory for further ecogeographic study of wild tomatoes and closely related species at the local population scale, especially focused *in situ* conservation reserves as well as in localities outside protected areas. Edaphoclimatic descriptors in addition with other abiotic or biotic factors could help to better estimate the species ecological niches and determine local ecotypes. Selected descriptors would be tested in models of current and future distribution considering the impact of climate change and anthropic activities along the distribution range of these valuable genetic resources.

Finally, this research can be used as a study model to replicate in other species.

DATA AVAILABILITY STATEMENT

The data sets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Solanacea source: http://solanaceaesource.org/; Tomato Genetic Resource Center: https://tgrc.ucdavis.edu; Global Biodiversity Information Facility: https://www.gbif.org.

AUTHOR CONTRIBUTIONS

Conceptualization: GR-O, IP, and JR-P; methodology: GR-O, IP, and JR-P; software: GR-O and JR-P; validation: JR-P, JS-C, ER-G, IP, JC-S, TM-H, JR-V, and LV-N; formal analysis: GR-O, JR-P, and JS-C; data curation: IP, TM-H, JR-V, and LV-N; writing—original draft preparation: GR-O, JR-P, and IP; writing—review and editing: GR-O, JR-P, JS-C, ER-G, IP, JC-S, TM-H, JR-V, and LV-N. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research received funding from Chapingo Autonomous University through project: D.G.I.P. 20001-C66.

ACKNOWLEDGMENTS

The authors wish to thank all institutions involved in this research: in Mexico: Chapingo Autonomous University, University of Guadalajara, and National Polytechnic Institute; in Peru: Ministry of Environment and National University Pedro Ruiz Gallo; in Argentina: University of Cuyo and National Scientific Technological Center (CONICET). We also want to thank Frontiers in Genetics editorial and support teams for their help and advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.748979/full#supplementary-material

REFERENCES

Aflitos, S., Aflitos, S., Schijlen, E., de Jong, H., de Ridder, D., Smit, S., et al. (2014). Exploring Genetic Variation in the Tomato (SolanumsectionLycopersicon) Clade by Whole-Genome Sequencing. *Plant J.* 80 (1), 136–148. doi:10.1111/tpj.12616

Arellano-Rodríguez, L. J., Rodríguez-Guzmán, E., Ron-Parra, J., Martínez-Ramírez, J. L., Lozoya-Saldaña, H., Sánchez-Martínez, J., et al. (2013). Evaluación de resistencia a *Phytophthora* en poblaciones silvestres de *Solanum lycopersicum* var. *cerasiforme*. *Rev. Mex. Cien. Agríc.* 4, 753–766. doi:10.29312/remexca.v4i5.1173

Balaguera-López, H. E., Álvarez-Herrera, J. G., Martínez-Arévalo, G. E., and Balaguera, W. A. (2009). Clay Content of Soil Influence on Yield of Tomato (*Solanum lycopersicum* L.) Crop. *Rev. Colomb. Cienc. Hort.* 3 (2), 199–209. doi:10.17584/rcch.2009v3i2.1213

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and Future Köppen-Geiger Climate Classification Maps at 1-km Resolution. *Sci. Data* 5, 180214. doi:10.1038/sdata.2018.214

Bergougnoux, V. (2014). The History of Tomato: from Domestication to Biopharming. *Biotechnol. Adv.* 32 (1), 170–189. doi:10.1016/j.biotechadv.2013.11.003

Burbano-Orjuela, H. (2016). El suelo y su relación con los servicios ecosistémicos y la seguridad alimentaria. *rcia* 33 (2), 117–124. doi:10.22267/rcia.163302.58

Cerda-Hurtado, I. M., Mayek-Pérez, N., Hernández-Delgado, S., Muruaga-Martínez, J. S., Reyes-Lara, M. A., Reyes-Valdés, M. H., et al. (2018). Climatic Adaptation and Ecological Descriptors of Wild Beans from Mexico. *Ecol. Evol.* 8 (13), 6492–6504. doi:10.1002/ece3.4106

Cervantes-Moreno, R., Rodríguez-Pérez, J. E., Rodríguez-Pérez, J. E., Carrillo-Fonseca, C., Sahagún-Castellanos, J., and Rodríguez-Guzmán, E. (2014). Tolerance of 26 Native Tomato Collections from Mexico to Nematode *Meloidogyne incognita* (KOFOID and WHITE) Chitwood. *rchsh* XX, 5–18. doi:10.5154/r.rchsh.2012.12.071

Chávez-Servia, J. L., Carrillo-Rodríguez, J. C., Vera-Guzmán, A. M., and Rodríguez-Guzmán, y. R. (2011). *Utilización actual y potencial del jitomate silvestre mexicano Subsistema Nacional de Recursos Fitogenéticos para la Alimentación y la Agricultura (SINAREFI), Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación*. México: CIIDIR-Unidad Oaxaca del Instituto Politécnico Nacional e Instituto Tecnológico del Valle de Oaxaca. Oaxaca.

Chen, H., Chen, X., Chen, D., Li, J., Zhang, Y., and Wang, A. (2015). A Comparison of the Low Temperature Transcriptomes of Two Tomato Genotypes that Differ in Freezing Tolerance: *Solanum lycopersicum* and *Solanum Habrochaites*. *BMC Plant Biol.* 15, 132. doi:10.1186/s12870-015-0521-6

Chetelat, R. T., Pertuzé, R. A., Faúndez, L., Graham, E. B., and Jones, C. M. (2009). Distribution, Ecology and Reproductive Biology of Wild Tomatoes and Related Nightshades from the Atacama Desert Region of Northern Chile. *Euphytica* 167, 77–93. doi:10.1007/s10681-008-9863-6

Córdova-Téllez, L., and Molina-Moreno, J. C. (2006). "Conservación *Ex Situ*," in *Recursos Fitogenéticos de México para la Alimentación y la Agricultura; Informe Nacional sobre el estado de los recursos fitogenéticos para la agricultura y la alimentaciónSecretaria de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación y Sociedad Mexicana de Fitogenética, A.C: Chapingo, México*. Editors J C M. Moreno and L C Téllez.

del Ambiente, M. (2020). *Línea de base de la diversidad del tomate peruano con fines de bioseguridad*. Lima: Perú.

Délices, G., Otto, R., , O. R., Andrés Meza, P., Serna-Lagunez, R., Gamez Pastrana, R., et al. (2019). Biogeografía del tomate Solanum lycopersicum var. cerasiforme (Solanaceae) en su centro de origen (sur de América) y de domesticación (México). *Rbt* 67 (4), 1023–1036. doi:10.15517/rbt.v67i4.33754

Dinh, Q.-D., Dechesne, A., Furrer, H., Taylor, G., Visser, R. G. F., Harbinson, J., et al. (2019). High-Altitude Wild Species Solanum arcanum LA385-A Potential Source for Improvement of Plant Growth and Photosynthetic Performance at Suboptimal Temperatures. *Front. Plant Sci.* 10, 1163. doi:10.3389/fpls.2019.01163

FAO/IIASA/ISRIC/ISSCAS/JRC (2009). *Harmonized World Soil Database 1.1*. Rome, Italy.

Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. *Int. J. Climatol* 37 (12), 4302–4315. doi:10.1002/joc.5086

Flores-Hernández, L. A., Lobato-Ortiz, R., García-Zavala, J. J., Molina-Galán, J. D., Sargerman-Jarquín, D. M., and Velasco-Alvarado, M. d. J. (2017). Parientes Silvestres Del Tomate Como Fuente De Germoplasma Para El Mejoramiento Genético De La Especie. *RevFitotecMex* 40 (1), 83–91. doi:10.35196/rfm.2017.1.83-91

Florido, M., Plana, D., Álvarez, M., Moya, C., Regla, M., and Caballero, A. (2009). Evaluación del germoplasma de tomate (*Solanum* L. sección Lycopersicon) conservado *Ex situ* en Cuba utilizando prueba *In Vitro* para predecir las respuestas a altas temperaturas. *Cult. Trop.* 30 (4), 57–61.

Foolad, M. R., and Lin, G. Y. (2000). Relationship between Cold Tolerance during Seed Germination and Vegetative Growth in Tomato: Germplasm Evaluation. *jashs* 125 (6), 679–683. doi:10.21273/JASHS.125.6.679

Fortuny-Fernández, N. M., Monserrat Ferrer, M., and Ruenes-Morales, M. d. R. (2017). Centros de origen, domesticación y diversidad genética de la ciruela mexicana, Spondias purpurea (Anacardiaceae). *Acta Bot. Mex.* 121, 7–38. doi:10.21829/abm121.2017.1289

GBIF (2021). *Global Biodiversity Information Facility*. . https://www.gbif.org/ (Accessed March 21, 2021).

Getis, A., and Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Anal.* 24 (3), 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x

Godoy-Bürki, A. C. (2016). Efectos del cambio climático sobre especies de plantas vasculares del sur de los Andes Centrales: un estudio en el noroeste de Argentina (NOA). *Ecología Austral* 26 (1), 83–94. doi:10.25260/EA.16.26.2.0.11010.25260/ea.16.26.1.0.110

González, P. (2013). Geographical Distribution of Wild Tomatoes (Solanum L. Sect. Lycopersicon (Mill.) Wettst. Solanaceae). *Arnaldoa* 20 (2), 301–314.

Grandillo, S., Chetelat, R., Knapp, S., Spooner, D., Peralta, I., Cammareri, M., et al. (2011). "Solanum Sect. Lycopersicon," in *Wild Crop Relatives: Genomic and Breeding Resources Vegetables*. Editor K. Chittaranjan (Germany: SpringerBerlin/Heidelberg), 129–215. doi:10.1007/978-3-642-20450-0_9

Hernández-Bautista, A., Lobato-Ortiz, R., Cruz-Izquierdo, S., and García-Zavala, Y. J. L. (2014). Variación fenotípica, heterosis y heredabilidad en una cruza interespecífica de tomate. *Interciencia* 39 (5), 327–332.

Hijmans, R. J., and Spooner, D. M. (2001). Geographic Distribution of Wild Potato Species. *Am. J. Bot.* 88 (11), 2101–2112. doi:10.2307/3558435

Hijmans, R. J., Spooner, D. M., Salas, A. R., Guarino, L., and de la Cruz, J. (2002). *Atlas of Wild Potatoes. Systematic and Ecogeographic Studies on Crop Genepools 10*. Rome, Italy: International Plant Genetic Resources Institute.

FAO (2019). "The State of the World's Biodiversity for Food and Agriculture," in *FAO Commission on Genetic Resources for Food and Agriculture Assessments*. Editors J. Bélanger and D. Pilling (Rome. doi:10.4060/CA3129EN

Kantar, M. B., Sosa, C. C., Khoury, C. K., Castañeda-Álvarez, N. P., Achicanoy, H. A., Bernau, V., et al. (2015). Ecogeography and Utility to Plant Breeding of the Crop Wild Relatives of sunflower (Helianthus Annuus L). *Front. Plant Sci.* 6, 841. doi:10.3389/fpls.2015.00841

Lin, Y.-P., Lu, C.-Y., and Lee, C.-R. (2020). The Climatic Association of Population Divergence and Future Extinction Risk of *Solanum Pimpinellifolium*. *AoB PLANTS* 12, plaa012. doi:10.1093/aobpla/plaa012

Lobo-Burle, M., Torres-Cordeiro, C. M., Fonseca, J. R., de Melo, L. A. M. P., de Belem-das, R. N. A., and Abadie, T. (2013). Characterization of Germplasm According to Environmental Conditions at the Collecting Site Using GIS: Two Case Studies from Brazil. *Plant Genet. Resour. Newsl.* 135, 1–11.

Luebert, F., and Weigend, M. (2014). Phylogenetic Insights into Andean Plant Diversification. *Front. Ecol. Evol.* 2 (27), 1–17. doi:10.3389/fevo.2014.00027

Magallanes-López, A. M., Martínez-Damián, M. T., Sahagún-Castellanos, J., Pérez-Flores, L. J., Marín-Montes, I. M., and Rodríguez-Pérez, J. E. (2020). Calidad Poscosecha De 40 Poblaciones De Tomate (Solanum Lycopersicum L.) Nativas De México. *agro* 54 (6), 779–796. doi:10.47163/agrociencia.v54i6.2184

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial Biogeography: Putting Microorganisms on the Map. *Nat. Rev. Microbiol.* 4 (2), 102–112. doi:10.1038/nrmicro1341

M. Causse, J. Giovannoni, M. Bouzayen, and M. Zouine (Editors) (2016). *The Tomato Genome* (Berlin, Germany: Springer). doi:10.1007/978-3-662-53389-5

Mittova, V., Guy, M., Tal, M., and Volokita, M. (2004). Salinity Up-Regulates the Antioxidative System in Root Mitochondria and Peroxisomes of the Wild Salt-Tolerant Tomato Species *Lycopersicon Pennellii*. *J. Exp. Bot.* 55 (399), 1105–1113. doi:10.1093/jxb/erh113

Nakazato, T., Warren, D. L., and Moyle, L. C. (2010). Ecological and Geographic Modes of Species Divergence in Wild Tomatoes. *Am. J. Bot.* 97 (4), 680–693. doi:10.3732/ajb.0900216

Nosenko, T., Böndel, K. B., Kumpfmüller, G., and Stephan, W. (2016). Adaptation to Low Temperatures in the Wild Tomato speciesSolanum Chilense. *Mol. Ecol.* 25 (12), 2853–2869. doi:10.1111/mec.13637

Parra-Quinajo, M., Iriondo, J. M., and Torres, E. (2012). Review. Applications of Ecogeography and Geographic Information Systems in Conservation and Utilization of Plant Genetic Resources. Span. *J. Agric. Res.* 10 (2), 419–429. doi:10.5424/sjar/2011102-303-11

Pease, J. B., Haak, D. C., Hahn, M. W., and Moyle, L. C. (2016). Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *Plos Biol.* 14 (2), e1002379. doi:10.1371/journal.pbio.1002379

Peralta, I. E., Spooner, D. M., and Knapp, S. (2008). Taxonomy of Wild Tomatoes and Their Relatives (*Solanum* Sect. Lycopersicoides, Sect. Juglandifolia, Sect. Lycopersicon; Solanaceae). *Syst. Bot. Monogr.* 84, 1–186.

Ramírez-Ojeda, G., Orozco-Gutiérrez, G., RuizSanchez, E., and Ruiz-Sánchez, E. (2021b). EDAPHOCLIMATIC DIVERSITY and ECOLOGICAL DESCRIPTORS of Guadua BAMBOO SPECIES (Poaceae: Bambusoideae) IN MEXICO. *Ijaeb* 06 (3), 228–244. doi:10.35410/IJAEB.2021.5641

Ramírez-Ojeda, G., Peralta, I. E., Rodríguez-Guzmán, E., Chávez-Servia, J. L., Sahagún-Castellanos, J., and Rodríguez-Pérez, J. E. (2021a). Climatic Diversity and Ecological Descriptors of Wild Tomato Species (Solanum Sect. Lycopersicon) and Close Related Species (Solanum Sect. Juglandifolia Y Sect. Lycopersicoides) in Latin America. *Plants* 10, 855. doi:10.3390/plants10050855

Razali, R., Bougouffa, S., Morton, M. J. L., Lightfoot, D. J., Alam, I., Essack, M., et al. (2018). The Genome Sequence of the Wild Tomato Solanum Pimpinellifolium Provides Insights into Salinity Tolerance. *Front. Plant Sci.* 9 (1402), 1–21. doi:10.3389/fpls.2018.01402

Razifard, H., Ramos, A., Della Valle, A. L., Bodary, C., Goetz, E., Manser, E. J., et al. (2020). Genomic Evidence for Complex Domestication History of the Cultivated Tomato in Latin America. *Mol. Biol. Evol.* 37 (4), 1118–1132. doi:10.1093/molbev/msz297

Rodríguez, F., Wu, F., Ané, C., Tanksley, S., and Spooner, D. M. (2009). Do potatoes and Tomatoes Have a Single Evolutionary History, and what Proportion of the Genome Supports This History? *BMC Evol. Biol.* 9 (9), 191. doi:10.1186/1471-2148-9-191

Ruiz Corral, J. A., Durán Puga, N., Sánchez González, J. d. J., Ron Parra, J., González Eguiarte, D. R., Holland, J. B., et al. (2008). Climatic Adaptation and Ecological Descriptors of 42 Mexican Maize Races. *Crop Sci.* 48 (4), 1502–1512. doi:10.2135/cropsci2007.09.0518

Ruiz-Corral, J. A., Sánchez-González, J. J., Hernández-Casillas, J. M., Willcox, M. C., Ramírez-Ojeda, G., Ramírez-Díaz, J. L., et al. (2013). Identificación de razas de maíz adaptadas a condiciones deficientes de humedad mediante datos biogeográficos. *Rev. Mex. Cienc. Agric.* 4 (6), 829–842. doi:10.29312/remexca.v4i6.1152

Sánchez González, J. d. J., Ruiz Corral, J. A., García, G. M., Ojeda, G. R., Larios, L. D. l. C., Holland, J. B., et al. (2018). Ecogeography of Teosinte. *PLoS ONE* 13, e0192676.. doi:10.1371/journal.pone.019267

Sandoval-Ceballos, M. G., Kalungwana, N. A., Griffin, J. H. C., Martínez-Guerra, G., Ramírez-Ramírez, I., Maldonado-Peralta, R., et al. (2021). The Importance of Conserving Mexico's Tomato Agrodiversity to Research Plant Biochemistry under Different Climates. *Plants People Planet.*, 1–7. doi:10.1002/ppp3.10218

SAS Institute (2011). *SAS/STAT User's Guide: Software Version 9.3*. Cary, NC, USA: Statistical Analysis System Institute.

Solanaceae Source (2021). A Global Taxonomic Source for the Nightshade Family. Available at: http://solanaceaesource.org/(Accessed March 18, 2021).

Sotomayor, D., Vilchez, D., and Zorrilla, C. (2019). Dataset on Occurrences of *Lycopersicum* Species from Peru. *Figshare. Dataset.* doi:10.6084/m9.figshare.9880073.v1

Spooner, D. M., Gavrilenko, T., Jansky, S. H., Ovchinnikova, A., Krylova, E., Knapp, S., et al. (2010). Ecogeography of Ploidy Variation in Cultivated Potato (Solanum Sect. Petota). *Am. J. Bot.* 97 (12), 2049–2060. doi:10.3732/ajb.1000277

Stam, R., Scheikl, D., and Tellier, A. (2017b). The Wild Tomato speciesSolanum Chilenseshows Variation in Pathogen Resistance between Geographically Distinct Populations. *PeerJ* 5 (5), e2910. doi:10.7717/peerj.2910

Stam, R., Silva-Arias, G. A., Tellier, A., Scheikl, D., Hörger, A., Stephan, W., et al. (2017a). Subsets of NLR Genes Drive Adaptation of Tomato to Pathogens during Colonisation of New Habitats. *bioRxiv*, 210559. doi:10.1101/210559

Steiner, J. J., and Greene, S. L. (1996). Proposed Ecological Descriptors and Their Utility for Plant Germplasm Collections. *Crop Sci.* 36 (2), 439–451. doi:10.2135/cropsci1996.0011183X003600020037x

TGRC (2021). *Tomato Genetics Resource Center*. Available at: https://tgrc.ucdavis.edu/(Accessed March 18, 2021).

Tofalo, R., Perpetuini, G., Schirone, M., Fasoli, G., Aguzzi, I., Corsetti, A., et al. (2013). Biogeographical Characterization of *Saccharomyces cerevisiae* Wine Yeast by Molecular Methods. *Front. Microbiol.* 4 (166), 1–13. doi:10.3389/fmicb.2013.00166

Trabucco, A., and Zomer, R. J. (2019). Global High-Resolution Soil-Water Balance. *Figshare. Fileset*. doi:10.6084/m9.figshare.7707605.v3

Tropicos.org. Missouri Botanical Garden (2021). Available at: https://tropicos.org/home. (Accessed March 17, 2021).

Unep-Wcmc (2016). *El estado de la biodiversidad en América Latina y el Caribe*. Cambridge, Reino Unido: UNEP-WCMC.

Venema, J. H., Linger, P., Heusden, A. W., Hasselt, P. R., and Brüggemann, W. (2005). The Inheritance of Chilling Tolerance in Tomato (Lycopersiconspp). *Plant Biol.* 7 (2), 118–130. doi:10.1055/s-2005-837495

Vilchez, D., Sotomayor, D. A., , D., and Zorrilla, C. (2019). *Ex Situ* Conservation Priorities for the Peruvian Wild Tomato Species (*Solanum* L. Sect. Lycopersicum (Mill.) Wettst). *Ecol. Apl.* 18 (2), 171–183. doi:10.21704/rea.v18i2.1335

Violle, C., and Jiang, L. (2009). Towards a Trait-Based Quantification of Species Niche. *J. Plant Ecol.* 2 (2), 87–93. doi:10.1093/jpe/rtp007

Zhao, L., Qiu, C., Li, J., Chai, Y., Kai, G., Li, Z., et al. (2005). Investigation of Disease Resistance and Cold Tolerance of *Solanum Lycopersicoides* for Tomato Improvement. *HortSci* 40 (1), 43–46. doi:10.21273/HORTSCI.40.1.43

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 December 2021
doi: 10.3389/fgene.2021.780793

S-RNase Alleles Associated With Self-Compatibility in the Tomato Clade: Structure, Origins, and Expression Plasticity

*Amanda K. Broz[1†], Christopher M. Miller[1†], You Soon Baek[1†], Alejandro Tovar-Méndez[2], Pablo Geovanny Acosta-Quezada[3], Tanya Elizabet Riofrío-Cuenca[3], Douglas B. Rusch[4] and Patricia A. Bedinger[1]**

[1]Department of Biology, Colorado State University, Fort Collins, CO, United States, [2]Department of Biochemistry, University of Missouri, Columbia, MO, United States, [3]Departamento de Ciencias Biológicas y Agropecuarias, Universidad Técnica Particular de Loja, Loja, Ecuador, [4]Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN, United States

OPEN ACCESS

Edited by:
Peter Poczai,
University of Helsinki, Finland

Reviewed by:
Edeline Gagnon,
Royal Botanic Garden Edinburgh, United Kingdom
Alice Y. Cheung,
University of Massachusetts Amherst, United States

***Correspondence:**
Patricia A. Bedinger
Patricia.Bedinger@colostate.edu

†These authors have contributed equally to this work

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

Received: 21 September 2021
Accepted: 09 November 2021
Published: 06 December 2021

Citation:
Broz AK, Miller CM, Baek YS, Tovar-Méndez A, Acosta-Quezada PG, Riofrío-Cuenca TE, Rusch DB and Bedinger PA (2021) S-RNase Alleles Associated With Self-Compatibility in the Tomato Clade: Structure, Origins, and Expression Plasticity. Front. Genet. 12:780793. doi: 10.3389/fgene.2021.780793

The self-incompatibility (SI) system in the Solanaceae is comprised of cytotoxic pistil S-RNases which are countered by S-locus F-box (SLF) resistance factors found in pollen. Under this barrier-resistance architecture, mating system transitions from SI to self-compatibility (SC) typically result from loss-of-function mutations in genes encoding pistil SI factors such as *S-RNase*. However, the nature of these mutations is often not well characterized. Here we use a combination of *S-RNase* sequence analysis, transcript profiling, protein expression and reproductive phenotyping to better understand different mechanisms that result in loss of S-RNase function. Our analysis focuses on 12 *S-RNase* alleles identified in SC species and populations across the tomato clade. In six cases, the reason for gene dysfunction due to mutations is evident. The six other alleles potentially encode functional S-RNase proteins but are typically transcriptionally silenced. We identified three *S-RNase* alleles which are transcriptionally silenced under some conditions but actively expressed in others. In one case, expression of the *S-RNase* is associated with SI. In another case, *S-RNase* expression does not lead to SI, but instead confers a reproductive barrier against pollen tubes from other tomato species. In the third case, expression of *S-RNase* does not affect self, interspecific or inter-population reproductive barriers. Our results indicate that *S-RNase* expression is more dynamic than previously thought, and that changes in expression can impact different reproductive barriers within or between natural populations.

Keywords: self-incompatibility (incompatible), self-compatibility (compatible), mating system transitions, S-RNase, reproductive barriers

INTRODUCTION

Self-incompatibility (SI) is a genetic mechanism that prevents self-fertilization in numerous plant species, usually by preventing "self" pollen tube germination on stigmas or self-pollen tube growth in styles (De Nettancourt, 1977; Takayama and Isogai, 2005; Fujii et al., 2016). Although relatively few SI systems have been extensively studied, all examined to date contain a complex *S*-locus which encodes both pistil- and pollen-expressed genes that are polymorphic within populations and act to

regulate the specificity of SI (Franklin-Tong, 2008; Fujii et al., 2016; Bedinger et al., 2017; Jany et al., 2019; Nasrallah, 2019). The *S*-locus in the Solanaceae contains a single gene encoding a pistil-expressed *S*-locus RNase (S-RNase) and 15–20 genes encoding pollen-expressed *S*-locus F-box proteins (SLFs) (McClure, 2004; Kubo et al., 2015; Li and Chetelat, 2015; Williams et al., 2015; Wu et al., 2020). Each specific combination of *S-RNase* and *SLF* genes at an *S*-locus constitutes a unique *S*-haplotype, and successful mating only occurs between plants with different *S*-haplotypes.

The SI mechanism operating in the Solanaceae is gametophytic, since it depends on post-meiotic pollen-expressed genes, and can be thought of in terms of a barrier-resistance architecture comprised of pistil-side cytotoxic S-RNases (the barriers) and pollen SLFs that act as resistance factors (Bedinger et al., 2017). In styles, S-RNases are secreted into the transmitting tissue and are taken up by growing pollen tubes. Active S-RNases degrade pollen tube RNA, resulting in pollen tube death, unless they are recognized and detoxified by SLF proteins. Under the non-self-recognition model, the constellation of SLFs produced in pollen tubes of each *S*-haplotype can detoxify all S-RNases except the one encoded by their own haplotype (Kubo et al., 2010). Phylogenetic evidence suggests that SI is the ancestral state in the Solanaceae (Allen and Hiscock, 2008; Igic et al., 2008), and there is frequently high conservation in *S-RNase* allele sequences between species (Ramanauskas and Igić, 2017).

Although SI systems enforce outcrossing and thus maintain genetic diversity within populations, mating system transitions from outcrossing to selfing are common evolutionary events (Darwin, 1876; Stebbins, 1974; Coyne and Orr, 2004; Igić et al., 2008; Wright et al., 2013), especially in circumstances where selfing individuals have a reproductive advantage (Baker, 1955; Baker, 1967; Busch and Schoen, 2008; Pannell et al., 2015). Given the barrier-resistance architecture of SI in the Solanaceae, female-side loss-of-function (pistil first) mutations that lead to self-compatibility (SC) are predicted to be more common than male-side gain-of-function mutations (Bedinger et al., 2017). For example, loss of S-RNase function would eliminate the barrier to self-pollen tube growth. These loss-of-function *S-RNase* mutations are codominant in the sense that plants heterozygous for the mutation will exhibit the SC phenotype and only the haplotype with the non-functional *S-RNase* gene will be transmitted in self-pollinations if the suite of pollen SLFs are intact. In this scenario, self-pollen tubes containing the mutant (SC) *S-RNase* haplotype will be successful in self-pollinations, because all non-self S-RNases will be detoxified, whereas pollen tubes harboring a functional SI haplotype will be destroyed by their self S-RNases. In addition, pollen with the SC haplotype can also be successful in outcross pollinations. These characteristics allow SC to rapidly spread to fixation within a population unless the SC phenotype is countered by other detrimental phenotypes (i.e., pollen/seed discounting; inbreeding depression) (Porcher and Lande, 2005; Busch and Delph, 2012). Alternatively, because of the non-self mode of recognition in S-RNase-based SI, male-side loss-of-function mutations in *SLFs* would not result in SC and could make pollen tubes vulnerable to non-self S-RNases. However, *SLF* gain-of-function mutations in that allow for detoxification of a self S-RNase could result in SC, and there is some evidence these types of mutations can occur at low frequency (Tsukamoto et al., 2003b; Kubo et al., 2015; Markova et al., 2017).

In systems of S-RNase-based SI, there is both functional and mechanistic overlap between SI and pollen-pistil incompatibilities in crosses between species that result in interspecific reproductive barriers (IRBs). Both S-RNase and SLFs have been found to play a role in unilateral IRBs, known as unilateral incompatibility (UI, wherein a cross is incompatible in one direction but the reciprocal cross is compatible) and mutations in the genes encoding these SI factors can alter both interspecific and inter-population reproductive barriers (Tovar-Méndez et al., 2014; Li and Chetelat, 2015; Markova et al., 2016; Broz et al., 2017). SI modifier genes, that are not located at the *S*-locus, have also been implicated in both SI and UI. For example, CUL1 is a pollen-expressed factor that is involved in both SI and UI (Li et al., 2010; Li and Chetelat, 2014). The pistil-expressed modifier HT-proteins are required for SI, and contribute to IRBs (Tovar-Méndez et al., 2014; Tovar-Mendez et al., 2017). However, it is important to note that IRBs can be produced by alternative mechanisms (S-RNase-dependent and S-RNase-independent), and recent studies have identified pistil and pollen factors that contribute to S-RNase-independent IRBs (Qin et al., 2018; Qin and Chetelat, 2021). Because mechanisms of SI and IRBs are only partially redundant, it is not possible to predict how mutation of a particular SI factor will affect interspecific pollen tube growth.

The 13-member tomato clade, *Solanum* section Lycopersicon, is particularly amenable to studying mating system shifts from SI to SC, as multiple independent transitions from SI to SC have occurred both in entire species and within populations of SI species. Six of the 13 tomato species (*S. lycopersicum, S. pimpinellifolium, S. galapagense, S. cheesemaniae, S. neorickii* and *S. chmielewskii*) are fully SC, and four predominately SI species (*S. pennellii, S. arcanum*, *S. habrochaites* and *S. peruvianum*) contain one or more SC populations. *S. chilense* has two segregating SI/SC populations (www.tgrc.ucdavis.edu), and the remaining two species (*S. corneliomulleri* and *S. huaylasense*) are fully SI.

In predominately SI wild tomato species, transitions to SC typically occur at species range margins. For example, the migration of *S. habrochaites* northward through the Amotape-Huancabamba Zone, which consists of microhabitats with widely varying altitudes and temperatures (Weigend, 2002; Weigend, 2004), provides a particularly striking example of multiple independent mating system transitions associated with migration and population differentiation (Landis et al., 2021). This is likely because the ability of a plant to reproduce through self-pollination can provide reproductive assurance to small locally adapted populations colonizing new environments (Baker, 1955, Baker, 1967; Pannell and Barrett, 1998; Pannell et al., 2015).

In many plant species, mating system transitions to SC are associated with changes in floral morphology, often referred to as the "selfing syndrome" (Wright et al., 2013). One prominent

phenotype associated with selfing syndrome is reduced flower size, which can evolve when the need for pollinator attraction has been abrogated due to high rates of self-pollination. In the tomato clade, the SC species *S. neorickii* exhibits extremely small flowers and is considered to be autogamous (Rick et al., 1976). Differences in both corolla diameter and stigma exsertion have also been identified between SI and SC populations of *S. habrochaites* (Rick et al., 1979; Broz et al., 2017); although SC populations have not been exhaustively examined.

In the tomato clade, transitions to SC can also be associated with changes in IRBs (Covey et al., 2010; Bedinger et al., 2011; Baek et al., 2015; Broz et al., 2017). In general, UI between tomato clade species follows the SI x SC rule wherein SI species reject pollen tubes of SC species, but the reciprocal cross is compatible, resulting in UI (Baek et al., 2015). However, there are exceptions, particularly in SC populations of typically SI species. For example, an SC population of the typically SI species *S. arcanum* shows decreases in pistil-side IRBs compared to its SI relatives, allowing interspecific pollen tubes to penetrate substantially further into the style (Baek et al., 2015). Self-compatible populations of predominately SI *S. habrochaites* can also show weakened pistil-side IRBs, some of which are associated with the loss of specific pistil-side proteins including S-RNase and HT-protein (Broz et al., 2017).

Here, using a combination of transcriptomics, degenerate PCR amplification, phenotyping and analysis of published sequence data, we characterized *S-RNase* alleles associated with SC across the tomato clade (**Table 1**). The main objectives of this work were to 1) provide a comprehensive survey of newly discovered and previously identified *S-RNase* alleles that are associated with SC in the tomato clade, 2) evaluate RNA and protein expression of SC-associated *S-RNase* alleles that have no apparent sequence defect, 3) identify putative progenitor (functional) *S-RNase* alleles in SI populations and species and to 4) better understand how SC-associated *S-RNase* alleles affect IRBs. We show that, in most cases, the transition to SC is associated with *S-RNase* mutations that either prevent S-RNase production, reduce S-RNase protein activity, or involve the transcriptional silencing of potentially functional *S-RNase* genes. We find that in some but not all cases, these *S-RNase* mutations affect IRBs in addition to mating system.

MATERIALS AND METHODS

Plant Material and Growth

Seeds were acquired from the C.M. Rick Tomato Genetic Resource Center (TGRC) at University of California, Davis (www.tgrc.ucdavis.edu) or collected in Loja Province in Ecuador (denoted as EC collections) in 2014. Representative accessions for all species and populations are listed in **Table 1** and refer to the populations used in our study. Details on additional populations used for study of *S. neorickii* can be found in the Section *S-RNase alleles LpfSRN-1 and LpfSRN-2 in SC S. neorickii*, and those for *S. habrochaites* are provided in the Section *SC accessions in S. habrochaites*. All EC collections, excepting EC40, have representative collections at TGRC (EC6˜LA2101, EC7˜LA2864 and EC10˜2099) and were verified to exist at the same sites in this study. Seed collections of EC populations are housed at the Departamento de Ciencias Biológicas y Agropecuarias, Universidad Técnica Particular de Loja, Loja, Ecuador. Seeds were sterilized according to recommendations from TGRC. For genotyping, seeds were planted in ProMix-HP and grown on a light shelf for 2 weeks. For experiments to assess pollen tube growth or to produce seeds, plants were grown in 4-inch pots containing ProMix-BX under greenhouse conditions (16 h light at 26°C and 8 h dark at 18°C) until they were 6–12 inches tall, then transplanted to outdoor agricultural fields at Colorado State University or placed in a growth chamber (10 h days) as needed to induce flowering. After performing crosses to obtain specific progeny, fruits were allowed to mature on plants for at least 2 months (or until soft and ripe).

Pollen Tube Growth Assessment and Reproductive Barrier Phenotyping

In a previously uncharacterized accession of SC *S. habrochaites* (LA2863), and for all *S. neorickii* accessions including F1 and F2 cross types (see *S-RNase Alleles LpfSRN-1 and LpfSRN-2 in SC S. neorickii* Section) we performed reproductive phenotyping. Pollen tube growth in styles was assessed as previously described (Covey et al., 2010). Briefly, emasculated flowers were pollinated, and after 48 h pistils were placed in fixative, softened with NaOH, stained using Aniline Blue Fluorochrome and examined with a fluorescence microscope. In field grown plants, inflorescences were covered with mesh bags to prevent pollinators from interacting with flowers to be used in crosses. Interspecific and inter-population barriers were examined using "tester" lines, as described more thoroughly in Broz et al., 2017. Briefly, to test for IRBs, pistils were pollinated using *S. lycopersicum* cultivars VF36, M82 or LA1221 as males. To test for pistil-side inter-population reproductive barriers in *S. habrochaites*, hand pollinations were performed using *S. habrochaites* SC accession LA0407 as male, and to test for pollen-side inter-population reproductive barriers, hand pollinations were performed using SI accession LA1777 as female.

Stylar Transcriptome Sequencing and Analysis

Transcriptome sequencing was utilized to identify *S-RNase* alleles in SC (LA2119, LA2863), mixed SI/SC (LA 2099, LA 2098, LA2175) and SI (LA2868, LA2864) accessions of *S. habrochaites*. Unpollinated styles from three individuals of each accession were separately collected into RNAlater solution (Qiagen), and total RNA was extracted using the Qiagen RNeasy Plant Mini Kit. Total RNA was submitted to Indiana University's Center for Genomics and Bioinformatics for cDNA library construction using a TruSeq Stranded mRNA LT Sample Prep Kit (Illumina) following the standard manufacturing protocol. In some cases, RNA from individuals within an accession were pooled, and sequencing of the unfragmented whole transcriptome libraries was performed on an Illumina MiSeq instrument to generate 250bp paired end reads. In all

TABLE 1 | Comprehensive list of SC species and populations that have been identified in the tomato clade and their associated *S-RNase* alleles. Representative accessions are listed for SC species and for groups of SC *S. habrochaites* populations.[a]Li and Chetelat, 2015; [b]This work; [c]Broz et al., 2017; [d]Kondo et al., 2002b; [e]Kondo et al., 2002a; [f]Covey et al., 2010; [g]Royo et al., 1994a; [h]Kowyama et al., 1994; [i]Landis et al., 2021; [j]Broz et al., 2021; [k]Markova et al., 2017; [l]HT-protein expression was confirmed with an antibody that binds to both HT-A and HT-B; [m]amino acid identity (id) or similarity (sim) of available sequences; *inferred from allele testing in *S. chilense* in Igić et al., 2007 or in *S. peruvianum* in Miller and Kostyun, 2011; ** *S. arcanum LpSC* and *S. chmielewskii LcwSRN-1* are 99.3% identical (**Supplementary Figure S2**); ***20 amino acids available for alignment prior to frame-shift mutation; # 96 amino acids available for alignment prior to nonsense mutation; NT = not tested, NA = not applicable.

SC species	*S-RNase* allele	GenBank	Represent-ative accession	Mutation or expression defect	RNA Y/N	Protein Y/N	*S*-locus Y/NT	Related *S. chilense* allele	Related functional S-RNase (% aa id/sim)	HTA/ HTB
S. lycopersicum	*SRN-red*[a,b]	AC246123.1, XM004229015	Tomato cultivars	Silenced[b]	N[b]	N[d]	Y[a]	*S20*	*S. chilense* S20 (95.5/97.7)	N[d]/N[d]
S. pimpinellifolium	*SRN-red*	KJ814947.1	LA1589	NT	NT	N[b]	NT	*S20*	*S. chilense* S20 (95.5/97.7)	NT/NT
S. galapagense	*SRN-orange*[b]	OK091157	LA0317	NT	NT	N[b]	NT	*S20*	*S. chilense* S20 (96.3/98.5)	NT/NT
S. cheesmanieae	*SRN-orange*[b]	OK091158	LA0522	NT	NT	N[b]	NT	*S20*	*S. chilense* S20 (96.3/98.5)	NT/NT
S. chmielewskii	*LcwSRN-1*[e]	AB072477.1	LA1316	Silenced[e]	N[e]	N[e]	Y*	*S11*	*S. chilense* S11 (100/100)**	Y[e]/N[e]
S. neorickii	*LpfSRN-1*[e]	AB072475.1	LA1322	Varies: Silenced or low RNase activity[e,b]	Y[e]/N[b]	Y[e]/N[b]	Y[b]	*S1*	*S. peruvianum* SP2 (96.7/98.3)	Y[be]/N[e]
S. neorickii	*LpfSRN-2*[e]	AB072476.1	LA0247	Frame-shift[e]	N[e]	N[e]	Y[b]	*S7*	*S. arcanum* S6 (95/95***)	Y[bl]
SC Populations										
S. pennellii	NA	NA	LA0716	Deletion[a]	NA	NA	Y[a]	NA	NA	Y[f]/Y[f]
S. arcanum	*LpSc*[g]	Z26581.1	LA2157	Missense, lacks active site histidine[g]	Y[g,h]	Y[g,h]	Y[h]	*S11*	*S. chilense* S11 (99.3/100)**	NT/NT
S. habrochaites SC-1	*hab-7*[i]	OK091159	LA2119	Silenced in SC-1 group with exceptions[b]	N[i]/Y[b]	N[i]/Y[b]	Y[b]	*S32*	*S. peruvianum* S13 (99.5/100)	Y[b]/N[f]
S. habrochaites SC-2	*LhgSRN-1*[e]	AB072478.1	LA0407	Silenced in SC-2 group[b,c,e,f]	N[e,f]	N[c]	Y[b]	*S6*	*S. habrochaites hab-16* (99.5/100)	Y[f]/N[f]
S. habrochaites SC-4	*hab-6*[f,j]	MW183811.1	LA1927	Missense, low RNase activity[f,j]	Y[f,j]	Y[f,j]	Y[j]	*S2*	*S. peruvianum* SP11 (98.3/99.2)	Y[f]/N[f]
S. habrochaites SC-5	*hab-8*[b]	OK091160	LA2101	Nonsense[b]	NT	N[c]	Y*	*S15*	*S. habrochaites hab-14*[#] (100/100)	Y[b]/N[f]
S. habrochaites SC-6	unknown	NA	LA4654	Unknown	NT	N[i]	NT	NA	NA	Y[i]/N[f]
S. habrochaites SC-7	*hab-12*[b]	OK091161	LA2863	Missense, lacks N-gylcosylation sites[b]	Y[b]	Y[b]	Y*	*S18*	*S. habrochaites hab-13* (99.5/100)	Y[b]/N[f]
S. peruvianum	unknown[k]	NA	LA4125	Unknown	NT	NT	NT	NA	NA	NT/NT

other cases, sequencing was performed using an Illumina NextSeq500 platform with 150 bp cycle module generating 60 bp paired-end reads. After the sequencing run, demultiplexing was performed with bcl2fastq v2.20.0.422. The raw transcriptome data are available on the NCBI SRA database PRJNA310635. Details on data processing, analysis and identification of *S-RNase* sequences are described in Broz et al., 2021. Transcriptome analysis led to the identification of new *S-RNase* alleles *hab-7*, *hab-12*, *hab-13*, *hab-14*, *hab-15*, *hab-16* and *hab-17* (GenBank numbers OK091159, OK091161- OK091166).

Degenerate PCR to Isolate *hab-8 S-RNase* Allele

A PCR-based strategy devised by Kondo *et al.* (2002a) was used for the isolation of the *S-RNase* allele from *S. habrochaites* accession LA2101. Briefly, we amplified unknown *S-RNase* sequences from the genomic DNA using degenerate primers based on conserved *S-RNase* sequences (Covey et al., 2010) and appropriately sized products were gel purified (Qiagen) and ligated to pJET1.2 (ThermoFisher). Colony PCR was performed, and the resulting PCR products were purified (Zymo) and sequenced (Genewiz). The sequence identified in LA2101 is *hab-8*, GenBank OK091160.

PCR Amplification and Sequence Analysis

We used PCR amplification and sequencing to obtain the *S-RNase* alleles for *S. galapagense* LA0317 and *S. cheesmaniae* LA0522, to verify alleles from *S. habrochaites* that were identified by transcriptome analysis (see *Stylar Transcriptome Sequencing and Analysis* Section), and to verify previously identified alleles from *S. neorickii.* Genomic DNA was extracted from leaf tissue of seedlings in 200 mM Tris-HCl pH 9.0, 250 mM NaCl, 25 mM EDTA, and 1% SDS, followed by precipitation in isopropanol. All primers are listed in **Supplementary Table S1**, including primers designed to amplify specific *S-RNase* alleles. PCR was performed using EconoTaq Plus Green Mastermix (Lucigen). Genomic DNA quality was assessed by amplifying single copy control genes (**Supplementary Table S1**). For genotyping, PCR products were analyzed on 1.2% agarose gels. For sequencing, PCR products were purified (Zymo) and both strands of amplicons were sequenced (GeneWiz). Genomic DNA and deduced amino acid sequences were aligned using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/ (Madeira et al., 2019). Signal peptide predictions were made using TargetP http://www.cbs.dtu.dk/services/TargetP/ and N-glycosylation site predictions were made using NetNGlyc 1.0 http://www.cbs.dtu.dk/services/NetNGlyc/.

Reverse Transcriptase-PCR

Solanum neorickii LpfSRN-1 expression was tested using RT-PCR. Total RNA was purified from both mature pistils and leaves using a Qiagen RNeasy Plant Mini Kit and treated with a Qiagen RNase-Free DNase Kit. First strand cDNA templates were synthesized using a Bio-Rad iScript cDNA Synthesis Kit (http://www.bio-rad.com) using cycling conditions of 25°C for 5 min, 40°C for 30 min, and 85°C for 5 min. EconoTaq plus Green Mastermix (Lucigen) was used to amplify cDNA with the *LpfSRN-1* (test) and *CAC* (positive control) primer sets (**Supplementary Table S1**). RT-PCR products were run on a 1.2% agarose gel to examine expression levels.

Immunostaining of Stylar S-RNase Proteins

Immunostaining was performed for all red and orange fruited species (see *S-RNase Alleles in Four SC Red/Orange-Fruited Tomato Species* Section), all *S. neorickii* accessions and cross types (see *S-RNase Alleles LpfSRN-1 and LpfSRN-2 in SC S. neorickii* Section), and for selected *S. habrochaites* accessions that had not been previously analyzed for S-RNase protein expression (see *SC Accessions in S. habrochaites* Section). Stylar proteins were extracted from at least 10 mature, post-anthesis unpollinated styles to test for S-RNase expression. Weighed styles were homogenized in 2x SDS buffer (0.125 M Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 50 mM dithiothreitol, and 0.01% Bromophenol blue) at 10 µL per mg fresh weight. After grinding styles in the buffer, samples were heated for 5 min at 90°C and centrifuged at 14,000 g for 10 min. The supernatant was collected and frozen until use.

For each individual tested, protein extract equivalent to 0.2 mg fresh weight (unless otherwise noted) was separated by electrophoresis, blotted, and immunostained as previously described (Covey et al., 2010). tSRNC2 antibodies raised against the S-RNase conserved C-2 domain, were used as probes for S-RNase (Chalivendra et al., 2013), and those raised against a conserved peptide in HT-A and HT-B were used as probes for HT-protein (Broz et al., 2017).

Segregation Analysis for *S*-Locus Localization

Since there are numerous *RNase* genes in plant genomes that resemble *S-RNase* genes, we assessed whether alleles from *S. neorickii* (*LpfSRN-1*, *LpfSRN-2*) and *S. habrochaites* (*LhgSRN-1* and *hab-7*) segregated as would be predicted for a gene at the *S*-locus. We crossed females that were homozygous for well-characterized loss-of-function *S-RNase* alleles with males that were heterozygous for the *S-RNase* allele being tested and an *S-RNase* allele known to be at the S-locus. Allele-specific PCR was used to identify *S-RNase* sequences in progeny, including the expected female allele as a DNA quality control. If the male allele being tested is at the *S*-locus, we expect that it would never be inherited with the male allele known to be at the *S*-locus. The Freeman Halton extension of Fishers exact test was used to determine whether observed (progeny genotype) values differed from expected values if the tested allele was at the *S*-locus (0AB:1A:1B) or was not linked to the *S*-locus (2AB:3A:3B).

Floral Characters in SI and SC *S. habrochaites* Populations

The transition to SC is often correlated with reductions in flower size, and we wanted to assess this trait in populations of *S.*

habrochaites. Flower size was measured with digital calipers *in situ* in Ecuador, but to increase the accuracy of measurements, flowers from plants grown in a common garden at Colorado State University in the summer of 2016 were first preserved using clear packing tape as previously described (Spooner and Van Den Berg, 2001). The reproductive whorls were removed by snipping them at their base using forceps, and the corolla lobes were rolled out to stick to the tape, with the calyx removed. All open flowers of three separate inflorescences were scanned at high resolution (1200dpi) and measured digitally using ImageJ (Schneider et al., 2012). Measurements included petal length (A), inter-petal distance (B), width (C), sepal length (E), anther length (F), and stigma exsertion (G). Corolla area was approximated by calculating the area of a 5-pointed star [5AB * sin (36°)], where A = petal length and B = inter-petal distance.

A mixed model was used to detect significant differences between collection regions while accounting for sources of environmental variation and experimental blocks. Field designation (north or south plot), field position (row and column), flower collection date, and days post anthesis (day 0, 1, etc.) were used as random effects to detect significant ($p < 0.05$) differences between geographical regions (modeled as a fixed effect) for each variable. Generalized linear models were similarly used to detect significant differences between regions of collection sites for the other morphological observations (both *in situ* and common garden).

RESULTS

S-RNase Alleles in SC Species

S-RNase Alleles in Four SC Red/Orange-Fruited Tomato Species

Four of the six SC tomato clade species group in a subclade of closely related species that produce red or orange fruits: *S. lycopersicum*, *S. pimpinellifolium*, *S. galapagense*, and *S. cheesmaniae*. The *S*-locus of *S. lycopersicum* is one of the few *S*-loci in the Solanaceae to be completely sequenced (Sato et al., 2012, https://solgenomics.net). Li and Chetelat (2015) analyzed the *S*-locus of cultivated tomato and reported the presence of a single *S-RNase*-related sequence associated with a cluster of *SLF* genes in the pericentric region of Chromosome 1, as predicted for the *S*-locus in *Solanum*. Originally, the *S-RNase*-like sequence was referred to as a pseudogene with a 93-bp insertion, and it was proposed that this insertion could explain the lack of RNase activity in *S. lycopersicum* styles (Kondo et al., 2002b). However, a closer examination of the sequence reveals that the putative insertion is actually the characteristic single intron found in all Solanaceous *S-RNase* genes (**Supplementary Figure S1**). Similar sequences are found in all four members of the SC red/orange-fruited subclade (**Figure 1A**; **Supplementary Figure S1**), suggesting that this allele became fixed in a common ancestor to the group.

The predicted amino acid sequences of the encoded S-RNases in the red/orange-fruited species (**Figure 1A**) contain the five known conserved sequences C1-C5 in known S-RNases and are closely related to the known functional S20 S-RNase in *S. chilense* (Igić et al., 2007). There are two non-conservative differences between the pair of red-fruited species (*S. lycopersicum* and *S. pimpinellifolium*) compared to the pair of orange-fruited species endemic to the Galapagos Islands (*S. galapagense* and *S. cheesmaniae*). Given these differences, we refer to the *S-RNase* alleles as *SRN-red* for the allele in red-fruited species and as *SRN-orange* in the orange-fruited species (**Table 1**). One of the non-conservative differences between the encoded S20 S-RNase in *S. chilense* and the S-RNase encoded in red-fruited species is a Thr→ Met substitution in the C2 conserved domain that would eliminate an N-glycosylation site that is highly conserved in Solanaceae S-RNases, and which may reduce RNase function but should not prevent expression (Williams et al., 2015). These results suggest that the *S-RNase*-like sequences at the *S*-locus in these species could be expressed and, at least in the orange-fruited species, encode a potentially functional S-RNase. Previous work has indicated that styles of *S. lycopersicum* do not express S-RNase protein (Tovar-Méndez et al., 2014). We confirmed this and found that styles of the other red and orange fruited species also do not express S-RNase protein (**Figure 1B**). Further, transcriptome data collected on Sol Genomics (https://solgenomics.net/) and stylar RNAseq studies of *S. lycopersicum* (Pease et al., 2016) show no expression of the *S-RNase* gene. At this time, there is no clear explanation for the lack of *SRN-red* or *SRN-orange* expression, and therefore we classify these alleles as silenced (**Table 1**). The red- and orange-fruited species lack HT expression in addition to S-RNase expression (Kondo et al., 2002b), and lack IRBs, which can be partially restored by the transgenic introduction of functional *S-RNase* and *HT* genes (Tovar-Méndez et al., 2014).

S-RNase Allele *LcwSRN-1* in SC *S. chmielewskii*

In addition to the four red/orange fruited tomato species, there are two additional SC species that group within a subclade known as the Arcanum group, which contains three species: SI *S. arcanum*, SC *S. chmielewskii* and SC *S. neorickii*. Recent data have shown that while both SC species are derived from SI *S. arcanum*, they are independently derived from distinct geographical subsets of *S. arcanum* (Florez-Rueda et al., 2021). The single known *S-RNase* allele in *S. chmielewskii*, *LcwSRN-1* (after the previous species name, *Lycopersicum chmielewskii*), is not expressed at the RNA level (Kondo et al., 2002a), and the transition from SI to SC in *S. chmielewskii* likely followed the typical pistil first pattern with a loss of S-RNase function (Markova et al., 2017). The predicted amino acid sequence of the encoded LcwSRN-1 protein is 100% identical to that of the known functional S11 S-RNase of *S. chilense* in the aligned region (Igić et al., 2007); **Supplementary Figure S2**). Thus, *LcwSRN-1* likely represents another example of a potentially functional *S-RNase* allele that is transcriptionally silenced by an as yet unknown mechanism (**Table 1**).

S-RNase Alleles *LpfSRN-1* and *LpfSRN-2* in SC *S. neorickii*

S. neorickii is the other SC species in the Arcanum group. Kondo *et al.*, 2002a isolated two *S-RNase* alleles from *S. neorickii*,

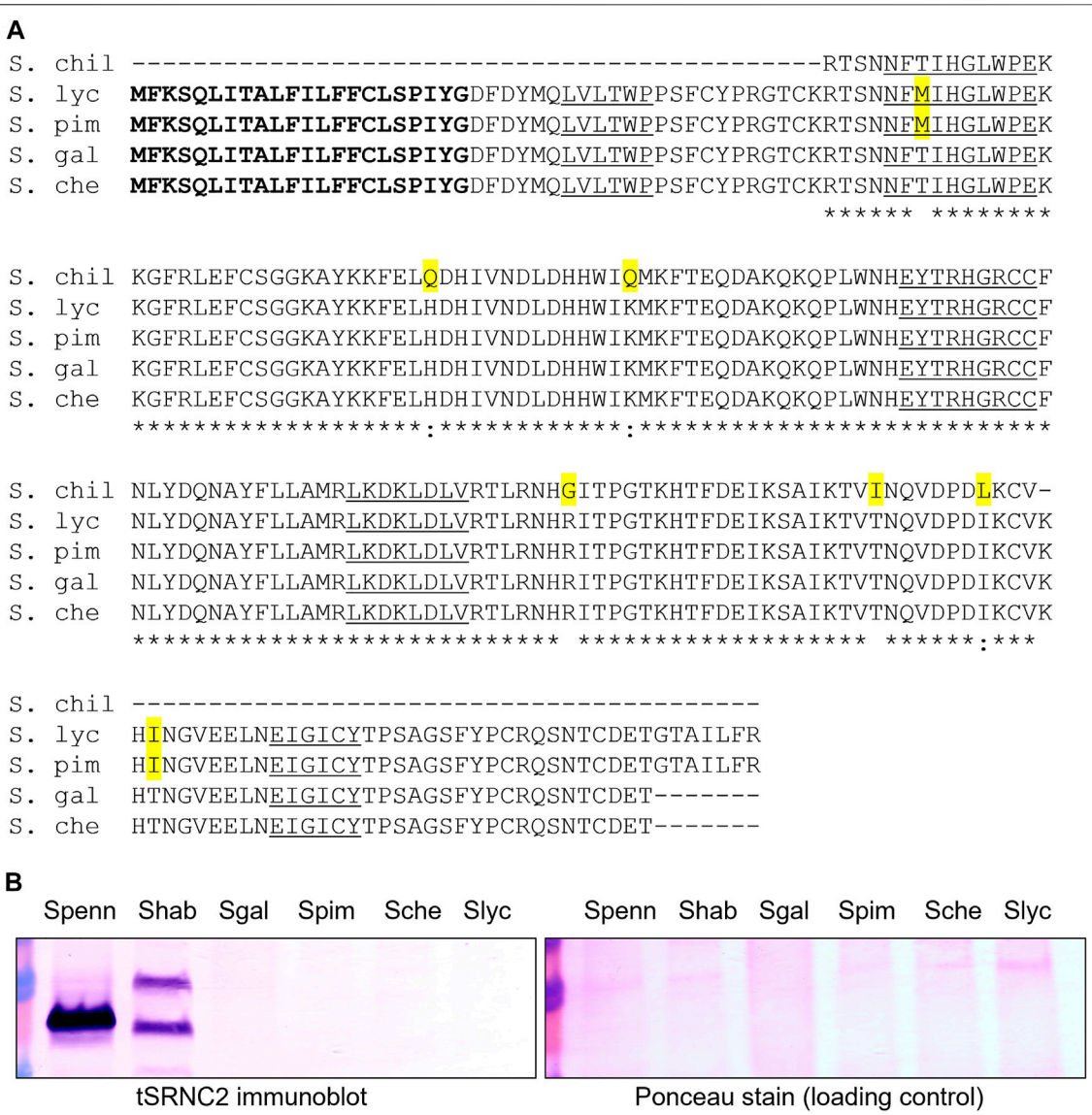

FIGURE 1 | S-RNase alignment and immunoblot from four red/orange-fruited tomato species. **(A)** alignment of deduced amino acid sequences S. chil = *S. chilense* S20 S-RNase, partial sequence (Igić et al., 2007) GenBank EF680098, S. lyc = *S. lycopersicum* cultivar Heinz 1706 (Li and Chetelat, 2015) Sol Genomics Network (https://solgenomics.net) Solyc01g055200.1, S. pim = *S. pimpinellifolium* accession LA1589 coding sequences Sol Genomics Network Sopim01g055200.0.1, S. gal = *S. galapagense* accession LA0317 PCR product, this study, GenBank OK091157, S. che = *S. cheesmaniae* accession LA0522 PCR product, this study, GenBank OK091158. The predicted signal peptide is bolded, amino acid substitutions are highlighted in yellow and conserved S-RNase sequences C1-C5 are underlined. Asterisks indicate conservation between all sequences. **(B)** Immunoblot of stylar extracts using the tSRNC2 antibody raised to the conserved C2 region of S-RNases. Ponceau Stain of the membrane is shown as a loading control. Extracts from two SI plants were run as positive controls. Spenn = SI *S. pennellii* LA1340, Shab = SI *S. habrochaites* LA1777, Sgal = SC *S. galapagense* LA0438, Spim = SC *S. pimpinellifolium* LA1245, Sche = SC *S. cheesmaniae* LA0522, Slyc = SC *S. lycopersicum* LA4444.

LpfSRN-1 and *LpfSRN-2* (after the previous species name, *Lycopersicum parviflorum*). *LpfSRN-2* is closely related to functional *S. arcanum* allele S6 (Royo et al., 1994a) but is non-functional due to a 1-bp insertion causing a frame shift that results in a premature stop codon (**Supplementary Figure S3**).

In contrast, *LpfSRN-1* has no obvious defect in its coding region (Kondo et al., 2002a) and the predicted protein is closely related to that of a known functional *S. peruvianum* S-RNase *SP2* [(Miller and Kostyun, 2011); **Supplementary Figure S4**]. *LpfSRN-1* is transcribed and translated in *S. neorickii* accession LA1322, although no RNase activity above background is detected in styles (Kondo et al., 2002a). Although it was initially assumed that SC in this species is due to the loss of S-RNase expression or function (Kondo et al., 2002a), recent evidence suggests SC may have resulted from the acquisition of a pollen-expressed *SLF23* gene whose encoded protein can detoxify the self LpfSRN-1 RNase (Markova et al., 2017). However,

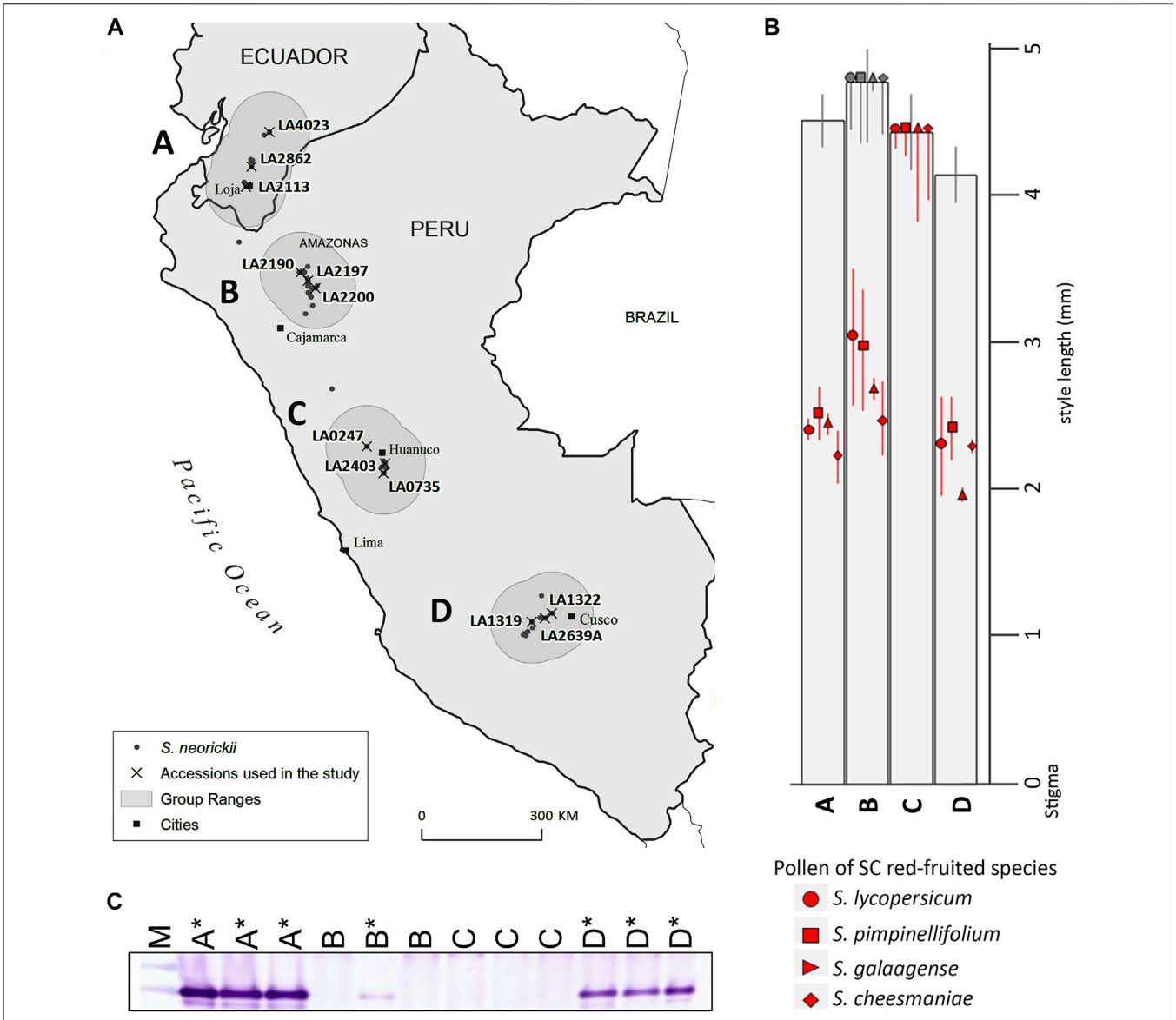

FIGURE 2 | Distribution, IRBs and S-RNase expression in *S. neorickii* geographic groups A-D**. (A)** Accessions used in this study are grouped according to geographical location (north to south) and range (shaded regions) within the species distribution. **(B)** Red/orange species pollen tube lengths 48 h post-pollination in *S. neorickii* styles are shown in millimeters and include the average length of the majority of pollen tubes (species symbols) and standard deviation (bars). Pollen tubes grew to ovaries in some individuals of geographic group B (gray symbols), but not in others (red symbols). **(C)** Immunoblot of stylar extracts using the tSRNC2 antibody raised to the conserved C2 region of S-RNases. Sample order is M, molecular weight marker; LA4023 (group A); LA2862 (A); LA2113 (A); LA2190 (B); LA2197 (B); LA2200 (B); LA0247 (C); LA0247 (C); LA0735 (C); LA1322 (D); LA2639A (D); LA1319 (D). * = accessions/individuals in which pollen tubes from red-fruited species are rejected in styles (presence of IRBs). The expression of S-RNase in Group B LA2197 was confirmed in a separate blot using a greater volume of extract (not shown).

because both pollen and pistil SI specificity genes have undergone mutation and are fixed in this species, it is difficult to know which mutation came first.

Previous work demonstrated that, although all *S. neorickii* accessions are SC, some accessions had functional IRBs (Baek et al., 2015). We therefore hypothesized that although LpfSRN-1 S-RNase cannot function in SI, it may still be able to function in interspecific pollen tube rejection. Because *S. neorickii* is distributed into four distinct geographic groups within its range [(**Figure 2A**) (groups A (Ecuador), B (Amazonas, Peru), C (Huánco, Peru) and D (Cusco and Apurimac, Peru)], we selected three accessions from each geographic group for further studies of IRBs.

We assessed whether the two different alleles previously identified in *S. neorickii* (*LpfSRN-1* and *LpfSRN-2*) are associated with different populations in the geographic distribution of the species using allele-specific PCR (**Supplementary Figure S5**). We found the *LpfSRN-1* allele in all accessions tested, consistent with the findings of Markova *et al.* (2017). However, we found the *LpfSRN-2* allele in group A and C

accessions, but not in all B and D group accessions. Upon sequencing, all *LpfSRN-2* alleles in positive accessions that were tested (data not shown) contained the same loss-of-function insertional mutation that was previously reported (Kondo et al., 2002a).

Because *S. neorickii* is highly autogamous, finding two *S-RNase* alleles in a presumably "heterozygous" state in multiple individuals from each of the A and C accessions was surprising. We hypothesized the two *S-RNase* alleles are linked to each other, putatively the result of transposition and/or genetic exchange near the *S*-locus as has been documented in Petunia (Wu et al., 2020). We tested for linkage of the two *S. neorickii* S-RNase alleles to each other and for *S*-locus localization using segregation analysis. Plants that contained the two *S. neorickii S-RNase* alleles (*LpfSRN-1* and *LpfSRN-2*) and *hab-7*, an *S-RNase* allele known to be at the *S*-locus (shown below), were used as males in crosses with female plants that were homozygous for a known *S-RNase* allele (*SRN-red* or *LhgSRN-1*, **Supplementary Table S2**). By analyzing *S-RNase* alleles in the progeny of this cross, we found that the two *S. neorickii* alleles were always inherited together, and never separately. Further, the two *S. neorickii* alleles were never inherited with the *hab-7* allele in progeny plants. These results are consistent with the two *S. neorickii* alleles being linked to each other and with these alleles being located at, or near, the *S*-locus (**Supplementary Table S2**).

We examined variation of IRBs in *S. neorickii* by pollinating pistils of accessions from each geographic group with pollen from red -fruited species and evaluating pollen tube growth in styles (**Figure 2B**; **Supplementary Figure S6**). We found that styles of geographic groups A and D accessions reject interspecific pollen tubes (IRBs present), whereas styles of group C accessions do not (IRBs absent), and styles of group B accessions varied depending on the individual being tested (IRBs segregating). Since previous work had demonstrated that S-RNase expression (with HT-protein) could constitute an IRB acting on pollen of red-fruited species (Tovar-Méndez et al., 2014) we next tested the same accessions for expression of S-RNase and HT by immunoblotting stylar extracts. Previously HT-A (but not HT-B) was identified in *S. neorickii* LA1322 (Group D) (Kondo et al., 2002a). We probed stylar extracts of *S. neorickii* accessions with an antibody designed to a peptide present in both HT-proteins and show that all accessions tested expressed HT-protein (**Supplementary Figure S7**). We found that accessions that were able to reject interspecific pollen also expressed S-RNase, whereas those that lacked IRBs did not express S-RNase at the mRNA or protein levels (**Figure 2C**, **Supplementary Figure S8**). Since the *LpfSRN-1* allele is silenced in some accessions and expressed in others, although no RNase activity is detected in styles (Kondo et al., 2002a), we classified this allele as both transcriptionally silenced and low S-RNase activity when expressed (**Table 1**).

To determine whether IRBs were dominant, we crossed a group C accession lacking IRBs (LA0247) and a group D accession possessing IRBs (LA1322). In all F_1 hybrid plants tested, all progeny expressed S-RNase protein and rejected interspecific pollen (**Figure 3**; **Supplementary Table S3**, **Supplementary Figure S9**). Four different F_1 plants were self-pollinated and F_2 progeny were phenotyped for both IRBs and S-RNase expression. We found that all F_2 plants that reject interspecific pollen tubes (possess IRBs) also express S-RNase protein (17/49, **Supplementary Table S3**; **Supplementary Figure S9**). However, a significant number of F_2 plants that accept interspecific pollen tubes (lack IRBs) also express S-RNase protein (12/49) (**Supplementary Table S3**; **Supplementary Figure S9**). These results suggest that expression of LpfSRN-1 S-RNase protein is necessary, but is not sufficient, for interspecific pollen tube rejection and therefore that another pistil factor is likely required for the observed IRBs.

S-RNase Alleles in SC Populations of SI Species

S. pennellii SC Accession LA0716 *S-RNase* Deletion

S. pennellii is a generally SI species, but a small number of accessions identified at the southern range margin exhibit SC. The best characterized of these is SC accession LA0716, which has been extremely useful for both basic research and applications to agriculture. For example, the ease of producing fertile F1 hybrids (Rick, 1960) allowed the construction of introgression lines (Eshed and Zamir, 1995) that have been used to generate a detailed physical map of the tomato genome and to isolate important agronomic factors (Lippman et al., 2007). The SC trait of this accession was also essential for the generation of the first complete genome sequence of a wild tomato species not closely related to the cultivated species (Bolger et al., 2014). S-RNase is undetectable in this accession either by activity (Covey et al., 2010) or by immunostaining (Chalivendra et al., 2013). Analysis of the *S*-locus in LA0716 failed to identify even a remnant *S-RNase* gene, suggesting that the gene has been deleted (Li and Chetelat, 2015; **Table 1**). It is thought that the complete lack of S-RNase caused a pistil first mating system transition in this accession, since male components that contribute to both SI and IRBs are active (Li and Chetelat, 2010; Li and Chetelat, 2015). Because LA0716 lacks S-RNase but exhibits robust IRBs (Baek et al., 2015), this accession has also been extremely useful for identifying genes involved in S-RNase-independent IRBs (Tovar-Méndez et al., 2017; Qin et al., 2018; Qin and Chetelat, 2021).

S. arcanum SC Accession LA2157 *LpSc* Allele

In *S. arcanum*, a single SC accession (LA2157) has been identified in this otherwise SI species. The *S-RNase* allele in LA2157 *LpSc* (after the previous species name, *Lycopersicum peruvianum*) is expressed at the protein level but a missense mutation eliminates a histidine residue essential for RNase activity (Royo et al., 1994b; **Table 1**). The protein encoded by the *LpSc* allele segregates with the SC phenotype, indicating that the allele resides at the *S*-locus and is responsible for the SC phenotype (Royo et al., 1994a). Except for the single amino acid substitution in the active site, the amino acid sequence of LpSc S-RNase is identical to both LcwSRN-1 (Markova et al., 2017) and functional *S. chilense* S11 S-RNase (**Supplementary Figure S2**). The transition from SI to SC in this *S. arcanum* accession likely followed the typical pistil first pattern of mutations with a loss of pistil S-RNase

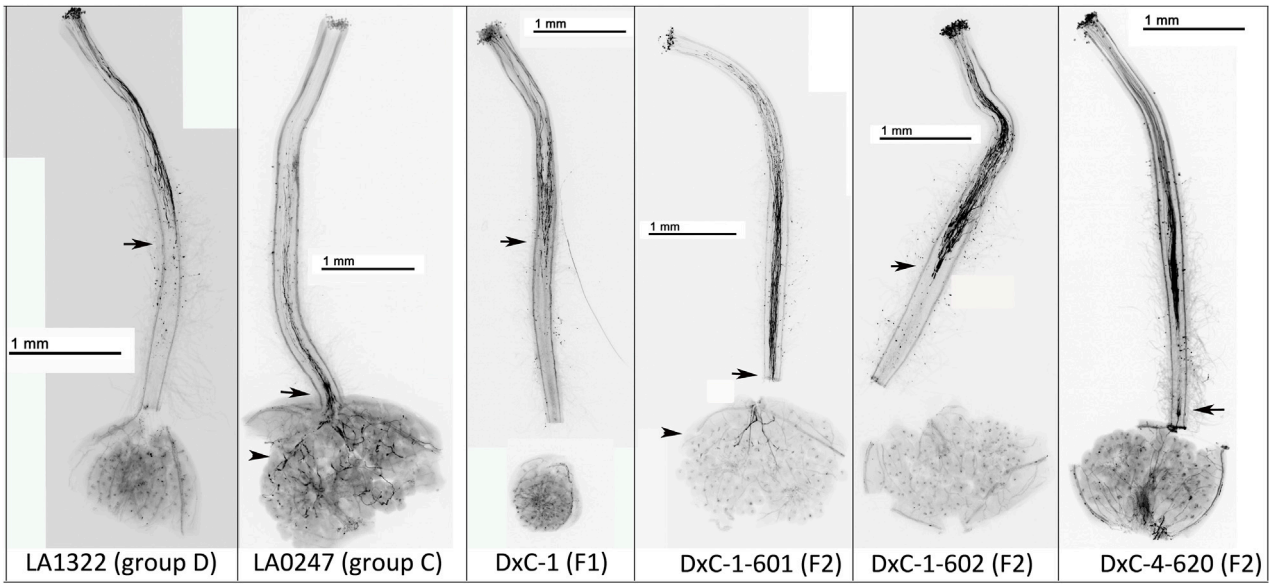

FIGURE 3 | IRBs in *S. neorickii* geographic groups D and C, F_1 and F_2 plants. Representative images of crosses in the F2 plants with pollen from red-fruited species (*S. lycopersicum* or *S. pimpinellifolium*). Plants used as the female in crosses are listed in the figures. Arrow indicates the point at which the majority of rejected pollen tubes stop growing and arrowhead indicates pollen tubes in the ovaries.

expression/activity (Markova et al., 2017). Pistil-side IRBs are greatly weakened in LA2157 compared to SI *S. arcanum* accessions (Baek et al., 2015; Tovar-Méndez et al., 2017) suggesting that loss of S-RNase activity causing a mating system transition to SC also affects IRBs.

SC Accessions in *S. habrochaites*

Remarkably, SC has arisen at least six times in the generally SI species *S. habrochaites* (**Table 1**). Five of the six known SI →SC transitions occurred in Ecuador at the northern species margin, and the SC-associated *S-RNase* alleles found in these SC accessions were likely derived from those present in ancestral SI populations in the region near the Ecuador-Peru border (**Figure 4**). The SC accessions of *S. habrochaites* have been categorized into groups (SC-1 to SC-6) based on distinct reproductive phenotypes (Broz et al., 2017; Broz et al., 2021; Landis et al., 2021) and specific *S-RNase* alleles (**Table 1**). Recent studies have demonstrated that these SC groups also display population differentiation (Landis et al., 2021).

Since reduced flower size is a character often associated with the selfing syndrome that can result from mating systems transitions to SC (Sicard and Lenhard, 2011; Wright et al., 2013), we measured corolla area across *S. habrochaites* SC and SI accessions in Ecuador (**Supplementary Figure S10**). Overall, we found that SC *S. habrochaites* accessions have not undergone floral size reduction, with the exception of SC-2 accessions at the far northern species margin, consistent with previous reports (Rick et al., 1979; Broz et al., 2017). However, we found a significant increase in the number of floral buds per inflorescence in SI versus SC populations (**Supplementary Figure S10**), which could potentially increase pollinator attraction in obligate outcrossers. We next examined the structure, origin and expression of *S-RNase* alleles involved in mating system transitions to SC in *S. habrochaites*.

The hab-8 S-RNase *Allele*

The newly discovered *hab-8 S-RNase* allele was identified in accession LA2101, collected in San Pedro de Cariamanga, Ecuador, in 1980 and was also found in wild populations at the same site in 2014 (site EC6, **Figure 4**). Transcripts of *hab-8* were not detected using RNA-seq, and S-RNase protein is not detected by immunoblotting (Broz et al., 2017). The predicted protein encoded by the *hab-8* allele is truncated due to a G→ A transition that creates a premature stop codon, i.e., a nonsense mutation (**Figure 5**; **Supplementary Figure S11**). The encoded hab-8 S-RNase is identical prior to the premature stop codon to that encoded by the *hab-14 S-RNase* allele segregating in SI accession LA2864 and in the mixed SI/SC accession LA 2098, both only ~40 km from the SC accessions containing *hab-8* (**Figure 4**; **Supplementary Table S4**). In turn, the *hab-14* allele encodes an S-RNase that is highly similar (a single non-conservative amino acid substitution) to that encoded by the previously reported *hab-11* allele found in SI plants in accessions LA0094 and LA2314 from Peru (Broz et al., 2021).

SC accessions near San Pedro de Cariamanga were designated as being group SC-5, and population structure analysis indicates a close relationship of group SC-5 with SI accessions in southern Ecuador (Landis et al., 2021; **Figure 4**). Previous studies have shown that SC-5 plants do not express S-RNase but do retain S-RNase-independent IRBs (Broz et al., 2017; Landis et al., 2021). We found that pistil-expressed HT-A, a protein involved in both SI and IRBs (Tovar-Méndez et al., 2014; Tovar-Méndez et al., 2017; **Table 1**) appears to be functional based on sequence

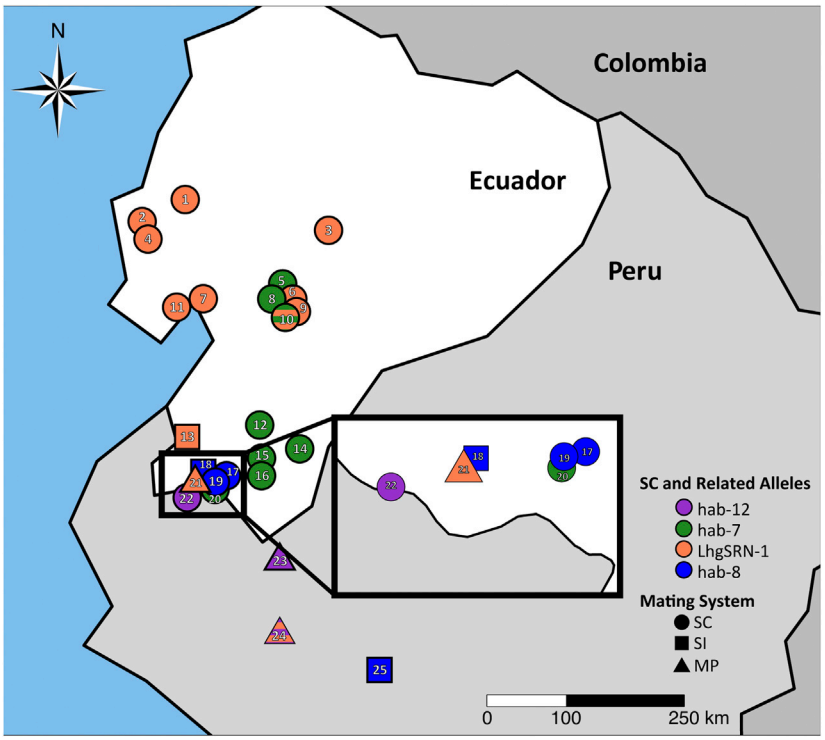

FIGURE 4 | SC-associated *S-RNase* alleles and ancestral SI-associated *S-RNase* alleles in *S. habrochaites* at the northern species margin. Alleles associated with self-compatibility (SC) are found in SC populations (circles), and their related putatively ancestral alleles are found in segregating SI (square) or SI/SC mixed populations (MP, triangles). Colors represent differentiated populations as described in Landis et al., 2021. For a key to accessions displayed see **Supplementary Table S4**. Orange = *LhgSRN-1* or *LhgSRN-1-like* alleles (SC-2), green = *hab-7* allele (SC-1), striped orange/green = segregating *LhgSRN-1* and *hab-7* alleles in a region of SC-1/SC-2 hybridization, blue = *hab-8* or *hab-8-like* alleles (SC-5), purple = *hab-12* or *hab-12-like* alleles (SC-7). *LhgSRN-1-like* alleles are also found in SI or MP populations in central Peru (**Supplementary Table S4**).

```
LA2098 hab-14  ------MSVLFLFLFVFSPVCGDFDYLQLVLQWPRSFCKSRYCPNPIPRNFTIHGLWPDK
LA2314 hab-11  MTKPQLMSVLFLFLFVFSPVCGDFDYLQLVLQWPRSFCKSRYCPNPIPRNFTIHGLWPDK
LA2101 hab-8   ------MSVLFLFLFVFSPVCGDFDYLQLVLQWPRSFCKSRYCPNPIPRNFTIHGLWPDK
EC6 hab-8      ------------------PVCGDFDYLQLVLQWPRSFCKSRYCPNPIPRNFTIHGLWPDK
                                  ****************************************
LA2098 hab-14  QRIMPINCPAKESYKSITDFKKIKLLEQHWPDLTSNQGSAEFWRYQYKKHGTCSVDLYNQ
LA2314 hab-11  QRIMPINCPAKESYKSITDSKKIKLLEQHWPDLTSNQGSAEFWRYQYKKHGTCSVDLYNQ
LA2101 hab-8   QRIMPINCPAKESYKSITDFKKIKLLEQHWPDLTSNQGSAEF------------------
EC6 hab-8      QRIMPINCPAKESYKSITDFKKIKLLEQRWPDLTSNQGSAEF------------------
               ****************** ********.*************
LA2098 hab-14  EQYFDLAIELKEKFDLLKTLKNHGITPSKTNTVIDVEKAIKAVTKEVPNLNCIGDSSQTM
LA2314 hab-11  EQYFDLAIELKEKFDLLKTLKNHGITPSKTNTVIDVEKAIKAVTKEVPNLNCIGDSSQTM
LA2101 hab-8   ------------------------------------------------------------
EC6 hab-8      ------------------------------------------------------------

LA2098 hab-14  ELLEIGICFNREGTTVIACRRRWNNHPNGNQKITLPP
LA2314 hab-11  ELLEIGICFNREGTTVIACRRRWNNHPNGNQKITLPP
LA2101 hab-8   ------------------------------------
EC6 hab-8      ------------------------------------
```

FIGURE 5 | Amino acid sequence of *S. habrochaites* SC-associated hab-8 S-RNase aligned with amino acid sequence of SI-associated hab-14 and hab-11 S-RNases. Predicted amino acids of hab-8 S-RNase (GenBank OK091160) from SC accessions LA2101 and EC6 are aligned with predicted hab-14 sequences (GenBank OK091163) from SI individuals from mixed SI/SC accession LA 2098 (an identical sequence was recovered from SI accession LA2864, not shown) and hab-11 in LA2314 (identical to partial codon sequence of hab-11 in accession LA0094, GenBank MW183817, not shown). The predicted signal peptide is bolded, amino acid substitutions are highlighted in yellow and conserved sequences C1-C5 are underlined. Asterisks indicate conservation between all sequences.

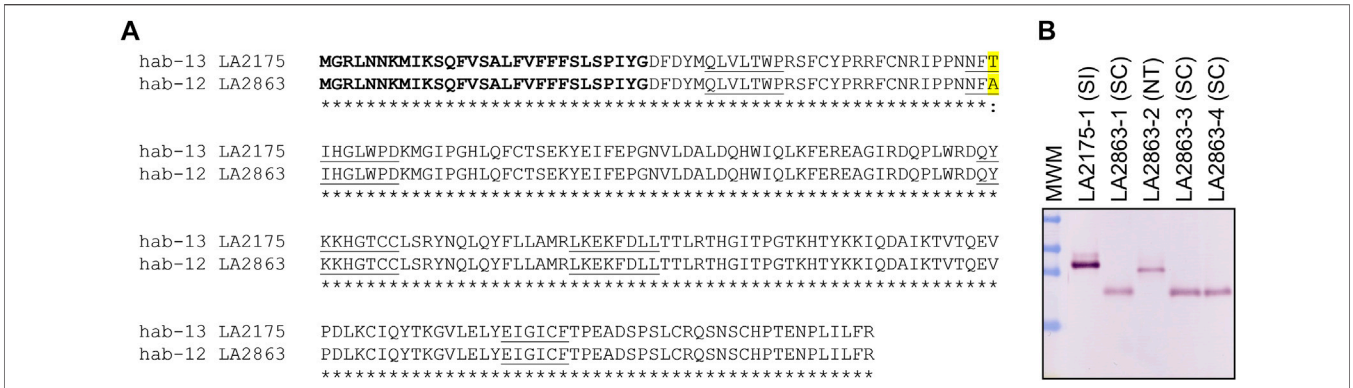

FIGURE 6 | Alignment and immunoblot of hab-12 S-RNase. **(A)** Deduced amino acid sequences of hab-12 (GenBank OK091161) and hab-13 (GenBank OK091162) from SC accession LA2863 and an SI plant from MP accession LA2175, respectively. A shorter but identical hab-13 sequence was identified in MP accession LA1391 (not shown). The predicted signal peptide is bolded, the single amino acid substitution is highlighted in yellow and conserved sequences C1-C5 are underlined. Asterisks indicate conservation between both sequences. **(B)** Immunoblot of stylar extracts using the tSRNC2 antibody raised to the conserved C2 region of S-RNases. MWM, molecular weight marker; SI, self-incompatible; SC, self-compatible; NT, mating system not tested.

analysis of RNA-seq data (**Supplementary Figure S12**) and is expressed at the protein level (Broz et al., 2017) in SC-5 accessions.

The hab-12 S-RNase *Allele*

The newly discovered *hab-12 S-RNase* allele (**Figure 6A**) was identified in SC accession LA2863 collected near Macará in southern Ecuador (**Figure 4**). In accession LA2863, SC plants express an S-RNase protein that appears smaller than normal on immunoblots (**Figure 6B**). RNA-seq data using RNA from styles of SC plants revealed high expression (33,000–46,000 FPKM) of a single *S-RNase* allele that we named *hab-12*. The nucleotide sequence (including the sequence of the single intron) of the *hab-12* allele was identical to that of the *hab-13 S-RNase* allele found in SI plants of the mixed SI/SC accession LA2175 and mixed SI/SC accession LA1391 from northern Peru except for a single A→ T transition that creates a missense mutation (**Supplementary Figure S13**) resulting in a single Thr→ Ala amino acid substitution within the conserved C2 region (**Figure 6A**). This substitution in the hab-12 protein would eliminate the only potential N-glycosylation site in the protein, a modification which is apparently not required for allele-specific S-RNase function (Karunanandaa et al., 1994; Soulard et al., 2013) but may affect S-RNase uptake, stability or targeting in pollen tubes (Williams et al., 2015). The lack of glycosylation could explain the apparent low molecular weight of the hab-12 S-RNase on immunoblots (**Figure 6B**). Interestingly, not all LA2863 individuals showed the same pattern on immunoblots (**Figure 6B**, individual 2), suggesting that *hab-12* is not fixed in this accession.

The SC type found segregating in accession LA2863 was designated as SC group SC-7 (**Table 1**). Since the reproductive phenotype of the SC-7 group had not been previously characterized, we performed test crosses to assess different types of reproductive barriers in this group (**Supplementary Figure S14**). We confirmed an SC mating system (self-pollen tubes reach ovaries and fruits are formed in self pollinations), determined that IRBs are intact in SC-7 (pistils reject pollen tubes of cultivar tomato and of *S. neorickii*) and that SC-7 plants do not have defects in pollen resistance factors (SC-7 pollen tubes are accepted by pistils of SI accession LA1777). Further, SC-7 pistils do not reject pollen tubes of accession SC-2 group accession LA0407, indicating that SC-7 pistils lack the inter-population barriers that are found in SI and in SC-4 accessions (Broz et al., 2017). *HT-A* sequences were identified in SC plants in the SC-7 accession LA2863 and appear to be expressed and functional (**Supplementary Figure S12**).

The LhgSRN-1 S-RNase *Allele*

In the most northern accessions of *S. habrochaites* (the SC-2 group), the *LhgSRN-1 S-RNase* allele is associated with SC (**Figure 4**). Segregation analysis indicated that the *LhgSRN-1* allele is at, or near, the *S*-locus (**Supplementary Table S2**). Previous studies showed that although this allele encodes a seemingly functional S-RNase, it is not expressed at the RNA or protein level (Kondo et al., 2002a; Covey et al., 2010; Broz et al., 2017). The SC-2 accessions all possess a Miniature Inverted-repeat Transposable Element (MITE) in the promoter of the *LhgSRN-1* gene (Kondo et al., 2002a; Broz et al., 2017), and this MITE was assumed to be responsible for the lack of expression. However, we identified highly similar, presumably ancestral, functional *LhgSRN-1*-like alleles in SI accessions throughout the species range that harbored the same MITE sequence in their promoter regions (**Supplementary Table S5**). Analysis of stylar RNA from several of these SI and mixed SI/SC accessions showed high levels of expression of the *LhgSRN-1*-like alleles (**Table 2**). This indicates that the presence of the MITE is not responsible for the lack of expression of *LhgSRN-1* in SC-2 accessions. Thus, *LhgSRN-1* is classified as a silenced allele (**Table 1**), but the silencing mechanism is currently unknown.

The *LhgSRN-1*-like alleles identified in SI and SI/SC accessions include *hab-4*, *hab-9*, *hab-16* and *hab-17*

TABLE 2 | Expression of silenced/expressed *S-RNase* alleles in SC and SI *S. habrochaites* populations. Transcriptional expression was analyzed by either RT-PCR or RNA-seq analysis using stylar RNA. S-RNase protein was analyzed by immunoblotting with stylar protein extracts. NT = not tested. [a]This study, [b]Covey et al., 2010, [c]Broz et al., 2017, [d]Broz et al., 2021, [e]Multiple individuals tested negative, and a single individual tested positive, [f]This population segregates for *hab-7* and *LhgSRN-1*, and all individuals tested that contained *hab-7* were positive for *hab-7* mRNA with RNA-seq and for S-RNase protein with immunoblotting (**Supplementary Figure S15**), [g]Two clones of this plant were grown in either the field or in a growth chamber and gave identical results, [h]This plant was a genetic sibling of plant (LA2119 x LA2175)-851 and was grown in the greenhouse.

Allele	Plant type/accession	SI/SC	*S-RNase* transcript expression	S-RNase protein
LhgSRN-1	LA0407	SC	Negative with RT-PCR[b]	Negative[c]
LhgSRN-1-like hab-16	LA2868	SI	NT	Positive[c]
LhgSRN-1-like hab-17	LA2099	SI/SC	FPKM = 22,840[a,f]	NT
LhgSRN-1-like hab-4	LA1353	SI	Positive with RT-PCR[b]	NT
LhgSRN-1-like hab-9	LA0094	SI	Positive with RT-PCR[d]	Positive[d]
hab-7	LA2119	SC	FPKM = ~ 40[a]	Negative[c,e]
hab-7	LA2119	SC	NT	Positive[a]
hab-7	EC40	SC	NT	Positive[a]
hab-7	PI250315	SC	FPKM = 30,000–56,000[a]	Positive[a,f]
hab-7/hab-15	(LA2119 x LA2175)-851[g]	SC	FPKM = ~ 40/35,000[a]	NT
hab-7/hab-15	(LA2119 x LA2175)-852[h]	SC	FPKM = 24,000/29,000[a]	NT

```
hab-16    MIKTQHTLSFFILLCALSDVYGTFNQLQLVLRWPASFCEGKKCERTPNNFTIHGLWPDIK
hab-17    MIKTQHTLAFFILLCALSDVYGTFNQLQLVLRWPASFCEGKKCERTPNNFTIHGLWPDIK
hab-9     ---------------ALSDVYGTFNQLQLVLRWPASFCEGKKCERTPNNFTIHGLWPDIK
LhgSRN-1  MIKTQHTLSFFILLCALSDVYGTFNQLQLVLRWPASFCEGKKCERTPNNFTIHGLWPDIK
                          *******************************************

hab-16    DTILNNCNPDAKYAPVTGGKFVKRNKHWPDLILTEAASLKRQGFWEYQFKKHGTCCSDLF
hab-17    GTILNNCNPDAKYAPVTGGKFVKRNKHWPDLILTEAASLKRQGFWEYQFKKHGTCCSDLF
hab-9     GTILNNCNPDAKYASVTGGKFVKRNKHWPDLILTEAASLKRQGFWEYQFKKHGTCCSDLF
LhgSRN-1  DTILNNCNPDAKYAPVTGGKFVKRNKHWPDLILTEAASLKRQGFWEYQFKKHGTCCSDLF
          .*************.*******************************************

hab-16    NQEKYFYLALILKDKFDLLTTFRNKGIIPKSTCTINKIQKTIRTVTGMVPNLSCTPTMEL
hab-17    NQEKYFDLALILKDKFDLLTTFRNKGIIPKSTCTINKIQKTIRTVTGVVPNLSCTPTMEL
hab-9     NQEKYFDLALILKDKFDLLTTFRNKGIIPKSTCTINKIQKTIRTVTGVVPNLSCTPTMEL
LhgSRN-1  NQEKYFYLALILKDKFDLLTTFRNKGIIPKSTCTINKIQKTIRTVTGMVPNLSCTPTMEL
          ****** ****************************************:************

hab-16    LEVGICFNRDTSKLIDCDQPKTCGTSGNTEIFFP-
hab-17    LEVGICFNRDASKLIDCDQPKTCGTSGNTEIFFP-
hab-9     LEVGICFNRDASKLIDCDQPKTCGTSGNTEIFFSL
LhgSRN-1  LEVGICFNRDASKLIDCDQPKTCGTSGNTEIFFP-
          **********:**********************.
```

FIGURE 7 | Amino acid sequence alignment of *S. habrochaites* SI-associated LhgSRN-1-like hab-16, hab17 and hab-9 S-RNases with SC-associated LhgSRN-1 S-RNase. Predicted amino acids of LhgSRN-1 S-RNase from SC-2 group accessions are aligned with predicted hab-16 sequences (GenBank OK091165) from SI individuals from SI accession LA2868, hab-17 (GenBank OK091166) from mixed SI/SC LA 2099 (identical sequences were recovered from SI EC7 and EC10 collections as well as mixed SI/SC accession LA1391, not shown) and hab-9 in LA0094, GenBank MW183816 (identical sequences were recovered from accession LA1648). The LhgSRN-1-like hab-4 sequence is not shown due to its relatively short length. The predicted signal peptide is bolded, amino acid substitutions are highlighted in yellow and conserved sequences C1-C5 are underlined. Asterisks indicate conservation between all sequences.

(**Table 2**), and their deduced amino acid sequences differ from LhgSRN-1 by between one and six amino acids (**Figure 7**). The closely related hab-16 sequence has a single Ala/Thr substitution relative to LhgSRN-1, and the SI accession from which this sequence is derived (LA2868) is geographically close to the SC-2 group accessions (**Figure 4**; **Supplementary Table S4**). In addition, LA2868 and SC-2 accessions display similar population structure (Landis et al., 2021). Together, the data strongly suggest that *hab-16* is the ancestral functional allele of *LhgSRN-1*.

Pistils of SC-2 accessions show a reduction in strength of IRBs against pollen tubes of both cultivated tomato and SC *S. neorickii* (Broz et al., 2017) compared to their SI counterparts which contain robust IRBs (**Supplementary Table S5**; Covey et al., 2010; Broz et al., 2017; Broz et al., 2021; Landis et al., 2021). This suggests that loss of S-RNase expression in the SC-2 group diminishes but does not eliminate IRBs. All SC-2 group accessions were found to contain HT-protein, except for a single accession (LA1223) which lacks all IRBs and was designated SC-3 (Broz et al., 2017).

The hab-7 S-RNase *Allele*

The recently described *hab-7* allele (Landis et al., 2021) resides in SC-1 accessions (Broz et al., 2017; Landis et al., 2021). SC-1 accessions are generally found in a north-south corridor centering on the town of Loja in southern Ecuador, and in 2014 SC populations with the *hab-7* allele were found to persist in this region, as well as near the town of Cariamanga (EC40) (**Figure 4**). Previous studies have shown that the SC-1 group possesses S-RNase independent IRBs (Broz et al., 2017), and population structure analysis indicates that the SC-1 group has differentiated from ancestral SI populations and other SC groups (Landis et al., 2021). Segregation analysis indicated that the *hab-7* allele is at, or near, the *S*-locus (**Supplementary Table S2**).

Although the *hab-7 S-RNase* appears to encode a functional S-RNase, with a single conservative Val/Leu amino acid difference with the *S. peruvianum* S-13 S-RNase (Landis et al., 2021), it is generally not expressed in SC-1 plants (Broz et al., 2017), and the silencing mechanism remains unknown. Unexpectedly, our RNA-seq and immunoblot studies indicate that the normally silenced *hab-7* allele can become activated, resulting in high levels of expression in styles (**Table 2**). For example, immunoblotting showed that hab-7 S-RNase protein is expressed in plants of the SC population EC40. Further, we found that hab-7 was expressed in a clone of a single LA2119 plant (an SC-1 accession containing *hab-7*) grown in a greenhouse at Colorado State University (**Supplementary Figure S15**; **Table 2**), but not in clones of the same plant grown in the field (**Table 2**), suggesting that environmental conditions may influence expression. Changes in genetic background may also activate *hab-7* expression. For example, accessions from a region of central Ecuador where SC-1/SC-2 hybridization may have occurred (represented by accession PI251305), segregate for both *hab-7* and *LhgSRN-1*. In these accessions, plants containing the *hab-7* allele also express *hab-7* mRNA to high levels in styles, about 1,000x higher than when the gene is silenced (**Table 2**). Styles of plants heterozygous for *hab-7* exhibit about half of the expression level compared to *hab-7* homozygotes, which is also suggested by immunostaining for S-RNase protein (**Supplementary Figure S15**). In another example of expression induced by hybridization, when we produced *hab-7/hab-15* hybrids for RNA-seq and segregations studies, we found that some, but not all, plants expressed the *hab-7* allele at a level comparable to that of the functional SI *S-RNase* allele *hab-15* (**Table 2**). Thus, although *hab-7* is classified as a silenced allele (**Table 1**), our data suggest that different genetic or environmental conditions can lead to robust transcriptional activation.

The hab-6 S-RNase *Allele*

In contrast to the multiple mating system transitions seen at the northern *S. habrochaites* species margin, there has been a single SI → SC transition at the southern species margin in central Peru, producing the SC-4 group, which represents nearly 25% of the species range (Covey et al., 2010; Broz et al., 2021). The *hab-6* allele associated with the SC-4 group is expressed but encodes a low activity protein which does not appear to function in SI. However, the SC-4 group retains robust IRBs (Covey et al., 2010; Broz et al., 2021).

DISCUSSION

A Pathway to SC – Loss of Function Mutations and Silencing of *S-RNase* Genes

Although mutations in *S-RNase* genes can drive mating system transitions from SI to SC, the specific nature of these mutations is often not well characterized. Here we examined the structure, origin and expression of 12 *S-RNase* alleles associated with SC species and populations in the tomato clade (**Table 1**; **Figure 8**). In three cases, the reason for *S-RNase* gene dysfunction due to mutations is quite clear: one species contains a gene deletion (*S. pennellii* SC accession LA0716), one allele has a frame-shift mutation (*S. neorickii LpfSRN-2*), and another allele contains a nonsense mutation (*S. habrochaites hab-8*). In three other cases, *S-RNase* alleles are expressed but produce proteins that are non-functional in SI (*S. arcanum LpSc*, and *S. habrochaites hab-6* and *hab-12*). Here SC is predicted to result from changes in critical amino acid residues that are likely important for S-RNase protein function. We found that five alleles are transcriptionally silenced (*SRN-red* in *S. lycopersicum* and *S. pimpinellifolium* and *SRN-orange* in *S. galapagense* and *S. cheesmaniae*, *S. chmielewskii LcwSRN-1*, and *S. habrochaites LhgSRN-1* and *hab-7*), but the silencing mechanisms remain unknown. Finally, one *S-RNase* allele (*LpfSRN-1*) can be either transcriptionally silenced or actively transcribed and translated to produce an S-RNase that does not function in SI. In this case, the S-RNase has very low activity (Kondo et al., 2002a), but it appears that the transition to SC is likely to have resulted from a gain-of-function in the male component of the *S*-locus (Markova et al., 2017). Still, although gain-of-function acquisitions of pollen SLFs are expected to be rare given the architecture of SI in the Solanaceae, it should be noted that without complete sequencing of *S*-loci and a better understanding of the interactions between S-RNases and their cognate SLFs, it is impossible to completely rule out pollen-first mutations even when *S-RNase* alleles are non-functional.

In addition to those in the tomato clade, SC populations have been detected in numerous SI species within the Solanaceae (Goldberg et al., 2010; Robertson et al., 2011). However, in most cases, the genes underlying these SC transitions remain unknown. Studies on SC populations of *S. chacoense* (Qin et al., 2001), *S. carolinense* (Mena-Ali and Stephenson, 2007) and *Petunia axillaris* (Tsukamoto et al., 1999; Tsukamoto et al., 2003a), have found low or no expression of specific *S-RNase* alleles, and genetic factors modifying *S-RNase* expression are likely the causative factors leading to SC (Qin et al., 2001; Tsukamoto et al., 2003a). In the Rosaceae and Rutaceae families, which also exhibit gametophytic SI with *S-RNase* and *SLF* alleles at the *S*-locus, there is evidence that SC can result from deletion (Sassa et al., 1997), mutation (Liang et al., 2020) or silencing (Fernandez I Marti et al., 2014) of *S-RNase* genes, or from pollen-side mutations (Vilanova et al., 2006). We believe the

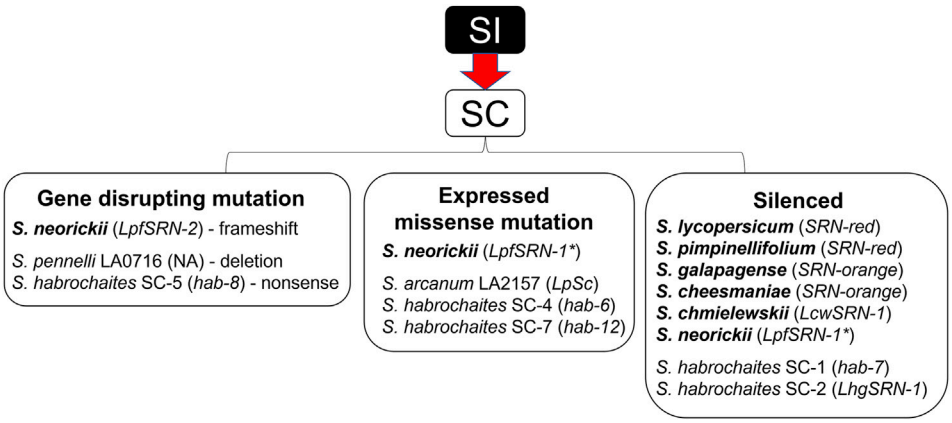

FIGURE 8 | Overview of *S-RNase* alleles associated with self-compatibility (SC) in the tomato clade. Species that are fully SC are shown in bold, for other predominately SI species, SC populations or population groups are listed. Names of *S-RNase* alleles are shown in parentheses. Missense mutations result in low S-RNase activity. **LpfSRN-1* is silenced in some accessions but is expressed and has low activity in others. For additional details see **Table 1** and manuscript text.

current study, combined with previous results, establishes the tomato clade, with its array of SC species and populations, as the premier study system for understanding the molecular basis of S-RNase-based SI to SC transitions. However, continued work in other plant species will be critical to understanding the diverse mechanisms by which SC arises.

Expression of SC-Associated S-RNases can Alter IRBs

When functional S-RNases are expressed (in conjunction with HT and other pistil SI factors), they can function in both SI and in S-RNase-dependent IRBs (Tovar-Méndez et al., 2014; Tovar-Méndez et al., 2017). When S-RNases are not expressed, IRBs can be entirely absent, as in SC red- and orange-fruited tomato species and SC *S. chmielewskii*, or severely weakened, as in the SC-2 group of *S. habrochaites* (Baek et al., 2015). Alternatively, when S-RNases are not expressed, S-RNase-independent IRBs can be robust, entirely replacing S-RNase-dependent IRBs. Examples include SC *S. pennellii* accession LA0716, in which the *S-RNase* gene is deleted (Covey et al., 2010), and the SC-1, SC-5 and SC-6 groups of *S. habrochaites* (Broz et al., 2017; Landis et al., 2021), in which S-RNases are not expressed but IRBs are intact.

Intriguingly, in cases where a low/no activity S-RNase protein is expressed and plants are SC, IRBs can vary. In the single *S. arcanum* SC accession LA2157, which expresses defective S-RNase LpSc, pistil-side IRBs are severely compromised but still weakly active against pollen tubes of red- and orange-fruited tomato species (Baek et al., 2015). In SC-4 and SC-7 groups of *S. habrochaites*, which express hab-6 and hab-12 S-RNases respectively, self-pollen tubes are not rejected, but interspecific pollen tubes are rejected [(Covey et al., 2010; Broz et al., 2021), **Supplementary Figure S14**]. In these cases, it is not clear whether the S-RNases that are defective in SI can still function in IRBs, or if S-RNase-independent mechanisms are responsible for rejection of interspecific pollen tubes.

In this study, we found that expression of low-activity LpfSRN-1 S-RNase was required for IRBs in *S. neorickii* populations (**Figures 2**, **3**; **Supplementary Figure S6**; **Supplementary Table S3**). However, our results also indicate that another pistil factor(s) is required along with LpfSRN-1 for fully functioning IRBs. We found that HT-proteins, which play a role in both S-RNase-dependent and S-RNase independent IRBs (Tovar-Méndez et al., 2014; Tovar-Méndez et al., 2017) were expressed in all *S. neorickii* accessions (**Supplementary Figure S7**), suggesting that a different factor is involved. Possibilities include the additional SI and UI factors that have been identified within the Solanaceae (Hancock et al., 2005; Jiménez-Durán et al., 2013; Garcia-Valencia et al., 2017; Qin and Chetelat, 2021). Clearly, more work will be required to identify this additional IRB factor(s).

Transcriptional Plasticity of *S-RNase* Alleles

Our study uncovered transcriptionally silenced *S-RNase* alleles, and the mechanism(s) underlying the silencing of apparently intact *S-RNase* alleles is currently unknown. Genomic sequencing could clarify whether expression depends on sequence variation of promoters or other regulatory sequences. In the two cases where there are virtually identical pairs of silenced and expressed alleles (*LcwSRN-1* and *LpSc*, *LhgSRN-1* and *hab-16*) direct sequence comparison of regulatory regions should be possible. However, *S-RNase* expression may depend on additional genetic factors, as indicated by our crossing experiments in *S. neorickii* showing heritable variation in *LpfSRN-1* expression. The activation of *hab-7* transcription in hybrids of *S. habrochaites* (*hab-7/hab-15* heterozygotes and SC-1/SC-2 hybrids, **Table 2**) also points to a genetic basis for transcriptional activation of *S-RNase*. Our results also suggest that there may be environmental influences on *S-RNase* expression, given the apparently spurious reactivation of *hab-7* in a greenhouse-grown versus field-grown clone from the same plant. Previous work indicates that levels of S-RNase can vary between plants of different genetic backgrounds, and even between styles on the

same plant, which may be due to differences in factors influencing expression, activity or turnover of S-RNase (Qin et al., 2006; Mena-Ali and Stephenson, 2007; Fernandez I Marti et al., 2014). Mapping studies combined with whole genome sequencing could shed light on the genetic mechanisms underlying silencing, while studies of DNA methylation, chromatin modification and small RNA expression can help clarify if silencing and reactivation result from epigenetic modifications.

It has been widely assumed that the transition from SI to SC is irreversible (Stebbins, 1974; Igić and Busch, 2013), in part because SC is often the result of one or more loss-of-function mutations; and functional reconstitution of these genes would be extremely rare. In this scenario, the genetic diversity normally maintained by outcrossing is lost, putting selfing populations at a higher risk for extinction than their SI counterparts. The plasticity of *S-RNase* expression that we observed suggests that mating system transitions may have the potential for reversibility. This would have important agronomic implications, as a lack of SC germplasm limits many plant breeding programs, and additional tools to manipulate mating system would be extremely valuable (Muñoz-Sanz et al., 2020). In addition, reversible *S-RNase* silencing could influence the evolution and spread of plant populations. Although highly speculative, it is interesting to consider a scenario under which a temporary pause on enforced outcrossing due to transient silencing of an *S-RNase* allele could promote the successful colonization of novel habitats. Plants carrying a single copy of the silenced *S-RNase* would exhibit the SC phenotype, but other *S*-haplotypes would be retained in heterozygous individuals. As long as a sufficient number of diverse *S*-haplotypes (>3) is preserved in a locally adapted founder population, SI systems could theoretically become reactivated with the reversal of *S-RNase* gene silencing, preventing a permanent loss of genetic diversity. Of course, if selfing syndrome characters such as reduced flower size evolve in conjunction with SC, a simple reversal of *S-RNase* expression may not be sufficient to reinstitute SI. Our finding that the transition to SC in *S. habrochaites* is typically not associated with reduced flower size may suggest that SC plants can continue to recruit pollinators, facilitating outcrossing even when selfing is possible. In this case, we would not expect changes in floral morphology to inhibit the putative reversion from SC to SI. A greater understanding of the frequency and mechanisms of *S-RNase* silencing will determine the potential for this scenario.

CONCLUSION

An analysis of SI to SC transitions in the tomato clade reveals a diverse array of mutations that can lead to the loss of *S-RNase* function. This likely represents only a fraction of the diversity that lies within the Solanaceae, and more broadly in S-RNase-based systems of SI. The nature of *S-RNase* mutations can also lead to changes in IRBs, influencing interactions between species. Intriguingly we identified a number of cases in which *S-RNases* can undergo transcriptional silencing, which in some cases can be reversed. Taken together, our results, suggest that *S-RNase* expression, and potentially mating system transitions, may be more dynamic than has previously been thought.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, OK091159 - OK091166 https://www.ncbi.nlm.nih.gov/genbank/, PRJNA310635.

AUTHOR CONTRIBUTIONS

AB, CM, YB, and PB conceived, designed and performed experiments, analyzed data, interpreted data and wrote sections of the manuscript. AT performed immunostaining experiments, analyzed data and interpreted data. PA and TR performed experiments and provided technical support. DR analyzed and interpreted RNA-seq data. PB and AB drafted the final version of the manuscript. All authors provided intellectual content, edited the manuscript and approved the final manuscript.

FUNDING

This study was supported by grant MCB-1127059 to PB from the National Science Foundation Plant Genome Research Program. Field studies in Ecuador were supported by a Fellowship to PB from the Fulbright Foundation.

ACKNOWLEDGMENTS

We thank Bruce McClure for help with immunostaining experiments, and Matthew Hahn for coordinating the RNA-seq analysis. We also thank Olivia Todd, Lauren Nalezny, Nicole Irace, Oliver Kassenbrock, Alex King, Quincy Cobb and Laura Hantzis for help with pollen tube imaging and plant care, and Emily Bernard, Tatum Hastings, Evan Hayden and Dakota Loe for assistance with genotyping.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.780793/full#supplementary-material

REFERENCES

Allen, A. M., and Hiscock, S. J. (2008). "Evolution and Phylogeny of Self-Incompatibility Systems in Angiosperms," in *Self-Incompatibility in Flowering Plants – Evolution, Diversity, and Mechanisms*. Editor V. E. Franklin-Tong (Berlin Heidelberg: Springer-Verlag), 73–101. doi:10.1007/978-3-540-68486-2_4

Baek, Y. S., Covey, P. A., Petersen, J. J., Chetelat, R. T., McClure, B., and Bedinger, P. A. (2015). Testing the SI × SC Rule: Pollen-Pistil Interactions in Interspecific Crosses between Members of the Tomato Clade (Solanum Section Lycopersicon , Solanaceae). *Am. J. Bot.* 102, 302–311. doi:10.3732/ajb.1400484

Baker, H. G. (1955). Self-compatibility and Establishment after "'Long-Distance" Dispersal. *Evolution* 9, 347–349. doi:10.1111/j.1558-5646.1955.tb01544.x

Baker, H. G. (1967). Support for Baker's Law-As a Rule. *Evolution* 21, 853–856. doi:10.2307/2406780

Bedinger, P. A., Broz, A. K., Tovar-Méndez, A., and McClure, B. (2017). Pollen-Pistil Interactions and Their Role in Mate Selection. *Plant Physiol.* 173, 79–90. doi:10.1104/pp.16.01286

Bedinger, P. A., Chetelat, R. T., McClure, B., Moyle, L. C., Rose, J. K. C., Stack, S. M., et al. (2011). Interspecific Reproductive Barriers in the Tomato Clade: Opportunities to Decipher Mechanisms of Reproductive Isolation. *Sex. Plant Reprod.* 24, 171–187. doi:10.1007/s00497-010-0155-7

Bolger, A., Scossa, F., Bolger, M. E., Lanz, C., Maumus, F., Tohge, T., et al. (2014). The Genome of the Stress-Tolerant Wild Tomato Species *Solanum pennellii*. *Nat. Genet.* 46, 1034–1038. doi:10.1038/ng.3046

Broz, A. K., Randle, A. M., Sianta, S. A., Tovar-Méndez, A., McClure, B., and Bedinger, P. A. (2017). Mating System Transitions in *Solanum habrochaites* Impact Interactions between Populations and Species. *New Phytol.* 213, 440–454. doi:10.1111/nph.14130

Broz, A. K., Simpson-Van Dam, A., Tovar-Méndez, A., Hahn, M. W., McClure, B., and Bedinger, P. A. (2021). Spread of Self-compatibility Constrained by an Intrapopulation Crossing Barrier. *New Phytol.* 231, 878–891. doi:10.1111/nph.17400

Busch, J. W., and Delph, L. F. (2012). The Relative Importance of Reproductive Assurance and Automatic Selection as Hypotheses for the Evolution of Self-Fertilization. *Ann. Bot.* 109, 553–562. doi:10.1093/aob/mcr219

Busch, J. W., and Schoen, D. J. (2008). The Evolution of Self-Incompatibility when Mates Are Limiting. *Trends Plant Sci.* 13, 128–136. doi:10.1016/j.tplants.2008.01.002

Chalivendra, S. C., Lopez-Casado, G., Kumar, A., Kassenbrock, A. R., Royer, S., Tovar-Mèndez, A., et al. (2013). Developmental Onset of Reproductive Barriers and Associated Proteome Changes in Stigma/styles of *Solanum pennellii*. *Exbotj* 64, 265–279. doi:10.1093/jxb/ers324

Covey, P. A., Kondo, K., Welch, L., Frank, E., Sianta, S., Kumar, A., et al. (2010). Multiple Features that Distinguish Unilateral Incongruity and Self-Incompatibility in the Tomato Clade. *Plant J.* 64, 367–378. doi:10.1111/j.1365-313x.2010.04340.x

Coyne, J. A., and Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.

Darwin, C. R. (1876). *The Effects of Cross and Self-Fertilisation in the Vegetable Kingdom*. London: J. Murray.

De Nettancourt, D. (1977). *Incompatibility in Angiosperms*. Berlin: Springer-Verlag.

Eshed, Y., and Zamir, D. (1995). An Introgression Line Population of *Lycopersicon pennellii* in the Cultivated Tomato Enables the Identification and fine Mapping of Yield-Associated QTL. *Genetics* 141, 1147–1162. doi:10.1093/genetics/141.3.1147

Fernández i Martí, A., Gradziel, T. M., and Socias I Company, R. (2014). Methylation of the S F Locus in almond Is Associated with S-RNase Loss of Function. *Plant Mol. Biol.* 86, 681–689. doi:10.1007/s11103-014-0258-x

Florez-Rueda, A. M., Scharmann, M., Roth, M., and Städler, T. (2021). Population Genomics of the "Arcanum" Species Group in Wild Tomatoes: Evidence for Separate Origins of Two Self-Compatible Lineages. *Front. Plant Sci.* 12, 624442. doi:10.3389/fpls.2021.624442

Franklin-Tong, V. E. (2008). "Self-incompatibility in *Papaver Rhoeas*: Progress in Understanding Mechanisms Involved in Regulating Self-Incompatibility in *Papaver*," in *Self-Incompatibility in Flowering Plants: Evolution, Diversity, and Mechanisms*. Editor V. E. Franklin-Tong (Berlin Heidelberg: Springer-Verlag), 237–258. doi:10.1007/978-3-540-68486-2_11

Fujii, S., Kubo, K., and Takayama, S. (2016). Non-self- and Self-Recognition Models in Plant Self-Incompatibility. *Nat. Plants* 2, 16130. doi:10.1038/nplants.2016.130

García-Valencia, L. E., Bravo-Alberto, C. E., Wu, H.-M., Rodríguez-Sotres, R., Cheung, A. Y., and Cruz-García, F. (2017). SIPP, a Novel Mitochondrial Phosphate Carrier, Mediates in Self-Incompatibility. *Plant Physiol.* 175, 1105–1120. doi:10.1104/pp.16.01884

Goldberg, E. E., Kohn, J. R., Lande, R., Robertson, K. A., Smith, S. A., and Igić, B. (2010). Species Selection Maintains Self-Incompatibility. *Science* 330, 493–495. doi:10.1126/science.1194513

Igic, B., and Busch, J. W. (2013). Is Self-fertilization an Evolutionary Dead End. *New Phytol.* 198, 386–397. doi:10.1111/nph.12182

Igic, B., Lande, R., and Kohn, J. R. (2008). Loss of Self-Incompatibility and its Evolutionary Consequences. *Int. J. Plant Sci.* 169, 93–104. doi:10.1086/523362

Igic, B., Smith, W. A., Robertson, K. A., Schaal, B. A., and Kohn, J. R. (2007). Studies of Self-Incompatibility in Wild Tomatoes: I. S-Allele Diversity in *Solanum chilense* Dun. (Solanaceae). *Heredity* 99, 553–561. doi:10.1038/sj.hdy.6801035

Jany, E., Nelles, H., and Goring, D. R. (2019). The Molecular and Cellular Regulation of Brassicaceae Self-Incompatibility and Self-Pollen Rejection. *Int. Rev. Cel Mol. Biol.* 343, 1–35. doi:10.1016/bs.ircmb.2018.05.011

Jiménez-Durán, K., McClure, B., García-Campusano, F., Rodríguez-Sotres, R., Cisneros, J., Busot, G., et al. (2013). NaStEP: A Proteinase Inhibitor Essential to Self-Incompatibility and a Positive Regulator of HT-B Stability in *Nicotiana alata* Pollen Tubes. *Plant Physiol.* 161, 97–107. doi:10.1104/pp.112.198440

Karunanandaa, B., Huang, S., and Kao, T. (1994). Carbohydrate Moiety of the *Petunia inflata S*3 Protein Is Not Required for Self-Incompatibility Interactions between Pollen and Pistil. *Plant Cell* 6, 1933–1940. doi:10.1105/tpc.6.12.1933

Kondo, K., Yamamoto, M., Itahashi, R., Sato, T., Egashira, H., Hattori, T., et al. (2002a). Insights into the Evolution of Self-Compatibility in Lycoperseiconfrom a Study of Stylar Factors. *Plant J.* 30, 143–153. doi:10.1046/j.1365-313x.2002.01275.x

Kondo, K., Yamamoto, M., Matton, D. P., Sato, T., Hirai, M., Norioka, S., et al. (2002b). Cultivated Tomato Has Defects in Both S-RNase and HT Genes Required for Stylar Function of Self-incompatibility. *Plant J.* 29, 627–636. doi:10.1046/j.0960-7412.2001.01245.x

Kowyama, Y., Kunz, C., Lewis, I., Newbigin, E., Clarke, A. E., and Anderson, M. A. (1994). Self-Compatibility in a *Lycopersicon peruvianum* Variant (LA2157) is Associated with a Lack of Style S-RNase Activity. *Theor. Appl. Genet.* 88, 859–864.

Kubo, K.-i., Entani, T., Takara, A., Wang, N., Fields, A. M., Hua, Z., et al. (2010). Collaborative Non-self Recognition System in S-RNase-Based Self-Incompatibility. *Science* 330, 796–799. doi:10.1126/science.1195243

Kubo, K.-I., Paape, T., Hatakeyama, M., Entani, T., Takara, A., Kajihara, K., et al. (2015). Gene Duplication and Genetic Exchange Drive the Evolution of S-RNase-Based Self-Incompatibility in *Petunia*. *Nat. Plants* 1, 14005. doi:10.1038/nplants.2014.5

Landis, J. B., Miller, C. M., Broz, A. K., Bennett, A. A., Carrasquilla-Garcia, N., Cook, D. R., et al. (2021). Migration through a Major Andean Ecogeographic Disruption as a Driver of Genetic and Phenotypic Diversity in a Wild Tomato Species. *Mol. Biol. Evol.* 38, 3202–3219. doi:10.1093/molbev/msab092

Li, W., and Chetelat, R. T. (2010). A Pollen Factor Linking Inter- and Intraspecific Pollen Rejection in Tomato. *Science* 330, 1827–1830. doi:10.1126/science.1197908

Li, W., and Chetelat, R. T. (2014). The Role of a Pollen-Expressed Cullin1 Protein in Gametophytic Self-Incompatibility in *Solanum*. *Genetics* 196, 439–442. doi:10.1534/genetics.113.158279

Li, W., and Chetelat, R. T. (2015). Unilateral Incompatibility Gene ui1.1 Encodes an S-Locus F-Box Protein Expressed in Pollen of *Solanum* Species. *Proc. Natl. Acad. Sci. USA* 112, 4417–4422. doi:10.1073/pnas.1423301112

Li, W., Royer, S., and Chetelat, R. T. (2010). Fine Mapping of ui6.1, a Gametophytic Factor Controlling Pollen-Side Unilateral Incompatibility in Interspecific *Solanum* Hybrids. *Genetics* 185, 1069–1080. doi:10.1534/genetics.110.116343

Liang, M., Cao, Z., Zhu, A., Liu, Y., Tao, M., Yang, H., et al. (2020). Evolution of Self-Compatibility by a Mutant Sm-RNase in Citrus. *Nat. Plants* 6, 131–142. doi:10.1038/s41477-020-0597-3

Lippman, Z. B., Semel, Y., and Zamir, D. (2007). An Integrated View of Quantitative Trait Variation Using Tomato Interspecific Introgression Lines. *Curr. Opin. Genet. Develop.* 17, 545–552. doi:10.1016/j.gde.2007.07.007

Madeira, F., Park, Y. m., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi:10.1093/nar/gkz268

Markova, D. N., Petersen, J. J., Qin, X., Short, D. R., Valle, M. J., Tovar-Méndez, A., et al. (2016). Mutations in Two Pollen Self-incompatibility Factors in Geographically Marginal Populations of *Solanum habrochaites* Impact Mating System Transitions and Reproductive Isolation. *Am. J. Bot.* 103, 1847–1861. doi:10.3732/ajb.1600208

Markova, D. N., Petersen, J. J., Yam, S. E., Corral, A., Valle, M. J., Li, W., et al. (2017). Evolutionary History of Two Pollen Self-incompatibility Factors Reveals Alternate Routes to Self-compatibility within Solanum. *Am. J. Bot.* 104, 1904–1919. doi:10.3732/ajb.1700196

McClure, B. (2004). S-RNase and SLF Determine S-haplotype-specific Pollen Recognition and Rejection. *Plant Cell* 16, 2840–2847. doi:10.1105/tpc.104.161130

Mena-Ali, J. I., and Stephenson, A. G. (2007). Segregation Analyses of Partial Self-Incompatibility in Self and Cross Progeny of *Solanum carolinense* Reveal a Leaky S-Allele. *Genetics* 177, 501–510. doi:10.1534/genetics.107.073775

Miller, J. S., and Kostyun, J. L. (2011). Functional Gametophytic Self-Incompatibility in a Peripheral Population of *Solanum peruvianum* (Solanaceae). *Heredity* 107, 30–39. doi:10.1038/hdy.2010.151

Muñoz-Sanz, J. V., Zuriaga, E., Cruz-García, F., McClure, B., and Romero, C. (2020). Self-(In)compatibility Systems: Target Traits for Crop-Production, Plant Breeding, and Biotechnology. *Front. Plant Sci.* 11, 195. doi:10.3389/fpls.2020.00195

Nasrallah, J. B. (2019). Self-incompatibility in the Brassicaceae: Regulation and Mechanism of Self-Recognition. *Curr. Top. Develop. Biol.* 131, 435–452. doi:10.1016/bs.ctdb.2018.10.002

Nathan Hancock, C., Kent, L., and McClure, B. A. (2005). The Stylar 120 kDa Glycoprotein Is Required for *S*-specific Pollen Rejection in *Nicotiana*. *Plant J.* 43, 716–723. doi:10.1111/j.1365-313x.2005.02490.x

Pannell, J. R., Auld, J. R., Brandvain, Y., Burd, M., Busch, J. W., Cheptou, P. O., et al. (2015). The Scope of Baker's Law. *New Phytol.* 208, 656–667. doi:10.1111/nph.13539

Pannell, J. R., and Barrett, S. C. H. (1998). Baker's Law Revisited: Reproductive Assurance in a Metapopulation. *Evolution* 52, 657–668. doi:10.1111/j.1558-5646.1998.tb03691.x

Pease, J. B., Guerrero, R. F., Sherman, N. A., Hahn, M. W., and Moyle, L. C. (2016). Molecular Mechanisms of Postmating Prezygotic Reproductive Isolation Uncovered by Transcriptome Analysis. *Mol. Ecol.* 25, 2592–2608. doi:10.1111/mec.13679

Porcher, E., and Lande, R. (2005). Loss of Gametophytic Self-Incompatibility with Evolution of Inbreeding Depression. *Evolution* 59, 46–60. doi:10.1111/j.0014-3820.2005.tb00893.x

Qin, X., and Chetelat, R. T. (2021). Ornithine Decarboxylase Genes Contribute to S-RNase-independent Pollen Rejection. *Plant Physiol.* 186, 452–468. doi:10.1093/plphys/kiab062

Qin, X., Li, W., Liu, Y., Tan, M., Ganal, M., and Chetelat, R. T. (2018). A Farnesyl Pyrophosphate Synthase Gene Expressed in Pollen Functions in S - RN Ase-independent Unilateral Incompatibility. *Plant J.* 93, 417–430. doi:10.1111/tpj.13796

Qin, X., Liu, B., Soulard, J., Morse, D., and Cappadocia, M. (2006). Style-by-style Analysis of Two Sporadic Self-Compatible *Solanum chacoense* Lines Supports a Primary Role for S-RNases in Determining Pollen Rejection Thresholds. *J. Exp. Bot.* 57, 2001–2013. doi:10.1093/jxb/erj147

Qin, X., Luu, D. T., Yang, Q., Maës, O., Matton, D. P., Morse, D., et al. (2001). Genotype-dependent Differences in S12-RNase Expression lead to Sporadic Self-Compatibility. *Plant Mol. Biol.* 45, 295–305. doi:10.1023/a:1006445120648

Ramanauskas, K., and Igić, B. (2017). The Evolutionary History of Plant T2/S-type Ribonucleases. *PeerJ* 5, e3790. doi:10.7717/peerj.3790

Rick, C. M., Fobes, J. F., and Tanksley, S. D. (1979). Evolution of Mating Systems in *Lycopersicum hirsutum* as Deduced from Genetic Variation in Electrophoretic and Morphological Characters. *Pl Syst. Evol.* 132, 279–298. doi:10.1007/bf00982390

Rick, C. M. (1960). Hybridization between *Lycopersicum esculentum* and *Solanum pennellii*: Phylogenetic and Cytogenetic Significance. *Proc. Natl. Acad. Sci.* 46, 78–82. doi:10.1073/pnas.46.1.78

Rick, C. M., Kesicki, E., Fobes, J. F., and Holle, M. (1976). Genetic and Biosystematic Studies on Two New Sibling Species of Lycopersicon from Interandean Perú. *Theoret. Appl. Genet.* 47, 55–68. doi:10.1007/bf00281917

Robertson, K., Goldberg, E. E., and Igić, B. (2011). Comparative Evidence for the Correlated Evolution of Polyploidy and Self-Compatibility in Solanaceae. *Evolution* 65, 139–155. doi:10.1111/j.1558-5646.2010.01099.x

Royo, J., Kowyama, Y., and Clarke, A. E. (1994a). Cloning and Nucleotide Sequence of Two S-RNases from *Lycopersicon peruvianum* (L.) Mill. *Plant Physiol.* 105, 751–752. doi:10.1104/pp.105.2.751

Royo, J., Kunz, C., Kowyama, Y., Anderson, M., Clarke, A. E., and Newbigin, E. (1994b). Loss of a Histidine Residue at the Active Site of S-Locus Ribonuclease Is Associated with Self-Compatibility in *Lycopersicon peruvianum*. *Proc. Natl. Acad. Sci.* 91, 6511–6514. doi:10.1073/pnas.91.14.6511

Sassa, H., Hirano, H., Nishio, T., and Koba, T. (1997). Style-specific Self-Compatible Mutation Caused by Deletion of the S-RNase Gene in Japanese Pear (*Pyrus serotina*). *Plant J.* 12, 223–227. doi:10.1046/j.1365-313x.1997.12010223.x

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., et al. (2012). The Tomato Genome Sequence Provides Insights into Fleshy Fruit Evolution. *Nature* 485, 635–641.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 Years of Image Analysis. *Nat. Methods* 9, 671–675. doi:10.1038/nmeth.2089

Sicard, A., and Lenhard, M. (2011). The Selfing Syndrome: a Model for Studying the Genetic and Evolutionary Basis of Morphological Adaptation in Plants. *Ann. Bot.* 107, 1433–1443. doi:10.1093/aob/mcr023

Soulard, J., Qin, X., Boivin, N., Morse, D., and Cappadocia, M. (2013). A New Dual-specific Incompatibility Allele Revealed by Absence of Glycosylation in the Conserved C2 Site of a *Solanum chacoense* S-RNase. *J. Exp. Bot.* 64, 1995–2003. doi:10.1093/jxb/ert059

Spooner, D. M., and Van Den Berg, R. (2001). "Quantitative Assessment of Corolla Shape Variation in Mexican Solanum Section Petota," in *Solanaceae V: Advances in Taxonomy and Utilization*. Editors R. Van Der Berg, G. W. W. Barendse, G. M. Can Der Weerden, and C. Mariani (Nikmegen, Netherlands: Nikmegen University Press), 61–71.

Stebbins, G. L. (1974). *Flowering Plants: Evolution above the Species Level.* Cambridge: Belknap Press.

Takayama, S., and Isogai, A. (2005). "Self-incompatibility in Plants," in *Annual Review of Plant Biology*, 56, 467–489. doi:10.1146/annurev.arplant.56.032604.144249*Annu. Rev. Plant Biol.*

Tovar-Méndez, A., Lu, L., and McClure, B. (2017). HT Proteins Contribute to S-RNase-independent Pollen Rejection in *Solanum*. *Plant J.* 89, 718–729. doi:10.1111/tpj.13416

Tovar-Méndez, A., Kumar, A., Kondo, K., Ashford, A., Baek, Y. S., Welch, L., et al. (2014). Restoring Pistil-Side Self-Incompatibility Factors Recapitulates an Interspecific Reproductive Barrier between Tomato Species. *Plant J.* 77, 727–736. doi:10.1111/tpj.12424

Tsukamoto, T., Ando, T., Kokubun, H., Watanabe, H., Masada, M., Zhu, X., et al. (1999). Breakdown of Self-Incompatibility in a Natural Population of *Petunia axillaris* (Solanaceae) in Uruguay Containing Both Self-Incompatible and Self-Compatible Plants. *Sex. Plant Reprod.* 12, 6–13. doi:10.1007/s004970050166

Tsukamoto, T., Ando, T., Kokubun, H., Watanabe, H., Sato, T., Masada, M., et al. (2003a). Breakdown of Self-Incompatibility in a Natural Population of *Petunia axillaris* Caused by a Modifier Locus that Suppresses the Expression of an S-RNase Gene. *Sex. Plant Reprod.* 15, 255–263. doi:10.1007/s00497-002-0161-5

Tsukamoto, T., Ando, T., Takahashi, K., Omori, T., Watanabe, H., Kokubun, H., et al. (2003b). Breakdown of Self-Incompatibility in a Natural Population of *Petunia axillaris* Caused by Loss of Pollen Function. *Plant Physiol.* 131, 1903–1912. doi:10.1104/pp.102.018069

Vilanova, S., Badenes, M. L., Burgos, L., Martinez-Calvo, J., Llacer, G., and Romero, C. (2006). Self-compatibility of Two Apricot Selections Is Associated with Two Pollen-Part Mutations of Different Nature. *Plant Physiol.* 142, 629–641. doi:10.1104/pp.106.083865

Weigend, M. (2004). Additional Observations on the Biogeography of the Amotape-Huancabamba Zone in Northern Peru: Defining the South-Eastern Limits. *Revista Peruana de Biologia* 11, 127–134. doi:10.15381/rpb.v11i2.2447

Weigend, M. (2002). Observations on the Biogeography of the Amotape-Huancabamba Zone in Northern Peru. *Bot. Rev.* 68, 38–54. doi:10.1663/0006-8101(2002)068[0038:ootbot]2.0.co;2

Williams, J. S., Wu, L., Li, S., Sun, P., and Kao, T.-H. (2015). Insight into S-RNase-Based Self-Incompatibility in *Petunia*: Recent Findings and Future Directions. *Front. Plant Sci.* 6, 41. doi:10.3389/fpls.2015.00041

Wright, S. I., Kalisz, S., and Slotte, T. (2013). Evolutionary Consequences of Self-Fertilization in Plants. *Proc. R. Soc. B.* 280, 20130133. doi:10.1098/rspb.2013.0133

Wu, L., Williams, J. S., Sun, L., and Kao, T. H. (2020). Sequence Analysis of the *Petunia inflata* S-locus Region Containing 17 S-Locus F-Box Genes and the S-RNase Gene Involved in Self-incompatibility. *Plant J.* 104, 1348–1368. doi:10.1111/tpj.15005

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, orclaim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 10 January 2022
doi: 10.3389/fgene.2021.791055

Genome-Wide Analysis of TIR-NBS-LRR Gene Family in Potato Identified *StTNLC7G2* Inducing Reactive Oxygen Species in Presence of *Alternaria solani*

Namo Dubey[1,2], *Anjali Chaudhary*[1,2] *and Kunal Singh*[1,2]*

[1]CSIR-Institute of Himalayan Bioresource Technology, Palampur, India, [2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

Resistance gene analogs (RGAs) comprising NBS-LRR gene family members are considered prominent candidates in the development of disease-resistant genotypes. NBS-LRR gene family comprised a very large number of genes; therefore, members of one subfamily TIR-NBS-LRR (TNL) are identified in the present study from *Solanum tuberosum* genome, followed by their bioinformatics characterization. The study identified a total of 44 genes encoding 60 TNL transcripts with two prominent clusters at chromosome 1 and chromosome 11. Expression analysis of 14 TNL genes after *Alternaria solani* infection at 1, 2, 3, 5, and 7 days post inoculation in two disease-tolerant varieties, Kufri Jyoti and Kufri Pukhraj, and one relatively susceptible variety, Kufri Chandramukhi, showed differential expression of many genes including a high expression (>15-fold) of *StTNLC6G2T1* and *StTNLC11G9T1*. Functional characterization of one such gene, *StTNLC7G2*, reveals involvement in the generation of reactive oxygen species under *A. solani* attack, implicating its putative role in plant defense *via* hypersensitive response.

Keywords: potato, *Alternaria solani*, TIR-NBS-LRR, early blight disease, plant defense

OPEN ACCESS

Edited by:
Peter Poczai,
University of Helsinki, Finland

Reviewed by:
Kalenahalli Yogendra,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India
Cheng Qin,
Zunyi Vocational and Technical College, China

***Correspondence:**
Kunal Singh
kunal@ihbt.res.in

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

Received: *07 October 2021*
Accepted: *26 November 2021*
Published: *10 January 2022*

Citation:
Dubey N, Chaudhary A and Singh K (2022) Genome-Wide Analysis of TIR-NBS-LRR Gene Family in Potato Identified StTNLC7G2 *Inducing Reactive Oxygen Species in Presence of* Alternaria solani. *Front. Genet. 12:791055. doi: 10.3389/fgene.2021.791055*

INTRODUCTION

Phytopathogens are microbes that have developed the capacity to suppress and overcome the host immune responses and inhabit the plant tissue for their survival, growth, and even multiplication (Chisholm et al., 2006). In response, plants have evolved too with time and developed a sophisticated immune system involving specialized resistance proteins to cope and combat a wide array of pathogens (Qi and Roger, 2013). Members of the NBS-LRR family are one such class of proteins consisting of an N-terminal CC or TIR domain, central NBS domain, and LRR motifs at C-terminal (Mchale et al., 2006). These proteins, also known as resistance gene analogs (RGAs), have been studied extensively in many plant systems with delineation of their role against a variety of pathogens including fungal, viral, nematode, and bacterial species (Dubey and Singh, 2018). In plants, they form a major line of defense under innate immunity. Plants' innate immunity is the first line of defense response to pathogen attack by recognition of pathogen-associated molecular patterns (PAMPs) by plant recognition receptors (PRRs) (Jones and Dangl, 2006). This is followed by the induction of reactive oxygen species (ROS) known as hypersensitive response (HR) and recognition of pathogen-associated virulence factors and effectors by members of the NBS-LRR protein family. The

recognition and sensing of pathogen attack along with downstream relay of a signal by defense-associated signaling pathway help the host to act decisively (Padmanabhan et al., 2009). One major subset of this protein family composed of members with toll-interleukin receptor (TIR) domain at the N-terminal is known as TIR-NBS-LRR (TNL) proteins (Nandety et al., 2013).

Many TNLs have been reported to provide resistance in plants against microbial pathogens (Whitham et al., 1994). One such protein identified is Y1 of potatoes against potato virus Y (PVY), a homolog of N protein of tobacco that confers cell death upon infection with PVY (Goyer et al., 2015). Therefore, TNL proteins are prime candidates in assisting plant defense against phytopathogens. Among the various pathogens infecting potatoes (*Solanum tuberosum* L.), the filamentous pathogens including oomycetes and fungi are considered the most devastating. Early blight disease caused by *Alternaria solani*, late blight of potatoes caused by *Phytophthora infestans*, and *Rhizoctonia solani* causing black scurf and stem canker of potatoes are prominent ones. *A. solani* can be highly devastating in regions with frequent rainfall with warm and dry conditions, leading to huge losses to farmers (Vloutoglou and Kalogerakis, 2000).

In the present study, TNL genes encoding putative proteins have been identified from the potato phureja genome and characterized by bioinformatics methodology. Selected transcripts were analyzed for expression behavior against fungal phytopathogen *A. solani*. One gene, *StTNLC7G2*, showing induced expression after pathogen attack and localization in the plasma membrane was further characterized *via* agroinfiltration in *S. tuberosum* as well as *Nicotiana benthamiana*.

MATERIALS AND METHODS

Plant Material and *Alternaria solani* Infection

Three potato genotypes, namely, Kufri Pukhraj (KP), Kufri Jyoti (KJ), and Kufri Chandramukhi (KC), were procured. All the three varieties were grown under an institutional farm field at CSIR-IHBT (latitude 32.0934°N; longitude 76.5439°E), Palampur, India. Among the three varieties used in the experiments, KJ and KP have been previously released as blight disease-tolerant varieties, while KC is considered more susceptible to blight diseases of potatoes (https://cpri.icar.gov.in/AddLink/getIMG/6068). The *A. solani* (MTCC-10690; accession no. HM484353.1) culture was procured from IMTECH, Chandigarh, India, maintained in a laboratory at 25°C, and used for infection purposes. The fungal culture was inoculated to 1-month-old susceptible potato plants with 100 ml of fungal suspension (2×10^5 conidia ml^{-1}) and sprayed to retain the virulence. Conidia were counted using hemocytometer under a microscope. Plants were kept at a relative humidity of 70% at 26°C with 12-h day/night condition. Fungal mycelia obtained from infected portion using single spore culture methodology (Choi et al., 1999) were used for further maintenance of culture along with subsequent infection work. The identity of fungus being *A. solani* was further confirmed with the use of Internal Transcribed Spacer (ITS) sequencing using universal primer pairs ITS1–ITS4. The 556-bp ITS sequence is shown in **Supplementary Figure S2A**. *A. solani* biomass was measured for infected plants of each variety using fungal-specific primer pairs (AS1 and AS2) by qRT-PCR. The methodology of biomass measurement was followed as described by Kumar et al. (2013).

Identification of Putative TIR-NBS-LRR Proteins

Total proteome (39,031) was downloaded from the PGSC database and used for the screening of TIR-NBS domain family (PGSC_DM_v3.4_pep_representative.fasta.zip:// potatogenomics.plantbiology.msu.edu/index.html). A set of putative TIR-NBS proteins was identified from the complete potato proteome using hidden Markov model (HMM) profiles as previously described (Arya et al., 2014). To improve the numbers of TNL proteins left out by HMM application, another strategy was applied that uses pfam ids and conserved domain (CD) information using the CD database (CDD) of the National Center for Biotechnology Information (NCBI) (**Supplementary Figure S1**). All the proteins obtained were also cross-checked using MyHits (Pagni et al., 2004) and ScanProsite (de Castro et al., 2006) along with manual curation to ensure the presence of all the three TNL domains. Each transcript was further assessed for LRR motifs manually, using LxxLxLxxN/CxL or LxxLxL consensus (Enkhbayar et al., 2004), where x denoted any amino acid and L signifies leucine, along with LRR pfam id-0562. To rule out the presence of CC domain, MARCOIL HMM with a threshold of 90 was used (Delorenzi and Speed, 2002). "PGSC_DM_v3.4_g2t2c2p2func.txt" file was used to retrieve information regarding TNLs encoding desired transcripts. To examine the structural motifs among the identified TNLs, the predicted protein sequences were submitted for the motif analysis by MEME version 4.9.0 (Bailey et al., 2009).

Physical Mapping and Gene Duplication Events

The scaffold for all twelve chromosomes was obtained from the PGSC database and used for mapping TNL genes. The General Feature Format (GFF) file (PGSC_DM_v3.4_gene.gff) was also retrieved. These TNL genes were graphically portrayed on the chromosome using "PhenoGram" tool of Ritchie Lab (http:// visualization.ritchielab.psu.edu/phenograms/plot). To check possible duplicated TNLs in potatoes, the MCScanX program (Wang et al., 2012) was applied with all-vs-all blastp using parameters B = 100, V = 10, E-value 1e−10, and tabular output format.

Phylogenetic Analysis of TIR-NBS-LRRs

A phylogenetic tree was prepared using the longest peptide of each TNL gene. The TNL protein sequences were aligned using Clustal W incorporated in MEGA v.7.0 (http://www. megasoftware.net), and a phylogenetic tree was prepared using

the neighbor-joining (NJ) method with a bootstrap of 1,000 iterations (Zhang et al., 2018).

Quantitative Real-Time PCR

About 100 ml of fungal suspension (2×10^5 conidia ml^{-1}) was sprayed on plants. The concentration of conidia and mycelium pieces was calculated using a hemocytometer (Sigma-Aldrich Corp., St. Louis, MO, USA) under a microscope. Sterile water (Mock)-treated plants were used as control. All the plants were kept in a growth chamber with a relative humidity of 70% and temperature of 26°C with 12-h day/night conditions. Leaf samples from each potato cultivars were collected at five time points, *viz*., 1, 2, 3, 5, and 7 days post inoculation (dpi), from three biological replicates of pathogen-inoculated (treatment) and mock-inoculated (control) plants. Total RNA was isolated as previously described (Ghawana et al., 2011) from leaf tissues collected at different time points. The RNA was reverse transcribed using High Capacity cDNA synthesis kit (Applied Biosystems, Foster City, CA, USA). The cDNA was diluted 10-fold with diethyl pyrocarbonate (DEPC)-treated water. Primer Express software version 3.0.1 (Thermo Fisher Scientific, Waltham, MA, USA) was used to design the primers. The primers were designed with an amplicon size range of 80–190 bp, primer length of 18–25 nucleotides, Tm of 50–60°C, and maximum guanine–cytosine (GC) content of 50%–60%. All primer sequences are presented in **Supplementary Table S1**. The Power Sybr Green PCR master mix (ABI) was utilized for the gene expression using an optical 96-well plate on MX3000P real-time PCR machine (Agilent Technologies, Waldbronn, Germany) as per the manufacturer's manual. Elongation factor (*StEf1a*) (accession no. PGSC0003DMG400023270, XM_006343393.2) was selected as reference gene to normalize the targeted gene expression (Tang et al., 2017). The relative expression was calculated by the $2^{-\Delta\Delta CT}$ method (Schmittgen and Livak, 2008). All the experiments were repeated thrice with the inclusion of no template control (NTC) for the detection of DNA contamination.

Localization Assessment of *StTNLC7G2* (*StTNL41857*)

The full-length gene was amplified and cloned in-frame with C-terminal of green fluorescent protein (GFP) in vector pCAMBIA1302 and recombinant vector carrying GFP: *StTNLC7G2* construct was introduced into *Agrobacterium* GV3101 and incubated with onion peel epidermis (Sun et al., 2007). After 2 days, the epidermal layer was thoroughly washed with distilled water and visualized under a fluorescence microscope (Evos, Thermo Fisher Scientific) at 40×.

Functional Characterization of *StTNLC7G2 via* Agroinfiltration

Conidial suspension of *A. solani* was prepared using 0.01% Tween-80, and 50 µl was pre-inoculated (painted) on the proposed infiltration site and dried afterward for 30 min. The gene was cloned in pCAMBIA1302 using *Nco*I and *Bst*EII (**Supplementary Figure S2B**) and confirmed by sequencing (accession no. MZ597472). The positive *Agrobacterium* transformants were used for culturing in 200 ml of yeast extract peptone broth (YEP) media supplemented with rifampicin (25 µg/ml) and kanamycin (50 µg/ml) at 28°C to an OD_{600} of ~1.0. The culture was further centrifuged at 4,000×*g* for 10 min, and the pellet was resuspended in 10 mM of MES [2-(*N*-morpholino)ethanesulfonic acid] pH 5.6 and 10 mM of $MgCl_2$ with 100 µM of acetosyringone (Bhaskar et al., 2008) and incubated for 4–8 h at room temperature (RT) in the dark, and 200 µl of the inoculum was used for infiltration of 30-day-old potato leaves of KJ, KC, and KP. The *Agrobacterium* transformed with a blank vector was used as vector control. The inoculated plants were kept in the dark for 12 h, then shifted in a plant growth chamber with conditions described earlier, and sampled at 3 dpi. For DAB staining, 50 mg/100 ml of working stock was prepared in 0.1 M of sodium phosphate buffer at pH 3.0 and was maintained by using concentrated HCl. Earlier collected samples were now dipped into the staining solution and incubated at RT for 6–8 h at 100 rpm. Then leaves were decolorized using the bleaching solutions (ethanol:acetic acid: glycerol, 3:1:1) for 30 min at 90°C, then replaced with fresh bleaching solutions, and photographed for any brown color appearance denoting ROS (Liu et al., 2014). Gene expression after agroinfiltration was confirmed in *N. benthamiana* leaves collected at 5 dpi. Samples were collected by an incision measuring $1 \times 1\ cm^2$ using a razor blade surrounding the infiltration zones (2–3 cm away from the infiltration zones). RT-PCR was performed as previously described (Bhaskar et al., 2008).

RESULTS

TIR-NBS-LRR Genes Were Identified Encoding 60 Full-Length Peptides

The application of Hmmsearch method using Pf00931-pfam domain (NBS) resulted in 644 NBS coding transcripts. The application of TIR Pfam id-1582 within these transcripts narrowed down the result to 110 TIR-NBS-containing transcripts. After assessment of LRR motifs, in total, 55 transcripts were identified containing all the three domains, namely, TIR, NBS, and LRR. As there is always a possibility to skip the detection of the targeted domain by HMM profile due to partial or truncated sequences, a second strategy of CD search using the entire proteome set from the PGSC database was submitted on the NCBI CDD search engine. The strategy provided a total of 60 full-length TNL peptides with a high confidence level for TIR, NBS, and LRR domain encoded by 44 TNL genes (**Supplementary Table S2**). When these 44 genes were used to identify all possible coded peptides, a total of 36 additional transcripts were identified including 29 truncated sequences. Twenty-seven genes were found to code for more than 1 transcript probably by alternative splicing events (**Supplementary Table S3**). In MEME analysis, conserved and signature motifs of the TIR domain and NBS domain were identified to be preserved (**Supplementary Table S4**).

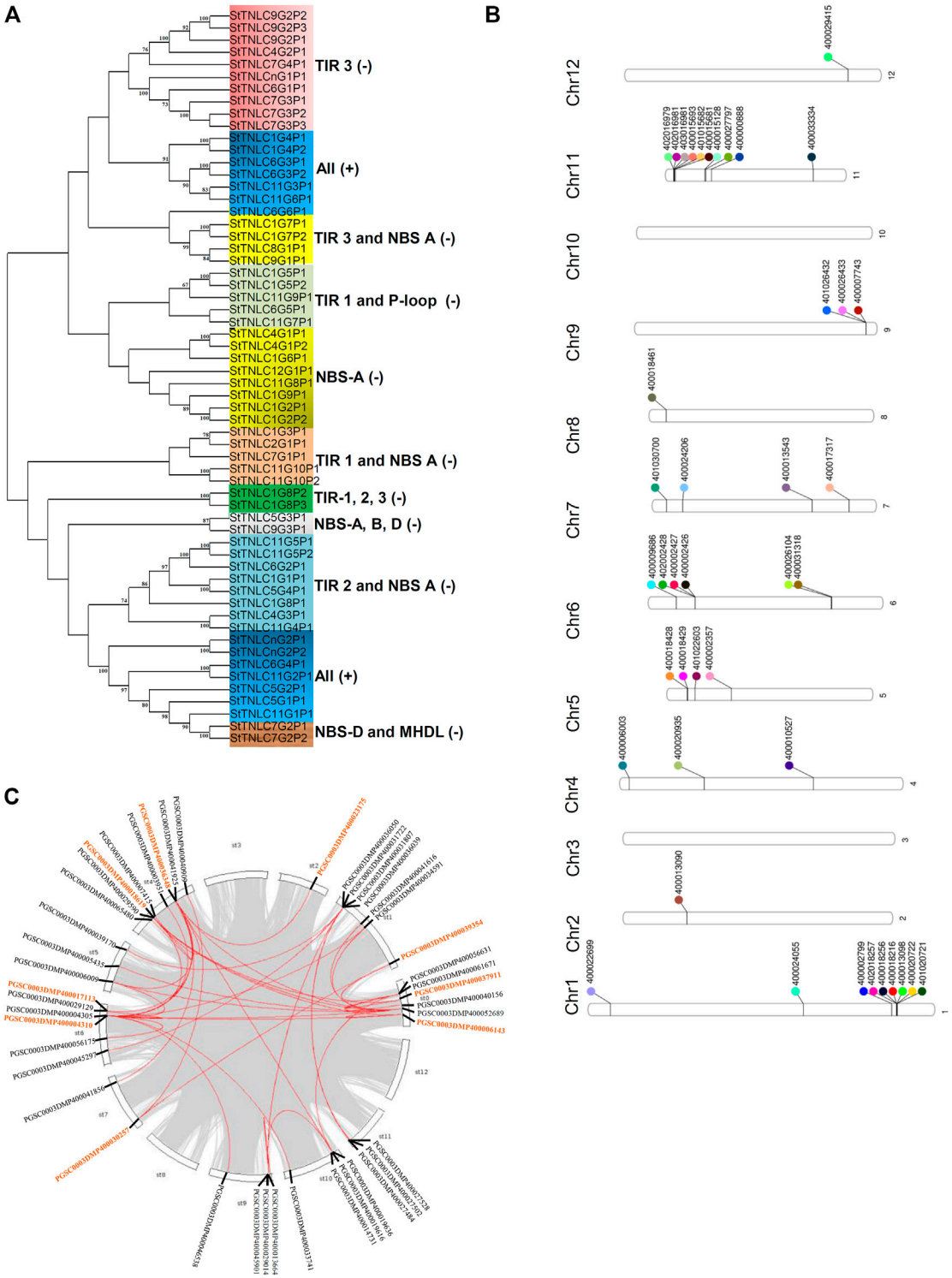

FIGURE 1 | (A) Phylogenetic tree of 60 TIR-NBS-LRR (TNL) peptides showing clades getting grouped and demarcated based on the presence of TIR and NBS motifs. **(B)** Phenogram representing the position of 44 TNLs on chromosomes of *Solanum tuberosum*. **(C)** *StTNL* collinear gene pairs within a genome are represented with red lines, while the genes involved in *StTNL–StTNL* duplication events are represented with orange color in the text.

Novel Methodology for Gene Family Member Nomenclature

Nomenclature has been assigned to these TNL genes for ease of understanding with an effort to bring uniformity of methodology in such study after retrieving relevant information in a smooth and easy manner. Here, the nomenclature of individual genes and their respective proteins is based on their physical location on the chromosome. For instance, *PGSC0003DMG400022699* localized on chromosome 1, at the first position, so the nomenclature of this gene and encoded protein is *StTNLC1G1* and StTNLC1G1P1, respectively. If a gene encodes more than one transcript, the nomenclature used here is based on their peptide length in descending order. For example, *PGSC0003DMG400024055* encodes 2 peptides: PGSC0003DMP400041617 and PGSC0003DMP400041616. The nomenclature of these peptides will be StTNLC1G2P1 and StTNLC1G2P2, respectively (**Supplementary Table S2**).

Phylogenetic Analyses

Phylogenetic grouping of the genes from the same chromosomes and the separate chromosomes was observed. Exon–intron structure tends to remain the same among genes in the same clade, indicating significant gene structure conservation over evolution, according to a phylogenetic tree–gene structure correlation. Clades were also formed on the basis of the presence and the absence of a particular motif in proteins and homology between them. Two major clades were formed, i.e., Clade I and Clade II, in which TNLs were demarcated on the basis of the domain and motif compositions (**Figure 1A**).

Chromosomal Localization of *StTNL* Genes

The physical location of the TNL genes was resolved on the basis of PGSC_DM_v3.4_gene.gff file as retrieved from the PGSC database. Forty-two out of forty-four TNL genes were successfully mapped on the twelve potato chromosomes using Ritchie Lab phenogram, while only 2 sequences remained unanchored. The chromosomal location of these genes has shown the uneven distribution of TNLs on different chromosomes. Two gene clusters were found, one each on chromosomes 1 and 11 (**Figure 1B**), both having 6 TNL genes. The number of TNLs mapped on chromosomes 1, 2, 4, 5, 6, 7, 8, 9, 11, and 12 was 9, 1, 3, 4, 6, 4, 1, 3, 10, and 1, respectively. There were no TNL genes found on chromosomes 3 and 10.

Duplication Events Observed

Across the potato genome, there were 7 pairs of TNL-TNL segmental duplication events contributed by a total of 9 TNL genes. *StTNL36348* was found to be involved in three TNL-TNL events with pairing to *StTNL17113*, *StTNL18619*, and *StTNL23175* in the genome. Two genes, *StTNL37911* and *StTNL4310*, are probably segmentally duplicated at least six times, including non-TNL duplications. The search for duplication events of TNL vs. other genes identified 35 duplication events/pair. Collinear pair was found on all chromosomes except chromosomes 3, 8, and 12 (**Figure 1C**). Only one gene, *StTNL45901*, physically located on chr9 was observed to be tandemly duplicated (**Supplementary Table S5**).

Expression of Multiple *StTNL* Genes Induced by Fungal Pathogen *Alternaria solani*

Before proceeding for expression assessment of transcripts, the virulence of *A. solani* was confirmed by visualization of symptoms using pathogen spray to plants in pots (**Supplementary Figure S4A**) and detached leaf assay (**Supplementary Figure S4B**) at 5 dpi. Both the experiments confirmed the susceptibility of KC to virulent *A. solani*, while KJ and KP showed tolerance. The results were even further supplemented with fungal biomass assessment (**Supplementary Figure S4C**). Fourteen genes were chosen randomly for expression analysis, but also keeping in mind that at least one gene from each chromosome may be represented. The differential expression pattern obtained for different transcripts can be summed up in three major patterns for TNL transcripts. The first groups of transcripts have their expression level constantly reaching more than 4.0-fold in expression across all the three varieties (**Supplementary Tables S6–S8**). Such expression pattern was obtained for seven transcripts, namely, *StTNLC1G3T1*, *StTNLC6G4T1*, *StTNLC7G1T1*, *StTNLC9G1T1*, *StTNLC11G9T1*, *StTNLC12G1T1*, and *StTNLCnG2T1*. All of them showed more than 10-fold expression in at least one disease-tolerant variety with the exception being *StTNLCnG2T1* and *StTNLC7G1T1* (**Figure 2A**). Strong and constant upregulation under pathogen treatment suggests that their expression modulates under pathogenic signals like pathogen-associated molecular pattern (PAMP). A few transcripts showed a difference in expression pattern between tolerant and susceptible varieties. Transcripts with such expression behavior show a constant decrease in expression in KC after the initial spike at 1 or 3 dpi. These transcripts were *StTNLC6G2T1*, *StTNLC7G2T1*, *StTNLC9G3T1*, *StTNLC11G7T1*, and *StTNLC11G8T1*. The expression of all these transcripts was decreased to <2.5-fold in KC at 7 dpi while maintaining >6.0-fold expression in KP and KJ.

Meanwhile, two transcripts showed expression of >15.00-fold under *A. solani* infection. These transcripts were *StTNLC6G2T1* and *StTNLC11G9T1*. The difference in expression value of these two genes at 7 dpi was almost 4 times in the comparison of KJ with KC. *StTNLC6G2T1* and *StTNLC11G9T1* must have prominent roles in plant defense under *A. solani* attack at a late stage. For the majority of the genes tested, the expression enhancement pattern was almost similar for KJ and KP with only a significant difference being obtained in *StTNLC12G1T1*. Furthermore, among all the candidates tested, the expression pattern of only one transcript (*StTNLC7G1T1*) was found to be constant across the varieties and time points.

StTNLC7G2 Induced More in Tolerant Variety and Localizes to Plasma Membrane

One gene, *StTNLC7G2*, showing homology with *N* gene of tobacco was selected for further characterization, as its transcript induction was greater than 10-fold at 5 and 7 dpi

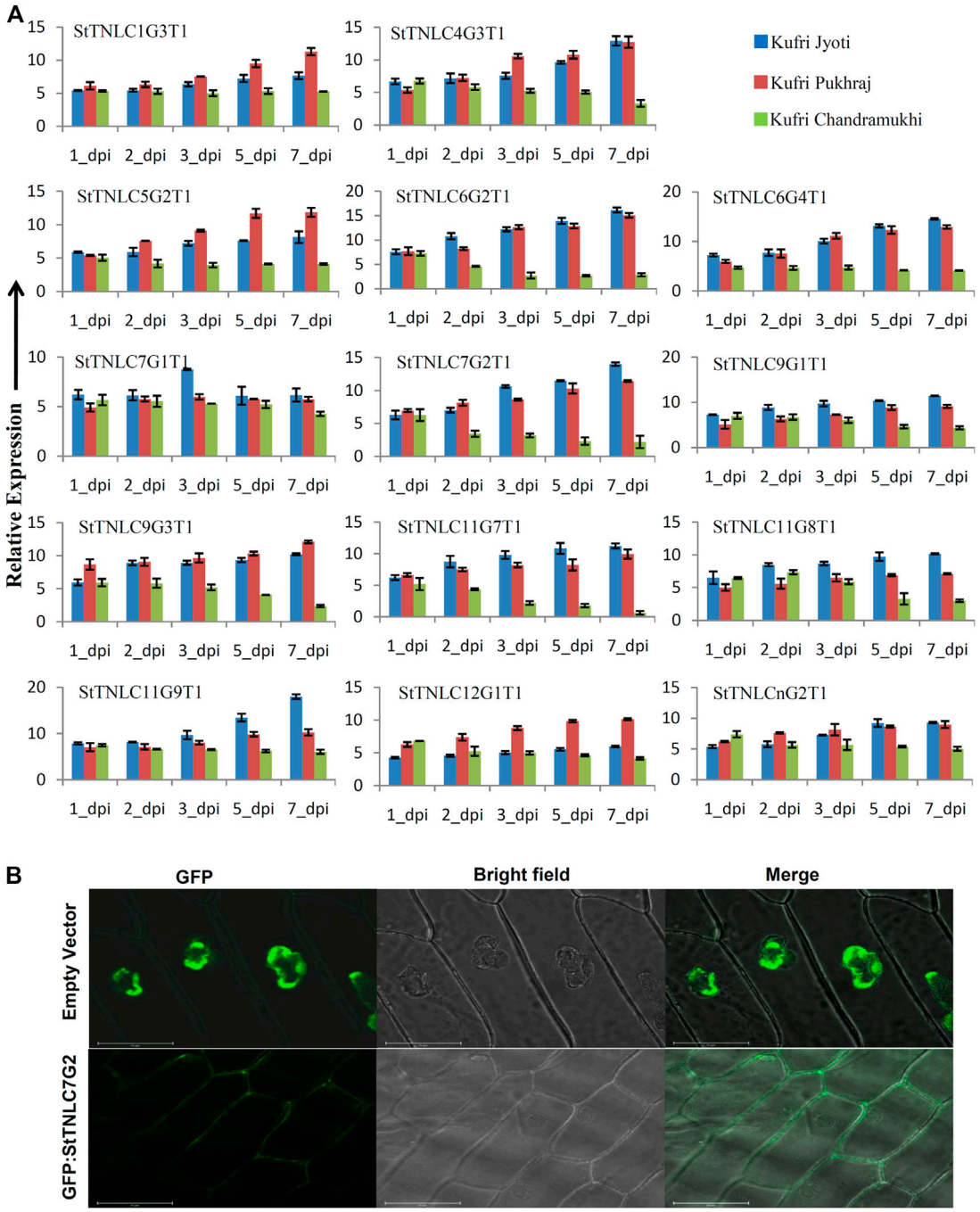

FIGURE 2 | (A) Differential expression of 14 TIR-NBS-LRRs (TNLs) following *Alternaria solani* infection. The expression values were normalized with *StELF1α* as an internal control. Relative expression values of each transcript were assessed in three varieties at five time points. The error bar shows the standard error (SE) with $n = 3$. **(B)** Fluorescence microscopy observation of green fluorescent protein (GFP): *StTNLC7G2* construct after expression in onion leaf epidermal cell. Vector with GFP only was taken as a control for localization comparison and assessment.

in KJ and KP, respectively, while expression in KC barely crosses 2-fold with respect to control, though immediately after pathogen inoculation at 1 dpi, the expression was similar in all three varieties. To check the localization of StTNLC7G2, the transcript was amplified and cloned in pCAMBIA1302 in the same frame as of GFP, and positive clones were used to inoculate onion peel. Amplification and cloning of full-length *StTNLC7G2* achieved 1,296 bp of amplicon encoding a protein of 431 amino acids (**Supplementary Figure S2C**). The localization of construct was compared with vector control (GFP only) that clearly showed nucleus localization, while StTNLC7G2 protein showed putative membrane

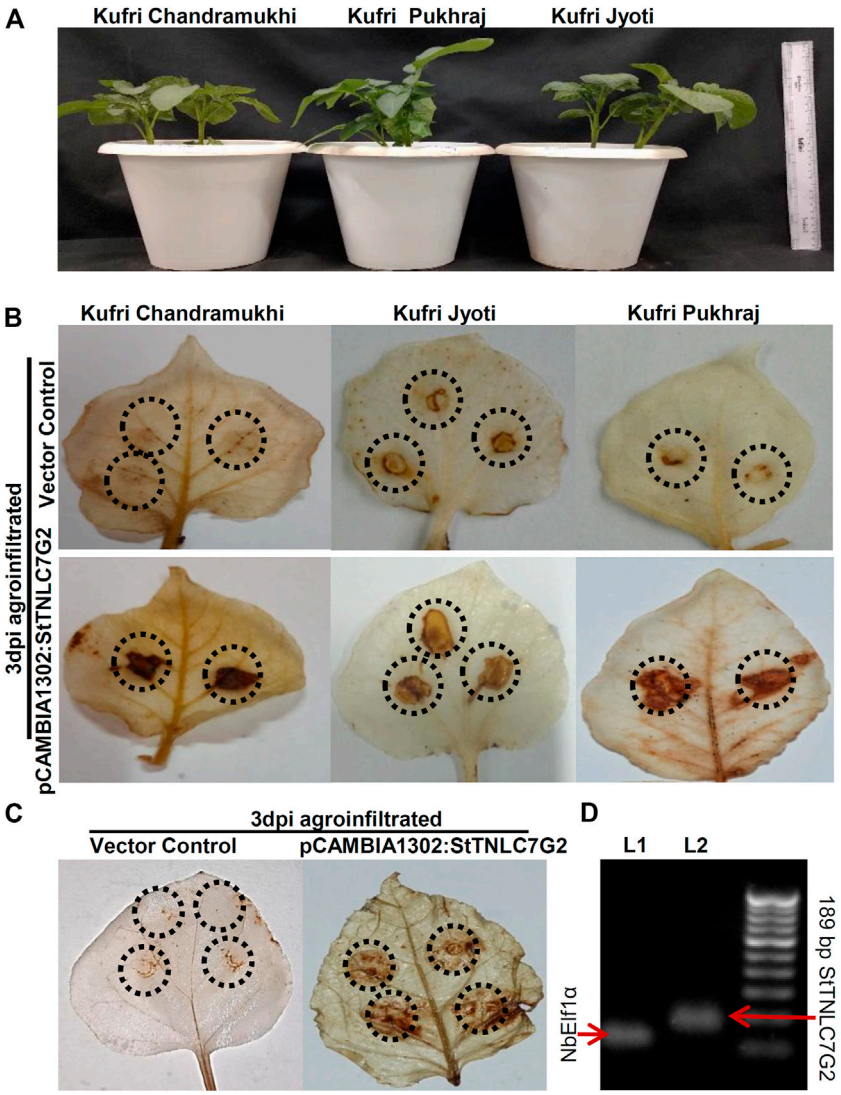

FIGURE 3 | Agroinfiltration assays of *StTNLC7G2* gene in potato plants. **(A)** One-month-old potato plants selected for each variety for agroinfiltration with nearly the same size. **(B)** Leaf images showing dark patches and spots at 3 dpi after DAB staining indicating induction of reactive oxygen species in response to *StTNLC7G2* under *Alternaria solani* attack in three varieties of potato. **(C)** Confirmation of reactive oxygen species generation under *StTNLC7G2* influence in *Nicotiana benthamiana* in comparison with vector control. **(D)** Gel image showing high expression of *StTNLC7G2* transcripts L2 (lane 2: 189 bp) from *Nicotiana* leaves 5-days post agroinfiltration. In L1 (lane 1: 117 bp), *NbElf1α* was taken as internal control.

localization (**Figure 2B**). StTNLC7G2 protein is probably involved in the reception of the pathogenic signal and their further transmission downstream. The high expression of the transcript at 7 dpi in a tolerant variety further indicates the same.

StTNLC7G2 Induces Reactive Oxygen Species Under *Alternaria solani* Attack

After inoculation of the fungal pathogen, *A. solani* suspension in the leaves, agroinfiltration was performed with both positive gene construct and vector control at the same site and left for visual observation followed by sampling. DAB staining of 3 dpi agroinfiltrated leaves showed dark patches in all the three varieties at 3 dpi (**Figure 3A**), indicating the prominent generation of ROS in the presence of StTNLC7G2 mimicking HR (**Figure 3B**; **Supplementary Figure S4D**). KJ showed light-dark patches in mock-treated (empty vector) leaves too, suggesting a stronger defense mechanism available to its genotype. An increase in ROS induction in the presence of StTNLC7G2 was more prominent in KP and KC. As ROS induction is the primary feature of HR during plant defense, the *StTNLC7G2* is probably co-related with innate immunity against fungal pathogen *A. solani*. The experiment was further conducted in *N. benthamiana* leaves with a similar result of ROS induction at 3 dpi (**Figure 3C**). Semi-quantitative RT-PCR confirmed the expression of *StTNLC7G2* in ROS-induced leaves (**Figure 3D**).

DISCUSSION

TNL genes are a subset of the NBS-LRR gene family. In the current work, a comprehensive analysis to identify the TNL members in potatoes utilizing *S. tuberosum* phureja genome resulted in 44 unique genes encoding 60 TNLs and 96 transcripts overall (including pseudogenes). The majority of genes was encoding more than one transcript. TNL gene frequency varies greatly across plant species, ranging from 0.27% in apples (Arya et al., 2014) to 0.000667% in common beans (Wu et al., 2017), while here we obtained a frequency of 0.117% in potatoes. Synteny and duplication assessment showed that all the genes are present in a single copy number except one (involved in tandem duplication). Probably higher fitness cost and more available transcripts per gene have reduced the duplication event occurrence. The analysis also identified the presence of the LRR domain in StTNLC7G2, based on CDD search that was omitted in the whole-genome analysis previously reported (Lozano et al., 2012). The physical distribution of *StTNL* genes identified two clusters at chr1 and chr11. As TNLs are known to form heterodimers to regulate the plant defense responses (Ashikawa et al., 2008), clustering of genes may help them to modulate simultaneously and co-expressed, if required. It has been reported that R-gene evolution is governed by cluster size, complexity, and duplication events (Friedman and Baker, 2007).

On average, TNL peptides were ~500 amino acids in length with a coding sequence ~3 kb. Exon–intron structures of these TNL varied significantly, and all of them were inclusive of introns (**Supplementary Figure S3**). Exons–introns exhibit a variety of structures that result in gene structural divergence due to three mechanisms: exon and intron distribution, insertion/deletion, and exonization/pseudo-exonization resulting in their gain or loss (Roy and Gilbert, 2005). In the present study, many transcripts were identified originating from a putative single gene, probably as a result of alternative splicing events due to the availability of multiple introns–exons. In one case, *StTNLC1G8* is encoding three transcripts (*StTNLG8T1–StTNLG8T3*). Such alternative splicing events have been reported in *Arabidopsis RPS4* (Borhan et al., 2004) and tobacco N (Whitham et al., 1994) due to the higher number of exons in their gene structure.

Many TNL transcripts showed higher expression in the presence of *A. solani*, including *StTNLC11G2*, which was also expressed in the early stages of resistant potato cultivar Beate against *Streptomyces turgidiscabies* (Dees et al., 2016). Another gene *StTNLC7G2* has been shown to express more in the resistant 'Premier Russet' variety of potatoes against PVY strain-O as compared with the susceptible 'Russet Burbank' variety (Goyer et al., 2015). The homolog of *StTNLC7G1* has been shown to upregulate in *Solanum cardiophyllum* and *Solanum pinnatisectum* against late blight attack in another case study (Gu et al., 2020).

The activation of H_2O_2-dependent responses in Solanaceae plants has been shown to successfully prevent the early growth of necrotrophic pathogens such as *Botrytis cinerea* (Asselbergh et al., 2007). Here, agroinfiltration of *StTNLC7G2* showed similar induction of ROS in response to the fungal pathogen as visualized by DAB staining, indicative of HR. Henceforth, all the data suggest functional involvement of *StTNLC7G2* in ROS-mediated plant innate immunity.

CONCLUSION

The potato genome revealed the presence of 44 TNL genes. Chromosome 1 and chromosome 11 carried one TNL cluster each. Many TNL genes including *StTNLC7G2* were found to be getting modulated under pathogen attack with much higher expression in a tolerant variety. StTNLC7G2 provides resistance to the pathogen by inducing ROS at the site of the attack. The mechanism showed *StTNLC7G2* having a role in potato innate immunity. This work is the first attempt to understand the structural and functional insight of TNLs along with their characterization in the presence of *A. solani*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

KS designed and conceived the experiments and wrote the manuscript. ND performed the experiment, analyzed the data, and wrote the manuscript. AC performed the experiment and analyzed the data.

FUNDING

This work was financially supported by Science and Engineering Research Board (SERB) (File No. YSS/2015/001036; Project code: GAP0205) between 2016 and 2019.

ACKNOWLEDGMENTS

The authors are thankful to the Director of CSIR-IHBT for providing a working facility. ND and AC are thankful to the University Grant Commission, India, and Council of Scientific and Industrial Research, India, for providing financial assistance in the form of fellowship. The IHBT communication number for this manuscript is 10080.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.791055/full#supplementary-material

REFERENCES

Arya, P., Kumar, G., Acharya, V., and Singh, A. K. (2014). Genome-wide Identification and Expression Analysis of NBS-Encoding Genes in *Malus X Domestica* and Expansion of NBS Genes Family in Rosaceae. *PLoS One* 9, e107987. doi:10.1371/journal.pone.0107987

Ashikawa, I., Hayashi, N., Yamane, H., Kanamori, H., Wu, J., Matsumoto, T., et al. (2008). Two Adjacent Nucleotide-Binding Site-Leucine-Rich Repeat Class Genes Are Required to Confer Pikm-specific Rice Blast Resistance. *Genetics* 180, 2267–2276. doi:10.1534/genetics.108.095034

Asselbergh, B., Curvers, K., França, S. C., Audenaert, K., Vuylsteke, M., Van Breusegem, F., et al. (2007). Resistance to *Botrytis Cinerea* in Sitiens, an Abscisic Acid-Deficient Tomato Mutant, Involves Timely Production of Hydrogen Peroxide and Cell wall Modifications in the Epidermis. *Plant Physiol.* 144, 1863–1877. doi:10.1104/pp.107.099226

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* 37, W202–W208. doi:10.1093/nar/gkp335

Bhaskar, P. B., Raasch, J. A., Kramer, L. C., Neumann, P., Wielgus, S. M., Austin-Phillips, S., et al. (2008). *Sgt1*, but Not *Rar1*, Is Essential for the RB-Mediated Broad-Spectrum Resistance to Potato Late Blight. *BMC Plant Biol.* 8, 1–9. doi:10.1186/1471-2229-8-8

Borhan, M. H., Holub, E. B., Beynon, J. L., Rozwadowski, K., and Rimmer, S. R. (2004). The Arabidopsis TIR-NB-LRR Gene *RAC1* Confers Resistance to *Albugo candida* (White Rust) and Is Dependent on *EDS1* but not *PAD4*. *Mpmi* 17, 711–719. doi:10.1094/MPMI.2004.17.7.711

Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. (2006). Host-microbe Interactions: Shaping the Evolution of the Plant Immune Response. *Cell* 124, 803–814. doi:10.1016/j.cell.2006.02.008

Choi, Y. W., Hyde, K. D., and Ho, W. H. (1999). Single Spore Isolation of Fungi. *Fungal Divers* 3, 29–38.

De Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., et al. (2006). ScanProsite: Detection of PROSITE Signature Matches and ProRule-Associated Functional and Structural Residues in Proteins. *Nucleic Acids Res.* 34, W362–W365. doi:10.1093/nar/gkl124

Dees, M. W., Lysøe, E., Alsheikh, M., Davik, J., and Brurberg, M. B. (2016). Resistance to *Streptomyces Turgidiscabiesin* Potato Involves an Early and Sustained Transcriptional Reprogramming at Initial Stages of Tuber Formation. *Mol. Plant Pathol.* 17, 703–713. doi:10.1111/mpp.12323

Delorenzi, M., and Speed, T. (2002). An HMM Model for Coiled-Coil Domains and a Comparison with PSSM-Based Predictions. *Bioinformatics* 18, 617–625. doi:10.1093/bioinformatics/18.4.617

Dubey, N., and Singh, K. (2018). "Role of NBS-LRR Proteins in Plant Defense," in *Molecular Aspects of Plant-Pathogen Interaction* (Singapore: Springer), 115–138. doi:10.1007/978-981-10-7371-7_5

Enkhbayar, P., Kamiya, M., Osaki, M., Matsumoto, T., and Matsushima, N. (2004). Structural Principles of Leucine-Rich Repeat (LRR) Proteins. *Proteins* 54, 394–403. doi:10.1002/prot.10605

Friedman, A. R., and Baker, B. J. (2007). The Evolution of Resistance Genes in Multi-Protein Plant Resistance Systems. *Curr. Opin. Genet. Dev.* 17, 493–499. doi:10.1016/j.gde.2007.08.014

Ghawana, S., Paul, A., Kumar, H., Kumar, A., Singh, H., Bhardwaj, P. K., et al. (2011). An RNA Isolation System for Plant Tissues Rich in Secondary Metabolites. *BMC Res. Notes* 4, 1–5. doi:10.1186/1756-0500-4-85

Goyer, A., Hamlin, L., Crosslin, J. M., Buchanan, A., and Chang, J. H. (2015). RNA-seq Analysis of Resistant and Susceptible Potato Varieties during the Early Stages of Potato Virus Y Infection. *BMC Genomics* 16, 1–13. doi:10.1186/s12864-015-1666-2

Gu, B., Cao, X., Zhou, X., Chen, Z., Wang, Q., Liu, W., et al. (2020). The Histological, Effectoromic, and Transcriptomic Analyses of *Solanum Pinnatisectum* Reveal an Upregulation of Multiple NBS-LRR Genes Suppressing *Phytophthora Infestans* Infection. *Ijms* 21, 3211. doi:10.3390/ijms21093211

Jones, J. D. G., and Dangl, J. L. (2006). The Plant Immune System. *Nature* 444, 323–329. doi:10.1038/nature05286

Kumar, S., Singh, R., Kashyap, P. L., and Srivastava, A. K. (2013). Rapid Detection and Quantification of *Alternaria solani* in Tomato. *Scientia Horticulturae* 151, 184–189. doi:10.1016/j.scienta.2012.12.026

Liu, Y.-H., Offler, C. E., and Ruan, Y.-L. (2014). A Simple, Rapid, and Reliable Protocol to Localize Hydrogen Peroxide in Large Plant Organs by DAB-Mediated Tissue Printing. *Front. Plant Sci.* 5, 745. doi:10.3389/fpls.2014.00745

Lozano, R., Ponce, O., Ramirez, M., Mostajo, N., and Orjeda, G. (2012). Genome-wide Identification and Mapping of NBS-Encoding Resistance Genes in *Solanum tuberosum* Group Phureja. *PLoS One* 7, e34775. doi:10.1371/journal.pone.0034775

Mchale, L., Tan, X., Koehl, P., and Michelmore, R. W. (2006). Plant NBS-LRR Proteins: Adaptable Guards. *Genome Biol.* 7, 1–11. doi:10.1186/gb-2006-7-4-212

Nandety, R. S., Caplan, J. L., Cavanaugh, K., Perroud, B., Wroblewski, T., Michelmore, R. W., et al. (2013). The Role of TIR-NBS and TIR-X Proteins in Plant Basal Defense Responses. *Plant Physiol.* 162, 1459–1472. doi:10.1104/pp.113.219162

Padmanabhan, C., Zhang, X., and Jin, H. (2009). Host Small RNAs Are Big Contributors to Plant Innate Immunity. *Curr. Opin. Plant Biol.* 12, 465–472. doi:10.1016/j.pbi.2009.06.005

Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C. V., and Falquet, L. (2004). MyHits: a New Interactive Resource for Protein Annotation and Domain Identification. *Nucleic Acids Res.* 32, W332–W335. doi:10.1093/nar/gkh479

Qi, D., and Innes, R. W. (2013). Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. *Front. Immunol.* 4, 348. doi:10.3389/fimmu.2013.00348

Roy, S. W., and Gilbert, W. (2005). Rates of Intron Loss and Gain: Implications for Early Eukaryotic Evolution. *Proc. Natl. Acad. Sci.* 102, 5773–5778. doi:10.1073/pnas.0500383102

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing Real-Time PCR Data by the Comparative CT Method. *Nat. Protoc.* 3, 1101–1108. doi:10.1038/nprot.2008.73

Sun, W., Cao, Z., Li, Y., Zhao, Y., and Zhang, H. (2007). A Simple and Effective Method for Protein Subcellular Localization Using Agrobacterium-Mediated Transformation of Onion Epidermal Cells. *Biologia* 62, 529–532. doi:10.2478/s11756-007-0104-6

Tang, X., Zhang, N., Si, H., and Calderón-Urrea, A. (2017). Selection and Validation of Reference Genes for RT-qPCR Analysis in Potato under Abiotic Stress. *Plant Methods* 13, 85. doi:10.1186/s13007-017-0238-7

Vloutoglou, I., and Kalogerakis, S. N. (2000). Effects of Inoculum Concentration, Wetness Duration and Plant Age on Development of Early Blight (*Alternaria solani*) and on Shedding of Leaves in Tomato Plants. *Plant Pathol.* 49, 339–345. doi:10.1046/j.1365-3059.2000.00462.x

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40, e49. doi:10.1093/nar/gkr1293

Whitham, S., Dinesh-Kumar, S. P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994). The Product of the Tobacco Mosaic Virus Resistance Gene *N*: Similarity to Toll and the Interleukin-1 Receptor. *Cell* 78, 1101–1115. doi:10.1016/0092-8674(94)90283-6

Wu, J., Zhu, J., Wang, L., and Wang, S. (2017). Genome-wide Association Study Identifies NBS-LRR-Encoding Genes Related with Anthracnose and Common Bacterial Blight in the Common Bean. *Front. Plant Sci.* 8, 1398. doi:10.3389/fpls.2017.01398

Zhang, Z., Wei, X., Liu, W., Min, X., Jin, X., Ndayambaza, B., et al. (2018). Genome-wide Identification and Expression Analysis of the Fatty Acid Desaturase Genes in Medicago Truncatula. *Biochem. Biophysical Res. CommunicationsBiochem. Biophys. Res. Commun.* 499, 361–367. doi:10.1016/j.bbrc.2018.03.165

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.

frontiers in Plant Science

MINI REVIEW
published: 28 January 2022
doi: 10.3389/fpls.2021.787590

Chromosome Evolution in the Family Solanaceae

Rocío Deanna[1,2†], María Cristina Acosta[1], Marisel Scaldaferro[1] and Franco Chiarini[1†]*

[1] Instituto Multidisciplinario de Biología Vegetal (CONICET-UNC), Córdoba, Argentina, [2] Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO, United States

This review summarizes and discusses the knowledge of cytogenetics in Solanaceae, the tomato family, its current applications, and prospects for making progress in fundamental systematic botany and plant evolution. We compile information on basic chromosome features (number, size, morphology) and molecular cytogenetics (chromosome banding and rDNA patterns). These data were mapped onto the Solanaceae family tree to better visualize the changes in chromosome features and evaluate them in a phylogenetic context. We conclude that chromosomal features are important in understanding the evolution of the family, especially in delimiting clades, and therefore it is necessary to continue producing this type of data. The potential for future applications in plant biology is outlined. Finally, we provide insights into understanding the mechanisms underlying Solanaceae's diversification that could substantially contribute to developing new approaches for future research.

Keywords: comparative methods, cytogenetics, karyotype, nightshades, phylogeny

OPEN ACCESS

Edited by:
Peter Poczai,
University of Helsinki, Finland

Reviewed by:
Sonia Garcia,
Botanical Institute of Barcelona, CSIC, Spain
Jaume Pellicer,
Spanish National Research Council (CSIC), Spain

****Correspondence:***
Franco Chiarini
chiarini@imbiv.unc.edu.ar;
franco.e.chiarini@gmail.com

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *30 September 2021*
Accepted: *22 December 2021*
Published: *28 January 2022*

Citation:
Deanna R, Acosta MC, Scaldaferro M and Chiarini F (2022) Chromosome Evolution in the Family Solanaceae. Front. Plant Sci. 12:787590. doi: 10.3389/fpls.2021.787590

INTRODUCTION

Chromosomes provide information for inferring phylogenetic relationships, since they are hereditary elements of the whole nuclear genome and discrete hereditary units of mutation. Within taxa, they may vary in number, size, morphology, and staining properties (Weiss-Schneeweiss and Schneeweiss, 2013). Chromosome number has always been a common character employed, since it is the most easily obtained information and the only one that is known for most plant groups. Karyotypes represent an important aspect in plant speciation since chromosomal differences establish immediate postzygotic crossing barriers. They provide diagnostic characters for plant systematics and evolution that are usually expected to be congruent with clade divergence (e.g., Baltisberger and Hörandl, 2016). In Solanaceae, chromosomes have played a relevant role: they have been useful to delimit taxa (Camadro et al., 2012), for genetic studies with model species (Blakeslee et al., 1920; Avery et al., 1959), and for the selection of cultivars in economically important species (Pringle and Murray, 1992). With the advent of molecular phylogenetic techniques, chromosome data acquired a new value, to the point that within the family, there is a major lineage named as the "x = 12 clade," whose members share such cytological synapomorphy (Olmstead et al., 2008). This review summarizes and discusses the knowledge of cytogenetics in Solanaceae, its current applications and prospects for making progress in fundamental systematic botany and plant evolution. Available chromosome data were reassessed by employing them in ancestral state reconstruction and thus revealing how many times traits changed over evolutionary time, inferring the degree of convergence, and suggesting relationships between chromosomes and other Solanaceae characteristics. Continuous cytogenetic characters (C, mean chromosome length; r,

mean arm ratio or r index; TL, total haploid chromosome length; and m_ratio, the ratio between the number of metacentric to the total number of chromosomes) were mapped and plotted onto the family phylogeny of Särkinen et al. (2013) with modifications by Dupin et al. (2017), after removing those species without data using the drop.tip function of the package {ape} in R (Paradis et al., 2004). The ancestral states were estimated assuming that the species evolved under a Brownian model and the mapping was performed using the ContMap function of the package {phytools} in R version 3.4.2 (Revell, 2012; **Supplementary Figures 1–4**). All the methodologies to gather the data are in **Supplementary Table 2**.

CHROMOSOME COUNTS

Knowing the basic number *x* for all Angiosperms and its clades has been a recurring question (Darlington, 1956). Initially, it was considered of little taxonomic utility at high hierarchical levels (orders, families) due to homoplasy. For Solanales, Raven (1975) proposed $x = 7$ while $x = 12$ may have derived from the tetraploid level ($n = 14$) by aneuploid reduction in the early history of Solanaceae. Acosta et al. (2006) suggested $x = 11$ as the ancestral number that best fits the available phylogenetic hypotheses (Olmstead et al., 1999). Recently, models of chromosome evolution were customized using statistical frameworks, allowing to leverage other tools, like Ancestral State Reconstruction (ASR, FitzJohn, 2012), and to infer character evolution across phylogenetic trees (Revell, 2012). The compilation and analysis of data in a phylogenetic context allow us to recalculate the first estimates and reconstruct the history of the changes. Although the basic number x is homoplastic in Solanaceae, it is more conserved than other karyotype features, which makes it useful to delimit taxa at lower hierarchical levels: tribes, subtribes, genera (e.g., Rodríguez et al., 2020). The availability of chromosome data is essential to delve into these analyses, but the information is still scattered and fragmentary (**Supplementary Tables 1**, **2**). Among the five largest genera of Solanaceae, the best studied (*Solanum*) has 48% of its species counted. There is a bias toward groups of species with economically valuable representatives (e.g., *Capsicum*, *Nicotiana*), while small clades and monotypic genera are karyologically unknown (**Supplementary Tables 1**, **2**). Applying model approaches to chromosome number would clarify this topic (Glick and Mayrose, 2014), although such methodologies have difficulties given the size of the matrix for the entire family (**Supplementary Table 1**) and discussing such downsides is beyond the scope of this work.

Solanaceae have undergone whole genome duplications occurring near the time of its origin, which could have contributed to the rise of key traits and drove species diversification (Schranz et al., 2012). Polyploidy, a phenomenon constantly reviewed (e.g., Van de Peer et al., 2021), because of the advantages it may confer, has been a subject since the beginning of cytogenetics in Solanaceae (e.g., Blakeslee et al., 1920) up to the present (e.g., Hijmans et al., 2007; Chiarini et al., 2018). Nowadays, the availability of data makes it possible to analyze polyploidy throughout the evolutionary history of the family and relate it to colonization of new habitats, speciation events and to the rising of adaptive traits (Hijmans et al., 2007; Scaldaferro et al., 2012; Zenil-Ferguson et al., 2018, 2019; McCarthy et al., 2019).

KARYOTYPES

Blakeslee's pioneering work on *Datura* mutants (Blakeslee et al., 1920, and subsequent papers) allowed to identify pairs of homologs by their shape, and correlate alterations in meiotic behavior to exomorphology. The chromosome morphology observable with classical technique is still essential to detect rearrangements involved in speciation, for example, detection of allopolyploids in *Nicotiana* (Goodspeed, 1954), Robertsonian translocations in *Solanum* (Chiarini and Bernardello, 2006) and in *Chamaesaracha* (Rodríguez et al., 2020), and centric fission in *Capsicum* (Jarret et al., 2019). In contrast, uniform chromosome morphology may be the explanation for crossability among Iochrominae species (Deanna et al., 2018) and suggests the phenomenon of "karyotype orthoselection" in *Lycieae* (Bernardello et al., 2008). Karyotype data integrated with molecular phylogenies provide characters for Solanaceae systematics and evolution (e.g., *Nierembergia*: Acosta et al., 2016; *Jaborosa*, Chiarini et al., 2017; *Iochroma*, Deanna et al., 2018; *Physalis*, Rodríguez et al., 2020). However, the distribution of cytological studies is uneven among the genera of the family (**Supplementary Table 2**). We registered 420 species of Solanaceae with some kind of data (karyotype formulae, idiograms, asymmetry indices), which makes up a percentage of ca. 15% of the family. The most frequently informed characters are the following.

Chromosome Size

Chromosome size (usually expressed as length in μm) is useful to single out individuals, samples, populations or species. DNA content is directly related to it, so the alleged factors affecting the former may indirectly affect the latter as the speed of DNA replication and the duration of the life cycle. Also, long chromosomes would undergo a higher number of chiasmata than small ones (Turney et al., 2004) and low recombination due to small chromosomes would be compensated by high ploidy levels (Nakazato et al., 2006). In Solanaceae, average chromosome size varies from ca. 1 μm (in *Solanum* and *Atropa*) to 6.5–11.5 μm (in *Cestrum* and the Cyphomandra clade of *Solanum,* **Figure 1**, **Supplementary Figure 1**, and **Supplementary Table 3**) although most species have small to medium-sized chromosomes, with an overall mean chromosome length of 2.95 ± 1.78 (**Supplementary Table 3**). In order to better visualize the changes in chromosome size and evaluate them in a phylogenetic context, we mapped C and TL onto the Solanaceae family tree (**Supplementary Tables 3, 4**). Although these two variables are homoplastic (**Supplementary Figures 1**, **2**) and have no significant phylogenetic signal, it is still useful to characterize some clades which are equivalent to medium-hierarchy taxonomic levels. Cestreae, Cyphomandra and *Capsicum* clades are outstanding because they have high C, conspicuously different from the remaining

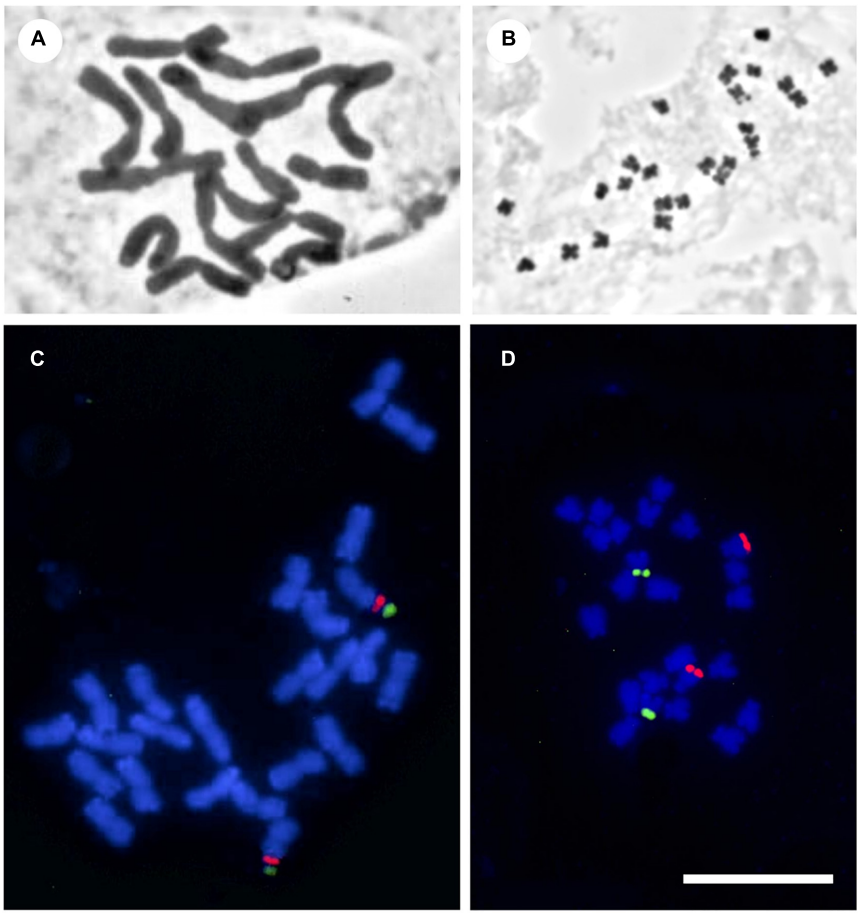

FIGURE 1 | Photomicrographs of mitotic metaphases of Solanaceae, illustrating the chromosome variety found in the family. **(A)** *Cestrum buxifolium* Kunth (2n = 16, *c* = 11.5 μm, *r* = 1.4). **(B)** *Solanum lidii* Sunding (2n = 24, *c* = 1.3 μm, *r* = 1.6). **(C)** *Iochroma edule* S. Leiva (2n = 24, *c* = 4.0, *r* = 1.2). **(D)** *Physalis lagascae* Roem. & Schult. (2n = 24, *c* = 2.5 μm, *r* = 2.8). Panels **(A,B)** with classical staining, panels **(C,D)** fluorescence *in situ* hybridization with 5S (red) and 18-5.8-26S (green) rDNA probes. All at the same scale. Bar = 10 μm.

clades (**Supplementary Figure 1**). These synapomorphic large chromosomes arose three times separately and are an interesting case to study phenomena such as "genome obesity" (Bennetzen and Kellogg, 1997), increases in DNA content through the incorporation of heterochromatin, repetitive sequences, and retrotransposons (Fregonezi et al., 2007; Miguel et al., 2012; Scaldaferro et al., 2013).

Chromosome Morphology

Asymmetry

There are a variety of indices to estimate chromosomal asymmetry and their accuracy is arguable (Peruzzi and Eroğlu, 2013). Here we have considered the r index because it is the most widespread, intuitive and easy to calculate, and also the most frequently informed (**Supplementary Table 5**). Asymmetry indices try to provide an idea of the general morphology of the karyotype through a single number, but logically not regarding the morphology of each chromosome. When all pairs of chromosomes undergo similar changes at the same time (like in *Nicotiana*, Kovarik et al., 2004), other measurements are needed to distinguish individual chromosome rearrangements. In a general survey of Solanaceae, Badr et al. (1997) reported values of r from 1.17 to 2.78, whereas in *Solanum* it ranges from 1.19 to 3.71 with an overall mean of 1.64 (Chiarini et al., 2018). Thus, karyotype asymmetry is not uniform in the family. According to ASR, symmetrical karyotypes are plesiomorphic (**Supplementary Figure 3**), being asymmetry a synapomorphy of some clades (*Physalis*, *Nicotiana*), while others would have conserved symmetry (*Lycium*). This supports the idea that in Solanaceae, shifts in chromosome morphology are frequent when considering broad frames of time (Wu and Tanksley, 2010; Chiarini et al., 2018).

Karyotype Formulae

The karyotype formula is a mean of expressing the result of a measurement process of the chromosome complement, i.e., the whole set of chromosomes in a nucleus, and it gives an overall idea of its morphology. Morphology provides the first clue to single out chromosome pairs (e.g., heteromorphic sex chromosomes). Chromosome morphology can be related to the

plant mating system: in *Rumex*, a species with sex chromosomes, the suppression of recombination varies between metacentric and acrocentric chromosomes (e.g., Rifkin et al., 2021). Also, bimodal karyotypes visible with classical technique evidence the presence of two genomes (Brandham and Doherty, 1998). According to the widespread Levan's classification (Levan et al., 1964), chromosomes can be, in increasing order of asymmetry, ***m***, ***sm***, ***st***, and ***t***. On this basis, karyotypes of Solanaceae are symmetrical, since only 17 out of the total species (4%) with known karyotypes contain 1–3 ***t*** chromosomes, while 100 (24%) present 1–7 ***sm*** chromosomes. Most Solanaceae have more than half of chromosomes per complement in the ***m*** and/or ***sm*** categories. These features are useful to characterize entire clades within the family: there are groups with a majority of ***m*** chromosomes, e.g., *Lycium* (Stiefkens et al., 2020), but there are also groups in which complements include ***st*** and ***t*** chromosomes, e.g., *Nicotiana*, *Capsicum*, *Jaborosa*, *Physalis* (**Figure 1D**), *Hyoscyamus*, *Nierembergia* (Sheidai et al., 1999; Moscone et al., 2006; Scaldaferro et al., 2013; Acosta et al., 2016; Chiarini et al., 2017; Rodríguez et al., 2020) and in some clades of *Solanum* (Bernardello et al., 1994; Acosta et al., 2005). Some *Nicotiana* species have karyotypes mostly with ***st*** chromosomes (Villa, 1984) and *Leptoglossis linifolia* has mostly ***sm*** chromosomes (Acosta et al., 2016).

In Solanaceae, establishing the direction of karyotype evolution is unlikely, as many reversals might have occurred (Stace, 2000), and karyotype asymmetry might be a temporary state rather than an evolutionary end (Mandáková and Lysak, 2008). Comparing karyotype formulae is arguable because it is difficult to establish homologies of chromosome pairs among different species. Here, we calculated the proportion of ***m*** chromosomes for each nightshade with the available karyotype formula and mapped this ratio onto the phylogeny (**Figure 2**, **Supplementary Figure 4**, and **Supplementary Table 6**). This reconstruction suggests the most probable ancestral karyotypes had ± 80% of ***m*** chromosomes (i.e., symmetric karyotype). Formulae with few or no ***m*** chromosomes are autapomorphic (e.g., *Nicotiana plumbaginifolia*) while complete formulae with ***m*** chromosomes is a synapomorphy for certain clades (*Lycium*, Iochrominae, **Figure 1C**). Transitions between karyotype formulae can be interpreted as evidence of chromosome rearrangements. Particularly, formulae with ***st*** or ***t*** chromosomes are seemingly the result of a deletion or translocation of the entire or part of one arm (Weiss-Schneeweiss and Schneeweiss, 2013) or centric fissions (Moscone et al., 2006; Jarret et al., 2019).

Chromosome Banding

Banding techniques are the next step in identifying chromosome pairs in a complement. They leverage on the physical properties of chromatin, using dyes that differentially stain euchromatin and fractions within heterochromatin (e.g., AT- or GC-rich), which are then visualized as bands across the chromatids. In Solanaceae, 218 spp. out of approx. 2,800 (ca. 7.7%) have been studied with any banding technique. The largest number of species studied belong to *Solanum* (69 out of 1,238 species, 5.6%), followed by *Capsicum* (25 out of 41, 60.9%) and *Lycium* (24 out of 105, 22.9%) (**Supplementary Table 2**). These last two genera are also the best studied in relative terms, together with *Nierembergia* (71.4% of 21 spp.), *Jaborosa* (65.2% of 23 spp.) and *Iochroma* (22.6% of 31 spp.) (**Supplementary Table 2**). A variety of banding techniques have been essayed in Solanaceae to answer a range of questions: silver impregnation detected active nucleolar organizing regions (NORs) (e.g., in *Cestrum*, Berg and Greilhuber, 1993; *Capsicum*, Scaldaferro et al., 2016; *Solanum*, Miguel et al., 2012; *Deprea*, Deanna et al., 2014); Giemsa C-banding revealed differences between taxa and contributed to taxonomic grouping of *Capsicum* (Moscone et al., 1993); Fluorescent banding, mostly with the double staining CMA/DAPI technique (Schweizer, 1976) demonstrated that chromosomes are mostly composed of non-coding chromatin, and also revealed strong phylogenetic signal, defining specific patterns in different clades (Acosta et al., 2016). CMA/DAPI technique helped to detect variability despite the morphological uniformity of the chromosomes, provided information on genetic variation at a population level regarding speciation, and revealed AT- and GC-rich regions of B chromosomes. Together with FISH (see below), this technique showed that B chromosomes possess nucleolar activity and nucleolar competition with the A chromosomes (Acosta and Moscone, 2011; Acosta et al., 2016; Montechiari et al., 2020).

Heterochromatin amount (H.a.) is a value often reported (usually as a percentage). It exhibits a remarkable variation among species of the same genus, for example in *Solanum* it varies from 1.86 to 35.43% (Brasileiro-Vidal et al., 2009; Chiarini et al., 2018). The data on H.a. have served to discuss whether the increases in genome size (either measured visually as TL, or as DNA content through flow cytometry) are due to increases in the amount of one or another fraction of chromatin (e.g., Karsburg et al., 2009). In some clades, H.a. is positively correlated with karyotype length (Moscone et al., 2006; Scaldaferro et al., 2013), while in others there is negative or no correlation, which is unexpected because heterochromatin is considered as an additional component of the genome (Pringle and Murray, 1993; Acosta et al., 2012, 2016). Heterochromatin patterns are more variable in some genera (*Jaborosa, Solanum*) than in other members of the $x = 12$ clade (e.g., *Lycium* and *Sclerophylax*), where heterochromatin is scarce and restricted to the NORs (Lujea and Chiarini, 2017; Stiefkens et al., 2020). Variable patterns might be evidence of intense chromosomal rearrangements (Evtushenko et al., 2016) associated with diversification and colonization of new habitats, since they function as species barriers (Hughes and Hawley, 2009).

rDNA and Repetitive DNA Patterns

Ribosomal DNAs are fundamental components of all cell types. In most plants, 5S and 18S-5.8S-26S rDNAs are present in high copy numbers to satisfy the cellular requirement for ribosomes. These rDNAs, highly conserved in plants, are the commonest markers for FISH (Heslop-Harrison and Schwarzacher, 2011). They are repetitive, tandemly arranged and generally clustered at different loci, either with a spatial separation or also linked in a single unit. FISH enables

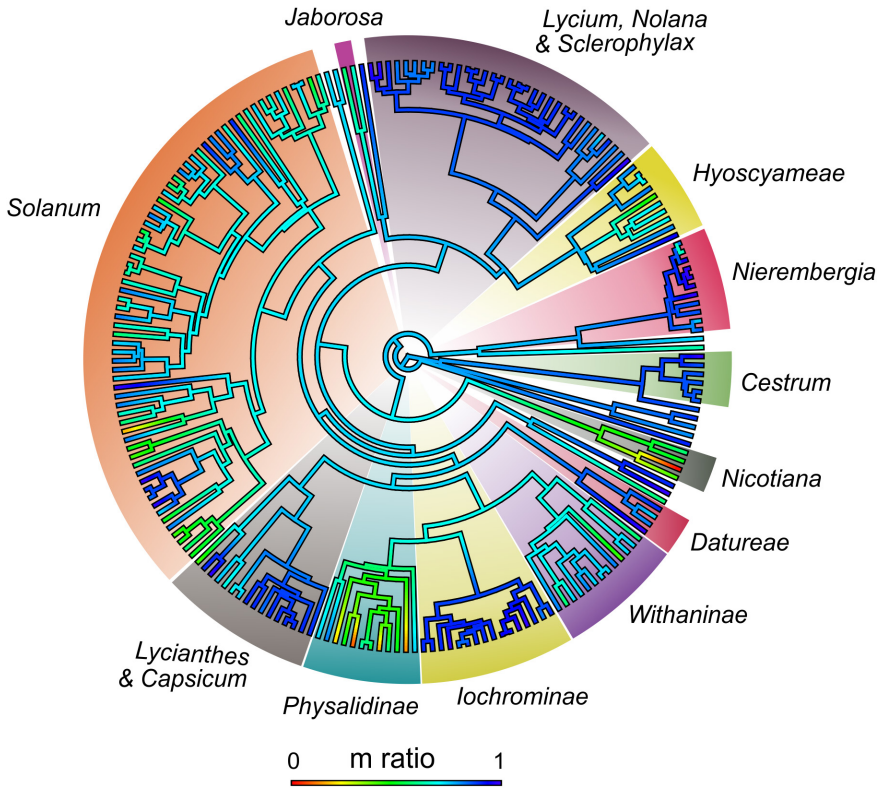

FIGURE 2 | Heatmap of m ratio (proportion of metacentric chromosomes over the total chromosomes of each karyotype formula) reconstructed on Solanaceae. Scale below indicates values of the ratio and its color guide. Values vary from 0 (no metacentric chromosomes) to 1 (all chromosomes metacentric).

direct visualization of rDNA loci on the chromosomes and determination of their rearrangements and organization (e.g., Srebniak et al., 2002; Melo and Guerra, 2003; Lou et al., 2010). Thus, homologous chromosomes in a complement can be identified and related species can be compared, answering evolutionary questions. Number, position and organization of rDNAs have been suitably employed in Solanaceae as systematic and evolutionary approaches (**Figures 1C,D** and **Supplementary Table 2**), for example in *Nierembergia* (Acosta et al., 2016), *Jaborosa* (Chiarini et al., 2017) and *Iochroma* (Deanna et al., 2018). Genera with economically important species (*Capsicum*, *Nicotiana*) have received special attention (Kitamura et al., 2005; Scaldaferro et al., 2016). Within the largest genera, *Lycium* is the most studied (Stiefkens et al., 2020), revealing uniformity of patterns, but little is known about the rest of the family. Chromosomal rearrangements between *Solanum* crops and several related species have been studied using tomato and potato bacterial artificial chromosomes (BACs) in multiple FISH essays, providing support for grouping of species into sections and suggesting interspecific hybridization events (Doganlar et al., 2002; Frary et al., 2016). Another type of repetitive DNA revealed by FISH are the telomeric sequences. Most plants present the Arabidopsis-type telomeres, but some Solanaceae are exceptional in lacking such sequences: the characterization of unusual telomeres in the three genera of Cestreae has shed light on patterns of telomere evolution, maintenance and function (Sýkorová et al., 2003; Peška et al., 2015).

CYTOGENETIC AND PHYLOGENETIC GAPS

There is a growing interdependence between chromosomal data and molecular phylogeny. Cytogenetics and molecular phylogeny feed into each other, providing evidence in both ways. To some extent, Solanaceae ASRs allow reconstructing the cytogenetic data when there is molecular data (e.g., Rodríguez et al., 2020), but the reverse is obviously not possible. Thus, the largest information gaps are found in *Cestrum*, (165 species without any data of a total of 233, neither counts nor sequences), *Lycianthes* (116 species without data of about 161), *Jaltomata* (32 species of 73), *Physalis* (47 of 109) and *Browallia* (18 species without data of about 22). Of the ca. 2,900 species in the family, 494 (17%) possess chromosome counts and are included in the phylogeny of Särkinen et al. (2013), while of these, only 207 (7% of the family) also possess karyotypes. At the same time, there are 137 species for which there are counts but no sequences, and of these, 40 also have karyotypes, but these data cannot be analyzed in a molecular phylogenetic context (**Supplementary Tables 7**, **8**). Important gaps in both cytogenetics and phylogeny still wait to be filled: in Goetzeoideae, a clade of which nothing is known;

in Solandreae, only chromosome studies in one *Trianaea* species and two *Dyssochroma*; *Cuatresia*, genus of Physalideae with an important number of species about which nothing is known (**Supplementary Table 8**).

CYTOGENETICS AND GENOMICS IN SOLANACEAE

Extensive pairwise comparative mapping studies have been performed for several major solanaceous crops relative to tomato (Tanksley et al., 1988, 1992; Livingstone et al., 1999; Doganlar et al., 2002; Wu et al., 2009; Wu and Tanksley, 2010). By using single-copy conserved ortholog sets (COSII), the outcomes of chromosomal evolution in the Solanaceae over 30 million years were deduced, estimating the rates and timing of chromosomal rearrangements as well as calculating the age of ancestral species and predicting their genome features (Wu and Tanksley, 2010). Thus, studies integrating cytogenetics and genomics, though focused on main crops and close relatives, generate knowledge that can be extrapolated to the rest of the species of the family.

Recently, Cytogenetics developed feedback with Genomics: the availability of species with complete chloroplast genome sequenced and some with complete nuclear genome, allows the finding of new markers for hybridization techniques. In return, these cytogenetic data become valuable information on chromosome identification and genome organization, enabling the spatial location of specific, single, or low-copy sequences. Next-generation sequencing speeds up cytogenetic research on Solanaceae. In species with available transcriptomes, like *S. dulcamara*, new markers like SNP SSR, AFLP and CAPS were developed, allowing the construction of a genome-wide genetic linkage map (D'Agostino et al., 2013). Based on gene orthologs, the markers were anchored to the genome of related *Solanum* species (tomato, potato and eggplant), revealing both conserved and novel chromosomal rearrangements. Estimates of the evolutionary moment of rearrangements were possible and showed that chromosomal breakpoints are regularly re-used.

Also, advances in NGS, such as the study of repetitive DNA, contributed to the design of more specific oligo-probes for FISH (Buggs et al., 2012), and hence bringing new chances to improve our understanding of systematics and genome organization at different taxonomic hierarchies. More recently, there has been progress in the phylogenomics of Solanaceae, particularly in some clades within the family, through the use of nuclear target capture and high-throughput sequencing of transcriptomic data (Gates et al., 2018; Poczai et al., 2018; Gagnon et al., 2021). These approaches provide more comprehensive and better resolved phylogenies which can be used in the study of chromosome evolution.

CONCLUDING REMARKS

- Evolutionary trends: The ancestor of the family had a karyotype formula with 80% of m chromosomes, each about 2 μm in size. Increased asymmetry and size are synapomorphies of some clades that arose in independent events.
- Although homoplastic at a higher hierarchical level, karyotype traits are useful to delimit taxa at lower levels: tribes, subtribes, genera.
- Important gaps in both cytogenetics and phylogeny still wait to be filled.
- Potential for future applications: once there is a genomic phylogenetic backbone of Solanaceae, including divergence time estimates, adding chromosomes to it would be illuminating, because relationships between major clades of the family and Solanoideae are still poorly resolved using Sanger sequence data alone.

AUTHOR CONTRIBUTIONS

FC and RD conceived the ideas, designed the methodology, analyzed the data, and wrote the manuscript. FC, RD, MS, and MA collected the data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Agency for Scientific Promotion and Technological grant PICT 2017-2370 and PICT 2016-1525 (FONCyT), the National Science Foundation (NSF-DEB 1902797), and National Council for Scientific and Technological Research (CONICET). Publication of this article was funded by the University of Colorado Boulder Libraries Open Access Fund.

ACKNOWLEDGMENTS

We would like to thank Stacey Smith for advice during the comparative analyses performed.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.787590/full#supplementary-material

REFERENCES

Acosta, M. C., Bernardello, G., Guerra, M., and Moscone, E. A. (2005). Karyotype analysis in several South American species of *Solanum* and *Lycianthes rantonnei* (Solanaceae). *Taxon* 54, 713–723. doi: 10.2307/25065428

Acosta, M. C., Guerra, M., and Moscone, E. A. (2012). Karyological relationships among some South American species of *Solanum* (Solanaceae) based on fluorochrome banding and nuclear DNA amount. *Plant Syst. Evol.* 298, 1547–1556. doi: 10.1007/s00606-012-0657-x

Acosta, M. C., and Moscone, E. A. (2011). B Chromosomes in *Nierembergia aristata* (Solanaceae): Nucleolar Activity and Competition with the A Chromosomes. *Cytogenet. Gen. Res.* 132, 105–112. doi: 10.1159/000320705

Acosta, M. C., Moscone, E. A., and Cocucci, A. A. (2016). Using chromosomal data in the phylogenetic and molecular dating framework: karyotype evolution and diversification in *Nierembergia* (Solanaceae) influenced by historical changes in sea level. *Plant Biol.* 18, 514–526. doi: 10.1111/plb.12430

Acosta, M. C., Ordoñez, A. D. V., Cocucci, A. A., and Moscone, E. A. (2006). Chromosome reports in South American Nicotianeae (Solanaceae) with particular reference to *Nierembergia*. *Ann. Mo. Bot. Gard.* 93, 634–646. doi: 10.3417/0026-6493(2006)93[634:crisan]2.0.co;2

Avery, A. G., Satina, S., and Rietsema, J. (1959). *Blakeslee: The Genus Datura*. New York: Ronald Press Co.

Badr, A., Khalifa, S. F., Aboel-Atta, A. I., and Abou-El-Enain, M. M. (1997). Chromosomal criteria and taxonomic relationships in the Solanaceae. *Cytologia* 62, 103–113. doi: 10.1508/cytologia.62.103

Baltisberger, M., and Hörandl, E. (2016). Karyotype evolution supports the molecular phylogeny in the genus *Ranunculus* (Ranunculaceae). *Perspect. Plant Ecol. Evol. Syst.* 18, 1–14. doi: 10.1016/j.ppees.2015.11.001

Bennetzen, J. L., and Kellogg, E. A. (1997). Do plants have a one-way ticket to genomic obesity? *Plant Cell* 9, 1509–1514. doi: 10.2307/3870439

Berg, C., and Greilhuber, J. (1993). Cold-sensitive chromosome regions and heterochromatin in *Cestrum* (Solanaceae): *C. strigillatum*, *C. fasciculatum*, and *C. elegans*. *Plant Syst. Evol.* 185, 133–151. doi: 10.1007/BF00937725

Bernardello, G., Stiefkens, L., and Las Peñas, M. L. (2008). Karyotype studies in *Grabowskia* and *Phrodus* (Solanaceae). *Plant Syst. Evol.* 275, 265–269. doi: 10.1007/s00606-008-0060-9

Bernardello, L. M., Heiser, C. B., and Piazzano, M. (1994). Karyotypic studies in *Solanum* section *Lasiocarpa* (Solanaceae). *Am. J. Bot.* 81, 95–103. doi: 10.1002/j.1537-2197.1994.tb15414.x

Blakeslee, A. F., Belling, J., and Farnham, M. E. (1920). Chromosomal duplication and Mendelian phenomena in *Datura* mutants. *Science* 52, 388–390. doi: 10.1126/science.52.1347.388

Brandham, P. E., and Doherty, M. J. (1998). Genome size variation in the Aloaceae, an angiosperm family displaying karyotypic orthoselection. *Ann. Bot.* 82, 67–73. doi: 10.1006/anbo.1998.0742

Brasileiro-Vidal, A. C., Melo-Oliveira, M. B., Carvalheira, G. M. G., and Guerra, M. (2009). Different chromatin fractions of tomato (*Solanum lycopersicum* L.) and related species. *Micron* 40, 851–859. doi: 10.1016/j.micron.2009.06.004

Buggs, R. J., Renny-Byfield, S., Chester, M., Jordon-Thaden, I. E., Viccini, L. F., Chamala, S., et al. (2012). Next-generation sequencing and genome evolution in allopolyploids. *Amer. J. Bot.* 99, 372–382. doi: 10.3732/ajb.1100395

Camadro, E. L., Erazzu, L. E., Maune, J. F., and Bedogni, M. C. (2012). A genetic approach to the species problem in wild potato. *Plant Biol.* 14, 543–554. doi: 10.1111/j.1438-8677.2012.00563.x

Chiarini, F., and Bernardello, G. (2006). Karyotype studies in South American species of *Solanum* subgen. *Leptostemonum* (Solanaceae). *Plant Biol.* 8, 486–493. doi: 10.1055/s-2006-923859

Chiarini, F., Moreno, N., Moré, M., and Barboza, G. (2017). Chromosomal changes and recent diversification in the Andean genus *Jaborosa* (Solanaceae). *Bot. J. Linn. Soc.* 183, 57–74. doi: 10.1111/boj.12493

Chiarini, F., Sazatornil, F., and Bernardello, G. (2018). Data reassessment in a phylogenetic context gives insight into chromosome evolution in the giant genus *Solanum* (Solanaceae). *Syst. Biodivers.* 16, 397–416. doi: 10.1080/14772000.2018.1431320

D'Agostino, N., Golas, T., Van de Geest, H., Bombarely, A., Dawood, T., Zethof, J., et al. (2013). Genomic analysis of the native European *Solanum species*, *S. dulcamara*. *BMC Gen*. 14:356. doi: 10.1186/1471-2164-14-356

Darlington, C. D. (1956). *Chromosome Botany*. London, United Kingdom: George Allen & Unwin

Deanna, R., Barboza, G. E., and Scaldaferro, M. A. (2014). First karyological report in *Larnax* and *Deprea* (Solanaceae). *Austral. J. Bot.* 62, 251–261. doi: 10.1071/BT14041

Deanna, R., Smith, S. D., Särkinen, T., and Chiarini, F. (2018). Patterns of chromosomal evolution in the florally diverse Andean clade Iochrominae (Solanaceae). *Perspect. Plant Ecol. Evol. Syst.* 35, 31–43. doi: 10.1016/j.ppees.2018.09.004

Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N., and Tanksley, S. D. (2002). A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the *Solanaceae*. *Genetics* 161, 1697–1711. doi: 10.1093/genetics/161.4.1697

Dupin, J., Matzke, N. J., Särkinen, T., Knapp, S., Olmstead, R. G., Bohs, L., et al. (2017). Bayesian estimation of the global biogeographical history of the *Solanaceae*. *J. Biogeogr.* 44, 887–899. doi: 10.1111/jbi.12898

Evtushenko, E. V., Levitsky, V. G., Elisafenko, E. A., Gunbin, K. V., Belousov, A. I., Šafáø, J., et al. (2016). The expansion of heterochromatin blocks in rye reflects the co-amplification of tandem repeats and adjacent transposable elements. *BMC Gen.* 17:337. doi: 10.1186/s12864-016-2667-5

FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* 3, 1084–1092. doi: 10.1111/j.2041-210X.2012.00234.x

Frary, A., Doganlar, S., and Frary, A. (2016). "Synteny among *solanaceae* genomes," in *The Tomato Genome*, eds M. Causse, J. Giovannoni, M. Bouzayen and M. Zouine (Berlin, Heidelberg: Springer), 217–243. doi: 10.1007/978-3-662-53389-5_12

Fregonezi, J. N., Vilas-Boas, L. A., Fungaro, M. H. P., Gaeta, M. L., and Vanzela, A. L. L. (2007). Distribution of a Ty3/gypsy-like retroelement on the A and B-chromosomes of *Cestrum strigilatum* Ruiz & Pav. and *Cestrum intermedium* Sendtn. (Solanaceae). *Genet. Mol. Biol.* 30, 599–604. doi: 10.1590/S1415-47572007000400017

Gagnon, E., Hilgenhof, R., Orejuela, A., McDonnell, A., Sablok, G., Aubriot, X., et al. (2021). Phylogenomic data reveal hard polytomies across the backbone of the large genus *Solanum* (Solanaceae). *bioRxiv* [Preprint]. doi: 10.1101/2021.03.25.436973

Gates, D. J., Pilson, D., and Smith, S. D. (2018). Filtering of target sequence capture individuals facilitates species tree construction in the plant subtribe Iochrominae (Solanaceae). *Mol. Phylogenet. Evol.* 123, 26–34. doi: 10.1016/j.ympev.2018.02.002

Glick, L., and Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* 31, 1914–1922. doi: 10.1093/molbev/msu122

Goodspeed, T. H. (1954). *The Genus Nicotiana; Origins, Relationships And Evolution Of Its Species In The Light Of Their Distribution, Morphology And Cytogenetics*. Waltham, Mass: Chronica Botanica Co.

Heslop-Harrison, J. S., and Schwarzacher, T. (2011). Organisation of the plant genome in chromosomes. *Plant J.* 66, 18–33. doi: 10.1111/j.1365-313X.2011.04544.x

Hijmans, R. J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A., and Spooner, D. M. (2007). Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section *Petota*). *Glob. Ecol. Biogeogr.* 16, 485–495. doi: 10.1111/j.1466-8238.2007.00308.x

Hughes, S. E., and Hawley, R. S. (2009). Heterochromatin: a rapidly evolving species barrier. *PLoS Biol.* 7:e1000233. doi: 10.1371/journal.pbio.1000233

Jarret, R. L., Barboza, G. E., da Costa Batista, F. R., Berke, T., Chou, Y. Y., Hulse-Kemp, A., et al. (2019). *Capsicum*—An abbreviated compendium. *J. Am. Soc. Hortic. Sci.* 144, 3–22. doi: 10.21273/JASHS04446-18

Karsburg, I. V., Carvalho, C. R., and Clarindo, W. R. (2009). Identification of chromosomal deficiency by flow cytometry and cytogenetics in mutant tomato (*Solanum lycopersicum*, Solanaceae) plants. *Aust. J. Bot.* 57, 444–449. doi: 10.1071/BT08223

Kitamura, S., Tanaka, A., and Inoue, M. (2005). Genomic relationships among *Nicotiana* species with different ploidy levels revealed by 5S rDNA spacer sequences and FISH/GISH. *Genes Genet. Syst.* 80, 251–260. doi: 10.1266/ggs.80.251

Kovarik, A., Matyasek, R., Lim, K. Y., Skalická, K., Koukalova, B., Knapp, S., et al. (2004). Concerted evolution of 18–5.8–26S rDNA repeats in *Nicotiana* allotetraploids. *Biol. J. Linn. Soc.* 8, 615–625. doi: 10.1111/j.1095-8312.2004.00345.x

Levan, A., Fredga, L., and Sandberg, A. (1964). Nomenclature for centromeric position on chromosomes. *Hereditas* 52, 201–220. doi: 10.1111/j.1601-5223.1964.tb01953.x

Livingstone, K. D., Lackney, V. K., Blauth, J. R., Van Wijk, R. I. K., and Jahn, M. K. (1999). Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics* 152, 1183–1202. doi: 10.1093/genetics/152.3.1183

Lou, Q., Iovene, M., Spooner, D. M., Buell, C. R., and Jiang, J. (2010). Evolution of chromosome 6 of *Solanum species* revealed by comparative fluorescence in situ hybridization mapping. *Chromosoma* 119, 435–442. doi: 10.1007/s00412-010-0269-6

Lujea, N. C., and Chiarini, F. E. (2017). Differentiation of *Nolana* and *Sclerophylax* (Solanaceae) by means of heterochromatin and rDNA patterns. *N. Z. J. Bot.* 55, 163–177. doi: 10.1080/0028825X.2016.1269812

Mandáková, T., and Lysak, M. A. (2008). Chromosomal phylogeny and karyotype evolution in x = 7 crucifer species (Brassicaceae). *Plant Cell* 20, 2559–2570. doi: 10.1105/tpc.108.062166

McCarthy, E. W., Landis, J. B., Kurti, A., Lawhorn, A. J., Chase, M. W., Knapp, S., et al. (2019). Early consequences of allopolyploidy alter floral evolution in *Nicotiana* (Solanaceae). *BMC Plant Biol.* 19:162. doi: 10.1186/s12870-019-1771-5

Melo, N. F., and Guerra, M. (2003). Variability of the 5S and 45S rDNA sites in *Passiflora* L. species with distinct base chromosome numbers. *Ann. Bot.* 92, 309–316. doi: 10.1093/aob/mcg138

Miguel, V., Acosta, M. C., and Moscone, E. A. (2012). Karyotype analysis in two species of *Solanum* (Solanaceae) Sect. Cyphomandropsis based on chromosome banding. *N. Z. J. Bot.* 50, 217–225. doi: 10.1080/0028825X.2012.662904

Montechiari, K. A., González, M. L., Yañez Santos, A. M., Hajduczyk Rutz, J. L., and Urdampilleta, J. D. (2020). Structure, behaviour and repetitive DNA of B-chromosomes in *Cestrum nocturnum* (Solanaceae). *Plant Biosyst.*, 29–37. doi: 10.1080/11263504.2018.1559249

Moscone, E. A., Lambrou, M., Hunziker, A. T., and Ehrendorfer, F. (1993). Giemsa C-banded karyotypes in *Capsicum* (Solanaceae). *Plant Syst. Evol.* 186, 213–229. doi: 10.1007/BF00940799

Moscone, E. A., Scaldaferro, M. A., Grabiele, M., Cecchini, N. M., Sánchez García, Y., Jarret, R., et al. (2006). "The evolution of chili peppers (*Capsicum*-Solanaceae): a cytogenetic perspective," in *VI International Solanaceae Conference: Genomics Meets Biodiversity*, (Leuven: International Society for Horticultural Science) 745, 137–170. doi: 10.17660/ActaHortic.2007.745.5

Nakazato, T., Jung, M. K., Housworth, E. A., Rieseberg, L. H., and Gastony, G. J. (2006). Genetic map-based analysis of genome structure in the homosporous fern *Ceratopteris richardii*. *Genetics* 173, 1585–1597. doi: 10.1534/genetics.106.055624

Olmstead, R. G., Bohs, L., Migid, H. A., Santiago-Valentin, E., Garcia, V. F., and Collier, S. M. (2008). A molecular phylogeny of the Solanaceae. *Taxon* 57, 1159–1181. doi: 10.1002/tax.574010

Olmstead, R. G., Sweere, J. A., Spangler, R. E., Bohs, L., and Palmer, J. D. (1999). "Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA," in Solanaceae *IV: advances in biology and utilization*, eds M. Nee, D. E. Symon, R. N. Lester, and J. P. Jessop (Kew: Royal Botanic Gardens), 257–274.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Peruzzi, L., and Eroğlu, H. E. (2013). Karyotype asymmetry: again, how to measure and what to measure? *Comp. Cytogenet.* 7, 1–9. doi: 10.3897/compcytogen.v7i1.4431

Peška, V., Fajkus, P., Fojtová, M., Dvořáčková, M., Hapala, J., Dvořáček, V., et al. (2015). Characterisation of an unusual telomere motif (TTTTTTAGGG)n in the plant *Cestrum elegans* (Solanaceae), a species with a large genome. *Plant J.* 82, 644–654. doi: 10.1111/tpj.12839

Poczai, P., Amiryousefi, A., Hyvönen, J. T., and Sablok, G. (2018). "Chloroplast and nuclear phylogenomics reveals concordant phylogenetic structure in wild tomatoes (*Solanum* sect. Lycopersicon, Solanaceae)," in *The 15th Solanaceae Conference: Applied Genomics, Accelerated Breeding, Gene Targeting*, (Chiang Mai: University of Helsinki)

Pringle, G. J., and Murray, B. G. (1992). Polyploidy and aneuploidy in the tamarillo, *Cyphomandra betacea* (Cav.) Sendt. (Solanaceae). I. Spontaneous polyploidy and features of the euploids. *Plant Breed.* 108, 132–138. doi: 10.1111/j.1439-0523.1992.tb00112.x

Pringle, G. J., and Murray, B. G. (1993). Karyotypes and C banding patterns in species of *Cyphomandra* Mart. ex Sendtn. (Solanaceae). *Bot. J. Linn. Soc.* 111, 331–342. doi: 10.1006/bojl.1993.1023

Raven, P. H. (1975). The bases of angiosperm phylogeny: cytology. *Ann. Mo. Bot. Gard.* 62, 724–764. doi: 10.2307/2395272

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Rifkin, J. L., Beaudry, F. E., Humphries, Z., Choudhury, B. I., Barrett, S. C., and Wright, S. I. (2021). Widespread recombination suppression facilitates plant sex chromosome evolution. *Mol. Biol. Evol.* 38, 1018–1030. doi: 10.1093/molbev/msaa271

Rodríguez, J., Deanna, R., and Chiarini, F. (2020). Karyotype asymmetry shapes diversity within the physaloids (Physalidinae, Physalideae, Solanaceae). *Syst. Biodivers.* 19, 168–185. doi: 10.1080/14772000.2020.1832156

Särkinen, T., Bohs, L., Olmstead, R. G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol.* 13, 1–15. doi: 10.1186/1471-2148-13-214

Scaldaferro, M. A., Grabiele, M., and Moscone, E. A. (2013). Heterochromatin type, amount and distribution in wild species of chili peppers (*Capsicum*-Solanaceae). *Genet. Res. Crop Evol.* 60, 693–709. doi: 10.1007/s10722-012-9867-x

Scaldaferro, M. A., Romero da Cruz, M. V., Cecchini, N. M., and Moscone, E. A. (2016). FISH and AgNor-mapping of the 45S and 5S rRNA genes in wild and cultivated *Capsicum* species (Solanaceae). *Genome* 59, 95–113. doi: 10.1139/gen-2015-0099

Scaldaferro, M., Chiarini, F., Santiñaque, F. F., Bernardello, G., and Moscone, E. A. (2012). Geographical pattern and ploidy levels of the weed *Solanum elaeagnifolium* (Solanaceae) from Argentina. *Genet. Res. Crop Evol.* 59, 1833–1847. doi: 10.1007/s10722-012-9807-9

Schranz, M. E., Mohammadin, S., and Edger, P. P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr. Opin. Plant Biol.* 15, 147–153. doi: 10.1016/j.pbi.2012.03.011

Schweizer, D. (1976). Reverse fluorescent chromosome banding with chromomycin and DAPI. *Chromosoma* 58, 307–324. doi: 10.1007/BF00292840

Sheidai, M., Mosallanejad, M., and Khatamsaz, M. (1999). Karyological studies in *Hyoscyamus* species of Iran. *Nord. J. Bot.* 19, 369–374. doi: 10.1111/j.1756-1051.1999.tb01130.x

Srebniak, M., Rasmussen, O., and Maluszynska, J. (2002). Cytogenetic analysis of an asymmetric potato hybrid. *J. Appl. Genet.* 43, 19–32.

Stace, C. A. (2000). Cytology and cytogenetics as a fundamental taxonomic resource for the 20th and 21st centuries. *Taxon* 49, 451–477. doi: 10.2307/1224344

Stiefkens, L., Las Peñas, M. L., Levin, R. A., Miller, J. S., and Bernardello, G. (2020). Chromosome evolution in the cosmopolitan genus *Lycium* (Solanaceae). *Taxon* 69, 124–141. doi: 10.1002/tax.12205

Sýkorová, E., Lim, K. Y., Chase, M. W., Knapp, S., Leitch, I. J., Leitch, A. R., et al. (2003). The absence of Arabidopsis-type telomeres in *Cestrum* and closely related genera *Vestia* and *Sessea* (Solanaceae): first evidence from eudicots. *Plant J.* 34, 283–291. doi: 10.1046/j.1365-313x.2003.01731.x

Tanksley, S. D., Bernatzky, R., Lapitan, N. L., and Prince, J. P. (1988). Conservation of gene repertoire but not gene ordeer in pepper and tomato. *Proc. Natl. Acad. Sci.* 85, 6419–6423. doi: 10.1073/pnas.85.17.6419

Tanksley, S. D., Ganal, M. W., Price, J. P., de Vicente, M. C., Bonierbale, M. W., Broun, P., et al. (1992). High density molecular linkage maps of the tomato and potato genomes; biological inferences and practical applications. *Genetics* 132, 1141–1160. doi: 10.1093/genetics/132.4.1141

Turney, D., de los Santos, T., and Hollingsworth, N. M. (2004). Does chromosome size affect map distance and genetic interference in budding yeast? *Genetics* 168, 2421–2424. doi: 10.1534/genetics.104.033555

Van de Peer, Y., Ashman, T. L., Soltis, P. S., and Soltis, D. E. (2021). Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* 33, 11–26. doi: 10.1093/plcell/koaa015

Villa, A. (1984). The chromosome idiogram of *Nicotiana plumbaginifolia*. *Genetica* 64, 145–148. doi: 10.1007/BF00120266

Weiss-Schneeweiss, H., and Schneeweiss, G. M. (2013). "Karyotype diversity and evolutionary trends in angiosperms," in *Plant Genome Diversity Volume 2*, ed. I. J. Leitch (Vienna: Springer), 209–230. doi: 10.1007/978-3-7091-1160-4_13

Wu, F., Eannetta, N. T., Xu, Y., and Tanksley, S. D. (2009). A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. *Theor. Appl. Genet.* 118, 927–935. doi: 10.1007/s00122-008-0950-9

Wu, F., and Tanksley, S. D. (2010). Chromosomal evolution in the plant family *Solanaceae*. *BMC Genomics* 11:182. doi: 10.1186/1471-2164-11-182

Zenil-Ferguson, R., Burleigh, J. G., Freyman, W. A., Igić, B., Mayrose, I., and Goldberg, E. E. (2019). Interaction among ploidy, breeding system and lineage diversification. *New Phytol.* 224, 1252–1265. doi: 10.1111/nph.16184

Zenil-Ferguson, R., Burleigh, J. G., and Ponciano, J. M. (2018). Chromploid: An R package for chromosome number evolution across the plant tree of life. *Appl. Plant Sci.* 6:e1037. doi: 10.1002/aps3.1037

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

REVIEW
published: 11 February 2022
doi: 10.3389/fpls.2022.802936

Germplasm Resources and Strategy for Genetic Breeding of *Lycium* Species: A Review

*Haiguang Gong[1,3], Fazal Rehman[1,3], Yun Ma[1,3], Biao A[1,3], Shaohua Zeng[1,3], Tianshun Yang[1], Jianguo Huang[1,3], Zhong Li[4], Dongpo Wu[5] and Ying Wang[1,2,3]**

[1] Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement, Provincial Key Laboratory of Digital Botanical Garden and Public Science, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, [2] School of Life Science, Gannan Normal University, Ganzhou, China, [3] School of Life Science, University of Chinese Academy of Sciences, Beijing, China, [4] Agricultural Comprehensive Development Center in Ningxia Hui Autonomous Region, Yinchuan, China, [5] Bairuiyuan Company, Yinchuan, China

Lycium species (goji), belonging to Solanaceae, are widely spread in the arid to semiarid environments of Eurasia, Africa, North and South America, among which most species have affinal drug and diet functions, resulting in their potential to be a superior healthy food. However, compared with other crop species, scientific research on breeding *Lycium* species lags behind. This review systematically introduces the present germplasm resources, cytological examination and molecular-assisted breeding progress in *Lycium* species. Introduction of the distribution of *Lycium* species around the world could facilitate germplasm collection for breeding. Karyotypes of different species could provide a feasibility analysis of fertility between species. The introduction of mapping technology has discussed strategies for quantitative trait locus (QTL) mapping in *Lycium* species according to different kinds of traits. Moreover, to extend the number of traits and standardize the protocols of trait detection, we also provide 1,145 potential traits (275 agronomic and 870 metabolic) in different organs based on different reference studies on *Lycium*, tomato and other Solanaceae species. Finally, perspectives on goji breeding research are discussed and concluded. This review will provide breeders with new insights into breeding *Lycium* species.

Keywords: *Lycium* species, molecular breeding, strategy of breeding, standardization, trait detection

OPEN ACCESS

Edited by:
Ezio Portis,
University of Turin, Italy

Reviewed by:
Jin Pei,
Chengdu University of Traditional Chinese Medicine, China
Lei Gao,
Wuhan Botanical Garden, Chinese Academy of Sciences (CAS), China

***Correspondence:**
Ying Wang
yingwang@scib.ac.cn

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *27 October 2021*
Accepted: *07 January 2022*
Published: *11 February 2022*

Citation:
Gong H, Rehman F, Ma Y, A B, Zeng S, Yang T, Huang J, Li Z, Wu D and Wang Y (2022) Germplasm Resources and Strategy for Genetic Breeding of Lycium Species: A Review. Front. Plant Sci. 13:802936. doi: 10.3389/fpls.2022.802936

INTRODUCTION

Goji, the general name of the plants in the genus *Lycium*, are members of the Solanaceae family found in arid to semiarid regions of Eurasia, Africa, and North and South America. Worldwide, 31 species of *Lycium* have been reported as medicines and/or foods, such as *L. ciliatum* (medicine for digestive inflammation), *L. cinereum* (fruit as food and medicine for rheumatism and headache; root as medicine for perfume, kidney disease, and anodyne), *L. pallidum* (fruit as food; root as medicine for chickenpox and toothache), *L. richii* (fruit as food), *L. intricatum* (seed and fruit as medicine for helminthiasis and eye diseases, respectively), among which *L. chinense* and *L. barbarum* have been widely regarded as superfoods with affinal drug and dietary

properties in recent years (Yao et al., 2018b). Goji (wolfberry), the fruit of *Lycium* plants, contains trace elements such as zinc, iron, calcium, and phosphorus as well as metabolic components such as flavonoids, carotenoids, and polysaccharides in red fleshy fruits with affinal drug and dietary functions (Yin and Dang, 2008; Wang C.C. et al., 2010; Wang J. et al., 2010; Qiu et al., 2014). Antioxidative compounds, polysaccharides (Dahech et al., 2013), and hydroxycinnamic acid amides (HCCAs) provide potential anti-inflammatory effects (Wang S. et al., 2017). *Lycium barbarum* polysaccharides (LBPs) have neuroprotective (Xing et al., 2016), liver-protective (Liu et al., 2015), anti-radiation (Duan et al., 2015), anti-fatigue (Reeve et al., 2010), antitumor (Peifei et al., 2015) and anti-aging properties (Ye et al., 2015), and they scavenge free radicals (Mocan et al., 2014), boost immunity (Su et al., 2014), reduce ischemia/reperfusion injury (Lu and Zhao, 2010), protect against cardiac poisoning (Xin et al., 2007, 2011), and improve reproductive function (Qian and Yu, 2016). Goji (*Lycium* Linn) is a significant medicinal plant, part of which goji fruit (gouqizi) and root (digupi) are two of the most important Chinese medicinal materials for diabetes prevention and treatment. Goji leaf tea, on the other hand, has a hypoglycemic effect on diabetes (Sun et al., 2018).

China is, in fact, the world's largest producer and exporter. Goji production areas in China are estimated at 88,000 ha, covering the entire northwest to central China, including Xinjiang, Ningxia, Gansu, Qinghai, Shanxi, Inner Mongolia and Hubei (Chen et al., 2018). The yield of dry goji berries in China reached 410,608 t in 2017, while the yield in the Ningxia region, which is considered the largest goji planting area, is expected to be approximately 108,473 t (State Forestry Administration of China [SFAC], 2018). As domestic and worldwide market demands grow, the manufacturing capacity of Ningxia Province cannot keep pace (Yao et al., 2018b). Thus, it is necessary to increase goji berry production in more favorable cultivation areas (Cao and Wu, 2015; Gong et al., 2019). Furthermore, all of the cultivars used in the goji industry are for dry fruit cultivation in Ningxia Province, China, and are rarely used for other purposes. The goji industry's constraint is a lack of cultivars for other uses (Nan et al., 2017). As a result, new cultivars with high production yield and for other uses in the goji industry must be developed.

Starting from the genetic origin of *Lycium* fruit, continuously researching and developing improved new varieties on the basis of the original varieties of *Lycium* species and improving the breeding technology system of good varieties of *Lycium* are effective methods to overcoming this bottleneck (Shi et al., 2016). However, there are few reports on the molecular markers and regulatory genes related to edible traits such as the fruit size, shape and hardness of *Lycium* species. With the development of the social economy, consumer demand for wolfberry has increased and includes products for medicinal uses, fresh food, processing, fruits, and vegetables. As such, traditional wolfberry varieties and industrial models have been unable to meet the constantly rich market demand; thus, wolfberry breeding work will be toward selection research on special types of wolfberry varieties to breed and the specificity to meet diverse usage requirements in industry (Nan et al., 2017).

In terms of breeding methods, the same problems that exist in the selection of special cultivars and ordinary cultivars of traditional *Lycium* breeding methods, especially individual plant selection methods, have been emphasized, while biological technology breeding methods have been neglected. With the development of biotechnology, molecular marker-assisted selection (MAS) breeding has been widely used in crop breeding. For example, researchers cloned for the first time a high-yield gene that increases the grain number per panicle after mapping QTLs using populations constructed by Koshihikari (cultivars with tall plants and low yields) and Habataki (cultivars with short plants and high yields). With this gene, a new type of super rice with high yield and lodging resistance was developed by MAS (Ashikari et al., 2005). The same MAS procedure was also successfully executed in disease resistance traits in tomato, taste and virus resistance selection in pepper, and bacterial wilt resistance in eggplant (Foolad and Panthee, 2012; Hanson et al., 2016; Khapte et al., 2018; Moodley et al., 2019). In the breeding of special varieties of *Lycium* species, we should strengthen the organic combination of traditional breeding technology and molecular marker-assisted breeding. It cannot only improve the selection efficiency but also greatly shorten the breeding time (fewer years) to implementing early assisted selection with molecular marker-based traditional breeding as the main method (Nan et al., 2017). The present attention of practical breeding is to develop an increasing number of markers using biotechnology tools to enable molecular-assisted selection to accelerate the delivery of improved varieties to agriculturalists (Rubiales et al., 2021).

As shown in **Figure 1**, the procedure of MAS combined with next-generation sequencing (NGS) breeding mainly includes genetic population construction, morphological trait detection, population sequencing and genotyping, statistical analysis between markers, and QTL/association mapping.

Thus, in this study, we summarize all factors in molecular MAS, including germplasm resources, compatibility between species with karyotypes, candidate traits and their protocols for detection, and the progress of molecular MAS in *Lycium*. Moreover, we also discuss research strategies for different traits and prospective directions of selective breeding in *Lycium* species.

GERMPLASM RESOURCES OF GOJI BERRY

By 2001, approximately 80 species of *Lycium* had been found and named from South America to North America, Australia to the Pacific Islands and Eurasia to the South (Fukuda et al., 2001; Shi et al., 2012), whereas in 2018, the number exceeded 97 (Yao et al., 2018a). *Lycium* species are found in a variety of habitats, primarily in temperate and subtemperate climates (Fukuda et al., 2001; Shi et al., 2012). According to statistics, there are 14 species of goji in Eurasia, 26 in South Africa, and 25 in the southern part of North America. Up to 32 species are widely distributed in South America, whereas there are only three in Australia.

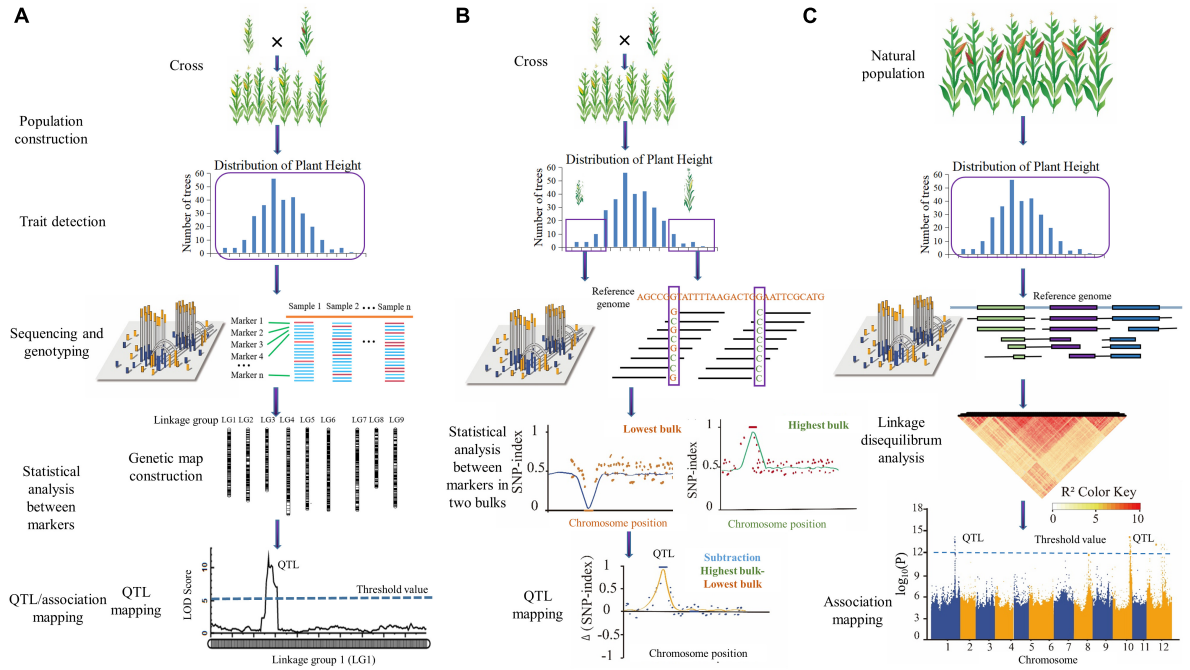

FIGURE 1 | The procedure of MAS combining next-generation sequencing (NGS). **(A)** Linkage mapping; **(B)** bulk segregation analysis (BSA); **(C)** association mapping. Genetic population construction, morphological trait detection, population sequencing and genotyping, statistical analysis between markers, and QTL/association mapping are common procedures in MAS combining NGS.

To date, as shown in **Figure 2**, the *Lycium* resources of 86 countries/regions have been reported, among which Mexico and the US have the most abundant *Lycium* species (Editorial Committee of Chinese Flora, Chinese Academy of Sciences, 1978; Yao et al., 2018a). Goji is primarily found in the northern parts of China (Yao et al., 2018a). There are seven species and three varieties of *Lycium* in China including *L. cylindricum*, *L. yunnanense*, *L. dasystemum* along with mutants (*L. dasystemum* var. rubricaulium), *L. truncatum*, *L. ruthenicum* (**Figure 3A**), *L. barbarum* (**Figures 3D,E**) along with mutants (*L. barbarum* L. var. *auranticarpum* K. F. Ching) (**Figure 3F**), and *L. chinense* Mill. (**Figures 3B,C**) along with mutants [*L. chinense* Mill. var. potaninii (Pojark.) A. M. Lu] (Editorial Committee of Chinese Flora, Chinese Academy of Sciences, 1978; Dong et al., 2008; Yao et al., 2018a), among which five species have medical functions and four can be regarded as food (Editorial Committee of Chinese Flora, Chinese Academy of Sciences, 1978; Li Y. C. et al., 2001; Azadi et al., 2007; Olatunji et al., 2018; Wang et al., 2018; Li et al., 2019a; Ye and Jiang, 2020).

However, among the *Lycium* species, according to the Chinese Pharmacopeia, only *L. barbarum* is recorded as a medicine (Zhang et al., 2001). Several cultivars of *L. barbarum* have been reported, including the Ningqi series (from Ningqi No. 1 to No. 10), Chaiqi series (Chaiqi No. 1, No. 2), Qingqi series (No. 1, No. 2), Jingqi series (from Jingqi No. 1 to No. 6) (Qin and Dai, 2017), and Zhongkelvchuan No. 1 (Yang et al., 2015).

In addition to new species, intercross plants have also been reported, such as *L. ciliatum* × *L. cestroides* (Bernardello et al., 1995), *L. barbarum* × *L. chinense* (Rehman et al., 2020; **Figure 3I**), and *L. barbarum* × *L. yunnanense* (Zhao et al., 2021). In the germplasm repository of the Chinese Academy of Sciences (E 106.04968°, 38.4398563°), offspring of *L. barbarum* and *L. ruthenicum* (**Figures 3J,L**) have also been created. Intercross-demonstration offspring can be used to map QTLs with interspecies crosses.

To provide a reference for germplasm collection in genetic breeding, we summarized the resource diversity in all countries reported (**Figure 2** and **Supplementary Table 1**). These resources with multiple trait variations will provide multiple materials for genetic research on traits related to fruits (color, size, and metabolism), leaves (length, width, and thickness), resistances and so on.

Currently, the *Lycium* germplasm repository at the Goji Engineering Technology Center mainly focuses on Ningxia Province, and it comprises seven species and three varieties distributed naturally in China as well as germplasm imported from the United States, South Korea, Malaysia, and other countries, totaling 60 varieties (lines) and more than 2,000 intermediate materials (1,500 red fruits, 120 yellow fruits, 300 black fruits and 150 other fruits) (NingXia Academy of Agricultural and Forestry Sciences, 2015). The second germplasm resource is in the Chinese Academy of Sciences, which has collected 35 core *Lycium* species resources with different characteristics from all over the world and obtained approximately 10,000 hybrid progenies through intraspecific and interspecific hybridization, providing abundant germplasm resources for breeding and parent plants with rich phenotypes

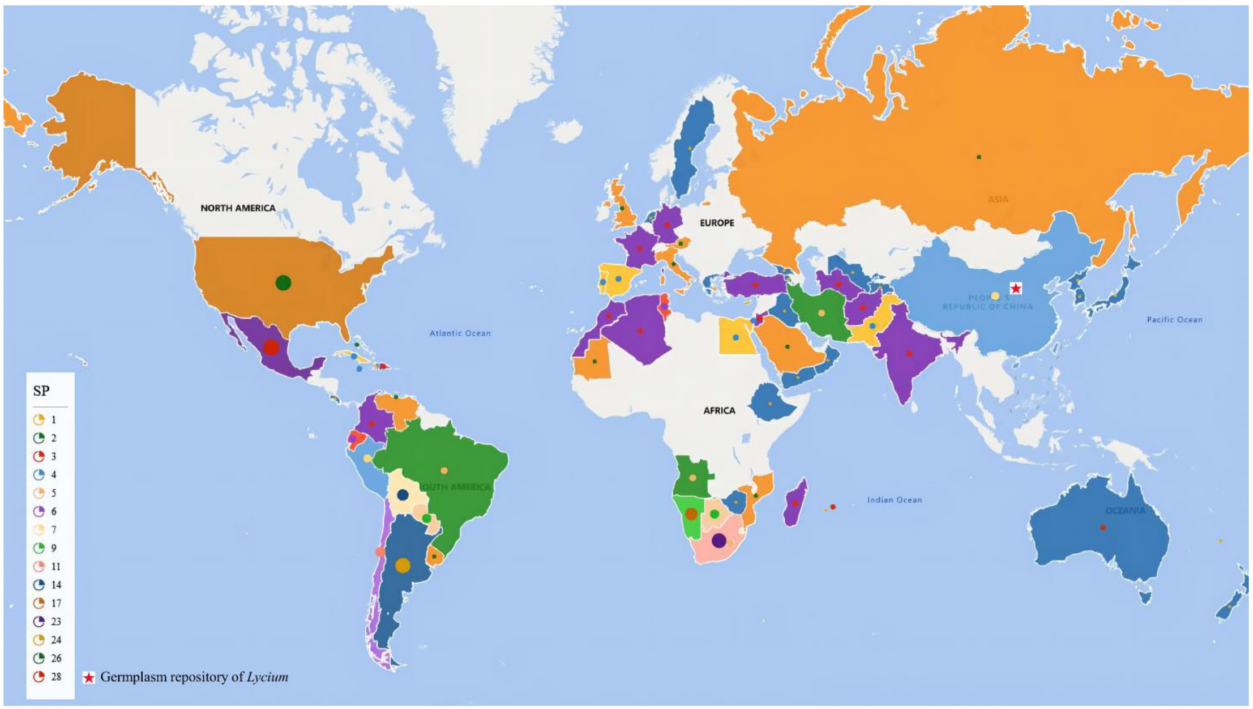

FIGURE 2 | Distribution of *Lycium* species worldwide. SP, species number. Dots with different colors and sizes indicate species numbers in the country marked by them. The red pentacle indicates the location of the germplasm repository of *Lycium*.

for the construction of NAM (nesting association mapping) populations of *Lycium* species.

KARYOTYPING

Cytological examination is a necessary procedure in molecular breeding (Xu, 2010), for validation of both intercrosses and intracrosses in relation to fertility and for avoiding individuals containing translocations and polyploid species containing monosomes or partial chromosomes from being used as mapping parents (Xu, 2010). The cross between tetraploid and diploid *L. barbarum* showed strong cross affinity or backcross affinity; however, the progeny were highly sterile, and the seed satiation rate was very low, making the breeding of triploid seedless plants difficult (Li J. et al., 2001), so it was not conducive to the construction of hybrid populations.

The karyotypes of all of the studied taxa can reveal differences among the species, including size, major visible chromosomal rearrangements, cryptic structural changes, paracentric inversions or reciprocal translocations of segments of similar length (Stiefkens and Bernardello, 2002). A karyotypic analysis showed that *L. barbarum* from Ningxia had 12 pairs of chromosomes ($2n = 24$) (Chen J. et al., 2013). In the karyotype study, 57 kinds of species, varieties, and hybrid plant were used out of a total of 97 species. Most of the *Lycium* species are diploid with 12 pairs of chromosomes (Laura et al., 2010). Some others are teraploid, octaploid, decaploid, and even hendecaploid (Bernardello et al., 1995; Stiefkens and Bernardello, 2005; Chen J. et al., 2013; Stiefkens et al., 2020; **Table 1**).

TRAITS FOR QUANTITATIVE TRAIT LOCUS MAPPING IN *LYCIUM*

Although *Lycium* species are found all over the world, only *L. barbarum* is frequently planted as fruit trees and medicinal plants, particularly in China (Shi et al., 2012). Thus, the majority of breeding selection research has been conducted in China. Although the Chinese government has recognized nearly 10 top varieties of *Lycium*, only "Ningqi No. 1," "Ningqi Cai No. 1," "Mengqi No. 1," "Ningqi No. 4," "Ningqi No. 7," and "Zhongkelvchuan No. 1" have been actively promoted and implemented in production (An et al., 2009; Chen et al., 2018), and the number of new cultivars is still growing. However, the majority of goji fruits on the market are dried fruits in medical use, and fruit juice and fresh fruits are rarely promoted to the market (Huang et al., 2013). As a result, single usage (only dry fruits for medical use) cannot adjust to the demands of industrial product diversification (An et al., 2009). The usage needs to be extended to food, tea, juice, fruit, and chemical extraction uses, such as goji leaf tea, wolfberry tea, wolfberry juice, and wolfberry wine.

Breeders have been grappling with how to standardize the evaluation of germplasm resources and new varieties in recent years, as goji plants have become more widely planted and the demands for germplasm resources have grown (Zhang and Sun, 2011). The State Forestry Administration and the Ministry of Agriculture are currently the main authorities in China for examining and approving new varieties. In anticipation of the

FIGURE 3 | Some collections of Lycium species in the resources. **(A)** *L. ruthenicum*; **(B,C)** *L. chinense*; **(D,E)** *L. barbarum*; **(F)** *L. barbarum* variant; **(G,H,K)** *L. ruthenicum* variant; **(I)** offspring between *L. barbarum* and *L. chinense*; and **(J,L)** offspring cross between *L. barbarum* and *L. ruthenicum*.

acceptance of new *Lycium* variants, the two departments have developed two sets of "Guidelines for the Conduct of Tests for Distinctness, Uniformity, and Stability of *Lycium* species" (DUS) (LY/T2099-2013 and NY/T2528-2013). The number of agricultural qualities covered by the two DUS guidelines is limited, and the breeding requirements cannot be fully met. The DUS testing guidelines for new plant varieties need to be further improved (Zhang and Sun, 2011). Tomato, eggplant, potato, and goji are all members of the Solanaceae family, which is one of the most significant families in agriculture. It possesses a highly conserved genomic sequence and a phenotypic trait set that is extremely diverse. Therefore, comparative genomics across different Solanaceae species can frequently lead to major findings (Bombarely et al., 2010). Currently, international studies on other Solanaceae species, such as tomato and potato, are quite in depth regardless of the variety of features, gene research, or

other themes. Plants have been investigated for more than 1,000 trait descriptions worldwide,[1] of which 387 are for the Solanaceae family (solgenomics.net). The Sol Genetic Network (SGN)[2] is a professional portal that provides genomic and phenotypic data on Solanaceae and allied species. It compiles information on the genomic and phenotypic characteristics of tomato, potato, pepper, eggplant, tobacco, and other species (Bombarely et al., 2010). In contrast to other crops, research on the resource base of *Lycium* species is still in its early stages, and systematic and in-depth research must be strengthened (An et al., 2009). We combine data from SGN, DUS, "Normative description and data standard of *L. barbarum* germplasm resources" (Shi et al., 2012), and other published papers to provide additional traits and their standardized detection methods and QTL mapping for breeders to facilitate standardization of trait research.

Whole-Plant Traits

Many genetic traits, such as ground diameter, crown width, plant height, branch hardness, and natural plant type, are dominated by whole-plant attributes, which can be utilized as the primary reference indices for *Lycium* species breeding selection (Yuan et al., 2013). Trunk diameter (TD), for example, is related to the yield, resistance and growth rate of the plant. Unlike the branch angle or tree height, the trunk diameter is unaffected by trimming and can properly depict the plant's growth state. It is a crucial index for precisely reflecting the plant's growth status, such as the amount of growth (Miller and Scorza, 2010; Wang et al., 2015b). Wang et al. (2015b) mapped four QTL loci influencing the growth of cherry trunk using a high-density genetic map of *Prunus avium* and discovered that the QTL loci controlling the growth of cherry trunk varied with distinct developmental stages (Wang et al., 2015b). These findings suggested that the QTLs governing trunk growth were stage-specific (Liu et al., 2008). To make production management easier, cultivars with a large branching angle, small tree body, and large trunk diameter have been produced (Wang et al., 2015b). Thus, trunk diameter is extremely important in production.

Whole-plant traits, such as plant height, trunk diameter, and internode length, are regulated by genes that can be detected by QTL mapping in *Lycium* and neighboring species (Frary et al., 2003; Barchi et al., 2009; Gong et al., 2019). Thus, mapping of QTLs and gene mining of plant-related traits are useful.

To standardize the detection of whole-plant-related traits, we summarized thirty kinds of traits and their references (**Supplementary Table 2**).

Resistance Traits

Plants of the *Lycium* genus attract a variety of pests due to their lush stems and leaves as well as their sweet fruit juice. According to research and investigation, there are approximately 38 pests in *L. barbarum* in Ningxia, including seven major pests and four diseases, the majority of which are unique to goji plants. If prevention and control are not strengthened in a timely manner, they frequently cause serious goji production problems, even failure, and severely affect yield (Cao and He, 2013). Planting resistant cultivars and spraying fungicides are two possible approaches to controlling pests (Villegas-Fernández et al., 2021). However, continuous use of pesticides can result in loss of biodiversity, environmental contamination, and a slew of ecological issues. Furthermore, widespread contamination of the environment and accumulation in the food chain can pose a risk of teratogenic, genotoxic, oncogenic, and carcinogenic effects on humans (Singh et al., 2017). Thus, stress-resistant cultivar breeding is required.

The first strategy to breeding cultivars with continuous resistance is to select plants with more than resistance genes with the help of robust molecular markers (Rubiales et al., 2021). Research on resistance to biotic stress on neighboring species in Solanaceae, including Capsicum (pepper), Solanum (potato), and Lycopersicon (tomato), showed that these resistances were regulated by resistance genes (R genes). Moreover, homologous genes in the R family (Pto, N, Sw-5, I2, and Prf) in pepper were discovered in syntenous sites in genomes of other solanaceous species and occasionally also discovered to extra sites neighboring phenotypically defined R genes in Solanaceae (Grube et al., 2000). Because resistance is regulated by resistance genes, QTL mapping can be conducted for resistance traits. QTL mapping can be used to further understand the genetic basis of various traits, including resistance to biotic stress (Iordăchescu et al., 2020).

The obtainability of a variety of resistant sources and of a dependable screening technique are crucial to execute a resistance selection project. Study on the molecular and inherited foundation of resistance can improve breeding efficiency (Rispail et al., 2020). To date, there are still no reports about genetic breeding in resistance research on *Lycium*. To facilitate genetic research on resistance in *Lycium* species and genetic resource conservation and evaluation, standardized methods to evaluate resistance to 11 kinds of insects/disease and references were summarized (**Supplementary Table 3**).

Phenology Traits

As domestic and worldwide market demand grows, the manufacturing capacity of Ningxia Province cannot keep pace (Yao et al., 2018b). Thus, it is necessary to increase goji berry output in other favorable agricultural areas (Barchi et al., 2009; Gong et al., 2019). The study of phenology aids in the development of *Lycium* species variants that are appropriate for various climate conditions. Higher latitudes and altitudes and shorter growing seasons than the current optimal areas limit the number of cultivars growers choose in outdoor production (Kevany et al., 2008). Qinghai Province, for example, is a plateau-climate producing area (Yao et al., 2018b). The harvest season for goji fruit is from July to November each year. However, in September in Qinghai, the temperature drops below 10°C, drastically shortening the harvest season and resulting in lower yield (Zheng et al., 2018). Growing early maturing cultivars is one way to address climate-related constraints. Growers benefit from these varieties as well because the earliest fruit to market in a season might command a higher price (Kevany et al., 2008). Thus, the ability to have fruits mature early is a very desirable and important feature (Kevany et al., 2008).

[1] www.cropontology.org

[2] http://solgenomics.net

TABLE 1 | Cytological characteristics of *Lycium* species.

	Taxon	Code	Chromosome number	Karyotype	TIb	c	r	A1	A2	St	R	References
1	*L. elongatum*	Diploid	24	11 m* + 1 sm	25.01	2.08	1.22	0.16	0.12	1 A	–	Stiefkens and Bernardello, 1996
2	*L. afrum*	Haploid	12	–	–	–	–	–	–	–	–	Minne et al., 1994
3	*L. ameghinoi*	Diploid	24	11 m* + 1 sm	20.2	1.7	1.23	0.17	0.16	2A	1.7	Stiefkens and Bernardello, 2005
4	*L. americanum*	Diploid	24	11 m* + 1 sm	22.74	1.89	1.22	0.16	0.12	1 A	–	Stiefkens and Bernardello, 1996
5	*L. amoenum*	Diploid	24	10 m* + 2 sm	26.76	2.23	1.29	0.2	0.1	1A	1.41	Laura et al., 2010
6	*L. arenicolum*	Triploid	36	–	–	–	–	–	–	–	–	Minne et al., 1994
7	*L. barbarum*	Diploid	24	12 m	49.99	4.17	1.23	–	–	–	1.51	Chen J. et al., 2013
8	*L. bosciifolium*	Diploid	24	11 m* + 1 sm	25.04	2.09	1.25	0.24	0.12	1A	1.49	Laura et al., 2010
9	*L. cestroides*	Diploid	24	11 m* + 1 sm	25.78	2.14	1.22	0.17	0.12	1 A	–	Stiefkens and Bernardello, 1996
		Diploid	24	11 mb + 1 sm	19.19	1.6	1.21	0.15	0.14	–	–	Bernardello et al., 1995
10	*L. chanar*	Diploid	24	11 m* + 1 sm	22.68	1.89	1.17	0.11	0.14	2A	–	Stiefkens and Bernardello, 2002
11	*L. chinense*	Diploid	24	12 m	50	4.17	1.31	–	–	–	1.93	Chen J. et al., 2013
12	*L. chilense* var. chilense	Diploid/ Tetraploid	24/48	11 m* + 1 sm	20	1.7	1.14	0.1	0.13	1A	1.52	Stiefkens and Bernardello, 2005
13	*L. chilense* var. confertifolium	Tetraploid	48	22 m + 2 sm	22.61	1.88	–	–	–	–	–	Stiefkens et al., 2020
14	*L. chilense* var. descolei	Diploid/ Tetraploid	24/48	11 m* + 1 sm	23.9	2	1.19	0.14	0.12	1A	1.46	Stiefkens and Bernardello, 2005
15	*L. chilense* var. Descolei	Tetraploid	48	22 m + 2 sm	26.32	2.19	–	–	–	–	–	Stiefkens et al., 2020
16	*L. chilense* var. filifolium	Diploid/ Tetraploid	24/48	11 m* + 1 sm	30	2.5	1.2	0.15	0.12	1A	1.51	Stiefkens and Bernardello, 2005
17	*L. ciliatum*	Diploid	24	11 m* + 1 sm	21.2	1.7	1.21	0.15	0.11	2A	1.4	Stiefkens and Bernardello, 2005
		Tetraploid	48	22 m + 2 sm	22.03	1.84	–	–	–	–	–	Stiefkens and Bernardello, 2002
18	*L. ciliatum*	Diploid	24	11 mb + 1 sm	20.57	1.71	1.24	0.17	0.11	–	–	Bernardello et al., 1995
19	*L. cuneatum*	Diploid	24	11 m* + 1 sm	20.34	1.69	1.18	0.12	0.14	2A	–	Stiefkens and Bernardello, 2002
20	*L. depressum*	Diploid	24	11 m + 1 sm	56.48	4.71	1.44	–	–	2A	–	Sheidai et al., 1999
		Diploid	24	12 m	64.08	5.34	1.3	–	–	1A	–	Sheidai et al., 1999
		Diploid	24	12 m	67.18	5.60	1.83	–	–	1A	–	Sheidai et al., 1999
21	*L. ferocissimum*	Diploid	24	10 m* + 2 sm	22.16	2.02	1.3	0.21	0.08	1A	1.29	Laura et al., 2010
22	*L. fremontii*	Octaploid	96	–	–	–	–	–	–	–	–	Chen J. et al., 2013
23	*L. fuscum*	Diploid	24	11 m* + 1 sm	22.7	1.89	1.22	0.16	0.14	2A	–	Stiefkens and Bernardello, 2002
24	*L. gilliesianum*	Diploid	24	11 m* + 1 sm	20.66	1.72	1.25	0.18	0.17	2A	–	Stiefkens and Bernardello, 2002
25	*L. horridum*	Diploid	24	–	–	–	–	–	–	–	–	Minne et al., 1994
26	*L. humile*	Diploid	24	11 m + 1 sm	21.45	1.79	1.34	0.16	0.11	-	1.71	Stiefkens et al., 2020
27	*L. infaustum*	Diploid	24	11 m* + 1 sm	21.52	1.79	1.25	0.18	0.13	1 A	-	Stiefkens and Bernardello, 1996
28	*L. kopetdaghi*	Diploid	24	12 m	50.91	4.24	1.38	–	–	1A	–	Sheidai et al., 1999
29	*L. makranicum*	Diploid	24	12 m	51.36	4.28	1.32	–	–	1A	–	Sheidai et al., 1999
30	*L. minutifolium*	Diploid	24	11 m* + 1 sm	24.93	2.07	1.21	0.16	0.08	1A	–	Stiefkens and Bernardello, 2002
31	*L. morongii*	Diploid	24	11 m* + 1 sm	23.23	1.94	1.2	0.14	0.12	1A	–	Stiefkens and Bernardello, 2002

(Continued)

TABLE 1 | (Continued)

	Taxon	Code	Chromosome number	Karyotype	Tlb	c	r	A1	A2	St	R	References
32	*L. nodosum*	Diploid	24	11 m* + 1 sm	20.03	1.67	1.27	0.19	0.13	2A	–	Stiefkens and Bernardello, 2002
33	*L. oxycarpum*	Diploid	24	10 m* + 2 sm	23.22	1.93	1.3	0.2	0.08	1A	1.3	Laura et al., 2010
34	*L. pallidum*	Diploid	24	10 m + 2 sm	49.99	4.17	1.22	–	–	–	1.55	Chen J. et al., 2013
35	*L. rachidocladum*	Diploid	24	11 m* + 1 sm	28.01	2.33	1.28	0.20	0.16	1 A	–	Stiefkens and Bernardello, 1996
36	*L. repens*	Hendecaploid	132	–	22.68	1.89	–	–	–	–	–	Stiefkens et al., 2020
37-1	*L. ruthenicum*	Diploid	24	9 m + 3 sm	70.87	5.91	1.5	–	–	2A	–	Sheidai et al., 1999
37-2	*L. ruthenicum*	Diploid	24	9 m + 3 sm	66.84	5.57	1.52	–	–	2A	–	Sheidai et al., 1999
37-3	*L. ruthenicum*	Diploid	24	12 m	50.18	4.18	1.25	–	–	–	1.53	Chen J. et al., 2013
38	*L. shawii*	Diploid	24	11 m + 1 sm	40.92	3.41	1.35	–	–	2A	–	Sheidai et al., 1999
39	*L. stenophyllum*	Diploid	24	11 m* + 1 sm	22.9	1.9	1.17	0.12	0.13	1A	–	Stiefkens and Bernardello, 2002
40	*L. tenue*	Diploid	24	10 m* + 2 sm	21.65	1.8	1.28	0.18	0.1	2A	1.43	Laura et al., 2010
41	*L. tenuispinosum*	Diploid	24	11 m* + 1 sm	–	–	–	–	–	–	–	Stiefkens and Bernardello, 1996
42	*L. tenuispinosum* var. friesii	Diploid	24	11 m* + 1 sm	25.29	2.10	1.21	0.15	0.12	1 A	–	Stiefkens and Bernardello, 1996
43	*L. tenuispinosum* var. tenuispinosum	Diploid	24	11 m* + 1 sm	23.63	1.96	1.21	0.16	0.1	1 A	–	Stiefkens and Bernardello, 1996
44	*L. tenuispinosum* var. calysinum	Diploid	24	11 m* + 1 sm	21.52	1.79	1.21	0.15	0.12	1 A	–	Stiefkens and Bernardello, 1996
45	*L. tetrandrum*	Triploid	36	–	–	–	–	–	–	–	–	Minne et al., 1994
46	*L. villosum*	Diploid	24	–	–	–	–	–	–	–	–	Minne et al., 1994
47	*L. vimineum*	Diploid	24	11 m* + 1 sm	20.01	1.67	1.24	0.17	0.13	2A	-	Stiefkens and Bernardello, 2002
48	*L. australe*	Diploid	24	10 m + 2 sm	31.43	2.62	1.25	0.16	0.12	–	1.61	Stiefkens et al., 2020
49	*L. berlandieri*	Diploid	24	10 m + 2 sm	38.7	3.23	1.36	0.22	0.12	–	1.55	Stiefkens et al., 2020
50	*L. californicum*	Diploid	24	11 m + 1 sm	29.8	2.48	1.19	0.14	0.1	–	1.46	Stiefkens et al., 2020
		Tetraploid	48	22 m + 2 sm	39.65	3.3	–	–	–	–	–	Stiefkens et al., 2020
51	*L. exsertum*	Tetraploid	48	20 m + 4 sm	40.26	3.35	–	–	–	–	–	Stiefkens et al., 2020
52	*L. fremontii*	Decaploid	120	–	29.24	2.44	–	–	–	–	–	Stiefkens et al., 2020
53	*L. intricatum*	Diploid	24	10 m + 2 sm	33.59	2.8	1.24	0.21	0.16	-	1.81	Stiefkens et al., 2020
54	*L. pallidum*	Diploid	24	10 m + 2 sm	29.98	2.5	1.37	0.25	0.11	-	1.49	Stiefkens et al., 2020
55	*L. parishii*	Diploid	24	11 m + 1 sm	30.67	2.55	1.15	0.12	0.08	-	1.36	Stiefkens et al., 2020
56	*Lycium* sp.	Tetraploid	48	20 m + 4 sm	24.52	2.04	–	–	–	–	–	Stiefkens et al., 2020
57	Hybrid (*L. ciliaturn* × *L. cestroides*)	Diploid	24	11 mb + 1 sm	22.23	1.85	1.22	0.16	0.13	–	–	Bernardello et al., 1995

Tl, mean total haploid chromosome length; C, mean chromosome length; r, mean arm ratio. Mean asymmetry indices: A1, intrachromosomic; A2, interchromosomic; St, Stebbins' (1971), category of asymmetry; R, ratio between largest and smallest chromosomes in complement. Lengths in μm. m, metacentric chromosome; sm, submetacentric chromosome. An asterisk indicates that the first chromosome pair has a satellite on the short arm.

Early ripening occurs when the ethylene receptor LeETR4 is suppressed, but fruit size, yield, and flavor-related chemical composition remain mostly intact. Moreover, gibberellins (GAs) were also found to have negative effects on fruit ripening. Restriction of GAs can activate ripening regulator genes such as CNR, RIN, and NOR, mediating the biosynthesis of ethylene.

Further study showed that the SlGA2ox1 gene regulated GA catabolism, resulting in reductions in GA levels in fruit tissues and leading to early ripening in tomato (Li et al., 2019b). Biotechnology could provide important tools for the creation of early ripening cultivars as our understanding of the molecular control of fruit ripening grows (Kevany et al., 2008). Thus, to apply MAS in phenology trait QTL mapping and standardize the detection method, we summarized 25 phenology traits in **Supplementary Table 4**.

Fruit Traits

The fruit size of *L. barbarum* impacts the price of dry goji fruits and is one of the most important indicators of goji fruit grading (Ma et al., 2017). The higher the wolfberry grade and the larger the fruit size, the higher the price. Therefore, in regard to goji cultivation, breeders are more likely to choose types with large fruit sizes and high yields. Fruit weight, fruit length, and fruit breadth are subdivided into indices for analyzing goji fruit size (An et al., 2007). Meanwhile, crop yield is considered positively correlated with fruit weight, fruit length, and fruit width, showing that superior farming varieties with large fruit length, width, and weight cannot only improve fruit quality but also increase yield (Henareh et al., 2015). Artificial domestication can boost the productivity and quality of numerous crops. When compared with predomestication, the size and weight of tomato fruits grew dramatically with favorable features following long-term artificial domestication (Tanksley, 2004).

Fruit weight in most tomato species is controlled by a major QTL (*fw2.2*) (Alpert et al., 1995). In *Lycium* species, QTLs for fruit traits were first detected by Zhao et al. (2019) with an interspecies genetic map, and 41 stable QTLs were detected (Zhao et al., 2019). QTL mapping for fruit traits was executed by Rehman et al. (2020), with two stable QTLs (Rehman et al., 2020). Genetics are a key factor for fruit traits, and it is feasible to map QTLs with a combination of genetic mapping data and morphological data. Standardized techniques were used to determine the 129 fruit-related traits (**Supplementary Table 5**).

Leaf Traits

Leaf traits of *Lycium* include leaf length, leaf width, leaf thickness, petiole length, leaf color et al. (State Forestry Administration of China [SFAC], 2013). The distance between the upper and lower surfaces of the blade is measured as the leaf thickness (LT) (Coneva et al., 2017). Leaf thickness is an important leaf shape parameter for crop plant type improvement and serves as a reference index in the breeding of high-yield crop varieties (Chen D. et al., 2015). Leaf thickness and crop yield have been discovered to have a substantial positive association in rice studies. Plants with thick leaves produce more fruit (Liu C. et al., 2014). As a result, genetic improvement of crop leaf thickness is critical for increasing crop output (Chen D. et al., 2015). Meanwhile, research has discovered that thicker leaves are better for plant growth in arid environments (Coneva et al., 2017).

Photosynthesis is a leaf feature that describes the process by which plants receive light energy and transform it into chemical energy, and it is the foundation of plant organic matter synthesis (Richards, 2000). It is critical to a plant's ability to synthesize its own chemicals (Long et al., 2006; Zhu et al., 2012). Photosynthetic features include the photosynthetic rate (Pn), stomatal conductance, intercellular carbon dioxide, the transpiration rate, the limiting value of the stoma, water usage efficiency, the chlorophyll content, and a variety of other variables (Qu et al., 2017). Research on other plants shows that polygenes play a major role in photosynthesis (Yang et al., 2007; Kumar et al., 2012; Czyczylo-Mysza et al., 2013). Moreover, Gong et al. (2019) detected 29 QTLs for photosynthetic traits, among which eight are stable QTLs (Gong et al., 2019), indicating that photosynthesis in *Lycium* species is regulated by genes and is applicable to mapping QTLs in *Lycium* species.

Thus, we provide 33 kinds of leaf trait descriptions and their references for standardized detection (**Supplementary Table 6**).

Flower Traits

Flowering transition, meristem identity choice, floral organ initiation, and floral organ morphogenesis are the four major steps of tomato flower development (Liu J. et al., 2014). (1) Plants respond to the external environment and their own signals during the flowering transition stage, transitioning from vegetative to reproductive growth. This process is controlled by a set of genes associated to flowering time (Liu J. et al., 2014). In tomato, two flowering control routes have been identified: photoperiod and autonomic. The photoperiod pathway is regulated by uniflora (*UF*) and compound inflorescence (*S*) gene, whereas the autonomous flowering pathway is regulated by single flower truss (*SFT*) gene and jointless (*J*) gene (Dielen et al., 2004; Molinero-Rosales et al., 2004; Liu J. et al., 2014). Flowering time-related genes *S* and *J* are positioned downstream of *UF*, which is a crucial gene in tomato that regulates flowering time (Quinet et al., 2006). Florigen signals are stimulated by SFT protein. The overexpression of *SFT* gene could stimulate flowering in tomatoes (Lifschitz and Eshed, 2006). (2) Plants respond to signals from diverse flowering time regulatory pathways to activate meristem characteristic genes and define meristem attributes during the meristem identity choice stage. After flowering induction, the vegetative meristem becomes the inflorescence meristem (IM) (Liu J. et al., 2014). Lateral organs replace leaves on the inflorescence meristem, which is followed by the flower meristem (FM) (Chandler, 2012). In tomato, the IM motifs SFT, UF, J, and Macrocalyx (MC) have been identified (Liu J. et al., 2014). (3) Meristem signature genes activate floral organ signature genes in various areas during the floral organ initiation stage. Meristem identity genes are critical genes that control the process of IM formation FM and determine the transition from flowering to floral organogenesis. Falsiflora (*FA*) and anantha (AN) were the FM-regulated genes in tomato (Liu J. et al., 2014). (4) Floral organ characteristic genes activate downstream organ morphogenesis genes during the floral organ morphogenesis stage, determining the precise cell types and tissues that make up each organ (Jack, 2004; Liu J. et al., 2014). The ABCE model is used to describe the mechanism of floral organ development (Krizek and Fletcher, 2005). The ABC genes are initially activated by floral meristem identification genes including leafy (*LFY*), apetala 1 (*AP1*), *AP2*, and unusual floral

organs (*UFO*) gene, but later expression patterns are adjusted through interactions between the ABC genes. Four activities are present in adjacent whorls of the flower: A, B, C, and E. These four activities are thought to work together with different combinations to determine the identity of the flower's four organs: sepals, petals, stamens, and carpels (Jack, 2004; Liu J. et al., 2014). In the whole steps, Auxin is also the key factor in determination of the discrimination and quantity of floral organs. The disorder of auxin signaling, transportation or biosynthesis always results in failure of floral discrimination (Cheng and Zhao, 2007).

The relative location of the pistil and stamen has a significant impact on plant fertility. The pistil is more exposed than the stamen in many plant species, and it is the main feature of cross-pollination (Chen and Tanksley, 2004). In hybrid systems of crops, particularly in the application of hybrid rice, the stigma exertion rate plays a vital function (Zhou et al., 2017). The stigma and the top of the stamen tube are almost equal in most wild tomatoes, indicating cross-pollination, whereas in domesticated tomatoes, the stigma and the top of the stamen tube being almost equal indicates self-pollination (Ranc et al., 2008).

The relative location of the pistil and stamen has a significant impact on plant fertility (Chen and Tanksley, 2004). Cross-pollination occurs in the majority of tomato species with exposed stigmas. On the other hand, most tomato species with similar lengths of pistils and stamens or with pistils shorter than the stamen display autocompatibility (Chen and Tanksley, 2004). The pistil of the sterile male varietal Ningqi No. 5 was substantially higher than the stamen among cross-pollinated *L. barbarum* varieties (Chen C. et al., 2017). Thus, the relative location of the pistil and stamen could be an indicator for identifying selfing fertility in *Lycium* plants.

In this section, we describe 27 flower-related traits and their detection references for standardizing the detection method.

Seed Traits

The seed is the primary site for endogenous hormone synthesis and accumulation, which is critical for fruit development. Previous studies in several fruits have revealed a strong link between fruit seed number and fruit size (Zheng et al., 2015). The incidence of the preharvest decline of fruits is associated with the loss of hormones from the seeds of almost ripe apples (Luckwill, 1948). Seed weight, length, and width have a substantial effect on fruit size development in *L. barbarum* according to previous studies. Seed number, on the other hand, has a greater impact on fruit size development than seed weight (Zheng et al., 2015; Wang J. et al., 2017). For the eight evaluated peach rootstocks, the fresh mass of fruits exhibits a strong positive association with the fresh mass of seeds (Souza et al., 2016). The fresh mass of seeds has a strong association with the germination speed index and mean germination time (Souza et al., 2016). Therefore, it can also serve as a predictor of plant reproductivity. Fruit yield was found to be positively associated with average seed yield and fruit weight per plant in a tomato study. The seed vigor index has the greatest direct effect on seed yield and is regarded as a useful measure for predicting seed lot performance in the field (Sharma, 2012).

According to research, there is a link between fruit development and seed development (Zheng et al., 2012). It reveals that the initial rapid growth phase of the fruit is also the first rapid growth period of the seed, but the seed's growth ratio is faster than the fruit's, and the seed's endosperm grows rapidly during this period. Seeds had slow development characteristics when fruits entered the slow growth stage, and the rate of rise in seed length and width was much lower than the rate of growth of seeds in the initial fast growth stage, during which seeds mostly underwent embryo differentiation. Fruit volume and weight increased fast during the second rapid development stage, although seed length and width increased very slightly. In this stage, only the embryo grew larger (Zheng et al., 2012). Goji seed research shows a similar trend to tomato seed development (Wm Hilhorst et al., 1998). In the early stages, a significant increase in ABA content is invariably followed by an increase in seed weight, however, in the later stages, the ABA content falls and the seed enters the sluggish development stage (Wm Hilhorst et al., 1998).

In seed development stage, a variety of metabolites and jasmonic acid (JA) accumulate in certain organs and tissues. Jasmonate-inducible genes, are expressed in tandem with the accumulation, including developmentally regulated genes and defense genes (Wasternack et al., 2013). In seed formation, coordinated expression of embryo and endosperm is necessary, which is aided by both sporophytic and gametophytic genes from parents (Chaudhury et al., 2001). In the early stages of seed formation, the *SlNAC3* gene (No apical meristem, *NAM*; *ATAF1*/*ATAF2*, and Cup-shaped cotyledon, *CUC* genes in tomato, Solanaceae) in tomato regulates embryo development, while factor for inversion stimulation (*FIS*) regulates endosperm development (Chaudhury et al., 2001; Han et al., 2014).

Although the chromosome ploidy of parents has an effect on seed fertility and the seed satiation rate (Li J. et al., 2001), in the case of good parental compatibility, the seed traits are mainly regulated by genetics. Research on Arabidopsis show that seed weight, size as well as the accumulation of seed oil and protein are regulated by APETALA2 (AP2) (Jofuku et al., 2005). Research on overexpression and mutation of da1-1 allele demonstrated that genes in this family play a key role as controller of organ and seed size in vegetation (Li et al., 2008). Thus, the feasibility of QTLs for seed traits was validated.

Here, eight kinds of traits and their standardized reference are provided (**Supplementary Table 8**).

Metabolites

The plant metabolome has a wide range of functions and is frequently viewed as a link between the genome and the phenome (Chen et al., 2016). It can make the link between genes and agronomic features easier. We found additional candidate genes potentially responsible for the variation in traits such as grain color and size using parallel mGWAS (metabolic GWAS) and pGWAS (phenotype GWAS) and provided evidence of a metabotype-phenotype linkage (Chen et al., 2016). Thus, research on mQTLs and mGWAS is crucial. With the

advancement of liquid chromatography–mass spectrometry (LC–MS) technology, a large number of metabolites can be detected on a large scale with a targeted or untargeted metabolic profile, relative content, or absolute content (Chen W. et al., 2013; Chen et al., 2016). Khan et al. (2012) discovered 418 metabolites in apple peel and 254 in apple pulp using untargeted LC–MS metabolic profiling, of which 50% in the peel and 44% in the pulp had their mQTLs successfully identified (Khan et al., 2012). Chen et al. (2016) discovered 587 metabolites in cereal using a targeted metabolic profile, of which 331 (56.4%) showed at least one significant relationship, indicating a link between metabotype and phenotype (Chen et al., 2016). Shi created a broad-targeted metabolome approach for detecting 612 compounds in goji fruits in 2019 (Shi et al., 2019). Thus, LC–MS may be used to detect a large number of metabolites for mQTLs (or mGWASs). As **Supplementary Table 9** shows, 870 metabolic traits and their reference detection methods have been described (**Supplementary Table 9**).

In addition, although the metabolic traits include the basic nutritional and medical functions of goji, according to *International Standard* of wolfberry (ISO 23193:2020), *American Herbal Pharmacopoeia*, *British Pharmacopoeia Commission*, *European Pharmacopoeia*, *The standards for wolfberry* in China (GB/T 18672-2014), *Chinese Pharmacopoeia*, *Korean Pharmacopoeia*, and *Vietnam Pharmacopoeia*, polysaccharide, total sugar, protein, zeaxanthin, zeaxanthin dipalmitate, betaine, and ash contents are regarded as quality indicators in wolfberry, and directly determine the level of quality (Korea Food and Drug Administration, 2014; Chinese Pharmacopoeia Commission, 2015; European Directorate for the Quality of Medicines, 2016; Japanese Pharmacopoeia Editorial Committee, 2016; British Pharmacopoeia Commission, 2017; Yao et al., 2018a; The American Herbal Pharmacopoeia, 2019). Thus, they can be key traits in breeding.

Postharvest Traits

The fruit processing and storage procedures for *L. barbarum* are determined by its postharvest quality. For example, Ningqi No. 1 and Ningqi No. 7 (elite *L. barbarum* cultivars) have excellent yield and disease resistance and are suitable for the production of dried fruits, but they are not suitable for the production of fruit juice due to their bitter and heavy medicinal flavor. Wolfberry with yellow fruits is ideal for making fruit juice because it has no unpleasant tastes, is bright in color, and leaves little residue after juicing (Wang et al., 2016). By examining the changes in the decay index and physiological indices of fresh *L. barbarum* fruit after harvest, Ge et al. (2008) discovered the ideal storage temperature for fresh fruits after harvest (Ge et al., 2008). Thus, mapping QTLs for postharvest traits can help to accelerate cultivar selection for the goji industry for different uses. In tomato, fruit firmness was validated to be regulated by GA2-oxidase, which is encoded by the *FIS1* gene (Li et al., 2020). In *Lycium* species, Rehman et al. (2020) also mapped one QTL (*qFF10–1*) for fruit firmness traits (Rehman et al., 2020), confirming that fruit firmness is a gene-regulated trait and that mapping QTLs by linkage mapping is feasible. Twenty-three kinds of postharvest traits are described in this section for further research (**Supplementary Table 10**).

PROGRESS OF MARKER-ASSISTED SELECTION TECHNOLOGY IN *LYCIUM*

Genetic Marker Development

Molecular markers are based on nucleotide sequence changes, directly reflecting DNA sequence variations. Because molecular markers are not affected by environmental changes, development, differentiation, or cellular defense, compared with traditional phenotypic markers, they are more stable and observable (Agarwal et al., 2008).

The ideal genetic markers have the following characteristics: high genetic polymorphism, high reproducibility (the results can be repeated in different laboratories), codominance (the heterozygosity and homozygosity can pass an experimental test), low price (low cost of marker development and genotype identification), simple operation (automated operation, procedure, and sequencing), neutral selection (no multiple gene effects), markers uniformly distributed throughout the genome, and clear differentiation of alleles (Xu, 2010). Traditional molecular marker techniques include restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), and NGS-based single nucleotide polymorphism (SNP) (Blenda et al., 2012).

In 2000, molecular markers such as RAPD were first applied in *Lycium* species for the validation of the genetic diversity of *L. barbarum* (Cheng et al., 2000), while RAPD was also validated for its feasibility to distinguish close species of *Lycium* (Zhang et al., 2001). To date, simple sequence repeat (SSR), conserved ortholog set II (COSII), characterized amplified region (SCAR), amplified fragment length polymorphism (AFLP), restriction site−associated DNA sequencing (RAD-seq), internal transcribed spacer (ITS), sequence-related amplified polymorphism (SRAP), intersimple sequence repeat (ISSR), and random amplified, microsatellite polymorphism (RAMP)-PCR markers have been applied in genetic diversity, population structure, morphological variation, phylogenetic inference, traceability, and cultivar/species identification and discrimination (Zhang et al., 2001, 2018; Yin et al., 2005; Sze et al., 2008; Chung et al., 2009; Kwon et al., 2009; Levin et al., 2009; Zhao et al., 2010; Balasubramani et al., 2011; Liu et al., 2012, 2020b,c; Tripathi, 2013; Xin et al., 2013; Chen and Zhong, 2014; Wang et al., 2015a; Chen C. et al., 2017; Chen J. et al., 2017; McCulloch et al., 2020; Jung et al., 2021). SSR markers can be used for the detection of genetic polymorphisms of species, calculation of genetic relationships (genetic distance) between varieties, identification of cultivars, and even construction of genetic maps and QTL mapping with sufficient SSR markers (Chen H. et al., 2015; Essid et al., 2015; Ibrahim et al., 2016; Thammina et al., 2017; Portis et al., 2018b). One hundred fifteen SSR and 12 ILP markers were applied to construct a genetic map with the F1 population (Hu, 2015). However, because of the limited number of markers and huge average genetic distance

between two adjacent markers, QTL mapping was not suitable (Hu, 2015). To address this problem, Chen C. et al. (2017) developed 22,537 EST-SSR markers using transcriptomes of different fruit ripening stages.

Linkage Mapping

Linkage mapping is based on the frequency of recombination between two markers, which can be used to determine their genetic distance on chromosomes. The greater the genetic distance between two markers on a chromosome, the higher their exchange frequency (Xu, 2010). A genetic map depicts the linear order of DNA markers on a chromosome based on their relative positions (Madhusudhana, 2015). The linkage mapping procedure mainly includes constructing mapping population(s), trait data collection, sequencing and genotyping with polymorphic markers, genetic map construction, and QTL mapping (marker-trait associations between genotypic and trait phenotypic data Madhusudhana, 2015; **Figure 1A**). Linkage mapping can be applied under QTL mapping for multiple traits at the same time (Meuwissen and Goddard, 2004). The linkage mapping (genetic map) population is always constructed with a cross of two parents (Yu et al., 2008). There are several advantages of linkage mapping: (1) The complexity of the population structure is reduced due to parental confirmation, which improves accuracy and reduces false-positives; and (2) the mapping population is constructed by combining the offspring of two parents. Even if the features are regulated by rare genes, the offspring contains a significant proportion of unusual genes. As a result, linkage mapping may be able to locate rare genes (Mackay and Powell, 2007; Yu et al., 2008).

Linkage mapping has been successfully applied in *Lycium* species (**Table 2**). The intraspecies genetic map consisted of 23,967 markers with a size of 964.03 cM (Gong et al., 2019). According to the genome published by Cao, the size of the genome in *L. barbarum* is 1.67 Gb. Thus, we determined that the average DNA content between two neighboring markers (genome size/marker number) was 69,679 bp. However, the two interspecies genetic maps had higher genetic distances (1702.45 and 1649.03 cM) and lower marker numbers (6,733 and 3,495 cM) (Zhao et al., 2019; Rehman et al., 2020), and the average DNA content between two neighboring markers was 248,032 and 477,825 bp. Obviously, intraspecies genetic maps have higher resolution than intraspecies genetic maps. However, all three maps still have higher resolution than most other species, such as maize (130,000 bp) and grape (676,000 and 606,000 bp in two parent maps) (Chen et al., 2014; Chen J. et al., 2015), which were successfully mapped to QTLs. Thus, mapping QTLs with linkage mapping is applicable to *Lycium* species.

OTHER MARKER-ASSISTED SELECTION TECHNIQUES WITH POTENTIAL VALUE

Bulk Segregation Analysis

Bulk segregant analysis (BSA) is a QTL mapping technique for locating genomic regions with genetic loci that impact a trait of interest (Michelmore et al., 1991; Magwene et al., 2011). Individuals are assayed for the focal characteristic in a segregating population derived from a genetic cross, and two pools (bulks) of segregants are established by choosing individuals from the phenotypic distribution tails (other sampling designs can also be used as discussed below). Individual genotyping or the production of pooled DNA samples from which allele frequencies are obtained are used to estimate genotype frequencies for the two bulks. In genomic regions with no loci affecting the trait, allele frequencies should be roughly comparable between the two bulks. Allele frequencies should differ between bulks in regions of the genome that include causative loci. High marker density and accurate allele frequency estimation inside bulks are the most effective features of BSA (Ehrenreich et al., 2009; Magwene et al., 2011; Lu et al., 2021).

Although utilization of BSA to map QTLs in *Lycium* species has not yet been reported, BSA, with its cost-effectiveness and efficiency, has been widely used to map QTLs in woody, herbaceous species, including apple and watermelon (Dong et al., 2018; Liu et al., 2021), as well as neighboring species in Solanaceae, such as fruit color in pepper, tomato, and eggplant (Lee et al., 2020; Lu et al., 2021; Tang et al., 2021). Thus, it has the potential for application in *Lycium* species.

The BSA procedure is performed by following key points: establishing a segregating population, collecting trait data and performing bulk selection. Individuals with high and low values for the trait of interest are phenotyped and selected. These individuals' DNAs are pooled into high and low bulks, which are then sequenced and SNP-called, obviating the requirements to design markers ahead of time and to statistically analyze linkage distances among markers of the two bulks. SNPs discovered in reads derived from regions not connected to the trait of interest should be approximately 50% of the reads in bulks selected from populations (**Figure 1B**). Depending on the bulk, SNPs in reads aligning to genomic regions directly associated with the trait should be over- or underrepresented. Under QTL mapping, QTL identification can be accomplished by comparing relative

TABLE 2 | Genetic map construction and QTL mapping in *Lycium* species.

Population type	Parents	Population size	Chr	Length (cM)	No. of markers	Marker interval (cM)	Max interval (cM)	Traits	QTL	References
Intraspecies	Lb × Lb	305	12	964.03	23,967	0.04	1.98	6	32	Gong et al., 2019
Interspecies	Lb × Lc	302	12	1702.45	6,733	0.31	11.7	7	55	Zhao et al., 2019
Interspecies	Lc × Lb	305	12	1649.03	3,495	0.47	16.99	11	117	Rehman et al., 2020

Lb, L. barbarum; Lc, L. chinense; Chr, chromosome.

allele depths or SNP indices (defined as the number of reads containing an SNP divided by the total sequencing depth at that SNP) between the bulks (Mansfeld and Grumet, 2018). BSA is a cost-effective and efficient approach to fine-mapping genes in unsequenced genomes by targeting SNPs to specific genomic intervals (Trick et al., 2012).

Association Mapping

Genome-wide association study (GWAS) is an analytical method based on linkage disequilibrium of genes or loci among biological populations, combining genotype and phenotypic data from mapping populations and analyzing the relationship between detection markers or loci and traits using statistical methods (Yu and Buckler, 2006). The procedure of association mapping comprises population selection, morphological data collection, NGS and genotyping, linkage disequilibrium analysis, and association mapping between markers and traits (Zhu et al., 2008; **Figure 1C**).

GWAS mapping technique is regarded as a substitute for linkage mapping for research on association between genotype and phenotype. Based on recombination to reorganize the genome, both techniques are for evaluating associations between genotype and phenotype. Linkage disequilibrium (LD), the term to evaluate the non-random connotation between markers, was the basis of GWAS (Barchi et al., 2019). GWAS offers three advantages over linkage analysis: (1) High precision and resolution. Linkage disequilibrium (LD) declines quickly after multiple hybridization, recombination, and mutation information from associated populations, and localization precision is considerably increased with high resolution. (2) Without the constraints of the parental population, multiple gene types can be discovered on the same locus. (3) Time efficiency is achieved by using natural populations as a source of material, which reduces the time required to produce genetic populations (Yu and Buckler, 2006; Madhusudhana, 2015; Hu et al., 2018). To date, there is still no report about the utilization of association mapping in *Lycium* species.

However, utilization in neighboring species in Solanaceae, such as QTLs for leaf-, plant-, and fruit-related traits in eggplant (Portis et al., 2015), virus resistance in pepper (Tamisier et al., 2020), and multiple root- and fruit quality-related traits in tomato (Tamisier et al., 2020), demonstrates potential utilization in the breeding of *Lycium* species.

MAPPING POPULATION AND STRATEGIES OF HIGH-DENSITY GENETIC MAP CONSTRUCTION IN *LYCIUM*

Classification of Mapping Populations

The mapping population can be classified into anther culture (AC), backcross population (BC), backcross inbred line (BIL), double haploid (DH), intermating (IM), near-isogenic line (NIL), recombinant inbred line (RIL), testcross (TC), and triple testcross (TTC) (Xu, 2010; **Figure 4A**).

Mapping With F_1 Population

Self-incompatibility is common among woody plants. After a lengthy period of evolution, *Lycium species*, like other woody plants, are genetically polymorphic and extremely heterozygous (Zhao et al., 2010). Goji are woody plants with a long development cycle, high genomic heterozygosity, and difficulty acquiring homozygous genes, making the genetic map of woody plants more difficult to generate than that of herbaceous species (Song et al., 2012; Guo et al., 2015; Zhang et al., 2015). However, the high heterozygosity of woody plants satisfies the requirements for genetic map construction and QTL mapping using the F_1 population. Thus, all genetic maps of woody plants and some highly heterozygous perennial herbs, such as *Prunus avium*, *Juglans regia*, *Camellia sinensis*, *Paeonia* Sect., *Prunus mume,* and *Cynara cardunculus*, are currently based on F_1 populations (Cai et al., 2015; Ma et al., 2015; Wang et al., 2015b; Zhang et al., 2015; Zhu et al., 2015; Portis et al., 2018a). To date, three genetic maps of *Lycium* species have been constructed using the F_1 population. The first high-density genetic map was constructed with an intraspecies cross population of *L. barbarum* (Gong et al., 2019), while the other two were *L. barbarum* and *L. chinense* interspecies cross populations (Zhao et al., 2019; Rehman et al., 2020).

Mapping With Other Population

Although the F_1 population is suitable for genetic map development, as shown in **Figure 4B**, several features, particularly qualitative features governed by recessive alleles, show no distribution in F_1 offspring, according to our research. Thus, selfing, sib mating and back-crossing to construct F_2 or BC1 populations are necessary. For example, Ningqi No. 5 (N5) is a *L. barbarum* (Chen C. et al., 2017) sterile male cultivar, and we crossed N5 with Zhongkeluchuan No. 1 to examine sterile male features. The F_1 offspring, on the other hand, are entirely reproductive males. As a result, we backcrossed with N5 and obtained 120 sterile males and 125 fertile plants. Thus, we obtained an F_2 population with male sterility segregation (sterile males: fertile plants = 1:1) that can be used for QTL mapping, and we discovered that male sterility is regulated by a single recessive gene. The segregations derived from diploid parents can be classified into eight segregations (ab × cd, ef × eg, hk × hk, lm × ll, nn × np, aa × bb, ab × cc and cc × ab), and only homozygous segregation pattern (aa × bb) is suitable for constructing the genetic map in BC1 and F2 populations (Fernando, 1982; Yu et al., 2013; Zhang et al., 2013). According to Zhao et al. (2019) and Rehman et al. (2020), the number of genotypes (aa × bb) in the parent (*L. barbarum* × *L. chinense*) is 17,513 (17.02%) and 24,329 (59.9%), respectively, indicating that the F_2 and BC1 populations can be used to generate genetic maps and map QTLs in *Lycium* species. Furthermore, using the newly published (Cao et al., 2021) *L. barbarum* genome by Cao et al. (2021) as a reference genome, we can map QTLs using a more cost-effective method, such as BSA.

Multiparental Population

When investigating natural populations, genome-wide association analysis offers numerous advantages, but it also

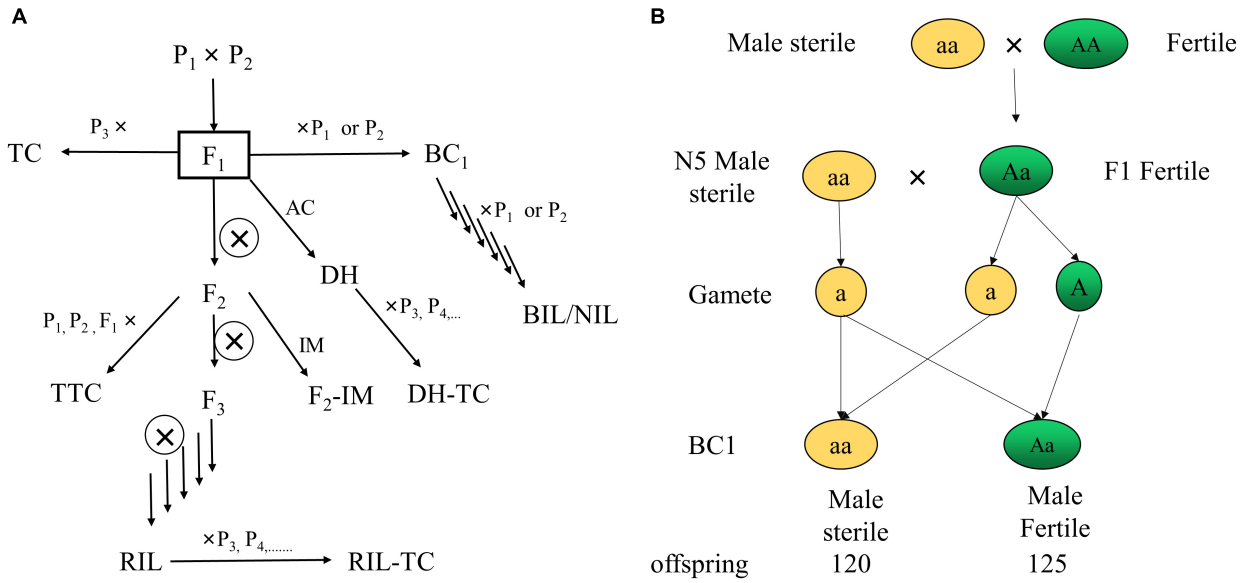

FIGURE 4 | Classification of mapping populations **(A)** (Xu, 2010) and strategy of population construction through sterile male QTL mapping **(B)**. P_n, parent; F1, filial generation 1; AC, mapping population can be classified into anther culture; BC, backcross population; BIL, backcross inbred line; DH, double haploid; IM, intermating; NIL, near-isogenic line; RIL, recombinant inbred line; TC, testcross; TTC, triple testcross; N5, a sterile male cultivar of *L. barbarum*.

has many drawbacks: (1) the population structure has an impact on the accuracy of GWAS analysis. Locus identification is prone to false-positives when the population structure is complicated. (2) Rare genes in populations are challenging to find using GWAS. Due to their low proportion and wide distribution in the population (linkage mapping), rare genes are easily filtered out of analyses (Lander and Schork, 1994). Linkage map analysis, on the other hand, has the following drawbacks: (1) due to the low frequency of marker recombination, the coverage of markers in broad chromosome regions is low (Yu et al., 2008; Hu et al., 2018); (2) a mapped population cannot be used for numerous traits because the population is only derived from the cross of two parents and the number of variant characteristics is limited compared with association mapping (Yu et al., 2008); and (3) it takes a long time to build the population because it is built by crossing (Madhusudhana, 2015).

However, in terms of prior knowledge, cross-validation, and statistical power, linkage analysis and association mapping are complementary to one another (Wilson et al., 2004). The genetic population of association mapping, on the other hand, is a natural population, whereas linkage mapping is invariably a biparental population. Yu et al. (2008) suggested the nested association mapping (NAM) technique to integrate the benefits of both linkage and association mapping strategies (Madhusudhana, 2015). The Buckler lab at Cornell University introduced nested association mapping (NAM) for the first time in 2008. Large populations with many small populations can be created by hybridizing multiple parents with a common parent (Buckler et al., 2009). The inbred line B73 was chosen as the common parent in the maize NAM population construction, and it was crossed with 25 other inbred lines with a wide genetic variety and then self-crossed to the F_5 generation using the single-grain transfer method. Finally, 25 RIL populations were acquired. There were 200 recombinant inbred lines (RILs) in each population (Madhusudhana, 2015). The use of the NAM population for QTL mapping has two obvious benefits: (1) because the NAM population has distinct parents, both association and linkage analyses can be performed at the same time, and mapping approaches are more diversified and versatile; (2) the population structure can be simplified, and the false-positives generated by the complicated population structure can be minimized because there is a common parent and a clear population structure (Madhusudhana, 2015). Minor changes are required to apply the NAM population to *Lycium*. Unlike herbaceous plants that have a brief lifespan, goji species are shrubs that live for decades (Cao and He, 2013) and can be bred by cuttage (Huang et al., 2015), which is convenient for germplasm resource conservation. As a result, there is no need to create a recombination inbred line. Furthermore, because goji species, like other shrub plants, are heterozygous (Zhao et al., 2010), they are excellent for mapping with F_1 populations. Conclusively, NAM populations can be created using F_1 populations crossed by a common parent with distinct cultivars/lines. After modest modifications, other multiparental populations, such as multiparental advanced generation inter-crosses (MAGIC) (Pascual et al., 2015), random-open-parent association mapping (ROAM) (Glowinski and Flint-Garcia, 2018), and complete-diallel plus unbalanced breeding-derived inter-crosses (CUBIC) (Liu et al., 2020a), can be used in *Lycium* breeding.

PERSPECTIVES OF *LYCIUM* SPECIES BREEDING

Mechanized Harvest

In China, the level of mechanization of Chinese medicinal materials is rather low; production is primarily through manpower, and the instruments are simple (Junfa et al., 2009), particularly for goji berries. Currently, the harvesting of goji is mainly reliant on manpower, resulting in low picking efficiency and high picking costs (Zheng et al., 2018). Furthermore, with the rapid migration of rural workers to urban areas, agricultural labor costs are rising. With a comprehensive understanding of farmers' labor (Junfa et al., 2009) costs, it is critical to build harvesting machinery systems (Zheng et al., 2018). A one-time picking technique, which requires crop growth periods and harvest times to be coordinated, can be utilized to cut labor costs and use mechanization (Zheng et al., 2018). However, due to the complicated growth features of goji fruit, mechanical harvesting is challenging due to the presence of ripe fruits, immature fruits, and blossoms on the same branch at the same time (Zheng et al., 2018). As a result, cultivating new varieties that are compact, tolerant to dense planting, and suitable for mechanized operation and have synchronized fruit ripening and compact plant architecture, has become the primary goal of agricultural genetic improvement (Guo et al., 2019). In coffee fruits, synchronized flowering and fruiting and a reduction in fruit removal force were shown to be ideal for mechanical harvesting (Tongumpai, 1993). Thus, we need to determine which characteristic makes fruit ripening more coordinated. The fruit ripening period of plants with a flower terminal is comparatively concentrated, according to tomato research, which is helpful to mechanical harvesting (Li et al., 2018; Guo et al., 2019). Flower production and limited growth at the apex of the branch can promote uniform fruit ripening, providing theoretical support and practical direction for plant type breeding suitable for mechanical harvesting operations (Guo et al., 2019). For many years, crop physiology, culture, and breeding have focused on improving flower terminal characteristics. Centroradialis (CEN) gene expression increases in the inflorescence, and the apex stops growing and blossoms, indicating apical capping (Bradley et al., 1996). The flower terminal of the tomato was found to be controlled by a gene, and gene editing might be used to domesticate it (Pnueli et al., 1998; Li et al., 2018).

The M1 strain developed in the early stages of our research group was used as the female parent, and Zhongkeluchuan1 was used as the male parent. A mutant plant with the flower terminal attribute was established in the established genetic population (population size, 425 plants). The branch tip stops growing when the new branches reach 40 cm, and the apical meristem develops into a cluster of flowers (**Figure 5**). The ripening rate of the fruits on the same branch is very consistent. The discovery of this germplasm will aid in the future identification of candidate genes for controlling flower terminal characteristics in *L. barbarum*. The directional breeding of congruent varieties of *L. barbarum*, as well as the functional verification of genes and the establishment of molecular markers, provides a solid theoretical and practical foundation.

Thus, to fulfill this aim, we could select approximately 300 offspring and their parents as genetic populations to construct a high-density genetic map with linkage mapping (SLAF-seq or ddRAD-seq) technology according to previous studies (**Figure 1A**, steps 1, 3, and 4) (Gong et al., 2019; Zhao et al., 2019; Rehman et al., 2020). For flower terminal trait detection, traits No. 8 (flower terminal) and No. 9 (flower terminal rate) in **Supplementary Table 7** can be reference methods (**Figure 1A**, step 2). Finally, as **Figure 1A** and Step 5 show, QTLs for flower terminals could be mapped in a genetic map with the combination of sequencing and phenotypic data. In addition, BSA-seq could also be used as an alternative, as shown in **Figure 1B**.

The detected QTLs could be developed as markers for determinate terminal trait selection in *Lycium* species, which will accelerate the selection of flower terminal cultivars. Moreover, the publication of the genome of *L. barbarum* (Cao et al., 2021) enables the mining of candidate genes for further study, such as candidate gene validation, rapid breeding of cultivars with determinate flower terminal using transgenic technology, and studies on the mechanism of terminals flower development.

Cultivars for Fruits

The goji products on the market right now are mostly dried fruits; however, the drying process causes losses of proteins, carotenoids, fatty acids, and other bioactive elements (Reeve et al., 2010; Hu et al., 2011). Fresh fruits, with their complete nutritional value and delicious flavor, can meet consumer demands for functional fruits. Therefore, fresh wolfberry, as a new functional fruit, has large market potential (Huang et al., 2013). The goji industry should be built on the basis of medicinal value and expanded on the basis of editable value. Thus, we should move beyond the research category of therapeutic impact and broaden the area of fresh food research to broaden the application and industrial scope (Qin and Dai, 2017). Fresh fruits, as opposed to dried fruits, must have a stunning appearance, a lengthy fruit, thick flesh, hard skin, a sweet and good taste, and other favorable traits. However, most modern cultivars have plump pulp, thin skin, and a high moisture content. They are easily damaged while being picked and during transportation, resulting in cracked peels and eventually moldy fruits (Li et al., 2010). Therefore, fruit color, size (especially length), pulp thickness, peel stiffness, and sensory taste should be prioritized in *Lycium* fruit breeding.

Obviously, there are many phenotypic requirements in fruit cultivars. To facilitate the breeding of fruit cultivars, we need a population with segregations of all traits required. However, it is not easy to construct a biparental population with so many distinguishing variations. Thus, the NAM population, constructed by crossing multiple parents and one common parent (**Figure 6**), with sharp segregations of multiple kinds of traits, could be a suitable option. NAM populations plus both association mapping and linkage mapping technologies could map QTLs for multiple fruit use related traits, with both high resolution and precision. Moreover, besides the contribution to QTL to facilitate breeding of fruit use cultivar (same to QTL for flower terminal trait), NAM population could cause more plant with more significant phenotypic variations, such as fruit color,

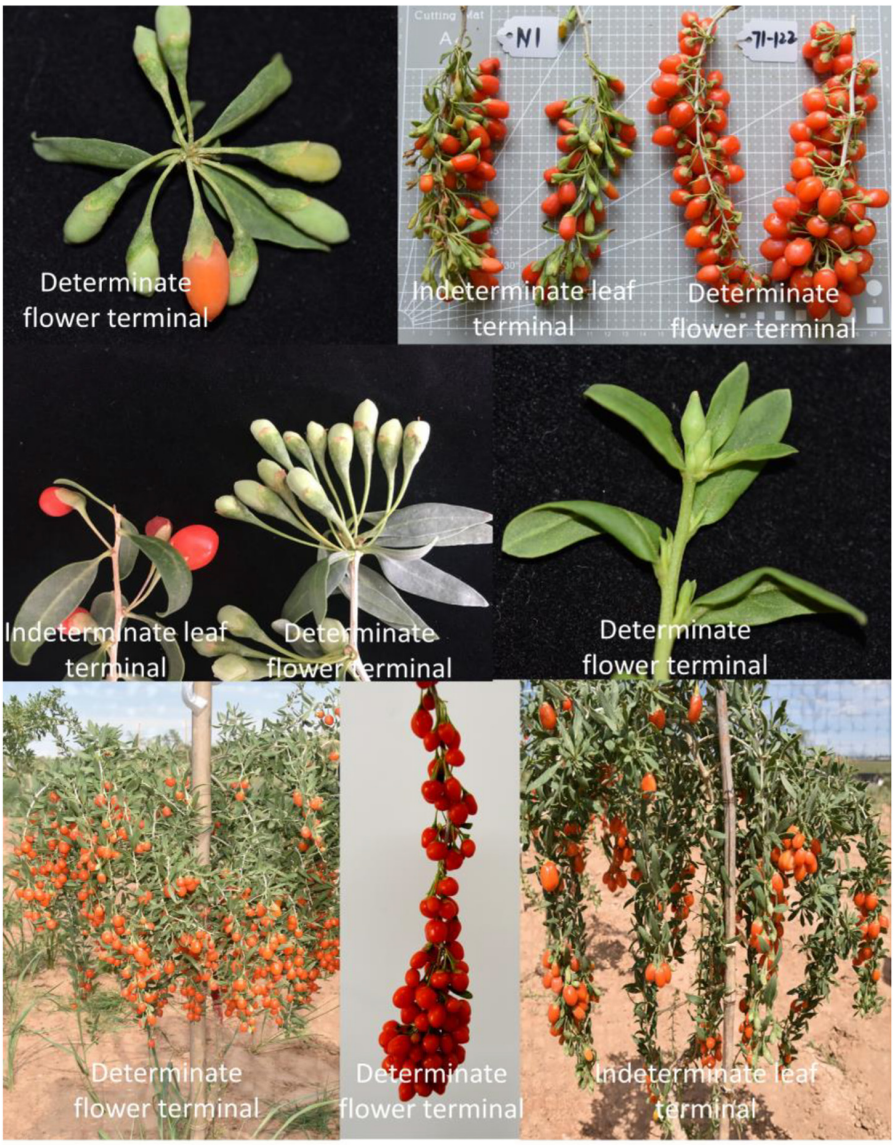

FIGURE 5 | Comparison of cultivars with indeterminate leaf terminal and determinate flower terminal.

fruit shape, leaf shape, plant shape, and et al. (**Figures 3G–J**), which facilitate the preservation of germplasm resources for ornamental horticulture and other potential uses.

Improving Light Use Efficiency

Global food crop production has doubled in the last two decades and continues to rise (Evans, 1996). According to Richard's research, the increases in grain yield in recent decades have two characteristics: (1) photosynthesis theory can be used to boost agricultural yields; (2) photosynthesis-related genetic theories have not been used to boost crop yields (Richards, 2000). Photosynthesis is genetically controlled and plays a crucial role in plant development, biomass buildup, and crop yield (Qu et al., 2017). Therefore, photosynthesis has not been completely utilized in modern agricultural production despite having a great development potential for increasing crop yields (Long et al., 2015). Improving the photosynthetic rate is thought to be one of the most essential ways to boost crop yield (Makino, 2011). There are a variety of techniques to boost the net photosynthetic rate, including increasing light gathering capacity, light energy conversion efficiency, CO_2 fixation ability, and CO_2 conversion efficiency in leaves (Ort et al., 2015). However, there have been few findings on genes that control the photosynthetic biosynthesis pathway (Osipova et al., 2016). The genes that affect the photosynthetic rate and water use efficiency are being studied to improve these qualities (Zheng et al., 2011).

There are sharp differences in photosynthesis between cultivars and species in Lycium, such as *L. cylindricum* (net photosynthetic rate, Pn, 9.40 μmol CO_2 s^{-1} m^{-2}), *L. yunnanense*

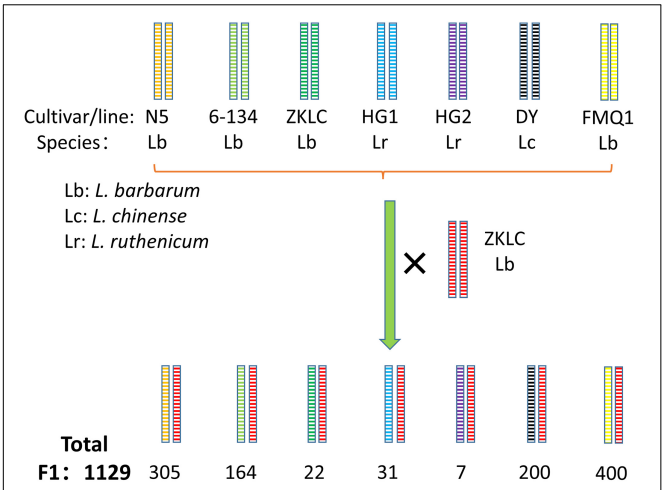

FIGURE 6 | Construction of NAM population in *Lycium*. cultivar/line of *L. barbarum*: N5, 6-134, ZKLC, FMQ1; cultivar/line of *L. chinense*: DY; cultivar/line of *L. ruthenicum*: HG1, HG2; × : cross.

(14.18 μmol CO_2 s^{-1} m^{-2}), *L. dasystemum* (19.46 μmol CO_2 s^{-1} m^{-2}), *L. ruthenicum* (7.29 μmol CO_2 s^{-1} m^{-2}), *L. barbarum* L. var. *auranticarpum* K. F. Ching (8.69 μmol CO_2 s^{-1} m^{-2}), *L. barbarum* "Ningqi-5" (9.20 μmol CO_2 s^{-1} m^{-2}) and *L. barbarum* L. "Zhongkelüchuan-1" (18.19 μmol CO_2 s^{-1} m^{-2}) (Cao and Wu, 2015; Gong et al., 2019), providing multiple resources for research on photosynthetic traits. Moreover, Gong et al. (2019) detected 29 QTLs for photosynthetic traits, including Pn, water use efficiency (WUE), limiting value of the stoma (Ls), transpiration rate (Trmmol), intracellular carbon dioxide (Ci), and stomatal conductance (Cond) (Gong et al., 2019), which will greatly shorten the breeding procedure. Existing photosynthetic QTLs could be developed as markers for photosynthetic trait selection in *Lycium* species, which will accelerate the selection of flower terminal cultivars. The other steps, including candidate gene mining, validation, rapid breeding of cultivars with highly efficient photosynthesis and study of the mechanism of photosynthesis, are the same as the procedure for research on flower terminal traits mentioned above.

Role of Marker-Assisted Selection in the Cultivation and Conservation of Germplasm Resources and Research on Multiple Traits

MAS with NGS consists of germplasm resource collection, population construction, trait detection, sequencing and genotyping, genetic statistical analysis, and QTL mapping (**Figure 1**). Thus, research on multiple traits with MAS could facilitate not only the breeding of elite cultivars but also cultivation and conservation of new germplasm resources.

Worldwide, there are 97 species with different variations and functions in *Lycium* (Yao et al., 2018a), and they provide a variety of germplasm resources for cultivation and breeding in goji plants. Here, we summarized the distribution of all species in *Lycium*, species richness in all countries worldwide, location of existing germplasm repositorys, and functions of different species according to existing published studies (**Figure 2** and **Supplementary Table 1**), simplifying and facilitating the collection and conservation of germplasm resources.

Most of the population in MAS was constructed by crossing. Different ploidies of chromosomes in parents may cause highly sterile progeny and a low rate of seed satiation, resulting in difficult breeding of seedless plants (Li J. et al., 2001). Validation of karyotype analysis in parents is necessary (Xu, 2010). To select intra- and interspecies parents for crossing and evaluate their fertilities, **Table 2** could be referenced. **Table 2** summarizes the karyotype analysis of 50 species, eight varieties, and one hybrid plant, among which 42 species had the diploid type. Crosses can occasionally create plants with new variations, such as fruit color (**Figure 3**) and flower terminals (**Figure 5**), which will facilitate the cultivation of new germplasms.

Compared with other crops, studies on the resource base of *Lycium* species, including systematic and in-depth research, are still lagging behind (An et al., 2009). Conversely, studies on other Solanaceae species, such as tomato and potato, are quite in depth, regardless of gene research, the variety of features, or other themes. Thus, the conclusion in our research on 1,144 traits (274 agronomic and 870 metabolic) and their standardized detection based on tomato, *Lycium*, and other Solanaceae species in the prospective trait supplying section (**Supplementary Tables 2–10**) will facilitate the morphological evaluation, conservation and breeding of *Lycium* germplasm.

QTLs derived from sequencing data and phenotypic data have two kinds of usages. One is for marker development, which could accelerate the selection for elite cultivars with specific trait selection. The present emphasis on practical breeding is incorporating biotechnological methods for the development of a variety of robust markers to execute MAS, in order to speed up the breeding of elite cultivars and delivery of them to the producer (Rubiales et al., 2021), such as the breeding of elite cultivars in grain, tomato, pepper, and eggplant (Foolad and Panthee, 2012; Hanson et al., 2016; Khapte et al., 2018; Moodley et al., 2019). The other is for genetic and mechanistic research of the traits, for example, rice grain. A gene (Gn1a) encoding cytokinin oxidase/dehydrogenase was detected after mapping QTLs for grain yield. The enzyme was found to restrict the synthesis of the phytohormone cytokinin. After a series of validations, the mechanism was successfully demonstrated as gene-cytokinin oxidase/dehydrogenase-phytohormone cytokinin-grain yield (Ashikari et al., 2005).

Finally, after reviewing many studies, we came to a conclusion on the future breeding research direction of *Lycium* species. This review will help to guide future *Lycium* breeding studies.

AUTHOR CONTRIBUTIONS

HG: conceptualization, methodology, software, writing—review and editing, and visualization. FR, SZ, ZL, and TY: validation

and formal analysis. BA: investigation. YM: resources. DW: data curation. JH: writing—original draft preparation. YW: supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Key Research and Development Program Project (Grant No. 2019YFC1711100), the Sub-Project of Chinese Academy of Sciences Pilot Project (Grant No. XDA24030502), the Guangdong Provincial Special Fund for Modern Agriculture Industry Technology Innovation Teams, China (Grant No. 2020KJ148), and the National Natural Science Foundation of China (Grant No. 31770334).

ACKNOWLEDGMENTS

We thank Qin Ken from NingXia Academy of Agricultural and Forestry Sciences for providing Pictures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.802936/full#supplementary-material

REFERENCES

Agarwal, M., Shrivastava, N., and Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* 27, 617–631. doi: 10.1007/s00299-008-0507-z

Alpert, K. B., Grandillo, S., and Tanksley, S. D. (1995). Fw 2.2: a major qtl controlling fruit weight is common to both red-and green-fruited tomato species. *Theor. Appl. Genet.* 91, 994–1000. doi: 10.1007/BF00223911

An, W., Zhang, H., He, J., Li, X., and Fan, Y. (2009). Research progress in goji breeding. *N. Hortic.* 5, 125–128.

An, W., Zhao, J., Shi, Z., and Jiao, E. (2007). Evaluation criteria of some fruit quantitative characteristics of wolfberry (*Lycium*) genetic r esources. *J. Fruit Sci.* 24, 172–175.

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., et al. (2005). Cytokinin oxidase regulates rice grain production. *Science* 309, 741–745. doi: 10.1126/science.1113373

Azadi, N., Nazeri, V., Namaki Shoushtari, A. A. H., and Kazem Pour Ousalou, S. (2007). *Lycium dasystemum* pojark. (Solanaceae), a new record from iran. *Iranian J. Bot.* 13, 109–111.

Balasubramani, S. P., Goraya, G. S., and Venkatasubramanian, P. (2011). Development of its sequence-based markers to distinguish berberis aristata dc. From b. *Lycium royle* and b. Asiatica roxb. *3 Biotech* 1, 11–19. doi: 10.1007/s13205-010-0001-5

Barchi, L., Lefebvre, V., Sage-Palloix, A., Lanteri, S., and Palloix, A. (2009). Qtl analysis of plant development and fruit traits in pepper and performance of selective phenotyping. *Theor. Appl. Genet.* 118, 1157–1171. doi: 10.1007/s00122-009-0970-0

Barchi, L., Portis, E., Toppino, L., and Rotino, G. L. (2019). *Molecular Mapping, Qtl Identification, and Gwa Analysis*. Cham: Springer International Publishing, 41–54.

Bernardello, L., Rodriguez, I., Stiefkens, L., and Galetto, L. (1995). The hybrid nature of *Lycium ciliatum* × *cestroides* (Solanaceae): experimental, anatomical, and cytological evidence. *Can. J. Bot.* 73, 1995–2005. doi: 10.1139/b95-214

Blenda, A., Fang, D. D., Rami, J., Garsmeur, O., Luo, F., and Lacape, J. M. (2012). A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. *PLoS One* 7:e0045739. doi: 10.1371/journal.pone.0045739

Bombarely, A., Menda, N., Tecle, I. Y., Buels, R. M., Strickler, S., Fischer-York, T., et al. (2010). The sol genomics network (solgenomics.net): growing tomatoes using perl. *Nucleic Acids Res.* 39, D1149–D1155. doi: 10.1093/nar/gkq866

Bradley, D., Carpenter, R., Copsey, L., Vincent, C., Rothstein, S., and Coen, E. (1996). Control of inflorescence architecture in antirrhinum. *Nature* 379, 791–797. doi: 10.1038/379791a0

British Pharmacopoeia Commission (2017). *British Pharmacopoeia*. London: British Pharmacopoeia Commission, 91–92.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., et al. (2009). The genetic architecture of maize flowering time. *Science* 325, 714–718. doi: 10.1126/science.1174276

Cai, C., Cheng, F., Wu, J., Zhong, Y., and Liu, G. (2015). The first high-density genetic map construction in tree peony (*Paeonia sect*. Moutan) using genotyping by specific-locus amplified fragment sequencing. *PLos One* 10:e0128584. doi: 10.1371/journal.pone.0128584

Cao, Y., and He, J. (2013). *Cultivation of Lycium*. Shenyang: Yangguang Press.

Cao, Y., Li, Y., Fan, Y., Li, Z., Yoshida, K., Wang, J. Y., et al. (2021). Wolfberry genomes and the evolution of *Lycium* (solanaceae). *Commun. Biol.* 4:671. doi: 10.1038/s42003-021-02152-8

Cao, Y., and Wu, P. (2015). *Wolfberry Germplasm Resources in China*. Beijing: China Forestry Publishing House.

Chandler, J. W. (2012). Floral meristem initiation and emergence in plants. *Cell. Mol. Life Sci.* 69, 3807–3818. doi: 10.1007/s00018-012-0999-0

Chaudhury, A. M., Koltunow, A., Payne, T., Luo, M., Tucker, M. R., Dennis, E. S., et al. (2001). Control of early seed development. *Annu. Rev. Cell Dev. Biol.* 17, 677–699. doi: 10.1146/annurev.cellbio.17.1.677

Chen, C., Xu, M., Wang, C., Qiao, G., Wang, W., Tan, Z., et al. (2017). Characterization of the *Lycium barbarum* fruit transcriptome and development of est-ssr markers. *PLoS One* 12:e187738. doi: 10.1371/journal.pone.0187738

Chen, D., Zhou, X., Li, L., Liu, C., and Chen, Y. (2015). Research progress on rice (*Oryza sativa*. L) leaf thickness. *J. Agric.* 5, 22–25.

Chen, H., Song, Y., Li, L., Khan, M. A., Li, X., Korban, S. S., et al. (2015). Construction of a high-density simple sequence repeat consensus genetic map for pear (*Pyrus* spp.). *Plant Mol. Biol. Rep.* 33, 316–325. doi: 10.1007/s11105-014-0745-x

Chen, H., and Zhong, Y. (2014). Microsatellite markers for *Lycium ruthenicum* (Solananeae). *Mol. Biol. Rep.* 41, 5545–5548. doi: 10.1007/s11033-014-3442-9

Chen, J., Chao, C. T., and Wei, X. (2018). *Gojiberry Breeding: Current Status and Future Prospects*. London: IntechOpen.

Chen, J., Liu, X., Zhu, L., and Wang, Y. (2013). Nuclear genome size estimation and karyotype analysis of *Lycium* species (Solanaceae). *Sci. Hortic.* 151, 46–50. doi: 10.1016/j.scienta.2012.12.004

Chen, J., Wang, N., Fang, L., Liang, Z., Li, S. H., and Wu, B. H. (2015). Construction of a high-density genetic map and QTLs mapping for sugars and acids in grape berries. *BMC Plant Biol.* 15:28. doi: 10.1186/s12870-015-0428-2

Chen, J., Zhang, D., Zhang, C., Xu, M., and Yin, W. (2017). Physiological characterization, transcriptomic profiling, and microsatellite marker mining of *Lycium ruthenicum*. *J. Zhejiang Univ. Sci. B* 18, 1002–1021. doi: 10.1631/jzus.B1700135

Chen, K. Y., and Tanksley, S. D. (2004). High-resolution mapping and functional analysis of se2.1: a major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus *Lycopersicon*. *Genetics* 168, 1563–1573. doi: 10.1534/genetics.103.022558

Chen, W., Gong, L., Guo, Z., Wang, W., Zhang, H., Liu, X., et al. (2013). A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant* 6, 1769–1780. doi: 10.1093/mp/sst080

Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J., et al. (2016). Comparative and parallel genome-wide association studies for metabolic and

agronomic traits in cereals. *Nat. Commun.* 7:12767. doi: 10.1038/ncomms 12767

Chen, Z., Wang, B., Dong, X., Liu, H., Ren, L., Chen, J., et al. (2014). An ultra-high density bin-map for rapid qtl mapping for tassel and ear architecture in a large f2 maize population. *BMC Genomics* 15:433. doi: 10.1186/1471-2164-15-433

Cheng, K. T., Chang, H. C., Huang, H., and Lin, C. T. (2000). Rapd analysis of *Lycium barbarum* medicine in Taiwan market. *Bot. Bull. Acad. Sin.* 41, 11–14. doi: 10.7016/BBAS.200001.0011

Cheng, Y., and Zhao, Y. (2007). A role for auxin in flower development. *J. Integr. Plant Biol.* 49, 99–104. doi: 10.1111/j.1744-7909.2006.00412.x

Chinese Pharmacopoeia Commission (2015). *Chinese Pharmacopoeia 1*. Beijing: China Medical Science Press, 124–249.

Chung, J., Lee, G., Bang, K., Park, C., and Park, Y. (2009). Cultivar discrimination of Korean and Chinese boxthorn (*Lycium chinense* mill. and *Lycium barbarum* L.) using SSR markers. *Korean J. Med. Crop Sci.* 17, 445–451.

Coneva, V., Frank, M. H., Balaguer, M. A. D. L., Li, M., Sozzani, R., Chitwood, D. H., et al. (2017). Genetic architecture and molecular networks underlying leaf thickness in desert-adapted tomato *Solanum pennellii*. *Plant Physiol.* 175, 376–391. doi: 10.1104/pp.17.00790

Czyczylo-Mysza, I., Tyrka, M., Marcinska, I., Skrzypek, E., Karbarz, M., Dziurka, M., et al. (2013). Quantitative trait loci for leaf chlorophyll fluorescence parameters, chlorophyll and carotenoid contents in relation to biomass and yield in bread wheat and their chromosome deletion bin assignments. *Mol. Breed.* 32, 189–210. doi: 10.1007/s11032-013-9862-8

Dahech, I., Farah, W., Trigui, M., Ben Hssouna, A., Belghith, H., Belghith, K. S., et al. (2013). Antioxidant and antimicrobial activities of *Lycium shawii* fruits extract. *Int. J. Biol. Macromol.* 60, 328–333. doi: 10.1016/j.ijbiomac.2013.05.020

Dielen, V., Quinet, M., Chao, J., Batoko, H., Havelange, A., and Kinet, J. M. (2004). Uniflora, a pivotal gene that regulates floral transition and meristem identity in tomato (*Lycopersicon esculentum*). *New Phytol.* 161, 393–400. doi: 10.1046/j.1469-8137.2003.00937.x

Dong, J., Yang, J., and Wang, Y. (2008). Species resources of *Lycium* in china and research progress domestic and overseas. *China J. Chin. Mater. Med.* 5, 2020–2027.

Dong, W., Wu, D., Li, G., Wu, D., and Wang, Z. (2018). Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Sci. Rep. U.K.* 8:2908. doi: 10.1038/s41598-018-21293-1

Duan, Y., Yao, X., Wang, C., Chen, X., Zhang, J., and Li, X. (2015). Protective effects of *Lycium ruthenicum* murr on x-radiation injured mice. *Nat. Product Res. Dev.* 27, 148–152.

Editorial Committee of Chinese Flora, Chinese Academy of Sciences (1978). *Flora of China*. Beijing: Science press.

Ehrenreich, I. M., Gerke, J. P., and Kruglyak, L. (2009). Genetic dissection of complex traits in yeast: insights from studies of gene expression and other phenotypes in the byxrm cross. *Cold Spring Harb. Symp. Quant. Biol.* 74, 145–153. doi: 10.1101/sqb.2009.74.013

Essid, A., Aljane, F., Ferchichi, A., and Ignacio Hormaza, J. (2015). Analysis of genetic diversity of tunisian caprifig (*Ficus carica* L.) accessions using simple sequence repeat (SSR) markers. *Hereditas* 152:1. doi: 10.1186/s41065-015-0002-9

European Directorate for the Quality of Medicines (2016). *European Pharmacopoeia*. Strasbourg: European Directorate for the Quality of Medicines, 1263–1264.

Evans, L. T. (1996). *Crop Evolution, Adaptation and Yield*. New York: Cambridge university press.

Fernando, C. (1982). New combinations and new variety of *Lycium* (Solanaceae) from north America. *Bot. Sci.* 43, 5–8. doi: 10.17129/botsci.1263

Foolad, M. R., and Panthee, D. R. (2012). Marker-assisted selection in tomato breeding. *Crit. Rev. Plant Sci.* 31, 93–123. doi: 10.1080/07352689.2011.616057

Frary, A., Doganlar, S., Daunay, M. C., and Tanksley, S. D. (2003). QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of *Solanaceous* species. *Theor. Appl. Genet.* 107, 359–370. doi: 10.1007/s00122-003-1257-5

Fukuda, T., Yokoyama, J., and Ohashi, H. (2001). Phylogeny and biogeography of the genus *Lycium* (Solanaceae): inferences from chloroplast DNA sequences. *Mol. Phylogenet. Evol.* 19, 246–258. doi: 10.1006/mpev.2001.0921

Ge, Y., Cao, Y., Xu, X., Zhao, J., and Zhang, B. (2008). Study on quality change of fresh fruit of *Lycium barbarum* after harvest. *N. Hortic.* 7, 227–229.

Glowinski, A., and Flint-Garcia, S. (2018). "Germplasm resources for mapping quantitative traits in maize," in *The Maize Genome*, eds J. Bennetzen, S. Flint-Garcia, C. Hirsch, and R. Tuberosa (Cham: Springer International Publishing), 143–159. doi: 10.1007/978-3-319-97427-9_10

Gong, H., Rehman, F., Yang, T., Li, Z., Zeng, S., Pan, L., et al. (2019). Construction of the first high-density genetic map and QTL mapping for photosynthetic traits in *Lycium barbarum* l. *Mol. Breed.* 39:106. doi: 10.1007/s11032-019-1000-9

Grube, R. C., Radwanski, E. R., and Jahn, M. (2000). Comparative genetics of disease resistance within the Solanaceae. *Genetics* 155, 873–887. doi: 10.1093/genetics/155.2.873

Guo, F., Yu, H., Tang, Z., Jiang, X., Wang, L., Wang, X., et al. (2015). Construction of a SNP-based high-density genetic map for pummelo using rad sequencing. *Tree Genet. Genomes* 11:18. doi: 10.1007/s11295-014-0831-0

Guo, P., Ren, G., Li, Z., Zeng, S., and De, T. (2019). Progress of determinate growth genes and gene editing breeding. *Chinese J. Trop. Crops* 40, 2016–2021. doi: 10.3969/j.issn.1000-2561.2019.10.014

Han, Q. Q., Song, Y. Z., Zhang, J. Y., and Liu, L. F. (2014). Studies on the role of the slnac3 gene in regulating seed development in tomato (*Solanum lycopersicum*). *J. Hortic. Sci. Biotech.* 89, 423–429. doi: 10.1080/14620316.2014.11513101

Hanson, P., Lu, S., Wang, J., Chen, W., Kenyon, L., Tan, C. W., et al. (2016). Conventional and molecular marker-assisted selection and pyramiding of genes for multiple disease resistance in tomato. *Sci. Hortic. Amsterdam* 201, 346–354. doi: 10.1016/j.scienta.2016.02.020

Henareh, M., Dursun, A., and Mandoulakani, B. A. (2015). Genetic diversity in tomato landraces collected from turkey and Iran revealed by morphological characters. *Acta Sci. Pol. Hortorum Cultus* 14, 87–96.

Hu, J., Guo, C., Wang, B., Ye, J., Liu, M., Wu, Z., et al. (2018). Genetic properties of a nested association mapping population constructed with semi-winter and spring oilseed rapes. *Front. Plant Sci.* 9:1740. doi: 10.3389/fpls.2018.01740

Hu, W. (2015). *Development and Utility of Est-Ssr and Est-Snp Markers for Four Medicinal Plants*. Huairou: University of Chinese Academy of Sciences.

Hu, Y. F., Hu, H. Y., Jiang, C. J., Zhu, Y. H., and Wang, J. M. (2011). The influence of different product process on the effect ingredient content of Ningxia *Lycium barbarum*. *Adv. Mat. Res.* 20, 2083–2086.

Huang, T., Liu, J., Qin, K., Dai, G., and Zhang, B. (2013). Study on quality requirements for fresh wolfberry. *Ningxia J. Agric. For. Sci. Technol.* 54, 151–154.

Huang, W., Liao, S., Lv, H., Khaldun, A. B. M., and Wang, Y. (2015). Characterization of the growth and fruit quality of tomato grafted on a woody medicinal plant, *Lycium chinense*. *Sci. Hortic. Amsterdam* 197, 447–453. doi: 10.1016/j.scienta.2015.10.005

Ibrahim, S. D., Adawy, S. S., Atia, M. A. M., Alsamman, A. M., and Mokhtar, M. M. (2016). Genetic diversity, variety identification and gene detection in some Egyptian grape varieties by SSR and SCOT markers. *Plant Omics* 9, 311–318. doi: 10.21475/poj.09.05.16.pne125

Iordăchescu, M., Udriște, A. A., and Bădulescu, L. (2020). Use of molecular markers in improving resistance to biotic stress in Solanaceae-a review. *Sci. Pap. Ser. B Hortic.* 64, 405–412.

Jack, T. (2004). Molecular and genetic mechanisms of floral control. *Plant Cell* 16, S1–S17. doi: 10.1105/tpc.017038

Japanese Pharmacopoeia Editorial Committee (2016). *Japanese Pharmacopoeia (17th)*. Tokyo: Ministry of Health, Labour and Welfare, 1909–1910.

Jofuku, K. D., Omidyar, P. K., Gee, Z., and Okamuro, J. K. (2005). Control of seed mass and seed yield by the floral homeotic gene apetala2. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3117–3122. doi: 10.1073/pnas.0409893102

Junfa, W., Xu, M., Liuxuan, M., and Chuanhua, Y. (2009). The current states and technology study of harvesting equipments of rhizome traditional Chinese medicinal materials. *J. Agric. Mech. Res.* 12, 242–246.

Jung, W., Chung, I., Kim, S., Chi, H., Yu, C. Y., Ghimire, B. K., et al. (2021). Direct shoot organogenesis from *Lycium chinense* miller leaf explants and assessment of genetic stability using ISSR markers. *Agronomy* 11:503. doi: 10.3390/agronomy11030503

Kevany, B. M., Taylor, M. G., and Klee, H. J. (2008). Fruit-specific suppression of the ethylene receptor leetr4 results in early-ripening tomato fruit. *Plant Biotechnol. J.* 6, 295–300. doi: 10.1111/j.1467-7652.2007.00319.x

Khan, S. A., Chibon, P. Y., de Vos, R. C., Schipper, B. A., Walraven, E., Beekwilder, J., et al. (2012). Genetic analysis of metabolites in apple fruits indicates an

mQTL hotspot for phenolic compounds on linkage group 16. *J. Exp. Bot.* 63, 2895–2908. doi: 10.1093/jxb/err464

Khapte, P., Singh, T. H., and Lakshmana, D. C. (2018). Screening of elite eggplant (*Solanum melongena*) genotypes for bacterial wilt (*Ralstonia solanacearum*) in field conditions and their genetic association by using ssr markers. *Indian J. Agr. Sci.* 88, 1502–1511.

Korea Food and Drug Administration (2014). *Korean Pharmacopoeia (11th)*. Seoul: Shinil Publishing Company, 1862–1863.

Krizek, B. A., and Fletcher, J. C. (2005). Molecular mechanisms of flower development: an armchair guide. *Nat. Rev. Genet.* 6, 688–698. doi: 10.1038/nrg1675

Kumar, S., Sehgal, S. K., Kumar, U., Prasad, P. V. V., Joshi, A. K., and Gill, B. (2012). Genomic characterization of drought tolerance-related traits in spring wheat. *Euphytica* 186, 265–276. doi: 10.1007/s10681-012-0675-3

Kwon, S., Lee, G., Lee, S., Park, Y., Gwag, J., Kim, T. S., et al. (2009). Isolation and characterization of 21 microsatellite loci in *Lycium chinense* and cross-amplification in *Lycium barbarum*. *Conserv. Genet.* 10, 1557–1560. doi: 10.1007/s10592-008-9792-x

Lander, E. S., and Schork, N. (1994). genetic dissection of complex traits. *Science* 5181, 2037–2048.

Laura, S., Peñas, M. L. L., Bernardello, G., Levin, R. A., and Miller, J. S. (2010). Karyotypes and fluorescent chromosome banding patterns in southern African *Lycium* (solanaceae). *Caryologia* 63, 50–61. doi: 10.1080/00087114.2010.10589708

Lee, S. B., Kim, J. E., Kim, H. T., Lee, G. M., Kim, B. S., and Lee, J. M. (2020). Genetic mapping of the c1 locus by GBS-based BSA-seq revealed pseudo-response regulator 2 as a candidate gene controlling pepper fruit color. *Theor. Appl. Genet.* 133, 1897–1910. doi: 10.1007/s00122-020-03565-5

Levin, R. A., Whelan, A., and Miller, J. S. (2009). The utility of nuclear conserved ortholog set II (COSII) genomic regions for species-level phylogenetic inference in *Lycium* (Solanaceae). *Mol. Phylogenet. Evol.* 53, 881–890. doi: 10.1016/j.ympev.2009.08.016

Li, H., Huang, M., Luo, Q., Hong, X., Ramakrishna, S., and So, K. F. (2019a). *Lycium barbarum* (wolfberry) increases retinal ganglion cell survival and affects both microglia/macrophage polarization and autophagy after rat partial optic nerve transection. *Cell Transplant.* 28, 607–618. doi: 10.1177/0963689719835181

Li, H., Wu, H., Qi, Q., Li, H., Li, Z., Chen, S., et al. (2019b). Gibberellins play a role in regulating tomato fruit ripening. *Plant Cell Physiol.* 60, 1619–1629. doi: 10.1093/pcp/pcz069

Li, J., Wang, J., Wang, L., and Huang, Z. (2001). New variety breeding of seedless Ningxia wolfberry. *Acta Bot. Boreali Occidentalia Sin.* 21, 446–450.

Li, R., Sun, S., Wang, H., Wang, K., Yu, H., Zhou, Z., et al. (2020). Fis1 encodes a ga2-oxidase that regulates fruit firmness in tomato. *Nat. Commun.* 11:5844. doi: 10.1038/s41467-020-19705-w

Li, T., Yang, X., Yu, Y., Si, X., and Zhai, X. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163. doi: 10.1038/nbt.4273

Li, X., He, J., and Cao, Y. (2010). The sales problems and solutions of fresh wolfberry fruit. *N. Hortic.* 6, 175–176.

Li, Y., Zheng, L., Corke, F., Smith, C., and Bevan, M. W. (2008). Control of final seed and organ size by the da1 gene family in *Arabidopsis thaliana*. *Gene Dev.* 22, 1331–1336. doi: 10.1101/gad.463608

Li, Y. C., Wu, P. E., and Zhou, J. (2001). *Standards Collection of Materia Medica*. Chengdu: Science and Technology Press.

Lifschitz, E., and Eshed, Y. (2006). Universal florigenic signals triggered by ft homologues regulate growth and flowering cycles in perennial day-neutral tomato. *J. Exp. Bot.* 57, 3405–3414. doi: 10.1093/jxb/erl106

Liu, C., Zhou, X., Chen, D., Li, L., Li, J., and Chen, Y. D. (2014). Natural variation of leaf thickness and its association to yield traits in Indica rice. *J. Integr. Agr.* 13, 316–325. doi: 10.1016/S2095-3119(13)60498-0

Liu, D., Sun, D., Liang, J., Dou, J., Yang, S., Zhu, H., et al. (2021). Characterization and bulk segregant analysis of 'moon and star' appearance in watermelon. *Sci. Hortic. Amsterdam* 285:110140. doi: 10.1016/j.scienta.2021.110140

Liu, G. F., Yang, J., Xu, H. M., Hayat, Y., and Zhu, J. (2008). Genetic analysis of grain yield conditioned on its component traits in rice (*Oryza sativa* L.). *Aust. J. Agr. Res.* 59, 189–195. doi: 10.1071/AR07163

Liu, H., Wang, X., Xiao, Y., Luo, J., Qiao, F., Yang, W., et al. (2020a). Cubic: an atlas of genetic architecture promises directed maize improvement. *Genome Biol.* 21:20. doi: 10.1186/s13059-020-1930-x

Liu, J., Liang, Y., Qin, L., and Wang, L. (2014). Research progress on the molecular biology of floral development in tomato (*Solanum lycopersicum* L.). *J. Agric. Biotechnol.* 22, 351–361.

Liu, X., Cheng, J., Mei, Z., Wei, C., Khan, M. A., Peng, J., et al. (2020b). Scar marker for identification and discrimination of specific medicinal *Lycium* chinense miller from *Lycium* species from RAMP-PCR RAPD fragments. *3 Biotech* 10:334. doi: 10.1007/s13205-020-02325-y

Liu, X., Du, J., Khan, M. A., Cheng, J., Wei, C., Mei, Z., et al. (2020c). Analysis of genetic diversity and similarities between different *Lycium* varieties based on ISSR analysis and RAMP-PCR markers. *World Acad. Sci. J.* 2, 83–90. doi: 10.3892/wasj.2020.39

Liu, Y., Cao, L., Du, J., Jia, R., and Wang, J. (2015). Protective effects of *Lycium barbarum* polysaccharides against carbon tetrachloride-induced hepatotoxicity in precision-cut liver slices in vitro and in vivo in common carp (*Cyprinus carpio* L.). *Comp. Biochem. Phys. C* 169, 65–72. doi: 10.1016/j.cbpc.2014.12.005

Liu, Z., Shu, Q., Wang, L., Yu, M., Hu, Y., Zhang, H., et al. (2012). Genetic diversity of the endangered and medically important *Lycium ruthenicum* Murr. Revealed by sequence-related amplified polymorphism (SRAP) markers. *Biochem. Syst. Ecol.* 45, 86–97. doi: 10.1016/j.bse.2012.07.017

Long, S. P., Marshall-Colon, A., and Zhu, X. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* 161, 56–66. doi: 10.1016/j.cell.2015.03.019

Long, S. P., Zhu, X. G., Naidu, S. L., and Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* 29, 315–330. doi: 10.1111/j.1365-3040.2005.01493.x

Lu, S., and Zhao, P. (2010). Chemical characterization of *Lycium barbarum* polysaccharides and their reducing myocardial injury in ischemia/reperfusion of rat heart. *Int. J. Biol. Macromol.* 47, 681–684. doi: 10.1016/j.ijbiomac.2010.08.016

Lu, Y., Luo, S., Li, N., Li, Q., Du, W., Zhang, W., et al. (2021). Candidate gene, smcpr1, encoding cpr1 related to plant height of the eggplant dwarf mutant dwf. *Horticulturae* 7:196. doi: 10.3390/horticulturae7070196

Luckwill, L. C. (1948). The hormone content of the seed in relation to endosperm development and fruit drop in the apple. *J. Hortic. Sci.* 24, 32–44. doi: 10.1080/00221589.1948.11513679

Ma, J., Huang, L., Ma, C., Jin, J., Li, C., Wang, R. K., et al. (2015). Large-scale snp discovery and genotyping for constructing a high-density genetic map of tea plant using specific-locus amplified fragment sequencing (SLAF-seq). *PLoS One* 10:e0128798. doi: 10.1371/journal.pone.0128798

Ma, J., Zhang, D., He, J., Kang, C., and Yi, W. (2017). Mechanical design of automatic conveying and grading device for wolfberry. *J. Agric. Mech. Res.* 39, 108–116. doi: 10.3969/j.issn.1003-188X.2017.02.023

Mackay, I., and Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* 12, 57–63. doi: 10.1016/j.tplants.2006.12.001

Madhusudhana, R. (2015). "Linkage mapping," in *Sorghum Molecular Breeding*, eds R. Madhusudhana, P. Rajendrakumar, and J. V. Patil (New Delhi: Springer India), 47–70.

Magwene, P. M., Willis, J. H., and Kelly, J. K. (2011). The statistics of bulk segregant analysis using next generation sequencing. *PLoS Comput. Biol.* 7:e1002255. doi: 10.1371/journal.pcbi.1002255

Makino, A. (2011). Photosynthesis, grain yield, and nitrogen utilization in rice and wheat. *Plant Physiol.* 155, 125–129. doi: 10.1104/pp.110.165076

Mansfeld, B. N., and Grumet, R. (2018). Qtlseqr: an r package for bulk segregant analysis with next-generation sequencing. *Plant Genome* 11:180006. doi: 10.3835/plantgenome2018.01.0006

McCulloch, G. A., Mauda, E. V., Chari, L. D., Martin, G. D., Gurdasani, K., Morin, L., et al. (2020). Genetic diversity and morphological variation in African boxthorn (*Lycium ferocissimum*) – characterising the target weed for biological control. *Biol. Control* 143:104206. doi: 10.1016/j.biocontrol.2020.104206

Meuwissen, T. H., and Goddard, M. E. (2004). Mapping multiple qtl using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol.* 36, 261–279. doi: 10.1186/1297-9686-36-3-261

Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method

to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9828–9832. doi: 10.1073/pnas.88.21.9828

Miller, S. S., and Scorza, R. (2010). Response of two novel peach tree growth habits to in-row tree spacing, training system, and pruning: effect on growth and pruning. *J. Am. Pomol. Soc.* 64, 199–217.

Minne, L., Spies, J. J., Venter, H. J. T., and Venter, A. M. O. F. (1994). Breeding systems in some representatives of the genus *Lycium* (Solanaceae). *Bothalia* 24, 107–110. doi: 10.4102/abc.v24i1.759

Mocan, A., Vlase, L., Vodnar, D. C., Bischin, C., Hanganu, D., Gheldiu, A. M., et al. (2014). Polyphenolic content, antioxidant and antimicrobial activities of *Lycium barbarum* l. And *Lycium chinense* Mill. Leaves. *Molecules* 19, 10056–10073. doi: 10.3390/molecules190710056

Molinero-Rosales, N., Latorre, A., Jamilena, M., and Lozano, R. (2004). Single flower truss regulates the transition and maintenance of flowering in tomato. *Planta* 218, 427–434. doi: 10.1007/s00425-003-1109-1

Moodley, V., Naidoo, R., Gubba, A., and Mafongoya, P. L. (2019). Development of potato virus y (pvy) resistant pepper (*Capsicum annuum* L.) lines using marker-assisted selection (MAS). *Physiol. Mol. Plant P* 105, 96–101. doi: 10.1016/j.pmpp.2018.12.002

Nan, X., Shao, Q., Wang, J., Chang, H., and Wang, H. (2017). *Prospect of Special Wolfberry Varieties Breeding*. Yinchuan: Ningxia Forestry News, 48–52.

NingXia Academy of Agricultural and Forestry Sciences (2015). The only *Lycium* germplasm resources garden in the word were built in Ningxia. *Ningxia J. Agric. For. Sci. Technol.* 56:58.

Olatunji, O. J., Chen, H., and Zhou, Y. (2018). *Lycium chinense* leaves extract ameliorates diabetic nephropathy by suppressing hyperglycemia mediated renal oxidative stress and inflammation. *Biomed. Pharmacother.* 102, 1145–1151. doi: 10.1016/j.biopha.2018.03.037

Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., et al. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8529–8536. doi: 10.1073/pnas.1424031112

Osipova, S., Permyakov, A., Permyakova, M., Pshenichnikova, T., Verkhoturov, V., Rudikovsky, A., et al. (2016). Regions of the bread wheat d genome associated with variation in key photosynthesis traits and shoot biomass under both well watered and water deficient conditions. *J. Appl. Genet.* 57, 151–163. doi: 10.1007/s13353-015-0315-4

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J. P., et al. (2015). Potential of a tomato magic population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol. J.* 13, 565–577. doi: 10.1111/pbi.12282

Peifei, L., Bingxiu, X., and Huilin, C. (2015). *Lycium Barbarum and Tumors in the Gastrointestinal Tract*. Dordrecht: Springer.

Pnueli, L., Carmel-Goren, L., Hareven, D., Gutfinger, T., and Alvarez, J. (1998). The self-pruning gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of cen and tfl1. *Development* 125, 1979–1989. doi: 10.1242/dev.125.11.1979

Portis, E., Acquadro, A., Tirone, M., Pesce, G. R., Mauromicale, G., and Lanteri, S. (2018a). Mapping the genomic regions encoding biomass-related traits in *Cynara cardunculus* L. *Mol. Breed.* 38:64. doi: 10.1007/s11032-018-0826-x

Portis, E., Cericola, F., Barchi, L., Toppino, L., Acciarri, N., Pulcini, L., et al. (2015). Association mapping for fruit, plant and leaf morphology traits in eggplant. *PLoS One* 10:e135200. doi: 10.1371/journal.pone.0135200

Portis, E., Lanteri, S., Barchi, L., Portis, F., and Valente, L. (2018b). Comprehensive characterization of simple sequence repeats in eggplant (*Solanum melongena* L.) genome and construction of a web resource. *Front. Plant Sci.* 9:401. doi: 10.3389/fpls.2018.00401

Qian, L., and Yu, S. (2016). Protective effect of polysaccharides from *Lycium barbarum* on spermatogenesis of mice with impaired reproduction system induced by cyclophosphamide. *Am. J. Reprod. Immunol.* 76, 383–385. doi: 10.1111/aji.12558

Qin, K., and Dai, G. (2017). Progress and prospect of wolfberry strains breeding. *Ningxia J. Agric. Forest* 58, 25–28.

Qiu, S., Chen, J., Chen, X., Fan, Q., Zhang, C., Wang, D., et al. (2014). Optimization of selenylation conditions for *Lycium barbarum* polysaccharide based on antioxidant activity. *Carbohyd. Polym.* 103, 148–153. doi: 10.1016/j.carbpol.2013.12.032

Qu, M., Zheng, G., Hamdani, S., Essemine, J., Song, Q., Wang, H., et al. (2017). Leaf photosynthetic parameters related to biomass accumulation in a global rice diversity survey. *Plant Physiol.* 175, 248–258. doi: 10.1104/pp.17.00332

Quinet, M., Dielen, V., Batoko, H., Boutry, M., and Havelange, A. (2006). Genetic interactions in the control of flowering time and reproductive structure development in tomato (*Solanum lycopersicum*). *New Phytol.* 170, 701–710. doi: 10.1111/j.1469-8137.2006.01717.x

Ranc, N., Muños, S., Santoni, S., and Causse, M. (2008). A clarified position for *Solanum lycopersicum* var. Cerasiformein the evolutionary history of tomatoes (Solanaceae). *BMC Plant Biol.* 8:130. doi: 10.1186/1471-2229-8-130

Reeve, V. E., Allanson, M., Arun, S. J., Domanski, D., and Painter, N. (2010). Mice drinking goji berry juice (*Lycium barbarum*) are protected from uv radiation-induced skin damage via antioxidant pathways. *Photoch. Photobio. Sci.* 9:601. doi: 10.1039/b9pp00177h

Rehman, F., Gong, H., Li, Z., Zeng, S., Yang, T., Ai, P., et al. (2020). Identification of fruit size associated quantitative trait loci featuring SLAF based high-density linkage map of goji berry (*Lycium* spp.). *BMC Plant Biol.* 20:474. doi: 10.1186/s12870-020-02567-1

Richards, R. A. (2000). Selectable traits to increase crop photosynthesis and yield of grain crops. *J. Exp. Bot.* 51, 447–458. doi: 10.1093/jexbot/51.suppl_1.447

Rispail, N., Prats, E., and Rubiales, D. (2020). "Medicago truncatula as a model to study powdery mildew resistance," in *The Model Legume Medicago Truncatula*, ed. F. J. de Bruijn (Hoboken: John Wiley & Sons), 390–397. doi: 10.1093/pcp/pcab021

Rubiales, D., Fondevilla, S., and Fernández-Aparicio, M. (2021). Development of pea breeding lines with resistance to *Orobanche crenata* derived from pea landraces and wild *Pisum* spp. *Agronomy* 11:36. doi: 10.3390/agronomy11010036

Sharma, B. (2012). Correlation and path coefficient analysis for quantitative and qualitative traits for fruit yield and seed yield in tomato genotypes. *Indian J. Hortic.* 69, 540–544.

Sheidai, M., Narengi, Z., and Khatamsaz, M. (1999). Karyotype and seed protein analyses of *Lycium* (Solanaceae) in Iran. *Edinburgh J. Bot.* 56, 253–264. doi: 10.1017/S0960428600001116

Shi, Y., Wei, C., Chen, Z., and Hou, Y. (2016). Research progress on postharvest physiology and storage technology of fresh fruit of *Lycium barbarum* L. *Storage Process* 16, 102–106. doi: 10.3969/j.issn.1009-6221.2016.03.021

Shi, Z., Men, H., and Du, H. (2012). *Description Specification and Data Standard of Lycium Germplasm Resources*. Beijing: China Forestry Publishing House.

Shi, Z., Wei, F., Wan, R., Li, Y., and Wang, Y. (2019). Impact of nitrogen fertilizer levels on metabolite profiling of the *Lycium barbarum* L. Fruit. *Molecules* 24:3879. doi: 10.3390/molecules24213879

Singh, N. S., Sharma, R., Parween, T., and Patanjali, P. K. (2017). *Pesticide Contamination and Human Health Risk Factor*. Cham: Springer International Publishing, 49–68.

Song, W., Li, Y., Zhao, Y., Liu, Y., Niu, Y., Pang, R., et al. (2012). Construction of a high-density microsatellite genetic linkage map and mapping of sexual and growth-related traits in half-smooth tongue sole (*Cynoglossus semilaevis*). *PLoS One* 7:e0052097. doi: 10.1371/journal.pone.0052097

Souza, A. D. G., Smiderle, O. J., Spinelli, V. M., Souza, R. O. D., and Bianchi, V. J. (2016). Correlation of biometrical characteristics of fruit and seed with twinning and vigor of prunus persica rootstocks. *J. Seed. Sci.* 38, 322–328. doi: 10.1590/2317-1545v38n4164650

State Forestry Administration of China [SFAC] (2013). *Guideline for the Conduct of Tests for Distinctness, Uniformity and Stability — Lycium*. LY/T 2099-2013. Beijing: Standards Press of China.

State Forestry Administration of China [SFAC] (2018). *China Forestry Statistics Yearbook-2017*. Beijing: China Forestry Publishing.

Stebbins, G. L. (1971). *Chromosomal Evolution in Higher Plants*. London: Edward Arnold.

Stiefkens, L., and Bernardello, G. (2002). Karyotypic studies in *Lycium* section mesocope (Solanaceae) from South America. *Caryologia* 55, 199–206. doi: 10.1080/00087114.2002.10589278

Stiefkens, L., and Bernardello, G. (2005). Karyotype studies in *Lycium* sections schistocalyx and sclerocarpellum (Solanaceae). *Edinburgh J. Bot.* 62, 53–67. doi: 10.1017/S0960428606000023

Stiefkens, L., and Bernardello, L. (1996). Karyotypic studies in South American *Lycium* (Solanaceae). *Cytologia* 61, 395–402. doi: 10.1508/cytologia.61.395

Stiefkens, L., Las Peñas, M. L., Levin, R. A., Miller, J. S., and Bernardello, G. (2020). Chromosome evolution in the cosmopolitan genus *Lycium* (solanaceae). *Taxon* 69, 124–141. doi: 10.1002/tax.12205

Su, C., Duan, X., Liang, L., Feng-Wang, Zheng, J., Fu, X. Y., et al. (2014). *Lycium barbarum* polysaccharides as an adjuvant for recombinant vaccine through enhancement of humoral immunity by activating tfh cells. *Vet. Immunol. Immunopathol.* 158, 98–104. doi: 10.1016/j.vetimm.2013.05.006

Sun, Y., Lei, Y., Wu, Z., and Yang, C. (2018). *The Intellectual Property Protection of Traditional Chinese Medicine Lycium for the Prevention and Treatment of Diabetes*. Chengdu: Sichuan Province, China.

Sze, S. C., Song, J., Wong, R. N., Feng, Y., Ng, T., Tong, Y., et al. (2008). Application of scar (sequence characterized amplified region) analysis to authenticate *Lycium barbarum* (wolfberry) and its adulterants. *Biotechnol. Appl. Biochem.* 51:15. doi: 10.1042/BA20070096

Tamisier, L., Szadkowski, M., Nemouchi, G., Lefebvre, V., and Szadkowski, E. (2020). Genome-wide association mapping of QTLs implied in potato virus y population sizes in pepper: evidence for widespread resistance QTL pyramiding. *Mol. Plant Pathol.* 21, 3–16. doi: 10.1111/mpp.12874

Tang, Y., Ren, J., Liu, C., Jiang, J., and Yang, H. (2021). Genetic characteristics and QTL analysis of the soluble sugar content in ripe tomato fruits. *Sci. Hortic. Amsterdam* 276:109785. doi: 10.1016/j.scienta.2020.109785

Tanksley, S. D. (2004). The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* 16, S181–S189. doi: 10.1105/tpc.018119

Thammina, C. S., Olsen, R. T., Kramer, M., and Pooler, M. R. (2017). Genetic relationships of boxwood (*Buxus* L.) Accessions based on genic simple sequence repeat markers. *Genet. Resour. Crop Evol.* 64, 1281–1293. doi: 10.1007/s10722-016-0436-6

The American Herbal Pharmacopoeia (2019). *Lycium (Goji) Berry*. Scotts Valley, CA: The American Herbal Pharmacopoeia.

Tongumpai, P. (1993). *Strategies for Machine Harvesting of Mature Coffee (Coffea Arabica l.) Fruits*. Dissertation. Corvallis, OR: Oregon State University.

Trick, M., Adamski, N. M., Mugford, S. G., Jiang, C. C., Febrer, M., and Uauy, C. (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* 12:14. doi: 10.1186/1471-2229-12-14

Tripathi, V. (2013). Assessment of genetic diversity in berberis *Lycium royle* complex using RAPD markers. *J. Cell Biol. Genet.* 3, 1–13. doi: 10.5897/JCBG11.003

Villegas-Fernández, ÁM., Amarna, A. A., Moral, J., and Rubiales, D. (2021). Crop diversification to control powdery mildew in pea. *Agronomy* 11:690. doi: 10.3390/agronomy11040690

Wang, C. C., Chang, S. C., Inbaraj, B. S., and Chen, B. H. (2010). Isolation of carotenoids, flavonoids and polysaccharides from *Lycium barbarum* L. and evaluation of antioxidant activity. *Food Chem.* 120, 184–192. doi: 10.1016/j.foodchem.2009.10.005

Wang, H., Li, J., Tao, W., Zhang, X., Gao, X., Yong, J., et al. (2018). *Lycium ruthenicum* studies: molecular biology, phytochemistry and pharmacology. *Food Chem.* 240, 759–766. doi: 10.1016/j.foodchem.2017.08.026

Wang, J., Chen, W., Chen, J., Zhou, X., Xu, D., Li, J. H., et al. (2015a). Population genetic diversity of wild *Lycium ruthenicum* in qaidam inferred from AFLP markers. *Chinese J. Plant Ecol.* 39, 1003–1011. doi: 10.17521/cjpe.2015.0097

Wang, J., Hu, Y., Wang, D., Zhang, F., Zhao, X., Abula, S., et al. (2010). *Lycium barbarum* polysaccharide inhibits the infectivity of newcastle disease virus to chicken embryo fibroblast. *Int. J. Biol. Macromol.* 46, 212–216. doi: 10.1016/j.ijbiomac.2009.11.011

Wang, J., Wang, Q., Liu, F., Li, Y., and Jiang, J. (2017). The correlation between seed characters and fruit characters of *Lycium ruthenicum* Murr. *Seed* 36, 80–83.

Wang, J., Zhang, K., Zhang, X., Yan, G., Zhou, Y., Feng, L., et al. (2015b). Construction of commercial sweet cherry linkage maps and QTL analysis for trunk diameter. *PLoS One* 10:e0141261. doi: 10.1371/journal.pone.0141261

Wang, S., Suh, J. H., Zheng, X., Wang, Y., and Ho, C. (2017). Identification and quantification of potential anti-inflammatory hydroxycinnamic acid amides from wolfberry. *J. Agric. Food Chem.* 65, 364–372. doi: 10.1021/acs.jafc.6b05136

Wang, Y., Guo, S., An, W., Liu, L., and Yin, Y. (2016). Study on the characters and major constituents in the fruits of five wolfberry varieties. *J. For. Environ.* 36, 367–372.

Wasternack, C., Forner, S., Strnad, M., and Hause, B. (2013). Jasmonates in flower and seed development. *Biochimie* 95, 79–85. doi: 10.1016/j.biochi.2012.06.005

Wilson, L. M., Whitt, S. R., Ibaénez, A. M., Rocheford, T. R., Goodman, M. M., and Buckler, E. S. IV (2004). Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16, 2719–2733. doi: 10.1105/tpc.104.025700

Wm Hilhorst, H., Pc Groot, S., and J Bino, R. (1998). The tomato seed as a model system to study seed development and germination. *Acta Bot. Neerlandica* 47, 169–183.

Xin, T., Yao, H., Gao, H., Zhou, X., and Ma, X. (2013). Super food *Lycium barbarum* (Solanaceae) traceability via an internal transcribed spacer 2 barcode. *Food Res. Int.* 54, 1699–1704. doi: 10.1016/j.foodres.2013.10.007

Xin, Y., Wan, L., Peng, J., and Guo, C. (2011). Alleviation of the acute doxorubicin-induced cardiotoxicity by *Lycium barbarum* polysaccharides through the suppression of oxidative stress. *Food Chem. Toxicol.* 49, 259–264. doi: 10.1016/j.fct.2010.10.028

Xin, Y., Zhou, G., Deng, Z., Chen, Y., Wu, Y., Xu, P. S., et al. (2007). Protective effect of *Lycium barbarum* on doxorubicin-induced cardiotoxicity. *Phytother. Res.* 21, 1020–1024. doi: 10.1002/ptr.2186

Xing, X., Liu, F., Xiao, J., and So, K. F. (2016). Neuro-protective mechanisms of *Lycium barbarum*. *Neuromol. Med.* 18, 253–263. doi: 10.1007/s12017-016-8393-y

Xu, Y. (2010). *Molecular Plant Breeding*. Cambridge, MA: CABI North American Office.

Yang, D., Jing, R., Chang, X., and Li, W. (2007). Quantitative trait loci mapping for chlorophyll fluorescence and associated traits in wheat (*Triticum aestivum*). *J. Integr. Plant Biol.* 49, 646–654. doi: 10.1111/j.1672-9072.2007.00443.x

Yang, T., Dong, J., Yue, J., and Wang, Y. (2015). A new wolfberry cultivar'zhongke lüchuan 1'. *Acta Hortic. Sin.* 42, 2557–2558. doi: 10.16420/j.issn.0513-353x.2014-1003

Yao, R., Heinrich, M., Zou, Y., Reich, E., Zhang, X., Chen, Y., et al. (2018b). Quality variation of goji (fruits of *Lycium* spp.) In china: a comparative morphological and metabolomic analysis. *Front. Pharmacol.* 9:151. doi: 10.3389/fphar.2018.00151

Yao, R., Heinrich, M., and Weckerle, C. S. (2018a). The genus *Lycium* as food and medicine: a botanical, ethnobotanical and historical review. *J. Ethnopharmacol.* 212, 50–66. doi: 10.1016/j.jep.2017.10.010

Ye, M., Moon, J., Yang, J., Lim, H. H., Hong, S. B., Shim, I., et al. (2015). The standardized *Lycium chinense* fruit extract protects against alzheimer's disease in 3xtg-ad mice. *J. Ethnopharmacol.* 172, 85–90. doi: 10.1016/j.jep.2015.06.026

Ye, X., and Jiang, Y. (2020). *Phytochemicals in Goji Berries*. Boca Raton, FL: CRC Press.

Yin, G., and Dang, Y. (2008). Optimization of extraction technology of the *Lycium barbarum* polysaccharides by box-behnken statistical design. *Carbohyd. Polym.* 74, 603–610. doi: 10.1016/j.carbpol.2008.04.025

Yin, X. L., Fang, K. T., Liang, Y. Z., Wong, R. N., and Ha, A. W. (2005). Assessing phylogenetic relationships of *Lycium* samples using rapd and entropy theory. *Acta Pharmacol. Sin.* 26, 1217–1224. doi: 10.1111/j.1745-7254.2005.00197.x

Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotech.* 17, 155–160. doi: 10.1016/j.copbio.2006.02.003

Yu, J., Holland, J. B., McMullen, M. D., and Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics* 178, 539–551. doi: 10.1534/genetics.107.074245

Yu, J., Zhang, K., Li, S., Yu, S., and Zhai, H. (2013). Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. *Theor. Appl Genet.* 126, 275–287. doi: 10.1007/s00122-012-1980-x

Yuan, H., An, W., Li, L., Cao, Y., and Weihua, L. (2013). The investigation and cluster analysis of main morphological characters for germplasm of Chinese wolfberry. *J. Plant Genet. Resour.* 14, 627–633.

Zhang, D., Xia, T., Dang, S., Fan, G., and Wang, Z. (2018). Investigation of Chinese wolfberry (*Lycium* spp.) germplasm by restriction site-associated DNA sequencing (rad-seq). *Biochem. Genet.* 56, 575–585. doi: 10.1007/s10528-018-9861-x

Zhang, J., Zhang, Q., Cheng, T., Yang, W., Pan, H., Zhong, J., et al. (2015). High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (*Prunus mume* sieb. Et zucc). *DNA Res.* 22, 183–191. doi: 10.1093/dnares/dsv003

Zhang, K. Y., Leung, H. W., Yeung, H. W., and Wong, R. N. (2001). Differentiation of *Lycium barbarum* from its related *Lycium* species using random amplified polymorphic dna. *Planta Med.* 67, 379–381. doi: 10.1055/s-2001-14310

Zhang, X., and Sun, Z. (2011). Development status of new plant variety protection and DUS testing. *For. Res.* 24, 247–252.

Zhang, Y., Wang, L., Xin, H., Li, D., and Ma, C. (2013). Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol.* 13:141. doi: 10.1186/1471-2229-13-141

Zhao, J., Li, H., Xu, Y., Yin, Y., Huang, T., Zhang, B., et al. (2021). A consensus and saturated genetic map provides insight into genome anchoring, synteny of Solanaceae and leaf- and fruit-related QTLs in wolfberry (*Lycium* linn.). *BMC Plant Biol.* 21:350. doi: 10.1186/s12870-021-03115-1

Zhao, J., Xu, Y., Li, H., Yin, Y., An, W., Li, Y., et al. (2019). A snp-based high-density genetic map of leaf and fruit related quantitative trait loci in wolfberry (*Lycium* linn.). *Front. Plant Sci.* 10:977. doi: 10.3389/fpls.2019.00977

Zhao, W., Chung, J., Cho, Y., Rha, W., Lee, G., Ma, K. H., et al. (2010). Molecular genetic diversity and population structure in *Lycium* accessions using SSR markers. *CR Biol.* 333, 793–800. doi: 10.1016/j.crvi.2010.10.002

Zheng, G., Su, X., Ma, Y., Qi, G., and Yang, J. (2015). Effect of seed characters on fruit size of *Lycium barbarum*. *N. Hortic.* 6, 134–137.

Zheng, G., Zhang, L., Wang, J., and Hu, Z. (2012). The morphology development of fruit and seed in *Lycium barbarum*. *Guangxi Zhiwu Guihaia* 32, 810–815.

Zheng, T. C., Zhang, X. K., Yin, G. H., Wang, L. N., Han, Y. L., Chen, L., et al. (2011). Genetic gains in grain yield, net photosynthesis and stomatal conductance achieved in Henan province of China between 1981 and 2008. *Field Crops Res.* 122, 225–233. doi: 10.1016/j.fcr.2011.03.015

Zheng, Z., Wang, R., Zhang, Y., Zhao, Z., and Gao, L. (2018). "Present situation and development of harvesting mechanization of chinese medicinal materials," in *Proceedings of the 2018 ASABE Annual International Meeting*, (St. Joseph, MI: ASABE).

Zhou, H., Li, P., Xie, W., Hussain, S., Li, Y., Xia, D., et al. (2017). Genome-wide association analyses reveal the genetic basis of stigma exsertion in rice. *Mol. Plant* 10, 634–644. doi: 10.1016/j.molp.2017.01.001

Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1, 5–20. doi: 10.3835/plantgenome2008.02.0089

Zhu, X., Song, Q., and Ort, D. R. (2012). Elements of a dynamic systems model of canopy photosynthesis. *Curr. Opin. Plant Biol.* 15, 237–244. doi: 10.1016/j.pbi.2012.01.010

Zhu, Y., Yin, Y., Yang, K., Li, J., Sang, Y., Huang, L., et al. (2015). Construction of a high-density genetic map using specific length amplified fragment markers and identification of a quantitative trait locus for anthracnose resistance in walnut (*Juglans regia* L.). *BMC Genom.* 16:614. doi: 10.1186/s12864-015-1822-8

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 February 2022
doi: 10.3389/fgene.2022.847328

Genome-Wide Characterization and Anthocyanin-Related Expression Analysis of the *B-BOX* Gene Family in *Capsicum annuum* L.

Jin Wang[1,2], *Guangbin Yang*[3], *Ying Chen*[3], *Yao Dai*[1,4], *Qiaoling Yuan*[1,4], *Qingyun Shan*[1,4], *Luzhao Pan*[1,2], *Li Dai*[1,4], *Xuexiao Zou*[1,4]*, *Feng Liu*[1,4]* and *Cheng Xiong*[1,4]*

[1]College of Horticulture, Hunan Agricultural University, Changsha, China, [2]College of Horticulture, Nanjing Agricultural University, Nanjing, China, [3]Hunan Vegetable Research Institute, Hunan Academy of Agricultural Sciences, Changsha, China, [4]Engineering Research Center for Horticultural Crop Germplasm Creation and New Variety Breeding, Ministry of Education, Changsha, China

The transcription factors, B-box (BBX), belong to a subfamily of the zinc finger family of proteins and exhibit multiple biological functions in plant growth, development, and abiotic stress response pathways. In this study, a total of 23 *CaBBX* members were identified using the pepper reference genome database. According to the gene structure, conserved domains, and the phylogenetic tree, 23 *CaBBX* genes were divided into four groups, wherein the analysis of the promoter region indicated the presence of *cis*-acting elements related to plant development, hormones, and stress response. Interspecies collinearity analysis showed that the *CaBBXs* had three duplicated gene pairs, and the highest gene density was found on chromosomes 2 and 7. Transcriptome RNA-seq data and quantitative polymerase chain reaction (qRT-PCR) analysis of pepper plants spanning the entire period showed that more than half of the *CaBBX* genes were widely expressed in diversity tissues of pepper. Co-expression network analysis indicated that the *CaBBXs* and the anthocyanin structural genes had a close co-expression relationship. Thus, it was reasonably speculated that the *CaBBX* genes may be involved in the regulation of anthocyanin biosynthesis. Overall, this study involved the genome-wide characterization of the *CaBBX* family and may serve as a solid foundation for further investigations on *CaBBX* genes involved in the anthocyanin synthesis mechanisms and development in pepper.

Keywords: pepper, CaBBXs, duplication, development expression, anthocyanin biosynthesis

OPEN ACCESS

Edited by:
Ezio Portis,
University of Turin, Italy

Reviewed by:
Ling Xu,
Zhejiang Sci-Tech University, China
Cheng Qin,
Zunyi Vocational and Technical College, China

***Correspondence:**
Feng Liu
jwszjx@hunau.edu.cn
Xuexiao Zou
zouxuexiao428@163.com
Cheng Xiong
xiongchenghzau@163.com

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

Received: *02 January 2022*
Accepted: *09 February 2022*
Published: *28 February 2022*

Citation:
Wang J, Yang G, Chen Y, Dai Y, Yuan Q, Shan Q, Pan L, Dai L, Zou X, Liu F and Xiong C (2022) Genome-Wide Characterization and Anthocyanin-Related Expression Analysis of the B-BOX Gene Family in Capsicum annuum L.. Front. Genet. 13:847328. doi: 10.3389/fgene.2022.847328

INTRODUCTION

Pepper (*Capsicum*), an important species of the Solanaceae family, has significant popularity and economic value. It is widely cultivated in almost all areas with arable land, owing to its high nutritional value and complex germplasm diversity (Qin et al., 2014). As an important trait in pepper, the color is a vital indicator for classification, maturity, and evolutionary characteristics of the pepper species (Liu et al., 2020a). A previous study shows that anthocyanins, carotenoids, and beet pigments are the main factors that determine the color of leaves, flowers, and fruits in natural plants and that their types and contents control the shades of different organs (Tang et al., 2020).

Anthocyanins, a flavonoid compound, is widely present in plants. They can not only impart color and attract pollinators and seed dispersers but also protect them from ultraviolet radiation. Anthocyanins play an important regulatory role in several plants through their antioxidants, phytoales, and antibacterial effects (Tang et al., 2020). Peppers rich in anthocyanins can reduce the risk of cancer, alleviate neurological dysfunction, and prevent cardiovascular diseases (Fang, 2015). The biosynthesis of anthocyanins is characterized by catalysis through an enzymatic cascade. First, malonyl-CoA and coumarin-CoA are used as substrates to produce chalcone through the catalysis of chalcone synthase (CHS). Chalcone is catalyzed by chalcone isomerase (CHI) to produce naringenin, which is then converted to dihydrokaempferol by flavanone 3-hydroxylase (F3H). Dihydroflavonol 4-reductase (DFR), UDP-glucose: flavonoid 3-glucosyltransferase (UFGT), and anthocyanin synthase (ANS) regulate the anthocyanin synthesis in different species. Anthocyanin reductase (ANR) and leucoanthocyanidin reductase (LAR) are the key factors in proanthocyanidin biosynthesis (Alappat and Alappat, 2020). The synthesis of various enzymes regulates the biosynthesis of anthocyanins.

The biosynthesis of anthocyanins is also regulated by several transcription factors. Among them, the MYB-bHLH-WD40 ternary complex (MBW complex) as a typical representative is known to positively regulate anthocyanin biosynthesis. In *Arabidopsis*, the WBM complex includes four *MYB* transcription factors, namely *AtPAP1*, *AtPAP2*, *AtMYB113*, and *AtMYB114*, three *bHLH* genes, *AtGL3*, *AtEGL3,* and *AtTT8*, and a typical WD40 protein, AtTTG1. In apple, the *MdMYB10* transcription factor was first identified as a member of the MBW complex and found to induce the accumulation of anthocyanins in heterologous and homologous species. It is positively correlated with anthocyanin levels during fruit development in apple plants. In carrot (*Daucus carota*), the *DcMYB7* gens regulate the inheritance of anthocyanin pigmentation in purple and non-purple carrot roots (Xu et al., 2019). In the subtropical deciduous tree species, Formosan sweetgum (*Liquidambar formosana* Hance), the *R2R3-MYB* transcription factor gene, *LfMYB113* is positively correlated with anthocyanin content in the leaves (Wen and Chu, 2017). In transgenic tomato fruits, the overexpression of two MYB genes, *SlANT1* or *SlAN2,* results in the accumulation of anthocyanins and the up-regulation of *EBGs*, *LBGs,* and *SlAN1*. The bHLH protein, SlANT1, influences the induction of anthocyanin biosynthesis (Meng et al., 2015). In pepper, the *CaAN2* gene is the main transcription factor that regulates and controls anthocyanin biosynthesis; the variations in the promoter non-LTR retrotransposon region caused the differences in colors of the pepper fruits (Jung et al., 2019). Simultaneously, the *CaAN2* gene can regulate the expression of structural genes (*CaF3′5′H*, *CaDFR*, *CaUFGT*, *CaCHS*, and *CaF3H*) and modulate the anthocyanin content in pepper (Zhang et al., 2015). In addition to the MBW complex that directly regulates anthocyanin biosynthesis, some other transcription factors also play a regulatory role in anthocyanin synthesis, directly or indirectly. For example, in petunia, the *WRKY* transcription factor, *PhPH3,* and MBW complex regulate the expression of the two proton pumps (PhPH1 and PhPH5) that are further responsible for vacuolar acidification, thereby imparting the red color to the anthocyanins (Verweij et al., 2016). An anthocyanin biosynthesis gene, *Ca3GT*, has been finely located in the pepper chromosome 10, and it is a strong candidate gene for anthocyanin biosynthesis in mature and immature pepper fruits (Liu J. et al., 2020).

In addition, environmental stresses, including extreme light intensity, mechanical damage, and/or low temperature can induce the accumulation of anthocyanins through the MBW complex and regulation by other transcription factors (Movahed et al., 2016). Light is an important environmental factor influencing the accumulation of anthocyanins. High light intensity stimulates the gene expression and production of anthocyanins in many species (Su et al., 2016). Among them, the B-box (BBX) zinc finger protein family transcription factors have attracted much attention in recent years. The N-terminus of the BBX proteins contain one or two conserved BBX domains (B-box1 and B-box2), and some members also possess a CCT (Constans, CO-like, and TOC1) structure at the C-terminal region (Huang et al., 2012). Several studies show that under light-mediated conditions, the B-box (BBX) protein interacts with *ELONGATED HYPOCOTYL 5* (*HY5*) and *CONSTITUTIVE PHOTOMORPHOGENIC 1* (*COP1*) and further, mediates their regulatory activities (Gangappa and Botto, 2016). In rice (*Oryza sativa*), anthocyanin biosynthesis is induced and fine-tuned by *OsBBX14* and *OsHY5*. The ectopic expression of *OsBBX14* in *Arabidopsis thaliana* results in a significant increase in the accumulation of anthocyanins in its seedlings (Kim et al., 2018). In apple, the BBX protein MdBBX37 plays a negative role in light signal transduction. *MdBBX37* inhibits the binding of *MdMYB1* and *MdMYB9*, two key positive regulators of anthocyanin biosynthesis, with their target genes, thereby contributing to the negative regulation of anthocyanin biosynthesis pathway (An et al., 2019). The expression of *PpBBX16* in pear (*Pyrus pyrifolia*) is highly induced by white light irradiation, and the transient overexpression of *PpBBX16* in the pear skin increases the accumulation of anthocyanins, while the virus-induced *PpBBX16* gene silencing reverses this trend substantially (Bai et al., 2019). In poplars (*Populus L.* spp.), *BBX23* directly binds to the promoter regions of proanthocyanidins and anthocyanin-specific genes and interacts with HY5 to enhance their activation (Li et al., 2021).

Although BBX-mediated flavonoid accumulation has been reported in multiple species, whether the candidate *CaBBX*s in pepper were related to anthocyanins, remains unknown. In this study, the *CaBBX* family were identified in pepper from the draft of the whole genome sequences, and the *CaBBX* family of genes in pepper were found to have a wide range of biological functions in the fruit development processes, with divergent regulatory effects in the regulatory processes of anthocyanin synthesis. In conclusion, this finding may provide central guidance for future research on the function of *CaBBX* genes in peppers, and laid a solid foundation for the study of the biosynthesis mechanism of pepper anthocyanins.

MATERIALS AND METHODS

Identification of *CaBBX* Transcription Factors in Pepper

To identify the BBX family of transcription factors, the hidden Markov model (HMM) profile of the B-box-type zinc finger domain (PF00643) was downloaded from the Pfam database (El-Gebali et al., 2019). The reference genomes of hot pepper, *C. annuum* cv. CM334 (Criollo de Morelos 334), and a Chinese inbred derivative, 'Zunla-1', were used in this study. The HMMER SEARCH3.0 with cutoff E value ≤0.01 and SMART database were used to confirm the conserved domains among the members of the BBX family (Finn et al., 2011). Theoretical pI and Mw values were computed using the ExPaSy online tool (Gasteiger et al., 2003). The SWISS-MODEL online tool was used for protein modeling (Schwede et al., 2003), and the softberry online software was used to predict the putative subcellular localization of genes (http://www.softberry.com).

Conserved Motifs, Gene Structures, and Phylogenetic Analysis

The conserved domains were estimated using the MEME online tool (Version 5.1.0, National Institutes of Health, Bethesda, MD, United States) and these domains were aligned using DNAMAN (Version 8.0.8, Lynnon Biosoft). The Gene Structure Display Server was used to predict the gene structures of *CaBBXs* (Hu et al., 2015). The full-length amino acid sequences of *CaBBXs* were acquired from the pepper reference genome, "Zunla-1" (http://peppersequence.genomics.cn/), the sequences of SlBBXs were obtained from the tomato reference genome "SL4.0" (https://solgenomics.net/organism/Solanum_lycopersicum/genome), and those of AtBBXs were obtained from the *Arabidopsis* reference genome (https://www.arabidopsis.org/). MEGA-X was used to construct an unrooted neighbor-joining phylogenetic tree of CaBBXs, SlBBXs, and AtBBXs proteins with the bootstrap test replicated 1,000 times (Kumar et al., 2018).

Synteny Analysis and *Cis*-Elements

The intra-species collinearity analysis of the *CaBBXs* was performed with MCScanx (Wang et al., 2012). The synonymous (Ks) and nonsynonymous (Ka) substitution rates were estimated using the KaKs_Calculator1.2 (Zhang et al., 2006). The Circos diagram showed the chromosomal location, collinearity, and Ka/Ks values for the genes (Krzywinski et al., 2009). 2.0 kbp upstream sequences of CDS were selected as the promoter region, and the PlantCARE web tool was used to predict the *cis*-elements in each gene (Lescot et al., 2002). The Tbtools were used for plotting (Chen et al., 2020).

Materials, and RNA-seq Analysis

The TransZol kit (TransGen Biotech, Inc., Beijing, China) was used to extract total RNA, and HiScript® IIQRT SuperMix for qPCR (+gDNA wiper) vazyme kit (Vazyme, Piscataway, NJ, United States) was used to synthesize the reverse cDNA. The raw data for the RNA-seq analysis were downloaded from Pepper Hub (Liu et al., 2017), and a high-generation inbred capsicum line "6421", was used. Fastqc was used for quality control of the sequencing data (Brown et al., 2017) and the low-quality sequences were removed using Trimmmatic-0.36 (Bolger et al., 2014). HISAT2 was used to align the sequencing reads to the reference genome, "Zunla" (Kim et al., 2014), and FeatureCounts was used to calculate the number of counts (Liao et al., 2014). Standardize counts data from DESeq2 package in R (Varet et al., 2016), and the fragments per kilobase of exon model per million mapped reads (FPKM) value represented the corresponding gene expression.

Co-Expression Analysis

The weighted Gene Co-Expression Network Analysis (WGCNA) was used to analyze the FPKM value from RNA-seq data. The WGCNA package in R v3.6.1 language was used for the analysis (Langfelder and Horvath, 2008). The TOM similarity algorithm calculated the adjacent ordered function formed by the gene network and the coefficient of difference between the different nodes; the weighted gene correlation network analysis method was used to analyze the gene expression patterns in multiple samples. The co-expression correlation matrix was calculated and the gene expression correlation in the network was estimated. Cytoscape v3.6.0 was used to draw the correlation network diagram for the non-weight coefficients (weights) of the relevant *CaBBXs* and anthocyanin structural genes in the extracted matrix (Shannon et al., 2003).

GO Enrichment Analysis and Real-Time Fluorescence Quantitative PCR

GO enrichment analyses were conducted using Plant Transcriptional Regulatory Map (Jin et al., 2017) and WEGO2.0 (Ye et al., 2006) online tool with the corrected *p*-value <0.05. LightCycle * 96 Real-Time PCR System (Roche, Basel, Switzerland) was used for performing qRT-PCR. A 50 μl reaction system was set and three biological repeats and three technical repeats were performed for each sample based on HiScript® IIQ RT SuperMix for qPCR (+gDNA wiper) vazyme kit (Vazyme, Piscataway, NJ, United States). The $2^{-\triangle\triangle Ct}$ formula was used to calculate the relative expressions. The primers for *CaBBXs* and *actin* control (*Capana04g001698*) were listed in **Supplementary Table S2**.

RESULTS

Genome-Wide Identification and Conserved Domains of *CaBBX* Genes

According to the hidden Markov model (HMM)-based profile (Bbox-type zinc-finger domain, PF00643), the genome-wide annotated proteins were screened and verified using both SMART and Pfam databases. A total of 23 *CaBBX* genes were identified in the pepper Zunla-1 genome, however, 22 *CaBBXs* were present in the pepper CM334 genome. Based on their conserved motifs, 23 *CaBBX* genes were named as *CaBBX1* to *CaBBX23* within the gene subfamilies. The length of the coding

TABLE 1 | The information on the *CaBBX* gene family.

GeneID	Zunla_2.0	CM334_1.55	Chr	Start	End	Strain	PI	Mw	Subcellular localization
CaBBX1	*Capana02g002620*	*CA00g70490*	Chr02	148180402	148183481	+	8.47	22972.52	Nuclear
CaBBX2	*Capana04g000266*	*CA04g21130*	Chr04	4091269	4092722	–	6.24	33860.88	Nuclear
CaBBX3	*Capana06g000735*	*CA06g21770*	Chr06	11697317	11699646	–	5.00	25979.56	Nuclear
CaBBX4	*Capana07g001114*	*CA00g25500*	Chr07	154056816	154057654	+	5.03	24831.96	Nuclear
CaBBX5	*Capana07g002062*	*CA07g16670*	Chr07	212563126	212570473	+	4.98	32012.77	Nuclear
CaBBX6	*Capana08g002611*	*CA01g19640*	Chr08	149745414	149746937	+	5.62	29831.52	Nuclear
CaBBX7	*Capana08g002625*	*CA01g19780*	Chr08	149926925	149929248	–	6.17	23515.68	Extracellular
CaBBX8	*Capana09g000394*	*CA07g16670*	Chr09	12877150	12887068	+	4.97	32173.20	Nuclear
CaBBX9	*Capana11g002294*	*CA11g01070*	Chr11	218148220	218150401	–	4.85	35202.56	Nuclear
CaBBX10	*Capana12g000659*	*CA12g17630*	Chr12	18213299	18215385	–	6.50	35220.55	Nuclear
CaBBX11	*Capana00g001486*	*CA11g11000*	Chr00	399610115	399614902	–	6.75	51670.91	Nuclear
CaBBX12	*Capana01g004030*	*CA01g28430*	Chr01	278243944	278245872	–	5.45	37932.28	Nuclear
CaBBX13	*Capana02g003199*	*CA02g26670*	Chr02	157107846	157110095	+	5.41	44556.35	Nuclear
CaBBX14	*Capana02g003200*	*CA02g26680*	Chr02	157118313	157120062	+	5.27	43889.44	Nuclear
CaBBX15	*Capana02g003201*	*CA02g26690*	Chr02	157124787	157126407	+	5.32	45394.43	Nuclear
CaBBX16	*Capana03g000377*	*CA03g33660*	Chr03	5273465	5275489	–	5.53	43464.09	Nuclear
CaBBX17	*Capana03g003558*	*CA03g08890*	Chr03	228734754	228738294	–	5.45	43342.99	Nuclear
CaBBX18	*Capana07g000030*		Chr07	1563552	1565426	–	6.56	42540.99	Nuclear
CaBBX19	*Capana07g001588*	*CA07g12780*	Chr07	191773542	191774577	+	4.60	27399.14	Extracellular
CaBBX20	*Capana12g000414*	*CA12g19270*	Chr12	8179392	8181909	–	5.29	39425.86	Nuclear
CaBBX21	*Capana00g004911*	*CA06g15580*	Chr00	672862631	672863374	–	9.17	27712.80	Extracellular
CaBBX22	*Capana05g001195*	*CA05g08470*	Chr05	84622805	84628926	+	6.00	54983.39	Extracellular
CaBBX23	*Capana10g000048*	*CA10g00340*	Chr10	574455	577401	–	4.29	20332.93	Extracellular

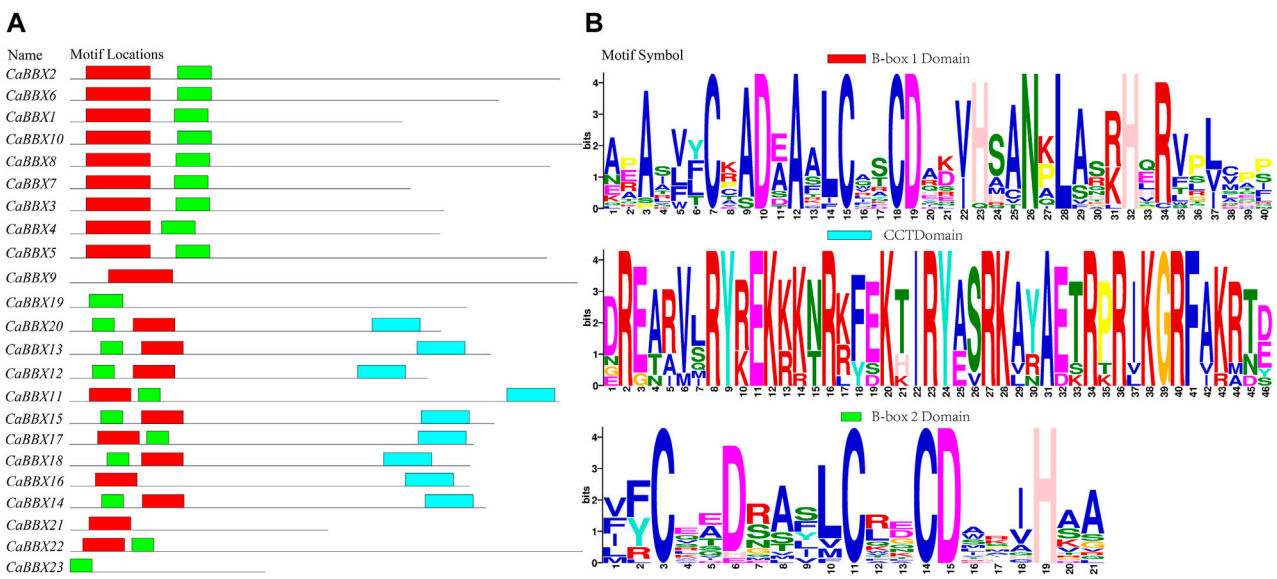

FIGURE 1 | The domains of the *CaBBX* gene family. **(A)** The distributions of conserved motifs in *CaBBXs*. **(B)** Typical logo of the *CaBBX* conserved domains.

sequences ranged between 564 (*CaBBX23*) to 1476 (*CaBBX22*) amino acid residues, while their molecular weights (Mw) varied greatly between 20.33293 KDa (*CaBBX23*) and 54.98339 KDa (*CaBBX22*). The isoelectric points (PI) of the *CaBBXs* ranged between 4.29 (*CaBBX23*) and 9.17 (*CaBBX21*), and most of them were weakly acidic proteins. Prediction software for subcellular localization suggested that these CaBBX proteins were mainly localized in the nucleus; five members, including *CaBBX7*, *CaBBX19*, *CaBBX21*, *CaBBX22*, and *CaBBX23* were found to localize to extracellular structures, such as chloroplasts and mitochondria (**Table 1**).

The conserved sequence of the zinc finger domain of B-box was found to exist in two forms, namely B-box1: C-X_2-C-X_{7-8}-C-X_2-D-X-A-X-L-C-X_2-C-D-X_3-H-X_2 -N-X_4-H (indicated in red color), and B-box2 of C-X_2-C-X_8-C-X_7-C-X_2- C-X_4-H(N)-X_{6-8}-H (indicated in green color). Protein sequence alignment showed that the two B-BOX domains had highly similar conserved sequences (**Supplementary Figure S1**). For example, the

cysteine (C) and aspartic acid (D) residues constituting the zinc fingers were highly conserved in the B-BOX domain. However, the conservative motif symbol, as shown in **Figure 1B**, was reversed owing to the conserved amino acid residues (Asn, Leu, His, and Arg) in the B-box1 domain; the B-box1 domain motifs were more conserved than those of the B-box2 domain. In addition, the BBX family also had a highly conserved CCT domain R-X_5-R-Y-X_2-K-X_3-R-X_3-K-X_2-R-Y-X_2-R-K-X_2-A-X_2-R-X-R-X_2-G-R-F-X-K, as shown as a blue bar in **Figure 1A**. Among all the 23 CaBBX proteins, only three members (*CaBBX*9, *CaBBX*16, *CaBBX*21) contained one B-box1 domain, while two genes (*CaBBX*19, *CaBBX*23) had only one B-box2 domain. The remaining 18 genes existed in both two B-box domains, and nine members contained the CCT-conserved domain.

Phylogenetic Relationships, Gene Structure, and *Cis*-Elements in the CaBBX Family

Based on conserved sequences in the proteins, all the 23 CaBBX proteins were categorized into five distinct phylogenetic classes. In evolutionary classification, CaBBX1, CaBBX2, CaBBX3, CaBBX5, CaBBX6, CaBBX7, CaBBX8, CaBBX10, CaBBX4, and CaBBX22 genes were found to be homologous (Group I), and they all contained two conserved B-box domains. CaBBX9, CaBBX16, CaBBX20, CaBBX12, CaBBX18, CaBBX15, CaBBX13, and CaBBX14 genes were in the Group II. CaBBX21, CaBBX19, and CaBBX23 were classified as Group III, while Group IV included CaBBX11 and CaBBX17 proteins (**Figure 2A**). The gene structure analysis indicated that the number of exons in the *CaBBX* family of genes in pepper was between one and five. Among them, 12 (52.17%) *CaBBXs* had three exons, five (21.74%) *CaBBXs* had two exons, two *CaBBXs* had four exons, and three *CaBBXs* had five exons. *CaBBX21* only had one exon (**Figure 2B**). To examine the phylogenetic relationship of the *CaBBX* gene family, a neighbor-joining (NJ) phylogenetic tree was constructed based on the CaBBX protein sequences in pepper, AtBBX protein sequences in Arabidopsis, and SlBBX protein sequences in tomato. According to the phylogenetic tree, all BBXs were divided into four clades, wherein, each clade contained a similar number of BBX genes from the three species. Almost all these four clades corresponded to their designated domains, while some CaBBX proteins with similar domains (such as CaBBX2 in Clade II and CaBBX6 in Clade III) were clustered into different clades (**Supplementary Figure S2**).

2,000 base pair (bp)-upstream of the coding region was selected as the promoter sequence in the *CBBXs* to predict the *cis*-regulatory elements. A total of three *cis*-element types were identified in the promoter regions of the *CaBBox* gene family. A few genes had *cis*-elements related to cellular development, such as flavonoid biosynthetic genes regulatory elements. In addition, many CaBBX genes contained *cis*-elements related to stress response pathways, including responsiveness to light, anaerobic induction, defense and stress responsiveness, and circadian control. Six typical hormone-related *cis*-regulatory elements in the promoter regions were identified, including abscisic acid response elements, salicylic acid, and gibberellin, and MeJA response elements. Specifically, there were 13 *cis*-elements for light responsiveness, including Box 4, G-Box, TCT-motif, GT1-motif, GATA-motif, I-box, AE-box, chs-CMA1a, Sp1, ATC-motif, MRE, GA-motif, and Box II (**Figure 2C**).

Chromosomal Distribution and Interspecies Synteny Analyses for CaBBXs

According to the physical positional information for the CaBBX gene family in the annotated zunla-genome, the 23 CaBBXs were found to be widely distributed in all 12 chromosomes, "01 g" to "12 g", of pepper. Two CaBBX genes (CaBBX11 and CaBBX21) were located on chromosome "00". Chromosomes 7 and 12 contained the most number of CaBBX genes, each harboring four CaBBXs. Chromosomes 8, 12, and 13 each contained two genes, and the remaining chromosomes each harbored one CaBBX member (**Figure 3**).

Only three segmental duplicated pairs and no tandem duplicates were observed for CaBBXs in pepper. In particular, CaBBX1-CaBBX2 and CaBBX2-CaBBX6 were pairs of mutually duplicated genes, while CaBBX1-CaBBX7 were present as duplicates. CaBBX6 and CaBBX7 were located on the same chromosome and were in close physical proximity. The selection pressure in duplicated genes was estimated using the values of non-synonymous mutation (Ka), synonymous mutation (Ks), and their ratio (Ka/Ks). The Ks values ranged between 1.76 and 3.16. The Ka/Ks values for all duplicated gene pairs were less than 1.00 and between 0.128 and 0.5. It was reversed that the three groups of CaBBX duplicated genes underwent purifying selection during the process of evolution. The minimum Ks and maximum Ka/Ks values were between the duplicate gene pair, CaBBX1-CaBBX2, and these two genes may have undergone greater purifying selection (**Figure 3**).

Expressions of CaBBX DEGs in Developmental Stages Analyzed by RNA-seq

To gain insight into the potential functions of the CaBBX gene family spanning the entire developmental stage of the pepper plant, the raw BAM data of RNA-seq for various organs and tissues at different developmental stages were downloaded from the Pepperhub online database and used for analyzing the expression profiles. As shown in **Figure 4**, most CaBBX genes showed differential expressions in the different organs and tissues. Based on the differential tissue expressions, the 23 CaBBXs were divided into five groups. Among them, five CaBBXs (CaBBX14, CaBBX15, CaBBX16, CaBBX17, and CaBBX9) were relatively highly expressed in the leaves, while CaBBX2 and CaBBX8 were specifically highly expressed in the late seed stages and were not expressed or showed low expression in other tissue types. CaBBX22, CaBBX5, CaBBX4, CaBBX3, CaBBX10, and CaBBX12 showed high expression in both

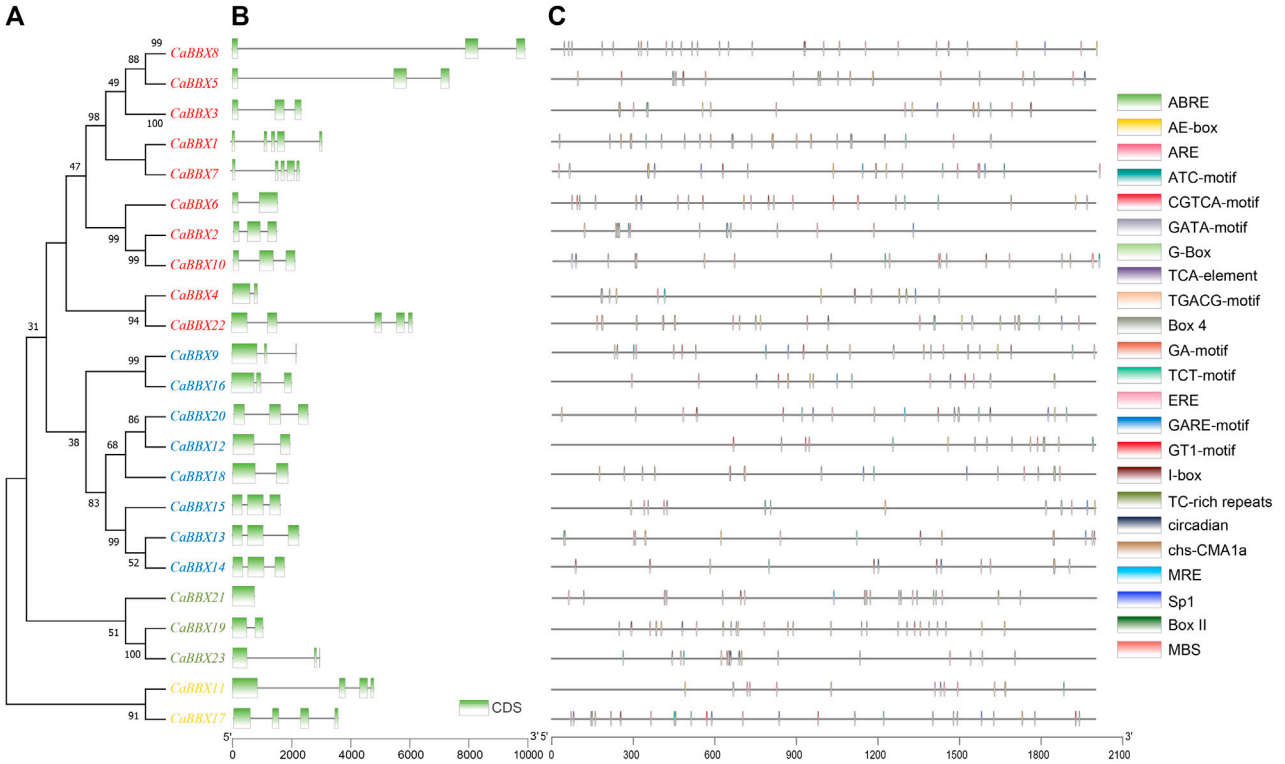

FIGURE 2 | Phylogenetic relationships, gene structure analysis, and predicted *cis*-elements of *CaBBXs*. **(A)** Phylogenetic tree for the 23 *CaBBXs*. Red genes were classified in Group I, blue members belonged to Group II, green genes to Group III, and yellow members to Group IV. **(B)** Green boxes represent the coding sequence (CDS) regions while intron regions are indicated by black lines. Exon lengths can be inferred from the scale at the bottom. **(C)** Promoter sequences (–2,000 bp) of *CaBBXs* were analyzed using PlantCARE. Different shapes and colors represent different elements. Annotated *cis*-elements are listed in **Supplementary Table S1**.

placenta and fruit, and CaBBX6, CaBBX18, CaBBX20, CaBBX7, CaBBX21, CaBBX23, CaBBX1, and CaBBX13 were highly expressed in all the flowering stages. Specifically, CaBBX11 and CaBBX19 showed low levels of expression across all the stages in leaf development, while CaBBX12 and CaBBX3 showed low levels of expression across all the stages of seed development. These results indicated that the CaBBX genes had diverse functions during the growth and development of pepper plants.

GO Enrichment and Co-Expression of CaBBXs and Anthocyanin Structural Genes

All 23 *CaBBXs* were used enrichment analysis and conducted on the assigned GO terms with the corrected *p*-value <0.05. Three GO categories were assigned, including cellular component, molecular function, and biological process. In cellular component, the cell part and cell were highly enriched in 23 *CaBBX* genes. In molecular function, *CaBBXs* were assigned in binding and cellular process. In biological process, *CaBBXs* were widely assigned in metabolic process, multicellular organismal process, developmental process, regulation of biological process, biological regulation, response to stimulus, reproduction, negative regulation of biological process, reproductive process, signaling, rhythmic process, cellular component organization or biogenesis, and positive regulation of biological process (**Figure 5A**).

According to the expression levels of 23 *CaBBXs* in different organs, the candidate *CaBBX* genes were predicted related to the anthocyanin synthesis pathway according to tissue specific expression. As shown in **Figure 5B**, all *CaBBX* genes were used for the construction of the co-expression network, along with key structural genes in the anthocyanin synthesis pathway. Among them, seven *CaBBX* members (*CaBBX1*, *CaBBX9*, *CaBBX13*, *CaBBX16*, *CaBBX20*, *CaBBX21*, and *CaBBX23*) (weighted TOM value >0.05) were mostly co-expressed with anthocyanin biosynthesis. In the early anthocyanin synthesis stages, *CaCHS*, and *CaCHI* were highly co-expressed with *CaBBX1*, *CaBBX20*, *CaBBX21*, and *CaBBX23*. *CaF3H* were co-expressed in the network but showed low co-expression with the genes in the anthocyanin synthesis pathway. In the late anthocyanin synthesis stages, *CaDFR* was also highly co-expressed with *CaBBX1*, *CaBBX20*, *CaBBX21*. And *CaANS* was highly co-expressed with *CaBBX9*, *CaBBX15*, and *CaBBX16*. Besides, *R2R3-MYB* transcription factors *CaAN1* and *CaAN2*, which involved in anthocyanin biosynthesis were also co-expressed with *CaBBXs*. *CaBBX1*,

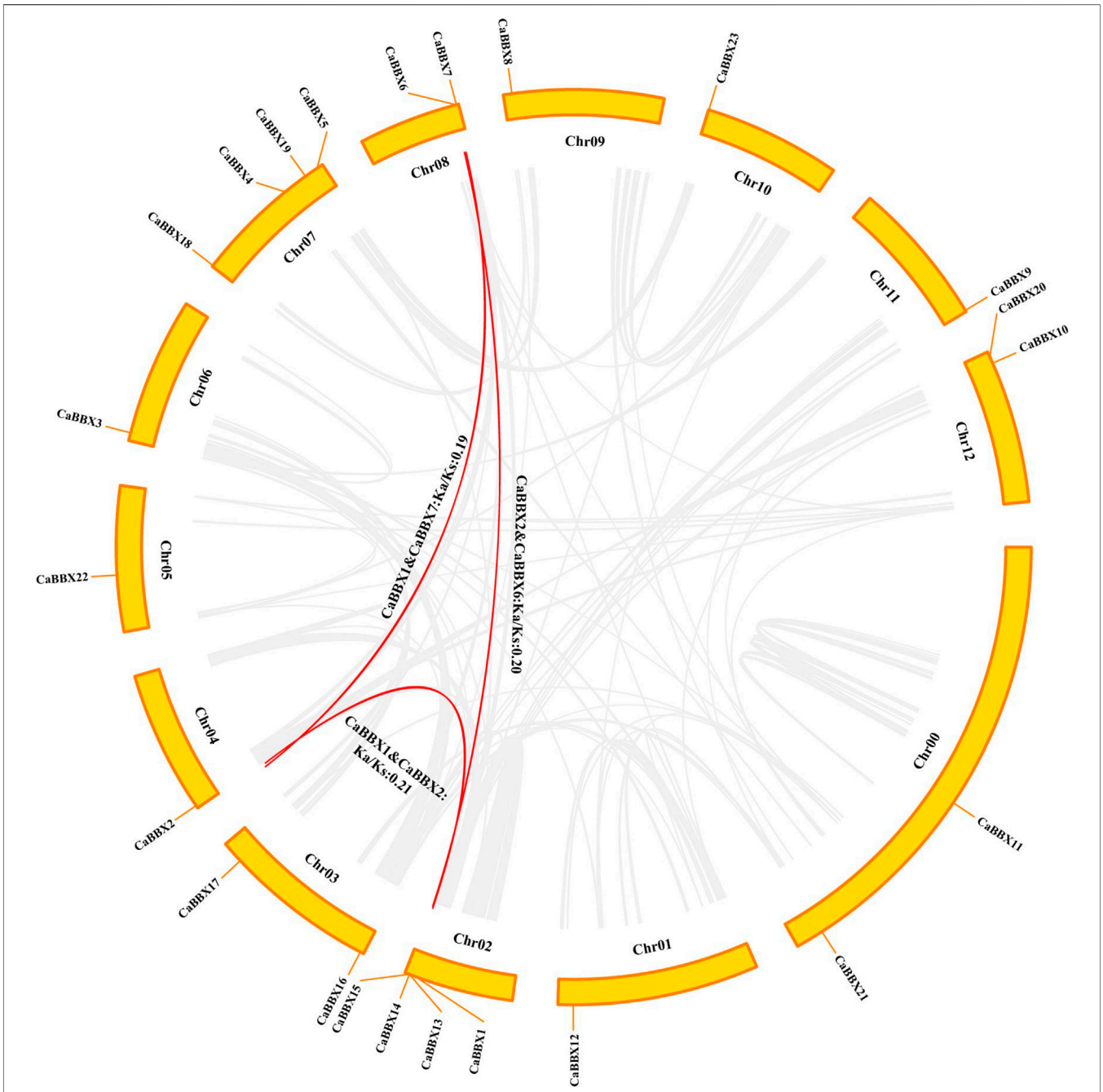

FIGURE 3 | Chromosomal location and interspecies gene synteny in *CaBBX*s. The chromosomal positions of *CaBBX* genes in the pepper genome. Red lines in the middle indicate duplicated gene pairs of *CaBBX*s. The positions of *CaBBX* genes in the pepper genome are marked on chromosomes; Chr refers to chromosomes.

CaBBX20, and *CaBBX21* were co-expressed with *CaAN2*, but *CaBBX9*, *CaBBX15*, and *CaBBX16* were co-expressed with *CaAN1*. In co-expression network, *CaBBX3*, *CaBBX4*, *CaBBX5*, *CaBBX6*, *CaBBX7*, *CaBBX11*, *CaBBX14*, *CaBBX15*, *CaBBX17*, *CaBBX18*, and *CaBBX29* were indirect co-expression with key seven *CaBBX* genes or anthocyanin structural genes to affect anthocyanin biosynthesis. In addition, there was a close co-expression relationship between the four *CaBBX* genes (*CaBBX2*, *CaBBX8*, *CaBBX12*, and *CaBBX22*).

qRT-PCR Analysis for Expressions of CaBBXs and Anthocyanin Structural Genes

Additionally, some typical CaBBXs were screened by qRT-PCR (Real-time polymerase chain reaction) and verification in the

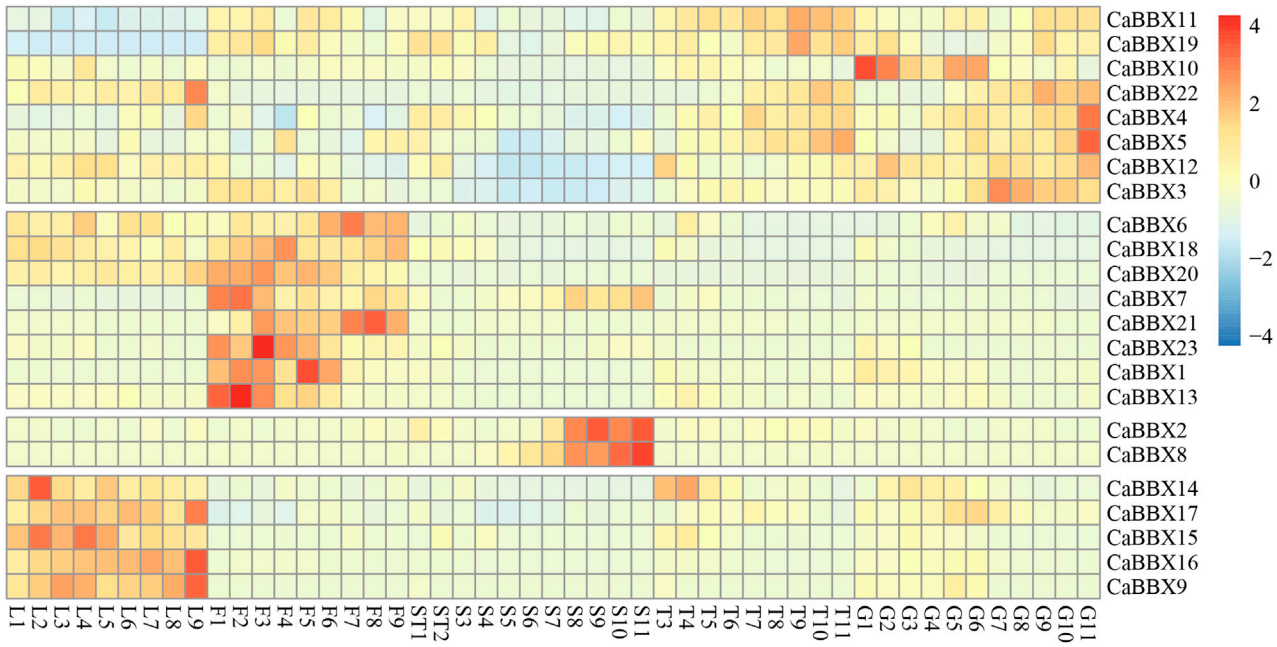

FIGURE 4 | A Expression patterns of *CaBBXs*. Expression profiles using RNA-seq fragments per kilobase million (FPKM) data in the leaves between ten DAP and 60 DAP (L1–L9), flowers between ten DAP and 50 DAP (F1-F9), placenta and seed between ten DAP and 15 DAP (ST1 and ST2), the placenta between 20 DAP and 60 DAP (T3–T10), and pericarp between ten DAP and 60 DAP (G1–G11).

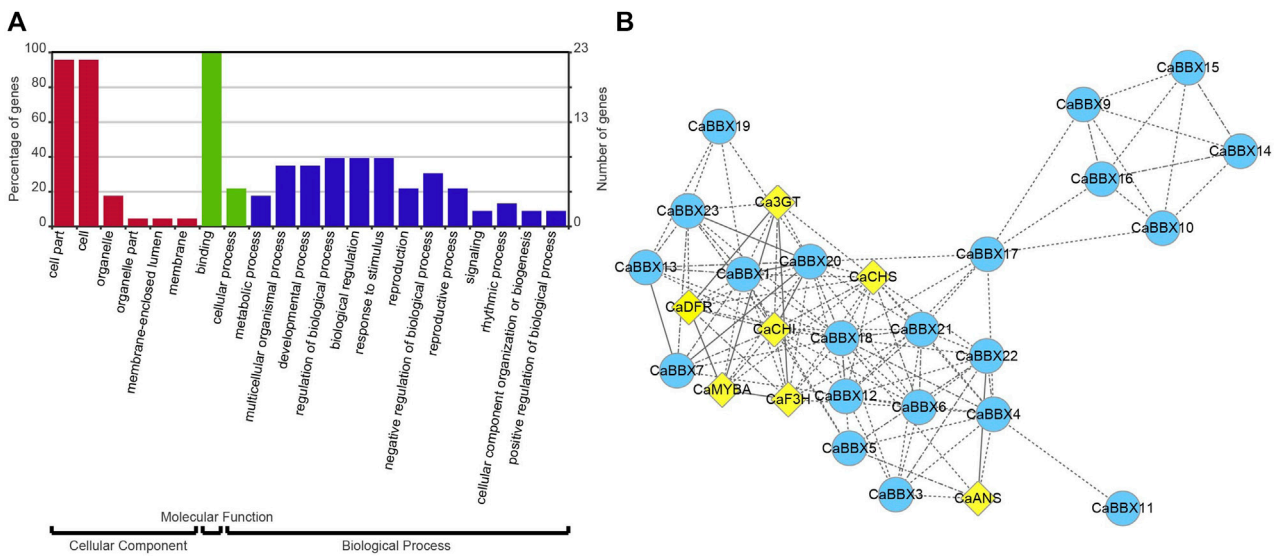

FIGURE 5 | (A) GO enrichment analysis of *CaBBXs*. The red bars represent cellular component, the green bars represent molecular function, and the blue bars represent biological process. **(B)** Co-expression analysis of *CaBBXs* and structural genes of the anthocyanin pathway. The yellow diamonds represent the structural genes of the anthocyanin pathway and the blue circles represent the *CaBBX* family of genes. The line indicates their correlation; the more solid is the line, the higher is the correlation.

cultivated pepper "6421". Actin gene (GenBank Accession No. Contig0023*+) was the endogenous control. Expressions of CaBBX3, CaBBX4, CaBBX5, CaBBX7, CaBBX8, and CaBBX13 were analyzed, which suggested that the qRT-PCR and RNA-seq expression level findings for CaBBXs were similar. Among them, CaBBX3, CaBBX4, and CaBBX8 were highly expressed in fruits but showed lower expression in flowers. Conversely, CaBBX7, and CaBBX13 were highly expressed in the flowers but showed

low expression in fruits. CaBBX5 was highly expressed in both flowers and fruits. The CaBBX8, and CaBBX13 were also expressed in other tissues such as leaves and stems. Moreover, expressions of four anthocyanin biosynthesis structural genes (CaF3H, CaANS, CaDFR, and CaDFR) and two typical regulation genes (CaAN1 and CaAN2) were also analyzed in different tissues. The related expressions of CaDFR, CaF3H, CaAN1, and CaAN2 were highly in flowers, while the CaCHI, and CaANS were highly expressed in leaves. The related expressions of CaANS, CaDFR, CaAN2 were also expressed in other tissues such as roots, fruits, and stems (**Figure 6**). This indicated that CaBBXs may exert important regulatory effects in the anthocyanin pathway during the stages of flowering, pigment accumulation in leaves and stems, and fruit development in pepper.

DISCUSSION

With the completion of the two germplasm reference genomes, "Zunla" and "CM334", of pepper (*Capsicum* spp.) in 2014, many families of transcription factors have been identified (Kim et al., 2014; Qin et al., 2014). For example, the WRKY transcription factors (Cheng et al., 2020), HD-ZIP gene family (Zhang et al., 2021), and MYB, one of the largest gene families, have all been characterized by genome-wide studies (Wang et al., 2020b). The B-box (BBX) transcription factors, a subfamily of the zinc finger family of proteins, have several biological functions, which have attracted research attention in recent years (Gangappa and Botto, 2014). In this study, a total of 23 *CaBBX*s were identified in pepper, along with the identification of one more member, *CaBBX18,* in the cultivated species, "Zunla", absent in the "CM334" reference genome. The *CaBBXs* were unevenly distributed across the 12 chromosomes on the pepper genome and had fewer members relative to *Arabidopsis* (32 *ATBBX*s) (Khanna et al., 2009). Among the Solanaceae plants, the tomato has a much smaller genome than pepper, however, 29 *SlBBXs* have been identified, to date (Chu et al., 2016). Potato contains 30 *StBBX* transcription factors owing to its highly heterozygous tetraploid genome (Talar et al., 2017). Species-specific duplication or deletion during evolution may have led to differences in the numbers of the BBX family of transcription factors across various species.

In pepper, the 23 members were divided into four categories according to the number and types of structural domains, including ten genes with two B-box domains, eight genes with two B-box domains and one CCT domain, four genes with only one B-box domain, and *CaBBX16* gene with one B-box and one CCT domains (**Figure 2**). The B-box conserved motifs play key roles in protein interaction and transcriptional regulation, while the CCT domain is mainly involved in nuclear transport and transcriptional regulation (Wang et al., 2021). In addition, the exon-intron structures also affected the functions of genes. The number of exons in the *CaBBX* family varied from two to five, and 4.3% of genes in the entire family had no introns. The phylogenetic tree constructed by cluster analysis divided the *CaBBXs* into four clades. Due to the highly conserved exon-intron characteristics, *CaBBXs* in the same clades may have similar functions. Another reason for functional diversity was due to the different *cis*-regulatory elements in the promoter regions (Haberer et al., 2004). A large number of *cis*-elements in the promoter region were related to light responsiveness, including Box 4, G-Box, TCT-motif, and GT1-motif. Many BBX proteins play a key role in light signaling pathways, for instance, *AtBBX20* is a positive regulator of light signaling transduction (Datta et al., 2007). Moreover, ABRE, TC-rich repeats, TCA-element, and ARE involved in responsiveness to abscisic acid or are *cis*-regulatory elements essential for the anaerobic induction. The hormone-related *cis*-regulatory elements, including ERE, CGTCA-motif, TGACG-motif, and GARE-motif were also presented in the *CaBBX* promoter region. Diversified domains, structures, and *cis*-regulatory elements in the promoter region may lead to diverse functions of the *CaBBX* family of transcription factors.

Gene duplication plays a key role in the expansion and functional diversity of gene families (Wang et al., 2020a). A total of three segmental duplicated pairs were found in the inter-species of pepper, a major contributor to gene expansion in the *CaBBX* family. Similar results have been reported in the grapevine *VviBBX* family[29], tomato *SlBBX* family (Chu et al., 2016), and the pear *PbBBX* family (Cao et al., 2017). No tandem duplications were found. *CaBBX1-CaBBX7* and *CaBBX2-CaBBX6* were two duplicated gene pairs, and *CaBBX1-CaBBX2* were mutual duplicates, in particular. This indicated that *CaBBX1* and *CaBBX2* were duplicated before *CaBBX6* and *CaBBX7*, and *CaBBX6* and *CaBBX7* were in close physical proximity on chromosomes in the genome. The selection pressures were estimated using the Ka/Ks value, and all the three duplicated *CaBBX* pairs had undergone purifying selection (Ka/Ks < 1.0). In addition, these four genes contained two conserved B-box domains, however, their gene structures were different. Both *CaBBX1* and *CaBBX7* contained five CDS regions, while *CaBBX2* had three introns, and *CaBBX6* had only two exons. Although these were duplicated genes, which may arise from the same ancestor, the functions could have diverged due to structural differentiation. Bioinformatic prediction of subcellular localization showed that the *CaBBX1*, *CaBBX2,* and *CaBBX6* genes were localized in the nucleus, while *CaBBX7* had extracellular localization. The localization of *CaBBX7* may support its ability to transport across membranes, however, the specific functions of the *CaBBXs* need further experimental proof. Simultaneously, nuclear localization of the *BBX* family of genes has been reported in other species (Feng et al., 2021). The homologous genes of the duplicated genes *CaBBX1, CaBBX2, CaBBX6,* and *CaBBX7,* belonging to the *CaBBX gene* family are involved in various signaling pathways and development processes. However, differential tissue-specific expression patterns are observed between duplicated gene pairs (Crocco and Botto, 2013). These four duplicated genes were in Clade II in the multi-species phylogenetic tree, which represented the characteristics of the *CaBBX gene* family. The RNA-seq data indicated that *CaBBX1*, *CaBBX6,* and *CaBBX7* were specifically expressed in

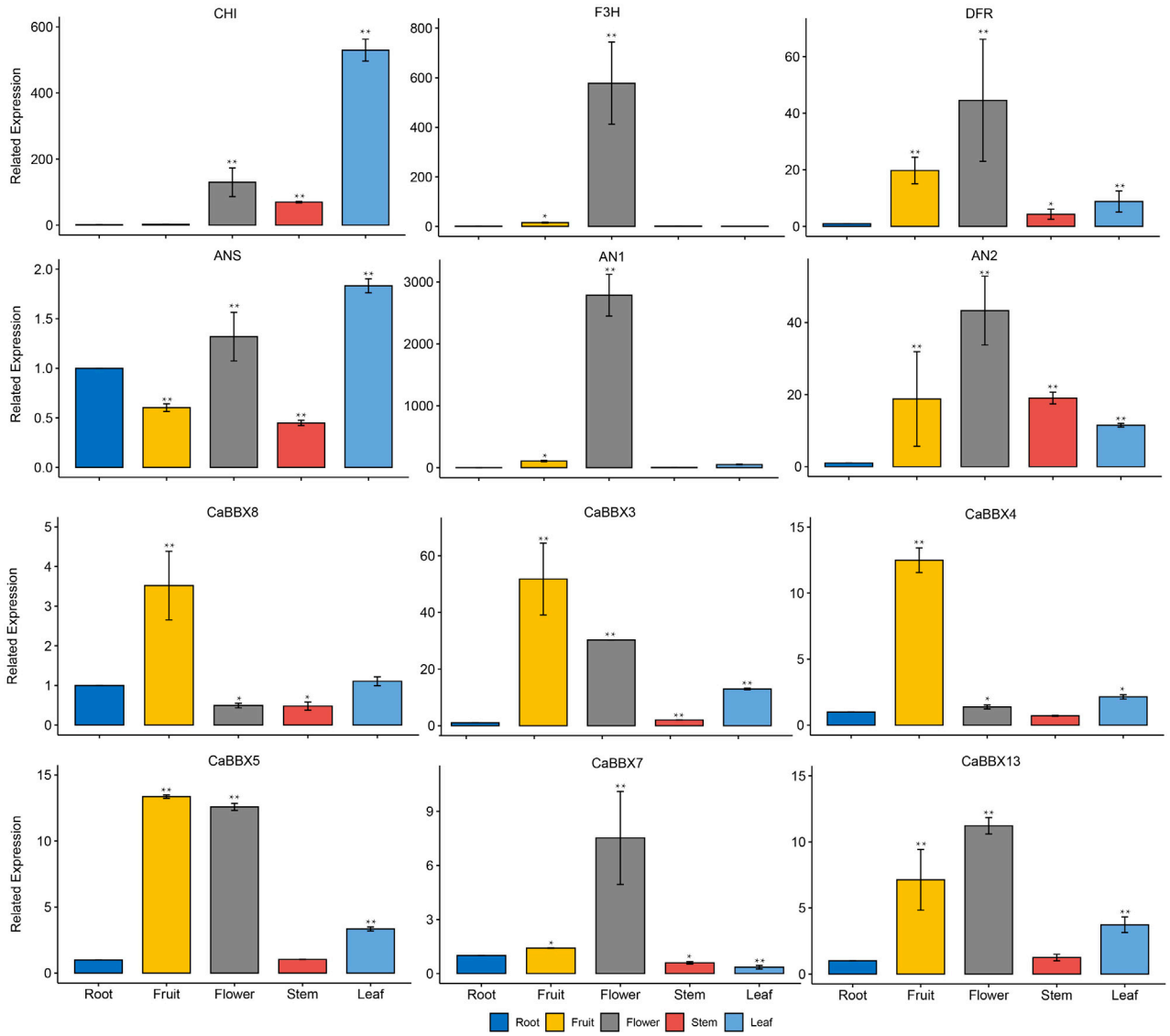

FIGURE 6 | qRT-PCR analysis of *CaBBXs* and the structural genes of anthocyanin biosynthesis pathway. Mean values and standard deviations (SDs) are indicated by error bars. Significant differences were determined by *t*-test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

flowers, while *CaBBX2* was specifically expressed in pepper seeds. Although there was lesser gene duplication in the pepper BBX family, the duplication had no direct effects on the functions of the genes. Through evolution, BBX duplicated genes have gained extensive differences in their structures, motifs, and promoter *cis*-acting elements, which directly contribute to the divergence in the expressions of duplicated genes.

Several studies show that BBX genes are widely involved in the growth and development of plants. For example, the B-box protein, STH2, in *Arabidopsis* is a positive regulator of photomorphogenesis and plays a direct role in transcriptional activation (Datta et al., 2007). Silencing *PpBBX16* reduces the accumulation of anthocyanins and participates in the regulation of anthocyanin accumulation in pear (*Pyrus pyrifolia*) (Bai et al., 2019). *ATBBX18* weakens the tolerance to high temperature in *Arabidopsis* (Wang et al., 2013), and *MdBBX10* enhances tolerance to ABA-mediated responses and ROS levels in apple plants (*Malus domestica* Borkh.) (Liu et al., 2019). However, to date, the specific roles of the BBX family of genes in pepper remain unclear. In this study, RNA-seq data from pepper seeds, flowers, leaves, fruits, and placenta at different stages were used to determine the expression profiles of the *CaBBX* family of genes. All 23 *CaBBXs* were involved in the regulation of fruit development and maturation in pepper, and different genes showed varying tissue-specific expression patterns. Among them, *CaBBX14*, *CaBBX15*, *CaBBX16*, *CaBBX17*, and *CaBBX9* were specifically expressed in the leaves and may be involved in the regulation of light signaling transduction and other pathways. *CaBBX6*, *CaBBX18*, *CaBBX20*, *CaBBX7*, *CaBBX21*, *CaBBX23*, *CaBBX1*, and *CaBBX13* were highly expressed in flowers.

SlBBX20, the tomato homologous gene of *CaBBX6 is highly* expressed in flowers and is a key regulator of carotenoid biosynthesis (Xiong et al., 2019). The homologous gene of *CaBBX1* and *CaBBX7, the ATBBX19* transcriptional regulator in *Arabidopsis* containing B-box domains, is involved in photomorphogenesis and flowering (Wang et al., 2014). In addition, *CaBBX2* and *CaBBX8* have a seed-specific expression in different organs, and the homologous gene of *CaBBX8, ABBX3,* produces a photoperiodic flowering switch in *Brassicaceae* (Simon et al., 2015). COP1-mediated degradation of the *ATBBX22* gene optimizes seedling development in *Arabidopsis* (Chang et al., 2011). The remaining *CaBBXs* were widely expressed specifically in the placenta or fruits of pepper. *CaBBX10* was only expressed in fruit. Its homologous gene, *ABBX21* of *Arabidopsis,* can interact with COP1 to regulate shade avoidance and long hypocotyl formation (Crocco et al., 2010). The *CaBBX11* and *CaBBX19* were under-expressed in leaves but over-expressed in other organs. While *ABBX28*, the homologous gene of *CaBBX28* encodes an atypical B-box domain protein that negatively regulates photomorphogenic development by interfering with the binding of the transcription factor, HY5, to its target gene promoters (Liu et al., 2020c). In this study, GO enrichment analysis indicated that *CaBBX*s of pepper were widely expressed and enriched spanning across all the stages of fruit development, and their functions were closely related to the differences in their domains, structures, promoters, and evolutionary processes.

In particular, eight family members were specifically expressed in flowers, and eight were specifically expressed in fruits, accounting for 69.56% of the total expression (**Figure 4**). Specific expression in different organs was considered a significant regulatory mechanism of flavonoid synthesis, growth and development, and fertility stages. In the co-expression network, the *CaBBXs* and anthocyanin structural genes were significantly associated. Seven *CaBBX* genes were closely co-expressed with anthocyanin structural genes (**Figure 5**). It was speculated that the *CaBBX* transcription factors in pepper play regulatory roles in the processes of anthocyanins biosynthesis in pepper. In other species, some BBXs were involved in regulating the synthesis of anthocyanins, such as the *BBX33/CONSTANS-like 11* (*COL11*) gene which controls the expression of *MYB10* in apple fruit, thereby affecting the synthesis of anthocyanins and regulating the color of the red skin (Plunkett et al., 2019). The overexpression of *BBX23* in poplars (*Populus trichocarpa*) activates the expression of MYB transcription factors and structural genes in the flavonoid pathway, thereby promoting the accumulation of procyanidins and anthocyanins (Li et al., 2021). In this study, the *CaBBX* gene family was systematically studied by both RNA-seq and qRT-PCR. *CaBBX3*, *CaBBX4*, and *CaBBX5* were identified as three candidate anthocyanins regulatory genes in fruits, while the *CaBBX6*, *CaBBX7*, and *CaBBX13* were three candidate anthocyanins regulatory genes in flowers. However, the specific regulatory mechanism and function of these candidate genes need to be verified by further molecular biology experiments. This study provided a strong reference for screening the transcription factors involved in the regulation of anthocyanin synthesis in pepper and laid a solid foundation for the further functional characterization of the *CaBBXs*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

XZ, FL, and JW designed the research. JW, GY, YC, YD, QY, and QS conducted the experiments. JW, LP, and LD analyzed the data. JW wrote the manuscript. FL, CX, and JW revised the manuscript and improved the English. XZ, CX, and FL acquired the funding. All authors have read, reviewed and approved the submitted version.

FUNDING

This work was supported by National Natural Science Foundation of China (U19A2028).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.847328/full#supplementary-material

REFERENCES

Alappat, B., and Alappat, J. (2020). Anthocyanin Pigments: beyond Aesthetics. *Molecules* 25, 5500–125629. doi:10.3390/molecules25235500

An, J.-P., Wang, X.-F., Espley, R. V., Lin-Wang, K., Bi, S.-Q., You, C.-X., et al. (2019). An Apple B-Box Protein MdBBX37 Modulates Anthocyanin Biosynthesis and Hypocotyl Elongation Synergistically with MdMYBs and MdHY5. *Plant Cel Physiol* 61, 130–143. doi:10.1093/pcp/pcz185

Bai, S., Tao, R., Tang, Y., Yin, L., Ma, Y., Ni, J., et al. (2019). BBX16, a B-box Protein, Positively Regulates Light-induced Anthocyanin Accumulation by activatingMYB10in Red Pear. *Plant Biotechnol. J.* 17, 1985–1997. doi:10.1111/pbi.13114

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Brown, J., Pirrung, M., and McCue, L. A. (2017). FQC Dashboard: Integrates FastQC Results into a Web-Based, Interactive, and Extensible FASTQ Quality Control Tool. *Bioinformatics* 33, 3137–3139. doi:10.1093/bioinformatics/btx373

Cao, Y., Han, Y., Meng, D., Li, D., Jiao, C., Jin, Q., et al. (2017). B-BOX Genes: Genome-wide Identification, Evolution and Their Contribution to Pollen Growth in Pear (Pyrus Bretschneideri Rehd.). *BMC Plant Biol.* 17, 156. doi:10.1186/s12870-017-1105-4

Chang, C.-S. J., Maloof, J. N., and Wu, S.-H. (2011). COP1-Mediated Degradation of BBX22/LZF1 Optimizes Seedling Development in Arabidopsis. *Plant Physiol.* 156, 228–239. doi:10.1104/pp.111.175042

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* 13, 1194–1202. doi:10.1016/j.molp.2020.06.009

Cheng, W., Jiang, Y., Peng, J., Guo, J., Lin, M., Jin, C., et al. (2020). The Transcriptional Reprograming and Functional Identification of WRKY Family Members in Pepper's Response to *Phytophthora Capsici* Infection. *BMC Plant Biol.* 20, 256. doi:10.1186/s12870-020-02464-7

Chu, Z., Wang, X., Li, Y., Yu, H., Li, J., Lu, Y., et al. (2016). Genomic Organization, Phylogenetic and Expression Analysis of the B-BOX Gene Family in Tomato. *Front. Plant Sci.* 7, 1552. doi:10.3389/fpls.2016.01552

Crocco, C. D., and Botto, J. F. (2013). BBX Proteins in green Plants: Insights into Their Evolution, Structure, Feature and Functional Diversification. *Gene* 531, 44–52. doi:10.1016/j.gene.2013.08.037

Crocco, C. D., Holm, M., Yanovsky, M. J., and Botto, J. F. (2010). AtBBX21 and COP1 Genetically Interact in the Regulation of Shade Avoidance. *Plant J.* 64, 551–562. doi:10.1111/j.1365-313X.2010.04360.x

Datta, S., Hettiarachchi, C., Johansson, H., and Holm, M. (2007). SALT TOLERANCE HOMOLOG2, a B-Box Protein in Arabidopsis that Activates Transcription and Positively Regulates Light-Mediated Development. *Plant Cell* 19, 3242–3255. doi:10.1105/tpc.107.054791

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam Protein Families Database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi:10.1093/nar/gky995

Fang, J. (2015). Classification of Fruits Based on Anthocyanin Types and Relevance to Their Health Effects. *Nutrition* 31, 1301–1306. doi:10.1016/j.nut.2015.04.015

Feng, Z., Li, M., Li, Y., Yang, X., Wei, H., Fu, X., et al. (2021). Comprehensive Identification and Expression Analysis of B-Box Genes in Cotton. *BMC Genomics* 22, 439. doi:10.1186/s12864-021-07770-4

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* 39, W29–W37. doi:10.1093/nar/gkr367

Gangappa, S. N., and Botto, J. F. (2014). The BBX Family of Plant Transcription Factors. *Trends Plant Sci.* 19, 460–470. doi:10.1016/j.tplants.2014.01.010

Gangappa, S. N., and Botto, J. F. (2016). The Multifaceted Roles of HY5 in Plant Growth and Development. *Mol. Plant* 9, 1353–1365. doi:10.1016/j.molp.2016.07.002

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: The Proteomics Server for In-Depth Protein Knowledge and Analysis. *Nucleic Acids Res.* 31, 3784–3788. doi:10.1093/nar/gkg563

Haberer, G., Hindemitt, T., Meyers, B. C., and Mayer, K. F. X. (2004). Transcriptional Similarities, Dissimilarities, and Conservation of Cis-Elements in Duplicated Genes of Arabidopsis. *Plant Physiol.* 136, 3009–3022. doi:10.1104/pp.104.046466

Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an Upgraded Gene Feature Visualization Server. *Bioinformatics* 31, 1296–1297. doi:10.1093/bioinformatics/btu817

Huang, J., Zhao, X., Weng, X., Wang, L., and Xie, W. (2012). The rice B-Box Zinc finger Gene Family: Genomic Identification, Characterization, Expression Profiling and Diurnal Analysis. *Plos ONE* 7, e48242. doi:10.1371/journal.pone.0048242

Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central Hub for Transcription Factors and Regulatory Interactions in Plants. *Nucleic Acids Res.* 45, D1040–D1045. doi:10.1093/nar/gkw982

Jung, S., Venkatesh, J., Kang, M.-Y., Kwon, J.-K., and Kang, B.-C. (2019). A Non-LTR Retrotransposon Activates Anthocyanin Biosynthesis by Regulating a MYB Transcription Factor in Capsicum Annuum. *Plant Sci.* 287, 110181. doi:10.1016/j.plantsci.2019.110181

Khanna, R., Kronmiller, B., Maszle, D. R., Coupland, G., Holm, M., Mizuno, T., et al. (2009). The Arabidopsis B-Box Zinc finger Family. *Plant Cell* 21, 3416–3420. doi:10.1105/tpc.109.069088

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. *Ijms* 19, 2190. doi:10.3390/ijms19082190

Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J. M., Lee, H.-A., et al. (2014). Genome Sequence of the Hot Pepper Provides Insights into the Evolution of Pungency in Capsicum Species. *Nat. Genet.* 46, 270–278. doi:10.1038/ng.2877

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An Information Aesthetic for Comparative Genomics. *Genome Res.* 19, 1639–1645. doi:10.1101/gr.092759.109

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a Database of Plant Cis-Acting Regulatory Elements and a portal to Tools for In Silico Analysis of Promoter Sequences. *Nucleic Acids Res.* 30, 325–327. doi:10.1093/nar/30.1.325

Li, C., Pei, J., Yan, X., Cui, X., Tsuruta, M., Liu, Y., et al. (2021). A poplar B-box Protein PtrBBX23 Modulates the Accumulation of Anthocyanins and Proanthocyanidins in Response to High Light. *Plant Cel Environ* 44, 3015–3033. doi:10.1111/pce.14127

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* 30, 923–930. doi:10.1093/bioinformatics/btt656

Liu, F., Yu, H., Deng, Y., Zheng, J., Liu, M., Ou, L., et al. (2017). PepperHub, an Informatics Hub for the Chili Pepper Research Community. *Mol. Plant* 10, 1129–1132. doi:10.1016/j.molp.2017.03.005

Liu, J., Ai, X., Wang, Y., Lu, Q., Li, T., Wu, L., et al. (2020b). Fine Mapping of the *Ca3GT* Gene Controlling Anthocyanin Biosynthesis in Mature Unripe Fruit of *Capsicum Annuum* L. *Theor. Appl. Genet.* 133, 2729–2742. doi:10.1007/s00122-020-03628-7

Liu, X., Li, R., Dai, Y., Yuan, L., Sun, Q., Zhang, S., et al. (2019). A B-Box Zinc finger Protein, MdBBX10, Enhanced Salt and Drought Stresses Tolerance in *Arabidopsis*. *Plant Mol. Biol.* 99, 437–447. doi:10.1007/s11103-019-00828-8

Liu, Y., Lin, G., Yin, C., and Fang, Y. (2020c). B-box Transcription Factor 28 Regulates Flowering by Interacting with Constans. *Sci. Rep.* 10, 17789. doi:10.1038/s41598-020-74445-7

Liu, Y., Lv, J., Liu, Z., Wang, J., Yang, B., Chen, W., et al. (2020a). Integrative Analysis of Metabolome and Transcriptome Reveals the Mechanism of Color Formation in Pepper Fruit (*Capsicum Annuum* L.). *Food Chem.* 306, 125629. doi:10.1016/j.foodchem.2019.125629

Meng, X., Wang, J.-R., Wang, G.-D., Liang, X.-Q., Li, X.-D., and Meng, Q.-W. (2015). An R2R3-MYB Gene, *LeAN2*, Positively Regulated the Thermo-Tolerance in Transgenic Tomato. *J. Plant Physiol.* 175, 1–8. doi:10.1016/j.jplph.2014.09.018

Movahed, N., Pastore, C., Cellini, A., Allegro, G., Valentini, G., Zenoni, S., et al. (2016). The grapevine *VviPrx31* Peroxidase as a Candidate Gene Involved in Anthocyanin Degradation in Ripening Berries under High Temperature. *J. Plant Res.* 129, 513–526. doi:10.1007/s10265-016-0786-3

Plunkett, B. J., Henry-Kirk, R., Friend, A., Diack, R., Helbig, S., Mouhu, K., et al. (2019). Apple B-Box Factors Regulate Light-Responsive Anthocyanin Biosynthesis Genes. *Sci. Rep.* 9, 17762. doi:10.1038/s41598-019-54166-2

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome Sequencing of Cultivated and Wild Peppers Provides Insights into Capsicum Domestication and Specialization. *Proc. Natl. Acad. Sci. U S A.* 111, 5135–5140. doi:10.1073/pnas.1400975111

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucleic Acids Res.* 31, 3381–3385. doi:10.1093/nar/gkg520

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Simon, S., Rühl, M., de Montaigu, A., Wötzel, S., and Coupland, G. (2015). Evolution of CONSTANS Regulation and Function after Gene Duplication Produced a Photoperiodic Flowering Switch in the Brassicaceae. *Mol. Biol. Evol.* 32, 2284–2301. doi:10.1093/molbev/msv110

Su, N., Lu, Y., Lu, Y., Wu, Q., Liu, Y., Xia, Y., et al. (2016). UV-B-induced Anthocyanin Accumulation in Hypocotyls of Radish Sprouts Continues in the Dark after Irradiation. *J. Sci. Food Agric.* 96, 886–892. doi:10.1002/jsfa.7161

Talar, U., Kiełbowicz-Matuk, A., Czarnecka, J., and Rorat, T. (2017). Genome-wide Survey of B-Box Proteins in Potato (Solanum tuberosum)-Identification, Characterization and Expression Patterns during Diurnal Cycle, Etiolation and De-etiolation. *Plos ONE* 12, e0177471. doi:10.1371/journal.pone.0177471

Tang, B., Li, L., Hu, Z., Chen, Y., Tan, T., Jia, Y., et al. (2020). Anthocyanin Accumulation and Transcriptional Regulation of Anthocyanin Biosynthesis in Purple Pepper. *J. Agric. Food Chem.* 68, 12152–12163. doi:10.1021/acs.jafc.0c02460

Varet, H., Brillet-Guéguen, L., Coppée, J.-Y., and Dillies, M.-A. (2016). SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *Plos One* 11, e0157022. doi:10.1371/journal.pone.0157022

Verweij, W., Spelt, C. E., Bliek, M., de Vries, M., Wit, N., Faraco, M., et al. (2016). Functionally Similar WRKY Proteins Regulate Vacuolar Acidification in Petunia and Hair Development in Arabidopsis. *Plant Cell* 28, 786–803. doi:10.1105/tpc.15.00608

Wang, C.-Q., Guthrie, C., Sarmast, M. K., and Dehesh, K. (2014). BBX19 Interacts with CONSTANS to Repress FLOWERING LOCUS T Transcription, Defining a Flowering Time Checkpoint in Arabidopsis. *Plant Cell* 26, 3589–3602. doi:10.1105/tpc.114.130252

Wang, J., Liu, Y., Chen, X., and Kong, Q. (2020a). Characterization and Divergence Analysis of Duplicated R2R3-MYB Genes in Watermelon. *J. Amer. Soc. Hort. Sci.* 145, 281–288. doi:10.21273/JASHS04849-19

Wang, J., Liu, Y., Tang, B., Dai, X., Xie, L., Liu, F., et al. (2020b). Genome-wide Identification and Capsaicinoid Biosynthesis-Related Expression Analysis of the R2R3-MYB Gene Family in *Capsicum Annuum* L. *Front. Genet.* 11, 598183. doi:10.3389/fgene.2020.598183

Wang, Q., Tu, X., Zhang, J., Chen, X., and Rao, L. (2013). Heat Stress-Induced *BBX18* Negatively Regulates the Thermotolerance in *Arabidopsis*. *Mol. Biol. Rep.* 40, 2679–2688. doi:10.1007/s11033-012-2354-9

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40, e49. doi:10.1093/nar/gkr1293

Wen, C.-H., and Chu, F.-H. (2017). A R2R3-MYB Gene *LfMYB113* is Responsible for Autumn Leaf Coloration in Formosan Sweet Gum (Liquidambar formosanaHance). *Plant Cel Physiol* 58, 508–521. doi:10.1093/pcp/pcw228

Xiong, C., Luo, D., Lin, A., Zhang, C., Shan, L., He, P., et al. (2019). A Tomato B-Box protein SlBBX20 Modulates Carotenoid Biosynthesis by Directly activating PHYTOENE SYNTHASE 1, and Is Targeted for 26S Proteasome-Mediated Degradation. *New Phytol.* 221, 279–294. doi:10.1111/nph.15373

Xu, Z.-S., Yang, Q.-Q., Feng, K., and Xiong, A.-S. (2019). Changing Carrot Color: Insertions in *DcMYB7* Alter the Regulation of Anthocyanin Biosynthesis and Modification. *Plant Physiol.* 181, 195–207. doi:10.1104/pp.19.00523

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WEGO: a Web Tool for Plotting GO Annotations. *Nucleic Acids Res.* 34, W293–W297. doi:10.1093/nar/gkl031

Zhang, X., Zhang, L., Ji, M., Wu, Y., Zhang, S., Zhu, Y., et al. (2021). Genome-wide Identification and Expression Analysis of the B-Box Transcription Factor Gene Family in grapevine (*Vitis vinifera* L.). *BMC Genomics* 22, 221. doi:10.1186/s12864-021-07479-4

Zhang, Z., Li, D.-W., Jin, J.-H., Yin, Y.-X., Zhang, H.-X., Chai, W.-G., et al. (2015). VIGS Approach Reveals the Modulation of Anthocyanin Biosynthetic Genes by *CaMYB* in Chili Pepper Leaves. *Front. Plant Sci.* 6, 500. doi:10.3389/fpls.2015.00500

Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G. K.-S., and Yu, J. (2006). KaKs_Calculator: Calculating Ka and Ks through Model Selection and Model Averaging. *Genomics, Proteomics & Bioinformatics* 4, 259–263. doi:10.1016/s1672-0229(07)60007-2

Zhang, Z., Zhu, R., Ji, X., Li, H. J., Lv, H., and Zhang, H. Y. (2021). Genome-wide Characterization and Expression Analysis of the HD-ZIP Gene Family in Response to Salt Stress in Pepper. *Int. J. Genomics* 2021, 1–14. doi:10.1155/2021/8105124

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 07 March 2022
doi: 10.3389/fpls.2022.847789

Newly Developed MAGIC Population Allows Identification of Strong Associations and Candidate Genes for Anthocyanin Pigmentation in Eggplant

Giulio Mangino[1†], Andrea Arrones[1†], Mariola Plazas[2], Torsten Pook[3], Jaime Prohens[1], Pietro Gramazio[2] and Santiago Vilanova[1*]*

[1] Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Valencia, Spain, [2] Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universitat Politècnica de València, Valencia, Spain, [3] Animal Breeding and Genetics Group, Department of Animal Sciences, Center for Integrated Breeding Research, University of Göttingen, Göttingin, Germany

OPEN ACCESS

Edited by:
Nunzio D'Agostino, University of Naples Federico II, Italy

Reviewed by:
Monica Rodriguez, University of Sassari, Italy
Mark Chapman, University of Southampton, United Kingdom

***Correspondence:**
Pietro Gramazio
piegra@upv.es
Santiago Vilanova
sanvina@upvnet.upv.es

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *03 January 2022*
Accepted: *20 January 2022*
Published: *07 March 2022*

Citation:
Mangino G, Arrones A, Plazas M, Pook T, Prohens J, Gramazio P and Vilanova S (2022) Newly Developed MAGIC Population Allows Identification of Strong Associations and Candidate Genes for Anthocyanin Pigmentation in Eggplant. Front. Plant Sci. 13:847789. doi: 10.3389/fpls.2022.847789

Multi-parent advanced generation inter-cross (MAGIC) populations facilitate the genetic dissection of complex quantitative traits in plants and are valuable breeding materials. We report the development of the first eggplant MAGIC population (S3 Magic EGGplant InCanum, S3MEGGIC; 8-way), constituted by the 420 S3 individuals developed from the intercrossing of seven cultivated eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) parents. The S3MEGGIC recombinant population was genotyped with the eggplant 5k probes SPET platform and phenotyped for anthocyanin presence in vegetative plant tissues (PA) and fruit epidermis (FA), and for the light-insensitive anthocyanic pigmentation under the calyx (PUC). The 7,724 filtered high-confidence single-nucleotide polymorphisms (SNPs) confirmed a low residual heterozygosity (6.87%), a lack of genetic structure in the S3MEGGIC population, and no differentiation among subpopulations carrying a cultivated or wild cytoplasm. Inference of haplotype blocks of the nuclear genome revealed an unbalanced representation of the founder genomes, suggesting a cryptic selection in favour or against specific parental genomes. Genome-wide association study (GWAS) analysis for PA, FA, and PUC detected strong associations with two myeloblastosis (MYB) genes similar to *MYB113* involved in the anthocyanin biosynthesis pathway, and with a *COP1* gene which encodes for a photo-regulatory protein and may be responsible for the PUC trait. Evidence was found of a duplication of an ancestral *MYB113* gene with a translocation from chromosome 10 to chromosome 1 compared with the tomato genome. Parental genotypes for the three genes were in agreement with the identification of the candidate genes performed in the S3MEGGIC population. Our new eggplant MAGIC population is the largest recombinant population in eggplant and is a powerful tool for eggplant genetics and breeding studies.

Keywords: multi-parent advanced generation inter crosses (MAGIC), eggplant (*Solanum melongena* L.), *S. incanum*, anthocyanins, pigmentation under calyx (PUC), genome wide association study (GWAS), candidate genes, SPET (single primer enrichment technology)

INTRODUCTION

Multi-parent experimental populations are of great interest for the genetic dissection of quantitative traits and for the development of new recombinant materials for plant breeding (Huang et al., 2015). Despite their complex management and resources requirement, multi-parent advanced generation intercross (MAGIC) populations represent powerful next-generation mapping tools by combining a high genetic diversity and recombination with a low population structure (Arrones et al., 2020; Scott et al., 2020). MAGIC populations are already available in model species, such as *Arabidopsis thaliana* and in several crops, such as cereals, pulses, and vegetables (Kover et al., 2009; Bandillo et al., 2013; Pascual et al., 2015; Huynh et al., 2018), and have demonstrated their power to dissect the structure of complex traits (Dell'Acqua et al., 2015; Stadlmeier et al., 2018).

Although the available MAGIC populations have become a useful resource for genetic studies and breeding, most of them have only exploited intraspecific variation. The incorporation of crop wild relatives (CWRs) as founders could be a way of including multiple wild genomic fragments or introgressions into cultivated background genomes (Arrones et al., 2020). Apart from being of great interest for genetic analysis, interspecific MAGIC populations can be useful for broadening the genetic base of crops and provide new variation for breeding multiple traits, including those related to adaption to climate change (Gramazio et al., 2020a). However, so far, the potential of interspecific MAGIC populations for plant breeding has largely remained unexploited (Arrones et al., 2020).

Eggplant (*Solanum melongena* L.) is a major vegetable crop of increasing importance, ranking fifth in global production among vegetables (FAOSTAT, 2019). Despite its economic importance, eggplant has lagged behind other major crops and little effort has been made to develop immortal experimental populations and genetic and genomic tools (Gramazio et al., 2018, 2019). So far, only one population of recombinant inbred lines (RIL) and one set of introgression lines (ILs) are publicly available (Lebeau et al., 2013; Gramazio et al., 2017a), while no multiparent population has been developed, so far. Conversely, in other related Solanaceae crops, such as tomato, several experimental populations have been developed, including MAGIC populations, which have allowed great advances in the genetic dissection of traits of interest (Pascual et al., 2015; Campanelli et al., 2019). For this reason, the development of this type of population would represent a landmark in eggplant breeding.

Anthocyanins are responsible for a set of specific and relevant traits in eggplant which can be used as a model plant for other crops (Moglia et al., 2020). Anthocyanins play a key role in plant defence mechanisms, and their synthesis and accumulation may vary in response to specific biotic and abiotic stresses (D'Amelia et al., 2018; Lv et al., 2019; Zhou et al., 2019). In addition, anthocyanins may prevent and alleviate human chronic diseases and provide health benefits (Toppino et al., 2020). In eggplant, anthocyanins are responsible for the purple colour of the peel, one of the traits of greatest interest for eggplant breeding (Daunay and Hazra, 2012). Purple coloured eggplant fruits are the most demanded in various markets (Li et al., 2018), and developing dark purple-coloured eggplants, which results from the combination of anthocyanins with chlorophylls, is a major objective in eggplant breeding programmes. Eggplants are variable for the presence of anthocyanins not only in fruits, but also in other plant organs and tissues such as hypocotyl, stem, leaves, leaf veins, prickles, flower calyx, or corolla (Toppino et al., 2020). Anthocyanin biosynthesis has been widely studied in Solanaceae species (Van Eck et al., 1994; Borovsky et al., 2004; Gonzali et al., 2009), but its genetic control in eggplant has not been fully clarified. The anthocyanin biosynthetic pathway is a very conserved network in many plant species, where many enzymes and regulatory transcription factors (TFs) are involved (Albert et al., 2014).

The major anthocyanins in the eggplant fruit epidermis are delphinidin-3-p-coumaroylrutinoside-5-glucoside and delphinidin-3-rutinoside (Mennella et al., 2012; Moglia et al., 2020). Fruit anthocyanins (FA) are highly dependent on light (Jiang et al., 2016), however, the genotypes that carry the pigmentation under the calyx (PUC) mutation are able to independently synthetise the anthocyanins in the fruit epidermis of the incidence of light (Tigchelaar et al., 1968). Quantitative trait locus (QTL)-related studies using family based or genome-wide association (GWA) mapping approaches evidenced that chromosome 10 harbours most of the QTL/genes involved in anthocyanin formation, distribution, and accumulation (Doganlar et al., 2002; Barchi et al., 2012; Cericola et al., 2014; Frary et al., 2014; Toppino et al., 2016, 2020; Wei et al., 2020). The availability of high-quality eggplant genome sequences and transcriptomic data allowed the identification of putative candidate genes belonging to the myeloblastosis (MYB) family which controls the variation in anthocyanin content and fruit colour in eggplant (Docimo et al., 2016; Li et al., 2017, 2018, 2021; Xiao et al., 2018; Moglia et al., 2020; Shi et al., 2021), highlighting their synteny with other Solanaceae. However, the genes underlying the PUC phenotype have not been identified, so far.

Here, we report on the first eggplant MAGIC population derived from an interspecific cross of seven accessions of *S. melongena* and its wild relative *S. incanum* (Gramazio et al., 2019). It represents the largest experimental population described, so far, in eggplant, with a similar population size to MAGIC populations in other solanaceous crops. The population has been genotyped by applying the Single Primer Enrichment Technology (SPET) to explore its genetic architecture and the contribution of founders to the final population, and phenotyped for the presence of anthocyanins in the fruit epidermis and other plant organs and for the PUC trait. These traits were chosen due to their physiological, agronomic, and morphological relevance, high stability, and heritability. An association analysis has been performed to locate the genomic regions and to identify the candidate genes involved in the traits under study.

MATERIALS AND METHODS

Multi-Parent Advanced Generation Inter-Cross Population Construction

The eggplant MAGIC population has been developed by intermating seven cultivated eggplants, i.e., MM1597 (A), DH ECAVI (B), AN-S-26 (D), H15 (E), A0416 (F), IVIA-371 (G) and ASI-S-1 (H), and the *S. incanum* accession MM577 (C) (**Figure 1A**). The wild relative founder was chosen for its tolerance to some biotic and abiotic stresses, mainly drought (Knapp et al., 2013), and for showing a high phenolic content (Prohens et al., 2013). The performance of the founders was comprehensively characterised in previous morphoagronomic and genetic diversity studies (Hurtado et al., 2014; Gramazio et al., 2017b; Kaushik et al., 2018). In addition, their genomes had been resequenced (Gramazio et al., 2019). The latter study highlighted that, in the founder parents, the residual heterozygosity was less than 0.06%.

In order to develop the eggplant S3 Magic EGGplant InCanum (S3MEGGIC) population, founder lines had been inter-crossed by following a simple "funnel" approach (Wang et al., 2017; Arrones et al., 2020; **Figure 1A**). The eight founders (A–H) had been pairwise inter-crossed to produce two-way or simple F1 hybrids (AB, CD, EF, and GH), which were subsequently inter-crossed in pairs (AB × CD and EF × GH) to obtain two four-way or double hybrids (ABCD and EFGH). In order to achieve a complete admixture of all founder genomes and to avoid assortative mating, the double hybrids were intercrossed following a chain pollination scheme, with each individual being used as female and male parents (Díez et al., 2002; **Supplementary Figure 1A**).

All the obtained eight-way or quadruple hybrids (S0 generation) presented all the eight randomly shuffled genomes and only differed on the cytoplasm inherited from the maternal parent. The S0 progenies obtained using the double hybrid ABCD, as a female parent carried the cytoplasm of the wild *S. incanum* MM577, while those derived using the double hybrid EFGH as a female parent carried the cytoplasm of the cultivated *S. melongena* ASI-S-1. Subsequently, the S0 progenies were selfed for three generations by a single seed descent (SSD) to obtain the S3 segregating individuals that were phenotyped and genotyped in this study. To ensure the continuity of the S0 progenies and to accelerate the self-fertilisation process, four plants of each S0 progeny were germinated, and only the first two that set a viable seed were selected for the next generation (S1; **Supplementary Figure 1B**). From each of the two S0 selected plants, two S1 plants were germinated, and only the first setting fruit was selected for the S2 generation. The same was done for the S3 generation so that for each progeny, two plants were germinated and phenotyped. Despite this, only one was used for originating the next generation. On the other hand, in the S3 progenies, if two individuals displayed some phenotypic differences, both of them were included in the S3MEGGIC population.

Cultivation Conditions

Seeds were germinated in Petri dishes, following the protocol developed by Ranil et al. (2015) and, subsequently, transferred to seedling trays in a climatic chamber under a photoperiod and temperature regime of 16 h light (25°C) and 8 h dark (18°C). After acclimatisation, plantlets were transplanted to 15 L pots and grown in a pollinator-free benched glasshouse of the Universitat Politècnica de València (UPV), Valencia, Spain (GPS coordinates: latitude, 39° 28′ 55″ N; longitude, 0° 20′ 11″ W; 7 m above sea level). Plants were spaced 1.2 m between rows and 1 m within the row, and fertirrigated using a drip irrigation system and trained with vertical strings. Pruning was done manually to regulate the vegetative growth and flowering. Phytosanitary treatments were performed when necessary. In order to shorten generation time of subsequent generations (S0–S3), plantlets were transplanted to individual thermoformed pots (1.3 L capacity) in a pollinator-free glasshouse, and selfings were stimulated by a mechanical vibration.

High-Throughput Genotyping

Young leaf tissue was sampled from 420 S3 individuals, the eight founders, and the four two-way hybrids. Genomic DNA was extracted using the silica matrix extraction (SILEX) extraction method (Vilanova et al., 2020) and checked for quality and integrity by agarose electrophoresis and NanoDrop ratios (260/280 and 260/230), while its concentration was estimated with Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States). After dilution, the samples were sent to Identity Governance and Administration (IGA) Technology Services (IGATech, Udine, Italy) for library preparation and sequencing with NextSeq500 sequencer (150 paired-end) for a high-throughput genotyping using the SPET technology using the 5k probes eggplant SPET platform (Barchi et al., 2019a). The latter comprises 5,093 probes, and was developed by filtering out the most informative and reliable polymorphisms (3,372 of them in coding sequences or CDS and 1,721 in introns and untranslated regions or UTR regions) from the set of over 12 million single-nucleotide polymorphisms (SNPs) identified among the MAGIC founders (Gramazio et al., 2019).

Raw reads were demultiplexed and the adapters were removed using a standard Illumina pipeline and Cutadapt (Martin, 2011), while trimming was performed by ERNE (Del Fabbro et al., 2013). Clean reads were mapped onto the eggplant reference genome "67/3" (Barchi et al., 2019b) using BWA-MEM (Li, 2013) with default parameters, and only the uniquely aligned reads were selected for the variant calling performed with GATK 4.0 (DePristo et al., 2011), following the best practice recommended by the Broad Institute.[1]

The SNPs identified by SPET were filtered using the Trait Analysis by Association, Evolution, and Linkage (TASSEL) software (ver. 5.0, Bradbury et al., 2007) in order to retain the most reliable ones (minor allele frequency > 0.01, missing data < 10% and maximum marker heterozygosity < 70%). In addition, a linkage disequilibrium (LD) k-nearest neighbour

[1] http://www.broadinstitute.org

genotype imputation method (LD KNNi) was performed to fill the missing calls or genotyping gaps.

Population Structure, Heterozygosity, and Haplotype Blocks Inferring

A principal component analysis (PCA) was performed to assess the population structure of S3MEGGIC using the R package vcfR (Knaus and Grünwald, 2017) and the function glPCA of the Adegenet package (Jombart, 2008). Finally, the PCA was graphically plotted with ggplot2 (Wickham, 2016). An Analysis of Molecular Variance (AMOVA) was performed to estimate the population differentiation according to the cytoplasm (cultivated vs. wild) of the individuals of the S3MEGGIC population by using the function poppr.amova of the poppr R package (Kamvar et al., 2014). The residual heterozygosity and its distribution were evaluated with TASSEL software (ver. 5.0, Bradbury et al., 2007). Parental contribution to S3MEGGIC individuals and haplotype blocks were estimated by using R-package HaploBlocker (Pook et al., 2019).

Phenotyping and Genome-Wide Association Study

Phenotypic data were collected from the 420 S3 individuals grown during the 2019/2020 season. Three traits were screened using a binary classification (presence/absence), namely, anthocyanins presence in vegetative plant tissues (PA) and fruit epidermis (FA), and anthocyanic PUC (**Supplementary Figure 4**). The presence of PA was phenotyped when the purple coloration was observed in any vegetative plant parts such as in stem, branches, leaf veins, or prickles. For FA, anthocyanins were considered as present when the purple colouration was observed in the fruit epidermis regardless of their distribution (uniform, listed, etc.) or intensity. The PUC trait could only be phenotyped in anthocyanic fruits by removing the calyx and observing the presence of anthocyanins under it. FA and PUC traits were screened at the stage of commercial maturity (e.g., when the fruit was physiologically immature) which is the best stage for phenotyping these traits. Phenotypic characteristics of the eight founders and two-way hybrids are described in **Figure 1A**.

Using the phenotypic and genotypic data collected from the S3MEGGIC individuals, a Genome-Wide Association Study (GWAS) was performed for the selected traits using the TASSEL software (ver. 5.0, Bradbury et al., 2007). For the association study, mixed linear model (MLM) analyses were conducted. The MLM analysis uses both fixed and random effects, which incorporates the kinship among the individuals. The multiple testing was corrected with the Bonferroni and the false discovery rate (FDR) methods (Holm, 1979; Benjamini and Hochberg, 1995) to identify candidate-associated regions at the significance level of 0.05 (Thissen et al., 2002). SNPs with limit of detection (LOD) [−log10(*p*-value)] over these thresholds or cut off values were declared to be significantly associated with the anthocyanin presence. The R qqman package (Turner, 2018) was used to visualise the Manhattan plots and LDBlockShow (Dong et al., 2020) to determine the LD and the plot haplotype block structure. The pattern of pairwise LD between SNPs was measured by LD correlation coefficient (r^2) by considering haplotype blocks with default r^2 values greater than 0.5 supported by the solid spine of LD method (Gabriel et al., 2002; Barrett et al., 2005). The LD was used to narrow down the genomic regions with significant associations. The genes underlying the associated regions were retrieved from the "67/3" eggplant reference genome (ver. 3) (Barchi et al., 2019b). Genes were considered as potential candidates in controlling the assessed traits when carrying homozygous allelic variants classified as "high impact" according to SnpEff software v 4.2 prediction (Cingolani et al., 2012) of the eight MAGIC founders (Gramazio et al., 2019). The Integrative Genomics Viewer (IGV) tool was used for visual exploration of founder genome sequences to validate SnpEff results and to confirm the presence of the so-called "high impact" variants (Robinson et al., 2020). In addition, eggplant and tomato syntenies were assessed by a BLASTx search of candidate genes sequences against the tomato genome (version SL4.0) in the Sol Genomics Network database.[2]

RESULTS

Multi-Parent Advanced Generation Inter-Cross Population Construction

Seven accessions of eggplant and one of the wild relative *S. incanum* were selected as founder parents (A-H) for the construction of the eggplant MAGIC population (S3MEGGIC). Following a funnel breeding scheme (**Figure 1A** and **Supplementary Figure 1**), a total of 420 individuals of the MAGIC populations were obtained. First, founders were pairwise inter-crossed to produce four two-way hybrids (AB, CD, EF, and GH), which were subsequently inter-crossed in pairs to obtain four-way hybrids (ABCD and EFGH). One-hundred and forty-nine individuals of each of the two four-way hybrids were inter-crossed using a chain pollination scheme (**Supplementary Figure 1**). Out of the theoretical maximum of 298 eight-way hybrid progenies (S0), seeds were obtained for 209 of them, of which 116 carried the *S. melongena* ASI-S-1 cytoplasm and 93 carried the *S. incanum* MM577 cytoplasm. Two plants per S0 progeny were used to advance the population, reaching 402 S1 progenies. These S1 progenies were advanced through single seed descend (SSD) to obtain the 391 S2 and 305 S3 MAGIC progenies. The final S3MEGGIC population was constituted by 420 S3 individuals, of which 348 individuals carried the cultivated cytoplasm and 72 the wild cytoplasm.

Single Primer Enrichment Technology Genotyping

The genotyping of the 420 S3 MAGIC individuals, the eight founders, and the four two-way hybrids by the eggplant SPET platform yielded 22,146 SNPs. After filtering, 7,724 high-confidence SNPs were retained for the subsequent analysis, and the low percentage of missing calls (0.53%) was imputed. Filtered SNPs were distributed across the entire

[2]http://www.solgenomics.net

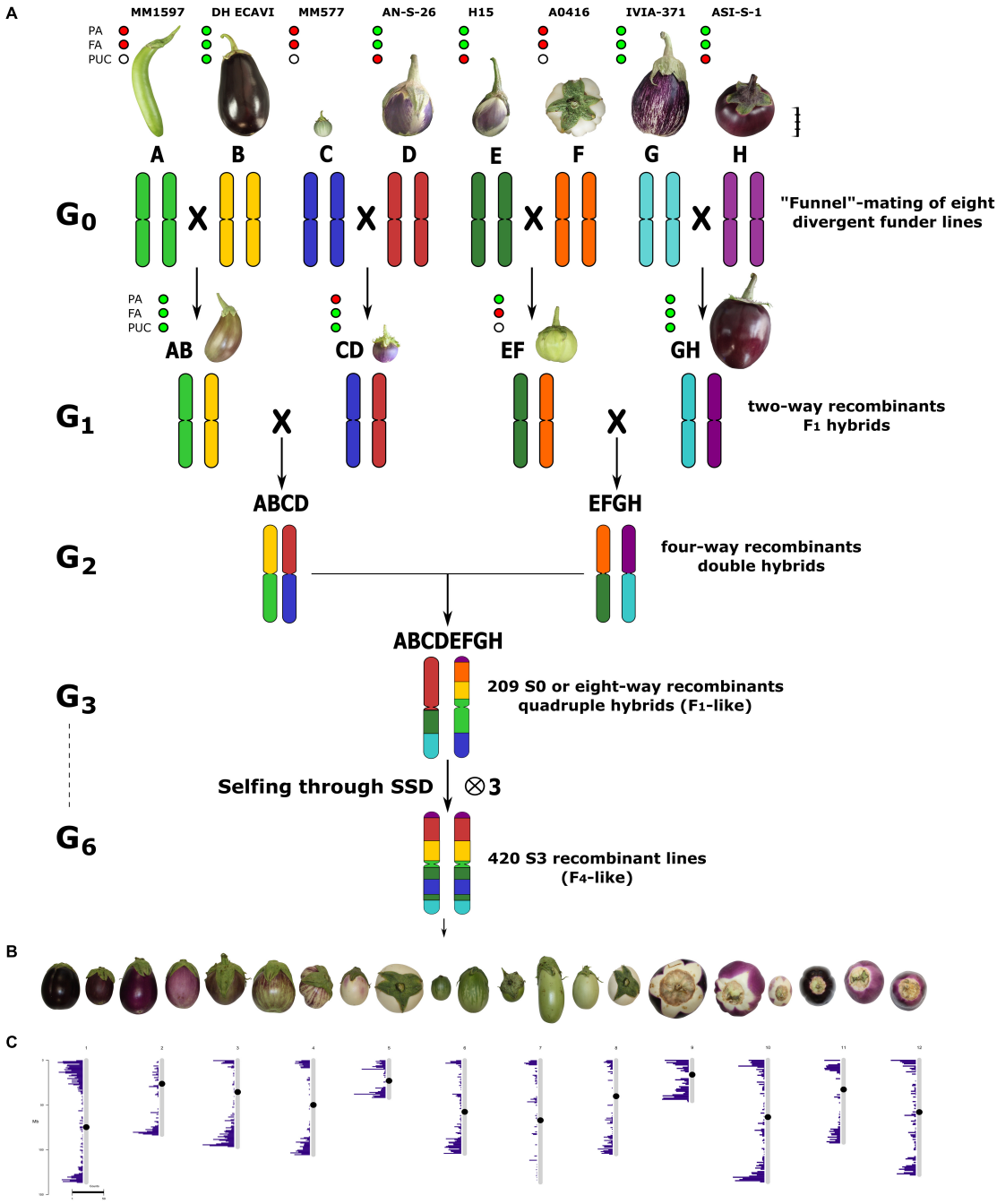

FIGURE 1 | (A) The funnel breeding design, used across the six generations (G1–G6), to develop the 420 S3 individuals of the S3MEGGIC population. The eight parents, coded from A to H and each with a different colour to represent their genomic background, are represented above at a scale based on the real fruit size. Scale bar represents 5 cm. The four two-way hybrids obtained in the G1 generation (AB, CD, EF, and GH) are also represented at the same scale as the founders. Phenotyping of founders and two-way hybrids for absence (red) or presence (green) of anthocyanins in vegetative plant tissues (PA) and fruit epidermis (FA), or anthocyanic pigmentation under the calyx (PUC). White dots for PUC mean uncertainty for non-anthocyanin fruits. **(B)** A representation of the phenotypic diversity of the S3 individuals found during the phenotyping. **(C)** Distribution of molecular markers across the chromosomes used for the genotyping.

eggplant genome, although the distribution of SNPs varied within and among chromosomes (**Table 1** and **Figure 1C**). Chromosome 9 had the highest average marker density after the SNP filtering, with 173.07 SNPs per Mb, while chromosome 7 had the lowest with an average of 23.03 SNPs per Mb. Generally, most of the SNPs were located in regions with a high gene density and decayed around the centromere (**Figure 1C**). The S3 MAGIC individuals exhibited

TABLE 1 | Statistics of the genotyping using the eggplant 5k probes Single Primer Enrichment Technology (SPET) platform of the 420 S3MEGGIC population individuals using the "67/3" eggplant reference genome (Barchi et al., 2019b).

Chr	Markers	Filtered markers	% Markers	% Filtered markers	Chr length (Mb)	Marker density (Mb)	Filtered marker density
1	3,234	1,205	14.60	15.60	13.66	236.82	88.24
2	1,119	424	5.05	5.49	8.34	134.21	50.85
3	2,004	758	9.05	9.81	9.71	206.41	78.07
4	1,567	588	7.08	7.61	10.57	148.23	55.62
5	1,438	571	6.49	7.39	4.39	327.80	130.16
6	2,108	794	9.52	10.28	10.90	193.38	72.84
7	2,483	328	11.21	4.25	14.24	174.34	23.03
8	1,392	550	6.29	7.12	10.96	126.98	50.17
9	1,841	625	8.31	8.09	3.61	509.80	173.07
10	2,069	798	9.34	10.33	10.67	193.86	74.77
11	1,260	453	5.69	5.86	7.23	174.23	62.64
12	1,631	630	7.36	8.16	10.05	162.33	62.70
Total	22,146	7,724	100.00	100.00	114.33		
Average						215.70	76.85

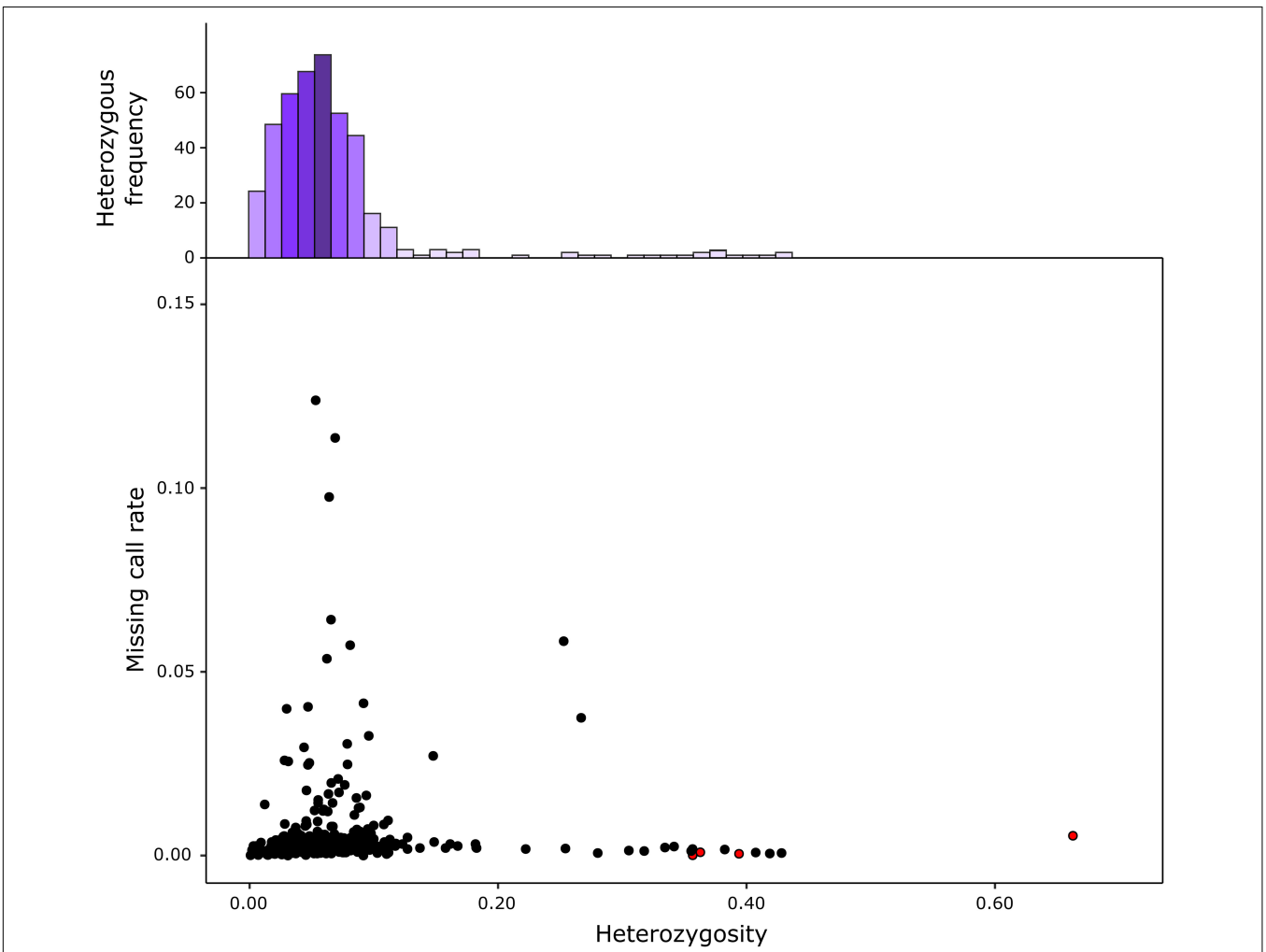

FIGURE 2 | Heterozygosity and proportion of missing data from the S3MEGGIC population. The top histogram represents the observed heterozygosity which is skewed to the left (mean 6.87%; mode 5.02%). The lower graph reports the residual heterozygosity of S3 individuals as black dots copared against two-way hybrids represented as red dots.

a heterozygosity average of 6.87%, with only 15 individuals (3.57%) with a proportion of residual heterozygosity higher than 20% (**Figure 2**).

Population Structure

Population stratification, performed by PCA, indicated the absence of subgroups in the S3 individuals since no clear clustering was observed (**Figure 3**). The first two principal components (PCs) account for 5.15% (PC1) and 3.30% of the total variation, respectively, while the first 10 PCs explain together only 25.14% of the total variation, revealing the absence of genetic structure in the S3MEGGIC population. No differentiated clusters among individuals carrying the wild (*S. incanum* MM577) or the cultivated (*S. melongena* ASI-S-1) cytoplasm were observed. An Analysis of Molecular Variance (AMOVA) was also performed, revealing that only 0.29% of the total sums of squares is accounted for the molecular variation among the *S. melongena* and *S. incanum* cytoplasm groups, thereby resulting in a very small phi-value of 0.0019, which indicates a low level of differentiation and supporting that no population structure exists.

The genome mosaics reconstruction of the S3 MAGIC individuals, in terms of the eight founder haplotypes, showed different haplotype block proportions depending on the genomic position for all chromosomes (**Figure 4**). The estimated contribution of some founders to the overall S3MEGGIC population differed from the expected value of 12.5%. Two of the founder genomes (A0416 and IVIA-371) had a high representation in the genome of the S3 individuals (32.6% and 23.6%, respectively), while two others (AN-S-26 and H15) had a small representation (0.3% in both cases). The wild founder, *S. incanum,* had an average haplotype representation of 5.8%.

Phenotypic Variation Among Multi-Parent Advanced Generation Inter-Cross Individuals and Association Analysis

The screening for absence or presence of PA, FA, and PUC in purple fruits of the 420 S3 MAGIC individuals revealed a considerable variation (**Figure 1B**). Out of the 420 S3 individuals, 57.6% displayed PA and 37.5% FA. Among the individuals displaying FA, 64.3% had the PUC phenotype.

Given that no population structure was observed for the S3MEGGIC population, the phenotypic data, together with the genotypic information, were used for GWAS analysis (**Figure 5**). The GWAS was performed taking into account the kinship in an MLM leading to the identification of significant associations for the evaluated traits.

Plant Anthocyanins

The Manhattan plot for PA revealed two major peaks on chromosome 9 and 10 (PA9 and PA10), with seven significant SNPs over the FDR threshold (LOD > 4.09), and three of them over the Bonferroni threshold (LOD > 5.16) (**Figure 5A**). One SNP (LOD = 4.25) was mapped on PA9 region (between 17.0 and 17.4 Mb; **Supplementary Figure 2A**) and six SNPs, with an LOD between 4.31 and 6.48, were mapped on a PA10 region (between 91.08 and 94.81 Mb; **Figure 5A**).

Fruit Anthocyanins

For FA, 22 significant SNPs above FDR threshold (LOD > 3.6), which included 11 SNPs above the Bonferroni threshold (LOD > 5.16), were plotted on three major peaks located on chromosomes 1, 9, and 10 (FA1, FA9, and FA10, respectively) (**Figure 5C**). One SNP (LOD = 4.07) was detected on FA1 region (between 5.11 and 5.88 Mb) in position 5,346,977 (**Figure 5D**). Five SNPs, with an LOD between 3.86 and 4.22, were located on FA9 region (between 16.2 and 17.4 Mb; **Supplementary Figure 2B**), which overlapped with the PA9 region (**Supplementary Figure 2A**). The other sixteen SNPs (LOD between 4.52 and 9.06) were detected on FA10 region between 91.08 and 94.81 Mb (**Figure 5E**), corresponding to the PA10 region, where significant associations were detected for the PA (**Figure 5B**).

Pigmentation Under the Calyx

Among the three traits evaluated, PUC was the one with the highest number of significant associations. The Manhattan plot for PUC revealed seven main peaks on chromosomes 1 (PUC1), 3 (PUC3.1 and PUC3.2), and 10 (PUC10.1–10.4), with 36 SNPs over the FDR threshold (LOD > 3.7), and 23 of them are above the Bonferroni threshold (LOD > 5.16) (**Figure 5F**). One SNP with an LOD of 4.21 was located on the PUC1 region (between 5.11 and 5.88 Mb) in position 5,480,282 (**Figure 5G**), 133,305 bp away of the significant SNP on the FA1 region (**Figure 5D**). On the chromosome 3, four SNPs (LOD between 4.41 and 6.18) were detected on the PUC3.1 region (between 7.22 and 8.65 Mb; **Supplementary Figure 2C**) and one SNP (LOD = 4.58) on the PUC3.2 region (between 31.54 and 40.09 Mb; **Supplementary Figure 2D**). Thirty SNPs were mapped on chromosome 10 in four genomic regions. One of the SNPs (LOD = 4.2) was identified on the PUC10.1 region (between 2.08 and 2.67 Mb) in position 2,388,375 (**Supplementary Figure 2E**), and other one (LOD = 4.2) on the PUC10.2 region (between 3.94 and 4.34 Mb; **Figure 5H**). Another twelve SNPs were located on the PUC10.3 region, between 91.08 and 94.81 Mb (LOD between 3.86 and 12.44; **Figure 5I**), as observed for PA (**Figure 5B**) and FA (**Figure 5D**). The 16 remaining SNPs (LOD between 3.71 and 7.93) were found in the PUC10.4 region, between 98.25 and 100.55 Mb (**Supplementary Figure 2F**).

Candidate Genes for Anthocyanin Biosynthesis

Based on the results of the GWAS analysis, putative candidate genes were identified close to or within the LD blocks defined in the genomic regions with significant associations (**Supplementary Table 1**). On chromosome 10, in the genomic region (91.08–94.81 Mb) associated with all the evaluated traits (PA10, FA10, and PUC10.3), a candidate gene was identified as similar to *MYB113* (SMEL_010g351850.1), a well-known regulatory transcription factor controlling anthocyanin synthesis in eggplant (Zhou et al., 2019; Shi et al., 2021). In addition, in the

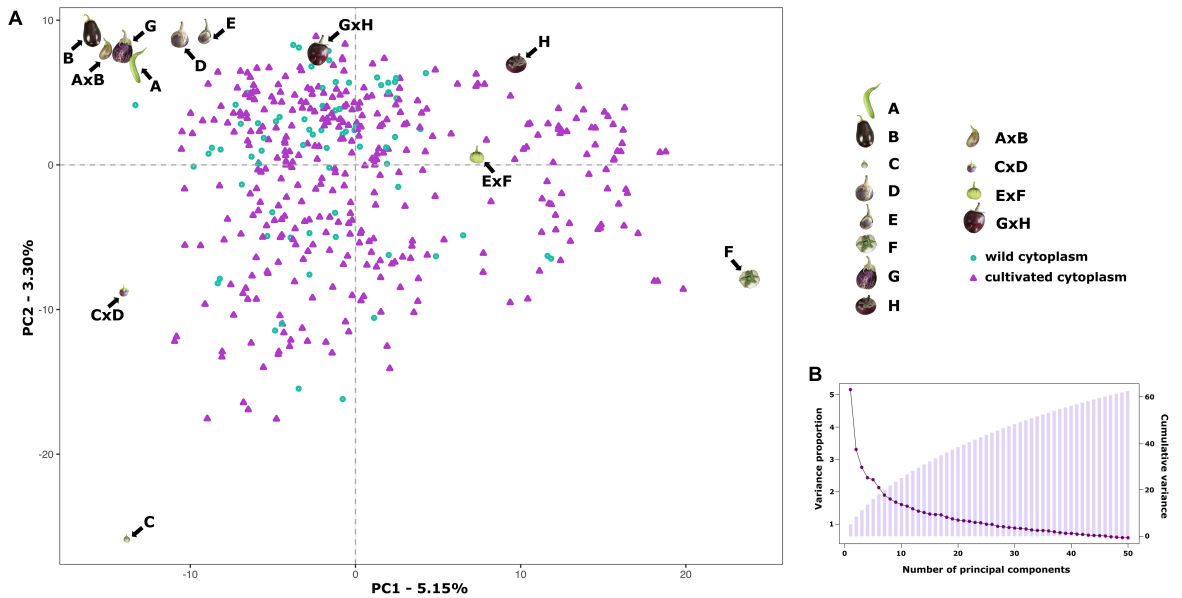

FIGURE 3 | Results of a Principal Component Analysis (PCA) on the S3MEGGIC population. **(A)** PCA of the first two principal components (PCs), including S3 individuals, founder lines, and two-way hybrids. S3 individuals with wild cytoplasm are represented with blue dots while those ones with cultivated cytoplasm are represented with purple triangles. **(B)** Scree plot of the PCs (x-axis) and their contribution to variance (left y-axis); bar blot of the PCs (x-axis) and the cumulative proportion of variance explained (right y-axis).

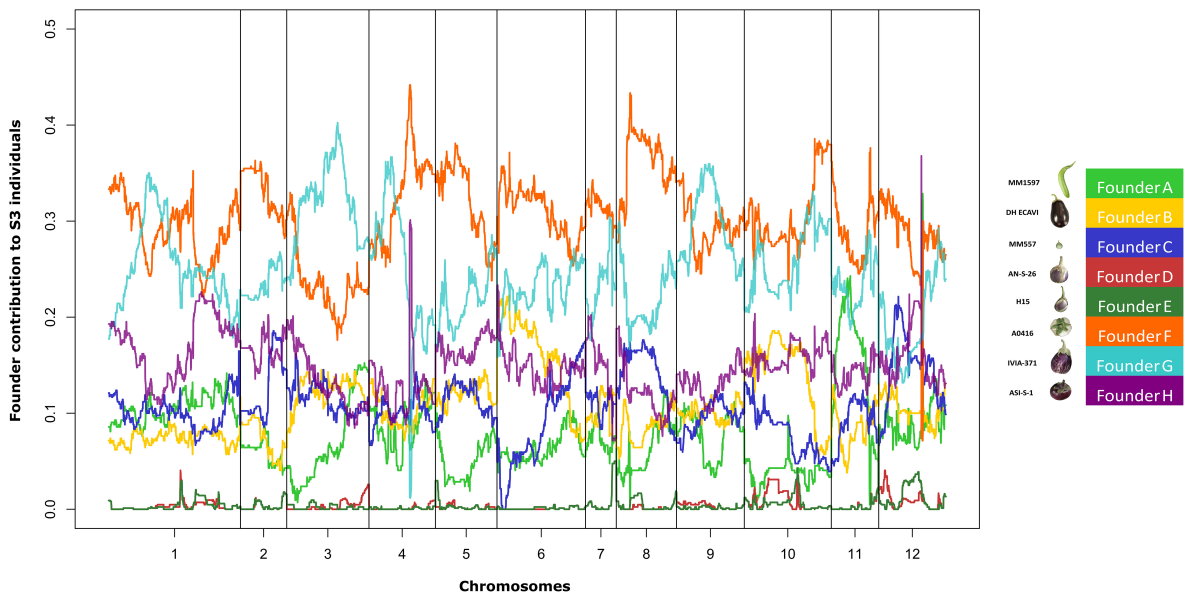

FIGURE 4 | Genome-wide founder haplotype blocks assignment across the entire S3MEGGIC population. In the x-axis, 12 eggplant chromosomes are represented and in the y-axis the average percentage of founders contribution for the 420 S3 individuals. In the legend, the colour code associated with each founder as in **Figure 1**.

genomic region associated with FA and PUC (FA1 and PUC1) on chromosome 1 (5.11–5.88 Mb), another candidate gene was identified similar to *MYB113* (SMEL_001g120500.1). Variants that predicted high impact effects on protein function were annotated by SnpEff for both *MYB113* genes in the population founders (A, C, and F) that do not present anthocyanins in plants and fruits, as confirmed by aligning the founder gene sequences (**Figures 5J,K**). Specifically, for the founders A and C, single frameshift variants were identified in two different positions while a disruptive inframe deletion and a slice region variant were identified for the founder F in a third region of the eggplant gene SMEL_010g351850.1 (**Figure 5J**). For SMEL_001g120500.1,

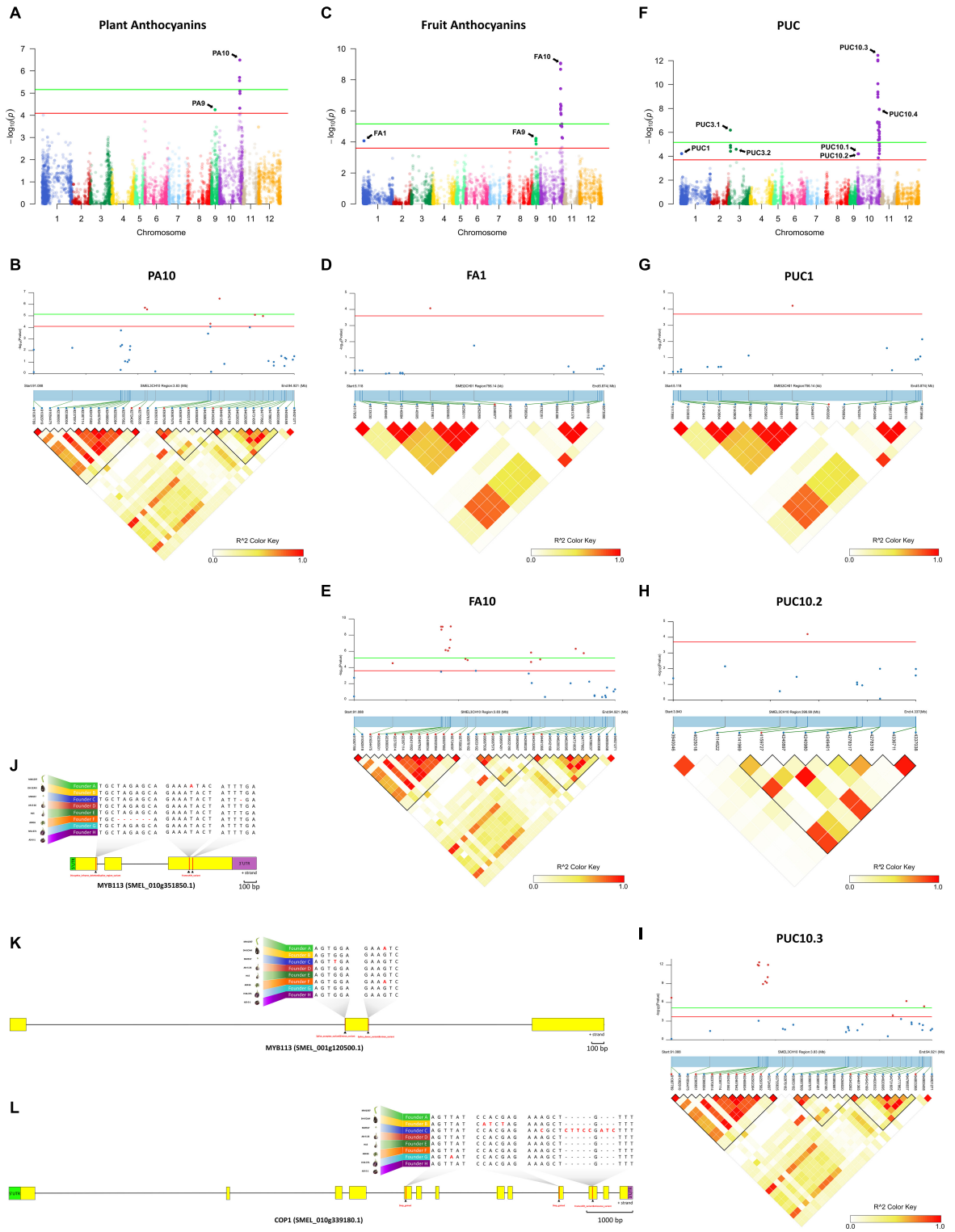

FIGURE 5 | Genome-wide association mapping with linkage disequilibrium (LD) block heatmap of significant regions and potential candidate genes controlling PA, FA, and PUC in eggplant. **(A,C,F)** Manhattan plot for PA **(A)**, FA **(C)**, and PUC **(F)**. Arrows indicate the position of main peaks detected for each trait. The red and green horizontal lines represent, respectively, FDR and Bonferroni significance thresholds at $p = 0.05$. **(B,D,E,G–I)** Local Manhattan plot (top) and LD heatmap (bottom) surrounding the peaks PA10 **(B)**, FA1 **(D)**, FA10 **(E)**, PUC1 **(G)**, PUC10.2 **(H)** and PUC10.3 **(I)**. Pairwise LD between SNPs is indicated as values of R^2 values: red indicates a value of 1 and white indicates 0. **(J–L)** Structure of candidate genes MYB113 (SMEL_010g351850.1) **(J)**, MYB113 (SMEL_001g120500.1) **(K)**, and COP1 (SMEL_010g339180.1) **(L)** and effect of the high-impact variants detected in those of the eight founders.

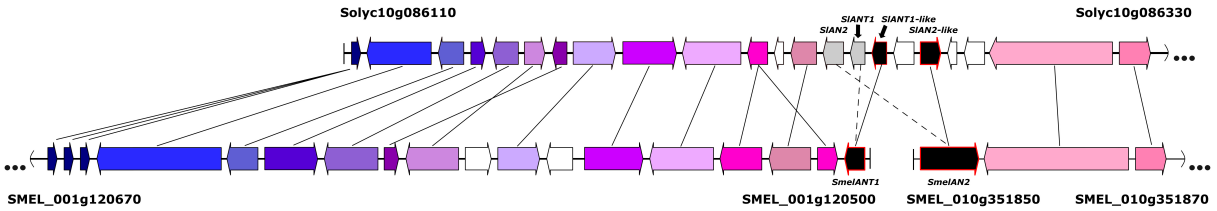

FIGURE 6 | Tomato-eggplant microsynteny representation of a tomato region from chromosome 10 (Solyc10), eggplant region of chromosome 10 (SMEL_010), and a 357 kb fragment of eggplant chromosome 1 (SMEL_001), where candidate genes for anthocyanins synthesis are located.

founders A and F exhibited the exact same splice donor, and the intron variant predicted as high impact, while in the founder C, a splice acceptor and intron variant were identified (**Figure 5K**). Founder reconstruction of the S3 individuals and haplotype blocks were estimated for both candidate gene regions (chromosome 10: 91.08–94.81 Mb; chromosome 1: 5.11–5.88 Mb) (**Supplementary Figures 3A,B**). Founders that contribute most to anthocyanin-related traits in these regions are IVIA-371 (111 and 97 S3 individuals, respectively) and ASI-S-1 (74 and 75 S3 individuals, respectively). Furthermore, a reciprocal best hit BLAST analysis of the SMEL_001g120500.1 onto the tomato genome indicated that this gene corresponded to the orthologue *SlANT1*-like with 70.99% identity. The eggplant SMEL_010g351850.1 corresponded to the tomato orthologue *SlAN2*-like with a 73.73% identity, which has been described as the best candidate gene for the anthocyanin fruit biosynthesis in tomato (Yan et al., 2020). These results suggest that a single duplication event and a translocation of a fragment of at least 357 kb from chromosome 10 to 1 ocurred during eggplant evolution (**Figure 6**).

A candidate gene that corresponds to a *COP1* (SMEL_010g339180.1), a gene that encodes for photo-regulatory proteins (Jiang et al., 2016; Li et al., 2018; He et al., 2019; Naeem et al., 2019), was reported in the genomic regions associated with PUC (PUC10.2) on chromosome 10 (3.94–4.34 Mb; **Supplementary Table 1**). High-effect variants were detected in *COP1* gene in those founders that are able to synthesise anthocyanin under the calyx (B and G) and in the green wild species *S. incanum* (C) (**Figure 5L**). The C founder mutation was confirmed by the F1 hybrid phenotype from the C × D inter-cross, *PUC* (anthocyanic pigmentation under the calyx) × *puc* (no anthocyanic pigmentation under the calyx), which showed the anthocyanin fruits with pigmentation under the calyx (**Figure 1A**). The wild founder presented multiple SNPs compared with the rest that were predicted to cause a frameshift and missed variants, while founders B and G exhibited variants that were predicted to produce stop codons (**Figure 5L**). Founders with a higher haplotype representation contributing to PUC in this region (chromosome 10: 3.94–4.34 Mb) are IVIA-371 (113 S3 individuals) and DH_ECAVI (71 S3 individuals) (**Supplementary Figure 3C**).

Furthermore, we found two candidate genes annotated as similar to *BHLH*, basic helix loop helix protein A (SMEL_009g326640.1), and, similar to *SPA3*, protein SPA1-RELATED 3 (SMEL_010g338090.1), respectively (data not shown), close to LD blocks in the PA9 and FA9 (between 17,862,090 and 17,872,428 Mb), and PUC10.1 (between 3,128,678 and 3,143,068 Mb) regions. Although they were described as genes related to the anthocyanin biosynthetic pathway (Maier et al., 2013; Li et al., 2018; Shi et al., 2021), no high-effects variants were identified in these genes for any of the founders.

DISCUSSION

Multi-parent advanced generation inter-cross (MAGIC) populations are outstanding genetic materials for identifying gene-trait associations with high resolution (Arrones et al., 2020; Scott et al., 2020). The introduction of multiple founders with an increased genetic and phenotypic diversity, together with the multiple rounds of inter-crossing and selfing, increases the number of accumulated recombinant events and, thus, improves the mapping accuracy (Scott et al., 2020). By introducing a wild relative as a founder parent, the genetic variability in the population increases, which is a key point for QTL identification (Gramazio et al., 2020a). Here, we present the first eggplant S3MEGGIC population of which one of the founders was one accession of the close wild relative *S. incanum*.

Large population sizes are essential to increase the power and mapping resolution in MAGIC populations (Collard et al., 2005; Valdar et al., 2006; Jaganathan et al., 2020). Following a simple "funnel" scheme design, the population was kept as large as possible to gather a large number of recombination events. However, a sharp reduction in the number of progenies was observed at the S0 generation, which might be related to the use of the wild species *S. incanum* (C) as a female founder and parent to obtain the simple (CD) and the double (ABCD) hybrids. This interspecific crossing dragged the maternal cytoplasmic background of the wild parent, which might have caused a partial sterility and bias in subsequent generations. Some studies confirmed a strong effect of wild *Solanum* cytoplasms in the reduction in pollen fertility of alloplasmic lines (Khan et al., 2020; Isshiki et al., 2021). However, the PCA highlighted the absence of a population structure, also confirmed by the lack of genetically differentiated cytoplasmic groups.

The genotyping of the S3MEGGIC population was carried out with the 5k probes eggplant SPET platform with a well-distributed marker density along all chromosomes (Barchi et al., 2019a). This genotyping strategy has already been used in the analysis of biparental populations (Herrero et al., 2020).

In this study, we have verified that its use can be extended to multiparent populations. The genotyping revealed a low heterozygosity in the S3MEGGIC individuals, similar to the expected value for an F5-like biparental inter-cross generation (6.25%). The contributions of each of the founder parents to the S3MEGGIC population revealed that some parents had a higher representation than others. Apart from drift effects, several biological reasons could potentially explain cryptic selection processes, causing the unbalanced representation of the genomes (Rockman and Kruglyak, 2008; Thépot et al., 2015), including seed dormancy, delayed germination, precocity, reduced fertility, and parthenocarpy associated to some genomes which have already been reported in eggplant and other crops (Barchi et al., 2010; Khan et al., 2015; Prohens et al., 2017). The rather limited contribution of the wild species *S. incanum* to the final S3 MAGIC individuals may have been caused by a selection pressure, as progenies bred from crosses involving the two different species tend to suffer from a reduced fertility and show a segregation distortion (Lefebvre et al., 2002; Barchi et al., 2010). In addition, *S. incanum* has a recalcitrant germination and a very erratic flowering and fruit set, which strongly depends on environmental conditions (Gisbert et al., 2011; Mangino et al., 2020, 2021). Other reasons for the segregation distortion could be the inability of the current genotyping density to efficiently distinguish between the founders that are genetically closer, like AN-S-26 and H15 genotypes. This phenomenon has already been observed in previous MAGIC populations (Dell'Acqua et al., 2015). A deeper genotyping, resulting in a better haplotype reconstruction, might shed light on the mechanisms that have led to the unbalanced representation of the founder genomes in the S3MEGGIC population. Due to this low contribution, care should be taken in the future when analysing the traits that are present only in these founders.

The anthocyanin biosynthetic pathway is one of the most studied biochemical routes in plants due to its physiological importance. Major structural genes of this pathway are under the control of a regulatory complex, where myeloblastosis (MYB) TFs are recognised as main regulators alone or in complexes with other TFs (Ramsay and Glover, 2005; Kiferle et al., 2015; Liu et al., 2018). Activator or repressor MYB proteins directly and competitively bind the basic-helix-loop-helix (bHLH) *via* the amino terminus domain and can act as positive or negative transcriptional regulators in a tissue-specific mode to modulate the anthocyanin synthesis (Barchi et al., 2019b; Moglia et al., 2020).

In eggplant, anthocyanin-related MYB protein-encoding genes have been reported to be related to fruit peel colouration (Zhang et al., 2014; Docimo et al., 2016; Xiao et al., 2018; Moglia et al., 2020; Toppino et al., 2020). In tomato, a cluster of four different MYB proteins was reported to be involved in the anthocyanin synthesis located on chromosome 10 and encoded by *SlAN2*, *SlANT1*, *SlANT1*-like, and *SlAN2*-like genes (Solyc10g086250, Solyc10g086260, Solyc10g086270, and Solyc10g086290, respectively). However, genetic associations to only two paralogue *MYB113* genes were detected in the S3MEGGIC population on chromosomes 1 and 10 (SMEL_001g120500.1 and SMEL_010g351850.1). The same occurs in potato, in which, only two anthocyanin genes have been identified, i.e., Sotub10g028550 and Sotub10g028540, both located on chromosome 10. Similarly, in pepper, in addition to CA10g11690 and CA10g11650, a third gene (CA10g11710) has been identified as an orthologue to *SlAN2*-like (Barchi et al., 2019b). In eggplant, these two tomato orthologs were previously described as *SmelANT1* and *SmelAN2* (Docimo et al., 2016; Barchi et al., 2019b), corresponding to *SlANT1* and *SlAN2*. However, the reciprocal best hit BLAST analysis of the eggplant coding proteins showed a stronger homology, respectively, to tomato *SlANT1*-like and *SlAN2*-like. Although previous studies in eggplant indicated that the overexpression of *SmelANT1* accounts for constitutive upregulation of most anthocyanin biosynthetic genes (Zhang et al., 2014; Shi et al., 2021), we considered the *SmelAN2* as the best candidate gene for different reasons. Previous studies in eggplant have located the major anthocyanin-related QTLs on chromosome 10 (Barchi et al., 2012; Cericola et al., 2014; Toppino et al., 2016), which are in agreement with the highest association signals for anthocyanin-related traits in our GWAS results. Furthermore, recent studies in tomato suggested that *SlAN2*-like functions as an activator to regulate biosynthesis genes, including *SlANT1*-like, and controls the accumulation of anthocyanins (Yan et al., 2020). However, in the S3MEGGIC population, high-impact variants on protein function were found in both *MYB113* genes for non-anthocyanin fruits. These results could suggest that a duplication of function occurred during an eggplant evolution, and both genes may be required for anthocyanin synthesis.

Although it has been demonstrated that the activation of the anthocyanin biosynthetic pathway in eggplant is strongly regulated by light (Li et al., 2018; Xiao et al., 2018), the PUC mutation confers to some genotypes the ability to synthesise anthocyanins under the calyx regardless of incidence of light (Tigchelaar et al., 1968). In this study, a candidate gene, related with light-dependent anthocyanin biosynthesis in fruits, was detected at the beginning of chromosome 10. The constitutive photomorphogenic1 (*COP1*) gene has been reported to be a regulatory TF responsible for mediating a light-regulated gene expression and development (Jiang et al., 2016; Li et al., 2018; He et al., 2019; Naeem et al., 2019). The *COP1* gene has been demonstrated to act as a light-inactivable repressor interacting with MYB TFs (Jiang et al., 2016), and is considered a "molecular switch" in metabolic processes, which are stimulated by light. In presence of light, *COP1* expression is inhibited, and its concentration decreases rapidly, promoting anthocyanins synthesis. Under dark conditions, *COP1* expression is induced and promotes the degradation of the photomorphogenesis-promoting TF and MYB inhibition by conforming the *COP1*/Suppressor of phya-105 (*SPA*) ubiquitin ligase complex (Maier et al., 2013; Jiang et al., 2016; Li et al., 2018). Although PUC can only be observed in anthocyanic fruits, high-impact variants were also identified in the wild species *S. incanum COP1* gene (founder C, green fruit). The PUC allele is more common in eggplants from western countries, where European markets demand for more homogeneously pigmentation eggplants. On the other hand, in Asian

eggplants, it is more common to find the *puc* phenotype, such as in the ASI-S-1 founder.

CONCLUSION

In conclusion, the S3MEGGIC population represents a landmark breeding material and a tool of great value, which allows the study and fine-mapping of complex traits due to (i) the highly phenotypically diverse founders; (ii) the large population size, being the largest developed eggplant experimental population so far; (iii) the high degree of homozygosity of the final individuals, which constitute a population of fixed "immortal" lines nearly homozygous at each locus; and (iv) the tailored genotyping SPET platform used for the genetic analysis of the population, which has been developed from the whole genome sequencing (WGS) of the founders and allows the comparison with the genotyped materials with the same setup (Gramazio et al., 2020b). In addition, the S3MEGGIC population has demonstrated its potential usefulness for association studies, allowing the establishment of marker associations to anthocyanin-related genes and identification of candidate genes for plant and fruit anthocyanins, including the first identification of a candidate gene for an economically relevant breeding trait in eggplant such as PUC.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA392603.

AUTHOR CONTRIBUTIONS

SV, PG, and JP conceived the idea and supervised the manuscript. GM, AA, MP, and PG performed the field trials. GM and AA prepared a first draft of the manuscript. All other authors reviewed and edited the manuscript.

FUNDING

This work was supported by the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional (grant RTI2018-094592-B-I00 from MCIU/AEI/FEDER, UE) and European Union's Horizon, 2020 Research and Innovation Programme under grant agreement no. 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops). AA is grateful to Spanish Ministerio de Ciencia, Innovación y Universidades for a pre-doctoral (FPU18/01742) contract. MP is grateful to Spanish Ministerio de Ciencia e Innovación for a post-doctoral grant (IJC2019-039091-I/AEI/10.13039/501100011033). PG is grateful to Spanish Ministerio de Ciencia e Innovación for a post-doctoral grant (FJC2019-038921-I/AEI/10.13039/501100011033). Funding for open access charge: Universitat Politècnica de València.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.847789/full#supplementary-material

Supplementary Figure 1 | (A) Chain pollination scheme of the four-way hybrids followed to obtain the eight-way hybrids. **(B)** For each S0 progeny, four plants were germinated, selecting for the next generation (S1) only the first two that set fruits with viable seed. For subsequent generations, two plants were germinated and only the first one that set fruit was selected for the next generation.

Supplementary Figure 2 | Local Manhattan plot (top) and LD heatmap (bottom) surrounding the peaks PA9 **(A)**, FA9 **(B)**, PUC3.1 **(C)**, PUC3.2 **(D)**, PUC10.1 **(E)**, and PUC10.4 **(F)**. The red and green horizontal lines represent, respectively, FDR and Bonferroni significance thresholds. Pairwise LD between SNPs is indicated as values of R^2 values: red indicates a value of 1 and white indicates 0.

Supplementary Figure 3 | Founder haplotype blocks representation predicted for each of the S3 individuals for the three anthocyanin-related candidate gene regions: **(A)** MYB113 (SMEL_001g120500.1) on chromosome 1 between 5.11 and 5.88 Mb identified by the FA1 and PUC1 associations; **(B)** COP1 (SMEL_010g339180.1.01) on chromosome 10 between 3.94 and 4.34 Mb identified by the PUC10.2 association; and **(C)** MYB113 (SMEL_010g351850.1) on chromosome 10 between 91.08 and 94.81 Mb identified by the PA10, FA10, and PUC10.3 associations.

Supplementary Figure 4 | Phenotyping of the S3MEGGIC population for presence or absence of PA, FA, and PUC.

REFERENCES

Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., et al. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in Eudicots. *Plant Cell* 26, 962–980. doi: 10.1105/tpc.113.122069

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., et al. (2020). The dawn of the age of multi-parent magic populations in plant breeding: novel powerful next-generation resources for genetic analysis and selection of recombinant elite material. *Biology* 9, 1–25. doi: 10.3390/biology9080229

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., et al. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6, 1–15. doi: 10.1186/1939-8433-6-11

Barchi, L., Acquadro, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O., et al. (2019a). Single Primer Enrichment Technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Front. Plant Sci.* 10:1005. doi: 10.3389/fpls.2019.01005

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., et al. (2019b). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Sci. Rep.* 9, 1–13. doi: 10.1038/s41598-019-47985-w

Barchi, L., Lanteri, S., Portis, E., Stàgel, A., Valè, G., Toppino, L., et al. (2010). Segregation distortion and linkage analysis in eggplant (*Solanum melongena* L.). *Genome* 53, 805–815. doi: 10.1139/g10-073

Barchi, L., Lanteri, S., Portis, E., Valè, G., Volante, A., Pulcini, L., et al. (2012). A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. *PLoS One* 7:e43740. doi: 10.1371/journal.pone.0043740

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

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300.

Borovsky, Y., Oren-Shamir, M., Ovadia, R., De Jong, W., and Paran, I. (2004). The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to Anthocyanin2 of *Petunia*. *Theor. Appl. Genet.* 109, 23–29. doi: 10.1007/s00122-004-1625-9

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

Campanelli, G., Sestili, S., Acciarri, N., Montemurro, F., Palma, D., Leteo, F., et al. (2019). Multi-parental advances generation inter-cross population, to develop organic tomato genotypes by participatory plant breeding. *Agronomy* 9:119.

Cericola, F., Portis, E., Lanteri, S., Toppino, L., Barchi, L., Acciarri, N., et al. (2014). Linkage disequilibrium and genome-wide association analysis for anthocyanin pigmentation and fruit color in eggplant. *BMC Genomics* 15:896. doi: 10.1186/1471-2164-15-896

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142, 169–196.

D'Amelia, V., Aversano, R., Chiaiese, P., and Carputo, D. (2018). The antioxidant properties of plant flavonoids: their exploitation by molecular plant breeding. *Phytochem. Rev.* 17, 611–625. doi: 10.1007/s11101-018-9568-y

Daunay, M. C., and Hazra, P. (2012). "Eggplant," in *Handbook of Vegetables*, eds K. V. Peter and P. Hazra (Houston, TX: Studium Press), 257–322.

Del Fabbro, C., Scalabrin, S., Morgante, M., and Giorgi, F. M. (2013). An extensive evaluation of read trimming effects on illumina NGS data analysis. *PLoS One* 8:e85024. doi: 10.1371/journal.pone.0085024

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., et al. (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biol.* 16, 1–23. doi: 10.1186/s13059-015-0716-z

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806

Díez, M. J., Picó, B., and Nuez, F. (2002). *Cucurbit Genetic Resources in Europe: Ad Hoc Meeting held in Adana, Turkey, 19 January 2002*. Rome: International Plant Genetic Resources Institute.

Docimo, T., Francese, G., Ruggiero, A., Batelli, G., De Palma, M., Bassolino, L., et al. (2016). Phenylpropanoids accumulation in eggplant fruit: characterization of biosynthetic genes and regulation by a MYB transcription factor. *Front. Plant Sci.* 6:1233. doi: 10.3389/fpls.2015.01233

Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N., and Tanksley, S. D. (2002). Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* 161, 1713–1726. doi: 10.1093/genetics/161.4.1713

Dong, S. S., He, W. M., Ji, J. J., Zhang, C., Guo, Y., and Yang, T. L. (2020). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* 22:bbaa227. doi: 10.1093/bib/bbaa227

FAOSTAT (2019). *FAOSTAT*. Available online at: http://www.fao.org (accessed February 20, 2020).

Frary, A., Frary, A., Daunay, M. C., Huvenaars, K., Mank, R., and Doğanlar, S. (2014). QTL hotspots in eggplant (*Solanum melongena*) detected with a high resolution map and CIM analysis. *Euphytica* 197, 211–228.

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

Gisbert, C., Prohens, J., and Nuez, F. (2011). Treatments for improving seed germination in eggplant and related species. *Acta Horticult.* 898, 45–52.

Gonzali, S., Mazzucato, A., and Perata, P. (2009). Purple as a tomato: towards high anthocyanin tomatoes. *Trends Plant Sci.* 14, 237–241. doi: 10.1016/j.tplants.2009.02.001

Gramazio, P., Pereira-Dias, L., Vilanova, S., Prohens, J., Soler, S., Esteras, J., et al. (2020a). Morphoagronomic characterization and whole-genome resequencing of eight highly diverse wild and weedy *S. pimpinellifolium* and *S. lycopersicum* var. cerasiforme accessions used for the first interspecific tomato MAGIC population. *Horticult. Res.* 7:174. doi: 10.1038/s41438-020-00395-w

Gramazio, P., Jaén-Molina, R., Vilanova, S., Prohens, J., Marrero, Á, Caujapé-Castells, J., et al. (2020b). Fostering conservation via an integrated use of conventional approaches and high-throughput SPET genotyping: a case study using the endangered canarian endemics *Solanum lidii* and *S. vespertilio* (Solanaceae). *Front. Plant Sci.* 11:757. doi: 10.3389/fpls.2020.00757

Gramazio, P., Prohens, J., Borràs, D., Plazas, M., Herraiz, F. J., and Vilanova, S. (2017a). Comparison of transcriptome-derived simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers for genetic fingerprinting, diversity evaluation, and establishment of relationships in eggplants. *Euphytica* 213:264.

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., and Vilanova, S. (2017b). Development and genetic characterization of advanced backcross materials and an introgression line population of *Solanum incanum* in a *S. melongena* background. *Front. Plant Sci.* 8:1477. doi: 10.3389/fpls.2017.01477

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., García-Fortea, E., et al. (2018). Genomic tools for the enhancement of vegetable crops: a case in eggplant. *Not. Bot. Horti Agrobot. Cluj Napoca* 46, 1–13. doi: 10.15835/nbha46110936

Gramazio, P., Yan, H., Hasing, T., Vilanova, S., Prohens, J., and Bombarely, A. (2019). Whole-genome resequencing of seven eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) accessions provides new insights and breeding tools for eggplant enhancement. *Front. Plant Sci.* 10:1220. doi: 10.3389/fpls.2019.01220

He, Y., Chen, H., Zhou, L., Liu, Y., and Chen, H. (2019). Comparative transcription analysis of photosensitive and non-photosensitive eggplants to identify genes involved in dark regulated anthocyanin synthesis. *BMC Genomics* 20:678. doi: 10.1186/s12864-019-6023-4

Herrero, J., Santika, B., Herrán, A., Erika, P., Sarimana, U., Wendra, F., et al. (2020). Construction of a high density linkage map in Oil Palm using SPET markers. *Sci. Rep.* 10, 1–9. doi: 10.1038/s41598-020-67118-y

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.

Huang, B. E., Verbyla, K. L., Verbyla, A. P., Raghavan, C., Singh, V. K., Gaur, P., et al. (2015). MAGIC populations in crops: current status and future prospects. *Theor. Appl. Genet.* 128, 999–1017. doi: 10.1007/s00122-015-2506-0

Hurtado, M., Vilanova, S., Plazas, M., Gramazio, P., Andújar, I., Herraiz, F. J., et al. (2014). Enhancing conservation and use of local vegetable landraces: the Almagro eggplant (*Solanum melongena* L.) case study. *Genet. Resourc. Crop Evol.* 61, 787–795.

Huynh, B. L., Ehlers, J. D., Huang, B. E., Muñoz-Amatriaín, M., Lonardi, S., Santos, J. R. P., et al. (2018). A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *Plant J.* 93, 1129–1142. doi: 10.1111/tpj.13827

Isshiki, S., Nakamura, I., Ureshino, K., and Khan, M. M. R. (2021). Pollen fertility differences in the progenies obtained from a cross between eggplant (*Solanum melongena* L.) as a seed parent and eggplant cytoplasmic substitution lines as pollen parents. *Austral. J. Crop Sci.* 15, 233–237. doi: 10.21475/ajcs.21.15.02.p2785

Jaganathan, D., Bohra, A., Thudi, M., and Varshney, R. K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. *Theor. Appl. Genet.* 133, 1791–1810. doi: 10.1007/s00122-020-03560-w

Jiang, M., Ren, L., Lian, H., Liu, Y., and Chen, H. (2016). Novel insight into the mechanism underlying light-controlled anthocyanin accumulation in eggplant (*Solanum melongena* L.). *Plant Sci.* 249, 46–58. doi: 10.1016/j.plantsci.2016.04.001

Jombart, T. (2008). Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129

Kamvar, Z. N., Tabima, J. F., and Grŕunwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2014, 1–14. doi: 10.7717/peerj.281

Kaushik, P., Plazas, M., Prohens, J., Vilanova, S., and Gramazio, P. (2018). Diallel genetic analysis for multiple traits in eggplant and assessment of genetic distances for predicting hybrids performance. *PLoS One* 13:e0199943. doi: 10.1371/journal.pone.0199943

Khan, M. M. R., Arita, T., Iwayoshi, M., Ogura-Tsujita, Y., and Isshiki, S. (2020). Development of the functional male sterile line of eggplant utilizing the cytoplasm of *Solanum kurzii* by way of the amphidiploid. *Environ. Control Biol.* 58, 79–83. doi: 10.2525/ecb.58.79

Khan, M. M. R., Hasnunnahar, M., Iwayoshi, M., Ogura-Tsujita, Y., and Isshiki, S. (2015). Pollen degeneration in three functional male-sterile lines of eggplant with the wild *Solanum cytoplasms*. *Horticult. Environ. Biotechnol.* 56, 350–357. doi: 10.1007/s13580-015-0015-3

Kiferle, C., Fantini, E., Bassolino, L., Povero, G., Spelt, C., Buti, S., et al. (2015). Tomato R2R3-MYB proteins SlANT1 and SlAN2: same protein activity, different roles. *PLoS One* 10:e0136365. doi: 10.1371/journal.pone.0136365

Knapp, S., Vorontsova, M. S., and Prohens, J. (2013). Wild relatives of the eggplant (*Solanum melongena* L.: Solanaceae): new understanding of species names in a complex group. *PLoS One* 8:e57039. doi: 10.1371/journal.pone.0057039

Knaus, B. J., and Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resourc.* 17, 44–53. doi: 10.1111/1755-0998.12549

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., et al. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* 5:e1000551. doi: 10.1371/journal.pgen.1000551

Lebeau, A., Gouy, M., Daunay, M. C., Wicker, E., Chiroleu, F., Prior, P., et al. (2013). Genetic mapping of a major dominant gene for resistance to *Ralstonia solanacearum* in eggplant. *Theor. Appl. Genet.* 126, 143–158. doi: 10.1007/s00122-012-1969-5

Lefebvre, V., Pflieger, S., Thabuis, A., Caranta, C., Blattes, A., Chauvet, J. C., et al. (2002). Towards the saturation of the pepper linkage map by alignment of three intraspecific maps including known-function genes. *Genome* 45, 839–854. doi: 10.1139/g02-053

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* [Preprint]. arXiv, 1303.3997,

Li, J., He, Y. J., Zhou, L., Liu, Y., Jiang, M., Ren, L., et al. (2018). Transcriptome profiling of genes related to light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.) before purple color becomes evident. *BMC Genomics* 19:201. doi: 10.1186/s12864-018-4587-z

Li, J., Ren, L., Gao, Z., Jiang, M., Liu, Y., Zhou, L., et al. (2017). Combined transcriptomic and proteomic analysis constructs a new model for light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). *Plant Cell Environ.* 40, 3069–3087. doi: 10.1111/pce.13074

Li, L., He, Y., Ge, H., Liu, Y., and Chen, H. (2021). Functional characterization of SmMYB86, a negative regulator of anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). *Plant Sci.* 302:110696. doi: 10.1016/j.plantsci.2020.110696

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., and Bovy, A. (2018). Anthocyanin biosynthesis and degradation mechanisms in Solanaceous vegetables: a review. *Front. Chem.* 6:52. doi: 10.3389/fchem.2018.00052

Lv, L. L., Feng, X. F., Li, W., and Li, K. (2019). High temperature reduces peel color in eggplant (*Solanum melongena*) as revealed by RNA-seq analysis. *Genome* 62, 503–512. doi: 10.1139/gen-2019-0021

Maier, A., Schrader, A., Kokkelink, L., Falke, C., Welter, B., Iniesto, E., et al. (2013). Light and the E3 ubiquitin ligase COP1/SPA control the protein stability of the MYB transcription factors PAP1 and PAP2 involved in anthocyanin accumulation in *Arabidopsis*. *Plant J.* 74, 638–651. doi: 10.1111/tpj.12153

Mangino, G., Plazas, M., Vilanova, S., Prohens, J., and Gramazio, P. (2020). Performance of a set of eggplant (*Solanum melongena*) lines with introgressions from its wild relative *S. incanum* under open field and screenhouse conditions and detection of QTLs. *Agronomy* 10:467. doi: 10.3390/agronomy10040467

Mangino, G., Vilanova, S., Plazas, M., Prohens, J., and Gramazio, P. (2021). Fruit shape morphometric analysis and QTL detection in a set of eggplant introgression lines. *Sci. Horticult.* 282:110006. doi: 10.1016/j.scienta.2021.110006

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.1089/cmb.2017.0096

Mennella, G., Lo Scalzo, R., Fibiani, M., D'Alessandro, A., Francese, G., Toppino, L., et al. (2012). Chemical and bioactive quality traits during fruit ripening in eggplant (*S. melongena* L.) and allied species. *J. Agric. Food Chem.* 60, 11821–11831. doi: 10.1021/jf3037424

Moglia, A., Francesco, E. F., Sergio, I., Alessandra, G., Milani, A. M., Cinzia, C., et al. (2020). Identification of a new R3 MYB type repressor and functional characterization of the members of the MBW transcriptional complex involved in anthocyanin biosynthesis in eggplant (*S. melongena* L.). *PLoS One* 15:e0232986. doi: 10.1371/journal.pone.0232986

Naeem, M., Muqarab, R., and Waseem, M. (2019). The *Solanum melongena* COP1 delays fruit ripening and influences ethylene signaling in tomato. *J. Plant Physiol.* 240:152997. doi: 10.1016/j.jplph.2019.152997

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J. P., et al. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol. J.* 13, 565–577. doi: 10.1111/pbi.12282

Pook, T., Schlather, M., De Los Campos, G., Mayer, M., Carolin Schoen, C., and Simianer, H. (2019). HaploBlocker: creation of subgroup-specific haplotype blocks and libraries. *Genetics* 212, 1045–1061. doi: 10.1534/genetics.119.302283

Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Kilian, B., Díez, M. J., et al. (2017). Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica* 213:158.

Prohens, J., Whitaker, B. D., Plazas, M., Vilanova, S., Hurtado, M., Blasco, M., et al. (2013). Genetic diversity in morphological characters and phenolic acids content resulting from an interspecific cross between eggplant, *Solanum melongena*, and its wild ancestor (*S. incanum*). *Ann. Appl. Biol.* 162, 242–257. doi: 10.1111/aab.12017

Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70. doi: 10.1016/j.tplants.2004.12.011

Ranil, R. H. G., Niran, H. M. L., Plazas, M., Fonseka, R. M., Fonseka, H. H., Vilanova, S., et al. (2015). Improving seed germination of the eggplant rootstock *Solanum torvum* by testing multiple factors using an orthogonal array design. *Sci. Horticult.* 193, 174–181.

Robinson, J. T., Thorvaldsdóttir, H., Turner, D., and Mesirov, J. P. (2020). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *bioRxiv* [Preprint]. bioRxiv, 075499,

Rockman, M. V., and Kruglyak, L. (2008). Breeding designs for recombinant inbred advanced intercross lines. *Genetics* 179, 1069–1078. doi: 10.1534/genetics.107.083873

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., et al. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity* 125, 396–416. doi: 10.1038/s41437-020-0336-6

Shi, S., Liu, Y., He, Y., Li, L., Li, D., and Chen, H. (2021). R2R3-MYB transcription factor SmMYB75 promotes anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). *Sci. Horticult.* 282:110020. doi: 10.1016/j.scienta.2021.110020

Stadlmeier, M., Hartl, L., and Mohler, V. (2018). Usefulness of a multiparent advanced generation intercross population with a greatly reduced mating design for genetic studies in winter wheat. *Front. Plant Sci.* 9:1825. doi: 10.3389/fpls.2018.01825

Thépot, S., Restoux, G., Goldringer, I., Hospital, F., Gouache, D., Mackay, I., et al. (2015). Efficiently tracking selection in a multiparental population: the case of earliness in wheat. *Genetics* 199, 609–623.

Thissen, D., Steinberg, L., and Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Stat.* 27, 77–83. doi: 10.3102/10769986027001077

Tigchelaar, E. C., Janick, J., and Erickson, H. T. (1968). The genetics of anthocyanin coloration in eggplant (*Solanum Melongena* L.). *Genetics* 60, 475–491. doi: 10.1093/genetics/60.3.475

Toppino, L., Barchi, L., Lo Scalzo, R., Palazzolo, E., Francese, G., Fibiani, M., et al. (2016). Mapping quantitative trait loci affecting biochemical and morphological fruit properties in eggplant (*Solanum melongena* L.). *Front. Plant Sci.* 7:256. doi: 10.3389/fpls.2016.00256

Toppino, L., Barchi, L., Mercati, F., Acciarri, N., Perrone, D., Martina, M., et al. (2020). A new intra-specific and high-resolution genetic map of eggplant based on a RIL population, and location of QTLs related to plant anthocyanin pigmentation and seed vigour. *Genes* 11, 1–29. doi: 10.3390/genes11070745

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* 3:731. doi: 10.21105/joss.00731

Valdar, W., Flint, J., and Mott, R. (2006). Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* 172, 1783–1797. doi: 10.1534/genetics.104.039313

Van Eck, H. J., Jacobs, J. M. E., Van Den Berg, P. M. M. M., Stiekema, W. J., and Jacobsen, E. (1994). The inheritance of anthocyanin pigmentation in potato (*Solatium tuberosum* L.) and mapping of tuber skin colour loci using RFLPs. *Heredity* 73, 410–421. doi: 10.1038/hdy.1994.189

Vilanova, S., Alonso, D., Gramazio, P., Plazas, M., García-Fortea, E., Ferrante, P., et al. (2020). SILEX: a fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. *Plant Methods* 16:110. doi: 10.1186/s13007-020-00652-y

Wang, H., Guo, X., Pandey, M. K., Ji, X., Varshney, R. K., Nwosu, V., et al. (2017). "History and impact of the international peanut genome initiative: the exciting journey toward peanut whole-genome sequencing," in *The Peanut Genome*, eds R. Varshney, M. Pandey, and N. Puppala (New York, NY: Springer.), 117–134. doi: 10.1007/978-3-319-63935-2_8

Wei, Q., Wang, W., Hu, T., Hu, H., Wang, J., and Bao, C. (2020). Construction of a SNP-based genetic map using SLAF-Seq and QTL analysis of morphological traits in eggplant. *Front. Genet.* 11:178. doi: 10.3389/fgene.2020.00178

Wickham, H. (2016). *ggplot2 Elegant Graphics for Data Analysis (Use R!)*. New York, NY: Springer.

Xiao, X. O., Li, K., Feng, X. F., and Jin, H. (2018). Transcriptome analyses reveal anthocyanins biosynthesis in eggplant. *PeerJ* 6:e27289v1. doi: 10.1186/s12870-019-1960-2

Yan, S., Chen, N., Huang, Z., Li, D., Zhi, J., Yu, B., et al. (2020). Anthocyanin Fruit encodes an R2R3-MYB transcription factor, SlAN2-like, activating the transcription of SlMYBATV to fine-tune anthocyanin content in tomato fruit. *New Phytol.* 225, 2048–2063. doi: 10.1111/nph.16272

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

Zhou, L., He, Y., Li, J., Liu, Y., and Chen, H. (2019). CBFs function in anthocyanin biosynthesis by interacting with MYB113 in eggplant (*Solanum melongena* L.). *Plant Cell Physiol.* 61, 416–426. doi: 10.1093/pcp/pcz209

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

MINI REVIEW
published: 17 May 2022
doi: 10.3389/fpls.2022.854372

Diversity and Conservation Gap Analysis of the Solanaceae of Southern South America

Andrés Moreira-Muñoz[1†], María Virginia Palchetti[2,3], Vanezza Morales-Fierro[4], Valeria Soledad Duval[5], Rudy Allesch-Villalobos[1] and Carlos E. González-Orozco[6†]*

[1] Instituto de Geografía, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile, [2] Instituto Multidisciplinario de Biología Vegetal - IMBIV, CONICET, Universidad Nacional de Córdoba, Córdoba, Argentina, [3] Departamento de Ciencias Farmacéuticas, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, Córdoba, Argentina, [4] Museo Nacional de Historia Natural, Interior Parque Quinta Normal S/N, Santiago, Chile, [5] Departamento de Geografía y Turismo, Universidad Nacional del Sur, Bahía Blanca, Argentina, [6] Corporación Colombiana de Investigación Agropecuaria- Agrosavia, Centro de Investigación La Libertad, Meta, Colombia

OPEN ACCESS

Edited by:
Rocio Deanna,
University of Colorado, Boulder, United States

Reviewed by:
Eduardo Ruiz-Sanchez,
University of Guadalajara, Mexico
Richard Miller,
Flower Diversity Institute, United States

****Correspondence:***
Andrés Moreira-Muñoz
andres.moreira@pucv.cl

†ORCID:
Andrés Moreira-Muñoz
orcid.org/0000-0002-9136-1391
Carlos E. González-Orozco
orcid.org/0000-0002-9268-5224

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *13 January 2022*
Accepted: *11 April 2022*
Published: *17 May 2022*

Citation:
Moreira-Muñoz A, Palchetti MV, Morales-Fierro V, Duval VS, Allesch-Villalobos R and González-Orozco CE (2022) Diversity and Conservation Gap Analysis of the Solanaceae of Southern South America. Front. Plant Sci. 13:854372. doi: 10.3389/fpls.2022.854372

There is a need to make substantial advances in the taxonomic, systematic, and distribution knowledge of plants, and find better ways of transmission of this information to society to surpass the general pattern described as "plant blindness." The diversity of the plant family Solanaceae reaches its peak in South America; however, many of its species are threatened due to the expansion of the human footprint. Here, we examine the diversity patterns of the family in southern South America (Argentina and Chile) by means of species richness (SR), weighted endemism (WE), and corrected weighted endemism (CWE). We also evaluated conservation gaps in relation to protected areas and the human footprint as a proxy for potential impacts on this biodiversity. Results show two richness centers in NW and NE Argentina, with a high degree of overlap with protected areas, which, on the other side, show a relative high index of human footprint. Comparatively, coastal Atacama (Chile) shows lower richness values, but outstanding CWE and WE values. The coast of Atacama harbors high values due the presence of species of the genus *Nolana* with restricted distributions. Protected areas in this tight coastal strip are sparse, and the human footprint is also relatively high. The degree of protection based on these parameters is then unbalanced, highlighting the need for a geographically explicit strategy for the conservation of the family at subcontinental scale. In doing so, it is likely that other representatives of these unique centers of richness and endemism will benefit.

Keywords: micro-hotspots, conservation biogeography, plant blindness, protected areas, biodiverse

INTRODUCTION

Plant conservation is limited by our knowledge of the diversity, distribution, and abundance of plant species (Gillson et al., 2020). This information is increasing but cannot keep pace with the threats plants suffer, leading to accelerated anthropogenic-caused extinctions and genetic erosion (Knapp, 2019; Wandersee and Schussler, 2001).

South America is one region where different approaches (e.g., taxonomy, phylogeny, biogeography, ethnobotany) are leading to the discovery of new species. Unfortunately, this diversity is dwindling across the continent (Ramirez-Villegas et al., 2012). Rapid land use changes, wildfires, and, in general, human footprint expansion (Zalles et al., 2021) are putting species and ecosystems increasingly under threat. Regional climate change amplifies these threats across biodiversity hotspots (Fuentes-Castillo et al., 2020).

Of special interest for plant conservation at a continental and subcontinental scale are several angiosperm groups that exhibit early diversification in South America, including the Bignoniaceae, Verbenaceae, Asteraceae, and Solanaceae (Olmstead, 2013; Dupin et al., 2017; Deanna et al., 2020). Obtaining a comprehensive understanding of the diversity and distribution of these taxa in South America is a challenge, but this effort is paramount to guide future conservation efforts.

The family Solanaceae encompasses approximately 2,800 species globally (98 genera). It is among the 10 families with the greatest diversity in countries considered to be megadiverse, such as Ecuador and Bolivia (Ulloa Ulloa et al., 2017). In addition, many solanaceous species are important food resources (Samuels, 2015).

Representatives of the Solanaceae family are distributed in the Americas from Alaska to Patagonia, from the sea level to the heights of the Andes (e.g., *Solanum acaule, Jaborosa squarrosa*, and *Lycium humile* (Barboza, 2013; Palchetti et al., 2021). Therefore, the family has inspired important biogeographic studies across the continent (Hijmans and Spooner, 2001; Anguiano-Constante et al., 2018).

The southern end of the continent is where the family reaches the greatest levels of diversity as the fifth largest family of the flora of the Southern Cone of South America after the Asteraceae, Poaceae, Fabaceae, and Orchidaceae (Zuloaga et al., 2019) (**Figure 1**).

The type genus of the family is *Solanum*, a genus of nearly cosmopolitan distribution, published by Linnaeus (1753) in Species Plantarum. It is the second most diverse genus of the vascular plants in the Southern Cone (216 spp.) after *Senecio* (Asteraceae, 415 spp.) (Zuloaga et al., 2019).

Solanaceae is the fourth family in species richness in Argentina (Palchetti et al., 2020) as it is in Chile (Moreira-Muñoz, 2011). With 52% of endemic species, Chile stands out as the most relative endemism-rich country for the Solanaceae, followed by the megadiverse Peru and Brazil (47%) (Palchetti et al., 2020). More diverse genera in Argentina are *Solanum* and *Lycium*, while, in Chile, the highest diversity is represented by *Solanum* and *Nolana*. In Argentina, ecoregions with highest diversity are Chaco, Andes, and Pampa, and highest endemism occurs in Chaco, Andes, Yungas, and Monte ecoregions (Oyarzabal et al., 2018; Del Valle Elías and Aagesen, 2019; Palchetti et al., 2020; Arana et al., 2021). In Chile, the most outstanding ecoregion is the Desert Scrub, as defined by Luebert and Pliscoff (2017). The southern Andes has played a central role in the early diversification of the Solanaceae. The history of elevation change in the Andes occurred concurrently with plant evolution and influenced it, the mountains acting as a corridor, a barrier or providing a geodiversity framework for species diversification (Luebert and Weigend, 2014; Moreira-Muñoz et al., 2020).

Recent advances in the knowledge of the taxonomy of this family have challenged us to update the overall understanding of diversity and conservation priorities. Ca. 30 species are considered as threatened in Argentina (Palchetti et al., 2020) and 14 species in Chile, but most species have not been assessed yet.[1]

Our main goal in this study was to map and analyze the diversity of the Solanaceae in southern South America (Argentina and Chile), overlaying regional richness and endemism with protected areas to identify conservation gaps. Additionally, the human footprint in the existing protected areas is evaluated as a proxy for the degree of effective protection of the family. In this way, the areas of geographic concentration (richness micro-hotspots and centers of endemism) of the family can be identified, which can guide future floristic prospecting and identify areas under threat from land uses incompatible with conservation.

METHODS

The distribution and richness analysis for the family was carried out through a compilation of a database, including different sources of information. Data for Argentina come mostly from the Documenta Florae Australis (2021), while the data for Chile collate specimen information from national (CONC and SGO) and international herbaria. The latter are available through the Global Biodiversity Information Facility (GBIF) platform. Specific status based on recent studies and reports published after April 2019 (not included in Palchetti et al., 2020) has been considered. This included studies, such as the revision of the genus *Schizanthus* (Morales-Fierro et al., 2020; Lavandero et al., 2021); updates in *Nolana* (Hepp and Dillon, 2018); the Morelloid clade of *Solanum* in Argentina (Knapp et al., 2020); cryptic species recently reported (Moreira-Muñoz and Muñoz-Schick, 2020); and new *Petunia* and *Nicotiana* species (Greppi et al., 2019; Santilli et al., 2021). After a first cleaning, data from *ex situ* living collections that introduced species, hybrid taxa, and records with doubtful or incomplete identification at the species level were excluded. Despite the fact that most of the records had georeferenced data in their originally source, errors were detected and corrected (19% of the total records). For this purpose, the Geonames[2] and Mapcarta[3] sites were used. Species nomenclature was based on the following sources: Flora del Cono Sur,[4] POWO,[5] and Solanaceae Source.[6] After eliminating duplicate coordinates for each species, the final database consists of 15,510 records, which include 35 genera and 423 species (**Supplementary Material**).

Diversity maps were carried out by means of the Biodiverse 3.1 software. (Laffan et al., 2010).[7] We used grid cells of 1

[1]https://clasificacionespecies.mma.gob.cl/
[2]https://www.geonames.org/
[3]https://mapcarta.com/
[4]http://www.darwin.edu.ar
[5]http://www.plantsoftheworldonline.org
[6]http://solanaceaesource.org/
[7]http://shawnlaffan.github.io/biodiverse/

FIGURE 1 | Floral morphology and habitat diversity of southern Solanaceae. **(A)** *Schizanthus porrigens* and **(B)** *Schizanthus parvula* from Central Chile; **(C)** *Fabiana denudata* in Catamarca, Argentina; **(D)** *Fabiana ramulosa* in Chilean Altiplano; **(E)** *Exodeconus flavus* in Tarapacá Precordillera, north Chile; **(F)** *Solanum trinominum* in Chilean coastal sand dunes; **(G)** *Salpichroa glandulosa* and **(H)** *Dunalia spinosa* in the heights of Parinacota, Chile; **(I,J)** *Nolana mollis* and *N. villosa* on the lomas formation, Chile; **(K)** *Lycium humile* on salt-rich environments of the Altiplano highlands. Photographs by M. Virginia Palchetti, Rocío Deanna, and Andrés Moreira-Muñoz.

degree (latitude and longitude) (the most suited resolution at subcontinental scale), and three diversity indices were computed from the grid-cell data. The species richness (SR) of a cell is defined as the total number of species within that grid cell. Weighted endemism (WE) is the sum, over all species present in the window for that grid cell, of the number of grid cells in the window with that species divided by the range of that species. The range is defined as the total number of all grid cells in which that species is present. Crisp et al. (2001) defined corrected weighted endemism CWE as the weighted endemism (WE) divided by the total number of species in that window. This last division adjusts the index for the effect of SR. To assess the confidence in the identified centers of endemism, we conducted a randomization test (Laffan and Crisp, 2003). The test generated 999 random iterations, each of which preserves the observed SR of each geographically located cell, and the total number of cells, or ranges, for each species. CWE was then calculated for each random iteration, and the original ranked against the randomizations. Cells with CWE randomization ranks in the top 5% are significantly different from random at a threshold of alfa = 0.05 (González-Orozco et al., 2011) (**Supplementary Material**).

Gap analysis was carried out by means of the superposition of the richness units with available information on protected areas. Argentina-protected units were a downloaded official site.[8] Chilean units are available on the SNIT Geoportal.[9] Additionally, as a proxy for the state of protection of species within protected areas, we overlayed them with the "human footprint" index, as developed by Sanderson et al. (2002) by means of ArcGis 10.3 (ESRI, 2015). The human footprint is a quantitative analysis and representation of human influence across the planetary surface based upon four types of data: population density, land transformation, accessibility, and electrical power infrastructure. Human impact is represented on a scale of 0 (minimum) to 100 (maximum) on a resolution of one square kilometer (Sanderson et al., 2002). A score of 1 indicates the least human influence. The shape file for South America was obtained from[10] and overlapped with the protected areas in Chile and Argentina. We calculated the mean of the human footprint in the set of pixels, encompassing a protected area (**Table 1** and **Supplementary Material**).

RESULTS

Taxonomic Diversity

The number of native Solanaceae species in Argentina and Chile is 430 and belongs to 35 genera. Both countries share 18 genera and 55 species, of which 27 are endemic to Argentina and Chile. Argentina has 315 native species, distributed in 32 genera, of which 80 are endemic species (25%). Chile has 170 native species, distributed in 21 genera, of which 89 are endemic species (52%). **Supplementary Material** considering Argentina and Chile, a total of 8 genera are endemic (i.e., *Benthamiella*, *Combera*, *Reyesia*, *Salpiglossis*, and *Schizanthus* shared between countries; the monotypic genus *Panthacantha* only grows in Argentina, and the monotypic genera *Latua* and *Vestia* in Chile). The most speciose genus in Chile is *Nolana*, with 49 native species, while in Argentina is *Solanum*, with 126 native species.

[8]https://www.argentina.gob.ar/ambiente/areas-protegidas

[9]http://www.geoportal.cl/geoportal/catalog/main/home.page

[10]https://earthdata.nasa.gov/learn/sensing-our-planet/the-human-footprint

TABLE 1 | Protected areas in Argentina and Chile with the higher number of Solanaceae species.

Argentina	No species	Footprint index
Reserva de Biosfera de las Yungas	77	18.3
Patrimonio Cultural de la Humanidad Quebrada de Humahuaca	54	17.5
Reserva Natural Provincial del Iberá	50	11.8
Sitio Ramsar Humedales Chaco	43	28.6
Parque Provincial Cumbres Calchaquíes	40	19.0
Chile	**No species**	**Footprint index**
Parque Nacional Pan de Azúcar	22	17.2
Monumento Natural Paposo Norte	19	26.7
Parque Nacional Fray Jorge	17	36.2
Parque Nacional Morro Moreno	12	26.8
Parque Nacional Llanos de Challe	9	19.7

Spatial Patterns of Biodiversity

According to Biodiverse 3.1 outputs, the primary centers of SR are in the northeast and northwestern regions of Argentina and coastal areas in the north of Chile (**Figure 2**). The SR scores ranged between 1 and 111 species, but SR maximum value of 47 at a threshold of 5–95% was found in a single-grid cell. We identified three main hotspots of WE: east and west northern corner of Argentina and the coastal areas in the north of Chile. WE scores ranged between 0.007 and 11.71, but a WE maximum value of 5.5 (5.5% of species are endemic to that grid cell) at a threshold of 5–95% was found in a single grid cell (**Figure 2**). Once species richness is being corrected, two of the main WE centers remained in the same location, and the one in the northwestern corner of Argentina tended to disappear. However, new areas of high CWE appeared in the south of Chile and Argentina as well as a few scattered grid cells in the central regions. Interestingly, the northern hotspot of endemism in Chile increases in size under the CWE. These changes are likely because of richness biases on specific grid cells. CWE scores ranged between 0 and 0.50, but a CWE maximum value of 0.15 (after a correction of richness, 15% of species are endemic to that grid cell) at a threshold of 5–95% was found in a single-grid cell. To test the validity of spatial CWE patterns, the randomization results show that all major identified centers of endemism were significantly different from random at a threshold of alfa = 0.05 (**Figure 2**).

In synthesis, NW and NE Argentina and the northern Atacama coast can be considered as richness/endemism centers, or micro-hotspots of biodiversity, at the margins of globally recognized biodiversity hotspots (**Supplementary Material**).

Conservation

Gap analysis shows that, from 331 protected areas in Argentina, 129 have a degree of spatial coincidence with the distribution of Solanaceae. In Chile, 38 from 102 protected units superpose with cells with the distribution of Solanaceae. Protected areas with highest number of species in Chile and Argentina are those summarized in **Table 1**. Protected areas in the northwest of Argentina show a spatial match with richness centers and, to some lower degree, in NE Argentina (**Figure 2E**). The presence of protected areas in Chile is sparse at the coast of Atacama, where the highest endemism appears. The human footprint is higher in Central Chile and Central/northern Argentina, around metropolitan central areas (**Supplementary Material**). The mean value of the human print index in each protected area (PA) varies between 1.1 and 87 in Argentina (20.1 total mean). The index varies between 1.8 and 71 in the case of PA in Chile (19.7 total mean). PA areas encompassing high numbers of Solanaceae and a relatively high index are Sitio Ramsar Humedales Chaco in Argentina and Parque Nacional Fray Jorge in Chile (**Table 1**).

DISCUSSION

One of the greatest current challenges in conservation biogeography is identifying areas of high species richness and

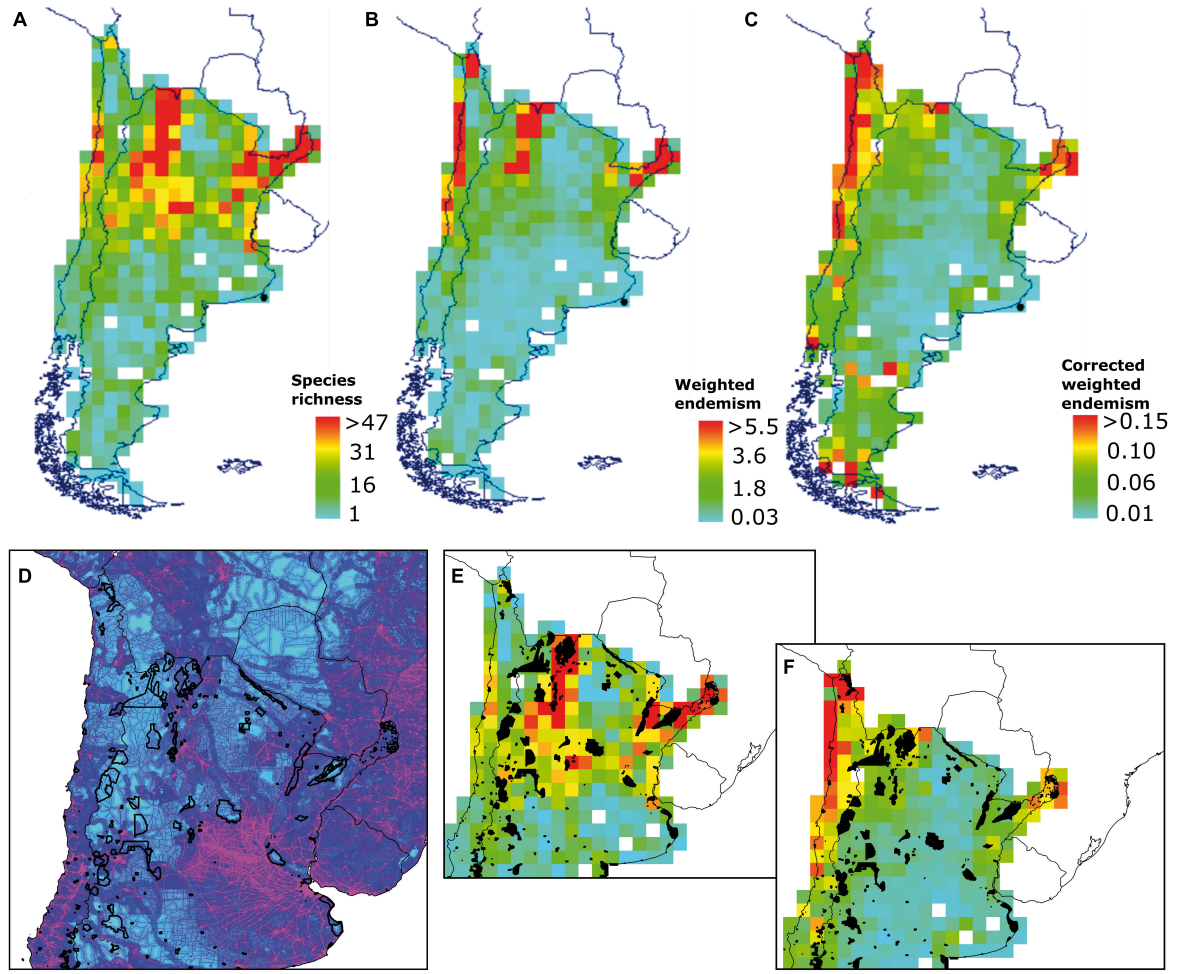

FIGURE 2 | Diversity of Solanaceae species in southern South America, mapped using Biodiverse 3.1 on 1 × 1 degree matrix: **(A)** Species richness; **(B)** Weighted endemism; **(C)** Corrected weighted endemism; **(D)** A detailed area showing a footprint index and protected areas; **(E)** Protected areas overlapped with species richness; **(F)** Protected areas overlapped with corrected weighted endemism. Detailed maps and tables are available as (**Supplementary Material**).

endemism, both to establish conservation priorities and to better understand the evolution of plant diversity. This is especially relevant in southern South America, a territory recognized as especially important in the evolution of diverse families, such as Solanaceae, Bignoniaceae, Verbenaceae, Asteraceae, Orchidaceae (Olmstead, 2013; Ulloa Ulloa et al., 2017). In southern South America, the wide geographic distribution of the family Solanaceae and its taxonomic richness is partly explained by a long evolutionary history since the early Eocene (Dupin et al., 2017; Deanna et al., 2020).

Diversity indexes applied to this wide study area show different regions as outstanding for Solanaceae species richness (SR) and endemism, respectively. Weighted endemism (WE) and corrected weighted endemism (CWE) are parameters that have shown great utility for analysis of restricted distributions (Sosa and de Nova, 2012; Rodríguez et al., 2018; Ruiz-Sánchez et al., 2020). SR shows two main areas in NW and NE Argentina. Richness areas show an important degree of protection mainly by "Reserva de Biosfera de las Yungas" and "Patrimonio Cultural de la Humanidad Quebrada de Humahuaca" in the NW, and "Reserva Natural Provincial del Iberá" and "Sitio Ramsar Humedales Chaco" in the NE. Both regions have been largely recognized as important sources of biodiversity and medicinal plants (Hilgert and Gil, 2006; Bernacki et al., 2015; Campanello et al., 2019). Especially Yungas has been recognized as an outstanding ecoregion for the conservation of different biotic groups (Grosso and Quintana, 2009; Arana et al., 2016; Torres and González-Reyes, 2017). When weighted endemism (WE) and corrected weighted endemism (CWE) are applied, another region clearly appears as outstanding at the regional scale: the coast of Atacama, mainly due the diversity and restricted distribution of species in genus *Nolana*. This genus is a main component of Lomas vegetation, a plant formation found along the coast from Peru to northern Chile, characterized by high endemism and richness maintained by the coastal fog reaching the coastal cliffs at an altitude around 1,000 m asl (Muñoz-Schick et al., 2001). The Solanaceae component of this remarkable environment has been emphasized by Dillon (2005, 2016), and other components of this

unique biota have been recently highlighted (Moat et al., 2021; Pizarro-Araya et al., 2021).

The three outstanding centers of richness and endemism (micro-hotspots) of southern Solanaceae species are part of three different biodiversity hotspots at a continental scale: the northwest of Argentina is at the margin of the Tropical Andes hotspot; the northeast is at the margin of the Atlantic Forest hotspot, and the coastal Atacama is adjacent to the northern end of the Central Chile hotspot (**Supplementary Material**). Hotspots are defined as large areas of high species richness subjected to intense threats and landscape modification, as it gets clear by the expansion of the human footprint at a continental scale (Zalles et al., 2021). At least in Chile, protected areas have huge deficits in effective protection (Petit et al., 2018), and, certainly, much more efforts shall be done for effective protection at the landscape level, including target families, such as the Solanaceae. Results presented here remark the need for a geographically explicit strategy for the conservation of the family Solanaceae at subcontinental scale. In doing so, it is likely that other representatives of these centers of richness and endemism will benefit, promoting conservation and restoration at the landscape scale (Ianni and Geneletti, 2010; Malizia et al., 2012), and hopefully contributing also to reduce the generalized "plant blindness."

AUTHOR CONTRIBUTIONS

AM-M and MP conceived the study and wrote the manuscript. VM-F compiled, cleaned, and updated the data base. VM-F and RA-V ran GIS analysis by means of ArcGis. VSD discussed the implications of biodiversity values in relation to protected areas. CG-O ran Biodiverse 3.1 analysis and analyzed results. All authors edited the manuscript and approved the submitted version.

FUNDING

AM-M has a grant from Agencia Nacional de Investigación y Desarrollo: Fondecyt-ANID 1180211. VSD was a CONICET Research Fellow.

ACKNOWLEDGMENTS

Two reviewers provided very helpful comments on earlier drafts of the manuscript. We gratefully recognize the big efforts done by many botanists in the compilation of Documenta Florae Australis. RD invited us to this special issue and kindly provided photos. Mélica Muñoz-Schick and Alicia Marticorena shared information from SGO and CONC herbaria, respectively. Gloria Barboza, Juan José Cantero, and Miguel Dillon have been of constant help regarding taxonomic and distributional queries.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.854372/full#supplementary-material

REFERENCES

Anguiano-Constante, M. A., Munguía-Lino, G., Ortiz, E., Villaseñor, J. L., and Rodríguez, A. (2018). Riqueza, distribución geográfica y conservación de *Lycianthes* serie Meizonodontae (Capsiceae, Solanaceae). *Rev. Mex. Biodivers.* 89, 516–529.

Arana, M. D., Larsen, C., and Ponce, M. M. (2016). Revisión y análisis panbiogeográfico de las Hymenophyllaceae de las Yungas meridionales de Argentina (Selva Tucumano-Boliviana). *Rodriguésia* 67, 55–75. doi: 10.1590/2175-7860201667105

Arana, M. D., Natale, E., Ferretti, N., Romano, G., Oggero, A., Martínez, G., et al. (2021). Esquema biogeográfico de la República Argentina. *Opera Lilloa.* 56:240.

Barboza, G. E. (Coord). (2013). "Solanaceae," in *Flora Argentina*, 1st Edn, Vol. 13, eds F. O. Zuloaga, M. Belgrano, and A. M. Anton (San Isidro: IBODA-IMBIV, CONICET), 1–350. doi: 10.2307/j.ctt16vj2hs.4

Bernacki, F., Albornoz, P., Valoy, M., and Ordano, M. (2015). Anatomía de flor y fruto de *Vassobia breviflora* (Solanaceae) en el sur de las Yungas australes (Argentina). *Phyton* 84, 478–487.

Campanello, P. I., von Below, J., Hilgert, N. I., Cockle, K., Villagra, M., di Francescantonio, D., et al. (2019). ¿Es posible el uso sostenible del bosque en Misiones? Necesidades de manejo a diferentes escalas, investigación, intervenciones de alto impacto y más recursos económicos. *Ecol. Aust.* 29, 122–137. doi: 10.25260/EA.19.29.1.0.756

Crisp, M. D., Laffan, S. W., Linder, P., and Monro, A. (2001). Endemism in the Australian flora. *J. Biogeogr.* 28, 183–198. doi: 10.1046/j.1365-2699.2001.00524.x

Deanna, R., Wilf, P., and Gandolfo, M. A. (2020). New physaloid fruit-fossil species from early Eocene South America. *Am. J. Bot.* 107, 1–14. doi: 10.1002/ajb2.1565

Del Valle Elías, G., and Aagesen, L. (2019). Areas of endemism and recent speciation in the Southern Cone of South America, using *Senecio* (Asteraceae) as a proxy. *Biol. J. Linn. Soc.* 128, 70–82.

Dillon, M. O. (2005). "Solanaceae of the Lomas formations of Coastal Peru and Chile," in *A Festschrift for William G. D'Arcy: The Legacy of a Taxonomist, Monographs*, eds V. Hollowell, T. Keating, W. Lewis, and T. Croa 131–155.

Dillon, M. O. (2016). "*Nolana* (Solanaceae)," in *The Families and Genera of Vascular Plants. Asterales*, et al Edn, Vol. 8, eds G. E. Barboza, A. T. Hunziker, and G. Bernardello (Berlin: Springer-Verlag), 343–344.

Documenta Florae Australis (2021). Available at http://www.darwin.edu.ar/iris/ (accessed July, 2021).

Dupin, J., Matzke, N. J., Särkinen, T., Knapp, S., Olmstead, R. G., Bohs, L., et al. (2017). Bayesian estimation of the global biogeographical history of the Solanaceae. *J. Biogeogr.* 44, 887–899. doi: 10.1111/jbi.12898

ESRI (2015). *ArcGis 10.3*. Redlands, CA: Esrilabs.

Fuentes-Castillo, T., Hernández, H. J., and Pliscoff, P. (2020). Hotspots and ecoregion vulnerability driven by climate change velocity in Southern South America. *Reg. Environ. Chang.* 20, 1–16. doi: 10.1007/s10113-020-01595-9

Gillson, L., Seymour, C. L., Slingsby, J. A., and Inouye, D. W. (2020). What are the grand challenges for plant conservation in the 21st Century? *Font. Conserv. Sci.* 1, 1–6. doi: 10.3389/fcosc.2020.600943

González-Orozco, A. C., Laffan, S. W., and Miller, J. T. (2011). Spatial distribution of species richness and endemism of the genus *Acacia* in Australia. *Austr. J. Bot.* 59, 600–608. doi: 10.1071/BT11112

Greppi, J. A., Hagiwara, J. C., and Stehmann, J. R. (2019). A new species of *Petunia* (Solanaceae) from Corrientes, Argentina. *Phytotaxa* 414, 289–295. doi: 10.11646/phytotaxa.414.6.3

Grosso, L. D. E., and Quintana, M. G. (2009). Can insect data be used to infer areas of endemism? An example from the Yungas of Argentina. *Rev. Chil. Hist. Nat.* 82, 507–522.

Hepp, J., and Dillon, M. O. (2018). A new endemic species of *Nolana* (Solanaceae-Nolaneae) from near Iquique, Chile. *Arnaldoa* 25, 323–338. doi: 10.22497/arnaldoa.252.25202

Hijmans, R. J., and Spooner, D. M. (2001). Geographic distribution of wild potato species. *Am. J. Bot.* 88, 2101–2112. doi: 10.2307/3558435

Hilgert, N. I., and Gil, G. E. (2006). Medicinal plants of the Argentine Yungas plants of the Las Yungas biosphere reserve, Northwest of Argentina, used in health care. *Biodivers. Conserv.* 15, 2565–2594. doi: 10.1007/s10531-005-3874-6

Ianni, E., and Geneletti, D. (2010). Applying the ecosystem approach to select priority areas for forest landscape restoration in the Yungas, Northwestern Argentina. *Environ. Manage.* 46, 748-760.doi: 10.1007/s00267-010-9553-8

Knapp, S. (2019). Are humans really blind to plants? *Plants People Planet* 1, 164–168. doi: 10.1002/ppp3.36

Knapp, S., Chiarini, F., Cantero, J. J., and Barboza, G. E. (2020). The Morelloid clade of *Solanum* L. (Solanaceae) in Argentina: nomenclatural changes, three new species and an updated key to all taxa. *PhytoKeys* 164, 33–66. doi: 10.3897/phytokeys.164.54504

Laffan, S. W., and Crisp, M. D. (2003). Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. *J. Biogeogr.* 30, 511–520. doi: 10.1046/j.1365-2699.2003.00875.x

Laffan, S. W., Lubarsky, E., and Rosauer, A. F. (2010). Biodiverse, a tool for the spatial analysis of biological and related diversity. *Ecography* 33, 643–647. doi: 10.1111/j.1600-0587.2010.06237.x

Lavandero, N., Chinga, J., Pinto, R., and Pérez, M. F. (2021). A new distinctive species of *Schizanthus* (Solanaceae) and the Reinstatement of *Schizanthus fallax*. *Syst. Bot.* 46, 456–469. doi: 10.1600/036364421X16231782047541

Linnaeus, C. (1753). *Species Plantarum. The Ray Society's Facsimile*, Vol. 2014. London.

Luebert, F., and Pliscoff, P. (2017). *Sinopsis Bioclimática y Vegetacional de Chile*, 2nd Edn. Santiago: Editorial Universitaria.

Luebert, F., and Weigend, M. (2014). Phylogenetic insights into Andean plant diversification. *Front. Ecol. Evol.* 2:27. doi: 10.3389/fevo.2014.00027

Malizia, L., Pacheco, S., Blundo, C., and Brown, A. (2012). Caracterización altitudinal, uso y conservación de las Yungas Subtropicales de Argentina. *Ecosistemas* 1-2, 53–73.

Moat, J., Orellana-Garcia, A., Tovar, C., Arakaki, M., Arana, C., Cano, A., et al. (2021). Seeing through the clouds – mapping desert fog oasis ecosystems using 20 years of MODIS imagery over Peru and Chile. *Int. J. Appl. Earth Obs. Geoinf.* 103:102468. doi: 10.1016/j.jag.2021.102468

Morales-Fierro, V., Muñoz-Schick, M., and Moreira-Muñoz, A. (2020). Synopsis of *Schizanthus* Ruiz & Pav. (Solanaceae), a genus endemic to the southern Andes. *PhytoKeys* 154, 57–102. doi: 10.3897/phytokeys.154.49615

Moreira-Muñoz, A. (2011). *Plant Geography of Chile*. Dordrecht: Springer.

Moreira-Muñoz, A., and Muñoz-Schick, M. (2020). Rediscovery and taxonomic placement of *Solanum polyphyllum* Phil. (Solanaceae), a narrow endemic from the Chilean Atacama Desert. *PhytoKeys* 156, 47–54. doi: 10.3897/phytokeys.156.53703

Moreira-Muñoz, A., Scherson, R. A., Luebert, F., Román, M. J., Monge, M., Diazgranados, M., et al. (2020). Biogeography, phylogenetic relationships and morphological analyses of the South American genus *Mutisia* L.f. (Asteraceae) shows early connections of two disjunct biodiversity hotspots. *Org. Divers. Evol.* 20, 639–656. doi: 10.1007/s13127-020-00454-z

Muñoz-Schick, M., Pinto, R., Mesa, A., and Moreira-Muñoz, A. (2001). Fog oases during the El Niño Southern Oscillation 1997-1998, in the coastal hills south of Iquique, Tarapacá region, Chile. *Rev. Chil. Hist. Nat.* 74, 1–15.

Olmstead, R. G. (2013). Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: a comparison of continental and intercontinental diversification patterns. *Bot. J. Linn. Soc.* 171, 80–102. doi: 10.1111/j.1095-8339.2012.01306.x

Oyarzabal, M., Clavijo, J., Oakley, L., Biganzoli, F., Tognetti, P., Barberis, I., et al. (2018). Unidades de vegetación de la Argentina. *Ecol. Austr.* 028, 40–63.

Palchetti, M. V., Cantero, J. J., and Barboza, G. E. (2020). Solanaceae diversity in South America and its distribution in Argentina. *An. Acad. Bras. Cienc.* 92, 1–17. doi: 10.1590/0001-3765202020190017

Palchetti, M. V., Cantero, J. J., Morales-Fierro, V., Barboza, G. E., and Moreira-Muñoz, A. (2021). Living in extreme environments: distribution of *Lycium humile* (Solanaceae), an endemic halophyte from the Altiplano-Puna region, South America. *PhytoKeys* 185, 1–15. doi: 10.3897/phytokeys.185.71377

Petit, I. J., Campoy, A. N., Hevia, M., Gaymer, C. F., and Squeo, F. A. (2018). Protected areas in Chile: are we managing them? *Rev. Chil. Hist. Nat.* 91, 1–8. doi: 10.1186/s40693-018-0071-z

Pizarro-Araya, J., Alfaro, F. M., Ojanguren-Affilastro, A., and Moreira-Muñoz, A. (2021). A Fine-Scale Hotspot at the Edge: epigean Arthropods from the Atacama Coast (Paposo-Taltal, Antofagasta Region, Chile). *Insects* 12, 1–15. doi: 10.3390/insects12100916

Ramirez-Villegas, J., Jarvis, A., and Touval, J. (2012). Analysis of threats to South American flora and its implications for conservation. *J. Nat. Conserv.* 20, 337–348. doi: 10.1016/j.jnc.2012.07.006

Rodríguez, A., Castro-Castro, A., Vargas-Amado, G., Vargas-Ponce, O., Zamora-Tavares, P., González-Gallegos, J., et al. (2018). Richness, geographic distribution patterns, and areas of endemism of selected angiosperm groups in Mexico. *J. System. Evol.* 56, 537–549. doi: 10.1111/jse.12457

Ruiz-Sánchez, E., Munguía-Lino, G., Vargas-Amado, G., and Rodríguez, A. (2020). Diversity, endemism and conservation status of native Mexican woody bamboos (Poaceae: Bambusoideae: Bambuseae). *Bot. J. Linn. Soc.* 192, 281–295.

Samuels, J. (2015). Biodiversity of food species of the solanaceae family: preliminary taxonomic inventory of subfamily solanoideae. *Resources* 4, 277–322. doi: 10.3390/resources4020277

Sanderson, E. W., Jaiteh, M., Levy, M. A., Redford, K. H., Wannebo, A. V., and Woolmer, G. (2002). The human footprint and the last of the wild. *Bioscience.* 52, 891–904. doi: 10.1641/0006-3568(2002)052[0891:thfatl]2.0.co;2

Santilli, L., Pérez, M. F., de Schrevel, C., Dandois, P., Mondaca, H., and Lavandero, N. (2021). *Nicotiana rupicola* sp. nov and *Nicotiana knightiana* (Paniculatae, Solanaceae), a new endemic and a new record for the flora of Chile. *ARPHA Preprints* 1:e73387. doi: 10.3897/arphapreprints.e73387

Sosa, V., and de Nova, J. A. (2012). Endemic angiosperm lineages in Mexico: hotspots for conservation. *Acta Bot. Mex.* 100, 293–315.

Torres, V. M., and González-Reyes, A. X. (2017). Diversidad taxonómica y funcional de arañas (Araneae) epigeas en bosques nativos de las Yungas (Salta. Argentina). *Caldasia* 39, 326–344.

Ulloa Ulloa, C., Acevedo-Rodríguez, P., Beck, S., Belgrano, M. J., Bernal, R., Berry, P. E., et al. (2017). An integrated assessment of the vascular plant species of the Americas. *Science* 358, 1614–1617. doi: 10.1126/science.aao0398

Wandersee, J., and Schussler, E. (2001). Toward a theory of plant blindness. *Plant Sci. Bull.* 47, 2–8.

Zalles, V., Hansen, M. C., Potapov, P. V., Parker, D., Stehman, S. V., Pickens, A. H., et al. (2021). Rapid expansion of human impact on natural land in South America since 1985. *Sci. Adv.* 7, 1–12. doi: 10.1126/sciadv.abg1620

Zuloaga, F. O., Belgrano, M. J., and Zanotti, C. A. (2019). Actualización del catálogo de las plantas vasculares del Cono Sur. *Darwin. Nueva Ser.* 7, 208–278. doi: 10.14522/darwiniana.2019.72.861

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

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