Genome-wide association study and candidate gene identification for agronomic traits in 182 upward-growing fruits of C. frutescens and C. annuum

Genying Fu
Hainan University
Shuang Yu
Hainan University
Kun Wu
Hainan University
Mengxian Yang
Hainan University
Muhammad Ahsan Altaf
Hainan University
Zhuo Wu
Hainan University
Qin Deng
Hainan University
Xu Lu
Hainan University
Huizhen Fu
Hainan University
Zhiwei Wang
Hainan University
Shanhan Cheng
990865@hainanu.edu.cn
Hainan University

Article

Keywords: pepper, agronomic traits, genome-wide association study(GWAS), single nucleotide polymorphisms (SNPs), whole genome resequencing

Posted Date: March 6th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3972890/v1

Additional Declarations: No competing interests reported.

Abstract

Pepper agronomic traits serve as pivotal indicators for characterizing germplasm attributes and correlations. Investigating genotypic disparities through phenotypic variations holds significant scientific merit Whole genome resequencing facilitates comprehensive examination of diverse individuals with known references, enabling subsequent differential analyses to pinpoint single nucleotide polymorphisms(SNPs) linked to pepper agrnomic. This study ,conducted a genome-wide association study(GWAS) encompassing 26 agronomic traits in 182 pepper specimens. Rigorous measures, including phylogenetic analysis, population structure analysis, population principal component analysis, kinship analysis, and linkage disequilibrium analysis, were employed to ensure the precision and reliability of GWAS results. The optimal statistical model. Was determined through these analyses. A total of 929 SNPs significantly associated with 26 agronomic traits, were identified, alongside the detection of 519 candidate genes within 100kb region adjacent to these SNPs. Additionally, through gene annotation and expression pattern scrutiny, genes such as SCPL13, extensin-1-like, and DDB1 correlated with fruit traits in *Capsicum frutescens* and *Capsicum annuum* were validated via qRT-PCR. This validation provides a robust reference for molecular marker-assisted breeding of pepper agronomic traits, offering both genetic resources and theoretical foundations for future endeavors in molecular marker-assisted breeding for pepper.

Introduction

Pepper (*Capsicum spp.*), belonging to the *Solanaceae* family, originated in tropical and subtropical regions of Central and South America[1]. It serves as a widely consumed fresh or seasoning vegetable, boasting a global cultivation area exceeding 2.23 million hectares[2]. Peppers are characterized by their nutritional richness, containing significant amounts of vitamin C, capsaicin, and antioxidants, with diverse applications in culinary, medicinal, and military fields[3–5]. The consequential demand for consumption and the broad spectrum of applications underscore its significance. With a storied cultivation history and an abunance of germplasm resources, *Capsicum* comprises five cultivated species: *C. annuum* L., *C. frutescens* L., *C. chinense Jacquin*, *C. baccatum* L., and *C. pubescens.* Additionally, there are over 30 related crop feral relatives. These species exhibit substantial genetic variation in plant structure, fruit characteristics, flowers, leaves, and other traits. Agronomic traits, encompassing fruit size, shape, color, weight, and features associated with biotic and abiotic stresses, sensory attributes and nutritional qualities, constitute the foundational descriptors for genetic diversity of germplasm resources. Simultaneously, these traits serve as key breeding objectives in horticultural plants. Therefore, the identification and exploration of genes governing these traits hold paramount importance in enhancing both yield and quality of horticultural products.

Traditionally, molecular markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), simple sequence repeats (SSR), and linkage maps, have been employed to map numerous quantitative trait loci(QTL) related to pepper traits onto have chromosomes. Examples include fs3.1 and fs10.1, fruit shape, and fw2.1 and fw6.1, governing fruit weight. However, achieving accurate detection and identification of QTLs necessitates higher-density genetic maps.Yarnes et al. (2013) utilized a pepper gene chip encompassing 30,815 EST sequences to perform gene typing on a recombinant inbred line (RIL) population, successfully mapping 96 QTLs for 38 traits.Unfortunately, the analysis did not extend to identifying genes in the QTL region[6].Han et al. (2016) constructed an ultra-high-density bin map of pepper using a sliding window approach, detecting 86 significant QTLs controlling 17 horticultural traits, including plant structure, leaf dimensions, flower size, fruit dimensions, and weight, Notably, they identified 32 major effect QTL sites governing 13 traits[7]. However, the limited number of gene loci influencing critical agronomic traits of pepper persists, primarily attribute to the utilization of a single material and low-coverage sequencing.

High-throughput resequencing of significant germplasm resources in both flora and fauna has emerged as a valuable tool for genome-wide genotyping and the execution of genome-wide association studies (GWAS). This approach involves the integration of targeted phenotype data to identify single nucleotide polymorphism(SNP) loci associated with specific phenotypes, facilitating the precise localization of genes linked to various traits. The advantages of this method include the rapid rapid identification of genetic variations, comprehensive analysis of population structure, and the generation of an extensive pool of candidate genes. Its successfully application extend across diverse crops, such as wheat[8], maize[9], soybean[10], grapes[11], cucumbers[12], and tomatoes[13]. Colonna et al. (2019) sequenced 1.8% of the genome of 373 materials from 11 pepper species sourced from 51

countries. Their investigation illuminated genomic variations at the population level, further delineating subdivisions at both group and species levels. The study confirmed the presence of the Longifolia 1-like gene, influencing fruit shape[14]. Ahn et al. (2016) conducted a comprehensive resequencing of the entire genomes of *C. baccatum* (PRH1 - a powdery mildew resistant line) and *C. annuum* (Saengryeg - a powdery mildew susceptible line) for GWAS analysis, successfully identifying 6281 SNPs associated with 46 powdery mildew resistance genes[15]. Wu et al. (2019) utilized specific-locus amplified fragment sequencing (SLAF-seq) to perform GWAS on 287 agronomic traits of 36 pepper resources. Notably, they accurately identified Capana06g002967 and Capana06g002969 as candidate genes for Rf[16]. Ro et al. (2023) employed genotyping-by-sequencing (GBS) and GWAS analysis to identify 57 SNPs significantly associated with anthracnose resistance through in 197 *C. chinense*[17]. In the study of important fruit traits, Du et al. (2019) employed the Target SNP-seq genotyping method to locate nine significantly associated loci with the fruit shape index among 271 pepper varieties, located on chromosomes 1, 2, 3, 4, 6 and 12[18]. Nimmakayala et al. (2021) identified 43081 SNPs related to various fruit traits, with 12 SNPs associated with the number of locules and 8 SNPs linked to other fruit shape traits using genotyping by sequencing. GWAS analysis unveiled that SNPs in genes such as CLAVAT1, WD-40, Auxin receptor, AAA type ATPase family protein, and RNA polymerase III serve as primary markers for the number of locules, and others such as subunit of exocyst complex 8, enhancer of ABA co-receptor 1, tetratricopeptide-repeat thioredoxin-like 3 and pleiotropic drug resistance proteins are associated with fruit shape. Notably, CLAVAT1, WD-40 and the auxin receptor gene are established as known genes influencing tomato fruit shape[19]. Lee et al. (2022) conducted a comprehensive candidate gene study on the main fruit traits of 351 pepper varieties using a combined approach of GWAS and QTL, successfully identifying 187 significant SNP loci related to fruit traits. Cross-validated of 17 QTLs related to fruit traits with GWAS results enabled the identification of 16 candidate genes associated with fruit traits[20].It is evident that the identification of SNP loci and the number of candidate genes substantially vary based on the materials employed and the sequencing methods applied. Importantly, numerous genes influencing key agronomic traits in peppers are yet to be discovered.

In Hainan, China, numerous feral peppers have been introduced, alongside cultivated varieties of *C. annuum* L. Preliminary identifications suggest variations in fruit length and size, yet all exhibit the characteristic traits of the *C. annuum* L. variety, including upright growth and heightened spiciness. Despite this,comprehensive genomic information, population relationships, and geographic distribution remain unclear[21,22]. This study endeavors to resequence 107 local feral peppers from Hainan and 75 cultivated *C. annuum* L. varieties. Through the analysis of the genetic diversity of these 182 materials, the research aimed to elucidate the distribution relationships of local feral peppers in Hainan and utilizing GWAS, pinpoint candidate genes associated with pivotal agronomic traits. This work lays the groundwork for future gene exploration and molecular breeding efforts.

Results

Diversity assessment of major agronomic traits

This study systematically investigated 26 agronomic traits in peppers, classifying them into 12 qualitative and 14 quantitative traits (Table S1-S3). The Shannon-Wiener index for qualitative traits, it ranged from 0.03 to 0.95 (average 0.49), while for quantitative traits, it ranged from 0.95 to 2.07 (mean 1.82). Except for locule number, the Shannon-Wiener index for quantitative traits surpassed that of qualitative traits, signifying a more extensive diversity in quantitative traits. Correlation analysis of the 14 quantitative traits (Fig. 1) revealed significant correlations, mostly highly significant, between quantitative traits of *C. frutescens* and *C. annuum* germplasm resources. A total of 57 pairs of traits exhibited highly significant correlations, with 39 pairs positively correlated and 18 pairs negatively correlated. Furthermore, 10 pairs of traits showed significant correlations, with 9 pairs positively correlated and 1 pair negatively correlated. Notably, the highest correlation coefficient observed was 0.93, corresponding to the relationship between weight per fruit and the longitudinal diameter of the commercial fruit (Table S4). The majority of the 26 agronomic traits demonstrated substantial phenotypic variations, affirming the suitability of the population for GWAS analysis.

Whole genome resequencing analysis

A total of 107 feral peppers sample and 75 bell pepper samples underwent re-sequencing, revealing a GC content ranging from 34.01–35.36%. The Q20 proportion fell between 96.14% and 98.6%, while Q30 exceeded 90%. Mapping rates varied from 96.85–99.53%, with an average of 99.02%, and an average sequencing depth of 9.62X. The average 1X coverage (minimum 1 base

coverage) was 90.47%, and the average 5X coverage (minimum 5 base coverage) was 76.79% (Table S5). These results attest to the high sequencing and gene mapping quality, establishing a robust foundation for subsequent analyses.

Employing the Genome Analysis Toolkit (GATK)for SNP detection yielded a total of 64,110,473 identified SNPs. The distribution of these variations was visually represented using a circos plot (Fig. 2). Uniform distribution of SNPs and Indels across chromosomes underscired indicating the reliability of the SNP and Indel data.

Population structure and chain imbalance analysis

The population structure analysis yields insights into the subdivision of the population into distinct subgroups. Integrating population structure analysis as a covariate proves effective in mitigating false positives resulting from population structure in GWAS analysis. Assuming the population can be divided into K = 2 ~ 10 subgroups, when K = 10 (Figure S2, Table S6), the lowest cv value is observed, signifying the optimal subdivision of all pepper germplasm resources into 10 groups (Fig. 3d). Conversely, at K = 1, the population remains undivided; at K = 2, clear stratification emerges, indicating the existence of two subgroups within the 182 pepper resources under study. Subgroup 1, denoted in red, comprises 107 materials, all of *C.frutescens*,while Subgroup, represented in blue, consists of 75 materials, exclusively of *C.annuum*. This aligns with the results of the phylogenetic tree and principal component analysis.

The construction of a phylogenetic tree utilizing the Neighbor-Joining method visually presents the distinctions within the population and the proximity of genetic relationships. The tree illustrates a division of the 182 individuals into 2 branches (Fig. 3b), consistent with the population structure analysis, predominantly segregating the population into two major groups: group 1 and group 2. A subsequent phylogenetic analysis (Fig. 3c) of 107 Hainan feral peppers (*C. frutescens*) reveals that all Hainan feral peppers from Baisha, Hainan (group3), Wenchang, Hainan (group7) and Haikou, Hainan (group6) from distinct branches. While feral peppercorns within the same city and county may share transmission routes, leading to minimal genetic diversity evolutionary distinctions persist among feral peppercorns from different cities and counties, undercoring the high genetic diversity of Hainan feral peppercorns in geographic distribution. Principal component analysis (PCA) classifies the 182 pepper germplasms into two major groups, *C. frutescens* and *C. annuum* (Fig. 3f). PC1, PC2, and PC3 elucidate 77.39%, 3.02%, and 2.09% of the genetic variation, respectively. *C.annuum* from external regions clusters densely, indicating a close genetic relationship among the pepper germplasms. Conversely, *C.frutescens* from diverse location within Hainan displays greater dispersion, reflecting a rich genetic background. Evaluation of genetic relationships based on selected SNP markers (Fig. 3e) reveals genetic relationship values, with meterials exhibiting of -0.2 accounting for 48.45%, those between 0 and 1.2 constituting 29.86%, and those exceeding 1.2 making up 21.68%. These findings affirm the high genetic diversity of the germplasm resources used, conducive to a robust whole-genome association study with minimal interference.

Calculation of the linkage disequilibrium (LD) decay of the 182 pepper materials (Fig. 3a) demonstrates a decline in R2 with increasing physical distance. At an R2 of 0.1, the LD decay distance for group 1 exceeds 1000 kb, while for group 2, it surpasses 500 kb. Notably, group 1 aligns with *C.frutescens*, and group 2 with *C.annuum*, indicating a faster LD decay rate for *C.annuum*, with the smallest LD decay distance observed in the *C.frutescens* population. This further substantiates the high genetic diversity within the *C.frutescens* population.

Genome-wide association study of 26 agronomic traits

Utilizing a subset of highly consistent SNPs derived from 182 pepper materials, we conducted a GWAS for 26 agronomic traits (Table 1). To enhance the reliability of GWAS results and mitigate the influence of population structure, we employed two statistical models: the Linear Mixed Model (LMM) and the Efficient Mixed Model Association eXpedited (EMMAX). After comparing their performance performance through Q-Q plots, we identified the most suitable statistical model for each trait. The LMM proved optimal for 9 traits, while EMMAX was optimal for 17 traits;for instance, fruit peduncle length analysis favored the LMM model (Fig. 4, Figure S3). These results underscore the necessity of selecting the most appropriate model for each trait in GWAS analysis, as the optimal model may vary.

Table 1
Screening loci and gene counts for pepper traits identified through GWAS.

Trait	Model	Chromosome	Number of SNPs	gene
Plant height	EMMAX	NC_029977.1	13	26
		NC_029978.1	3	
		NC_029980.1	1	
		NC_029981.1	5	
		NC_029982.1	2	
		NC_029984.1	1	
		NC_029985.1	13	
		NC_029986.1	4	
		NC_029987.1	1	
		NC_029988.1	1	
		NW_015960393.1	1	
		NW_015960461.1	1	
		NW_015960498.1	2	
		NW_015961641.1	1	
Plant breadth	LMM	NC_029978.1	2	41
		NC_029979.1	24	
		NC_029980.1	2	
		NC_029982.1	11	
		NC_029985.1	6	
		NC_029986.1	2	
		NW_015961225.1	8	
Leaf length	EMMAX	NC_029987.1	10	6
		NC_029988.1	1	
Leaf width	EMMAX	NC_029980.1	10	10
Petiole length	EMMAX	-	-	-
First stem node length/cm	LMM	NC_029978.1	1	20
		NC_029979.1	1	
		NC_029982.1	10	
		NC_029988.1	10	
		NW_015960959.1	1	
Stem thickness	EMMAX	NC_029977.1	3	7
		NC_029980.1	3	
		NC_029981.1	3	

Trait	Model	Chromosome	Number of SNPs	gene
		NC_029985.1	2	
		NC_029986.1	1	
Weight per fruit	LMM	NC_029977.1	5	46
		NC_029979.1	12	
		NC_029980.1	5	
		NC_029981.1	1	
		NC_029982.1	4	
		NC_029983.1	22	
		NC_029987.1	3	
		NC_029988.1	188	
		NW_015961202.1	1	
		NW_015961228.1	2	
Longitudinal diameter of fruit	LMM	NC_029977.1	7	62
		NC_029979.1	8	
		NC_029982.1	5	
		NC_029983.1	7	
		NC_029985.1	2	
		NC_029987.1	6	
		NC_029988.1	37	
		NW_015960681.1	3	
		NW_015960900.1	1	
Transverse diameter of fruit	LMM	NC_029979.1	12	28
		NC_029980.1	11	
		NC_029982.1	2	
		NC_029988.1	3	
Fruit peduncle length	LMM	NC_029977.1	5	64
		NC_029979.1	18	
		NC_029980.1	27	
		NC_029981.1	3	
		NC_029982.1	1	
		NC_029983.1	10	
		NC_029985.1	2	
		NC_029986.1	1	
		NC_029987.1	18	
		NC_029988.1	2	

Trait	Model	Chromosome	Number of SNPs	gene
		NW_015960759.1	7	
Thickness of flesh	EMMAX	-	-	-
Placenta size	LMM	NC_029979.1	7	50
		NC_029987.1	114	
Number of locules	EMMAX	NC_029977.1	2	31
		NC_029979.1	4	
		NC_029980.1	1	
		NC_029981.1	4	
		NC_029983.1	3	
		NC_029984.1	10	
		NC_029985.1	2	
		NC_029986.1	9	
		NC_029988.1	4	
		NW_015960458.1	1	
		NW_015961180.1	6	
		NW_015961815.1	1	
Branching type	EMMAX	-	-	-
Leaf surface	EMMAX	-	-	-
Leaf shape	EMMAX	-	-	-
Style length	EMMAX	NC_029977.1	3	26
		NC_029979.1	5	
		NC_029983.1	2	
		NC_029984.1	15	
		NC_029987.1	1	
		NW_015961188.1	1	
Flower color	EMMAX	-	-	-
Anther color	LMM	NC_029977.1	10	94
		NC_029979.1	5	
		NC_029980.1	8	
		NC_029981.1	3	
		NC_029982.1	14	
		NC_029983.1	11	
		NC_029985.1	30	
		NC_029986.1	9	
		NC_029987.1	2	

Trait	Model	Chromosome	Number of SNPs	gene
		NC_029988.1	4	
		NW_015960838.1	8	
		NW_015961196.1	2	
		NW_015961290.1	5	
		NW_015961339.1	9	
Pendage at blossom end	EMMAX	NC_029979.1	3	8
		NC_029981.1	1	
		NC_029983.1	1	
		NC_029984.1	5	
		NC_029985.1	2	
		NC_029987.1	1	
Color of immature fruit	EMMAX	-	-	-
Color of mature fruit	LMM	-	-	-
Persistent calyx at base of fruit	EMMAX	-	-	-
Fruit shoulder shape	EMMAX	-	-	-
Fruit shape	EMMAX	-	-	-

The number of highly significant SNPs and associated genes varied across traits and chromosome. Among the 26 agronomic traits, 15 traits exhibited 929 significantly associated loci, while the remaining 11 traits showed no significant associations, resulting in a total of 519 genes (Table S7).

For instance, plant height was associated with a total of 49 SNPs (26 genes), predominantly distributes across chromosomes NC_029977.1, NC_029978.1, NC_029980.1, NC_029981.1, NC_029982.1, NC_029984.1, NC_029985.1, NC_029986.1, NC_029987.1, NC_029988.1, NW_015960393.1, NW_015960461.1, NW_015960498.1, and NW_015961641.1. Similarly, plant breadth exhibited 55 SNPs (41 genes), primarily located on chromosomes NC_029978.1, NC_029979.1, NC_029980.1, NC_029982.1, NC_029985.1, NC_029986.1, and NW_015961225.1. Stem thicknessness was associated with 12 SNPs(7 genes), mainly distributed on chromosomes NC_029977.1, NC_029980.1, NC_029981.1, NC_029985.1, and NC_029986.1. The first stem node length ,showed association with 23 SNPs (20 genes), primarily distributed on chromosomes NC_029978.1, NC_029979.1, NC_029982.1, NC_029988.1, and NW_015960959.1. Leaf length exhibited association with 11 SNPs (6 genes), primarily located on chromosomes NC_029987.1 and NC_029988.1, while leaf width displayed association with 10 SNPs (10 genes), mainly distributed on chromosome NC_029980.1. The style length was associated with 27 SNPs (26 genes), predominantly kurtosis on chromosomes NC_029977.1, NC_029979.1, NC_029983.1, NC_029984.1, NC_029987.1, and NW_015961188.1. Anther color exhibited association with 120 SNPs (94 genes), mainly distributed on chromosomes NC_029977.1, NC_029979.1, NC_029980.1, NC_029981.1, NC_029982.1, NC_029983.1, NC_029985.1, NC_029986.1, NC_029987.1, NC_029988.1, NW_015960838.1, NW_015961196.1, NW_015961290.1, and NW_015961339.1.

Moreover, SNP analysis identified a total of 243 SNPs (46 genes) associated with weight per fruit, mainly showing kurtosis on chromosomes NC_029977.1, NC_029979.1, NC_029980.1, NC_029981.1, NC_029982.1, NC_029983.1, NC_029987.1, NC_029988.1, NW_015961202.1, and NW_015961228.1. Additionally, 76 SNPs (62 genes) were associated with the longitudinal diameter of fruits, primarily distributed on chromosomes NC_029977.1, NC_029979.1, NC_029982.1, NC_029983.1, NC_029985.1, NC_029987.1, NC_029988.1, NW_015960681.1, and NW_015960900.1. Further, 28 SNPs (28 genes) were associated with the transverse diameter of fruits, predominantly distributed on chromosomes NC_029979.1, NC_029980.1, NC_029982.1, and NC_029988.1. Additionally, 94 SNPs (64 genes) were associated with fruit peduncle length, mainly distributed on chromosomes

NC_029977.1, NC_029979.1, NC_029980.1, NC_029981.1, NC_029982.1, NC_029983.1, NC_029985.1, NC_029986.1, NC_029987.1, NC_029988.1, and NW_015960759.1. Furthermore, 121 SNPs (50 genes) were associated with placenta size, primarily distributed on chromosomes NC_029979.1 and NC_029987.1. Additionally, 47 SNPs (31 genes) were associated with the number of locules, mainly distributed on chromosomes NC_029977.1, NC_029979.1, NC_029980.1, NC_029981.1, NC_029983.1, NC_029984.1, NC_029985.1, NC_029986.1, NC_029988.1, NW_015960458.1, NW_015961180.1, and NW_015961815.1. Similarly, 13 SNPs (8 genes) were associated with Pendage at blossom end, mainly distributed on chromosomes NC_029979.1, NC_029981.1, NC_029983.1, NC_029984.1, NC_029985.1, and NC_029987.1.

Gene Function Annotation

In the screened genes, we identified 26, 41, 6, 10, 20, 7, 46, 62, 28, 64, 50, 31, 26, 94, and 8 genes associated with plant height, plant width, leaf length, leaf width, first stem node length, stem thickness, weight per fruit, longitudinal diameter of fruits, transverse diameter of fruits, fruit peduncle length, placenta size, number of locules, style length, anther color, and pendage at blossom end, respectively (Table S8). Subsequently, gene functional annotation was conducted for these identified genes.

Genes associated with stem-related traits were primarily implicated in functions such as signal transduction mechanisms, replication, recombination and repair, transcription, posttranslational modification, protein turnover, chaperones, intracellular trafficking, secretion, and vesicular transport. Among these, 26 genes had an unknown function, and 10 genes lacked functional annotations (Fig. 5a). Regarding leaf-related traits, the annotated genes were predominantly involved in functions such as translation, ribosomal structure and biogenesis, signal transduction mechanisms, and posttranslational modification, protein turnover, chaperones, with 2 genes of unknown function and 3 genes not functionally annotated (Fig. 5b). Genes associated with flower-related traits were predominantly involved in functions such as posttranslational modification, protein turnover, chaperones, replication, recombination and repair, transcription, and signal transduction mechanisms, with 38 genes of unknown function and 12 genes lacking functional annotations (Fig. 5c). Finally, genes associated with fruit-related traits were mainly implicated in functions such as replication, recombination and repair, signal transduction mechanisms, posttranslational modification, protein turnover, chaperones, transcription, and carbohydrate transport and metabolism, with 92 genes of unknown function and 48 genes lacking functional annotations(Fig. 5d).

Candidate gene screening and validation

The boxplots depicting four traits—weight per fruit, longitudinal diameter of fruits, transverse diameter of fruits, and placenta size—are illustrated in Fig. 6a-d. The numerical values of these four traits in *C. annuum* surpass those in *C. frutescens*. This, in conjunction with the qRT-PCR results from both *C. annuum* and *C. frutescens*, facilitates the analysis functional trends in genes.

Leveraging the outcome of gene functional annotation, we randomly selected 11 candidate genes associated with weight per fruit, longitudinal diameter of fruits, transverse diameter of fruits, and placenta size for qRT-PCR validation experiments shown, as depicted in Fig. 6e-h. These aimed to substantiate the influence of these genes on the growth of various traits. Notably, the expression level of the TPX2 gene in *C. frutescens*, characterized by smaller weight per fruit, significantly exceeded that in *C. annuum*, indicating its potential role in inhibiting weight per fruit in both species. Similarly, nsLTPs13 and NUTCRACKER displayed significantly higher expression levels in the variety with smaller weight per fruit, suggesting their potential as genes inhibiting weight per fruit in both *C. frutescens* and *C. annuum*. In the context of the longitudinal diameter of commercial fruits, the expression levels of extensin-2-like and COP10 genes were significantly higher in *C. annuum* compared to *C. frutescens*, whereas SCPL13 exhibited significantly higher expression in *C. frutescens.* This implies that these extensin-2-like and COP10 may promote longitudinal diameter of fruits in both species, while SCPL13 may have an inhibitory role. Notably, no significant difference in the expression level of the extensin-1-like gene was observed between the two varieties. Examining the transverse diameter of fruits, the expression of the DDB1 gene in *C. annuum* with a larger diameter was significantly higher than that in *C. frutescens* with a smaller diameter, suggesting its potential role in promoting the transverse diameter of fruits in both species Conversely, the gene GAUT1 exhibited opposite expression levels in the two varieties, indicating its potential as a gene inhibiting this trait. No significant differences in the expression levels of the RKD4 gene were observed between the two varieties. For placenta size, SCPL51 displayed significantly higher expression in the smaller seed cavity of *C. frutescens* compared to the larger seed cavity of

C. annuum.In contrast, HDT2 and HDAC9 showed significantly higher expression in *C. frutescens*, indicating their potential roles as genes inhibiting placenta size in both varieties.

Discussion

Chili peppers exhibit notable variation in flower, leaf, and fruit-related traits, as well as diverse agronomic characteristics[23,24] Key yield and quality attributes in chili breeding are typically influenced by multiple genes or QTL[25]. primarily used genetic mapping of agronomic traits in chili peppers has mostly focused on QTL methods[26], the inherent low sequencing depth of this approach may lead to overlooking significant genotypes[27]. With advancements in high-throughput sequencing technology, GWAS has emerged as a potent tool for localizing candidate genes[28]. Despite the application of GWAS in chili breeding, most studies have focused on a limited number of traits or disease resistance[29–31]. In our study, GWAS was conducted on 26 agronomic traits of chili peppers, revealing significant SNPs associated with these traits. The results highlight the substantial impact of GWAS on efficiently localizing genes linked to multiple agronomic traits in chili peppers.

Agronomic traits in chili peppers encompass various aspects, including yield, quality, and resistance. Determinants of chili pepper yield include fruit longitudinal and transverse diameters, weight per fruit, and the fruit shape index. GWAS analysis is a valuable tool for localizing QTL and predicting candidate genes associated with fruit yield-related agronomic traits. This approach provides a theoretical basis for molecular marker-assisted breeding of chili peppers. In tomato research, next-generation sequencing sequencing technology has facilitated the identification of representative gene families controlling fruit size. Notably, genes such as CNR, SUN, and OVATE, belonging to these families, have been successfully cloned. Their roles in regulating fruit length and shape have been validated[32–35]. The insights gained from tomato studies have been applied to QTL localization in chili peppers, revealing the complex genetic structure controlling these quantitative genetic traits[36,37]. Conducting a multi-trait QTL analysis in chili pepper has identified 40 candidate genes associated with *C. annuum* traits. These genes are implicated in diverse functions, including defense response, metabolic processes oxidation-reduction, and phosphorylation[38]. Understanding the genetic basis of these traits is critical, as cell number and size are pivotal determinants of plant organ size, and any variation in these parameters can impact organ size significantly[39]. Fruit peduncle length, a crucial trait, is intricately regulated by cell number or size, influenced indirectly by hormones and multiple pathways. Kinases play essential roles in this regulation, affecting plant growth and development. In our study, SNPs related to fruit stalks were identified in key genes, including leucine-rich repeat receptor-like kinases (LRR-RLKs), serine/threonine protein kinase, ABC transporter gene, and RING finger protein. These genes are likely to play pivotal roles in growth, development, cell wall integrity, and elongation rate[40–43].Our gene functional annotation identified genes related to plant height, first internode length, fruit length, and fruit stalk length. For instance, nsLTPs13 plays a crucial role in cell development and organogenesis[44,45], while TPX2 family members are involved in various plant developmental processes[46] NUTCRACKER is implicated in defining asymmetric cell division and stabilizing tissue boundaries [47], and genes like SCPL13, regulate cell elongation and locule numbers in rice[48], Arabidopsis[49] and tobacco[50].

The extensin-2-like protein actively participates in plant cell wall formation and assembly, influencing plant cell elongation functions[51]. COP10 primarily regulates plant growth and development,with a role in light signal transduction[52,53]. The gene DDB1, associated with commercial fruit transverse diameter, acts as a damage sensor maintaining the balance between genome integrity and cell cycle progression[6]. Galacturonosyltransferase (GAUT)1 and GAUT7 are core components of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex[54]. SCPL51, associated with seeds per fruit, regulates cell elongation and carpel numbers[49]. Members of the SCP and SCPL families, expressed in major tissue types, play roles in various biochemical and cellular processes in plants[55–57]. Studies have revealed that overexpression of SCPL41 reduces membrane lipid content, while is absence of SCPL41 increases membrane lipid content[58]. HDT1/2 influences the transition from cell division to expansion in root tips by inhibiting the expression of GA2ox2[59]. The regulation of histone acetylation involves the antagonism between HDAC and histone acetyltransferases (HAT), modulating enzyme activity[60].

Our study involved the investigation of 929 significantly associated SNPs in 182 *C. frutescens* and *C. annuum*. chili pepper resources, resulting in the identification of 519 candidate genes after meticulous screening. Through qRT-PCR validation, genes such as SCPL13, extensin-1-like, and DDB1, associated with fruit-related traits in *C. frutescens* and *C. annuum*, were confirmed.

Future endeavors will focus on identifying candidate genes and the subsequent cloning and functional analysis of genes regulating fruit length. These efforts aim to provide a solid foundation for the development of high-quality chili pepper varieties.

Conclusions

This study conducted a comprehensive genome-wide analysis of a population comprising 182 chili pepper resources, encompassing both *C. frutescens* and *C. annuum*, through whole-genome resequencing. The analysis yieded a set of 64,110,473 high-quality SNP markers, addressing 26 key agronomic traits, including plant height, weight per fruit, fruit length, and seed cavity size. Utilizing GWAS, we identified 929 significant SNP-associated loci, unveiling 519 candidate genes. These genes play pivotal roles in plant growth and development, regulation of light signal transduction, and responses to salt stress. Furthermore, through meticulous gene annotation and assessment of expression patterns, certain candidate genes, namely SCPL13, extensin-1-like, and DDB1,were delineated. While these findings contribute significantly to understanding of genes related to chili pepper fruit traits, elucidating the precise mechanisms governing these traits requires further investigation. This study provides valuable genetic resources and theoretical foundation for future endeavors in molecular marker-assisted breeding of chili peppers.

Methods and material

Plant materials and treatment

In this study, the pepper research group at the School of Tropical Agriculture and Forestry, Hainan University. Experimental samples of hot peppers, including the collection of plant materials, were collected according to relevant institutions, national and international guidelines and legislation, with the appropriate permission of Hainan Provincial Department of Science and Technology. Systematically acquired a total of 182 samples comprising both *Capsicum annuum* and *Capsicum frutescens* (Table S9). Specifically, 107 instances of *Capsicum frutescens* were procured from diverse cities and counties in Hainan Province, It is a pre-collection of chili peppers distributed in villages, forests and ravines all over Hainan, while the remaining 75 specimens of *Capsicum annuum* originated from cultivated resources in other provinces. Notably, all specimens exhibited distinctive features of tall plants and upward growth of fruits. To ensure methodological consistency,all pepper plants were cultivated in a randomized design with three replicates at Sanjia Town, Dongfang City, Hainan Province, during September 2022. Individual plant samples were tagged and tender leaves were wrapped in aluminum foil, frozen in liquid nitrogen, and stored for further use.

DNA extraction and sequencing

The genomic DNA from tender leaves was extracted using the cetyltrimethylam-monium bromide (CTAB) method. Subsequent to extraction, the concentration and quality of the genomic DNA were determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). Following the DNA quality assessments, libraries were constructed. These qualified libraries then underwent sequencing on the NovaSeq 6000 platform. The raw reads obtained from the sequencing process underwent a series of essential data filtering steps: (1) the removal of reads containing adapters, (2) filtering out reads with more than 10% N content, and (3) elimination of reads with low-quality bases exceeding 50% at a quality value below 20. This filtering process yielded clean reads, which were utilized for subsequent data analysis.

SNP and Indel calls

The Zunla-1 genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_000710875.1), characterized by a size of approximately 2.9 Gb and a GC content of 34.5%, was used as the reference genome. Alignment of Clean Reads to the reference genome employed the Burrows-Wheeler Alignment (bwa-mem2 v2.2)[61] with the MEM algorithm. The identification and removal of redundant reads were executed based on the alignment results, employing samtools (v1.9)[62]. Subsequent to alignment, the calling of SNPs and Indels was performed using the HaplotypeCaller module of GATK (v3.8). Filtering criteria encompassed parameters: QUAL < 30, QD < 2.0, MQ < 40, FS > 60.0, MQrankSum < -12.5, ReadPosRankSum < -8.0, -clusterSize 2, -clusterWindowSize 5. Further filtering of the identified SNPs was conducted based on minor allele frequency (MAF: 0.05) and site integrity (INT: 0.8), ensuring the acquisition of highly consistent SNP positions.

Population evolutionary analysis

The MEGA X[63] software was utilized to construct phylogenetic trees for 182 species using the neighbor-joining method, and applying the Kimura 2-parameter model with 1000 bootstrap repetitions. The analysis of population structure for the investigated materials utilized the admixture[64] software based on SNP data. Clustering, spanning a predefined range of subgroups (K values: 1–10), determined the optimal number of subgroups based on the valley of the cross-validation error (CV error).To facilitate evolutionary analysis by revealing clustering patterns, principal component analysis was executed using the EIGENSOFT25 software, leveraging SNP data. Estimation of relateness between individuals in natural population was carried out using the GCTA[65] software. Additionally, the calculation of linkage disequilibrium between pairwise SNPs within a specified distance range (1,000 kb) on the same chromosome was accomplished with the PopLDdecay(v3.41)[66] software,

GWAS

Filtered SNPs underwent GWAS. Wherein all traits underwent comprehensive analysis utilizing GEMMA (LMM model) and EMMAX (EMMAX model) software, incorporating both phenotype and genotype data. The formulations for these models are as follows:

LMM model formula: $$y = W\alpha + X\beta + X\mu + \mathrm{e}\ (1)$$

EMMAX model formula: $$y_i = \beta_0 + \beta_k X_{ik} + \eta_{i\bar{k}} (2)$$

Validation of candidate genes using real-time quantitative PCR (qRT-PCR)

Statistical analysis of data,encompassing box plots and qRT-PCR, was conducted aceoss a spectrum of 20 varieties. This selection comprised both the top and bottom 10 with respect to the numerical values associated with each trait. Total RNA underwent reverse transcription to cDNA following the protocols outlined in the Roche reverse transcription reagent kit. Primer sequences corresponding to genes such as SCPL13, COP10, and DDB1 (Table 2) were designed using Premier 5.0 software, with Actin (Accession number: AY486137.1) serving as the reference gene. Subsequent qRT-PCR reactions utilized SYBR® Premix Ex TaqTM, and the ensuring data analysis adhered to the $2^{-\Delta\Delta CT}$ method.

Table 2
qRT-PCR primer information

Gene ID	Name	Forward (5’ to 3’)	Reverse (5’ to 3’)
gene-LOC107878351	nsLTPs13	TGTAGATGGAACCGTTGAGAA	TACGAGACTGGCGATAATGAG
gene-LOC107878353	TPX2	TGAGGAGGAAATGATGGCTA	GGAGGTCGTGCTCTAAGTGA
gene-LOC107878392	NUTCRACKER	AATATGGGTGTGGGTCAAGAG	GATGGAATAAGGGTAGTCGTGT
gene-LOC107851634	SCPL13	CTTCCTTTTTATCTTGAGACCG	AGCCAATGGACCTACTTCGT
gene-LOC107854462	extensin-1-like	TTGTAGCCAGCCATGTTGTT	TTGAAGGAGCGGGTGATTT
gene-LOC107873714	extensin-2-like	AGCTTTGGAAATTTAGGGCA	GGGAGCATCATAGTTGGACG
gene-LOC107873729	COP10	CGTCGGTATCATCATCAGGT	CAAGAAAATAAATGCCACCC
gene-LOC107867065	DDB1	AAGTATCTCCGCTGCTATGTG	TAGTAGGTGAAGACCCCCGT
gene-LOC107867061	GAUT1	TCCATCTCTTTCGCTCACAT	CCATTCCTCGATTTTCCTAGT
gene-LOC107865044	RKD4	CAATCACTTGCCGATTTTTG	GCTCTTCTTATGCCTCCCAG
gene-LOC107847544	SCPL51	GGTTGGAAAACTAAAGTGGGA	GCTGGTGATTGTGTGATGCTA
gene-LOC107848037	HDT2	CCTAAAAGGGTTGAGGAGAAG	GTTGTTACCATTGGCTGCTG
gene-LOC107847639	HDAC9	ATGGAAAATCTTCGTTGCCT	TCGCCATCATAATACTCGTCA
LOC107873556	Actin	GCTGGAGGTGTATTTTTGGTT	TTGGCCCTGTCAGTCTTGTA

Statistical analysis

Data analysis involved the utilization of Microsoft Excel 2007 for data organization and the calculation of the frequency distribution for qualitative traits. Additionally, the maximum, minimum, and coefficient for variation of quantitative traits were calculated. Standard deviation (SD) and mean (Mean) served as classification criteria, with each 0.5 SD interval designated as a level "i." Levels were systematically categorized from i = 1 to 10, and the frequency distribution P_i for each level was determined. The computation of the Shannon-Weaver diversity index (H’) was undertaken using the following formula:

Data availability

The data stored in our system is accessible, but requires the author's permission. Information of relevant data can be obtained by contacting S.C. via email, 990865 @ hainanu.edu.cn.

Declarations

Acknowledgements

This research was funded by Key R&D Projects in Hainan Province (ZDYF2022XDNY159) and Hainan Provincial Natural Science Foundation of China (320RC474).

Author Contributions

G.F.: investigate, collate and analyze data ,writing-original draft, and writing—review and editing. S. Y., K.W. and M.Y.: methodology and software. Muhammad Ahsan Altaf, Q. D., Z. W., H. F., X. L.: English writing polishing. S.C.: conceptualization, supervision, funding acquisition, validation, resources, and writing—review and editing. All authors have read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional Information

Supplementary Information The online version version contains supplementary material available at

Correspondence and requests for materials should be addressed to S.C.

Reprints and permissions information is available at www.nature.com/reprints

Publisher' note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Callahan, K. L. Ancient mesoamerican civilization. Sci. 143, 531–537 (1964).
2. Qiao L., Zhao B., Zong Y., Kou C. & Dong Y. Development current situation,tendency,and countermeasure for China's pepper industry. China Vegetables. 11, 9–15 (2023)
3. Hwang, I. G., Shin, Y. J., Lee, S., Lee, J. & Yoo, S. M. Effects of different cooking methods on the antioxidant properties of red pepper (*Capsicum annuum* L.). Prev. Nutr. Food Sci. 17, 286–292 (2012).
4. Saleh, B. K., Omer, A. & Teweldemedhin, B. Medicinal uses and health benefits of chili pepper (*Capsicum spp.*): a review. MOJ Food Process Technol. 6, 325–328 (2018).
5. Persson, A. An unintended side effect of pepper spray: Gender trouble and "repair work" in an armed forces unit. Men Masc. 15, 132–151 (2012).
6. Yarnes, S. C. et al. Identification of QTLs for capsaicinoids, fruit quality, and plant architecture-related traits in an interspecific *Capsicum* RIL population. Genome. 56, 61–74 (2013).
7. Han, K. et al. An ultra-high-density bin map facilitates high-throughput QTL mapping of horticultural traits in pepper (*Capsicum annuum*). DNA Res. 23, 81–91 (2016).
8. Pang, Y. et al. High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. Mol. Plant. 13, 1311–1327 (2020).
9. Zhang, X. et al. Combined GWAS and QTL analysis for dissecting the genetic architecture of kernel test weight in maize. Mol. Genet. Genomics. 295, 409–420 (2020).
10. Wang, L. et al. GWAS reveals two novel loci for photosynthesis-related traits in soybean. Mol. Genet. Genomics. 295, 705–716 (2020).
11. Zhang, C., Wu, J. Y., Cui, L. W. & Fang, J. G. Mining of candidate genes for grape berry cracking using a genome-wide association study. J. Integr. Agric. 21, 2291–2304 (2022).
12. Liu, X. et al. Identification of novel loci and candidate genes for cucumber downy mildew resistance using GWAS. Plants (Basel). 9, 1659 (2020).
13. Yang, J. et al. Genome-wide association study of eigenvectors provides genetic insights into selective breeding for tomato metabolites. BMC Biol. 20, 120 (2022).
14. Colonna, V. et al. Genomic diversity and novel genome-wide association with fruit morphology in *Capsicum*, from 746k polymorphic sites. Sci Rep. 9, 10067 (2019).
15. Ahn, Y. K. et al. Whole genome resequencing of *Capsicum baccatum* and *Capsicum annuum* to discover single nucleotide polymorphism related to powdery mildew resistance. Sci Rep. 8, 5188 (2018).
16. Wu, L. et al. Genome-Wide Correlation of 36 agronomic traits in the 287 pepper (*Capsicum*) accessions obtained from the SLAF-seq-based GWAS. Int. J. Mol. Sci. 20, 5675 (2019).
17. Ro, N. et al. Genome-wide association study of resistance to anthracnose in pepper (*Capsicum chinense*) germplasm. Bmc Plant Biol. 23, 389 (2023).

18. Du, H. et al. Target sequencing reveals genetic diversity, population structure, core-SNP markers, and fruit shape-associated loci in pepper varieties. BMC Plant Biol. 19, 578 (2019).
19. Nimmakayala, P. et al. Exploration into natural variation for genes associated with fruit shape and size among *Capsicum chinense* collections. Genomics. 113, 3002–3014 (2021).
20. Lee, H. Y. et al. Uncovering candidate genes controlling major fruit-related traits in pepper via genotype-by-sequencing based QTL mapping and genome-wide association study. Front Plant Sci. 11, 1100 (2020).
21. Yang, L. et al. feral pepper resources in western rural areas of Hainan province. Res Sci. 32, 1608–1614 (2010).
22. Nong, S. et al. Investigation of *Capsicum annuum var*. conoidesin east central rural areas of Hainan province. Res Sci. 32, 2400–2406 (2010).
23. Sudré, C. P. et al. Genetic variability in domesticated *Capsicum spp* as assessed by morphological and agronomic data in mixed statistical analysis. GMR, Genet. Mol. Res. 9, 283–294 (2010).
24. Usman, M. G., Rafii, M. Y., Ismail, M. R., Malek, M. A. & Abdul Latif, M. Heritability and genetic advance among chili pepper genotypes for heat tolerance and morphophysiological characteristics. *Sci. World J.* 2014, 308042 (2014).
25. Chaim, A. B. et al. QTL mapping of fruit-related traits in pepper (*Capsicum annuum)*. Theor Appl Genet. 102, 1016–1028 (2001).
26. Barchi, L., Lefebvre, V., Sage-Palloix, A.M., Lanteri, S. & Palloix, A. QTL analysis of plant development and fruit traits in pepper and performance of selective phenotyping. Theor Appl Genet. 118, 1157–1171 (2009).
27. Pasaniuc, B. et al. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. Nat Genet. 44, 631–635 (2012).
28. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. Am. J. Hum. Genet. 90, 7–24 (2012).
29. Wu, L. et al. Mapping of CaPP2C35 involved in the formation of light-green immature pepper (*Capsicum annuum* L.) fruits via GWAS and BSA. Theor. Appl. Genet. 135, 591–604 (2022).
30. Nimmakayala, P. et al. Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum annuum* L. Sci. Rep. 6, 38081 (2016).
31. Tamisier, L. et al. 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 (2020).
32. Guo, M. & Simmons, C. R. Cell number counts–the fw2.2 and CNR genes and implications for controlling plant fruit and organ size. Plant Sci. 181, 1–7 (2011).
33. Chakrabarti, M. et al. A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proc. Natl. Acad. Sci. U. S. A.* 110, 17125–17130 (2013).
34. Rodríguez, G. R. et al. Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. Plant Physiol. 156, 275–285 (2011).
35. Soyk, S. et al. Bypassing negative epistasis on yield in tomato imposed by a domestication gene. Cell. 169, 1142–1155.e1112 (2017).
36. Ramchiary, N., Kehie, M., Brahma, V., Kumaria, S. & Tandon, P. Application of genetics and genomics towards *Capsicum* translational research.Plant Biotechnol. Rep. 8, 101–123 (2014).
37. Chunthawodtiporn, J., Hill, T., Stoffel, K. & Van Deynze, A. Quantitative trait loci controlling fruit size and other horticultural traits in bell pepper (*Capsicum annuum*). Plant Genome. 11, 1 (2018).
38. Lozada, D. N., Barchenger, D. W., Coon, D., Bhatta, M. & Bosland, P. W. Multi-locus association mapping uncovers the genetic basis of yield and agronomic traits in chile pepper (*Capsicum spp.*). *CBGG*. 4, 2 (2022).
39. Hwang, W. W. et al. A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. Mol Cell. 11, 261–266 (2003).
40. Guo, H. et al. Three related receptor-like kinases are required for optimal cell elongation in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7648–7653 (2009).
41. Gish, L. A. & Clark, S. E. The RLK/Pelle family of kinases. Plant J. 66, 117–127 (2011).

42. Gaedeke, N. et al. The *Arabidopsis thaliana* ABC transporter AtMRP5 controls root development and stomata movement. EMBO J. 20, 1875–1887 (2001).
43. Nieuwland, J. et al. Lipid transfer proteins enhance cell wall extension in tobacco. Plant Cell. 17, 2009–2019 (2005).
44. Ambrose, C., DeBono, A. & Wasteneys, G. Cell geometry guides the dynamic targeting of apoplastic GPI-linked lipid transfer protein to cell wall elements and cell borders in *Arabidopsis thaliana*. PLoS One. 8, e81215 (2013).
45. Du, P. et al. Genome-wide analysis of the TPX2 family proteins in *Eucalyptus grandis*. BMC Genomics. 17, 967 (2016).
46. Welch, D. et al. *Arabidopsis* JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action.*Genes Dev*. 21, 2196–2204 (2007).
47. Li, Y. et al. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat Genet. 43, 1266–1269 (2011).
48. Wen, J., Li, J. & Walker, J. C. Overexpression of a serine carboxypeptidase increases carpel number and seed production in *Arabidopsis thaliana*. Food Energy Secur. 1, 61–69 (2012).
49. Bienert, M. D., Delannoy, M., Navarre, C. & Boutry, M. NtSCP1 from tobacco is an extracellular serine carboxypeptidase III that has an impact on cell elongation. Plant Physiol. 158, 1220–1229 (2012).
50. Lamport, D. T., Kieliszewski, M. J., Chen, Y. & Cannon, M. C. Role of the extensin superfamily in primary cell wall architecture. Plant Physiol. 156, 11–19 (2011).
51. Suzuki, G., Yanagawa, Y., Kwok, S. F., Matsui, M. & Deng, X. W. *Arabidopsis* COP10 is a ubiquitin-conjugating enzyme variant that acts together with COP1 and the COP9 signalosome in repressing photomorphogenesis. *Genes Dev*. 16, 554–559 (2002).
52. Wei, N. et al. *Arabidopsis* COP8, COP10, and COP11 genes are involved in repression of photomorphogenic development in darkness. *Plant Cell*. 6, 629–643 (1994).
53. Iovine, B., Iannella, M. L. & Bevilacqua, M. A. Damage-specific DNA binding protein 1 (DDB1): a protein with a wide range of functions. Int. J. Biochem. Cell Biol. 43, 1664–1667 (2011).
54. Atmodjo, M. A. et al. Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. *Proc. Natl. Acad. Sci. U. S. A*. 108, 20225–20230 (2011).
55. Fraser, C. M., Rider, L. W. & Chapple, C. An expression and bioinformatics analysis of the *Arabidopsis* serine carboxypeptidase-like gene family. Plant Physiol. 138, 1136–1148 (2005).
56. Lehfeldt, C. et al. Cloning of the SNG1 gene of *Arabidopsis* reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism. Plant Cell. 12, 1295–1306 (2000).
57. Granat, S. J., Wilson, K. A. & Tan-Wilson, A. L. New serine carboxypeptidase in mung bean seedling cotyledons. J. Plant Physiol. 160, 1263–1266 (2003).
58. Chen, J., Li, W. Q. & Jia, Y. X. The serine carboxypeptidase-like gene SCPL41 negatively regulates membrane lipid metabolism in *Arabidopsis thaliana*. Plants (Basel). 9, 696 (2020).
59. Li, H. et al. Plant-specific histone deacetylases HDT1/2 regulate GIBBERELLIN 2-OXIDASE2 expression to control *Arabidopsis* root meristem cell number. Plant Cell. 29, 2183–2196 (2017).
60. Milazzo, G. et al. Histone deacetylases (HDACs): evolution, specificity, role in transcriptional complexes, and pharmacological actionability. Genes. 11, 556 (2020).
61. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 25, 1754–1760 (2009).
62. Danecek, P. et al. Twelve years of SAMtools and BCFtools. GigaScience. 10, giab008 (2021).
63. .Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 35, 1547–1549 (2018).
64. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664 (2009).
65. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82 (2011).

66. Zhang, C., Dong, S.S., Xu, J.Y., He, W.M. & Yang, T.L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 35, 1786–1788 (2018).

Figures

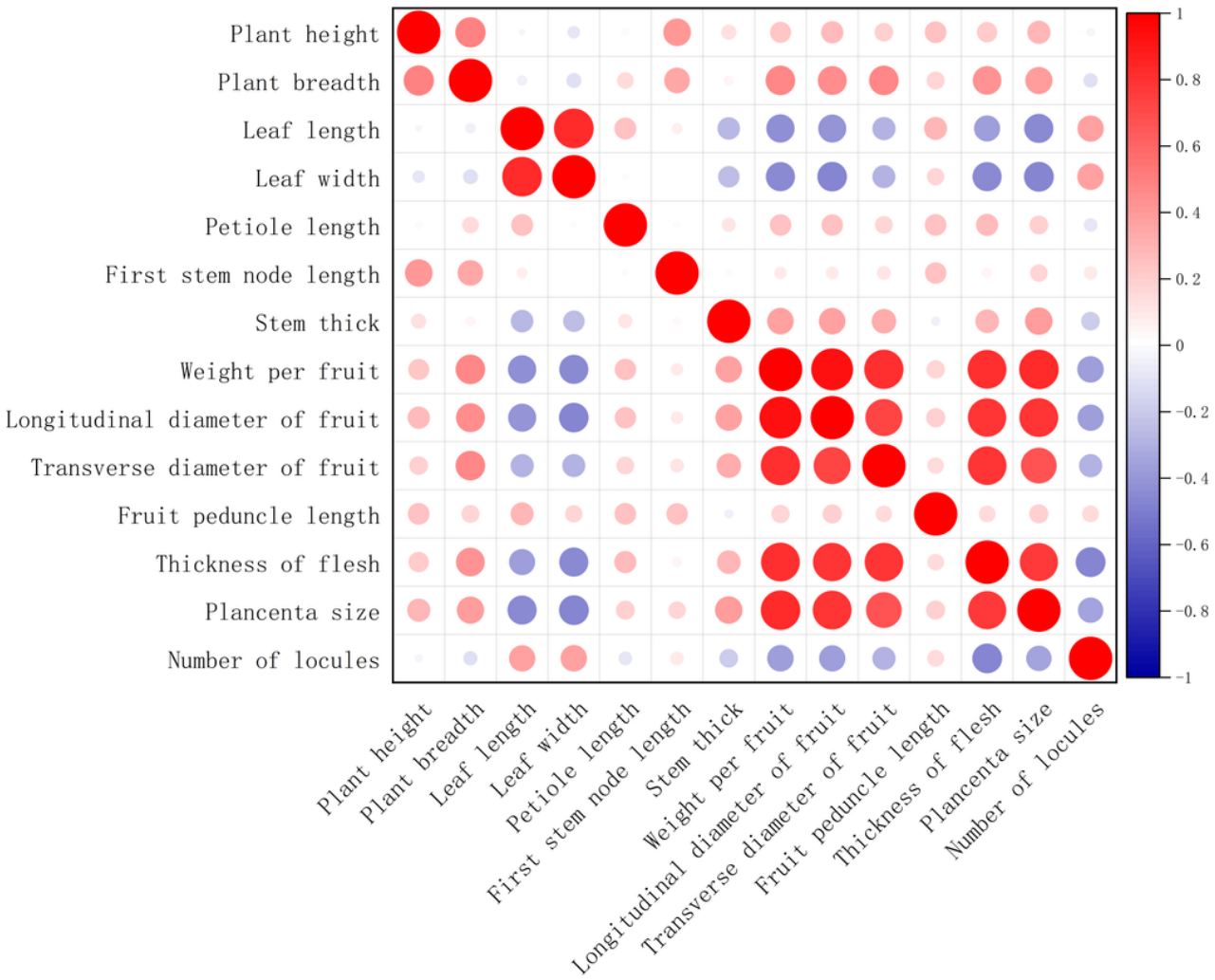

Figure 1

Correlation analysis of quantitative traits. Red indicates positive correlation, while blue indicates negative correlation. The intensity of color reflects strength of the correlation.

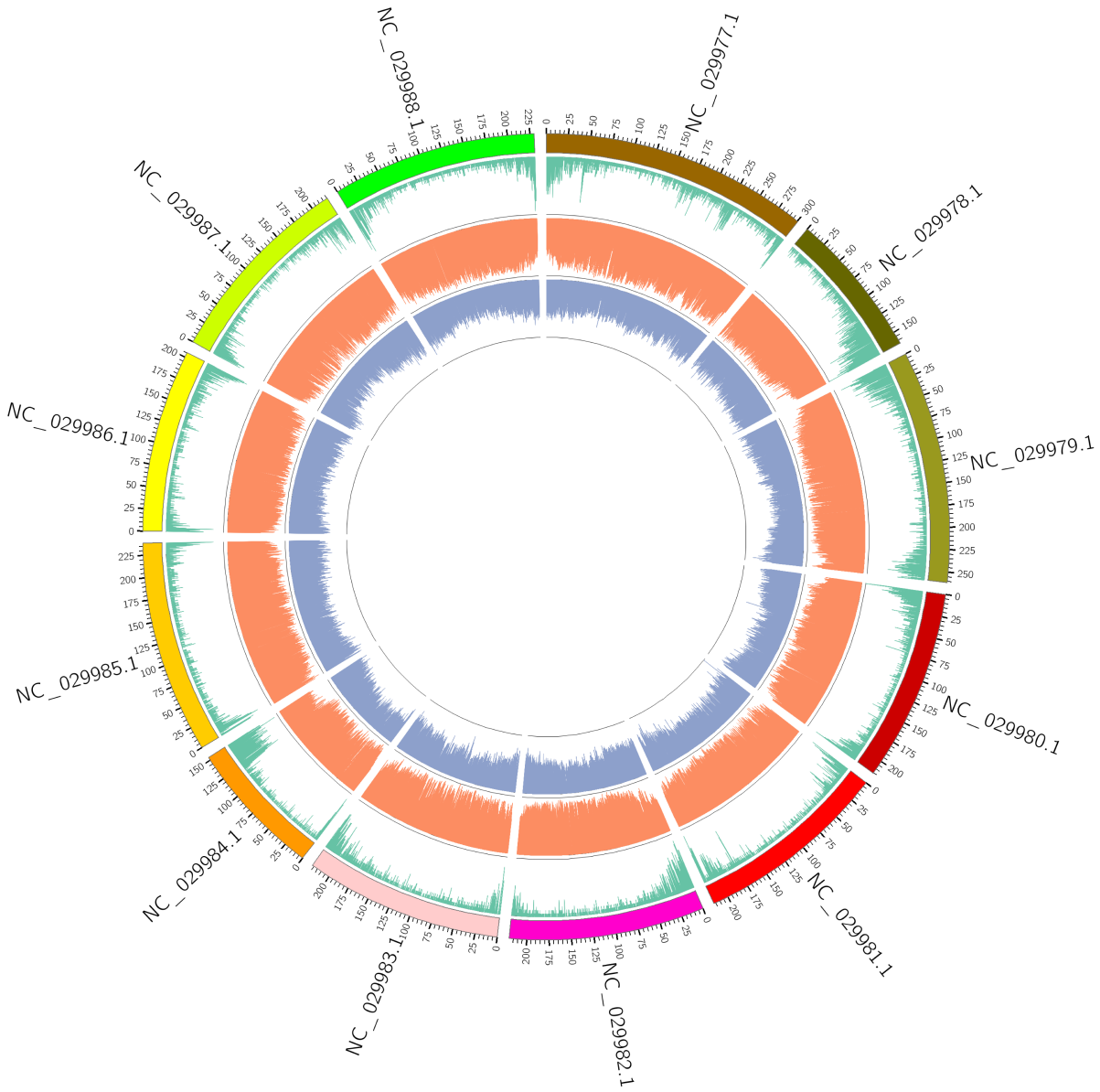

Figure 2

Distribution of variant types on chromosomes for each sample. Outer to inner:chromosome coordinates (colored squares), SNP density distribution (orange), and Indel density distribution (blue).

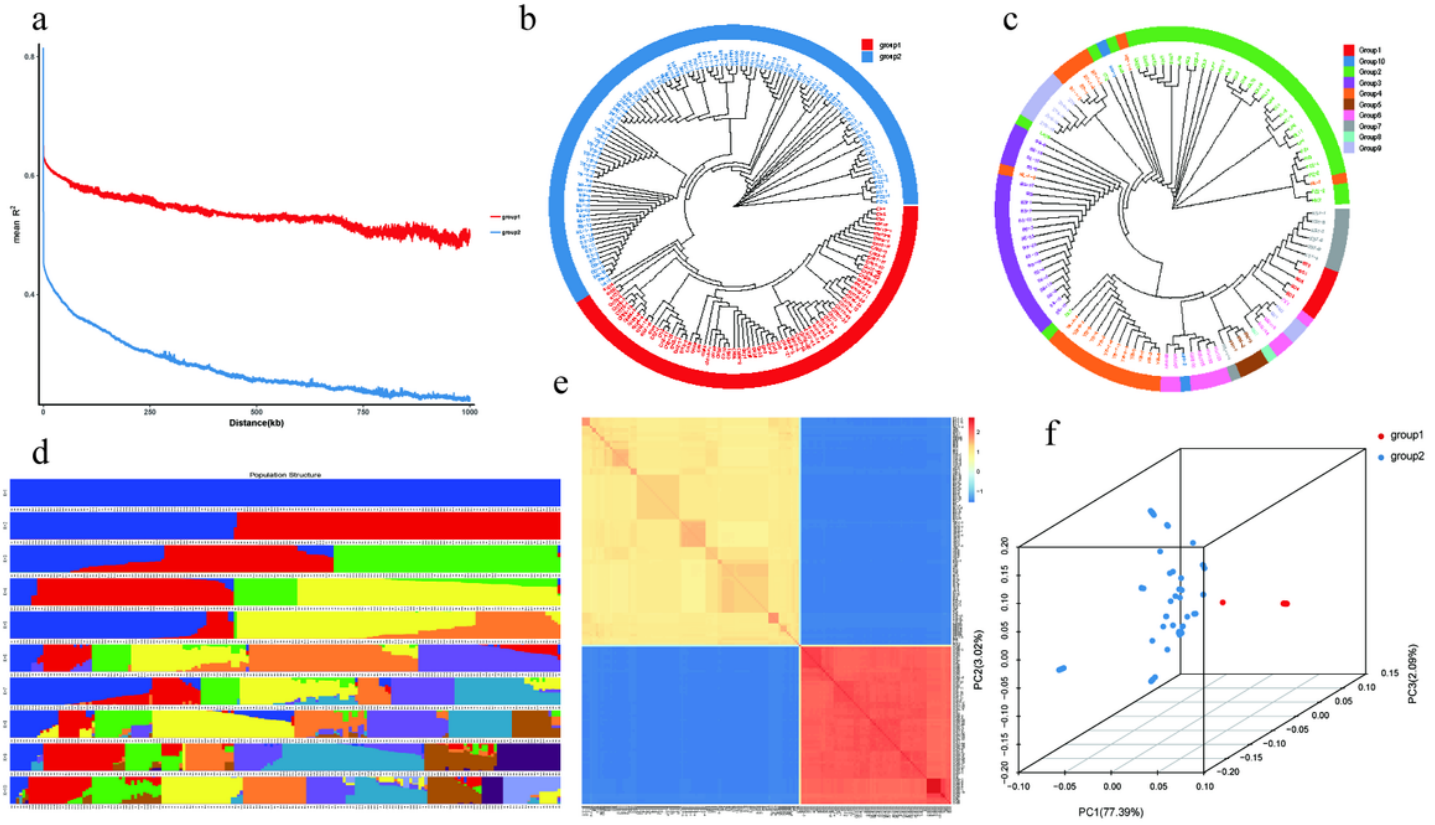

Figure 3

Population structure and chain imbalance analysis of *C. annuum* and *C. frutescens*. (a) Linkage disequilibrium analysis. (b) Phylogenetic analysis of 182 *C. annuum* and *C. frutescens*. (c) Phylogenetic analysis of 107 Hainan native escaped pepper populations (*C. frutescens*). (d) Population structure analysis. (e) Kinship analysis (f) Principal component analysis. Blue in A, B and F represents group 1; red represents group 2.

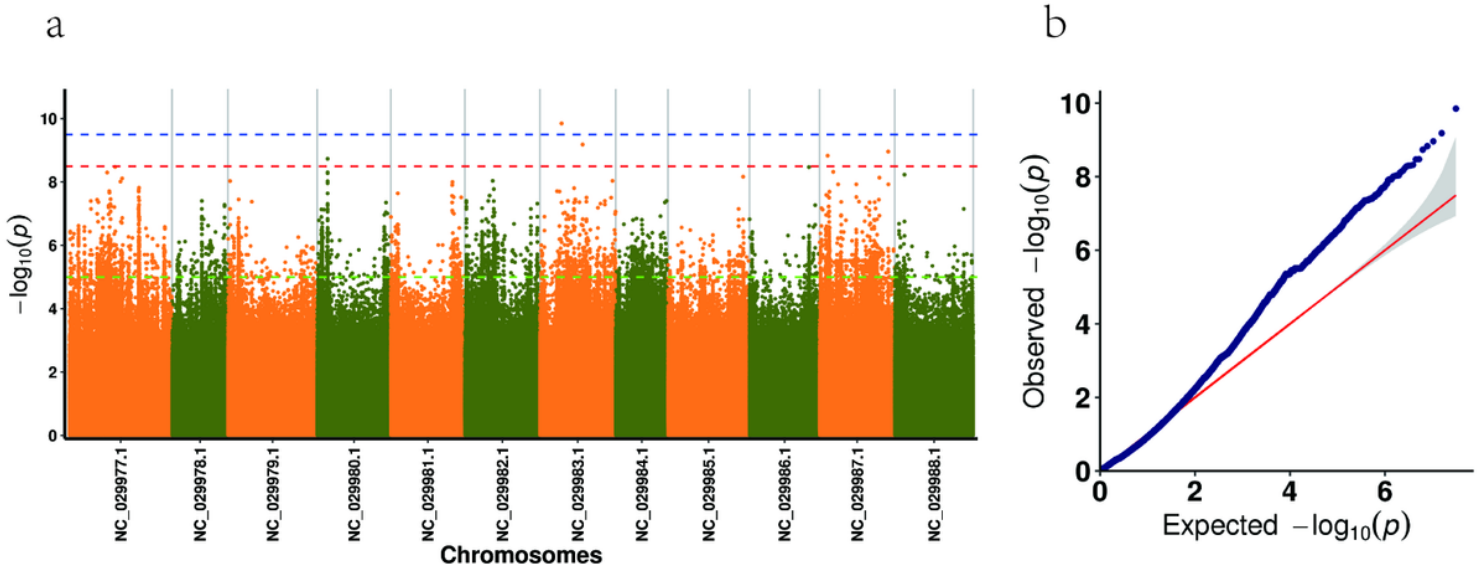

Figure 4

Manhattan and Q-Q plots of the LMM association model for fruit peduncle length. Horizontal coordinates represent chromosome positions, vertical coordinates represent -log10(p) values, with green line representing a value of 5.Blue or red denotes 0.1/labeled

amount and 0.01/labeled amount. Loci above the threshold line are candidate loci.

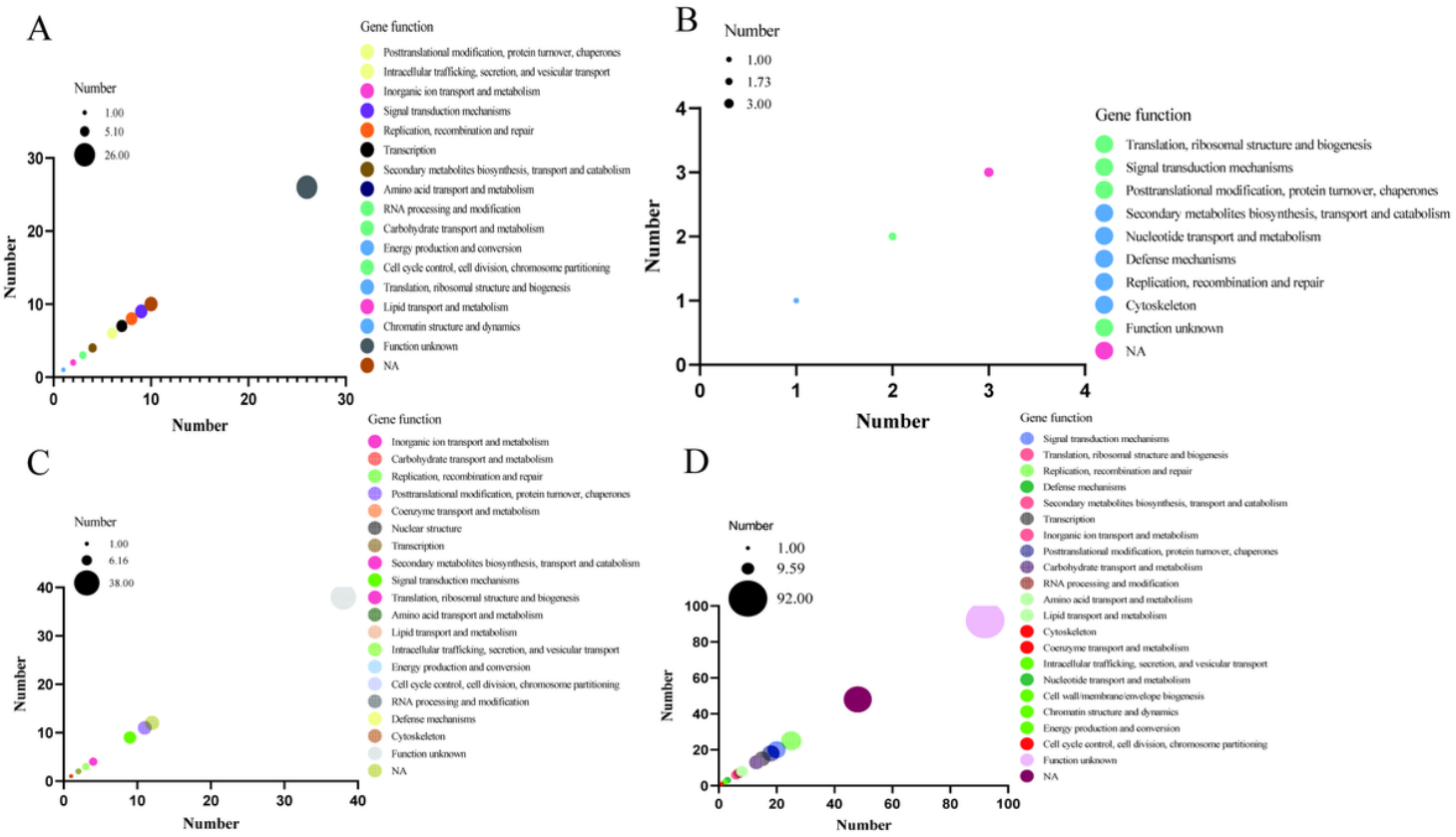

Figure 5

Functional annotation of 519 candidate genes. (a) Stem-related candidate genes. (b) Leaf-related candidate genes (c) Flower-related candidate genes. (d) Fruit-related candidate genes.

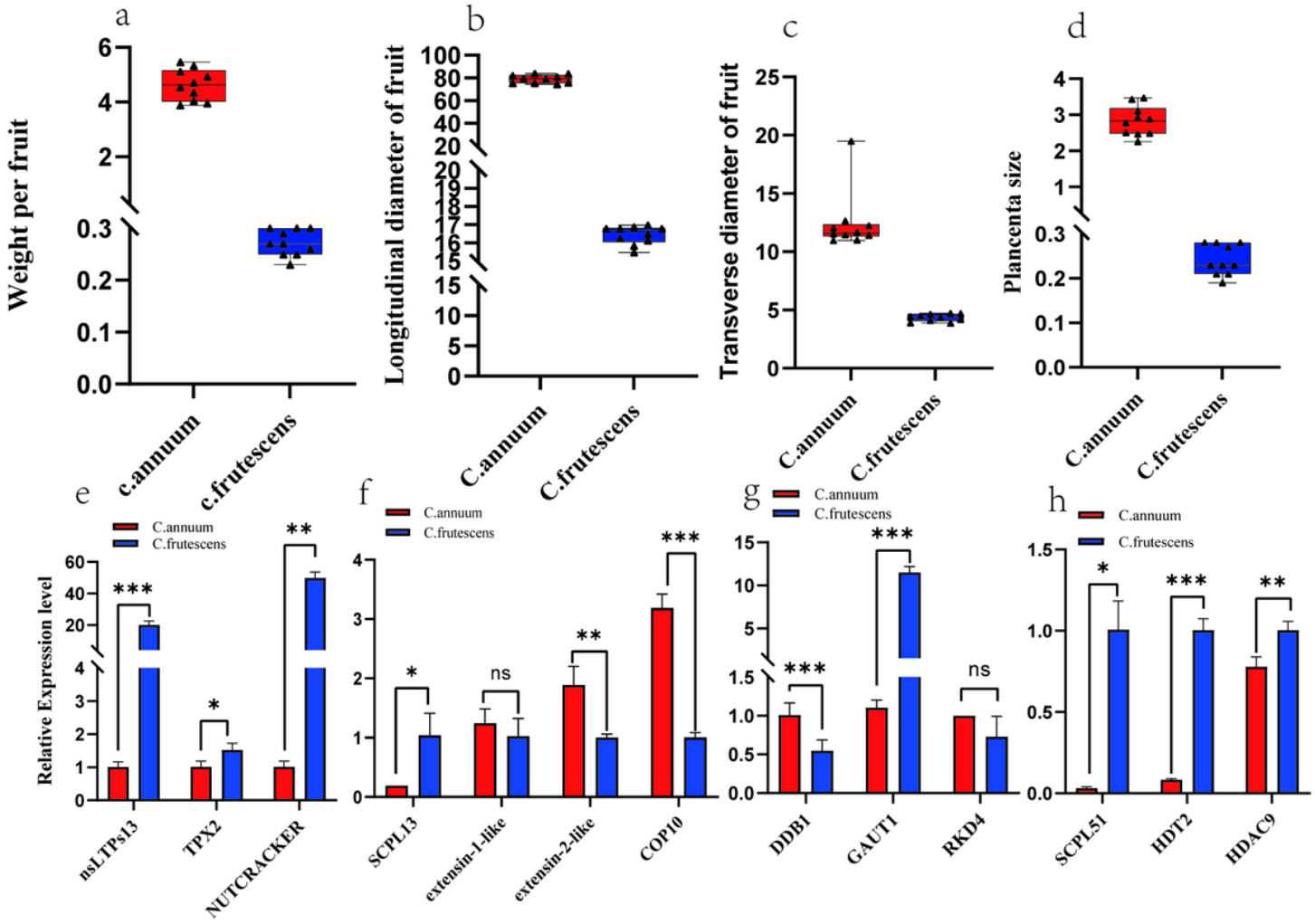

Figure 6

Box plots for four traits (a-d) and qRT-PCR validation results for 11 candidate genes (e-h) (a) Weight per fruit. (b) Longitudinal diameter of fruits. (c) Transverse diameter of fruits. (d) Placenta size. (e) Weight per fruit candidate genes. (f) longitudinal diameter of fruit candidate genes. (g) Transverse diameter of fruit candidate genes. (h) Placenta size candidate genes. * denotes significant correlation (P<0.05), ** (0.001<P<0.01), and *** (P<0.001) highly significant correlations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial0.docx
- Supplementarytable5Sequencingqualityandcomparisonresultsstatistics.of182pepperaccessions..xlsx
- Supplementarytable6Sampleclusteringinformationateachvalueofkandadmixturecrossvalidationerrorrateforeachvalueofk..xlsx
- Supplementarytable7Significantlociandcandidategeneinformationfor24agronomictraits..xlsx
- Supplementarytable8Annotationofimportantcandidategenesfor24agronomictraits..xlsx
- Supplementarytable9Thespeciesof182peppergermplasmswerecollectedinthisstudy..xlsx