

Article

Multi-Locus Association Mapping Uncovers the Genetic Basis of Yield and Agronomic Traits in Chile Pepper (*Capsicum* spp.)

Dennis N. Lozada^{1,2,*}, Derek W. Barchenger³, Danise Coon^{1,2}, Madhav Bhatta⁴, Paul W. Bosland^{1,2}

¹ Department of Plant and Environmental Sciences, New Mexico State University, Las Cruces, NM 88003, USA

² Chile Pepper Institute, New Mexico State University, Las Cruces, NM 88003, USA

³ World Vegetable Center, Shanhua, Tainan 74151, Taiwan

⁴ Bayer Crop Science, Chesterfield, MO 63017, USA

* Correspondence: Dennis N. Lozada, Email: dlozada@nmsu.edu; Tel.: +1-575-646-5171

ABSTRACT

Genome-wide association studies (GWAS) are one of the foundations for modern molecular breeding for plant genetic improvement. Single-locus GWAS methodologies impose constraints in association mapping due to the complex nature of quantitative traits and the often-stringent thresholds used. Multi-locus models, therefore, serve as alternative approaches in identifying significant marker-trait associations. This study used six multi-locus models to determine quantitative trait loci (QTL) associated with yield and agronomic traits in a diverse population of chile pepper evaluated in New Mexico, USA. Using the GWAS models, eight genotyping-by-sequencing (GBS)-derived single nucleotide polymorphism (SNP) markers across six chromosomes were identified to be associated with multiple traits. Plant height shared similar genetic control with plant width, whereas total yield per plant had common QTL with yield components. Epigenetic mechanisms such as methylation and chromatin remodeling and organization were predicted for first pod date and flowering time. The diverse functions of candidate genes identified reflect the complex genomic architecture of the evaluated traits. Allele specific assays for the identified QTL will be developed and validated for marker-assisted selection and genome-wide selection towards the genetic improvement of chile pepper.

Open Access

Received: 08 April 2022

Accepted: 16 May 2022

Published: 19 May 2022

Copyright © 2022 by the author(s). Licensee Hapres, London, United Kingdom. This is an open access article distributed under the terms and conditions of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

KEYWORDS: genome-wide association study; genotyping-by-sequencing; marker-assisted selection; quantitative trait loci; single nucleotide polymorphisms

ABBREVIATIONS

GBS, genotyping-by-sequencing; GWAS, genome-wide association study; LD, linkage disequilibrium; NM-CAMP, New Mexico *Capsicum* association mapping panel; QTL, quantitative trait loci; SNP, single nucleotide polymorphism

INTRODUCTION

Chile pepper (*Capsicum* spp.) is among the most important vegetable and spice crops in the world due to its important diverse uses and cultural significance [1]. Global chile pepper production reached ~61.6 million tons in 2019 [2]. However, chile pepper production is continuously constrained by different factors such as the rapidly changing climate, presence of biotic and abiotic stress factors, labor availability, and costs associated with manual harvesting [3–5]. In the state of New Mexico in the United States, chile peppers are regarded as a major economic and cultural crop [6]. Chile pepper production in New Mexico was 51,000 tons in 2021, a 22% decrease from the previous year [7]. Genetic improvement is crucial to improve production and to meet the increasing demands for chile pepper.

Marker-assisted selection (MAS) has the potential to accelerate crop genetic improvement. The availability of reference genomes for the *Capsicum* spp. [8–10] facilitates effective molecular breeding, marker discovery, genetic mapping, and candidate gene analysis. GWAS is one of the tools to dissect complex traits and identify trait-marker relationships based on linkage disequilibrium (LD) [11]. The development of a mapping population is faster, as GWAS uses a diverse germplasm instead of a biparental population for genetic mapping, eliminating the need to perform experimental hybridizations [12]. Models for GWAS also evaluate multiple traits simultaneously. Nevertheless, spurious associations can arise from genetic and population stratification [13]. Incorporating factors such as relatedness matrices (e.g., principal components) derived from molecular marker data in the GWAS model is therefore necessary to correct for the presence of population structure [14].

The general model for GWAS employs a mixed linear model (MLM) that includes kinship (K) and population structure (Q) information (i.e., $K + Q$ model) that effectively controls Type I and II error rates [15]. More advanced MLM-based methods including the efficient mixed model association (EMMA; [16]), compressed mixed linear model (CMLM; [17]), efficient mixed model association expedited (EMMAX; [18]), and genome-wide efficient mixed model association (GEMMA; [19]), were later developed for association mapping. These approaches feature a one-dimensional survey of the genome by testing a single marker at a time, making the approach computationally efficient especially when using a large number of markers [20].

Previous GWAS in chile pepper used single-locus approaches to understand the genetic basis of various traits with agronomic and

economic importance. The genetic basis of fruit morphology, for example, has been examined by using 373 accessions belonging to 11 *Capsicum* species across 746K polymorphic single nucleotide polymorphism (SNP) sites resulting in the identification of four novel loci associated with fruit shape [21]. Association mapping for fruit weight and capsaicinoid content in *C. annuum* identified 16 SNP markers related to variation for fruit weight located within known genes [22]. In another study, Lee et al. [23] identified 17 fruit-related regions and 16 causal genes controlling major fruit-related traits using genotyping-by-sequencing [GBS]-derived SNP markers implemented for GWAS and biparental mapping. The genetic architecture of *Phytophthora capsici* resistance has also been examined using GWAS, and 117 significant SNP markers across the whole genome were identified to be associated with resistance to the disease [24]. Altogether, these studies demonstrated the utility of single-locus GWAS methods in identifying genomic regions associated with complex traits in chile pepper. The single-locus models, nevertheless, may not render an accurate estimate of marker effects if a trait were controlled by multiple loci [20], especially for large experimental error typical in plant breeding field experiments [25]. The Bonferroni multiple test correction for declaring significant thresholds is also often too stringent, consequently causing many important loci to be declared as non-significant [26]. Moreover, implementing strict thresholds decreases the probability of accepting false positives, but does not minimize the potential of rejecting true positives because of setting very high significance thresholds [27]. Multi-locus GWAS models, therefore, serve as good alternatives to the existing single-locus approaches and has been recommended for the genetic dissection of complex traits.

Multi-locus association mapping involves two major steps in detecting significant marker-trait associations. In the first step, markers are scanned and selected based on a low significance level; after which, a multiple locus method is implemented for markers that pass the initial screening in the second step, and the true quantitative trait loci are confirmed using a likelihood test [20]. Contrary to single-locus association mapping, all potential loci are fitted to a single linear model and their effects are calculated and tested simultaneously in multi-locus GWAS [25], consequently improving power over the single-locus methods [28]. The first multi-locus association mapping approach, multi-locus mixed model [MLMM], was implemented in human and *Arabidopsis thaliana* datasets [29]. The MLMM considers the loci as fixed effects and features a simple, forward-backward stepwise approach where variance components are re-estimated at each step. Random effect multi-locus models for GWAS were subsequently developed due to the observed inflation of the true marker effects from the predicted values resulting from accounting markers as fixed effects [30]. In recent years, the application of multi-locus GWAS methodologies assuming random marker effects to complement or as alternatives to single-locus approaches has expanded across different

plant species. Multi-locus random effect GWAS models have been implemented in major crops including bread wheat (*Triticum aestivum* L.; [31,32]), maize (*Zea mays* L.; [33]), rice (*Oryza sativa* L.; [34,35]), soybean (*Glycine max* L.; [36,37]), and upland cotton (*Gossypium hirsutum*; [38]). Altogether, these studies demonstrated the efficiency and power of multi-locus GWAS in dissecting the genetic basis of different complex traits resulting from using less stringent significant criteria. To our knowledge, there are currently no known reports on the application of multi-locus models for association mapping of complex traits in chile pepper.

In the present study, we demonstrated the potential of using multi-locus GWAS approaches in dissecting the genetic architecture of yield, phenology, and plant morphology-related traits in chile pepper (*Capsicum* spp.) The objectives of this study were to (1) explore the relationships among yield, yield components (green and red mature fruit yield per plant, ten pod weight), and agronomic traits (plant height and width) in a chile pepper mapping population; (2) identify significant genomic regions associated with different yield and agronomic traits using multi-locus GWAS models; and (3) determine candidate genes related to these significant loci. Results from this study will be important for the development and validation of markers for molecular breeding and genetic improvement of current chile pepper germplasm.

MATERIALS AND METHODS

Chile Pepper Germplasm

The New Mexico *Capsicum* association mapping panel (NM-CAMP) consisted of 203 accessions belonging to four cultivated and one wild species of chile pepper (*Capsicum* spp.; Supplementary Table S1). The majority (156; 77%) of the NM-CAMP comprised of members of *C. annuum*, including their wild progenitor, ‘chiltepin’ (*C. annuum* var. *glabriusculum*). The second largest group consisted of members of *C. chinense* (34; 17%); followed by *C. frutescens* (7; 3%), and *C. baccatum* (5; 2%). A single accession of the wild chile pepper *C. chacoense* was also included in the NM-CAMP population. The *C. annuum* group included cultivars previously released by the New Mexico State University (NMSU) Chile Pepper Breeding and Genetics Program such as the ‘NuMex Big Jim’ [39], ‘NuMex Heritage Big Jim’ [40], ‘NuMex Sandia Select’ [41], ‘NuMex Joe E. Parker’ [42], and ‘NuMex Heritage 6-4’ [43], all classified as New Mexican pod type chile pepper. Additionally, jalapeño types such as ‘NuMex Vaquero’ [44], the cayenne, ‘NuMex Las Cruces’ [45], the paprika, ‘NuMex R. Vince Hernandez’ [46], the serrano, ‘NuMex Cajohns’, and ornamental chile peppers ‘NuMex Christmas’, ‘NuMex Easter’, and ‘NuMex Chinese New Year’ [47], and NMSU breeding lines derived from single plant selections were included in the NM-CAMP. The *C. baccatum* comprised of the ajis from South America, whereas the *C. chinense* consisted of habaneros and the ‘Superhots’, chile peppers with more than 1 million Scoville Heat Units,

such as the ‘Carolina Reaper’, ‘Trinidad Moruga Scorpion’ [48], and ‘7 Pot Primo’. The NM-CAMP members of *C. frutescens* comprised of the tabasco, including the ‘NuMex Nobasco’, a no-heat type [49] and ‘Siling Labuyo’, a high-heat cultivar from the Philippines.

Collection and Analysis of Phenotypic Data

Yield and agronomic traits were collected from NM-CAMP for the 2021 growing season at the Leyendecker Plant Science Research Center, Las Cruces, NM (32.32° N, 106.76° W), and at the NMSU Los Lunas Agricultural Science Center, Los Lunas, NM (34.81° N, 106.73° W; hereafter referred to as the CRU and LUN locations, respectively), 320 kms (200 miles) N of CRU. Seedlings with 8–10 true leaves were transplanted 30 cm apart in 4.5 m plots with 0.30 m distance between plots in raised beds. The plants were cultivated under standard cultural and management practices including fertilizer application and furrow irrigation for growing chile pepper in New Mexico. Transplanting was done in April (CRU) and May (LUN), whereas a single manual harvest of pepper fruit samples was conducted from September–October 2021. The CRU location has a Belen clay loam class, whereas the LUN environment has a mixture of sandy clay loam (Gila) and Belen soil types. The mean temperatures between April and October and between May and October 2021 for the CRU and LUN locations were 21.5 °C and 19.5 °C, respectively, with the highest average temperature recorded in July at 25.7 °C in CRU (National Climatic Data Center; <https://www.ncdc.noaa.gov/cdo-web/search>; data retrieved 10 May 2022) [50].

The NM-CAMP was evaluated in an augmented complete block field design for the CRU and LUN environments, where each block consisted of replicated checks, namely ‘Charger’ (New Mexican type), ‘Centella’ (jalapeño), and ‘C2 792’ (cayenne) and un-replicated entries. The total yield per plant (TYP; in kgs.) was represented as the fresh weight of the mature pods (green and red fruits) divided by the total number of plants of each plot; whereas the yield component trait, 10-pod weight (TPW; in kgs.), was measured as the total weight of randomly selected five mature green and five mature red pepper fruits. Other yield components such as green yield per plant (GYP) and red yield per plant (RYP) represented the fresh weight for green and red mature chile pepper fruits, respectively, divided by the number of individual plants of each plot, in kgs. Plant height (PHT) was the average measurement (in cm) from the ground to the top of the plant for five individual plants in each plot. Plant width (PWIDTH; in cm) was measured as the widest point of five individual plants on the top of the canopy. Flowering time (FT) and first pod date (FPD) represented the number of days when the flowers and the pods (fruits) start to develop from the day of transplanting for each entry of the NM-CAMP. Phenotypic data collection was performed on a plot basis for each genotype using a maximum of ten plants for yield and yield components, and five individual plants for the plant morphology traits.

Phenotypic field data were analyzed under an augmented type complete block using a custom script for an augmented complete block design in the statistics program, R [51]. Best linear unbiased estimates (BLUEs) for the phenotypes were calculated in individual locations, whereas the Best Linear Unbiased Predictors (BLUPs) were estimated for the combined analysis across environments. Effects were considered as fixed and random when calculating BLUE and BLUP values, respectively. The broad-sense heritability, H^2 at an individual location was calculated as: $H^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_E^2)$; where σ_G^2 and σ_E^2 represent the variance due to genotype and residual, respectively. Across locations, H^2 was represented as $H^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_{GE/n}^2 + \sigma_E^2)$; where σ_{GE}^2 is the variance due to genotype-by-environment interaction and n is the number of environments.

Adjusted values for the trait data were calculated using the following statistical model for analysis for individual environment, where the effects were regarded as fixed:

$$Y_{ij} = \mu + Gen + Ck + ID_{ck} + Blk_i + \varepsilon_{ij} \quad (1)$$

where Y_{ij} is the phenotype of interest; μ is the mean effect; Gen is the un-replicated entry effect; Ck and ID_{ck} are the effects of the replicated check on each block and the check identifier, respectively; Blk_i is the effect of the i^{th} block; and ε_{ij} is the residual effect. Phenotypic values across combined locations were calculated using the following model, where the effects were considered random:

$$Y_{ikj} = \mu + Gen + Ck + ID_{ck} + Blk_i + Loc_i + Loc_i \times Gen + Loc_i \times Ck + Loc_i \times ID_{ck} + Blk_k (Loc_i) + \varepsilon_{ijkl} \quad (2)$$

where Y_{ij} is the phenotype of interest; μ is the mean effect; Gen is effect of un-replicated entry; Ck and ID_{ck} are the effects of the replicated checks on each block and the check identifier, respectively; Blk_i is the effect of the i^{th} block; Loc_i is the effect of location i ; $Loc_i \times Gen$, $Loc_i \times Ck$, $Loc_i \times ID_{ck}$ are the effects of the interactions between environment and the entry, check, and check identifiers, respectively; $Blk_k (Loc_i)$ is the effect of block nested into the location; and ε_{ij} is the residual effect. The BLUE and BLUP trait values were calculated using models (1) and (2), respectively. Analysis of variance for an augmented design was conducted using the ‘augmentedRCBD’ [52] package in R [53]. Principal components analysis and Pearson correlation coefficient calculation were performed using the software ‘DeltaGen’ [54] and the multivariate function in JMP® 13.2.1 [55], respectively.

DNA Isolation and Genotyping-by-Sequencing

Seeds were planted in F1020 multi-cell trays (American Horticultural Supply, Inc., CA, USA) at the Fabián García Science Center, Las Cruces, NM, USA and were grown in standard greenhouse conditions with temperatures maintained between 26 and 30 °C for cultivating chile pepper [56]. Tissue samples from seedlings (30–45-day old) of single plants per entry were collected using 1.2 mL Qiagen® collection microtubes

(Qiagen, MD, USA). Genomic DNA extraction was conducted using ~50 mg of fresh leaf tissue using the Qiagen DNEasy® 96 plant extraction kits through the University of Minnesota Genomics Center DNA Extraction Facility (<https://genomics.umn.edu/service/dna-extraction>). The total DNA isolated was quantified using Picogreen® (ThermoFisher Scientific, MA, USA) and samples were normalized to 10 ng/μL for genotyping-by-sequencing (GBS).

The GBS procedure for the NM-CAMP was conducted through the University of Minnesota Genomics Center (<https://genomics.umn.edu/services/gbs>) as described previously in Lozada et al. [57]. Briefly, a single enzyme digestion protocol was implemented using *ApeKI* (New England Biolabs, Inc., MA, USA) for ~100 ng of DNA per sample. Enzyme incubation at 75 °C for 2 h was performed and heat-inactivation at 80 °C for 20 min followed. Ligation of the DNA samples were then conducted with 200 units of T4 ligase (New England Biolabs, Inc. MA, USA). Purification of the ligated samples with solid phase reversible immobilization (SPRI) beads, and subsequent amplification (18 cycles) with 2X NEB Taq Master Mix were performed to add the barcodes. Libraries were purified, quantified, and pooled using SPRI beads. Fragments (300–744 bp) were selected and diluted to 1 nM for single end 1 × 100 sequencing using the Illumina NovaSeq™ 6000 sequencer (Illumina, CA, USA). Illumina ‘bcl2fastq’ software (Illumina, CA, USA) was used to demultiplex the raw FASTQ files. For the beginning of each sequence read, the first 12 bases were removed to exclude the adapters. Adapter sequences at the 3’ ends of the reads were removed using Trimmomatic [58]. Burrows-Wheeler Aligner [59] aligned the FASTQ files to the *Zunla-1* (*C. annuum*) reference genome [9]; *Zunla-1* reference genome GCA_000710875.1, PRJNA186921, v.1.0; (<https://www.ncbi.nlm.nih.gov/nuccore/ASJU00000000.1/>). Simultaneous joint calling of variants across all samples was done using the FreeBayes Bayesian genetic variant identifier [60]. VCFtools processed raw variant call format (VCF) files to remove variants with the following parameters: (a) minor allele frequency <1%; (b) genotype rates <95%, and (c) samples with genotype rates <50%. TASSEL 5.2.77 [61] was used to convert the VCF files to the HapMap format using where loci with MAF < 0.05 and unmapped markers were further excluded. Names of SNP markers were designated as "Scaffold/Chromosome name Position of the marker (in bp) in the scaffold". Missing data imputation was accomplished using the LD *k*-nearest neighbor genotype imputation function [62] in TASSEL 5.2.77.

Genetic Diversity and Population Structure

Different indices were used to calculate the genetic diversity for the NM-CAMP. Tajima’s D statistic [63] was calculated using a ‘sliding window’ analysis (step size = 100 (0.10 Kb) and window size = 500 (0.50 Kb)) in TASSEL 5.2.77. GenoDive program [64] calculated heterozygosity within populations (H_s), total heterozygosity (H_t), and inbreeding coefficient (G_{is}).

Analysis of molecular variance (AMOVA) [65] was performed through the package ‘poppr’ [66] in R. ‘GeneticSubsetter’ [67] package in R calculated polymorphism information content (PIC). Principal component analysis using genome-wide SNP markers were conducted using the “PCA” function in TASSEL v.5.2.77 and the first two PCs were fitted in a bivariate plot in JMP 13.2.1.

Population genetic stratification of the NM-CAMP was evaluated using STRUCTURE [68]. An admixture model was applied using a burn-in of 10,000 iterations, 10,000 Monte Carlo Markov Chain replicates, and number of clusters, K , between 1 and 10 with the number of replicates per K equal to 5. An ad-hoc statistic for ΔK based on the magnitude of changes in the log probability of data between the values of K [69] was used to infer the true number of K that best represent the entries in STRUCTURE HARVESTER [70]. ‘StructuRly’ program [71] visualized the admixture values derived from STRUCTURE for each entry through bar plots.

Linkage Disequilibrium Analysis

Analysis of linkage disequilibrium (LD) was conducted in TASSEL v.5.2.77, where LD coefficients for pairwise intrachromosomal marker pairs were calculated using the “sliding window” type function and a window size of 50 (0.05 Kb). The intrachromosomal LD values (r^2) were represented as the square of allele frequency correlations between pairs of loci [72]. Distance at which LD starts to decay was determined by fitting a non-linear regression model [73,74] in the r^2 against physical distance (Mb) biplot and identifying the intersection between the regression curve and a critical value ($r^2 = 0.25$) in the plot. Analysis of LD was also performed using the markers identified to be associated with yield and agronomic traits. Intrachromosomal marker pairs with $P < 0.05$ were declared to be in significant LD.

Multi-locus Association Mapping and Analysis of Candidate Genes

Multi-locus GWAS was conducted using the multi-locus random-SNP-effect Mixed Linear Model (‘mrMLM.GUI’) package [75] in R for the BLUE and BLUP trait values. A total of six GWAS models were implemented to identify QTL: (1) mrMLM [20]; (2) FASTmrMLM [76]; (3) FASTmrEMMA [77]; (4) pLARM EB [78]; (5) pKWmEB [79]; and (6) ISIS EM-BLASSO [80] using default parameters (Table 1). A restricted maximum likelihood (REML) approach was used for FASTmrEMMA and the bootstrap was set to ‘FALSE’ for pLARM EB. The search radius (in kb) of candidate genes for mrMLM and FASTmrMM was set to 20, whereas the number of potentially associated variables selected by LARS for pLARM EB was set to 50. To minimize the confounding effects of a small number of individuals for some representative species, the *C. chacoense* accession was excluded for the final GWAS analyses. A kinship-principal component (K -PC) model was implemented for the identification of QTL, where the K matrix and the first three PCs were included in the model to correct for population structure.

Principal components derived from marker data were calculated using TASSEL v.5.2.77. The threshold for identifying a significant marker-trait association using the multi-locus GWAS models was set to an LOD score of >3.0 to balance high power and low false positive rate for the detection of QTL [25]. Candidate genes within the LD regions for the significant multi-trait QTL were identified by comparing sequences against the genome of ‘Criollo de Morellos 334’ (CM-334; Genome assembly (GA): ASM512225v2) (*C. annuum* L.) using BLASTn in EnsemblPlants ([81]; <https://plants.ensembl.org/index.html>; accessed on 01 December 2021) with the following criteria: (a) $\geq 95\%$ sequence identity and (b) an e-value cut-off of 10. Genes were annotated based on their biological functions using a gene ontology-based sequence annotation.

Table 1. Overview of the multi-locus association mapping models used to identify QTL using the New Mexico *Capsicum* association mapping panel.

Model ¹	Features	Reference
1. Multi-locus random-SNP-effect mixed linear model (mrMLM)	Implements a ‘population parameters previously determined’ (P3D) algorithm and residual variance is estimated after QTL variance effect	Wang et al. [20]
2. Fast multi-locus random-SNP-effect mixed linear model (FASTmrMLM)	Accelerates mrMLM function by using GEMMA, matrix transformation, and identities	Tamba and Zhang [76]
3. Fast multi-locus random-SNP-effect efficient mixed model association (FASTmrEMMA)	Model chooses potential QTL with $P < 0.005$ and includes them in a multi-locus model for identification of significant associations	Wen et al. [77]
4. Polygenic-background-control-based least angle regression plus empirical Bayes (pLARM EB)	Least angle regression (LARS) model is implemented to select for significant markers	Zhang et al. [78]
5. Polygenic Kruskal-Wallis method with Empirical Bayes (pKWmEB)	Kruskal-Wallis test with LARS is used to determine markers associated with a trait	Ren et al. [79]
6. Iterative modified-sure independence screening Expectation-Maximization-Bayesian least absolute shrinkage and selection operator (ISIS EM-BLASSO)	Number of markers are reduced via correlation learning; EM-BLASSO estimates effects and likelihood ratio identifies true QTL	Tamba et al. [80]

¹All the models consider marker effects as random and controls polygenic background and population structure.

RESULTS

Phenotypic Diversity for Yield and Agronomic Traits

Analysis of variance for augmented design revealed significant differences ($P < 0.0001$) among entries for FPD, FT, PHT, and PWDTH within environments for the NM-CAMP (Supplementary Table S2). Overall, the total yield per plant (TYP) for Las Cruces (CRU) was higher than the Los Lunas (LUN) location, although no significant differences were observed between the environments (Supplementary Table S3). Entries in the CRU location were taller and wider, on average, than the LUN environment. Conversely, at the LUN location, earlier flowering time (FT) and first pod date (FPD) (~3 days earlier, on average) occurred as compared to the CRU. Across individual environments, plant height (PHT), FT, and FPD were the most heritable traits, with PHT reaching a broad-sense heritability (H^2) value of 0.97 for CRU, whereas TYP was moderately heritable for CRU (0.50) and LUN (0.45). For the combined multi-location analyses, H^2 values ranged between 0.20 (TYP and red yield per plant (RYP)) and 0.88 (ten pod weight, TPW). Phenotypic correlations among yield and agronomic traits showed wide range of variation for Best Linear Unbiased Prediction (BLUP) trait values across environments (Table 2). The yield components TPW, RYP, and green yield per plant (GYP) were significantly correlated with TYP with correlation coefficient, r values of 0.15 ($P < 0.05$), 0.74 ($P < 0.0001$); and 0.61 ($P < 0.0001$), respectively, whereas there was a negative, non-significant correlation between GYP and RYP ($r = -0.03$, $P > 0.05$). Time to flowering and first pod date (FT and FPD) were highly significantly correlated ($r = 0.98$; $P < 0.0001$). A significant correlation was likewise observed for PHT and PWDTH ($r = 0.65$; $P < 0.0001$). The principal component (PC) biplot for phenotypic traits support the values for phenotypic correlation. The total yield and yield components formed a separate cluster with the plant morphology and phenology-related traits (Figure 1A). The first and second PCs were related to 32.4 and 25.2% of phenotypic variation, respectively.

Table 2. Phenotypic correlations between different yield and other agronomic traits for the New Mexico *Capsicum* association mapping panel.

Trait ¹	GYP	RYP	TYP	TPW	FPD	FT	PWDTH
GYP	-						
RYP	-0.03	-					
TYP	0.61***	0.74***	-				
TPW	0.22**	0.06	0.15*	-			
FPD	0.02	0.04	0.05	-0.17*	-		
FT	0.02	0.04	0.05	-0.15*	0.98***	-	
PWDTH	-0.03	0.12*	0.10	-0.10	0.38***	0.38***	-
PHT	3.71×10^{-05}	0.02	0.04	-0.06	0.26***	0.27***	0.65***

¹ GYP—Green yield per plant; RYP—Red yield per plant; TYP—Total yield per plant; TPW—Ten pod weight; FPD—First pod date; FT—Flowering time; PWDTH—Plant width; PHT—Plant height; ***—Significant, $P < 0.0001$; **—Significant, $P < 0.001$; *—Significant, $P < 0.05$.

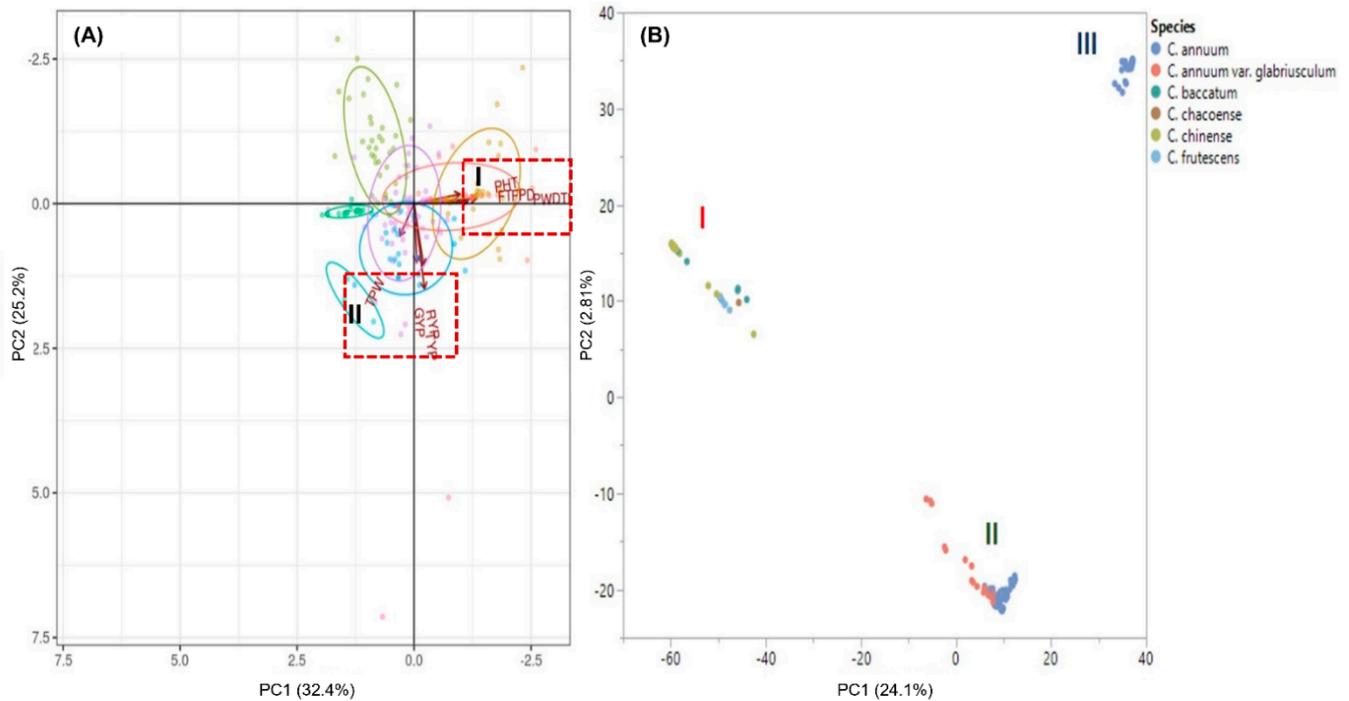


Figure 1. Principal component (PC) biplots from the first two PCs for (A) yield, plant morphology, and phenology-related traits. Group I consisted of PHT, PWDTH, FPD, and FT, whereas Group II comprised the yield (TYP) and yield components (GYP, RYP, and TPW). (B) New Mexico *Capsicum* association mapping panel (NM-CAMP) from GBS-derived SNP markers showing genetic differentiation based on species. FPD—First pod date; FT—Flowering time; GYP—green yield per plant; PHT—Plant height; PWDTH—Plant width; RYP—Red yield per plant; TPW—Ten pod weight; TYP—Total yield per plant.

GBS-Derived SNP Markers

A total of 25,867 GBS-derived SNP markers were identified across the 12 chromosomes of Chile pepper for the NM-CAMP. After removing variants with various filtering criteria (minor allele frequency <1%; genotype rates <95%; samples with genotype rates <50%), 14,922 SNP loci remained for further analysis. Chromosome P3 had the greatest number of SNPs identified (1965), followed by chromosomes P1 and P2, with 1692 and 1505 markers, respectively. Chromosomes P11 (871), P5 (911), and P9 (917) had the least number of SNP markers discovered. In terms of marker density (SNP/Mb), chromosome P8 had the highest SNP density (9.2), followed by P2 (9.1), and P3 (7.5); whereas chromosomes P9, P11, and P7 had the least, with 3.8, 3.9, and 4.1, respectively (Supplementary Table S4). Overall, the whole genome had a marker density of 5.6 SNP/Mb. Across the SNP sites, the most common nucleotide was thymine (T) (29.8%), followed by adenine (A) (24.9%), guanine (G) (23.1%), and cytosine (C) (22.2%). Ambiguous nucleotide calls comprised 133,858 (4.4%) of the total sites. Transition substitutions were identified in 8,149 sites (54.6%), where the ‘A/G’ was the most common (14.3%) followed by the ‘T/C’ (13.9%) type. A total of 6773 sites (45.4%) were transversion substitutions with the ‘T/A’ being the most common (6.8%) and the ‘G/C’ type the least (4.1%), across

the whole genome. The proportion of heterozygous SNP sites was 0.04 and the average minor allele frequency across all loci was 0.22.

Analysis of Principal Components, Genetic Diversity, and Population Structure

Principal component analysis (PCA) revealed three distinct clusters based on species (Figure 1B). The first two PCs accounted for 24.1% (PC1) and 2.8% (PC2) of the total genotypic variance. Cluster I was comprised of the *C. baccatum*, *C. chinense*, *C. frutescens*, and *C. chacoense* complex (46 entries), whereas Clusters II and III consisted of the *C. annuum*. The chiltepins formed a complex with the ornamentals, jalapeno, serrano, cayenne, and New Mexican pod types previously released by the NMSU Chile Pepper Breeding Program (Cluster II; 112 lines). The last group (Cluster III; 45 genotypes) consisted of the NMSU breeding lines derived from single plant selections of several ‘NuMex’ cultivars including the ‘NuMex Joe E. Parker’, ‘NuMex Big Jim’, and ‘NuMex Sandia Select’.

Tajima’s D value for the NM-CAMP was 2.97. Negative Tajima’s values were observed for all groups, with -1.17, -0.19, and -0.20 for Clusters I, II, and III, respectively. Analysis of molecular variance revealed that most of the genetic variation was a consequence of variation between the *Capsicum* populations (PCA groups) (73.4%; Table 3). Variation between samples within a population and variation within samples accounted for 17.3 and 9.3% of variation, respectively.

Inference for population stratification using STRUCTURE identified $K = 5$ ($\Delta K = 14,014.61$) as the optimal number of genetic groups for the NM-CAMP (Figures 2A, B). These clusters corresponded to the different *Capsicum* species, where the *C. annuum* and *C. annuum* var. *glabriusculum* (chiltepins) were divided into four different clusters. Accordingly, Group I consisted of the *C. annuum* and chiltepins, whereas the *C. chinense*, *C. baccatum*, *C. frutescens*, and *C. chacoense* formed a complex (Group II). Clusters III and IV comprised exclusively of the *C. annuum* and chiltepins, respectively. Group 5 was an admixture of *C. annuum* and *C. annuum* var. *glabriusculum*. The number of cluster $K = 3$ ($\Delta K = 5696.21$) also showed a high delta K value, demonstrating that this could serve as an alternative clustering to describe the population differentiation among the different species for the NM-CAMP (Figure 2B). The grouping for $K = 3$ for the population was consistent with the PC biplot with differentiation based on species, where *C. annuum* formed Cluster II and III, whereas the rest of the *Capsicum* species comprised a complex for Group I.

Table 3. Analysis of molecular variance using genome-wide SNP markers for the New Mexico *Capsicum* association mapping panel.

Source of variation	Degrees of freedom	Sum of squares	Mean square	Standard deviation	%
Between populations	2	619,197.9	309,598.9	2543.8	73.3
Between samples within population	201	306,071.1	1522.7	599.7	17.3
Within samples	204	65,949.0	323.3	323.3	9.3
Total	407	991,218.0	2435.4	3466.8	100.0

Linkage Disequilibrium

Analysis of linkage disequilibrium (LD) identified 681,178 intrachromosomal marker pairs for the NM-CAMP; of which 530,921 (77.9%) were in significant ($P < 0.05$) LD (Supplementary Table S4). The analysis of LD for QTL associated with yield and other agronomic traits further identified 777 intrachromosomal marker pairs in significant LD (0.11%) (Supplementary Table S5). Chromosome P1 had the highest number of pairs in significant LD (61,708; 80.1%) whereas chromosome P4 had the least (41,763; 78.9%). The mean coefficient of LD (r^2) across all pairs of SNP markers was 0.25. A total of 30,111 (4.4%) pairs of loci were in complete LD ($r^2 = 1.0$). Across the individual chromosomes, the mean distance of marker pairs in significant LD ranged between 2.8 (chromosome P2) and 7.25 Mb (chromosome P9). The average distance (in Mb) of marker pairs in complete LD was highest for chromosome P9 (8.88), followed by P5 (8.55), and P11 (8.07). The LD began to decay at a physical distance of ~2.82 Mb (critical value, $r^2 = 0.25$) across the whole genome for the NM-CAMP (Figure 2C). Decay for LD varied across the different chromosomes, ranging between 0.81 Mb (chromosome P8), and extended up to 8.94 Mb (chromosome P5). Across the different PCA groups, distances at which LD decay began were highest for Cluster III (2.74 Mb), followed by Cluster I (1.69 Mb). Cluster II, comprised of NMSU chile pepper cultivars and chiltepin, had the least LD decay distance among the groups at 0.24 Mb.

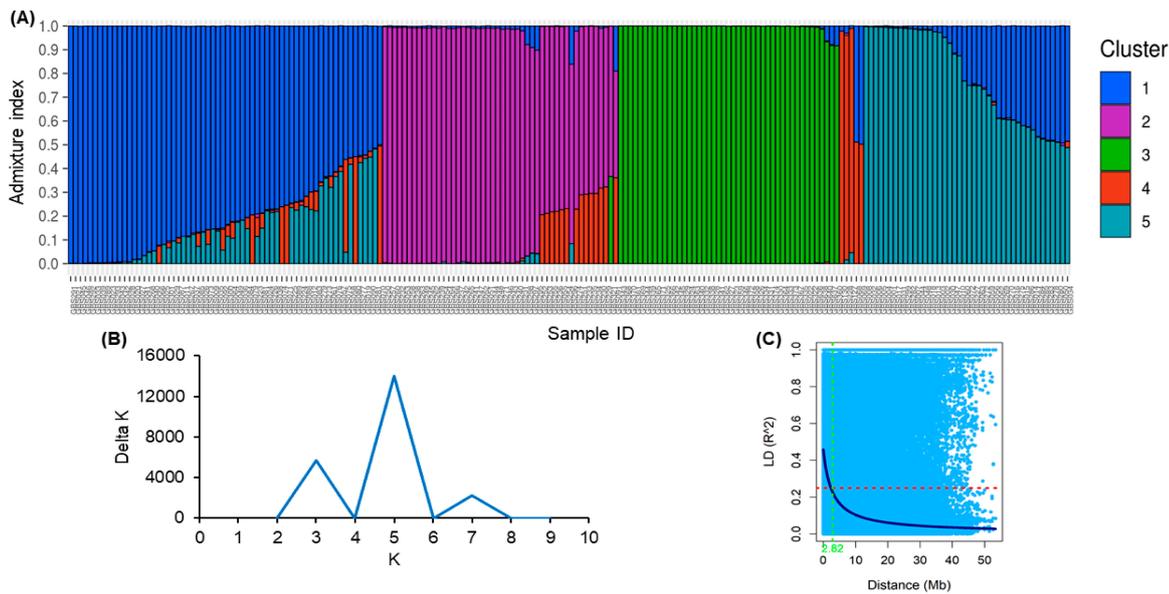


Figure 2. Bar plots for the admixture indices for each individual in the New Mexico *Capsicum* association mapping panel (NM-CAMP) for (A) $K = 5$ clusters. Clusters 1, 3, and 5 comprised of the *C. annuum* and *C. annuum* var. *glabriusculum* (chiltepins), whereas cluster 2 consisted of the *C. baccatum*, *C. chacoense*, *C. chinense*, and *C. frutescens* complex. Cluster 4 comprised exclusively of the chiltepins. (B) Inference for the optimal number of clusters for the NM-CAMP using the Evanno method revealed $K = 5$ with the highest value for ΔK ; (C) Decay plot for linkage disequilibrium (LD). The red dashed line represents the critical value for LD ($r^2 = 0.25$) and the blue solid line corresponds to the non-linear regression curve. The intersection between the critical value and the regression curve is the distance at which LD starts to decay (~ 2.82 Mb).

Quantitative Trait Locus and Candidate Gene Identification

Using six multi-locus GWAS models, 215 QTL distributed in 12 chromosomes were identified to be associated with yield and other agronomic traits for the NM-CAMP based on an LOD score threshold of > 3.0 (Supplementary Table S6). Of the identified QTL, 86 (40%) were detected using at least two different models, traits, and/or environments. The multi-locus GWAS model, pKWmEB, resulted in the greatest number of significant SNPs identified (75), whereas the FASTmrEMMA had the fewest QTL (31). Chromosome P3 had the highest number of identified QTL (28), followed by P2 (24), and P1 (23). Overall, 60 QTL were associated with TPW, whereas 29 and 23 QTL were related with GYP and TYP, respectively. Among the plant morphology-related traits, PHT had a greater number of identified QTL (54) compared to PWDTH (27); whereas for the phenology-related traits, FPD had more QTL (23) than FT (20). A total of eight multi-trait, multi-model QTL were detected in at least two environments (Table 4). The traits PHT and PWDTH shared a common QTL mapped on chromosomes P2 and P11, whereas TYP and GYP had multi-trait QTL on chromosomes P2 and P6. Time to flowering and first pod date also shared common QTL on P1, P6, P7 (Figure 3). All six multi-locus models identified the multi-trait QTL *SCM002817.1_37397491* (P6), associated with FPD and FT and related with 2.9–11.9% of phenotypic variation. The QTL *SCM002815.1_213170635* (P4) was associated with FPD and PHT and

explained ~6.0% of variation for the traits. A total of 40 candidate genes for *C. annuum* were identified for the multi-trait QTL (Supplementary Table S7). These candidate genes were associated with at least 60 diverse biological functions related to defense response, metabolic processes, oxidation-reduction, phosphorylation, and gene silencing, among others. A candidate gene for the multi-trait QTL *SCM002817.1_37397491*, *T459_13341*, mapped on chromosome P6 and associated with FPD and FT, is predicted for epigenetic mechanisms namely chromatin organization and methylation in *C. annuum*. Another candidate gene for *SCM002817.1_37397491*, *T459_07853*, has function associated with the remodeling of chromatin.

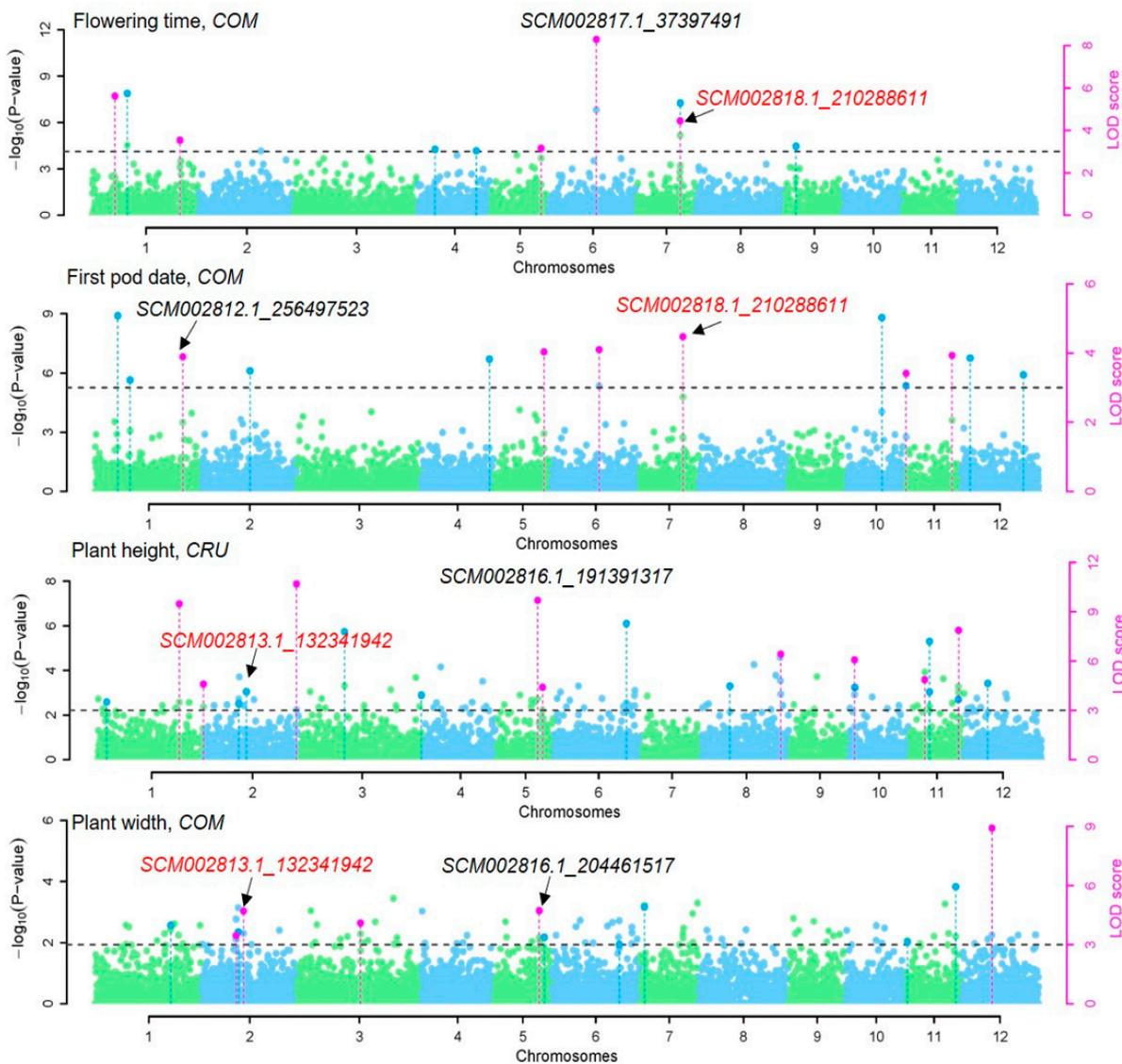


Figure 3. Genome-wide quantitative trait loci (QTL) associated with different phenology and plant morphology-related traits using multi-locus models in chili pepper (threshold LOD score >3.0; horizontal dashed line). Loci in red text are multi-trait QTL. The QTL identified by multiple GWAS models are indicated by pink dots above the significance threshold. *COM*- combined values across environments; *CRU*- Las Cruces, NM.

Table 4. Multi-trait quantitative trait loci (QTL) identified for the New Mexico *Capsicum* association mapping panel.

QTL	Traits ¹	Chr.	Marker position (Mb)	Env. ²	Model ³	QTL effect	LOD	R ² ⁴
SCM002812.1_256497523	FPD, FT	P1	256.49	COM	1	2.30	3.89	4.39
				COM	2	2.25	4.84	4.38
				COM	4	1.75	3.78	2.63
				LUN	6	5.50	4.93	10.62
				COM	4	1.79	3.97	2.49
				COM	5	1.73	3.11	9.87
SCM002818.1_210288611	FPD, FT	P7	210.29	COM	1	-4.28	4.47	32.13
				COM	2	-2.80	4.03	14.35
				COM	5	-3.37	16.63	38.31
				LUN	1	-8.21	5.90	20.65
				COM	1	-3.28	4.22	17.79
				COM	3	-6.79	4.68	17.71
SCM002817.1_37397491	FPD, FT	P6	37.39	LUN	4	-7.31	3.69	14.25
				COM	2	2.46	3.09	2.91
				COM	3	6.59	3.25	5.59
				COM	2	4.50	11.30	8.72
				COM	4	3.47	5.29	5.19
				COM	6	5.05	14.50	10.98
SCM002815.1_213170635	FPD, PHT	P4	213.17	LUN	1	10.42	10.59	6.12
				LUN	2	9.24	8.24	9.02
				LUN	3	18.54	5.93	9.41
				LUN	5	8.78	7.48	11.87
				COM	5	2.06	3.83	6.22
				LUN	2	5.76	4.33	6.53
SCM002813.1_138962498	GYP, TYP	P2	138.96	COM	6	-0.004	6.28	25.54
				LUN	4	-0.58	7.13	8.72
				LUN	4	-0.88	4.62	19.63
SCM002817.1_189395041	GYP, TYP	P6	189.35	COM	5	-0.001	3.71	4.12
				LUN	2	-0.22	3.32	5.73
				LUN	4	-0.18	4.19	1.31
				LUN	6	-0.24	4.19	6.80
SCM002822.1_213369314	PHT, PWIDTH	P11	213.36	CRU	1	-16.94	9.10	18.34
				CRU	2	-10.44	7.87	6.96
				CRU	6	-9.99	4.61	6.38
				COM	5	-4.25	5.94	9.96
SCM002813.1_132341942	PHT, PWIDTH	P2	132.34	CRU	6	7.12	4.14	4.26
				COM	1	5.03	4.86	16.44
				COM	2	3.88	4.71	9.85
				COM	6	3.66	4.55	8.78

¹ FPD- First pod date; FT- Flowering time; GYP- Green yield per plant; PHT- Plant height; PWIDTH- Plant width; TYP- Total yield per plant; ² COM- Combined values across environments; CRU- Las Cruces, NM; LUN- Los Lunas, NM; ³ Multi-locus GWAS models: (1) mrMLM; (2) FASTmrMLM; (3) FASTmrEMMA; (4) pLARM; (5) pKWmEB; (6) ISIS EM-BLASSO (Table 1); ⁴ Percent variation explained for each QTL.

DISCUSSION

Since its first application in crops two decades ago [74,82], GWAS has become one of the foundational approaches for modern molecular plant breeding. The method itself complements traditional linkage mapping based on biparental populations by providing a higher mapping resolution, often to the gene level, because of higher recombination events that occur in natural populations [83]. The genetic architecture of complex traits such as yield and the often-stringent thresholds in declaring significant associations in single-locus GWAS models, nevertheless, can be limiting when implementing these mapping approaches. In the current study, multi-locus association mapping uncovered the genetic architecture of major yield and agronomic traits in chile pepper. Candidate genes underlying these QTLs were further identified, rendering deeper insights into the genetic basis of these complex traits in *Capsicum*.

Population Structure in Chile Pepper was Related to Genetic Differentiation Based on Species

Population stratification for the NM-CAMP was inferred using a Bayesian iterative algorithm approach implemented in STRUCTURE. Accordingly, the optimal number of clusters K derived from using an ad-hoc Evanno criterion was observed to be at $K = 5$, where the *C. frutescens*, *chinense*, *baccatum*, and *chacoense* formed a complex, whereas the *C. annuum* was further divided into four distinct clusters. Our results support recent phenotypic and phylogenetic analyses by Parry et al. [84] that demonstrated a more complex genetic relatedness within the *C. annuum*, whereas the other *Capsicum* species are more clustered together. These observations highlight the high genetic similarity present in the different species grouped in the same cluster, and the potential for crop improvement by using the wild relatives having close genetic relationships for hybridizations [84]. As a species, the genetic variability of the *C. annuum* was demonstrated on another study where the annuum genotypes clustered into two different groups based on pod types (New Mexican, jalapeno, ornamental types, etc.) [57]. The “biological significance” of the number of inferred clusters is of prime relevance when interpreting results from STRUCTURE, as the optimal K might not essentially reflect the ideal number of groups, a consequence of it being determined exclusively by a predetermined sampling scheme [68]. Differences in terms of the inferred optimal K value and the number of groups based on other approaches, such as PCA, have been observed in the *Capsicum*. For instance, in a recent genetic diversity study using a different population of New Mexican chile pepper consisting of 165 genotypes, we observed that $K = 2$ is the optimal K ; however, PCA revealed four different groups based on species [57]. Analysis of linkage disequilibrium showed a moderate LD decay (~2.82 Mb) for the population indicating that a higher number of markers might be necessary in implementing association

mapping for the whole population. In the present study, the *C. annuum* group (PCA Cluster III) showed the highest extent of LD (2.74 Mb), whereas in other studies, a rapid LD decay was observed for the annuum [57,85]. It should be noted that LD, as a phenomenon itself, is affected by multiple factors including the degree of inbreeding, population structure, recombination, rare alleles, etc., where one factor is not considered more dominant over the other [11,74,86]. The presence of potential selective sweeps, as demonstrated by the Tajima's D values, could have also affected the genome-wide LD on the respective species. Among the PCA groups, Cluster I, comprised of the *baccatum*, *chinense*, and *frutescens* complex, had the lowest value for Tajima's D. Similarly, the *C. frutescens* and *C. chinense* have been observed previously to have low Tajima's D [57,87], indicative of the presence of rare alleles and differences in allele frequencies, which could have resulted in a more extensive LD compared to the other species. A kinship-principal component (*K-PC*) model (number of PC = 3) was utilized to correct for genetic (population) stratification, that can cause genome-wide LD between unlinked loci consequently resulting to confounding effects for GWAS [29], to identify QTL for yield and agronomic traits in chile pepper.

Co-localization of Quantitative Trait Loci with Previously Identified Significant Marker-trait Associations

In the past, only few single-locus association mapping studies that focused exclusively on a set of traits such as capsaicinoid content, fruit morphology, and fruit weight have been conducted in chile pepper, where QTL in chromosomes P1, P2, P3, P4, P7, P8, P9, and P10, among others, have been identified [21–23,88,89]. Among these traits, five candidate genes, namely, *pAMT*, *C4H*, *4Cl*, *CSE*, and *FatA*, involved in various biosynthetic pathways, were identified for capsaicinoid content [88]. Phenotypes such as total yield, yield components, and phenology-related traits, including FT, have not been previously evaluated. In the present work, 215 QTL were identified across the 12 chromosomes of chile pepper for yield and agronomic traits using six multi-locus methodologies. Among these QTL, eight multi-trait loci distributed in six chromosomes were identified across multiple GWAS models and environments, which can be targeted for future molecular breeding. These identified SNP markers can be used for designing allele specific assays to be used in MAS.

Yield and yield components

In this study, the TYP and its component trait, GYP shared genetic control on chromosomes P2 and P6, whereas the phenology-related traits, FPD and FT, had common QTL on chromosomes P1, P6, and P7. These QTL colocalized with markers previously identified to be associated with various agronomic traits in chile pepper, indicating potential effects of linkage due to genomic location or pleiotropy. The QTL *SCM002817.1_189395041*, associated with both GYP and TYP, for example,

is located upstream of a cluster of SNP markers associated with fruit weight in chromosome P6 of *C. annuum* [21].

Single-trait QTL such as *SCM002819.1_133495639*, associated with TPW in chromosome P8 (Supplementary Table S6), was mapped closed (~0.09 Mb) to a SNP marker previously related to weight per fruit in a pepper mapping population [90]. Two other QTL for TPW, namely, *SCM002816.1_169617895* and *SCM002823.1_197062622*, were mapped in proximity (~1.08 and 3.73 Mb, respectively) to loci associated with weight per fruit in chromosomes P5 and P12 [90]. A study by Lee et al. [23] identified a fruit weight-related locus, *S06_194967541*, on chromosome P6 downstream of the QTL *SCM002817.1_189395041*, associated with TYP and GYP in the NM-CAMP. More similar to yield, significant QTLs for total fruit weight were reported on chromosomes P1, P2, and P3 by Dwivedi et al. [91] using biparental mapping. High broad-sense heritability ($H^2 = 0.97$) has been reported for fruit shape (i.e., fruit length divided by fruit width) in chile pepper [92], and a QTL on chromosome P3 has been reported to contribute to this trait [93], in addition to QTLs on P8 and P11. We observed high H^2 value (0.88; COM) for TPW, which is supported by previous work demonstrating a strong genotype contribution to the variability of fruit weight, as compared to the environment [94]. A single-trait QTL, *SCM002819.1_133495639*, associated with TPW on chromosome P8, was mapped closed (~0.09 Mb) to a SNP marker previously related to weight per fruit in a pepper mapping population [90], but was in contrast to previous findings [91,93,95].

Plant architecture and morphology

Plant height for the NM-CAMP has been found to have moderate to high broad-sense heritability, ranging between 0.61 and 0.97. These results were consistent with previous observations for PHT, where an H^2 value of 0.91 and narrow-sense heritability ($h^2 = 0.55$) suggesting additive effects, were reported [92]. Utilizing a RIL population, Han et al. [93] identified plant height associated QTLs on chromosomes P2, P4, P6, P7, and P8 as well as QTLs on P2 and P5 associated with plant width, supporting our findings of significant QTL on chromosomes P2 and P4 for plant height and width. Conversely, Dwivedi et al. [91] identified a QTL on P5 contributing to 47% of the total phenotypic variance for plant height. Altogether, our results indicated that QTL controlling these traits across different pepper populations are located on similar genomic regions suggesting pleiotropy and/or linkage due to genetic distance and location. Moreover, we demonstrated the potential of multi-trait genetic improvement in chile pepper by targeting breeding and selection across these specific regions of the genome.

Epigenetic Factors Might Play a Role for Phenology-Related Traits in Chile Pepper

The genetic complexity of the evaluated traits was further revealed by the diverse functions of the candidate genes predicted for the identified QTL. Epigenetic mechanisms including methylation and chromatin remodeling and organization were predicted for the candidate genes for the QTL *SCM002817.1_37397491* mapped on chromosome P6 and associated with FPD and FT. The role of epigenetic factors in gene expression in chile pepper is currently not well established, although a denser methylation profile for chile pepper compared to the potato (*Solanum tuberosum*) and tomato (*S. lycopersicum*) genomes has been documented [96]. Recent evidence on the potential involvement of epigenetic mechanisms on different traits (e.g., disease resistance) in *Capsicum* have been reported. Results from transcriptomic, QTL mapping, and meta-analyses supported the putative effects of these epigenetic factors, particularly of their association with a cluster of major *Phytophthora capsici* resistance loci in the short arm of chromosome P5 of chile pepper [57,97,98]. The involvement of non-coding RNAs (ncRNAs) and cytosine methylation in regulating fruit development and ripening in chile pepper has also been suggested [99,100]. It might be relevant, therefore, to examine the epigenomic landscape of the *Capsicum* spp. to gain a better understanding of gene expression and how this process is regulated by different epigenetic mechanisms to affect the terminal phenotype.

CONCLUSIONS

The potential of using multi-locus GWAS approach in dissecting the genetic architecture of complex traits in chile pepper was demonstrated in the current study. Genome-wide QTL were identified for yield, plant morphology, and phenology-related traits. Candidate gene analysis revealed diverse functions related to different biological processes including oxidation-reduction, phosphorylation, defense response, and epigenetic mechanisms demonstrating the genetic complexity of the evaluated traits. The QTL identified in this study will assist in developing molecular markers and genome-wide prediction for breeding and selection of traits toward genetic improvement in *Capsicum*. Validation of the QTL identified from GWAS will be performed using biparental and other diverse populations in the NMSU chile pepper breeding program.

SUPPLEMENTARY MATERIALS

The following supplementary materials are available online, Supplementary Table S1: Genotypes for the New Mexico *Capsicum* association mapping population and their inferred clusters based on hierarchical and Bayesian-based model clustering for number of clusters, $K = 5$. The inferred group designation for each genotype is highlighted.

Supplementary Table S2: Mean of squares from analysis of variance (ANOVA) for yield and agronomic traits in chile pepper. Supplementary Table S3: Phenotypic variation for the New Mexico *Capsicum* association mapping panel. Supplementary Table S4: Overview of linkage disequilibrium for the intrachromosomal marker pairs in the New Mexico *Capsicum* association mapping population. Supplementary Table S5: Coefficient of linkage disequilibrium (r^2) for intrachromosomal SNP loci associated with yield and other agronomic traits in significant LD ($P < 0.05$). Supplementary Table S6: Significant SNP markers identified for the New Mexico *Capsicum* association mapping panel using multi-locus GWAS approaches. Supplementary Table S7: Candidate genes for the multi-trait quantitative trait loci (QTL) and their potential biological functions for *C. annuum*.

DATA AVAILABILITY

The dataset of the study is available from the authors upon reasonable request.

AUTHOR CONTRIBUTIONS

DNL conceived this research, analyzed the genotype and phenotype data, and wrote the first draft of the manuscript. DWB co-wrote the manuscript. MB performed population structure and genetic diversity analyses. DC collected phenotypic data and processed samples for DNA extraction and genotyping-by-sequencing. PWB edited the manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

This research was funded by USDA-Hatch Program, Accession # 1025360 and USDA-NIFA Grant No. 2022-67014-37078.

ACKNOWLEDGMENTS

The authors would like to thank Samuel Hernandez, Zachary Edwards, Guillermo Nunez, and Charles Havlik for their assistance in collecting phenotypic data for the NM-CAMP.

REFERENCES

1. Lozada DN, Bosland P, Barchenger DW, Haghshenas-Jaryani M, Sanogo S, Walker S. Chile Pepper (*Capsicum*) Breeding and Improvement in the “Multi-Omics” Era. *Front Plant Sci.* 2022;13:879182.
2. Food and Agriculture Organization of the United Nations. FAOSTAT statistics database. Available from: <http://www.fao.org/faostat/en/#data/QCL>. Accessed 2022 May 16.

3. Culler K. New Mexico Chiles: History, Legend and Lore. Charleston (US): Arcadia Publishing; 2015.
4. Walker SJ, Funk P, Joukhadar I, Place T, Havlik C, Tonnessen B. 'NuMex Odyssey', a New Mexico-type Green Chile Pepper for Mechanical Harvest. *HortScience*. 2021;56(12):1605-7.
5. Barchenger DW, Lamour KH, Bosland PW. Challenges and strategies for breeding resistance in *Capsicum annuum* to the multifarious pathogen, *Phytophthora capsici*. *Front Plant Sci*. 2018;9:628.
6. Sammis TW, Shukla MK, Mexal JG, Bosland PW, Daugherty LA. Improving the chile industry of New Mexico through industry, agriculture experiment station, and Cooperative Extension Service collaboration: A case study. *J Ext*. 2009;47(1):1RIB4.
7. USDA National Agricultural Statistics Service. 2021 New Mexico Chile Production. Available from: https://www.nass.usda.gov/Statistics_by_State/New_Mexico/Publications/Special_Interest_Reports/NM-2021-Chile-Production.pdf. Accessed 2022 May 16.
8. Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat Genet*. 2014;46(3):270-8.
9. Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc Natl Acad Sci*. 2014;111(14):5135-40.
10. Hulse-Kemp AM, Maheshwari S, Stoffel K, Hill TA, Jaffe D, Williams SR, et al. Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hort Res*. 2018;5(1):1-13.
11. Zhu C, Gore M, Buckler ES, Yu J. Status and prospects of association mapping in plants. *Plant Genome*. 2008;1(1):5-20.
12. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994;265(5181):2037-48.
13. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
14. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-9.
15. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*. 2006;38(2):203-8.
16. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient control of population structure in model organism association mapping. *Genetics*. 2008;178(3):1709-23.
17. Zhang Z, Ersoz E, Lai C, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet*. 2010;42(4):355-60.
18. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. 2010;42(4):348-54.

19. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 2012;44(7):821-4.
20. Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ, et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci Rep.* 2016;6(1):1-10.
21. Colonna V, D'Agostino N, Garrison E, Albrechtsen A, Meisner J, Facchiano A, et al. Genomic diversity and novel genome-wide association with fruit morphology in *Capsicum*, from 746k polymorphic sites. *Sci Rep.* 2019;9(1):1-14.
22. Nimmakayala P, Abburi VL, Saminathan T, Alaparathi SB, Almeida A, Davenport B, et al. Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum annum* L. *Sci Rep.* 2016;6(1):38081.
23. Lee HY, Ro NY, Patil A, Lee JH, Kwon JK, Kang BC. 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.* 2020;11:1100.
24. Siddique MI, Lee HY, Ro NY, Han K, Venkatesh J, Solomon AM, et al. Identifying candidate genes for *Phytophthora capsici* resistance in pepper (*Capsicum annum*) via genotyping-by-sequencing-based QTL mapping and genome-wide association study. *Sci Rep.* 2019;9(1):9962.
25. Zhang YM, Jia Z, Dunwell JM. The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front Plant Sci.* 2019;10:100.
26. Su J, Ma Q, Li M, Hao F, Wang C. Multi-locus genome-wide association studies of fiber-quality related traits in Chinese early-maturity upland cotton. *Front Plant Sci.* 2018;9:1169.
27. Li C, Fu Y, Sun R, Wang Y, Wang Q. 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.* 2018;9:1083.
28. Tibbs-Cortes L, Zhang Z, Yu J. Status and prospects of genome-wide association studies in plants. *Plant Genome.* 2021;14(1):e20077.
29. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet.* 2012;44(7):825-30.
30. Goddard ME, Wray NR, Verbyla K, Visscher PM. Estimating effects and making predictions from genome-wide marker data. *Stat Sci.* 2009;24(4):517-29.
31. Peng Y, Liu H, Chen J, Shi T, Zhang C, Sun D, et al. Genome-wide association studies of free amino acid levels by six multi-locus models in bread wheat. *Front Plant Sci.* 2018;9:1196.
32. Malik P, Kumar J, Singh S, Sharma S, Meher PK, Sharma MK, et al. Single-trait, multi-locus and multi-trait GWAS using four different models for yield traits in bread wheat. *Mol Breed.* 2021;41(8):1-21.

33. Zhang Y, Liu P, Zhang X, Zheng Q, Chen M, Ge F, et al. Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. *Front Plant Sci.* 2018;9:611.
34. Misra G, Badoni S, Domingo CJ, Cuevas RPO, Llorente C, Mbanjo EGN, et al. Deciphering the genetic architecture of cooked rice texture. *Front Plant Sci.* 2018;9:1405.
35. Zhong H, Liu S, Sun T, Kong W, Deng X, Peng Z, et al. Multi-locus genome-wide association studies for five yield-related traits in rice. *BMC Plant Biol.* 2021;21(1):1-12.
36. Lü H, Yang Y, Li H, Liu Q, Zhang J, Yin J, et al. Genome-wide association studies of photosynthetic traits related to phosphorus efficiency in soybean. *Front Plant Sci.* 2018;9:1226.
37. Fang Y, Liu S, Dong Q, Zhang K, Tian Z, Li X, et al. Linkage analysis and multi-locus genome-wide association studies identify QTNs controlling soybean plant height. *Front Plant Sci.* 2020;11:9.
38. Liu R, Gong J, Xiao X, Zhang Z, Li J, Liu A, et al. GWAS analysis and QTL identification of fiber quality traits and yield components in upland cotton using enriched high-density SNP markers. *Front Plant Sci.* 2018;9:1067.
39. Nakayama R. Notice of the Naming and Release of 'NuMex Big Jim,' a Semi-Mild Pungent Chile Variety for New Mexico. Las Cruces (US): New Mexico State University Agricultural Experiment Station; 1975.
40. Bosland PW, Coon D. 'NuMex Heritage Big Jim' New Mexican Chile Pepper. *HortScience.* 2013;48(5):657-8.
41. Bosland PW, Coon D. 'NuMex Sandia Select' New Mexican Chile Pepper. *HortScience.* 2014;49(5):667-8.
42. Bosland PW, Iglesias J, Gonzalez MM. 'NuMex Joe E. Parker' Chile. *HortScience.* 1993;28(4):347-8.
43. Bosland PW. 'NuMex Heritage 6-4' New Mexican Chile Pepper. *HortScience.* 2012;47(5):675-6.
44. Bosland PW. 'NuMex Vaquero' Jalapeno. *HortScience.* 2010;45(10):1552-3.
45. Bosland PW, Strausbaugh CA. 'NuMex Las Cruces' Cayenne Pepper. *HortScience.* 2010;45(11):1751-2.
46. Walker SJ. 'NuMex R. Vince Hernandez' Paprika-type Chile. *HortScience.* 2015;50(12):1820-1.
47. Coon D, Bosland PW. The chile cultivars of New Mexico State University released from 1913 to 2016. Available from: <https://www.cabdirect.org/cabdirect/abstract/20183214356>. Accessed 2022 May 17.
48. Bosland PW, Coon D, Reeves G. 'Trinidad Moruga Scorpion' pepper is the world's hottest measured chile pepper at more than two million Scoville heat units. *Horttechnology.* 2012;22(4):534-8.
49. Bosland PW, Coon D. 'NuMex NoBasco': A No-heat Tabasco-type Chile Pepper. *HortScience.* 2020;55(5):741-2.
50. National Oceanic and Atmospheric Administration. National Climatic Data Center Climate Data Online Search. Available from: <https://www.ncdc.noaa.gov/cdo-web/search>. Accessed 2022 May 10.

51. Rodríguez F, Alvarado G, Pacheco Á, Burgueño J. “ACBD-R. Augmented Complete Block Design with R for Windows. Version 4.0”. Available from: <https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10855>. Accessed 2022 May 16.
52. Aravind J, Mukesh Sankar S, Wankhede DP, Kaur V. augmentedRCBD: Analysis of Augmented Randomised Complete Block Designs (R package version 0.1.3). Available from: <https://mran.microsoft.com/snapshot/2020-09-11/web/packages/augmentedRCBD/readme/README.html>. Accessed 2022 May 17.
53. R Development Core Team. R: A Language and Environment for Statistical Computing (Internet). Available from: <https://www.r-project.org/>. Accessed 2022 May 17.
54. Jahufer MZZ, Luo D. DeltaGen: A comprehensive decision support tool for plant breeders. *Crop Sci.* 2018;58(3):1118-31.
55. JMP® 13.2.1. Cary (US): SAS Institute Inc.; 1989-2021.
56. Sharma H, Shukla MK, Bosland PW, Steiner R. Soil moisture sensor calibration, actual evapotranspiration, and crop coefficients for drip irrigated greenhouse chile peppers. *Agric Water Manag.* 2017;179:81-91.
57. Lozada DN, Bhatta M, Coon D, Bosland PW. Single nucleotide polymorphisms reveal genetic diversity in New Mexican chile peppers (*Capsicum* Spp.). *BMC Genom.* 2021;22:356.
58. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-20.
59. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25(14):1754-60.
60. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. Available from: <https://arxiv.org/pdf/1207.3907.pdf>. Accessed 2022 May 16.
61. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007;23(19):2633-5.
62. Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong GY, Myles S. LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3.* 2015;5(11):2383-90.
63. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123(3):585-95.
64. Meirmans PG. Genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Mol Ecol Resour.* 2020;20(4):1126-31.
65. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* 1992;131(2):479-91.
66. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ.* 2014;2:e281.

67. Graebner RC, Hayes PM, Hagerty CH, Cuesta-Marcos A. A comparison of polymorphism information content and mean of transformed kinships as criteria for selecting informative subsets of barley (*Hordeum vulgare* L. sl) from the USDA Barley Core Collection. *Genet Resour Crop Evol.* 2016;63(3):477-82.
68. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945-59.
69. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005;14(8):2611-20.
70. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 2012;4(2):359-61.
71. Criscuolo NG, Angelini C. StructuRly: A novel shiny app to produce comprehensive, detailed and interactive plots for population genetic analysis. *PLoS One.* 2020;15(2):e0229330.
72. Weir BS, Cockerham C. Genetic data analysis II: Methods for discrete population genetic data. Sunderland (US): Sinauer Associates Inc.; 1996.
73. Hill WG, Weir BS. Variances and covariances of squared linkage disequilibria in finite populations. *Theor Popul Biol.* 1988;33(1):54-78.
74. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci.* 2001;98(20):11479-84.
75. Zhang YW, Tamba CL, Wen YJ, Li P, Ren WL, Ni YL, et al. mrMLM v4. 0: An R platform for multi-locus genome-wide association studies. *Genomics Proteomics Bioinformatics.* 2020;18(4):481-87.
76. Tamba CL, Zhang YM. A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv* 341784 [Preprint]. 2018 Jun 7. doi: 10.1101/341784
77. Wen YJ, Zhang H, Ni YL, Huang B, Zhang J, Feng JY, et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief Bioinform.* 2018;19(4):700-12.
78. Zhang J, Feng JY, Ni YL, Wen YJ, Niu Y, Tamba CL, et al. pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity.* 2017;118(6):517-24.
79. Ren WL, Wen YJ, Dunwell JM, Zhang YM. pKWmEB: integration of Kruskal-Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity.* 2018;120(3):208-18.
80. Tamba CL, Ni YL, Zhang YM. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput Biol.* 2017;13(1):e1005357.
81. Bolser D, Staines DM, Pritchard E, Kersey P. Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. In: Edwards D, editors. *Plant bioinformatics.* New York (US): Humana Press; 2016. p. 115-40.
82. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet.* 2001;28(3):286-9.

83. Brachi B, Morris GP, Borevitz JO. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 2011;12(10):1-8.
84. Parry C, Wang YW, Lin S, Barchenger DW. Reproductive compatibility in *Capsicum* is not necessarily reflected in genetic or phenotypic similarity between species complexes. *PLoS One.* 2021;16(3):e0243689.
85. Taranto F, D'Agostino N, Greco B, Cardi T, Tripodi P. Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. *BMC Genom.* 2016;17(1):943.
86. Gaut BS, Long AD. The lowdown on linkage disequilibrium. *Plant Cell.* 2003;15(7):1502-6.
87. Pereira-Dias L, Vilanova S, Fita A, Prohens J, Rodríguez-Burruezo A. Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). *Hort Res.* 2019;6(1):54.
88. Han K, Lee H, Ro N, Hur O, Lee J, Kwon J, et al. QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. *Plant Biotechnol J.* 2018;16(9):1546-58.
89. Tripodi P, Rabanus-Wallace MT, Barchi L, Kale S, Esposito S, Acquadro A, et al. Global range expansion history of pepper (*Capsicum* spp.) revealed by over 10,000 genebank accessions. *Proc Natl Acad Sci.* 2021;118(34):e2104315118.
90. Wu L, Wang P, Wang Y, Cheng Q, Lu Q, Liu J, 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.* 2019;20(22):5675.
91. Dwivedi N, Kumar R, Paliwal R, Kumar U, Kumar S, Singh M, et al. QTL mapping for important horticultural traits in pepper (*Capsicum annuum* L.). *J Plant Biochem Biotechnol.* 2015;24(2):154-60.
92. Stommel JR, Griesbach RJ. Inheritance of fruit, foliar, and plant habit attributes in *Capsicum*. *J Am Soc Hortic Sci.* 2008;133(3):396-407.
93. Han K, Jeong HJ, Yang HB, Kang SM, Kwon JK, Kim S, et al. An ultra-high-density bin map facilitates high-throughput QTL mapping of horticultural traits in pepper (*Capsicum annuum*). *DNA Res.* 2016;23(2):81-91.
94. Barchenger DW, Clark RA, Gniffke PA, Ledesma DR, Lin S, Hanson P, et al. Stability of yield and yield components of pepper (*Capsicum annuum*), and evaluation of publicly available predictive meteorological data in East and Southeast Asia. *HortScience.* 2018;53(12):1776-83.
95. Lu FH, Kwon SW, Yoon MY, Kim KT, Cho MC, Yoon MK, et al. SNP marker integration and QTL analysis of 12 agronomic and morphological traits in F₈ RILs of pepper (*Capsicum annuum* L.). *Mol Cells.* 2012;34(1):25-34.
96. Rawoof A, Ahmad I, Ramchiary N. Epigenome landscape in *Capsicum* genome. In: Ramchiary N, Kole C, editors. *The Capsicum Genome*. Cham (Switzerland): Springer; 2019. p. 187-99.
97. Du JS, Hang LF, Hao Q, Yang HT, Ali S, Badawy RSE, et al. The dissection of R genes and locus Pc5. 1 in *Phytophthora capsici* infection provides a novel view of disease resistance in peppers. *BMC Genom.* 2021;22(1):1-16.

98. Lozada DN, Whelpley M, Acuña-Galindo A. Genetic Architecture of Chile Pepper (*Capsicum* spp.) QTLome Revealed Using Meta-QTL Analysis. *Horticulturae*. 2021;7(8):227.
99. Kumar A, Kumar S, Anju T, Ramchiary N. Genetic, Epigenetic, and Hormonal Regulation of Fruit Development and Ripening in *Capsicum* L. Species. *Annu Plant Rev*. 2018;4(2):295-356.
100. Yang S, Yang T, Tang Y, Aisimutuola P, Zhang G, Wang B, et al. Transcriptomic profile analysis of non-coding RNAs involved in *Capsicum chinense* Jacq. fruit ripening. *Sci Hortic*. 2020;264:109158.

How to cite this article:

Lozada DN, Barchenger DW, Coon D, Bhatta M, Bosland PW. Multi-locus Association Mapping Uncovers the Genetic Basis of Yield and Agronomic Traits in Chile Pepper (*Capsicum* spp.). *Crop Breed Genet Genom*. 2022; 4(2):e220002. <https://doi.org/10.20900/cbgg20220002>