the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2023) doi: 10.1111/tpj.16425

RESOURCE

Multi-environment association study highlights candidate genes for robust agronomic quantitative trait loci in a novel worldwide *Capsicum* core collection

Louis McLeod[1,2], Lorenzo Barchi[3,*], Giorgio Tumino[4], Pasquale Tripodi[5], Jérémy Salinier[1,†], Christophe Gros[1], Hatice Filiz Boyaci[6,‡], Ramazan Ozalp[6], Yelena Borovsky[7], Roland Schafleitner[8], Derek Barchenger[8], Richard Finkers[4,§], Matthijs Brouwer[4], Nils Stein[9,10], Mark Timothy Rabanus-Wallace[9,¶], Giovanni Giuliano[11], Roeland Voorrips[4], Ilan Paran[7] and Véronique Lefebvre[1,*]

[1]*INRAE, GAFL, Montfavet, France,*
[2]*INRAE, A2M, Montfavet, France,*
[3]*Department of Agricultural, Forest and Food Sciences (DISAFA), Plant Genetics, University of Torino, Grugliasco, Italy,*
[4]*Plant Breeding, Wageningen University and Research (WUR), Wageningen, The Netherlands,*
[5]*Research Centre for Vegetable and Ornamental Crops, Council for Agricultural Research and Economics (CREA), Pontecagnano Faiano, Italy,*
[6]*Bati Akdeniz Agricultural Research Institute (BATEM), Antalya, Türkiye,*
[7]*The Volcani Center, Institute of Plant Sciences, Agricultural Research Organization (ARO), Rishon LeZion, Israel,*
[8]*Vegetable Diversity and Improvement, World Vegetable Center, Shanhua, Taiwan,*
[9]*Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland,Corre, Gatersleben, Germany,*
[10]*Department of Crop Sciences, Center for Integrated Breeding Research, Georg-August-University, Göttingen, Germany,*
[11]*Casaccia Research Centre, Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), Rome, Italy*

Received 27 April 2023; revised 13 July 2023; accepted 4 August 2023.
*For correspondence (e-mail veronique.lefebvre@inrae.fr; lorenzo.barchi@unito.it).
†Present address: CIRAD, UMR PVBMT, Saint-Pierre, La Réunion, France
‡Present address: Department of Horticulture, Faculty of Agriculture, University of Recep Tayyip Erdogan, Rize, Türkiye
§Present address: GenNovation B.V., Wageningen, The Netherlands
¶Present address: Faculty of Science, University of Melbourne, Parkville, Australia

SUMMARY

Investigating crop diversity through genome-wide association studies (GWAS) on core collections helps in deciphering the genetic determinants of complex quantitative traits. Using the G2P-SOL project world collection of 10 038 wild and cultivated *Capsicum* accessions from 10 major genebanks, we assembled a core collection of 423 accessions representing the known genetic diversity. Since complex traits are often highly dependent upon environmental variables and genotype-by-environment (G × E) interactions, multi-environment GWAS with a 10 195-marker genotypic matrix were conducted on a highly diverse subset of 350 *Capsicum annuum* accessions, extensively phenotyped in up to six independent trials from five climatically differing countries. Environment-specific and multi-environment quantitative trait loci (QTLs) were detected for 23 diverse agronomic traits. We identified 97 candidate genes potentially implicated in 53 of the most robust and high-confidence QTLs for fruit flavor, color, size, and shape traits, and for plant productivity, vigor, and earliness traits. Investigating the genetic architecture of agronomic traits in this way will assist the development of genetic markers and pave the way for marker-assisted selection. The G2P-SOL pepper core collection will be available upon request as a unique and universal resource for further exploitation in future gene discovery and marker-assisted breeding efforts by the pepper community.

Keywords: multi-environment trials, association study, core collection, pepper, *Capsicum annuum* L., agronomic traits, candidate genes.

INTRODUCTION

Most yield- and quality-related traits are inherited quantitatively and governed by multiple genes (Bertin et al., 2010; Rao et al., 2003). Genetic dissection of quantitative traits has mostly been performed by mapping quantitative trait loci (QTLs) underlying their variation (Phan & Sim, 2017). Traditional QTL analysis approaches are based on dedicated mapping progenies segregating for the traits of interest, constructed from controlled crosses between two or more parental lines exhibiting diverse phenotypes (Korte & Farlow, 2013). These mapping studies are therefore time- and resource-intensive and are limited to the genetic and phenotypic variation existing between the founder lines (Phan & Sim, 2017). The emergence of genome-wide association studies (GWAS) based on collections of ideally distantly-related accessions and the development of cutting-edge next-generation sequencing technologies are now ways to circumvent these limitations (Phan & Sim, 2017). Additionally, GWAS increases the accuracy of QTL localization by tracking a higher number of historically accumulated recombinations, as compared to methods that rely on the meiosis of parental lines only (Perez-de-Castro et al., 2012). However, relatedness between accessions and population stratification can lead to spurious associations (Price et al., 2006). To correct this, GWAS models often incorporate principal components (PCs), a kinship matrix, or an estimation of global ancestry, all derived from large sets of markers (Price et al., 2010; Zhu et al., 2008).

The concept of core collection was first introduced by Frankel (1984). It refers to limited sets of accessions representing the greatest possible portion of genetic variation found in a crop species, including its wild progenitors. They are typically subsets of larger collections, designed for use in experiments and analyses requiring a survey of genetic diversity in situations where working on the entire set would be prohibitively laborious or computationally demanding. Owing to their compactness and diversity, core collections can be phenotyped for a wider range of traits than traditional mapping progenies, and contribute to the conservation and accessibility of plant genetic resources (Brown & Spillane, 1999).

The pepper fruit is a seeded fleshy berry grown in all populated continents, mainly between the two tropics. Over 40 *Capsicum* species have been recognized so far (Barboza et al., 2022). Five of these species are domesticated, including the widely-cultivated and comparatively well-studied *Capsicum annuum* (Pereira-Dias et al., 2019). Two major types dominate the diversity in *C. annuum*: the small-fruited and pungent chili peppers and the large-fruited non-pungent sweet peppers (Pereira-Dias et al., 2019; Tripodi et al., 2021). While domestication and breeding have been mainly focused on yield improvement of these types, numerous additional traits contribute to the production of desirable cultivated varieties, including shoot architecture, phenology, flowering time, fruit morphology, fruit quality, metabolite composition, resistance to biotic stresses and tolerance to abiotic stresses (IPGRI et al., 1995; Junior e Silva et al., 2013). The diverse natural variation observed within *Capsicum* has enabled the characterization of inherited beneficial variants for such traits, using a range of mapping approaches and genetic marker technologies (e.g., Lefebvre et al., 2002; Lippert et al., 1965; Paran & Fallik, 2011; Wang & Bosland, 2006; see Table S1 for an up-to-date summary). Large numbers of QTLs to be implemented for pepper improvement were located throughout the genome (Arpaci & Karataş, 2020; Jeong et al., 2015; Lin et al., 2015; Parsons et al., 2012; Rani et al., 2021; Thabuis et al., 2004).

High-throughput single nucleotide polymorphism (SNP) technologies such as genotyping-by-sequencing (GBS) have recently been used to describe the genetic structure of the pepper germplasm (Colonna et al., 2019; Hill et al., 2013; Lee, Ro, et al., 2020; Lozada et al., 2022; Ro et al., 2022; Tamisier et al., 2020; Taranto et al., 2016; Wu et al., 2019). Most recently, the G2P-SOL project produced a comprehensive survey of pepper diversity, describing the genetic structure of over 10 000 wild and cultivated accessions housed in 10 genebanks (Tripodi et al., 2021). Several core collections and GWAS have been reported in *C. annuum* (for an up-to-date summary, see Table S2). However, none has yet leveraged this extent of diversity to define a core collection suitable to perform a maximally powerful GWAS. Moreover, GWAS to date has relied upon phenotypic data gathered in a single or a limited number of trials within a given country (Table S2). Complex traits, however, are generally modulated by the environment and the interaction between genotype (*sensu* accession) and environment (G × E) (Kang, 2004). The latter especially impact crop research and breeding and necessitate replication of the trials in different locations in order to efficiently and reliably develop locally adapted varieties (Kang, 2020).

We report the compilation and analysis of a novel pepper core collection of 423 accessions, representing the accessible global pepper diversity. Twenty-three plant growth and fruit traits were phenotyped in 350 *C. annuum* accessions of this core collection in multiple environments. GWAS led to the identification of novel and previously reported QTLs, which may be used for breeding and for functional genomics studies identifying the causal genes underlying trait variation. These results demonstrated the usefulness and universality of the core collection as a worldwide research and breeding resource designed for mining the pepper genome and germplasm.

RESULTS

The G2P-SOL core collection captures the diversity of the worldwide germplasm

The reference G2P-SOL pepper core collection

A core collection of *Capsicum* spp. consisting of 423 accessions was designed to capture the genetic variability present in the *Capsicum* panel of 10 038 accessions analyzed by Tripodi et al. (2021) by minimizing a repetitiveness criterion (Table S3). These 423 accessions comprise 391 *C. annuum* accessions and 32 accessions from other cultivated *Capsicum* species, including several likely derived from interspecific hybridizations. Principal component analyses (PCA), either based on the 10038 *Capsicum* accessions panel or on its 7848 *C. annuum* accessions subset, show that the core collection and its subset of 391 *C. annuum* accessions spanned the whole *Capsicum* and *C. annuum* germplasms, respectively (Figure 1a,b).

C. annuum *accessions and markers used for GWAS*

The SNP calling and filtering parameters applied to the core collection and the GBS markers resulted in 350 *C. annuum* accessions and 10 195 markers (9546 SNPs and 649 Indels) exploitable for GWAS (Table S3). The post-filtering marker set had an average density of 3.2 markers/Mb with a mean minor allele frequency (MAF) of 12.8%, a 1.5% rate of heterozygosity, and 4.9% missing data at the marker level (Figure S1). The accessions also had on average a 1.5% rate of heterozygosity and 4.9% missing data. On average, the linkage disequilibrium (LD) estimated for each chromosome decayed below $r^2 = 0.2$ at 0.4 Mb (Figure S2).

The 350 *C. annuum* accessions used for GWAS were, like the entire set of 391 *C. annuum* accessions, representative of the whole *C. annuum* germplasm (Figure 1b), and there was no evidence of strong stratification in the PCA conducted solely on these 350 accessions (Figure 2). However, the negative correlation between the multi-environment best linear unbiased predictions (BLUPs) for fruit weight (FWe) and pungency (FP) confirms the existence of two major types (large-sweet versus small-spicy) dominating the *C. annuum* diversity (Figure 2a,b). Investigation of the population structure by Admixture revealed that $K = 11$ was a sensible model choice, according to the cross-validation (CV) error (Figure S3). The PCA colored according to the geographical origin and to the observed K subpopulations highlighted that the 350 accessions were partly structured according to their geographical origin (Figure 2c,d), with clusters 7 and 11 roughly corresponding to the European accessions, clusters 3 and 6 to the Eastern Asian and Middle Eastern accessions, respectively, clusters 2 and 4 to part of the Eastern European accessions, clusters 8 and 9 to part of the Southern-Southeastern Asian and Central American accessions, respectively, and finally cluster 5 to part of the Southern-Southeastern Asian and Middle Eastern accessions.

GWAS identified QTLs for 23 plant and fruit development traits

Field traits variation and correlations

The 350 *C. annuum* accessions were phenotyped for 23 traits in six independent trials performed in five countries, thus showing a rich phenotypic variability suitable for GWAS (Table 1; Tables S4 and S5). Linear models were fitted to assess the contributions of environment, genotype (*sensu* accession), and G × E interactions to the measured trait values. The residual variances were comprised between 2% and 13% of the total variance for all traits except locule number (LN, 32% of the total variance), indicating a sound fit (Figure 3). For most traits, the amount of total variance attributed to the genotype (accession), was largely predominant (from 27 to 93% of the total variance), while for Brix (BX) and for fruit load (FLo), it was lower than the proportion of variance attributed to the environment and to the G × E interactions, respectively. Fruit maximum width (FWi) was the least influenced by the environment and G × E interactions, while FLo was the most influenced. Broad-sense heritability ranged from 0.75 for total fruit weight (TFWe) to 0.99 for fruit maximum width (FWi), apart from FLo and total fruit number (TFN) with low heritability values of 0.40 and 0.25, respectively (Figure 3). Overall, heritability values were above 0.90 for 15 out of 23 traits.

A PCA plot based on the accession means of the 23 traits illustrates a high level of agreement between trial results (Figure S4). The average correlation values between pairs of trials (here referred to as environments) were high for most traits, especially for fruit maximum width (FWi, 0.95), fruit weight (FWe, 0.94), fruit shape index (FShI, 0.93), fruit maximum length (FLe, 0.92) and pericarp thickness (PTh, 0.88) (Figure S5). Some trials with lower correlations were omitted from downstream analyses for five traits (FF, fruit fasciation; LN, locule number; FlT, flowering time; BX, Brix; FLo, fruit load) (Figure 3).

To correct for environment and G × E interaction effects, adjusted means per accession were estimated using a BLUP model. These BLUPs, used as GWAS phenotypic input, were calculated both for single environments and across highly correlated environments. Since the correlations between single- and multi-environment BLUPs were high for all traits, with an average by trait ranging from 0.82 to 0.98 (Table 2), we focused on the multi-environment BLUPs to investigate the correlation between traits (Figure S6a). Based on the observed correlations, all traits except IFCG (external immature fruit color green) could be grouped into two negatively correlated clusters. On the one hand, fruit size and shape traits (FWe, fruit weight; FWi, fruit maximum width; PTh, pericarp thickness;

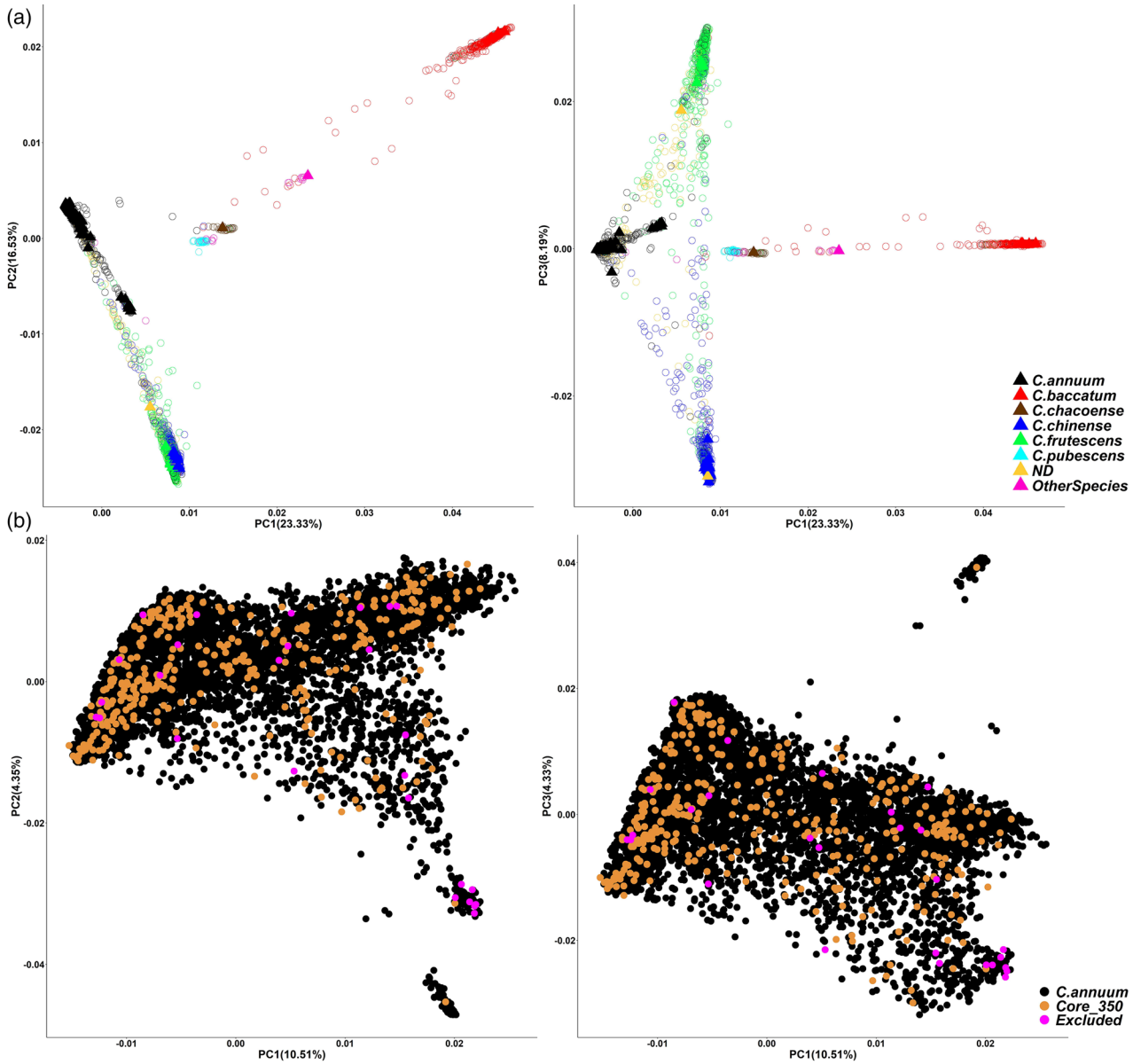

Figure 1. The worldwide genetic diversity of pepper based on single nucleotide polymorphisms.
(a) Plots of the first three components of the principal component analysis of the 10 038 *Capsicum* accessions from Tripodi et al. (2021), with the whole core collection as triangles.
(b) Plots of the first three components of the principal component analysis of the 7848 *Capsicum annuum* accessions, using 1073 linkage disequilibrium-pruned markers with a minor allele frequency superior to 0.05. The 350 *C. annuum* accessions of the core collection that were used for genome-wide association studies are colored in orange, and the rest of the *C. annuum* accessions of the core collection that were filtered out are colored in pink.

FF, fruit fasciation; LN, locule number; FLe, fruit maximum length; FShO, fruit predominant shape oblate) and plant productivity traits (TFWe, total fruit weight; Flo, fruit load) were positively correlated. On the other hand, fruit external color L, a and b traits (FCL, FCa, and FCb), flavor traits (BX, Brix; FP, fruit pungency), a fruit shape trait (FShl, fruit shape index), plant vigor and earliness traits (FlT, flowering time; AL, axis length; PHe, total plant height) and a plant productivity trait (TFN, total fruit number) were positively correlated.

Marker-phenotype associations and QTLs

We scanned for marker-phenotype associations using three models from the GAPIT R package: Mixed Linear Model (MLM), Multi-Locus Mixed-Model (MLMM), and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK). We identified a total of 704 significant marker-phenotype associations corresponding to 453 distinct quantitative trait nucleotides (QTNs) for the 23 studied traits. Associations differed between environments including the multi-environment (Figure 4a). They

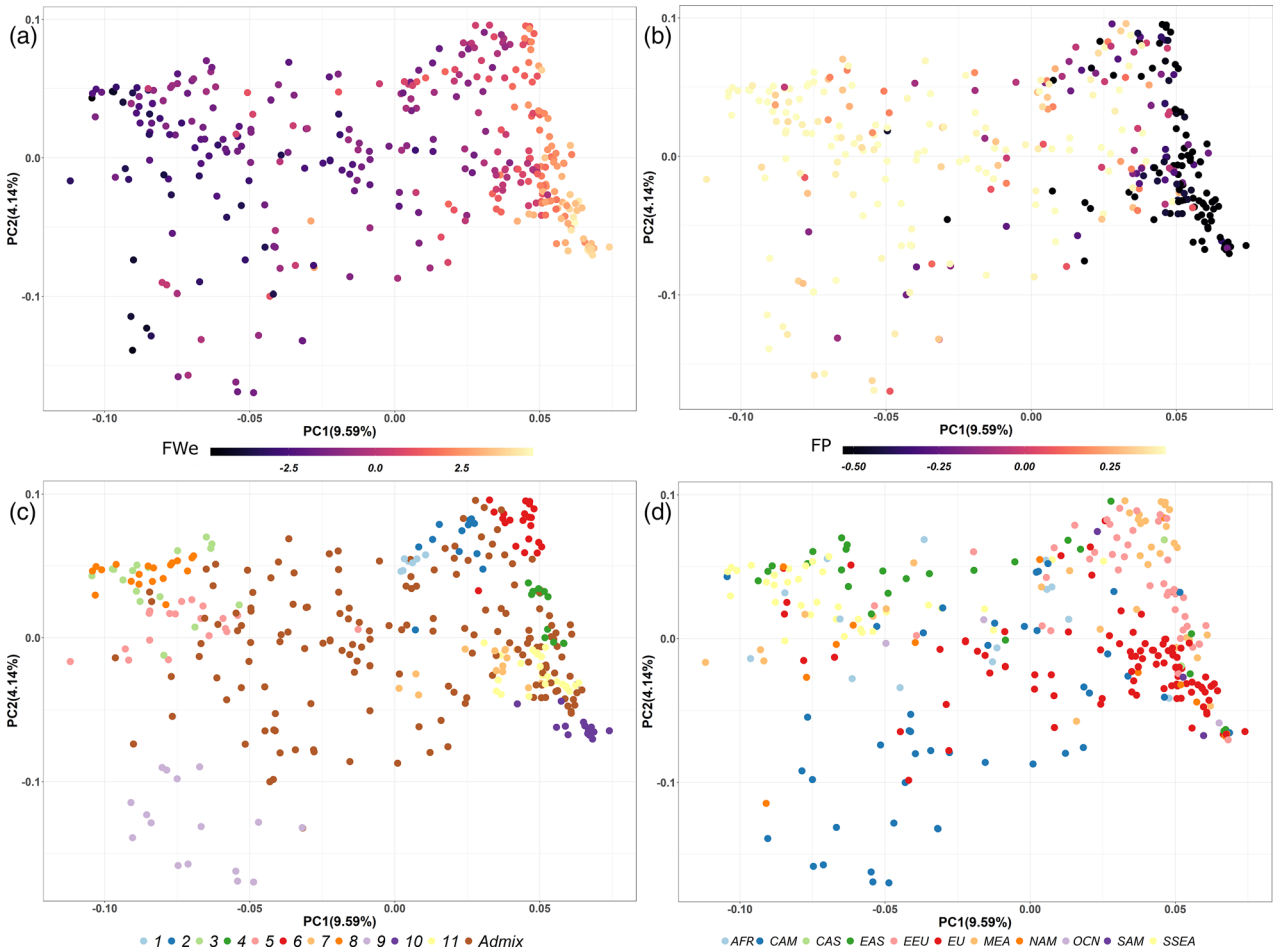

Figure 2. Population structure of the 350 *Capsicum annuum* accessions used for genome-wide association studies. Plot of the first and second component of a principal component analysis based on 2357 pruned single nucleotide polymorphism markers.
(a) Accessions are colored according to their fruit weight (FWe) multi-environment BLUP values.
(b) Accessions are colored according to their fruit pungency (FP) multi-environment BLUP values.
(c) Accessions are colored according to their cluster of structure identified with ADMIXTURE, from K1 to K11.
(d) Accessions are colored according to their geographical origin (AFR, Africa; CAM, Central America; CAS, Central Asia; EAS, Eastern Asia; EEU, Eastern Europe; EU, Europe; MEA, Middle East; NAM, Northern America; OCN, Oceania; SAM, South America; SSEA, Southern-Southeastern Asia).

were primarily found with the BLINK model (499 associations), considerably less with the MLM model (110 associations), and with the MLMM model (95 associations) (Table S6). The 453 QTNs were grouped into 423 QTLs, using the chromosome-specific LD extent value (Figure S2): 332 QTLs for the quantitative traits and 91 QTLs for the ordinal and binary categorical traits (Table S7; Figure S7). Overall, QTL size varied from 600 to 1819 kb (mean: 821 kb). They explained up to 45% of the multi-environment BLUPs variation (R^2), with 19 QTLs explaining more than 20%, 61 QTLs explaining between 10 and 20%, 106 QTLs explaining between 5 and 10%, and 237 QTLs explaining less than 5% (Table S7).

Some of these QTLs were particularly robust, i.e., detected in several environments (Figure 4a; Figure S7). Among the most robust QTLs, *FLe-P10.1* (fruit maximum length), *FShO-P5.6* (fruit predominant shape oblate), *FShI-P10.1* (fruit shape index), and *IFCG-P1.2* (external immature fruit color green) were detected in all investigated environments (*FShI-P10.1* and *IFCG-P1.2* by all three GWAS models in each environment), *FWe-P2.4* (fruit weight) in six of the seven investigated environments, *FShO-P5.3* and *FShO-P5.4* in five of the six investigated environments, and *FShO-P3.3*, *FShO-P6.4*, *FShO-P9.1*, *FShO-P10.1* in four of the six investigated environments. Overall, 411 out of the 423 QTLs controlled traits phenotyped in several environments, of which 74 were robust (18%), 73 (18%) were detected with the multi-environment BLUPs only, and the remaining 264 (64%) were detected in one trial. The proportion of investigated environments in which a QTL was detected was only slightly correlated to the effect of the QTL (R^2) on the multi-environment BLUPs (Pearson correlation = 0.12, P = 0.011), meaning that the QTLs having a major effect on a trait were not systematically robust.

Table 1 Descriptive statistics on the accession means of the 23 traits

Trait			Descriptive statistics[a]						
Code	Name	Unit	Nb. val	Min	Max	Median	Mean	SEM	SD
AL	Axis length	cm	350	8.54	76.54	28.50	30.04	0.47	8.78
BX	Brix	°Bx	350	5.27	13.33	7.75	7.91	0.06	1.21
FCa	Fruit external color a	-	349	−2.25	44.63	34.08	33.36	0.34	6.43
FCb	Fruit external color b	-	349	2.28	70.64	20.40	22.23	0.46	8.64
FCL	Fruit external color L	-	349	26.67	78.25	38.15	39.25	0.30	5.60
FLe	Fruit maximum length	cm	350	1.07	21.08	8.25	8.59	0.23	4.37
FlT	Flowering time	days	350	13.71	63.39	27.16	28.64	0.43	8.11
FShI	Fruit shape index	-	350	0.56	19.53	3.04	3.59	0.14	2.54
FWe	Fruit weight	g	350	0.30	201.55	17.89	37.39	2.38	44.44
FWi	Fruit maximum width	cm	350	0.43	8.55	2.53	3.22	0.11	2.04
IFCa	Immature fruit external color a	-	329	−20.68	5.81	−13.66	−13.43	0.19	3.49
IFCb	Immature fruit external color b	-	329	3.65	49.46	20.53	22.46	0.44	7.90
IFCL	Immature fruit external color L	-	329	28.78	82.60	39.46	42.84	0.53	9.68
LN	Locule number	count	350	1.94	4.31	2.61	2.72	0.03	0.50
PHe	Total plant height	cm	350	43.98	154.22	91.61	92.91	1.11	20.75
PTh	Pericarp thickness	mm	350	0.72	5.97	2.26	2.61	0.07	1.39
TFN	Total fruit number	count	346	0.00	226.25	17.54	24.15	1.32	24.52
TFWe	Total fruit weight	g	347	0.00	1273.33	413.21	431.07	13.86	258.18
FF	Fruit fasciation	1–4	350	1.00	3.77	1.79	1.85	0.03	0.60
FLo	Fruit load	1–3	350	1.40	3.00	2.31	2.30	0.02	0.31
FP	Fruit pungency	0 versus 1	348	0.00	1.00	0.73	0.57	0.02	0.43
FShO	Fruit predominant shape oblate	0 versus 1	350	0.00	1.00	0.00	0.03	0.01	0.13
IFCG	External immature fruit color green	0–4	349	0.58	4.00	3.07	2.80	0.04	0.68

max, maximal value; min, minimal value; Nb. val, number of values; SD, standard deviation; SEM, standard error on the mean.
[a]Outliers were removed before computing the statistics.

Highest-confidence QTLs

Among the 423 QTLs, 114 QTLs which were located on the 12 pepper chromosomes, and 5 QTLs which were located on the artefactual chromosome P0 were considered particularly promising for plant breeding, as they were either robust (i.e., identified in several environments), or identified by several GWAS models, or with a particularly high −log10(*P*-value) (Table S7; Figure 5). Fifty-one of these 114 QTLs affected six of the eight traits never analyzed by GWAS (AL, BX, FCa, FCL, TFN, and FShO). Twenty-one out of the 114 most promising QTLs overlapped on eight distinct chromosomal regions. Three of these regions, on chromosome P1, P2, and P10, included the most robust QTLs for traits related to immature fruit color, to fruit yield and to fruit size and shape, respectively (Table S8; Figure 5). As an example, the fruit maximum length QTL *FLe-P10.1* was detected in all environments (Figure 4a), and accessions of the AA and GG genotypes at the corresponding marker exhibited contrasting phenotypes (Student test, $P = 5.632e\text{-}06$; Figure 4b,c).

Candidate genes were identified for 53 of the highest-confidence QTLs

There was a total of 2790 genes in the intervals of the 114 most promising QTLs mapped on the 12 chromosomes, with an average of 137 genes per trait and 24.5 genes per QTL. We identified 97 candidate genes likely to determine the traits for 53 of these QTLs, based on previously reported biological function (Table S9).

Among the eight candidate genes potentially affecting fruit flavor, the *Pun1* gene and homologs of two MYB transcription factors (*MYB61* and *MYB63*) underlie fruit pungency QTLs *FP-P2.1* and *FP-P7.3*, while a homolog of *SNF1-RELATED PROTEIN KINASE REGULATORY SUBUNIT BETA-2* (*KINB2*) underlies the Brix QTL *BX-P5.2*.

The 28 candidate genes potentially affecting fruit size and shape include three Ovate Family Proteins transcription factors: *C. annuum OVATE FAMILY PROTEIN 20* (*CaOFP20*) for colocalized QTLs *FLe-P10.1*, *FShI-P10.1*, *FShO-P10.1*, *FF-P10.1*, and *PTh-P10.1*; a homolog of transcriptional repressor *OVATE FAMILY PROTEIN 3* (*OFP3*) for *FLe-P6.4*; and a homolog of *OVATE* for *FWe-P2.4*. Among the other strong candidates, a homolog of *FANTASTIC FOUR 3* (*FAF3*) underlies *FLe-P6.2*, a homolog of *BEL1-LIKE HOMEODOMAIN PROTEIN 2* (*BLH2*) underlies *FWe-P2.4*, a homolog of a probable leucine-rich repeat receptor-like protein kinase of the CLAVATA family *At5g63930* underlies *LN-P3.1*, and homologs of *KATANIN P60 ATPASE-CONTAINING SUBUNIT A1* (*AAA1*) and *PURINE PERMEASE 3* (*PUP3*) underlie *FShI-P1.1* and *FShI-P6.3*, respectively.

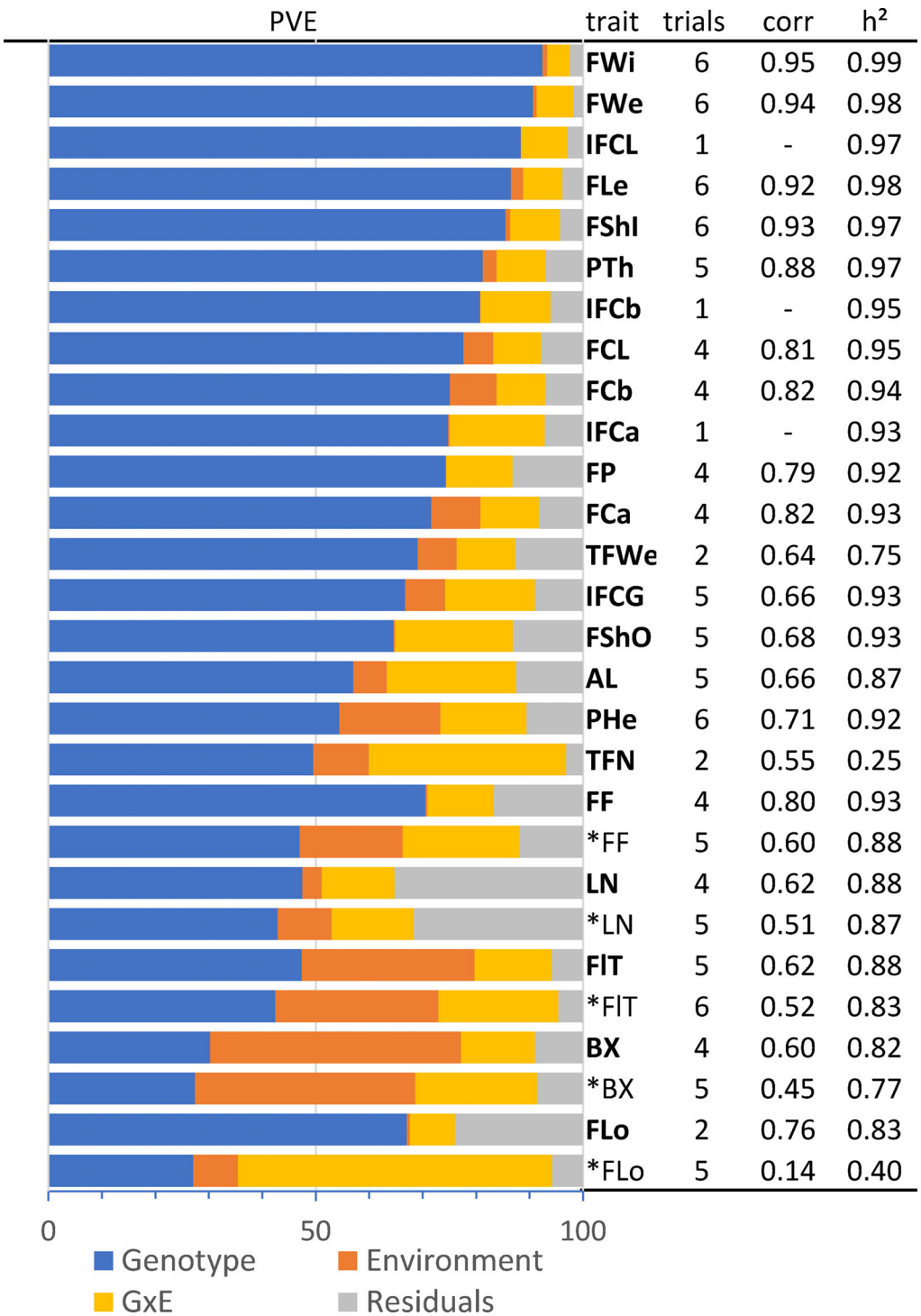

Figure 3. Proportion of phenotypic variance explained by the effects of genotype, environment and their interaction, and heritability values for the 23 agronomic traits.

Environment corresponds to the trial (or to the block for IFCa, IFCb and IFCL only). Bold characters indicate that only highly correlated trials were included, asterisks (*) indicate that lesser-correlated trials were included as well. PVE, Proportion of phenotypic variance explained; trials, number of trials; corr, average of the 2-by-2 correlations across trials; h^2, broad-sense heritability; G × E, genotype-by-environment interaction; AL, axis length; BX, brix; FCa, fruit external color a; FCb, fruit external color b; FCL, fruit external color L; FF, fruit fasciation; FLe, fruit maximum length; FLo, fruit load; FlT, flowering time; FP, fruit pungency; FShI, fruit shape index; FShO, fruit predominant shape oblate; FWe, fruit weight; FWi, fruit maximum width; IFCa, immature fruit external color a; IFCb, immature fruit external color b; IFCG, immature fruit color green; IFCL, immature fruit external color L; LN, locule number; PHe, total plant height; PTh, pericarp thickness; TFN, total fruit number; TFWe, total fruit weight.

The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd.,
The Plant Journal, (2023), doi: 10.1111/tpj.16425

Table 2 Pearson correlation between single-environment BLUPs and multi-environment BLUPs

	Trial						
Trait code	ARO Israel April 2020–August 2020	WorldVeg Taiwan October 2020–January 2021	BATEM Türkiye August 2020–March 2021	CREA Italy June 2019–December 2019	INRAE 2019 France June–November	INRAE 2020 France June–November	Average
AL	0.68	NA	0.87	0.84	0.93	0.87	0.84
BX	0.66	NA	0.85	NA	0.88	0.87	0.82
FCa	0.82	NA	NA	0.93	0.96	0.96	0.92
FCb	0.90	NA	NA	0.94	0.93	0.92	0.92
FCL	0.89	NA	NA	0.93	0.93	0.91	0.92
FLe	0.97	0.97	0.96	0.96	0.95	0.98	0.97
FIT	0.77	0.87	0.88	NA	0.83	0.81	0.83
FShl	0.96	0.97	0.96	0.97	0.97	0.98	0.97
FWe	0.98	0.98	0.98	0.98	0.96	0.99	0.98
FWi	0.98	0.98	0.98	0.97	0.97	0.99	0.98
LN	0.75	NA	0.87	NA	0.89	0.90	0.85
PHe	0.73	0.84	0.85	0.94	0.94	0.89	0.87
PTh	0.93	NA	0.94	0.92	0.95	0.98	0.94
TFN	NA	0.95	0.84	NA	NA	NA	0.90
TFWe	NA	0.85	0.97	NA	NA	NA	0.91
FF	NA	NA	0.85	0.96	0.96	0.91	0.92
FLo	NA	NA	NA	NA	0.94	0.93	0.94
FP	NA	NA	0.86	0.95	0.96	0.92	0.92
FShO	0.77	NA	0.65	1.00	0.99	0.96	0.87
IFCG	0.87	NA	0.92	0.86	0.93	0.93	0.90
minimum	0.66	0.84	0.65	0.84	0.83	0.81	0.82
maximum	0.98	0.98	0.98	1.00	0.99	0.99	0.98
average	0.84	0.93	0.89	0.94	0.94	0.93	0.91

NA indicates that the trait was not phenotyped by the partner, or was removed for multi-environment BLUPs computation. All BLUPs are used to compute correlations (no phenotypic outlier removed). All correlations are significant ($P < 0.001$).

We identified 32 candidate genes potentially affecting fruit color. Regarding immature fruit color, a homolog of *ARABIDOPSIS PSEUDO RESPONSE REGULATOR2* (*APRR2*) underlies colocalizing QTLs *IFCa-P1.1*, *IFCb-P1.1* and *IFCG-P1.2*, and *CAPSANTHIN-CAPSORUBIN SYNTHASE* (*CCS*) underlies *IFCL-P6.1*. Regarding ripe fruit color, QTLs *FCa-P1.2*, *FCa-P3.1*, *FCa-P6.1*, and *FCa-P6.2* are underlain by two homologs of a *LICODIONE SYNTHASE* (*CYP93B1*), homologs of three MYB transcription factors (*MYB35*, *MYB59,* and *MYB98*), of two E3 ubiquitin-protein ligases (*UPL3* and a putative UPL3 LIN-like isoform X1), of a *STEROL 3-BETA-GLUCOSYLTRANSFERASE* (*UGT80B1*), of an F-box/kelch-repeat protein (*At3g18720*) and by a putative *FLAVONOL SYNTHASE/FLAVANONE 3-HYDROXYLASE* (*CA06g12570*).

We also found 15 candidate genes potentially affecting plant growth and earliness. Homologs of *GROWTH-REGULATING FACTOR 12* (*GRF12*) and *PLAT DOMAIN-CONTAINING PROTEIN 1* (*PLAT1*) underlie axis length QTLs *AL-P1.2* and *AL-P5.4*, respectively. A homolog of *NARROW LEAF 1* (*NAL1*) underlies the total plant height QTL *PHe-P2.1*. Regarding flowering time, *FIT-P9.2* is underlain by homologs of *SNW/SKI-INTERACTING PROTEIN* (*SKIP*) and of putative *GATA TRANSCRIPTION FACTOR 22* (*GATA22*), and *FIT-P9.2* by a homolog of Zinc finger protein *CONSTANS-LIKE 6* (*COL6*).

Finally, we identified 14 candidate genes potentially affecting plant productivity. Among these, a NAC transcription factor (*CA03g28100*) and a homolog of *AINTEGUMENTA* (*ANT*) underlie total fruit number QTLs *TFN-P12.2* and *TFN-P6.1*, respectively.

DISCUSSION

The G2P-SOL core collection is a valuable new resource for pepper genetics

Until recently, no global comprehensive survey of pepper phenotypic and genotypic diversity had been reported. Association panels and core collections mainly originated from rather limited numbers of accessions from one or few genebanks (Colonna et al., 2019; Lozada et al., 2022; Nimmakayala et al., 2014, 2016; Tamisier et al., 2020; Wu et al., 2019; Zewdie et al., 2004), and were designed based on the phenotypic diversity for the traits of interest (Hanson et al., 2004; Quenouille et al., 2016; Thies & Fery, 2002), and on limited genotypic data (Fan et al., 2004; Lee et al., 2016; Nicolaï et al., 2013). This limitation has now been lifted owing largely to the G2P-SOL project's survey of the global

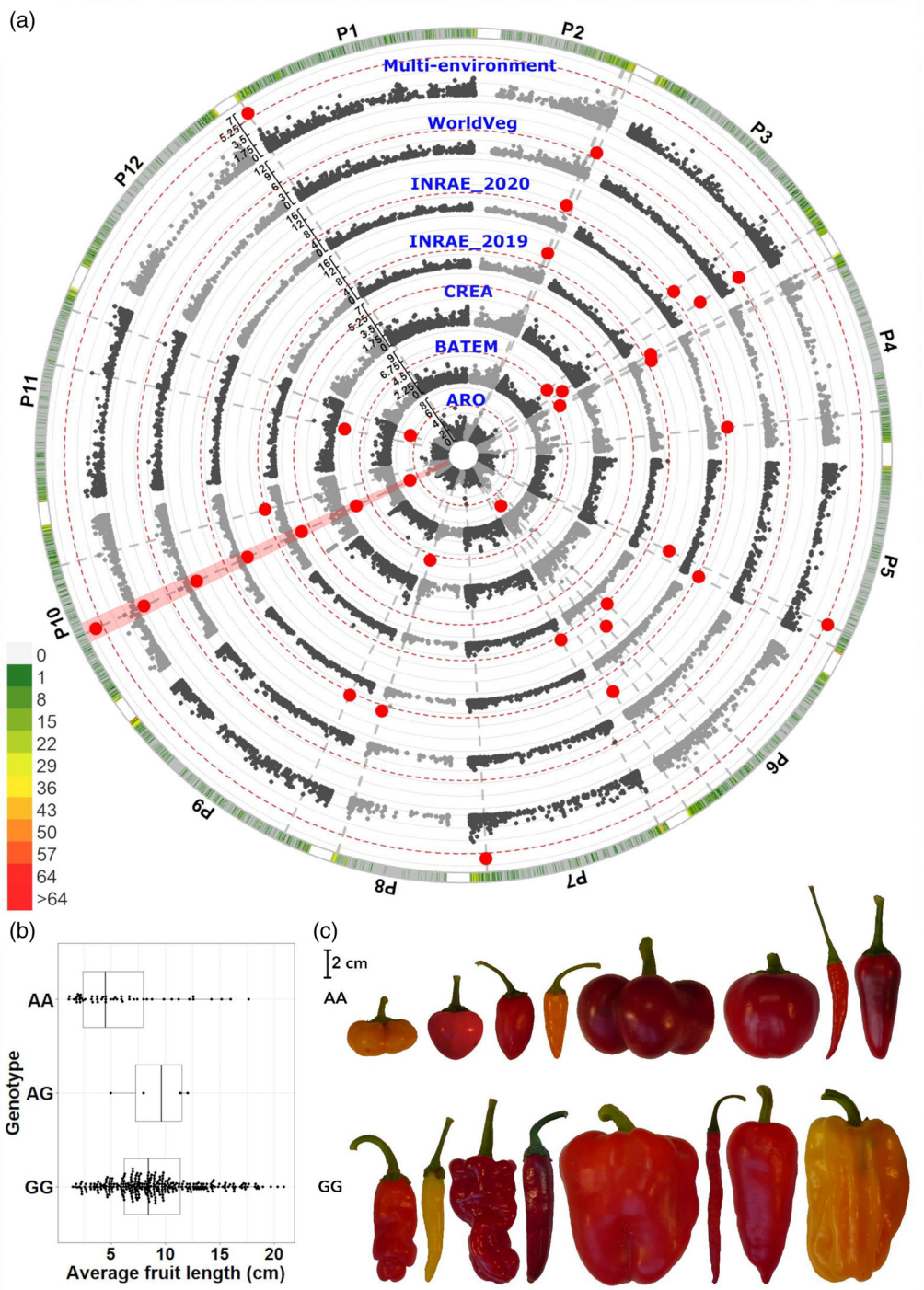

Figure 4. *FLe-P10*, a highly robust fruit length QTL on chromosome P10.
(a) Circular Manhattan-plot depicting the fruit maximum length QTNs detected with the multi-locus BLINK model. The seven internal circles show Manhattan plots for all single environments and the multi-environment, and the external circle displays the marker density (markers/megabase, according to the color scale). Red dashed circles represent the Bonferroni-adjusted significance threshold for each environment, and red dots the −log10(*P*-value) of significant marker-trait associations. Gray radial dashed lines indicate the projection of significant associations on all circles. QTL *FLe-P10* is highlighted in red on chromosome P10.
(b) Boxplots representing the fruit maximum length of the 350 accessions averaged over the six trials, according to their genotype at QTL *FLe-P10*. AA genotype: 49 accessions. AG genotype: 4 accessions. GG genotype: 283 accessions. AA and GG genotypes exhibited contrasting phenotypes (Student test, *P*-value = 5.632e-06).
(c) Pictures of eight accessions spanning the first to third quartiles of fruit maximum length for the AA and GG genotypes at QTL *FLe-P10*. QTL, quantitative trait loci; QTN, quantitative trait nucleotide.

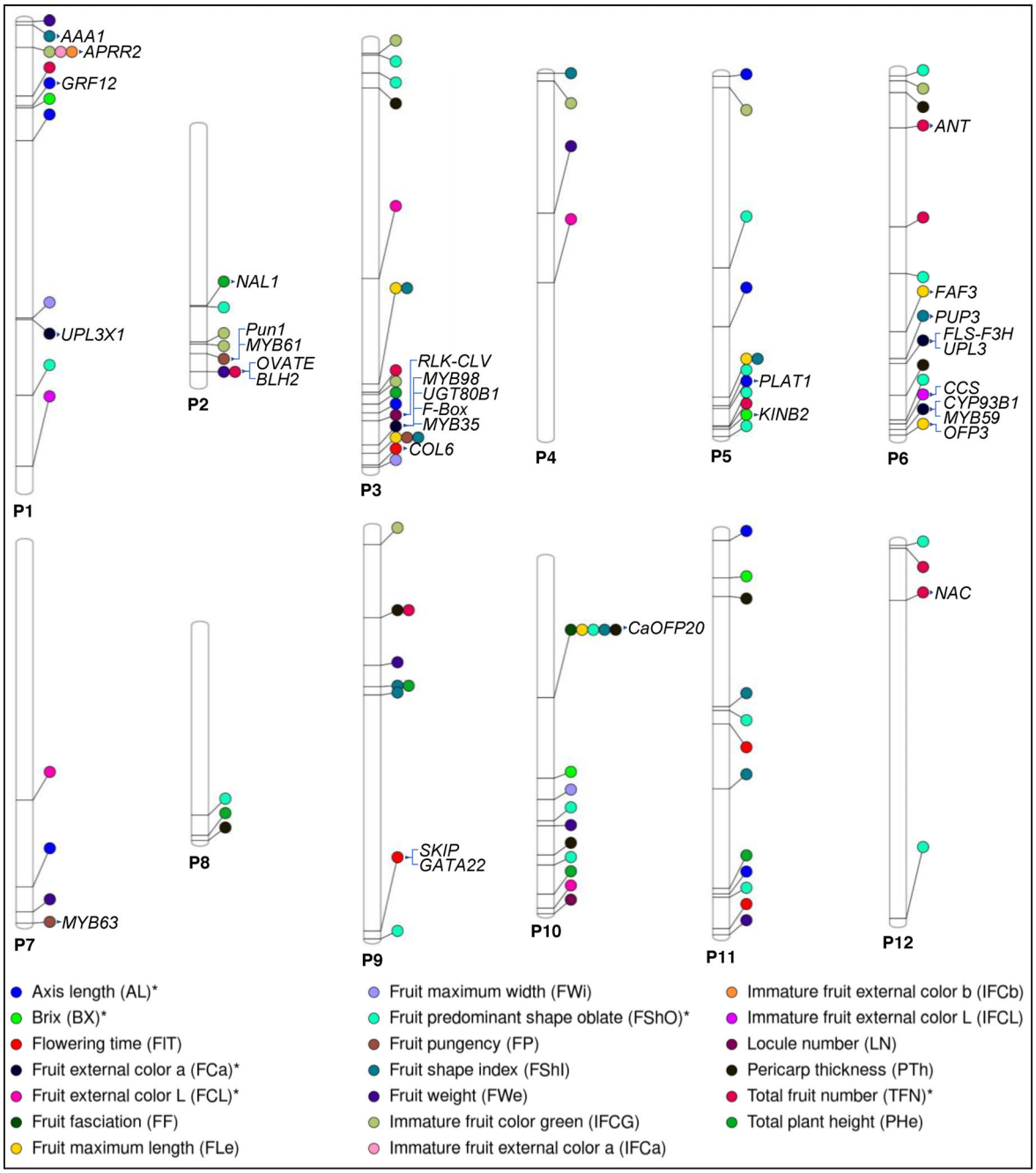

Figure 5. Physical position of the 114 highest-confidence QTLs and their best candidate genes mapped on the 12 pepper chromosomes.
The 31 candidate genes discussed in the manuscript are indicated next to the QTL positions. Candidate gene names are detailed in Table S9, along with the name and position of their corresponding QTLs. Asterisks (*) indicate which traits were never analyzed by GWAS before. *AAA1*, homolog of *KATANIN P60 ATPASE-CONTAINING SUBUNIT A1*; *ANT*, homolog of *AINTEGUMENTA*; *APRR2*, homolog of *ARABIDOPSIS PSEUDO RESPONSE REGULATOR2*; *BLH2*, homolog of *BEL1-LIKE HOMEODOMAIN PROTEIN 2*; *CaOFP20*, *C. annuum OVATE FAMILY PROTEIN 20*; *CCS*, *CAPSANTHIN-CAPSORUBIN SYNTHASE*; *COL6*, homolog of *CONSTANS-LIKE 6*; *CYP93B1*, homologs of a *LICODIONE SYNTHASE* (*CYP93B1*); *FAF3*, homolog of *FANTASTIC FOUR 3*; *F-Box*, homolog of the F-box/kelch-repeat protein *At3g18720*; *FLS-F3H*, homolog of Flavonol synthase and flavanone 3-hydroxylase; *GATA22*, homolog of *GATA TRANSCRIPTION FACTOR 22*; *GRF12*, homolog of *GROWTH-REGULATING FACTOR 12*; *KINB2*, homolog of *SNF1-RELATED PROTEIN KINASE REGULATORY SUBUNIT BETA-2*; *MYB35*, homolog of *MYB35* transcription factor; *MYB59*, homolog of *MYB59* transcription factor; *MYB61*, homolog of *MYB61* transcription factor; *MYB63*, homolog of *MYB63* transcription factor; *MYB98*, homolog of *MYB98* transcription factor; *NAC*, possible NAC transcription factor *CA03g28100*; *NAL1*, homolog of *NARROW LEAF 1*; *OFP3*, homolog of *OVATE FAMILY PROTEIN 3*; *OVATE*, homolog of *OVATE*; *PLAT1*, homolog of *PLAT DOMAIN PROTEIN 1*; *Pun1*, *Pun1* gene; *PUP3*, homolog of *PURINE PERMEASE 3*; *RLK-CLV*, homolog of leucine-rich repeat receptor-like protein kinase of the CLAVATA family *At5g63930*; *SKIP*, homolog of *SNW/SKI-INTERACTING PROTEIN*; *UGT80B1*, homolog of a *STEROL 3-BETA-GLUCOSYLTRANSFERASE* (*UGT80B1*); *UPL3*, homolog of *E3 UBIQUITIN-PROTEIN LIGASE 3*; *UPL3X1*, homolog of UPL3 LIN-like isoform X1.

pepper germplasm across 10 038 accessions from collections conserved in 10 major genebanks using 26 566 SNPs (Tripodi et al., 2021).

To fully exploit this unprecedented diversity survey, we constructed a core collection of 423 pepper accessions of rich genotypic and phenotypic variability. Representatives from the secondary and tertiary genepools were included in the core collection, *inter alia* for an exploratory screening of non-*annuum* species for specific traits such as responses to biotic and abiotic stresses. The core collection shows a species-wise genetic structure, and the stratification among its 350 *C. annuum* accessions used for GWAS is partly correlated to their geographical origin (Figure 2c,d), which is consistent with the previous results on different sets of *C. annuum* accessions from Nicolaï et al. (2013), Lee et al. (2016), Taranto et al. (2016) and Tripodi et al. (2021). Therefore, the inherent structure of the core collection needs to be addressed in future genotype–phenotype associations studies, as it was addressed here. This core collection paves the way for a broad range of studies. For instance, our suite of genomic trait association scans revealed candidate genes for relevant traits for pepper agriculture.

Multi-environment GWAS delivers robust QTLs

Various studies in pepper have shown that traits related to morpho-agronomic performance and phytochemical composition can be impacted by G × E interactions (Alimi et al., 2013; Cabral et al., 2017; Tripodi et al., 2019). To assess the influence of the environment and of G × E interactions, envirotyping of our core collection was performed with trials from five geographically and climatically contrasted locations, and for two consecutive years in one of them (Table S4). By addressing the environment and G × E interaction effects, we refined the estimations of the part of phenotypic variance explained by the genotype and of the heritability of traits in various environments (Malosetti et al., 2013). The use of multiple environments also differentiated environment-specific QTLs from more robust QTLs. The ANOVA models explaining the phenotypic variance fitted the data satisfactorily, and heritability values of the traits were high despite the G × E interaction effects (Figure 3), suggesting that the QTLs detected could be used efficiently in pepper breeding through marker-assisted selection, particularly for fruit morphology (FWe, FLe, FWi, FShl, and PTh), color (FCL, FCa, and FCb) and pungency (FP) traits. Eight robust QTLs were located in the vicinity of the already reported pepper genes *Pun1*, *CCS*, *OFP20*, and *OVATE*. In addition, we found that two robust QTLs may correspond to already reported QTLs from GWAS (Table S1). QTL *FW-P10.1* on chromosome P10 (interval 175 298 168–175 898 168 bp) was in the vicinity of a QTL determining fruit weight at 173 645 289 bp (Lozada et al., 2022). QTL *PHe-P2.1* on chromosome P2 (interval 116 761 596–117 561 596 bp) was located between two QTLs at 111 478 856 bp and 124 942 237 bp (Lozada et al., 2022). This underlines that QTLs detected by GWAS depend not only on the environment, but also on other factors, such as the phenotyping methodology, the SNP coverage over the whole genome, and the nature of the phenotyped core collection influencing the proportion of rare alleles and the extent of LD decay.

Altogether, phenotyping a representation of the global *C. annuum* diversity for a large number of traits in multiple environments allowed an unprecedented in-depth insight into the genetic architecture of the major agronomical traits of cultivated pepper. The observed correlations between two main antagonistic clusters of traits suggest how these traits, especially fruit morphology and fruit flavor, have been predominant in the breeding history of pepper which led to the development of divergent types. The colocalization on a few chromosomal regions of *a priori* unrelated traits (Table S8; Figure 5) could stem from pleiotropic genes or from several linked QTLs, which should be determined by further fine mapping. As an example, the intervals of the *FShl-P3.4* and *FP-P3.3* QTLs, whose most strongly associated markers are 17 kb apart, overlapped by 783 kb (Figure S8a,b). It is therefore not possible to conclude whether these two QTLs are determined by a single pleiotropic gene or by two physically linked genes. We observed that the GG haplotype, combining the alleles for narrow and elongated fruits (high FShl) and for low pungency (low FP), was absent from the set of 350 *C. annuum* (Figure S8c,d). The presence of the other three haplotypes indicates recombination between these two markers and possibly between the two associated causal genes, while the absence of the fourth haplotype may arise from a negative effect on another trait, and therefore its counter-selection. However, FShl and FP are not completely dependent, as the core collection includes a few accessions with long, thin and sweet fruits, indicating the influence of other QTLs on these traits. The possible physiological antagonism of the clusters of traits should be taken into consideration for breeding.

Candidate genes were identified for 53 of the highest-confidence QTLs

We defined a set of candidate genes within the most promising QTL regions (Table S9; Figure 5), proven or suspected to be involved in molecular pathways determining the trait, or governing the same or similar traits in pepper or other species (Pflieger et al., 2001). QTL intervals were defined according to chromosome-averaged LD decay assessments, although it is known that there are some local LD decay variations along chromosomes, which mostly show suppression of recombination in pericentromeric regions and high recombination rates in the distal chromosome regions (Fuentes et al., 2022). Thus, some QTL intervals

could be slightly misestimated. Here, four reported well-known candidate genes (*Pun1*, *CCS*, *OFP20*, and *OVATE*) are somewhat outside the calculated QTL interval (Table S9). Additionally, the lack of markers in some regions reduces the chances to find associations near a known gene. In some cases, this may also be due to the difference in type of studied population: alleles with large effects that are segregating in a biparental population may be rare in the germplasm, hence difficultly identified by GWAS (Korte & Farlow, 2013). To conclude, the candidate genes may be valuable targets for marker-assisted selection.

Candidate genes for fruit flavor

Among the eight candidate genes for fruit pungency and Brix, we highlighted four strong candidates for three QTLs. The well-known *Pun1* gene, corresponding to the locus *C* (Deshpande, 1935), determines pungency qualitatively through the presence or absence of capsaicinoids (Blum et al., 2002; Stewart et al., 2005) and is linked to our strongest marker-pungency association. This putative acyltransferase is indeed less than 50 kb apart from the QTL *FP-P2.1* on chromosome P2. Furthermore, various members of the MYB transcription factor family, and particularly genes of the R2R3-MYB subfamily, were demonstrated to play key regulatory roles in capsaicinoid biosynthetic pathway (Arce-Rodríguez & Ochoa-Alejo, 2017; Han et al., 2019; Ju et al., 2020; Sun et al., 2019, 2020; Wang, Liu, et al., 2020). In our study, homologs of R2R3-MYB family genes *MYB61* and *MYB63*, underlying QTLs *FP-P2.1* and *FP-P7.3*, respectively, could therefore control pungency.

A homolog of the gene encoding the *SNF1-RELATED PROTEIN KINASE REGULATORY SUBUNIT BETA-2* (*KINB2*) underlies *BX-P5.2* and may control Brix. Indeed, the Snf1-related protein kinase (SnRK) complex plays a central role in the regulation of carbohydrate metabolism to deal with sugar limitation during stress conditions in Arabidopsis (Guérinier et al., 2013), and the inactivation of the SnRK complex subunit SIP1 leads to reduced levels of sugars (Anderson & Kohorn, 2001).

Candidate genes for fruit size and shape

Regarding the 28 candidate genes probably affecting fruit size and shape, we highlighted eight strong candidates for 11 QTLs. Ovate Family Proteins (OFP) are transcriptional repressors controlling multiple aspects of plant growth and development in multiple species including pepper (Wang et al., 2016). *OVATE* was the first gene identified as responsible for a tomato fruit shape QTL (Liu et al., 2002), and *CaOvate*'s down-regulation in pepper is likely a promoter of fruit elongation (Tsaballa et al., 2011). The *OVATE* homolog positioned at 155 Mb on chromosome P2, is closely linked to the fruit weight QTL *fw2.1* (Zygier et al., 2005; Table S1) and close to the QTL *FWe-P2.4*, detected with the strongest marker-FWe association at position 160 Mb. Moreover, *C. annuum OVATE FAMILY PROTEIN 20* (*CaOFP20*), known to control fruit elongation through cell expansion and replication (Borovsky et al., 2022), is 1.2 Mb distant from our strongest associations for FLe, FShl, PTh, and FF (*FLe-P10.1*, *FShl-P10.1*, *PTh-P10.1*, and *FF-P10.1*) and one of our strongest associations for FShO (*FShO-P10.1*), which are all colocalized on chromosome P10. In peach, a chromosomal inversion downstream of an OFP gene is responsible for the flat fruit shape, comparable to the oblate pepper phenotype (Zhou et al., 2021). A homolog of the gene encoding the transcription repressor Ovate Family Protein 3 (*OFP3*) underlying the QTL *FLe-P6.4* could also determine fruit maximum length (FLe).

Transcription factors from the BEL1-LIKE HOMEODOMAIN (BLH) family were found to interact with OFPs in Arabidopsis, cotton, rice, and banana (Hackbusch et al., 2005; Li et al., 2011; Liu & Douglas, 2015; Pagnussat et al., 2007; Schmitz et al., 2015; Wang et al., 2016; Zhang et al., 2016). Moreover, Bernard et al. (2021) found that a BLH protein was likely associated with walnut kernel weight. A homolog of *BEL1-LIKE HOMEODOMAIN PROTEIN 2* (*BLH2*) underlying the QTL *FWe-P2.4*, detected with the strongest marker-FWe association, may thus affect fruit weight.

In tomato, *CELL SIZE REGULATOR* (*CSR*) encodes a FAF-like protein controlling fruit weight (Mu et al., 2017). In pepper, a FAF protein interacting with protein WUSCHEL might regulate the size of shoot meristems as in Arabidopsis (Wahl et al., 2010), as well as fruit weight (Nimmakayala et al., 2016). A homolog of *FANTASTIC FOUR 3* (*FAF3*) underlying the QTL *FLe-P6.2* may thus determine FLe.

Two members of the purine permeases (PUPs) transporter family independently induce a remarkable increase in rice grain size and length, probably by cytokinin transport and allocation into vascular bundle cells (Qi & Xiong, 2013; Xiao et al., 2019). The KATANIN microtubule-severing enzymes are AAA ATPases also involved in the progression of cell division, or cytokinesis, and cell division plane orientation (Luptovčiak et al., 2017). Wang et al. (2021) showed that a cucumber homolog of *KATANIN P60 ATPASE-CONTAINING SUBUNIT A1* (*CsKTN1*) is involved in fruit elongation. Therefore, homologs of *PURINE PERMEASE 3* (*PUP3*) and *KATANIN P60 ATPASE-CONTAINING SUBUNIT A1* (*AAA1*) may govern fruit shape index, as they underlie QTLs *FShl-P6.3* and *FShl-P1.1*, respectively.

In tomato, the *lc* locus encoding WUSCHEL and the *fas* locus encoding the receptor-like kinase CLAVATA3 (CLV3) synergistically control the locule number (Chu et al., 2019; Muños et al., 2011). The cucumber homolog of *CLV3* also impacts carpel number (Che et al., 2020) and *CLAVATA1* (*CLV1*) may impact fruit size in *C. chinense* (Nimmakayala et al., 2021). A homolog of the probable leucine-rich repeat receptor-like protein kinase of the

CLAVATA family *At5g63930*, underlying the QTL *LN-P3.1*, may thus determine locule number.

Candidate genes for fruit color

Among the 16 candidate genes for immature fruit color, we highlighted two strong candidate genes for four QTLs. Homologs of transcription factor *ARABIDOPSIS PSEUDO RESPONSE REGULATOR2* (*APRR2*) determine mature or immature fruit color in numerous cucurbits and solanaceous crops including melon (Oren et al., 2019), watermelon (Oren et al., 2019), cucumber (Liu et al., 2016; Tang et al., 2018), zucchini (Montero-Pau et al., 2017), wax gourd (Ma et al., 2021), tomato (Pan et al., 2013), eggplant (Arrones et al., 2022; Fang et al., 2023), as well as stem color in squash (Zhu et al., 2022). The pepper homolog of *APRR2* on chromosome P1 corresponds to the well-known mature fruit color locus *c1* (Kormos & Kormos, 1960), which was also shown to regulate chlorophyll content and immature fruit color (Jeong et al., 2020; Lee, Kim, et al., 2020; Pan et al., 2013) via regulating thylakoid numbers and chloroplast size (Lee, Kim, et al., 2020). Nonsense mutations resulting in premature stop codons were reported for this gene in all these crops, except zucchini (Arrones et al., 2022; Fang et al., 2023; Jeong et al., 2020; Lee, Kim, et al., 2020; Liu et al., 2016; Ma et al., 2021; Oren et al., 2019; Pan et al., 2013; Song et al., 2022; Tang et al., 2018; Zhu et al., 2022). Therefore, the *APRR2* homolog nonsense mutation identified in our findings is a strong candidate for the colocalizing QTLs *IFCG-P1.2*, *IFCa-P1.1*, and *IFCb-P1.1*, corresponding to the strongest marker associations with three immature external fruit color traits.

The *CAPSANTHIN-CAPSORUBIN SYNTHASE* gene (*CCS*), involved in the synthesis of red carotenoid pigments in the mature fruit, corresponds to the *y* locus determining red versus yellow ripe fruit color in pepper (Lefebvre et al., 1998; Li et al., 2013; Popovsky & Paran, 2000). This gene is positioned less than 300 kb apart from our strongest marker association for immature fruit external color L (*IFCL-P6.1*), suggesting that it might also determine the intensity of the immature fruit color.

Within the 16 candidate genes pertaining to ripe fruit color, we highlighted 10 strong candidates for four QTLs, potentially related to pepper pigments. Flavonol synthase and flavanone 3-hydroxylase are involved in flavonoid synthesis (Preuß et al., 2009; Winkel-Shirley, 2001), while a sterol glucosyltransferase (*UGT80B1*) is required for flavonoid accumulation in the Arabidopsis seed coat (DeBolt et al., 2009). Their respective pepper homologs are thus plausible candidates for QTLs *FCa-P6.1* and *FCa-P3.1*, the latter being detected by the strongest marker-FCa association. Moreover, two homologs of a *LICODIONE SYNTHASE* (*CYP93B1*) underlie *FCa-P6.2* and may govern fruit color, as Mizuno et al. (2016) suggested that the encoded enzyme was related to the synthesis of flavones and associated with dark or light brown coloration in tan-colored injured leaves of sorghum.

Transcription factors of the MYB family regulate the expression of structural genes of the anthocyanin biosynthetic pathway in plants (Yan et al., 2021) and more specifically in *C. annuum* (Stommel et al., 2009), leading to changes in fruit coloration through anthocyanin accumulation (Borovsky et al., 2004). Moreover, *CaMYB306* expression was higher in red fruit and its homolog's activity in tomato increases carotenoid accumulation and decreases anthocyanin and chlorophyll content (Ma et al., 2022). The two QTLs *FCa-P3.1* (detected by the strongest marker-FCa association) and *FCa-P6.2* are underlain by homologs of the three genes *MYB35*, *MYB59* and *MYB98* which may determine FCa. Additionally, the ubiquitin E3 ligase *CONSTITUTIVE PHOTOMORPHOGENIC1* (*MdCOP1*) modulates the degradation of a MYB transcription factor in order to regulate light-induced anthocyanin biosynthesis in apple and probably in other species, as well as the red coloration of apple fruit (Li et al., 2012). Homologs of *E3 UBIQUITIN-PROTEIN LIGASE 3* (*UPL3*) and a putative UPL3 LIN-like isoform X1 respectively underlie QTLs *FCa-P6.1* and *FCa-P1.2*, and may thus determine FCa.

Lastly, *FLAVIN-BINDING*, *KELCH REPEAT*, *F-BOX 1* (*SlFKF1*) may impact lycopene concentrations in tomato (Shibuya et al., 2021), suggesting that the pepper homolog of the F-box/kelch-repeat protein *At3g18720* underlying *FCa-P3.1* could impact FCa.

Candidate genes for plant vigor and earliness

Among the seven plant vigor candidate genes, we highlighted three strong candidates for three QTLs. The *PLAT DOMAIN PROTEIN 1* (*PLAT1*), a potential downstream target of the abscisic acid signaling pathway, is involved in plant growth in Arabidopsis (Hyun et al., 2014), while members of the growth-regulating factor (GRF) family are involved in rice stem elongation (Cao et al., 2017; van der Knaap et al., 2000). Homologs of *PLAT1* and *GRF12* may therefore govern axis length as they underlie QTLs *AL-P5.4* and *AL-P1.2*, respectively. Moreover, the total plant height QTL *PHe-P2.1* is underlain by a homolog of *NARROW LEAF 1* (*NAL1*) which encodes a putative trypsin-like serine and cysteine protease affecting plant height in rice (Qi et al., 2008; Subudhi et al., 2020).

Regarding the eight flowering time candidate genes, we highlighted three strong candidates for two QTLs. *CONSTANS* and *CONSTANS*-like genes are crucial mediators of the photoperiodic induction of flowering in numerous plants including Arabidopsis, aspen trees, rice, Ipomoea, chrysanthemum and tomato (Böhlenius et al., 2006; Fu et al., 2015; Liu et al., 2001; Suárez-López et al., 2001; Yang et al., 2020; Yano et al., 2000). Their function may thus be conserved across angiosperms, making the homolog of Zinc finger protein *CONSTANS-LIKE 6* (*COL6*) a strong

candidate for the QTL *FIT-P3.3*. Moreover, *SNW/SKI-INTERACTING PROTEIN* (*SKIP*) affects both circadian clock and flowering time in Arabidopsis through alternative splicing (Cao et al., 2015; Cui et al., 2017; Wang et al., 2012), and two paralogous genes encoding GATA-type transcription factors repress flowering time, probably by repressing gibberellin and auxin signaling (Hudson et al., 2011; Richter et al., 2010, 2013). Homologs of *SKIP* and of putative *GATA22* underlie *FIT-P9.2* and could thus affect FIT.

Candidate genes for plant productivity

Among the 14 candidate genes for total fruit number, we highlighted two strong candidates for two QTLs. The QTL *TFN-P6.1*, corresponding to the strongest marker association with TFN, is underlain by a homolog of *AINTEGUMENNTA* (*ANT*). In Arabidopsis, this gene and its homolog *AINTEGUMENTA-LIKE 6* (*AIL6*) have well-established importance in flower organs initiation and development, and affect flower number (Elliott et al., 1996; Krizek, 2009; Krizek et al., 2020; Long & Barton, 2000). A possible NAC transcription factor (*CA03g28100*) underlying *TFN-P12.2* may also affect TFN, given that different transcription factors from the NAC gene family influence fruit number in tomato (Lira et al., 2017; Ma et al., 2018; Wang, Zheng, et al., 2020).

EXPERIMENTAL PROCEDURES

Plant material

A core collection of *Capsicum* spp. was constructed using the GBS dataset described in Tripodi et al. (2021), corresponding to 10 038 accessions originated from 10 genebanks (Ozalp et al., 2020; Salinier et al., 2022). A Euclidean distance matrix was computed using the function dist from the stats R package (R Core Team, 2021). A core collection of 449 accessions was created with the coreCollection (v0.9.5) package (https://cran.r-project.org/package=coreCollection), using the Accession-to-Nearest-Entry "A-NE" criterion as introduced by Jansen and van Hintum (2007), the "split" adjusted group method, the "randomDescent" algorithm, and default values for the remaining parameters, with a preselected set of 26 wild and 84 cultivated accessions used as controls by the G2P-SOL project partners. For phytosanitary reasons, 26 of the 449 accessions could not be sent by the genebanks, reducing the G2P-SOL core collection to 423 *Capsicum* spp. accessions. From this core collection, 350 *C. annuum* accessions with less than 5% heterozygosity and less than 30% of missing data were retained for GWAS purposes. Accessions with a higher heterozygosity level were probably mixed genotypes since they were often scored as phenotypically heterogeneous in more than one trial.

Genotyping

DNA was extracted from leaf tissue samples of a single plant of each accession of the core collection only, and a second genotyping experiment was performed following the same *Msp*I/*Pst*I GBS protocol as reported in Tripodi et al. (2021) for the primary genotyping. Reads were aligned with *C. annuum* cv. CM334 reference genome version 1.6 (Kim et al., 2017), using the Burrows-Wheeler Aligner (BWA) program (v0.7.17-r1188) and the 'mem' command with default parameters (Li, 2013). BAM files were processed and used for the SNP calling using GATK (v4.1.4.1; DePristo et al., 2011). SNPs were filtered using VariantFiltration, with QualByDepth, FisherStrand, RMSMappingQuality, MappingQualityRankSumTest, and ReadPosRankSumTest arguments (QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0). The markers that were polymorphic within the 350 *C. annuum* accessions used for GWAS were combined from both genotyping experiments. A final dataset of 10 195 markers was obtained after filtering with Bcftools (v1.15; Li, 2011) to retain bi-allelic SNPs with a MAF greater than 0.01, a mean read depth greater than 8, less than 30% of missing data and heterozygous for less than 20% of the accessions. Finally, accessions showing a heterozygosity higher than 2% were removed.

Field trials

Six trials (independent experiments from different years and/or locations) were organized. Under a Mediterranean climate, the National Research Institute for Agriculture, Food and Environment (INRAE) conducted two trials in France in 2019 and 2020, the Council for Agricultural Research and Economics (CREA) conducted one trial in Italy in 2019, the Agricultural Research Organization (ARO) conducted one trial in Israel in 2020, and Batı Akdeniz Agricultural Research Institute (BATEM) conducted one trial in Türkiye over 2020 and 2021. The World Vegetable Center (WorldVeg) conducted one trial in Taiwan over 2020 and 2021, under a subtropical climate. Depending on the trial, the plants were transplanted 5–7 weeks after sowing, in a randomized complete block design with two to three blocks, and two to four plants per accession per block. Metadata for each trial and each trait are listed in Tables S4 and S5, respectively.

Phenotyping

A total of 23 agronomic traits mainly based on pepper descriptors described by the International Plant Genetic Resources Institute (IPGRI) were assessed (Table 1; Table S5). Fifteen of the 18 quantitative traits had a continuous measurement (axis length, AL; Brix, BX; fruit external color, FCL, FCa, and FCb; fruit maximum length, FLe; fruit shape index, FShI; fruit weight, FWe; fruit maximum width, FWi; immature fruit external color, IFCL, IFCa, and IFCb; total plant height, PHe; pericarp thickness, PTh; total fruit weight, TFWe) and three had discrete values (flowering time, FIT; locule number, LN; total fruit number, TFN). Three of the five qualitative traits were ordinal traits visually assessed following categorical values in a range (fruit fasciation, FF; fruit load, FLo; external immature fruit color green, IFCG), as opposed to the two binary traits assessed as presence/absence (fruit pungency, FP; fruit predominant shape oblate, FShO).

Data analysis

Statistical analyses using R software were performed with version 4.1.2 (R Core Team, 2021).

Population structure

To investigate the population structure, the SNP panel (without Indels) was pruned based on LD using plink2 (Purcell & Chang, 2019) with the command --indep-pairwise 50 10 0.2. PCA was performed using the SNPRelate R package (v3.13; Zheng et al., 2012) with default parameters. In addition, the population structure was analyzed by ADMIXTURE (v1.3.0; Alexander et al., 2009) with the cluster number K ranging from 1 to 15, using a CV

value of 15 (--cv = 15). Each *K* was run with 20 replicates and the outputs were analyzed using pong (v1.5; Behr et al., 2016) and plotted with pophelper (v2.3.1; Francis, 2017). CV error was used to determine a suitable value for *K*. Individuals were assigned to one of the *K* populations when their membership coefficient in that group was ≥0.70.

LD decay study

LD was estimated as pairwise squared allele-frequency correlation (r^2) between SNP markers, correcting for relatedness using a kinship matrix by the R package LDcorSV (v1.3.3; Mangin et al., 2012). The kinship matrix was estimated as covariance between accessions. The overall LD decay estimated per chromosome was likely not significantly affected by non-recombinant heterochromatic regions, since gene-rich euchromatic regions were preferentially sequenced because of the use of methylation-sensitive *Msp*I/*Pst*I restriction enzymes for the GBS protocol. The r^2 values were binned according to the physical distances (Mb) between SNP pairs, and the 95th percentile of each bin was fitted by a polynomial curve. The extent of LD decay was estimated as the distance (Mb) where LD decayed below the arbitrarily chosen baseline of 0.2.

Phenotypic data processing

The proportions of explained variance were calculated following a fixed-effects linear model: $Y_{ijk} = \mu + G_i + E_j + G_i \times E_j + \varepsilon_{ijk}$, where Y_{ijk} represents the phenotypic value of the plant *k* from the genotype *i* (G_i) in the environment *j* (E_j), $G_i \times E_j$ the interaction between genotype and environment, and ε_{ijk} the residual error. The model was fitted using the function lm from the stats R package (R Core Team, 2021). The proportions of explained variance were calculated as the ratio of the sum of squares of the genetic, environmental, or G × E effect considered to the total sum of squares.

The broad-sense heritability was computed following the approach proposed by Cullis et al. (2006) implemented in the inti R package (v0.5.5; Lozano-Isla, 2022). It uses genotypic BLUPs calculation and is particularly suitable for unbalanced datasets (Piepho & Möhring, 2007). First, a random-effects linear model was fitted: $Y_{ijk} = \mu + G_i + E_j + G_i \times E_j + \varepsilon_{ijk}$, where Y_{ijk} represents the phenotype value of the plant *k* from the genotype *i* (G_i) in the environment *j* (E_j), $G_i \times E_j$ the interaction between genotype and environment and ε_{ijk} the residual error. Then, heritability was calculated as $h^2 = 1 - \overline{v}_{\Delta}^{\mathrm{BLUP}} / \left(2 \times \sigma_g^2\right)$, where $\overline{v}_{\Delta}^{\mathrm{BLUP}}$ corresponds to the mean variance of a difference of two BLUPs for the genotypic effect and σ_g^2 to the genotypic variance. Environment *j* (E_j) represents the trial *j* in the models used for computing the proportion of explained variance and the heritability, except for external immature fruit color parameters measured at BATEM only (for IFCL, IFCa and IFCb, E_j represents the block *j*).

Quantitative traits were screened for the presence of outliers within each trial using a median absolute deviation (MAD)-based Bonferroni-Holm test (Bernal-Vasquez et al., 2016) with $\alpha = 0.01$, applied on the MAD-standardized residuals of a linear model correcting for genotype, block and genotype-block interaction effects (Table S5). The model was fitted using the function lm from the stats R package (R Core Team, 2021). Locule number was treated as an ordinal trait. When the degrees of freedom were insufficient due to a lower number of measurements, the correction for interaction effect was not applied. When the block information was not available for a particular trial, only the correction for genotype effect was applied. Simple descriptive statistics were computed for accession means of each trait after outlier removal, with the stat.desc function from the pastecs R package (v1.3.21; Grosjean et al., 2018). Between-trial Pearson correlations were computed for each trait on the accession means per environment, to detect less correlated trials. The Box-Cox family of power transformations was then applied to the most correlated trials together to optimize normality of residuals for each trait, after fitting a model including all the main effects for genotype, trial and block (and trial-block interaction when the number of measurements was sufficient) (Table S5). Prior to Box-Cox transformation, traits containing negative values were transformed into only-positive ones adding an appropriate constant. The less correlated trials were transformed separately. This procedure was performed using the functions powerTransform and bcPower from the R package car (v3.0.13; Fox & Weisberg, 2019). Adjusted means per accession to be used as phenotypic input for GWAS were estimated using a BLUP model with the 'lme4' R package (Bates et al., 2015). These BLUPs were calculated both for single environments (correcting for block and genotype-block interaction effects when the number of measurements was sufficient) and across highly correlated environments (correcting for block nested in the trial effects and for genotype-block interaction effects when the number of measurements was sufficient). The lesser correlated trials were analyzed through single environment GWAS only, not to bias the multi-environment GWAS conducted on the higher correlated trials. Accessions with extreme adjusted means that were more than three MADs apart from the median were then removed, as they could bias the estimation of marker effects (Table S5).

Ordinal traits and locule number were visually screened for outliers, and between-trial Spearman correlations were computed on the accession means per environment, in order to detect less correlated trials. The Box-Cox family of power transformations was applied and adjusted means per accession were calculated as for the quantitative traits. Single-environment BLUPs were not computed for the less correlated trials. The lower correlations for these ordinal and more subjective traits may have arisen from divergent appreciations of the different trial assessors. Accessions with extreme adjusted means were removed by visual inspection, using the same lower and upper bounds for the single-environment BLUPs and the multi-environment BLUPs (Table S5).

The fruit pungency (FP) and fruit predominant shape oblate (FShO) binary traits were coded as 0 versus 1. Between-trial Spearman correlations were computed to detect less correlated trials. Adjusted means per accession were then calculated as for the other traits.

Genome-wide association study

GWAS were conducted on both single-environment and multi-environment BLUPs, using the R package "Genomic Association and Prediction Integrated Tool" (GAPIT Version 3; Wang & Zhang, 2021). The mean dosage value per marker across accessions was used for imputing the remaining missing values (5% of the genotypic matrix). Three GWAS models were used: MLM (Yu et al., 2006), MLMM (Segura et al., 2012), and BLINK (Huang et al., 2019). The first four PCs of a PCA of marker genotypes were added to all the GWAS models as covariate variables, to reduce the number of false positives due to population stratification. In addition, MLM and MLMM included a random term to correct for relatedness according to a kinship matrix based on the VanRaden formula, calculated by GAPIT3. To control family-wise error rate, a significance threshold of 5.31 was calculated using the Bonferroni procedure (i.e., dividing the desired significance of 0.05 by the number of tested markers 10 195). Circular Manhattan plots were generated using the R package CMplot (v4.2; Yin et al., 2021).

QTLs and candidate genes

QTL intervals were calculated in order to group together highly linked QTNs (i.e., the SNP and Indel markers significantly associated to a trait) identified using the three different GWAS models (MLM, MLMM, and BLINK) on up to six environments or across all environments. Intervals were defined as the first and last QTN position ± the extent of specific chromosomal LD decay. Each QTL name was defined as the associated trait code, followed by the chromosome number and the QTL's order among all QTLs for this trait along the chromosome. Some of the QTLs were categorized as more promising, when they were robust (i.e., significant in at least two environments), or detected with at least two GAPIT models, or with a −log10(*P*-value) higher than five above the threshold (i.e., higher than 10.31). Since quantitative external immature fruit color traits were measured in one trial only, their QTLs with a −log10(*P*-value) higher than 2 above the threshold (i.e., higher than 7.31) were also included. Annotations for genes within the most promising QTL bounds were retrieved from *C. annuum* cv. CM334 reference genome version 1.55 (Kim et al., 2014) and version 1.6 (Kim et al., 2017). Their function in plants was then investigated in the literature, and genes involved in traits linked to the pepper associated trait were therefore proposed as putative causal genes.

Outlook

We constructed the G2P-SOL core collection, a unique and valuable resource for the pepper community. Its great genotypic and phenotypic variability, representative of the worldwide diversity, will accelerate the comprehensive understanding of the genetic basis of multiple traits. The core collection will be available on request, along with the genotypic and phenotypic data collected as part of the G2P-SOL project. As research resources, we also provided comprehensive up-to-date summaries of previously detected QTLs for agronomic traits in pepper and previously published GWAS in *C. annuum*. We conducted multi-environment GWAS on a great variety of plant and fruit traits, and GWAS for biotic stress resistance to multiple pathogens, heat stress tolerance and metabolic compounds is underway. We identified 97 new or already known genes potentially underlying 53 of the 119 highest-confidence QTLs. Further validation of specific genes will be needed to confirm their incidence on the phenotype. Nevertheless, these results demonstrate the usefulness and reliability of our core collection, phenotypic dataset and methodology. Traditionally, candidate genes identified for traits with simple inheritance can be used as diagnostic markers or as part of trait-specific marker-assisted selection. In addition, the genomic and phenotypic matrices provided by this study should improve multivariate genomic prediction at the whole-genome level, and should speed up the identification of candidate accessions that are phenotypically closest to a predefined genetic ideotype, adapted to a targeted environment or to several environments (Yang et al., 2022). The gene pool formed by these candidate lines, whose genetic backgrounds may be different, would greatly influence the simultaneous pre-breeding of multiple quantitative traits, and contribute to the development of genetically improved climate-resilient pepper crops.

AUTHOR CONTRIBUTIONS

VL, IP, PT, LB, and GG conceived the study. LM, VL, JS, CG, PT, HFB, RO, IP, YB, RS, and DB performed the agronomic trials and phenotyping. NS and MTR-W coordinated the genotyping-by-sequencing. LM, VL, LB, GT, MB, RF, and PT realized the bioinformatic and statistical analyses. LM, VL, LB, IP, PT, GT, and RV wrote the manuscript. All authors approved the manuscript.

ACKNOWLEDGMENTS

Authors gratefully acknowledge ARO, BATEM, CREA, INRAE, WorldVeg, IPK (Andreas Boerner), MVCRI (Maritsa Vegetable Crops Research Institute, Gancho Pasev), UPV (Universitat Politècnica de València, Maria-Jose Diez), CGN (Centre for Genetic Resources, Willem van Dooijeweert) for providing plant seeds to build the core collection. In this matter, the World Vegetable Center obtained additional support from their long-term strategic donors Taiwan, UK aid from the UK government, United States Agency for International Development (USAID), Australian Centre for International Agricultural Research (ACIAR), Germany, Thailand, Philippines, Korea and Japan. The G2P-SOL pepper core collection is now maintained and distributed by the INRAE GAFL's genebank, CRB-Lég. We are grateful to the members of the CRB-Lég and ReDD teams of INRAE GAFL unit and of the INRAE Mediterranean Fruit and Vegetables Experimental Facility (A2M, https://doi.org/10.17180/nxpc-kv48) for their technical assistance for the trials. This work was funded by the G2P-SOL project (Linking genetic resources, genomes, and phenotypes of Solanaceous crops) financed by the European Union's Horizon 2020 Research and Innovation Program under the grant agreement number 677379, and by the PRO-GRACE project (Promoting a Plant Genetic Resource Community for Europe) financed by the Horizon Europe Program under the grant agreement number 101094738. LM received a 3-year fellowship from the project G2P-SOL and from INRAE Plant Biology and Breeding Division.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The raw phenotypic and VCF genotypic datasets generated for this study are accessible at https://doi.org/10.57745/S7JVEM. Seeds from the G2P-SOL core collection will be available upon reasonable request at INRAE GAFL CRB-Lég, considering the sharing of costs for their production and their shipping (e-mail veronique.lefebvre@inrae.fr).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Genotypic dataset characteristics.

Figure S2. Average linkage disequilibrium (LD) decay plots for the 12 chromosomes.

Figure S3. A cluster number *K* of 11 derived from an ADMIXTURE population structure analysis of the 350 accessions used for GWAS.

Figure S4. Principal component analysis (PCA) using accession means of all 23 traits for each of the six independent trials.

Figure S5. Visualization of between-trials correlation matrices for the 23 traits.

Figure S6. Between-traits pairwise-complete Pearson correlation heatmap for each environment.

Figure S7. Physical position of all 423 quantitative trait loci (QTLs) detected for each of the 23 traits, mapped on the 12 pepper chromosomes.

Figure S8. *FShl-P3.4* and *FP-P3.3* fruit shape index (FShl) and fruit pungency (FP) overlapping QTLs on chromosome P3 lack a haplotype combining the alleles for narrow and elongated fruit with low pungency.

Table S1. Summary of *Capsicum* spp. QTLs in the literature from 1998 to 2022.

Table S2. Summary of genome-wide association studies on *Capsicum annuum* from 2014 to 2023.

Table S3. List of *Capsicum* spp. accessions included in the G2P-SOL core collection.

Table S4. Metadata of the six phenotypic trials.

Table S5. Characteristics of the 23 traits phenotyped in up to six trials.

Table S6. Number of marker-phenotype associations found according to the trial and to the genome-wide association study model.

Table S7. Full list of the 423 quantitative trait loci detected for the 23 traits.

Table S8. List of highest-confidence quantitative trait loci colocalized for several traits.

Table S9. Full list of candidate genes for the highest-confidence QTLs and their annotations.

REFERENCES

Alexander, D.H., **Novembre, J.** & **Lange, K.** (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Alimi, N.A., **Bink, M.C.A.M.**, **Dieleman, J.A.**, **Magán, J.J.**, **Wubs, A.M.**, **Palloix, A.** ***et al.*** (2013) Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. *Theoretical and Applied Genetics*, **126**, 2597–2625.

Anderson, C.M. & **Kohorn, B.D.** (2001) Inactivation of Arabidopsis SIP1 leads to reduced levels of sugars and drought tolerance. *Journal of Plant Physiology*, **158**, 1215–1219.

Arce-Rodríguez, M.L. & **Ochoa-Alejo, N.** (2017) An R2R3-MYB transcription factor regulates capsaicinoid biosynthesis. *Plant Physiology*, **174**, 1359–1370.

Arpaci, B.B. & **Karataş, K.** (2020) Comparison of chili pepper breeding populations for agronomic traits and polygenic resistance to *Phytophthora* blight. *Horticultura Brasileira*, **38**, 12–20.

Arrones, A., **Mangino, G.**, **Alonso, D.**, **Plazas, M.**, **Prohens, J.**, **Portis, E.** ***et al.*** (2022) Mutations in the SmAPRR2 transcription factor suppressing chlorophyll pigmentation in the eggplant fruit peel are key drivers of a diversified colour palette. *Frontiers in Plant Science*, **13**, 1025951.

Barboza, G.E., **García, C.C.**, **Bianchetti, L.d.B.**, **Romero, M.V.** & **Scaldaferro, M.** (2022) Monograph of wild and cultivated chili peppers (*Capsicum* L., Solanaceae). *PhytoKeys*, **200**, 1–423.

Bates, D., **Mächler, M.**, **Bolker, B.** & **Walker, S.** (2015) Fitting linear mixed-effects models using lme4. *The Journal of Statistical Software*, **67**, 1–48. Available from: https://doi.org/10.18637/jss.v067.i01

Behr, A.A., **Liu, K.Z.**, **Liu-Fang, G.**, **Nakka, P.** & **Ramachandran, S.** (2016) pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics*, **32**, 2817–2823.

Bernal-Vasquez, A.-M., **Utz, H.-F.** & **Piepho, H.-P.** (2016) Outlier detection methods for generalized lattices: a case study on the transition from ANOVA to REML. *Theoretical and Applied Genetics*, **129**, 787–804.

Bernard, A., **Crabier, J.**, **Donkpegan, A.S.L.**, **Marrano, A.**, **Lheureux, F.** & **Dirlewanger, E.** (2021) Genome-wide association study reveals candidate genes involved in fruit trait variation in Persian walnut (*Juglans regia* L.). *Frontiers in Plant Science*, **11**, 607213.

Bertin, N., **Martre, P.**, **Genard, M.**, **Quilot, B.** & **Salon, C.** (2010) Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits. *Journal of Experimental Botany*, **61**, 955–967.

Blum, E., **Liu, K.**, **Mazourek, M.**, **Yoo, E.Y.**, **Jahn, M.** & **Paran, I.** (2002) Molecular mapping of the *C* locus for presence of pungency in *Capsicum*. *Genome*, **45**, 702–705.

Böhlenius, H., **Huang, T.**, **Charbonnel-Campaa, L.**, **Brunner, A.M.**, **Jansson, S.**, **Strauss, S.H.** ***et al.*** (2006) *CO/FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science*, **312**, 1040–1043.

Borovsky, Y., **Oren-Shamir, M.**, **Ovadia, R.**, **De Jong, W.** & **Paran, I.** (2004) The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to Anthocyanin2 of Petunia. *Theoretical and Applied Genetics*, **109**, 23–29.

Borovsky, Y., **Raz, A.**, **Doron-Faigenboim, A.**, **Zemach, H.**, **Karavani, E.** & **Paran, I.** (2022) Pepper fruit elongation is controlled by *Capsicum annuum* ovate family protein 20. *Frontiers in Plant Science*, **12**, 815589.

Brown, A.H.D. & **Spillane, C.** (1999) Implementing core collections - principles, procedures, progress, problems and promise. In: Johanson, R.C. & Hodgkin, T. (Eds.) *Core collections for today and tomorrow*. Rome, Italy: International Plant Genetic Resources Institute, pp. 1–9.

Cabral, N.S.S., **Medeiros, A.M.**, **Neves, L.G.**, **Sudré, C.P.**, **Pimenta, S.**, **Coelho, V.J.** ***et al.*** (2017) Genotype × environment interaction on experimental hybrids of chili pepper. *Genetics and Molecular Research*, **16**, 29551.

Cao, A., **Jin, J.**, **Li, S.** & **Wang, J.** (2017) Integrated analysis of mRNA and miRNA expression profiling in rice backcrossed progenies (BC2F12) with different plant height. *PLoS One*, **12**, e0184106.

Cao, Y., **Wen, L.**, **Wang, Z.** & **Ma, L.** (2015) SKIP interacts with the Paf1 complex to regulate flowering via the activation of FLC transcription in Arabidopsis. *Molecular Plant*, **8**, 1816–1819.

Che, G., **Gu, R.**, **Zhao, J.**, **Liu, X.**, **Song, X.**, **Zi, H.** ***et al.*** (2020) Gene regulatory network of carpel number variation in cucumber. *Development*, **147**, dev.184788.

Chu, Y., **Jang, J.**, **Huang, Z.** & **van der Knaap, E.** (2019) Tomato locule number and fruit size controlled by natural alleles of *lc* and *fas*. *Plant Direct*, **3**, e00142. Available from: https://doi.org/10.1002/pld3.142

Colonna, V., **D'Agostino, N.**, **Garrison, E.**, **Albrechtsen, A.**, **Meisner, J.**, **Facchiano, A.** ***et al.*** (2019) Genomic diversity and novel genome-wide association with fruit morphology in *Capsicum*, from 746k polymorphic sites. *Scientific Reports*, **9**, 10067.

Cui, Z., **Tong, A.**, **Huo, Y.**, **Yan, Z.**, **Yang, W.**, **Yang, X.** ***et al.*** (2017) SKIP controls flowering time via the alternative splicing of SEF pre-mRNA in Arabidopsis. *BMC Biology*, **15**, 80.

Cullis, B.R., **Smith, A.B.** & **Coombes, N.E.** (2006) On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics*, **11**, 381–393.

DeBolt, S., **Scheible, W.-R.**, **Schrick, K.**, **Auer, M.**, **Beisson, F.**, **Bischoff, V.** ***et al.*** (2009) Mutations in UDP-glucose:sterol glucosyltransferase in Arabidopsis cause transparent testa phenotype and suberization defect in seeds. *Plant Physiology*, **151**, 78–87.

DePristo, M.A., **Banks, E.**, **Poplin, R.**, **Garimella, K.V.**, **Maguire, J.R.**, **Hartl, C.** ***et al.*** (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Deshpande, R.B. (1935) Studies in Indian chillies. 4. Inheritance of pungency in *Capsicum annuum* L. *Indian Journal of Agricultural Sciences*, **5**, 513–516.

Elliott, R.C., **Betzner, A.S.**, **Huttner, E.**, **Oakes, M.P.**, **Tucker, W.Q.**, **Gerentes, D.** ***et al.*** (1996) AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. *Plant Cell*, **8**, 155–168.

Fan, M.J., **Chen, S.** & **Engle, M.** (2004) The study on genetic diversity of the core collection for *Capsicum* using random amplified polymorphic DNA. *Journal of Agricultural Research of China*, **53**, 165–178.

Fang, H., **Wang, P.**, **Wang, W.**, **Peng, J.**, **Zheng, J.**, **Zhu, G.** ***et al.*** (2023) Fine mapping and identification of SmAPRR2 regulating rind color in eggplant (*Solanum melongena* L.). *International Journal of Molecular Sciences*, **24**, 3059.

Fox, J. & **Weisberg, S.** (2019) *An R companion to applied regression*, 3rd edition. Thousand Oaks, CA: Sage Publications.

Francis, R.M. (2017) pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, **17**, 27–32.

Frankel, O.H. (1984) Genetic perspectives of germplasm conservation. In: Arber, W., Llimensee, K., Peacock, W.J. & Starlinger, P. (Eds.) *Genetic manipulation: impact on man and society*. Cambridge: Cambridge University Press, pp. 161–170.

Fu, J., **Yang, L.** & **Dai, S.** (2015) Identification and characterization of the CONSTANS-like gene family in the short-day plant *Chrysanthemum lavandulifolium*. *Molecular Genetics and Genomics*, **290**, 1039–1054.

Fuentes, R.R., **De Ridder, D.**, **van Dijk, A.D.J.** & **Peters, S.A.** (2022) Domestication shapes recombination patterns in tomato. *Molecular Biology and Evolution*, **39**, msab287.

Grosjean, P., **Ibanez, F.** & **Etienne, M.** (2018) *Pastecs: package for analysis of space-time ecological series*. R package version 1.3.21. Available from: https://cran.r-project.org/web/packages/pastecs [Accessed September 2022]

Guérinier, T., **Millan, L.**, **Crozet, P.**, **Oury, C.**, **Rey, F.**, **Valot, B. *et al.*** (2013) Phosphorylation of $p27^{KIP1}$ homologs KRP6 and 7 by SNF1-related protein kinase-1 links plant energy homeostasis and cell proliferation. *The Plant Journal*, **75**, 515–525.

Hackbusch, J., **Richter, K.**, **Müller, J.**, **Salamini, F.** & **Uhrig, J.F.** (2005) A central role of *Arabidopsis thaliana* ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 4908–4912.

Han, K., **Jang, S.**, **Lee, J.-H.**, **Lee, D.-G.**, **Kwon, J.-K.** & **Kang, B.-C.** (2019) A MYB transcription factor is a candidate to control pungency in *Capsicum annuum*. *Theoretical and Applied Genetics*, **132**, 1235–1246.

Hanson, P.M., **Yang, R.**, **Lin, S.**, **Tsou, S.C.S.**, **Lee, T.-C.**, **Wu, J. *et al.*** (2004) Variation for antioxidant activity and antioxidants in a subset of AVRDC —the World Vegetable Center *Capsicum* core collection. *Plant Genetic Resources*, **2**, 153–166.

Hill, T.A., **Ashrafi, H.**, **Reyes-Chin-Wo, S.**, **Yao, J.**, **Stoffel, K.**, **Truco, M.-J. *et al.*** (2013) Characterization of *Capsicum annuum* genetic diversity and population structure based on parallel polymorphism discovery with a 30K unigene pepper genechip. *PLoS One*, **8**, e56200.

Huang, M., **Liu, X.**, **Zhou, Y.**, **Summers, R.M.** & **Zhang, Z.** (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*, **8**, giy154. Available from: https://doi.org/10.1093/gigascience/giy154/5238723

Hudson, D., **Guevara, D.**, **Yaish, M.W.**, **Hannam, C.**, **Long, N.**, **Clarke, J.D. *et al.*** (2011) GNC and CGA1 modulate chlorophyll biosynthesis and glutamate synthase (GLU1/Fd-GOGAT) expression in Arabidopsis. *PLoS One*, **6**, e26765.

Hyun, T.K., **van der Graaff, E.**, **Albacete, A.**, **Eom, S.H.**, **Großkinsky, D.K.**, **Böhm, H. *et al.*** (2014) The Arabidopsis PLAT domain protein1 is critically involved in abiotic stress tolerance. *PLoS One*, **9**, e112946.

IPGRI, AVRDC & CATIE. (1995) *Descriptors for Capsicum (Capsicum spp.)*. Rome, Italy: International Plant Genetic Resources Institute.

Jansen, J. & **van Hintum, T.** (2007) Genetic distance sampling: a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. *Theoretical and Applied Genetics*, **114**, 421–428.

Jeong, H.B., **Jang, S.J.**, **Kang, M.Y.**, **Kim, S.**, **Kwon, J.K.** & **Kang, B.C.** (2020) Candidate gene analysis reveals that the fruit color locus *C1* corresponds to *PRR2* in pepper (*Capsicum frutescens*). *Frontiers in Plant Science*, **11**, 399.

Jeong, H.S., **Jang, S.**, **Han, K.**, **Kwon, J.K.** & **Kang, B.C.** (2015) Marker-assisted backcross breeding for development of pepper varieties (*Capsicum annuum*) containing capsinoids. *Molecular Breeding*, **35**, 226.

Ju, L., **Lei, X.**, **Zhao, C.**, **Shu, H.**, **Wang, Z.** & **Cheng, S.** (2020) Identification of MYB family genes and its relationship with pungency of pepper. *Acta Horticulturae Sinica*, **47**, 875.

Junior e Silva, W.C., **de Carvalho, S.I.C.** & **Duarte, J.B.** (2013) Identification of minimum descriptors for characterization of *Capsicum* spp. germplasm. *Horticultura Brasileira*, **31**, 190–202.

Kang, M.S. (2004) Breeding: genotype-by-environment interaction. In: Goodman, R.M. (Ed.) *Encyclopedia of plant and crop science*. New York, USA: Marcel-Dekker, pp. 218–221.

Kang, M.S. (2020) Genotype-environment interaction and stability analyses: an update. In: *Quantitative genetics, genomics and plant breeding*. UK: CABI, pp. 140–161. Available from: https://doi.org/10.1079/9781789240214.0140

Kim, S., **Park, J.**, **Yeom, S.-I.**, **Kim, Y.-M.**, **Seo, E.**, **Kim, K.-T. *et al.*** (2017) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biology*, **18**, 210.

Kim, S., **Park, M.**, **Yeom, S.-I.**, **Kim, Y.M.**, **Lee, J.M.**, **Lee, H.A. *et al.*** (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature Genetics*, **46**, 270–278.

Kormos, J. & **Kormos, K.** (1960) Die Genetischen Typen der Carotinoid-Systeme der Parpikafrucht. *Acta Botanica Academiae Scientiarum Hungaricae*, **6**, 305–319.

Korte, A. & **Farlow, A.** (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, **9**, 29.

Krizek, B.A. (2009) *AINTEGUMENTA* and *AINTEGUMENTA-LIKE6* act redundantly to regulate Arabidopsis floral growth and patterning. *Plant Physiology*, **150**, 1916–1929.

Krizek, B.A., **Blakley, I.C.**, **Ho, Y.**, **Freese, N.** & **Loraine, A.E.** (2020) The Arabidopsis transcription factor AINTEGUMENTA orchestrates patterning genes and auxin signaling in the establishment of floral growth and form. *The Plant Journal*, **103**, 752–768.

Lee, H.-Y., **Ro, N.-Y.**, **Jeong, H.-J.**, **Kwon, J.K.**, **Jo, J.**, **Ha, Y. *et al.*** (2016) Genetic diversity and population structure analysis to construct a core collection from a large *Capsicum* germplasm. *BMC Genetics*, **17**, 142.

Lee, H.-Y., **Ro, N.-Y.**, **Patil, A.**, **Lee, J.-H.**, **Kwon, J.-K.** & **Kang, B.-C.** (2020) Uncovering candidate genes controlling major fruit-related traits in pepper via genotype-by-sequencing based QTL mapping and genome-wide association study. *Frontiers in Plant Science*, **11**, 1100.

Lee, S.B., **Kim, J.E.**, **Kim, H.T.**, **Lee, G.-M.**, **Kim, B.-S.** & **Lee, J.M.** (2020) Genetic mapping of the *c1* locus by GBS-based BSA-seq revealed Pseudo-Response Regulator 2 as a candidate gene controlling pepper fruit color. *Theoretical and Applied Genetics*, **133**, 1897–1910.

Lefebvre, V., **Kuntz, M.**, **Camara, B.** & **Palloix, A.** (1998) The capsanthin-capsorubin synthase gene: a candidate gene for the y locus controlling the red fruit colour in pepper. *Plant Molecular Biology*, **36**, 785–789.

Lefebvre, V., **Pflieger, S.**, **Thabuis, A.**, **Caranta, C.**, **Blattes, A.**, **Chauvet, J.-C. *et al.*** (2002) Towards the saturation of the pepper linkage map by alignment of three intraspecific maps including known-function genes. *Genome*, **45**(5), 839–854.

Li, E., **Wang, S.**, **Liu, Y.**, **Chen, J.-G.** & **Douglas, C.J.** (2011) OVATE FAMILY PROTEIN4 (OFP4) interaction with KNAT7 regulates secondary cell wall formation in *Arabidopsis thaliana*: KNAT7-OFP complex regulates secondary wall formation. *The Plant Journal*, **67**, 328–341.

Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.

Li, H. (2013) *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. Available from: https://arxiv.org/abs/1303.3997 [Accessed 11th January 2023]

Li, Y.-Y., **Mao, K.**, **Zhao, C.**, **Zhao, X.-Y.**, **Zhang, H.-L.**, **Shu, H.-R. *et al.*** (2012) MdCOP1 ubiquitin E3 ligases interact with MdMYB1 to regulate light-induced anthocyanin biosynthesis and red fruit coloration in apple. *Plant Physiology*, **160**, 1011–1022.

Li, Z., **Wang, S.**, **Gui, X.-L.**, **Chang, X.-B.** & **Gong, Z.-H.** (2013) A further analysis of the relationship between yellow ripe-fruit color and the capsanthin-capsorubin synthase gene in pepper (*Capsicum* sp.) indicated a new mutant variant in *C. annuum* and a tandem repeat structure in promoter region. *PLoS One*, **8**, e61996.

Lin, S.-W., **Shieh, H.**, **Wang, Y.-W.**, **Tan, C.-W.**, **Schafleitner, R.**, **Yang, W.-J. *et al.*** (2015) Restorer breeding in sweet pepper: introgressing Rf allele from hot pepper through marker-assisted backcrossing. *Scientia Horticulturae*, **197**, 170–175.

Lira, B.S., **Gramegna, G.**, **Trench, B.A.**, **Alves, F.R.R.**, **Silva, E.M.**, **Silva, G.F.F. *et al.*** (2017) Manipulation of a senescence-associated gene improves fleshy fruit yield. *Plant Physiology*, **175**, 77–91.

Liu, H., **Jiao, J.**, **Liang, X.**, **Liu, J.**, **Meng, H.**, **Chen, S. *et al.*** (2016) Map-based cloning, identification and characterization of the w gene controlling white immature fruit color in cucumber (*Cucumis sativus* L.). *Theoretical and Applied Genetics*, **129**, 1247–1256.

Liu, J., **van Eck, J.**, **Cong, B.** & **Tanksley, S.D.** (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 13302–13306.

Liu, J., **Yu, J.**, **McIntosh, L.**, **Kende, H.** & **Zeevaart, J.A.D.** (2001) Isolation of a *CONSTANS* ortholog from *Pharbitis nil* and its role in flowering. *Plant Physiology*, **125**, 1821–1830.

Liu, Y. & **Douglas, C.J.** (2015) A role for OVATE FAMILY PROTEIN1 (OFP1) and OFP4 in a BLH6-KNAT7 multi-protein complex regulating secondary cell wall formation in *Arabidopsis thaliana*. *Plant Signaling & Behavior*, **10**, e1033126.

Lippert, L.F., **Bergh, B.O.** & **Smith, P.G.** (1965) Gene list for the pepper. *Journal of Heredity*, **56**(1), 30–34.
Long, J. & **Barton, M.K.** (2000) Initiation of axillary and floral meristems in Arabidopsis. *Developmental Biology*, **218**, 341–353.
Lozada, D.N., **Barchenger, D.W.**, **Coon, D.**, **Bhatta, M.** & **Bosland, P.W.** (2022) Multi-locus association mapping uncovers the genetic basis of yield and agronomic traits in Chile pepper (*Capsicum* spp.). *Crop Breeding, Genetics and Genomics*, **4**(2), e220002.
Lozano-Isla, F. (2022) *inti: tools and statistical procedures in plant science*. R package version 0.6.0. Available from: https://CRAN.R-project.org/package=inti [Accessed September 2022]
Luptovčiak, I., **Komis, G.**, **Takáč, T.**, **Ovečka, M.** & **Šamaj, J.** (2017) Katanin: a sword cutting microtubules for cellular, developmental, and physiological purposes. *Frontiers in Plant Science*, **8**, 1982.
Ma, L., **Liu, Z.**, **Cheng, Z.**, **Gou, J.**, **Chen, J.**, **Yu, W.** ***et al.*** (2021) Identification and application of BhAPRR2 controlling peel colour in wax gourd (*Benincasa hispida*). *Frontiers in Plant Science*, **12**, 716772.
Ma, X., **Yu, Y.-N.**, **Jia, J.-H.**, **Li, Q.-H.** & **Gong, Z.-H.** (2022) The pepper MYB transcription factor CaMYB306 accelerates fruit coloration and negatively regulates cold resistance. *Scientia Horticulturae*, **295**, 110892.
Ma, X., **Zhang, Y.**, **Turečková, V.**, **Xue, G.-P.**, **Fernie, A.R.**, **Mueller-Roeber, B.** ***et al.*** (2018) The NAC transcription factor SlNAP2 regulates leaf senescence and fruit yield in tomato. *Plant Physiology*, **177**, 1286–1302.
Malosetti, M., **Ribaut, J.-M.** & **van Eeuwijk, F.A.** (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology*, **4**, 44.
Mangin, B., **Siberchicot, A.**, **Nicolas, S.**, **Doligez, A.**, **This, P.** & **Cierco-Ayrolles, C.** (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, **108**, 285–291.
Mizuno, H., **Yazawa, T.**, **Kasuga, S.**, **Sawada, Y.**, **Kanamori, H.**, **Ogo, Y.** ***et al.*** (2016) Expression of flavone synthase II and flavonoid 3′-hydroxylase is associated with color variation in tan-colored injured leaves of sorghum. *Frontiers in Plant Science*, **7**, 10.3389/fpls.2016.01718/full.
Montero-Pau, J., **Blanca, J.**, **Esteras, C.**, **Martínez-Pérez, E.M.**, **Gómez, P.**, **Monforte, A.J.** ***et al.*** (2017) An SNP-based saturated genetic map and QTL analysis of fruit-related traits in Zucchini using genotyping-by-sequencing. *BMC Genomics*, **18**, 94.
Mu, Q., **Huang, Z.**, **Chakrabarti, M.**, **Illa-Berenguer, E.**, **Liu, X.**, **Wang, Y.** ***et al.*** (2017) Fruit weight is controlled by cell size regulator encoding a novel protein that is expressed in maturing tomato fruits. *PLoS Genetics*, **13**, e1006930.
Muños, S., **Ranc, N.**, **Botton, E.**, **Bérard, A.**, **Rolland, S.**, **Duffé, P.** ***et al.*** (2011) Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near *WUSCHEL*. *Plant Physiology*, **156**, 2244–2254.
Nicolaï, M., **Cantet, M.**, **Lefebvre, V.**, **Sage-Palloix, A.-M.** & **Palloix, A.** (2013) Genotyping a large collection of pepper (*Capsicum* spp.) with SSR loci brings new evidence for the wild origin of cultivated *C. annuum* and the structuring of genetic diversity by human selection of cultivar types. *Genetic Resources and Crop Evolution*, **60**, 2375–2390.
Nimmakayala, P., **Abburi, V.L.**, **Abburi, L.**, **Alaparthi, S.B.**, **Cantrell, R.**, **Park, M.** ***et al.*** (2014) Linkage disequilibrium and population-structure analysis among *Capsicum annuum* L. cultivars for use in association mapping. *Molecular Genetics and Genomics*, **289**, 513–521.
Nimmakayala, P., **Abburi, V.L.**, **Saminathan, T.**, **Alaparthi, S.B.**, **Almeida, A.**, **Davenport, B.** ***et al.*** (2016) Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum annuum* L. *Scientific Reports*, **6**, 38081.
Nimmakayala, P., **Lopez-Ortiz, C.**, **Shahi, B.**, **Abburi, V.L.**, **Natarajan, P.**, **Kshetry, A.O.** ***et al.*** (2021) Exploration into natural variation for genes associated with fruit shape and size among *Capsicum chinense* collections. *Genomics*, **113**, 3002–3014.
Oren, E., **Tzuri, G.**, **Vexler, L.**, **Dafna, A.**, **Meir, A.**, **Faigenboim, A.** ***et al.*** (2019) The multi-allelic APRR2 gene is associated with fruit pigment accumulation in melon and watermelon. *Journal of Experimental Botany*, **70**, 3781–3794.
Ozalp, R., **Celik, I.**, **Boyaci, H.F.**, **Kabas, A.**, **Golukcu, M.** & **Unlu, A.** (2020) Collection, characterization and conservation for sustainability of traditional local pepper genotypes. *Acta Horticulturae*, **1282**, 141–148.
Pagnussat, G.C., **Yu, H.-J.** & **Sundaresan, V.** (2007) Cell-fate switch of synergid to egg cell in *Arabidopsis eostre* mutant embryo sacs arises from misexpression of the BEL1-like homeodomain gene *BLH1*. *The Plant Cell*, **19**, 3578–3592.
Pan, Y., **Bradley, G.**, **Pyke, K.**, **Lu, C.**, **Fray, R.**, **Marshall, A.** ***et al.*** (2013) Network inference analysis identifies an *APRR2-like* gene linked to pigment accumulation in tomato and pepper fruits. *Plant Physiology*, **161**, 1476–1485.
Paran, I. & **Fallik, E.** (2011) Breeding for Fruit Quality in Pepper (*Capsicum* spp.). In: Jenks, M.A. & Bebeli, P.J. (Eds.) *Breeding for fruit quality*. Hoboken, NJ, USA: John Wiley & Sons Inc, pp. 307–322.
Parsons, E.P., **Popopvsky, S.**, **Lohrey, G.T.**, **Lü, S.**, **Alkalai-Tuvia, S.**, **Perzelan, Y.** ***et al.*** (2012) Fruit cuticle lipid composition and fruit post-harvest water loss in an advanced backcross generation of pepper (*Capsicum* sp.). *Physiologia Plantarum*, **146**, 15–25.
Pereira-Dias, L., **Vilanova, S.**, **Fita, A.**, **Prohens, J.** & **Rodríguez-Burruezo, A.** (2019) 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). *Horticultural Research*, **6**, 54.
Perez-de-Castro, A.M., **Vilanova, S.**, **Canizares, J.**, **Pascual, L.**, **Blanca, J.**, **Diez, M.** ***et al.*** (2012) Application of genomic tools in plant breeding. *Current Genomics*, **13**, 179–195.
Pflieger, S., **Lefebvre, V.** & **Causse, M.** (2001) The candidate gene approach in plant genetics: a review. *Molecular Breeding*, **7**, 275–291.
Phan, N.T. & **Sim, S.-C.** (2017) Genomic tools and their implications for vegetable breeding. *Korean Journal of Horticultural Science and Technology*, **35**, 149–164.
Piepho, H.-P. & **Möhring, J.** (2007) Computing heritability and selection response from unbalanced plant breeding trials. *Genetics*, **177**, 1881–1888.
Popovsky, S. & **Paran, I.** (2000) Molecular genetics of the y locus in pepper: its relation to capsanthin-capsorubin synthase and to fruit color. *Theoretical and Applied Genetics*, **101**, 86–89.
Preuß, A., **Stracke, R.**, **Weisshaar, B.**, **Hillebrecht, A.**, **Matern, U.** & **Martens, S.** (2009) *Arabidopsis thaliana* expresses a second functional flavonol synthase. *FEBS Letters*, **583**, 1981–1986.
Price, A.L., **Patterson, N.J.**, **Plenge, R.M.**, **Weinblatt, M.E.**, **Shadick, N.A.** & **Reich, D.** (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**, 904–909.
Price, A.L., **Zaitlen, N.A.**, **Reich, D.** & **Patterson, N.** (2010) New approaches to population stratification in genome-wide association studies. *Nature Reviews. Genetics*, **11**, 459–463.
Purcell, S. & **Chang, C.** (2019) *PLINK2*. R package version 2.00. Available from: https://www.cog-genomics.org/plink/2.0/ [Accessed November 2022]
Qi, J., **Qian, Q.**, **Bu, Q.**, **Li, S.**, **Chen, Q.**, **Sun, J.** ***et al.*** (2008) Mutation of the rice *narrow leaf1* gene, which encodes a novel protein, affects vein patterning and polar auxin transport. *Plant Physiology*, **147**, 1947–1959.
Qi, Z. & **Xiong, L.** (2013) Characterization of a purine permease family gene *OsPUP7* involved in growth and development control in rice. *Journal of Integrative Plant Biology*, **55**, 1119–1135.
Quenouille, J., **Saint-Felix, L.**, **Moury, B.** & **Palloix, A.** (2016) Diversity of genetic backgrounds modulating the durability of a major resistance gene. Analysis of a core collection of pepper landraces resistant to *Potato virus Y*: genetic backgrounds shape *R*-gene durability. *Molecular Plant Pathology*, **17**, 296–302.
R Core Team. (2021) *R: a language and environment for statistical computing, version 4.1.2*. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.r-project.org/ [Accessed September 2022]
Rani, M., **Jindal, S.K.**, **Vikal, Y.** & **Meena, O.P.** (2021) Genetic male sterility breeding in heat tolerant bell pepper: introgression of ms10 gene from hot pepper through marker-assisted backcrossing. *Scientia Horticulturae*, **285**, 110172.
Rao, G.U., **Ben Chaim, A.**, **Borovsky, Y.** & **Paran, I.** (2003) Mapping of yield-related QTLs in pepper in an interspecific cross of *Capsicum annuum* and *C. frutescens*. *Theoretical and Applied Genetics*, **106**, 1457–1466.
Richter, R., **Behringer, C.**, **Müller, I.K.** & **Schwechheimer, C.** (2010) The GATA-type transcription factors GNC and GNL/CGA1 repress gibberellin

signaling downstream from DELLA proteins and PHYTOCHROME-INTERACTING FACTORS. *Genes & Development*, **24**, 2093–2104.

Richter, R., **Behringer, C.**, **Zourelidou, M.** & **Schwechheimer, C.** (2013) Convergence of auxin and gibberellin signaling on the regulation of the GATA transcription factors *GNC* and *GNL* in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 13192–13197.

Ro, N., **Haile, M.**, **Hur, O.**, **Geum, B.**, **Rhee, J.**, **Hwang, A.** ***et al.*** (2022) Genome-wide association study of resistance to *Phytophthora capsici* in the pepper (*Capsicum* spp.) collection. *Frontiers in Plant Science*, **13**, 902464.

Salinier, J., **Lefebvre, V.**, **Besombes, D.**, **Burck, H.**, **Causse, M.**, **Daunay, M.-C.** ***et al.*** (2022) The INRAE Centre for Vegetable Germplasm: geographically and phenotypically diverse collections and their use in genetics and plant breeding. *Plants*, **11**, 347.

Schmitz, A.J., **Begcy, K.**, **Sarath, G.** & **Walia, H.** (2015) Rice Ovate Family Protein 2 (OFP2) alters hormonal homeostasis and vasculature development. *Plant Science*, **241**, 177–188.

Segura, V., **Vilhjálmsson, B.J.**, **Platt, A.**, **Korte, A.**, **Seren, Ü.**, **Long, Q.** ***et al.*** (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, **44**, 825–830.

Shibuya, T., **Nishiyama, M.**, **Kato, K.** & **Kanayama, Y.** (2021) Characterization of the FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 homolog SIFKF1 in tomato as a model for plants with fleshy fruit. *International Journal of Molecular Sciences*, **22**, 1735.

Song, Z., **Zhong, J.**, **Dong, J.**, **Hu, F.**, **Zhang, B.**, **Cheng, J.** ***et al.*** (2022) Mapping immature fruit colour-related genes via bulked segregant analysis combined with whole-genome re-sequencing in pepper (*Capsicum annuum*). *Plant Breeding*, **141**, 277–285.

Stewart, C., **Kang, B.-C.**, **Liu, K.**, **Mazourek, M.**, **Moore, S.L.**, **Yoo, E.Y.** ***et al.*** (2005) The Pun1 gene for pungency in pepper encodes a putative acyltransferase. *The Plant Journal*, **42**, 675–688.

Stommel, J.R., **Lightbourn, G.J.**, **Winkel, B.S.** & **Griesbach, R.J.** (2009) Transcription factor families regulate the anthocyanin biosynthetic pathway in *Capsicum annuum*. *The Journal of the American Society for Horticultural Science*, **134**, 244–251.

Suárez-López, P., **Wheatley, K.**, **Robson, F.**, **Onouchi, H.**, **Valverde, F.** & **Coupland, G.** (2001) CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. *Nature*, **410**, 1116–1120.

Subudhi, P.K., **Garcia, R.S.**, **Coronejo, S.** & **De Leon, T.B.** (2020) A novel mutation of the NARROW LEAF 1 gene adversely affects plant architecture in rice (*Oryza sativa* L.). *International Journal of Molecular Sciences*, **21**, 8106.

Sun, B., **Zhou, X.**, **Chen, C.**, **Chen, C.**, **Chen, K.**, **Chen, M.** ***et al.*** (2020) Coexpression network analysis reveals an MYB transcriptional activator involved in capsaicinoid biosynthesis in hot peppers. *Horticultural Research*, **7**, 162.

Sun, B., **Zhu, Z.**, **Chen, C.**, **Chen, G.**, **Cao, B.**, **Chen, C.** ***et al.*** (2019) Jasmonate-inducible R2R3-MYB transcription factor regulates capsaicinoid biosynthesis and stamen development in *Capsicum*. *Journal of Agricultural and Food Chemistry*, **67**, 10891–10903.

Tamisier, L., **Szadkowski, M.**, **Nemouchi, G.**, **Lefebvre, V.**, **Szadkowski, E.**, **Duboscq, R.** ***et al.*** (2020) Genome-wide association mapping of QTLs implied in potato virus Y population sizes in pepper: evidence for widespread resistance QTL pyramiding. *Molecular Plant Pathology*, **21**, 3–16.

Tang, H.-Y., **Dong, X.**, **Wang, J.-K.**, **Xia, J.-H.**, **Xie, F.**, **Zhang, Y.** ***et al.*** (2018) Fine mapping and candidate gene prediction for white immature fruit skin in cucumber (*Cucumis sativus* L.). *International Journal of Molecular Sciences*, **19**, 1493.

Taranto, F., **D'Agostino, N.**, **Greco, B.**, **Cardi, T.** & **Tripodi, P.** (2016) Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. *BMC Genomics*, **17**, 943.

Thabuis, A., **Palloix, A.**, **Servin, B.**, **Daubèze, A.M.**, **Signoret, P.**, **Hospital, F.** ***et al.*** (2004) Marker-assisted introgression of 4 *Phytophthora capsici* resistance QTL alleles into a bell pepper line: validation of additive and epistatic effects. *Molecular Breeding*, **14**, 9–20.

Thies, J.A. & **Fery, R.L.** (2002) Evaluation of a core of the U.S. Capsicum Germplasm Collection for reaction to the northern root-knot nematode. *Hortscience*, **37**, 805–810.

Tripodi, P., **Ficcadenti, N.**, **Rotino, G.L.**, **Festa, G.**, **Bertone, A.**, **Pepe, A.** ***et al.*** (2019) Genotypic and environmental effects on the agronomic, health-related compounds and antioxidant properties of chilli peppers for diverse market destinations. *Journal of the Science of Food and Agriculture*, **99**, 4550–4560.

Tripodi, P., **Rabanus-Wallace, M.T.**, **Barchi, L.**, **Kale, S.**, **Esposito, S.**, **Acquadro, A.** ***et al.*** (2021) Global range expansion history of pepper (*Capsicum* spp.) revealed by over 10,000 genebank accessions. *Proceedings of the National Academy of Sciences of the United States of America*, **118**, e2104315118.

Tsaballa, A., **Pasentsis, K.**, **Darzentas, N.** & **Tsaftaris, A.S.** (2011) Multiple evidence for the role of an ovate-like gene in determining fruit shape in pepper. *BMC Plant Biology*, **11**, 46.

van der Knaap, E., **Kim, J.H.** & **Kende, H.** (2000) A novel gibberellin-induced gene from rice and its potential regulatory role in stem growth. *Plant Physiology*, **122**, 695–704.

Wahl, V., **Brand, L.H.**, **Guo, Y.-L.** & **Schmid, M.** (2010) The FANTASTIC FOUR proteins influence shoot meristem size in *Arabidopsis thaliana*. *BMC Plant Biology*, **10**, 285.

Wang, D. & **Bosland, P.W.** (2006) The genes of capsicum. *HortScience*, **41** (5), 1169–1187.

Wang, H., **Sun, J.**, **Yang, F.**, **Weng, Y.**, **Chen, P.**, **Du, S.** ***et al.*** (2021) CsKTN1 for a katanin p60 subunit is associated with the regulation of fruit elongation in cucumber (*Cucumis sativus* L.). *Theoretical and Applied Genetics*, **134**, 2429–2441.

Wang, J. & **Zhang, Z.** (2021) GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics & Bioinformatics*, **19**, 629–640.

Wang, J., **Zheng, C.**, **Shao, X.**, **Hu, Z.**, **Li, J.**, **Wang, P.** ***et al.*** (2020) Transcriptomic and genetic approaches reveal an essential role of the NAC transcription factor SINAP1 in the growth and defense response of tomato. *Horticultural Research*, **7**, 209.

Wang, J., **Liu, Y.**, **Tang, B.**, **Dai, X.**, **Xie, L.**, **Liu, F.** ***et al.*** (2020) Genome-wide identification and capsaicinoid biosynthesis-related expression analysis of the R2R3-MYB gene family in *Capsicum annuum* L. *Frontiers in Genetics*, **11**, 598183.

Wang, S., **Chang, Y.** & **Ellis, B.** (2016) Overview of OVATE FAMILY PROTEINS, a novel class of plant-specific growth regulators. *Frontiers in Plant Science*, **7**, 417. Available from: https://doi.org/10.3389/fpls.2016.00417

Wang, X., **Wu, F.**, **Xie, Q.**, **Wang, H.**, **Wang, Y.**, **Yue, Y.** ***et al.*** (2012) SKIP is a component of the spliceosome linking alternative splicing and the circadian clock in *Arabidopsis*. *Plant Cell*, **24**, 3278–3295.

Winkel-Shirley, B. (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology*, **126**, 485–493.

Wu, L., **Wang, P.**, **Wang, Y.**, **Liu, J.**, **Cheng, Q.**, **Lu, Q.** ***et al.*** (2019) Genome-wide correlation of 36 agronomic traits in the 287 pepper (*Capsicum*) accessions obtained from the SLAF-seq-based GWAS. *International Journal of Molecular Sciences*, **20**, 5675.

Xiao, Y., **Liu, D.**, **Zhang, G.**, **Gao, S.**, **Liu, L.**, **Xu, F.** ***et al.*** (2019) *Big Grain3*, encoding a purine permease, regulates grain size via modulating cytokinin transport in rice. *Journal of Integrative Plant Biology*, **61**, 581–597.

Yan, H., **Pei, X.**, **Zhang, H.**, **Li, X.**, **Zhang, X.**, **Zhao, M.** ***et al.*** (2021) MYB-mediated regulation of anthocyanin biosynthesis. *International Journal of Molecular Sciences*, **22**, 3103.

Yang, T., **He, Y.**, **Niu, S.**, **Yan, S.** & **Zhang, Y.** (2020) Identification and characterization of the CONSTANS (CO)/CONSTANS-like (COL) genes related to photoperiodic signaling and flowering in tomato. *Plant Science*, **301**, 110653.

Yang, W., **Guo, T.**, **Luo, J.**, **Zhang, R.**, **Zhao, J.**, **Warburton, M.L.** ***et al.*** (2022) Target-oriented prioritization: targeted selection strategy by integrating organismal and molecular traits through predictive analytics in breeding. *Genome Biology*, **23**, 80.

Yano, M., **Katayose, Y.**, **Ashikari, M.**, **Yamanouchi, U.**, **Monna, L.**, **Fuse, T.** ***et al.*** (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. *Plant Cell*, **12**, 2473–2483.

Yin, L., **Zhang, H.**, **Tang, Z.**, **Xu, J.**, **Yin, D.**, **Zhang, Z.** ***et al.*** (2021) rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics & Bioinformatics*, **19**, 619–628.

Yu, J., **Pressoir, G.**, **Briggs, W.H.**, **Vroh Bi, I.**, **Yamasaki, M.**, **Doebley, J.F.** ***et al.*** (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, **38**, 203–208.

Zewdie, Y., **Tong, N.** & **Bosland, P.** (2004) Establishing a core collection of Capsicum using a cluster analysis with enlightened selection of accessions. *Genetic Resources and Crop Evolution*, **51**, 147–151.

Zhang, L., **Zhang, X.**, **Ju, H.**, **Chen, J.**, **Wang, S.**, **Wang, H.** ***et al.*** (2016) Ovate family protein1 interaction with BLH3 regulates transition timing from vegetative to reproductive phase in Arabidopsis. *Biochemical and Biophysical Research Communications*, **470**, 492–497.

Zheng, X., **Levine, D.**, **Shen, J.**, **Gogarten, S.M.**, **Laurie, C.** & **Weir, B.S.** (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.

Zhou, H., **Ma, R.**, **Gao, L.**, **Zhang, J.**, **Zhang, A.**, **Zhang, X.** ***et al.*** (2021) A 1.7-Mb chromosomal inversion downstream of a *PpOFP1* gene is responsible for flat fruit shape in peach. *Plant Biotechnology Journal*, **19**, 192–205.

Zhu, C., **Gore, M.**, **Buckler, E.S.** & **Yu, J.** (2008) Status and prospects of association mapping in plants. *Plant Genome*, **1**, 89.

Zhu, L., **Wang, Y.**, **Zhang, Z.**, **Hu, D.**, **Wang, Z.**, **Hu, J.** ***et al.*** (2022) Chromosomal fragment deletion in APRR2-repeated locus modulates the dark stem color in *Cucurbita pepo*. *Theoretical and Applied Genetics*, **135**, 4277–4288.

Zygier, S., **Chaim, A.B.**, **Efrati, A.**, **Kaluzky, G.**, **Borovsky, Y.** & **Paran, I.** (2005) QTLs mapping for fruit size and shape in chromosomes 2 and 4 in pepper and a comparison of the pepper QTL map with that of tomato. *Theoretical and Applied Genetics*, **111**, 437–445.