Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative analysis of Cucurbita pepo genomes shed light in agronomic traits variation

Clara Perez-Moro
COMAV, Instituto de Conservación y Mejora de la Agrodiversidad, Universitat Politècnica de València
Daniela D'Esposito
University of Naples Federico II
Ana Pérez-de-Castro
COMAV, Instituto de Conservación y Mejora de la Agrodiversidad, Universitat Politècnica de València
Maria Ercolano
ercolano@unina.it
University of Naples Federico II
Claudio Capuozzo
University of Naples Federico II
Anna Guadagno
University of Naples Federico II

Research Article

Keywords: C. Pepo, MAGIC population, resequencing, SNP, IN/DEL, morphological traits, ZYMV, PM

Posted Date: April 4th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4162066/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Cucurbita pepo L. cultivars display high morphological traits variation. In addition, *C. pepo* faces numerous threats, such as viral and fungal infections, which significantly influence crop cultivation. Recent genomic advancements improved the understanding of genetic diversity and stress responses in this crop. We investigated genetic variations related to plant morphology and quality traits. Additionally, the inclusion of both powdery mildew (PM) and Zucchini yellow mosaic virus (ZYMV) susceptible and tolerant varieties facilitated the examination of genetic diversity concerning biotic stress.

Results

The sequencing of eight *Cucurbita pepo* varieties produced an average of 40 million raw reads with a coverage of reference genome ranging from 22 to 40X. More than 4.7 million genomic variants were identified in all genomes. Based on admixture and PCA analysis, the eight *C. pepo* genotypes resulted organized in two clusters belonging to Cocozelle and Zucchini groups, with “Whitaker” separated from the rest of the accessions. Genes involved in pathways related to gibberellin regulation, leaf development, and pigment accumulation resulted highly affected by variation suggesting that the diversity observed among varieties in plant and fruit morphology could be related to variants identified in such genes. Each variety showed its own set of genetic differences. The genomic comparison of 381e, 968Rb and SPQ allowed the identification of variants in chromosome regions affecting response to Zucchini yellow mosaic virus (ZYMV) and powdery mildew (PM). Variants in key genes associated with resistant traits were identified, suggesting potential pathways and mechanisms involved in biotic stress response and plant immunity.

Conclusions

Genetic variations affecting morphology and fruit quality in *C. pepo* emphasize their significance for breeding efforts. Furthermore, the genomic comparison of 381e, 968Rb and SPQ highlighted variants in chromosomal regions influencing zucchini's response to PM and ZYMV. These findings could pave the way for more targeted and effective genetic improvement strategies, thereby potentially leading to increased agricultural productivity and quality.

BACKGROUND

Cucurbita pepo L. (2n = 2x = 40) is the most widely grown and polymorphic of the *Cucurbita* species [1] and it is an economically important crop in many countries [2].Two distinct subspecies of *C. pepo* originated in two independent events, C. *pepo* ssp. *pepo*, and C. *pepo* ssp. *ovifera* [3]. Within the two subspecies, cultivars varying for growth habit, leaves size and shape and fruit morphology and texture [4,

5], are classified in eight horticultural groups. The Pumpkin, Vegetable marrow, Cocozelle and Zucchini groups belong to C. *pepo* ssp. *pepo* L. and the Acorn, Scallop, Crookneck and Straightneck groups belong to C. *pepo* ssp. *texana* (Scheele) Filov (syn ssp. *ovifera* (L.) Decker) [4–7].

Metabolic pathways regulating the synthesis of specific compounds, including phytohormones and pigments, influence important traits. For instance, plant dwarfism is related to the biosynthesis and signalling of gibberellins [8]. On the other hand, cell wall expansins [9] are involved in root growth, stem elongation and leaf enlargement. Metabolic pathways that control the accumulation of pigments such as carotenoids, flavonoids and anthocyanins modulate colour of fruit [5]. Genes controlling fruit shape, locule number and overall fruit morphology are highly conserved across species. SUN controls fruit elongation both in tomato (*Solanum lycopersicum* L.) [10] and cucumber (*Cucumis sativus* L.) [11] as well as the SF1 gene, encoding a RING-type E3 ligase [12]. LONGIFOLIA 1 and LONGIFOLIA 2 have been found to produce slender grains in rice [13]. CLAVATA (CLV) and WUSCHEL (WUS) genes affect locule number in various species [14, 15].

Cucurbita pepo is affected by several treats, including virus, fungi and pests. Viruses, such as Zucchini yellow mosaic virus (ZYMV), Cucumber mosaic virus (CMV), Tomato leaf curl New Delhi virus (ToLCNDV), transmitted by aphids and whiteflies can significantly limit zucchini cultivation [16]. One of most destructive viral diseases affecting *C. pepo* crops is ZYMV that belongs to the highly virulent single-stranded RNA potyvirus group and is transmitted by various aphid species in a non-persistent manner [17]. Symptoms in *C. pepo* affected by ZYMV include the development of green blisters, leaf deformation, severe stunting and a mosaic pattern on leaves [17–19]. Beside viral infections, other pathogens also contribute to pumpkin and squash production decrease. Powdery mildew, caused by *Podosphaera xanthii* (Castagne) U. Braun and N. Shish (PM) and *Golovinomyces orontii* (Castagne) V.P Heluta, can cause extensive damage and is challenging to control [20, 21].

Despite the significance of genetic and genomic resources for this crop species, the advancement of trait mapping and gene discovery has been relatively delayed compared to other cucurbits species [22, 23]. The sequencing of the *C. pepo* reference genome, MU-CU-16, classified as zucchini type, was released in 2018 [24]. Subsequently, the genome resequencing of several *C. pepo* varieties was performed to investigate the genetic diversity present in different morphotypes [25–27]. The availability of the zucchini genome also allowed the exploration of the molecular basis of ZYMV tolerance through omics approaches. For example, a recent genomic scanning performed to identify the candidate genes involved in ZYMV tolerance [28] showed the presence of SNPs in genes located on chromosome 8 (LG08SNP4) and chromosome 1 (LG01SNP1) and correlated with the ZYMV tolerance.

In this work, we re-sequenced eight varieties that were employed as founders of a MAGIC population, analysing the genetic diversity and the population structure of the selected cultivars. Moreover, we explored the variation in genes implicated in plant morphological and quality traits. Finally, the presence of varieties tolerant and susceptible to virus and tolerant to fungus [29] within our dataset, allowed us to explore the genetic diversity related to biotic stress.

RESULTS

The sequencing of the eight zucchini varieties resulted in an average of 40 million raw reads with a minimum value of 28,804,339 reads in “Nano verde di Milano” (NV) and a maximum value of 52,915,127 reads in “Whitaker”. The average mapping rate of the trimmed and filtered reads ranged from 98.7–90.54%. The higher mapping value was provided by “Lungo Bianco di Sicilia” (BDS) and the lower value by “Striato d'Italia” (SI). We identified an average of 1.3 million of raw variants, including SNPs and IN/DEL, in each sample, and 4.7 million raw variants considering the eight varieties all together. After filtering, the number of SNPs was reduced to an average of 485,120 per sample and 3,831,907 in the all-varieties dataset and the number of IN/DEL was reduced to an average of 50,968 per sample and 387,243 in the all-varieties dataset (Table S1). The nucleotide changes identified in our analysis were classified based on their presence in the population. Specifically, 150,217 variants (141,714 SNPs, 4,596 insertions and 3,907 deletions) resulted common to 95% of individuals. Additionally, 2,111,951 variants (1,915,792 SNPs, 96,602 insertions and 99,557 deletions) were observed in the 10 to 95% of individuals. Lastly, the 10% of the individuals was affected by 1,302,967 unique variants (1,180,581 SNPs, 56,966 insertions and 65,420 deletions).

By comparing the distribution of variants along the chromosomes (Figure S1) we inferred the different density of variants between varieties. “Whitaker” showed the higher frequency, while 381e and 968Rb showed the lowest number of variants. The higher frequency observed in “Whitaker” may due by the higher number of mapped reads and to its genetic background [30]. In general, the distribution of variants found along the different chromosomes was homogenous. (Fig. 1; Figure S1). The average read depth across the eight varieties was 29.7, with “Whitaker” exhibiting the lowest value (22.3) and 968Rb the highest (45.6) (Fig. 1, Table S1).

Population structure and genetic variability

In order to infer the genetic structure of the population an admixture-based clustering analysis was performed. Two suitable subpopulation structures (K = 4 and K = 3) were obtained (Fig. 2). The structure based on four subgroups showed a first cluster, including BDS, SPQ and SI and a second cluster composed of 381e and 968Rb and NV, showing mixed background, “Whitaker” and OF remained ungrouped (Fig. 2A). Applying a value of K = 3, OF was merged with the first subgroup (BDS, SPQ and SI) while NV was included within the 381e and 968Rb subgroup (Fig. 2B).

PCA built up on genotypic data displayed that the varieties were divided into three major clusters according to the morphological characteristics. The first two principal components (PC1 and PC2), explained 38% and 18.4% of the total variance, respectively (Fig. 3A). The clustering obtained through PCA is in accordance with the results obtained in the previously admixture-based analysis. Zucchini varieties clustered together, NV clustered with zucchini varieties, close also to the Cocozelle group. In the latter group, SI and BDS are located very close and OF was located at bottom right side. “Whitaker” was on the opposite quadrant close to the left horizontal axis.

The consensus phylogenetic tree (Fig. 3B) confirmed that the varieties were organized in two clusters belonging to Cocozelle and Zucchini group and that “Whitaker” was separated from the rest of the accessions.

The association of alleles at different loci was estimate using LD decay (Fig. 4). Appling the test to the whole varieties collection, LD was low (r2 < 0.3), approximately 60 kb. Hereafter, LD decay was calculated also separately for Zucchini and Cocozelle group, showing to be variable among the different groups. Overall, the results showed a slightly faster decay in Cocozelle group in comparison with Zucchini group. Within groups, the LD decay of Zucchini group was r2 = 0.6, 100kbp, it was lower for Cocozelle group (60kbp; r2 = 0.28) (Fig. 4).

Private SNPs identification and functional annotation

High quality homozygous and heterozygous alternative sites, differing from the homozygous reference allele were identified in all varieties. This resulted in 1.18 million unique SNPs across the eight varieties, although 815 thousand of these private sites were identified in “Whitaker” (Table 1). A SnpEff analysis was performed on the variants in all eight varieties together (all-varieties variant calling set) to identify a set of genetic variants having a potential biological effect. The majority of potential functional SNPs displayed a modifier impact, followed by the low impact, moderate impact and high impact (Table 1). In our study, we identified a total of 49,828 high-effect SNPs that are specific to only one variety, in total, 983 genes have been significantly affected by these SNPs across all eight varieties examined (Table S2). These SNPs may play a crucial role in the observed phenotypic differences among the different varieties. Each zucchini variety had its own set of distinct genetic differences in comparison to the others. “Whitaker” and 381e stood out as they possessed the highest number of unique or private SNPs. In the case of “Whitaker”, this is likely due to its derivation from multiple parental sources, which could account for the increased genetic diversity and unique SNPs. Additionally, 381e possesses introgressions linked to its resistance against ZYMV, where three associated QTLs (SNP1, SNP2, and SNP3) were identified [31, 32]. This finding could explain the abundance of private SNPs observed in this variety. We observed an uneven distribution of unique SNP proportions across different chromosomes, with a notable higher proportion of unique SNPs in variety 381e on chromosomes 1, 8, and 18. Also, 968Rb possesses an introgression associated with PM resistance [33] and exhibits a higher proportion of unique SNPs on chromosome 10 (Table S3). This finding suggested that 381e and 968Rb genomes hold specific regions with higher diversity, potentially related with their ZYMV and PM resistances respectively.

Table 1
Unique SNPs in the eight varieties. BDS (Lungo Bianco di Sicilia), NV (Nano Verde di Milano), SPQ (San Pasquale), OF (Ortolano di Faenza), SI (Striato di Italia), "Whitaker", 381E and 968Rb.

	BDS	NV	SPQ	OF	SI	WHITAKER	381E	968Rb
Nº of SNPs	52,928	20,129	39,323	70,163	47,207	815,131	128,336	7,364
% high effect SNPs	0.043	0.037	0.063	0.051	0.036	0.04	0.048	0.039
% low effect SNPs	2.085	2.207	2.15	2.144	2.015	2.31	3.049	1.798
% moderate effect SNPs	1.42	1.55	1.489	1.514	1.337	1.531	1.779	1.223
% modifier effect SNPs	96.452	96.206	96.298	96.291	96.612	96.119	95.124	96.939

Genes involved in plant and fruit morphological traits

As a result of the SNP calling performed for the eight varieties, a multitude of variants related to the morphological traits were discovered. All genotypes tested showed a bush growth habit phenotype, except BDS and NV that exhibited a semi-trailing growth habit. Gene candidates for the dwarf phenotype (Table S4), involved to the gibberellin metabolic pathway were assesses to identify variants. Interestingly, the *C. pepo* gene *Cp4.1LG10g05910*, involved in the dwarf growth habit [34] not exhibited variants either in BDS than in NV. By contrast, variants in *Cp4.1LG12g07890*, an orthologue of a gibberellin 3-beta-dioxygenase 2-like gene related to the dwarf phenotype in *C. moschata* [35] were found only in BDS and NV. Likewise, *Cp4.1LG18g03150*, annotated as gibberellin regulated protein had variants both in BDS and NV, suggesting that the semi-trailing habitus could be related to mutated genes in the gibberellin pathway. In addition, variants in 6 genes related to the expansin family were identified in the same genotypes (*Cp4.1LG02g13490, Cp4.1LG04g14850, Cp4.1LG04g04730* and *Cp4.1LG15g06270*). Finally, two SNPs in the purine permease 3 and nitrate regulatory gene 2 protein-like (*Cp4.1LG11g04650* and *Cp4.1LG11g08000*), reported as related to growth habit traits [36, 37] were also observed in semi-trailing genotypes. Four different orthologues genes, described as involved in leaf shape, showed SNPs with modifier effect only in the medium lobed leaf genotypes (BDS, NV and SPQ). While the deeply lobed leaf genotypes OF, "Whitaker", SI, 381e and 968Rb showed the same allele as the reference genome MU-CU-16, classified as deeply incised leaves [38]. Variants in BDS, NV, and SPQ were observed in key genes associated with leaf development. Specifically, in the class I KNOTTED1-like homeobox proteins (Cp4.1LG02g06180), involved in regulation of the leaf shape in tomato [39], and in Cp4.1LG07g06690 and Cp4.1LG09g01450, Homeobox-leucine zipper proteins, transcription factors related to deeply lobed leaf structures [40]. Additionally, the AP2-like ethylene-responsive transcription factor ANT (Cp4.1LG11g01170), involved in palmately lobed leaf traits [41], exhibited variations (Table S4).

The eight varieties clearly showed phenotypic variability in fruit shape. The four Cocozelle varieties (SI, BDS, SPQ, OF) displayed variants with moderate or modifier effect in six genes encoding a WUSCHEL-

related homeobox like, involved in the LOCULE NUMBER (LC) trait [14, 42] (Table S4). Likewise, five genes annotated as calmodulin-binding protein showed SNPs with moderate modifier effect (Table S4). SNPs with modifier effect have also been observed, in these four accessions for genes involved in fruit morphology, such as receptor protein kinase CLAVATA1-like (*Cp4.1LG02g08230*) [43]; protein LONGIFOLIA 1-like (*Cp4.1LG04g14520* and *Cp4.1LG20g00830*) [13, 44]; OVATE FAMILY PROTEINS (OFPs) (*Cp4.1LG19g08040*) [45] and E3 ubiquitin-protein ligase RING1-like (*Cp4.1LG11g11400*) [12] (Table S4).

The fruit color is another important factor involved in fruit quality. Seven of the eight varieties have the same rind color, being BDS the only white variety, which allowed us to search for SNPs with for alternative allele respect the reference genome, (MU-CU-16, green rind). By filtering for SNPs with high and moderate effect, we found BDS variants in two genes related to the carotenoids pathway (*Cp4.1LG03g0478* and *Cp4.1LG05g11360*) and in two genes related to the chlorophyll pathway (*Cp4.1LG04g08850* and *Cp4.1LG13g11170*) [46]. In addition, three genes previously reported as up-regulated during the synthesis of carotenoids in *C. pepo* with orange and yellow peel [47] such as the transcription factor bHLH82-like (Cp4.1LG04g09000), the two-component response regulator-like APRR2 (*Cp4.1LG05g02070*) and the transcription termination factor MTERF4 (*Cp4.1LG07g08650*) showed variants in BDS. Furthermore, BDS variants with high or moderate effects were found in *Cp4.1LG05g01000; Cp4.1LG05g10310; Cp4.1LG14g02760* and *Cp4.1LG14g02300* (Table S4).

Variants in genes involved in biotic stress

Among the zucchini varieties analyzed in this study, there were two isogenic lines to TF, 968Rb and 381e, one resistant to powdery mildew (PM) and one tolerant to ZYMV, respectively, as well the variety SPQ reported as susceptible to both diseases [32]. In order to identify putative chromosomal regions harboring genes responsible for the different response to the fungus and virus, the variants distribution of the three genotypes was compared along the zucchini genome. A non-uniform pattern of SNP and IN/DEL distribution was observed for some chromosomes, especially for the chromosome 1, the chromosome 8 and the chromosome 10.

A higher variants density was identified in 968Rb, in a region ranging from 1.0 to 2.8 Mb at the beginning of the chromosome 10 (Fig. 5A). By analyzing in the detail this region, we found putative genes related to biotic stress affected by variants. Many genes with high or moderate effect were identified related to the receptor kinase, TF, cell wall metabolism etc. (Fig. 5B). Genes specifically affected in 968Rb included: a MAP3K5-like isoform X1 (Cp4.1LG10g02640) with an IN/DEL with disruptive inframe deletion, the dof zinc finger protein DOF1.4 (Cp4.1LG10g04940), the serine/threonine-protein kinase-like protein CCR1 (Cp4.1LG10g04780). In addition, a moderate effect IN/DEL influenced the 968Rb peroxidase (Cp4.1LG10g04730), located at the beginning of the chromosome 10 (at 1192530 Mb), implicated in defense to fungus. Among the cell wall related genes, a pectate lyase (Cp4.1LG10g04500) was altered by two IN/DELs with modifier effect. Interestingly, genes belonging to a cluster composed of TFs and resistance proteins were affected by variants (Fig. 5B, blue inset and table). In addition, 968Rb and SPQ showed the presence of variant with modifier effects in the susceptibility gene, MLO gene

(Cp4.1LG10g05200, CpeMLO11), but in a different location. The Cp4.1LG10g03270 WAT1-related protein, Cp4.1LG10g03390 and Cp4.1LG10g03520 showed IN/DELs with moderate effect in 968Rb and in SPQ but not in 381e. The receptor kinase Cp4.1LG10g01720 and the kinase Cp4.1LG10g12160 showed in the three genotypes moderate and modifier variations with different location.

A relevant difference between the variant distribution in 381e and SPQ and n 968Rb was also identified on the chromosome 8. For what concerned the SNPs, the region ranging from 4.0 to 4.2 Mb, and the region around the 6 Mb showed a higher density in 381e. On the other hand, SPQ showed a high density of SNPs around 3 Mb (Fig. 6A, 6B).

Among the genes affected by variations, some are putatively related to biotic stress being involved in cell wall metabolism, small RNA mediated resistance, transcription, regulation of transcription, RNA processing, chromatin remodeling, and translation repression (Fig. 6B). The gene Cp4.1LG08g00040, harboring the LG08SNP4 reported by Amoroso et al [28], was affected in both 381e and SPQ by mutations, as well as the gene Cp4.1LG08g01460, containing LG08SNP1 [28].

On the chromosome 1, two regions with SNP density differing among the genotypes were identified. A region around 17-17.7 Mb showed a higher number of variants in 381e while a region around 6-6.4 Mb had a higher number of variants in SPQ (Fig. 7). In the region around 17 Mb, (Fig. 7A, red inset, Fig. 7B, zoom of the red inset), were highlighted genes affected by mutations only in 381e such as the N-acetyltransferase domain-containing protein (Cp4.1LG01g24380), the Histone acetyltransferase (Cp4.1LG01g20150), the leukotriene A-4 hydrolase homolog (Cp4.1LG01g24170.1), the protein RNA-directed DNA methylation 3-like (Cp4.1LG01g24630) and RNA binding proteins (Cp4.1LG01g19740.1, Cp4.1LG01g20260). In addition, Cp4.1LG01g20160 harbored high effect mutations in 381e as well as modifier mutations in all genotypes (16 in 381e and 5 in SPQ and 10 in 968Rb). The LG01SNP1, associated to ZYMV resistance [28] was located in the gene Cp4.1LG01g18580, that showed 42 modifier mutations in 381e. Near this gene there were two RNA polymerases (Cp4.1LG01g18460 and Cp4.1LG01g19120) with moderate and high variants in 381e.

By focusing on the region around 6 Mb of the chromosome 1 (Fig. 7A, blue inset), only SPQ showed a SNP peak (Fig. 7C, zoom of blue inset), including a cluster of four cucumisin-like genes (Cp4.1LG01g10880, Cp4.1LG01g10970, Cp4.1LG01g11000 and Cp4.1LG01g10900) with a series of moderate and modifiers mutations only in this genotype. In this area, among the others, the Cytochrome P450 (Cp4.1LG01g11290) was affected by a high number of variations.

DISCUSSION

In the present work, a total number of 3,831,907 high quality SNPs and 387,243 high quality IN/DELs were obtained in an average of 40 million raw reads of eight *C. pepo* accessions, in accordance with number of variants found in Xanthopoulou et al. [26], which sequenced eight cultivars of *C. pepo*, with BGISEQ-500 platform identifying 3,823,977 high confidence SNPs using a similar pipeline. By contrast Martínez-González [27], sequenced a total of 95 Mexican *C. pepo* landraces from various subspecies (*C.*

pepo ssp. *pepo, C. pepo* ssp. *fraterna, C. pepo* ssp. *ovifera* var. *ovifera*, and commercial cultivars) using Ion Torrent technology, identifying approximatively 2000 SNPs.

Genetic variability

The genetic variability captured in our dataset was explored using different tools. The population structure analysis showed that the varieties analyzed are organized in 3–4 groups. Similar results were obtained in previous studies conducted on *C. pepo* traditional varieties collected in Guatemala and assessed by microsatellite markers [25]. Admixture analysis showed that the seven *C. pepo* cultivars were organized in four genetic clusters. Martínez-González [27] also reported that the genome structure of 88 accessions allowed to obtain from 2 to 4 clusters. Capuozzo et al. [32] used the Illumina GoldenGate for genotyping nine *C. pepo* accessions, observing, through the PCA, two distinct Zucchini and Cocozelle groups, with tight clustering of near-isogenic lines 968Rb and 381e. Formisano et al. [29], comparing 17 *C. pepo* subsp. *pepo* accessions using AFLP, obtained a phylogenetic tree with three main branches, including Zucchini cultivars cluster, a Cocozelle cluster and a hybrid cluster, including Zucchini and Cocozelle shapes. Our data, through PCA and phylogenetic analysis, also showed the four Cocozelle accessions (“Lungo bianco di Sicilia”, “Striato di Italia”, “Ortolano di Faenza”, “San Pasquale”) clustering together as well as the zucchini nearly isogenic lines 381e and 968Rb [33] and “Nano Verde di Milano”, consistently with previous studies [29, 32]. On the contrary, Formisano et al. [29] reported higher genetic similarity between SPQ and OF than between BDS and OF. The higher proximity observed between OF and BDS in our data may due to the different markers discriminatory power. The polymorphism identified by AFLP markers is related to combination of primers used. Instead, SNP markers are distributed across the genome and have a higher resolution ability [48].

Finally, “Whitaker” exhibits a complex pedigree, with four different species among its parents, including *Cucurbita ecuadorensis* (resistant to ZYMV and Papaya ringspot virus) [49, 50] *C. okeechobeensis* spp. *martinezii* (resistant to CMV, Tobacco ringspot virus and PM) [51, 52] and *C. moschata* [30, 53]. These multiple introgressions accounted for a significant number of SNPs when compared to the *C. pepo* reference genome and a large number of unique SNPs, likely derived from wild species.

The unique SNPs discovered in each variety could be the result both of natural genetic variation and specific breeding efforts that have led to genetic divergence among the varieties. Identifying exclusive SNPs provides valuable information about the genomic diversity within the zucchini population and can have implications for understanding trait variations, genome evolutionary history and genome-wide tracking. Specific variants in only one individual suggests possible association with specific traits. On the other hand, the variants affecting the range of 10 to 95% of individuals indicates moderate diversity, with potential implications for adaptability and phenotypic variability. SNPs common to 95% of individuals suggest more conserved genomic regions that could be related to key biological functions. The distribution of the identified variants along the chromosomes was homogenous consistent with [26] findings. However, discrepancies in the SNP counts among the eight different varieties, particularly in “Whitaker”, 381e, and 98Rb, were observed. Notably, “Whitaker” exhibited a higher number of SNPs,

which as reported above can be attributed to wild species introgressions [30]. Conversely, NV, 381e, and 968Rb displayed the lowest number of SNPs, likely attributable to their closer genomic resemblance to the reference genome MU-CU-16, as they all belong to the Zucchini group.

Variation in genes related to horticultural traits

Among genes with potential functional variants, we found variation in genes involved in pathways implicated in plant morphological traits, fruit shape and color among the eight genotypes analyzed. Plant architecture traits are important for improving agricultural efficiency and reducing the production costs [54]. Interestingly, *Cp4.1LG12g07890* (gibberellin 3-beta-dioxygenase 2-like gene) showed several variants in both semi-trailing genotypes, BDS and NV. This gene is orthologue to *CmoCh08G006170* (Gibberellin 3-beta hydroxylase), reported as candidate gene for the dwarf phenotype in *C. moschata* [35]. In addition, other genes involved in the gibberellin pathway such as *Cp4.1LG12g07890* and *Cp4.1LG18g03150*, were affected by variants in both BDS and NV. Furthermore, expansin genes associated with height in oil palm [55] and long stem internode in wheat [56, 57] harbored several variants (i.e. *Cp4.1LG02g13490*, *Cp4.1LG04g14850*, *Cp4.1LG04g04730* and *Cp4.1LG15g06270*) in semi-trailing genotypes. SNPs with moderate effect were also observed in purine permease 3 *(Cp4.1LG11g04650)* reported to be involved in plant height increasing [36]. It is important to point out that the nitrate regulatory gene2 protein-like (*Cp4.1LG11g08000)*, that promotes shoot organogenesis in cytokinin-dependent manner [58], was affected by variants in BDS and NV. The combinate action of cytokinin and nitrate reductase can affect the modulation and the coordination of cell cycle activity and cell wall integrity [59].

Leaf shape and size can affect sunlight penetration, photosynthetic supply of nutrients, air exchange, light energy absorption, water transport, disease occurrence and crop quality and yield [40]. It is worth to observe that three homeobox genes, previously reported as associated to leaf shape [39, 40], presented variants in the medium lobed leaf genotypes, BDS, NV and SPQ. *Cp4.1LG02g06180* belongs to the class I KNOTTED1-like homeobox (KNOX) proteins, involved in auxin efflux transporter and promoting leaflet initiation [39]. REVOLUTA-like (*Cp4.1LG07g06690*) and HAT5-like (*Cp4.1LG09g01450*) are homeobox-leucine zipper proteins, essential to obtain simple serrated leaves and deeply lobed leaf [40]. Finally, the AP2-like ethylene-responsive transcription factor ANT (*Cp4.1LG11g01170*) showed variation that could be involved in palmately lobed leaf (pll) [41].

Interestingly, the analyzed varieties showed variants in orthologous genes reported to be related to fruit shape in other cucurbit species. In Cocozelle varieties (BDS, SI, OF and SPQ) and “Whitaker”, variants affected *Cp4.1LG02g08230*, encoding receptor protein kinase CLAVATA1-like (CVL), involved in carpel number variation in melon [43]. Variants in Cocozelle varieties were also found in the WUSCHEL-related homeobox genes, such as in the tomato LOCULE NUMBER (LC) that influences fruit shape and size [14, 42, 60]. Some SNPs in Cocozelle varieties were found in genes belonging to protein families related to calmodulin-binding such as SUN gene, encoding a IQD protein [42, 61], responsible for elongated fruit shape in tomato and watermelon [62, 63]. Proper subcellular localization of the TRMs (Tubulin-Related Proteins) in conjunction with the TTP (Tubulin Tyrosine Ligase-Like Protein) is of utmost importance for

ensuring accurate cell division and appropriate deposition of cellular components [64]. TONNEAU1-recruiting motif (TRM), such as LONGIFOLIA1 (LNG1) and LONGIFOLIA2 (LNG2), were reported to regulate leaf morphology by positively promoting longitudinal polar cell elongation [44]. SNPs with modifier effect in both LNG1 and LNG2 orthologues (*Cp4.1LG20g00830* and *Cp4.1LG04g14520)* were found in SI, BDS, SPQ and OF. *Cp4.1LG19g08040* encoding a probable transcription repressor OFP9, belonging to OVATE FAMILY PROTEINS (OFPs) also showed variants in Cocozelle varieties. OVATE proteins are reported key regulator of fruit shape in several species [45, 65]. Finally, mutations found in Cocozelle varieties in Cp4.1LG11g11400, encoding E3 ubiquitin-protein ligase RING1-like can be involved the development of short fruits as reported in cucumber for the mutation in the SF1 (Short Fruit 1) gene, which encodes a cucurbit-specific RING-type E3 ligase [12].

Pigment accumulation in zucchini peel is primarily determined by the proportion of chlorophyll and carotenoid contents, the absence of carotenoids will result in white skin [5], while a high chlorophyll content would mask the content of lutein and β-carotene, resulting in green skin [5]. Three genes previously reported as co-expresed up-regulated during the synthesis of carotenoids or degradation of chlorophyll in *C. pepo* [47] were affect by variants in the light green genotype BDS as well as bHLH128 (Cp4.1LG04g09000), involved in AtPSY (Phytoene synthase) expression [66] and pigment accumulation in zucchini peel [47]. Variants in BDS were also discovered in the two-component response regulator-like APRR2 (*Cp4.1LG05g02070*), homologous to the candidates for light green genotype in zucchini stem *Cp4.1LG15g03420* and *Cp4.1LG15g03360* [67], and in a transcription termination factor MTERF4 (*Cp4.1LG07g08650*) which, in *Arabidopsis* leads to dysplasia and altered gene expression in chloroplasts and mitochondria when knock down [68]. In addition, BDS showed variants with high or moderate effects in: Geraniol 8-hydroxylase-like (*Cp4.1LG05g01000*), involved in the modification of unstable anthocyanidin and the formation of eriodyctiol [69]; Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (*Cp4.1LG05g10310*), that in *Nicotiana tabacum L*. in involved in plant size, Rubisco activity, leaf soluble protein and chlorophyll content [70]; C2H2-type domain-containing protein (*Cp4.1LG14g02760*) and 1-aminocyclopropane-1-carboxylate synthase-like (*Cp4.1LG14g02300*), associated to the biosyntethic pathway ethylene [71, 72] and involved chlorophyll degradation [73].

Genes related to biotic stress: a focus on powdery mildew and ZYMV

The comparison of variants distribution along the genomes of 381e, 968Rb and SPQ highlighted chromosomic regions that can affect genes involved in the response of zucchini to PM and ZYMV, in fact 381e is tolerant to ZYMV, 968Rb is tolerant to PM and SPQ is susceptible to both [32].

The observation of the private SNP distribution across 968Rb chromosomes highlighted a notable prevalence of unique SNPs on chromosome 10. Among these genes, three (Cp4.1LG10g02780, Cp4.1LG10g02800 and Cp4.1LG10g02750) contained RPW8 domains, one (Cp4.1LG10g02750) also contained a LRR domain and an NB-ARC domain, and the last (Cp4.1LG10g02760) contained a MYB domain. It is worth to note that the activation of Cp4.1LG10g02780 after PM infection was higher in

resistant individuals [74]. In addition, only in 968Rb we found variants affecting the gene floral homeotic protein APETALA 2-like isoform X1 (Cp4.1LG10g02330) similar to the *Cucurbita moschata* gene *CmoCh3G009850* proved to be involved in the resistance to PM in pumpkin. The pectate lyase Cp4.1LG10g04500 with variants in 968Rb in turn can affect plant cell wall thickening and PM resistance [75] as reported in Arabidopsis [76, 77]. Finally, our data outlined two WAT1-related proteins (Cp4.1LG10g03270 and Cp4.1LG10g03390), affected by variants that can provide broad spectrum resistance [78, 79].

The Mildew Locus O (MLO) gene (Cp4.1LG10g05200), was affected by one modifier mutation with different position in both 968Rb and SPQ. Moreover, high effect variants were identified in 968Rb on Cp4.1LG14g10200 and Cp4.1LG19g08380 orthologues to melon genes with proved role in the PM response [80].

Notably, the analysis of 381e chromosome 8 revealed some genes affected by variants previously identified as correlated to ZYMV resistance such as the gene Cp4.1LG08g00040, harboring LG08SNP4, and Cp4.1LG08g01460, harboring the LG08SNP1 [28] and the RNA helicases of the DEAD box family gene (Cp4.1LG08g08000), reported by Capuozzo et al. [32]. The last gene encodes for an important player of RNA metabolism and chromatin remodeling [81–83]. RNA-based regulation has emerged as a critical layer of control in plant immunity also in the case of virus infection [84]. Interestingly, other important RNA metabolism actors were affected by variants only in 381e, such as the guard cell S-type anion channel SLAC1 (Cp4.1LG08g07810), RNA polymerase II transcription factor B subunit 5 (Cp4.1LG08g07700) histone deacetylase complex subunit SAP18-like (Cp4.1LG08g07910) and a thaumatin-like protein (Cp4.1LG08g07930). Among the others, it is worth to point out, that the Dicer (Cp4.1LG08g00150), involved in the RNAi-mediated antiviral immunity [84] and the transducin beta-like protein 2 (Cp4.1LG08g01940) having a potential role in virus translation repression [85], were affected by mutation in 381. On the same chromosome region, the NBS-LRR gene (Cp4.1LG08g07980), resulted affect by 34 modifiers in 381e, 7 in SPQ and 4 in 968Rb.

A different pattern of variant density was also highlighted on the chromosome 1. By starting from the Cp4.1LG01g18580, on which was found the SNP (LG01SNP1), associated to the ZYMV resistance [28], additional specific 381e modifier mutations were identified. This gene belongs to THO/TREX complex subunit proteins that have a role in small interfering RNA-dependent processes in plants [86]. Near this gene were located two RNA-dependent RNA polymerases (Cp4.1LG01g18460 and Cp4.1LG01g19120) that only in 381e showed high and moderate variants. It is worth to note that the Cp4.1LG01g18460 ortholog in *A*rabidopsis, (AT1G14790; AtRDR1) is involved in the turnover of viral RNAs in infected plants [87]. The presence of specific variants in a N-acetyltransferase domain-containing protein (Cp4.1LG01g24380) involved in in chromatin remodeling [88] were also of interest. In addition, an high effect mutation was observed in Cp4.1LG01g20160, a structural protein of ribosome involved in the regulation of defense response that could repress the virus translation [82]. By contrast, the susceptible SPQ showed specific mutations in the Cytochrome P450 gene (Cp4.1LG01g11290) involved in disease

response [89] and in cucumisins, subtilisin-like serine proteases (subtilases) genes, that regulated the hypersensitive response, pathogen recognition, priming and peptide hormone release [90].

CONCLUSION

The present study underscores the importance of genetic variations discovery for improving crop traits. The variants exhibited by the different varieties can contribute to better explore the genetic variability for plant architecture, fruit quality, and disease resistance, particularly against Zucchini yellow mosaic virus (ZYMV) and powdery mildew (PM). The identification of specific genetic variants in genes related to disease resistance and plant architecture holds substantial implications for crop improvement and trait selection in plant breeding. Our work improves knowledge of the potential genes underlining morphological traits and plant immune responses, promising applications in agriculture and food security. These findings can provide valuable insights into crop management strategies thank to the development of more resilient zucchini varieties.

MATERIAL AND METHODS

Plant material and DNA extraction

Seeds of eight zucchini varieties, four belonging to Zucchini group (“Nano Verde di Milano”, NV; “Whitaker”; 381e and 968Rb) and four belonging to Cocozelle group (“Lungo Bianco di Sicilia”, BDS; “San Pasquale”, SPQ; “Ortolano di Faenza”, OF; “Striato di Italia”, SI), kindly provided by Semiorto Sementi s.r.l. (Sarno, Italy), were sown at UNINA greenhouse.

Leaf DNA was extracted from each variety, using the DNeasy Plant Mini kit (QIAGEN, Valencia, CA). The quality and the quantity of DNA was checked by agarose gel and Nanodrop spectrophotometer (Thermo Fisher Scientific, Wilmington, USA).

Library construction and Sequencing

Indexed libraries were prepared from 1 µg purified DNA with the TruSeq DNA Sample Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Libraries were quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies) and pooled such that each index-tagged sample was present in equimolar amounts, with 2 nM final concentration. Pooled samples were subject to cluster generation and sequencing using the Illumina HiSeq 1500 platform (Illumina Inc, San Diego, CA, USA) at the Genomix life (Salerno, Italy) to generate 100-bp paired-end reads.

Reads processing, mapping and variants calling

The raw sequencing reads were pre-processed by removing sequencing adapters, poor-quality bases (quality < 3) from the beginning and end of the reads; low quality bases (quality < 20 in a 4-bp window) as well as reads shorter than 36 bp using Trimmomatic [91]. FastQC was used to check the quality before and after the trimming [92]. The filtered, high-quality paired reads were aligned to the last version of the

C. pepo genome v.4.1 available at cucurbitgenomics.org [24, 46] with BWA mem [93]. The resulting SAM files were converted into BAM files and sorted with SAMtools [94].

SAMtools flagstats was used to determinate the number of total reads, secondary reads, and mapped reads [95] and SAMtools view was used to remove unmapped reads [96]. Duplicated reads were marked by Picard [97, 98].

Variants were called by Bcftools mpileup [99, 100], a minimum quality mapping of 20 and minimum base quality of 20 was required. These raw variants file was additionally filtered with a minor allele frequency > 0.05, quality > 20 and read deep > 5. Density of variants was obtained by plotting read depth (DP) of all the variants across the chromosome using R package “ggplot2” [101]. Bcftools view was used to filter private SNPs, as well as SNPs affecting individuals ranging from 10–95%, and those affecting 95% of the individuals[99].

Population structure and genetic variability

To investigate the population structure, the analyses were conducted following the procedure described in Flores-León et al. [102]. Admixture analysis was performed on the eight accessions using admixture-linux-1.3.0 [103] with default parameters in an unsupervised mode of K = 1 to 10. The cross-validation error for each K was calculated with the -cv option, K = 4 and k = 3. “ggplot2” R package was used to plot the Cross-validation error K graph. A principal component analysis (PCA) was performed using the “snpgdsPCA” function of the “SNPRelate” R package [104] and 3D visualization of the data was performed by CurlyWhirly Software for Windows. SNPcalling result was transformed to fasta format USING tool vcf-to-tab and Perl script vcf_tab_to_fasta_alignment.pl [105]. The resulting data was aligned and analyzed to construct a tree using MEGA11 [106]. The Kimura two-parameter model was employed for the analysis, and bootstrap testing was conducted to assess the confidence of each node. Linkage disequilibrium (r2) was calculated for the eight accessions, as well as for each group separately using PopLDdecay and drawing by Perl script [107].

SNP and IN/DEL functional annotation

The effect of the SNPs and IN/DEL was analyzed using SnpEff program [108]. The variants predicted effects were categorized by impact, as high (disruptive impact on the protein), moderate (non-synonymous substitution), low (synonymous substitution) or modifier (with impact on non-coding regions) [109]. Gene Ontology (GO) annotation available in cucurbitgenomics.org [46] was used to characterize SNPs with putative effects predicted to be moderate or high, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) available in cucurbitgenomics.org [46] was used to identify functions in metabolic pathways in the candidate genes that we found. MAPMAN software [110] was used to identify genes affected by variants related to biotic stress response.

Abbreviations

BAM

Binary Alignment/Map
BDS
Lungo Bianco di Sicilia
BWA
Burrows-Wheeler Aligner
CMV
Cucumber mosaic virus
FAOSTAT
Food and Agriculture Organization of the United Nations statistics
GO
Gene Ontology
IN/DEL
Insertions/Deletions
KEGG
Kyoto Encyclopedia of Genes and Genomes
LC
Locule Number
LD
Linkage Disequilibrium
MEGA
Molecular Evolutionary Genetics Analysis
NV
Nano verde di Milano
OF
Ortolano di Faenza
PM
powdery mildew
RNAi
RNA Interference
SAM
Sequence Alignment/Map
SI
Striato d'Italia
SNP
Single nucleotide polymorphism
SPQ
San Pasquale
TF
Transcription Factor
ToLCNDV

Tomato leaf curl New Delhi virus
UPGMA
Unweighted Pair Group Method with Arithmetic Mean
ZYMV
Zucchini yellow mosaic virus

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials.

All the data supporting our findings are contained within the manuscript, in text, tables and figures and in the supplementary files. Moreover, the sequence variants have been deposited in the European Variation Archive (EVA) at EMBL-EBI under accession number PRJEB74211 (https://www.ebi.ac.uk/eva/?eva-study=PRJEB74211)

The authors declare that they have no competing interests

Funding

This research was funded by the Italian Ministry of University and Research (GenHORT project). CPM is a recipient of the predoctoral fellowship “Programa de Ayudas de Investigación y Desarrollo (PAID-01-19) de la Universitat Politècnica de València (FPI-UPV)” and mobility grant “Programa para la Formación de Personal investigador (FPI) de la Universitat Politècnica de València (UPV)”.

Authors´contributions

Design of the study: MRE. Genome mapping and data analysis: CPM and DDE with help of AG. Data interpretation, writing: CPM, DDE, MRE. Reviewing APDC. All authors contributed to the article and approved the submitted version. All authors have read and agreed to the published version of the manuscript.

Acknowledgment

We would like to thank La Semiorto Sementi S.r.l. for plant material and Flores-León, A. for his help in the development of the genetic variability script.

References

1. Paris HS. History of the Cultivar-Groups of Cucurbita pepo. In: Horticultural Reviews. 2001. pp. 71–8.
2. FAO. FAOSTAT Database. Food and Agriculture Organization of the United Nations. 2021. http://faostat3.fao.org. Accessed 20 Jan 2023.
3. Smith BD. Eastern North America as an independent center of plant domestication. Proc Natl Acad Sci. 2006;103:12223–8.
4. Lim TK. Cucurbita pepo. Edible Medicinal And Non-Medicinal Plants. Dordrecht: Springer; 2012. pp. 281–94.
5. Obrero Á, González-Verdejo CI, Die JV, Gómez P, Del Río-Celestino M, Román B. Carotenogenic gene expression and carotenoid accumulation in three varieties of Cucurbita pepo during fruit development. J Agric Food Chem. 2013;61:6393–403.
6. Lust TA, Paris HS. Italian horticultural and culinary records of summer squash (Cucurbita pepo, Cucurbitaceae) and emergence of the zucchini in 19th-century Milan. Ann Bot. 2016;118:53–69.
7. Formisano G, Roig C, Esteras C, Ercolano MR, Nuez F, Monforte AJ, et al. Genetic diversity of Spanish Cucurbita pepo landraces: an unexploited resource for summer squash breeding. Genet Resour Crop Evol. 2012;59:1169–84.
8. Davière JM, Achard P. Gibberellin signaling in plants. Development. 2013;140:1147–51.
9. Marowa P, Ding A, Kong Y. Expansins: roles in plant growth and potential applications in crop improvement. Plant Cell Rep. 2016;35:949–65.
10. van der Knaap E, Chakrabarti M, Chu YH, Clevenger JP, Illa-Berenguer E, Huang Z, et al. What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. Front Plant Sci. 2014;5:227.
11. Pan Y, Liang X, Gao M, Liu H, Meng H, Weng Y, et al. Round fruit shape in WI7239 cucumber is controlled by two interacting quantitative trait loci with one putatively encoding a tomato SUN homolog. Theor Appl Genet. 2017;130:573–86.
12. Xin T, Zhang Z, Li S, Zhang S, Li Q, Zhang ZH, et al. Genetic Regulation of Ethylene Dosage for Cucumber Fruit Elongation. Plant Cell. 2019;31:1063–76.
13. Zhou Y, Miao J, Gu H, Peng X, Leburu M, Yuan F, et al. Natural Variations in SLG7 Regulate Grain Shape in Rice. Genetics. 2015;201:1591–9.
14. Muños S, Ranc N, Botton E, Bérard A, Rolland S, Duffé P, et al. Increase in Tomato Locule Number Is Controlled by Two Single-Nucleotide Polymorphisms Located Near WUSCHEL. Plant Physiol. 2011;156:2244–54.
15. Cheng F, Song M, Zhang M, Cheng C, Chen J, Lou Q. A SNP mutation in the CsCLAVATA1 leads to pleiotropic variation in plant architecture and fruit morphogenesis in cucumber (Cucumis sativus L). Plant Sci. 2022;323:111397.
16. Paris HS. Germplasm enhancement of Cucurbita pepo (pumpkin, squash, gourd: Cucurbitaceae): progress and challenges. Euphytica. 2016;208:415–38.

17. Lisa V. Characterization of a Potyvirus That Causes Zucchini Yellow Mosaic. Phytopathology. 1981;71:667.
18. Zechmann B, Zellnig G, Urbanek-Krajnc A, Müller M. Artificial elevation of glutathione affects symptom development in ZYMV-infected Cucurbita pepo L. plants. Arch Virol. 2007;152:747–62.
19. LECOQ H, PITRAT M. Identification et caractérisation d'un potyvirus provoquant la maladie du rabougrissement jaune du melon. Agronomie. 1981;1:827–34.
20. Zhang S, Liu J, Xu B, Zhou J. Differential Responses of Cucurbita pepo to Podosphaera xanthii Reveal the Mechanism of Powdery Mildew Disease Resistance in Pumpkin. Front Plant Sci. 2021;12:633221.
21. Pirondi A, Kitner M, Iotti M, Sedláková B, Lebeda A, Collina M. Genetic structure and phylogeny of Italian and Czech populations of the cucurbit powdery mildew fungus Golovinomyces orontii inferred by multilocus sequence typing. Plant Pathol. 2016;65:959–67.
22. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, et al. The genome of the cucumber, Cucumis sativus L. Nat Genet. 2009;41:1275–81.
23. Navot N, Zamir D. Isozyme and seed protein phylogeny of the genus Citrullus (Cucurbitaceae). Plant Syst Evol. 1987;156:61–7.
24. Montero-Pau J, Blanca J, Bombarely A, Ziarsolo P, Esteras C, Martí-Gómez C, et al. De novo assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the Cucurbita genus. Plant Biotechnol J. 2018;16:1161–71.
25. Enríquez E, Landaverde-González P, Lima-Cordón R, Solórzano-Ortíz E, Tapia-López R, Nuñez-Farfán J. Population genetics of traditional landraces of Cucurbita pepo L., 1753 in the cloud forest in Baja Verapaz, Guatemala. Genet Resour Crop Evol. 2018;65:979–91.
26. Xanthopoulou A, Montero-Pau J, Mellidou I, Kissoudis C, Blanca J, Picó B, et al. Whole-genome resequencing of Cucurbita pepo morphotypes to discover genomic variants associated with morphology and horticulturally valuable traits. Hortic Res. 2019;6:94.
27. Martínez-González C, Castellanos-Morales G, Barrera-Redondo J, Sánchez-de la Vega G, Hernández-Rosales HS, Gasca-Pineda J, et al. Recent and Historical Gene Flow in Cultivars, Landraces, and a Wild Taxon of Cucurbita pepo in Mexico. Front Ecol Evol. 2021;9:656051.
28. Amoroso CG, Andolfo G, Capuozzo C, Di Donato A, Martinez C, Tomassoli L, et al. Transcriptomic and genomic analysis provides new insights in molecular and genetic processes involved in zucchini ZYMV tolerance. BMC Genomics. 2022;23:371.
29. Formisano G, Paris HS, Frusciante L, Ercolano MR. Commercial Cucurbita pepo squash hybrids carrying disease resistance introgressed from Cucurbita moschata have high genetic similarity. Plant Genet Resour. 2010;8:198–203.
30. Robinson RW, Reiners S. Parthenocarpy in summer squash. HortScience. 1999;34:715–7.
31. Paris HS, Cohen S. Oligogenic inheritance for resistance to Zucchini yellow mosaic virus in Cucurbita pepo. Ann Appl Biol. 2000;136:209–14.

32. Capuozzo C, Formisano G, Iovieno P, Andolfo G, Tomassoli L, Barbella MM et al. Inheritance analysis and identification of SNP markers associated with ZYMV resistance in Cucurbita pepo. Mol Breed. 2017;37.
33. Cohen R, Hanan A, Paris HS. Single-gene resistance to powdery mildew in zucchini squash (Cucurbita pepo). Euphytica. 2003;130:433–41.
34. Ding W, Wang Y, Qi C, Luo Y, Wang C, Xu W, et al. Fine mapping identified the gibberellin 2-oxidase gene CpDw leading to a dwarf phenotype in squash (Cucurbita pepo L). Plant Sci. 2021;306:110857.
35. Min Z, Hu X, Han X, Li Y, Li J, Wang D, et al. Genetic Mapping and Identification of the Gibberellin 3-Oxidase Gene GA3ox Leading to a GA-Deficient Dwarf Phenotype in Pumpkin (Cucurbita moschata D). Agronomy. 2022;12:1779.
36. Qi Z, Xiong L. Characterization of a Purine Permease Family Gene OsPUP7 Involved in Growth and Development Control in Rice. J Integr Plant Biol. 2013;55:1119–35.
37. Crawford NM. Nitrate: Nutrient and signal for plant growth. Plant Cell. 1995;7.
38. Montero-Pau J, Blanca J, Esteras C, Martínez-Pérez EM, Gómez P, Monforte AJ, et al. An SNP-based saturated genetic map and QTL analysis of fruit-related traits in Zucchini using Genotyping-by-sequencing. BMC Genomics. 2017;18:94.
39. Wei C, Chen X, Wang Z, Liu Q, Li H, Zhang Y, et al. Genetic mapping of the LOBED LEAF 1 (ClLL1) gene to a 127.6-kb region in watermelon (Citrullus lanatus L). PLoS ONE. 2017;12:e0180741.
40. Bo K, Duan Y, Qiu X, Zhang M, Shu Q, Sun Y, et al. Promoter variation in a homeobox gene, CpDll, is associated with deeply lobed leaf in Cucurbita pepo L. Theor Appl Genet. 2022;135:1223–34.
41. Gao X, Ning X, Wang Y, Wang X, Yan W, Zhang Z, et al. Fine mapping of a gene that confers palmately lobed leaf (pll) in melon (Cucumis melo L). Euphytica. 2014;200:337–47.
42. Pan Y, Wang Y, McGregor C, Liu S, Luan F, Gao M, et al. Genetic architecture of fruit size and shape variation in cucurbits: a comparative perspective. Theor Appl Genet. 2020;133:1–21.
43. Liu S, Gao P, Zhu Q, Zhu Z, Liu H, Wang X, et al. Resequencing of 297 melon accessions reveals the genomic history of improvement and loci related to fruit traits in melon. Plant Biotechnol J. 2020;18:2545–58.
44. Lee YK, Kim GT, Kim IJ, Park J, Kwak SS, Choi G, et al. LONGIFOLIA1 and LONGIFOLIA2, two homologous genes,regulate longitudinal cell elongation in Arabidopsis. Development. 2006;133:4305–14.
45. Wang S, Chang Y, Ellis B. Overview of OVATE FAMILY PROTEINS, A Novel Class of Plant-Specific Growth Regulators. Front Plant Sci. 2016;7:417.
46. Yu J, Wu S, Sun H, Wang X, Tang X, Guo S, et al. CuGenDBv2: an updated database for cucurbit genomics. Nucleic Acids Res. 2023;51:1457–64.
47. Xu X, Lu X, Tang Z, Zhang X, Lei F, Hou L, et al. Combined analysis of carotenoid metabolites and the transcriptome to reveal the molecular mechanism underlying fruit colouration in zucchini (Cucurbita pepo L). Food Chem Mol Sci. 2021;2:100021.

48. Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S. SNP Markers and Their Impact on Plant Breeding. Int J Plant Genomics. 2012;2012:1–11.
49. Robinson RW, Weeden NF, Provvidenti R. Inheritance of Resistance to Zucchini Yellow Mosaic Virus in the Interspecific Cross Cucurbita maxima x C. ecuadorensis. Cucurbit Genet Coop. 1988;11:74–5.
50. Herrington M, Byth D, Teakle D, Brown P. Inheritance of resistance to papaya ringspot virus type W in hybrids between Cucurbita ecuadorensis and C. maxima. Aust J Exp Agric. 1989;29:253.
51. Martín-Hernández AM, Picó B. Natural Resistances to Viruses in Cucurbits. Agronomy. 2020;11:23.
52. Kates HR. Pumpkins, squashes, and gourds (Cucurbita L.) of North America. North Am Crop Wild Relat Important Species. 2019;2:195–224.
53. Cornell Research Foundation. Whitaker. 1998;:1–18.
54. Xiang C, Duan Y, Li H, Ma W, Huang S, Sui X, et al. A High-Density EST-SSR-Based Genetic Map and QTL Analysis of Dwarf Trait in Cucurbita pepo L. Int J Mol Sci. 2018;19:3140.
55. Somyong S, Phetchawang P, Bihi AK, Sonthirod C, Kongkachana W, Sangsrakru D, et al. A SNP variation in an expansin (EgExp4) gene affects height in oil palm. PeerJ. 2022;10:e13046.
56. Xing SC, Li F, Guo QF, Liu DR, Zhao XX, Wang W. The involvement of an expansin geneTaEXPB23 from wheat in regulating plant cell growth. Biol Plant. 2009;53:429–34.
57. Zhang Y, Ni Z, Yao Y, Nie X, Sun Q. Gibberellins and heterosis of plant height in wheat (Triticum aestivum L). BMC Genet. 2007;8:40.
58. Landrein B, Formosa-Jordan P, Malivert A, Schuster C, Melnyk CW, Yang W, et al. Nitrate modulates stem cell dynamics in Arabidopsis shoot meristems through cytokinins. Proc Natl Acad Sci. 2018;115:1382–7.
59. Gigli-Bisceglia N, Engelsdorf T, Strnad M, Vaahtera L, Khan GA, Jamoune A, et al. Cell wall integrity modulates Arabidopsis thaliana cell cycle gene expression in a cytokinin- and nitrate reductase-dependent manner. Development. 2018;145:dev166678.
60. Niu H, Liu X, Tong C, Wang H, Li S, Lu L, et al. The WUSCHEL-related homeobox1 gene of cucumber regulates reproductive organ development. J Exp Bot. 2018;69:5373–87.
61. Wu S, Xiao H, Cabrera A, Meulia T, van der Knaap E. SUN Regulates Vegetative and Reproductive Organ Shape by Changing Cell Division Patterns. Plant Physiol. 2011;157:1175–86.
62. Kaźmińska K, Hallmann E, Korzeniewska A, Niemirowicz-Szczytt K, Bartoszewski G. Identification of Fruit-Associated QTLs in Winter Squash (Cucurbita maxima Duchesne) Using Recombinant Inbred Lines. Genes (Basel). 2020;11:419.
63. Legendre R, Kuzy J, McGregor C. Markers for selection of three alleles of ClSUN25-26-27a (Cla011257) associated with fruit shape in watermelon. Mol Breed. 2020;40:19.
64. Snouffer A, Kraus C, van der Knaap E. The shape of things to come: ovate family proteins regulate plant organ shape. Curr Opin Plant Biol. 2020;53:98–105.
65. Tsaballa A, Pasentsis K, Darzentas N, Tsaftaris AS. Multiple evidence for the role of an Ovate-like gene in determining fruit shape in pepper. BMC Plant Biol. 2011;11:1–16.

66. Stanley L, Yuan Y-W. Transcriptional Regulation of Carotenoid Biosynthesis in Plants: So Many Regulators, So Little Consensus. Front Plant Sci. 2019;10:1017.
67. Zhu L, Wang Y, Zhang Z, Hu D, Wang Z, Hu J, et al. A novel chromosomal fragment deletion in the APRR2 repeated locus modulates dark sterm color in Cucurbita pepo. Theor Appl Genet. 2022;135:4277–88.
68. Tang B, Xie L, Yi T, Lv J, Yang H, Cheng X, et al. Genome-Wide Identification and Characterization of the Mitochondrial Transcription Termination Factors (mTERFs) in Capsicum annuum L. Int J Mol Sci. 2019;21:269.
69. Yang W, Xu H, Xiao Q, Li X, Shao Q. Combined analysis of metabolome and transcriptome provides insights into metabolisms of chlorophylls, carotenoids, and flavonoids in the yellowing leaves of 'HAES344' macadamia. Sci Hortic (Amsterdam). 2023;308:111600.
70. Price GD, Evans JR, von Caemmerer S, Yu JW, Badger MR. Specific reduction of chloroplast glyceraldehyde-3-phosphate dehydrogenase activity by antisense RNA reduces CO2 assimilation via a reduction in ribulose bisphosphate regeneration in transgenic tobacco plants. Planta. 1995;195:369–78.
71. Han Ychao, Fu C, chun, Kuang J, fei, Chen J ye, Lu W. jin. Two banana fruit ripening-related C2H2 zinc finger proteins are transcriptional repressors of ethylene biosynthetic genes. Postharvest Biol Technol. 2016;116:8–15.
72. Dong T, Hu Z, Deng L, Wang Y, Zhu M, Zhang J, et al. A Tomato MADS-Box Transcription Factor, SlMADS1, Acts as a Negative Regulator of Fruit Ripening. Plant Physiol. 2013;163:1026–36.
73. de Araújo FF, da Costa LC, da Silva TP, Puiatti M, Finger FL. Action of ethylene on postharvest of summer squash Menina Brasileira. Rev Ceres. 2017;64:360–7.
74. Wang Y, Qi C, Luo Y, Zhang F, Dai Z, Li M, et al. Identification and mapping of CpPM10.1, a major gene involved in powdery mildew (race 2 France of Podosphaera xanthii) resistance in zucchini (Cucurbita pepo L). Theor Appl Genet. 2021;134:2531–45.
75. Nie J, Wang Y, He H, Guo C, Zhu W, Pan J, et al. Loss-of-Function Mutations in CsMLO1 Confer Durable Powdery Mildew Resistance in Cucumber (Cucumis sativus L). Front Plant Sci. 2015;6:1155.
76. Vogel JP, Raab TK, Schiff C, Somerville SC. PMR6, a pectate lyase-like gene required for powdery mildew susceptibility in Arabidopsis. Plant Cell. 2002;14.
77. Vogel JP, Raab TK, Somerville CR, Somerville SC. Mutations in PMR5 result in powdery mildew resistance and altered cell wall composition. Plant J. 2004;40:968–78.
78. Denancé N, Ranocha P, Oria N, Barlet X, Rivière M, Yadeta KA, et al. Arabidopsis wat1 (walls are thin1)-mediated resistance to the bacterial vascular pathogen, Ralstonia solanacearum, is accompanied by cross-regulation of salicylic acid and tryptophan metabolism. Plant J. 2013;73:225–39.
79. Hanika K, Schipper D, Chinnappa S, Oortwijn M, Schouten HJ, Thomma BPHJ, et al. Impairment of Tomato WAT1 Enhances Resistance to Vascular Wilt Fungi Despite Severe Growth Defects. Front Plant Sci. 2021;12:721674.

80. Zhu L, Li Y, Li J, Wang Y, Zhang Z, Wang Y, et al. Genome-wide identification and analysis of the MLO gene families in three Cucurbita species. Czech J Genet Plant Breed. 2021;57:119–23.
81. Linder P, Jankowsky E. From unwinding to clamping — the DEAD box RNA helicase family. Nat Rev Mol Cell Biol. 2011;12:505–16.
82. Sanfaçon H. Plant Translation Factors and Virus Resistance. Viruses. 2015;7:3392–419.
83. Jia R, Lin J, You J, Li S, Shan G, Huang C. The DEAD-box helicase Hlc regulates basal transcription and chromatin opening of stress-responsive genes. Nucleic Acids Res. 2022;50.
84. Leonetti P, Stuttmann J, Pantaleo V. Regulation of plant antiviral defense genes via host RNA-silencing mechanisms. Virol J. 2021;18:194.
85. Vermeulen A, Takken FLW, Sánchez-Camargo VA. Translation Arrest: A Key Player in Plant Antiviral Response. Genes (Basel). 2023;14:1293.
86. Francisco-Mangilet AG, Karlsson P, Kim M, Eo HJ, Oh SA, Kim JH et al. THO 2, a core member of the < scp > THO / TREX complex, is required for micro < scp > RNA production in Arabidopsis. Plant J. 2015;82:1018–29.
87. Mosharaf MP, Rahman H, Ahsan MA, Akond Z, Ahmed FF, Islam MM, et al. In silico identification and characterization of AGO, DCL and RDR gene families and their associated regulatory elements in sweet orange (Citrus sinensis L). PLoS ONE. 2020;15:e0228233.
88. Vlachonasios K, Poulios S, Mougiou N. The Histone Acetyltransferase GCN5 and the Associated Coactivators ADA2: From Evolution of the SAGA Complex to the Biological Roles in Plants. Plants. 2021;10:308.
89. Pandian BA, Sathishraj R, Djanaguiraman M, Prasad PVV, Jugulam M. Role of Cytochrome P450 Enzymes in Plant Stress Response. Antioxidants. 2020;9:454.
90. Figueiredo J, Sousa Silva M, Figueiredo A. Subtilisin-like proteases in plant defence: the past, the present and beyond. Mol Plant Pathol. 2018;19:1017–28.
91. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
92. Simon Andrews. FastQC A Quality Control tool for High Throughput Sequence Data. Babraham Bioinformatics. 2020;5. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 20 Sep 2022.
93. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013.
94. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10.
95. Robinson KM, Hawkins AS, Santana-Cruz I, Adkins RS, Shetty AC, Nagaraj S et al. Aligner optimization increases accuracy and decreases compute times in multi-species sequence data. Microb Genomics. 2017;3.
96. Altmann A, Weber P, Bader D, Preuß M, Binder EB. Mü Ller-Myhsok B. A beginners guide to SNP calling from high-throughput DNA-sequencing data. Springer. 2012;131:1541–54.

97. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr Protoc Bioinforma. 2013;43:11–10.
98. Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, et al. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Front Genet. 2015;6:235.
99. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–93.
100. Yao Z, You FM, N'Diaye A, Knox RE, McCartney C, Hiebert CW, et al. Evaluation of variant calling tools for large plant genome re-sequencing. BMC Bioinformatics. 2020;21:360.
101. Wickham H. Data Analysis. ggplot2. Use R! Cham: Springer; 2016. pp. 189–201.
102. Flores-León A, Peréz Moro C, Martí R, Beltran J, Roselló S, Cebolla-Cornejo J, et al. Spanish Melon Landraces: Revealing Useful Diversity by Genomic, Morphological, and Metabolomic Analysis. Int J Mol Sci. 2022;23:7162.
103. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.
104. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28:3326–8.
105. Chen J, vcf-tab-to-fasta. Github. 2014. https://github.com/JinfengChen/vcf-tab-to-fasta. Accessed 9 Nov 2023.
106. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38:3022–7.
107. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35:1786–8.
108. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff Fly (Austin). 2012;6:80–92.
109. Valdisser PAMR, Pereira WJ, Almeida Filho JE, Müller BSF, Coelho GRC, de Menezes IPP, et al. In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. BMC Genomics. 2017;18:423.
110. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, et al. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 2004;37:914–39.

Figures

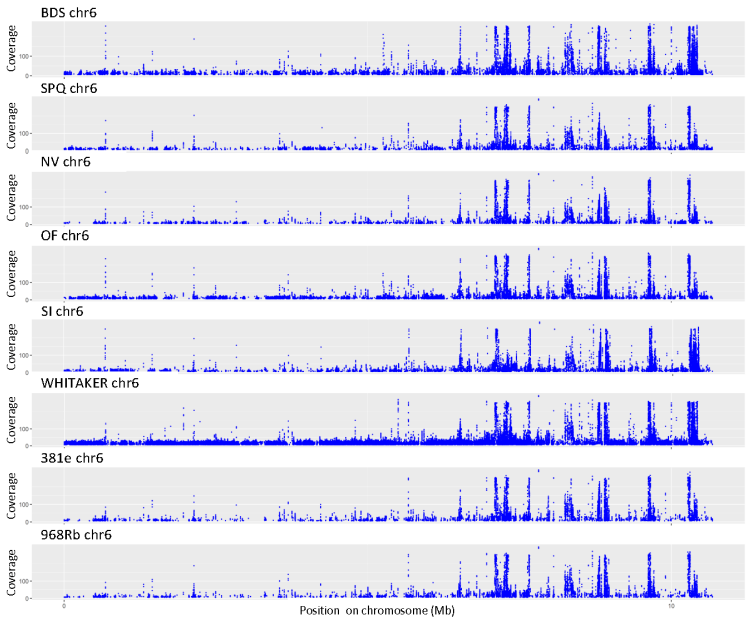


Figure 1

Single-nucleotide polymorphisms (SNPs) coverage and density along the chromosome VI in eight accessions of *Cucurbita pepo*. Up to down, BDS (“Lungo Bianco di Sicilia”), SPQ (“San Pasquale”), NV (“Nano Verde di Milano”), OF (“Ortolano di Faenza”), SI (“Striato di Italia”), “Whitaker”, 381e, 968Rb.

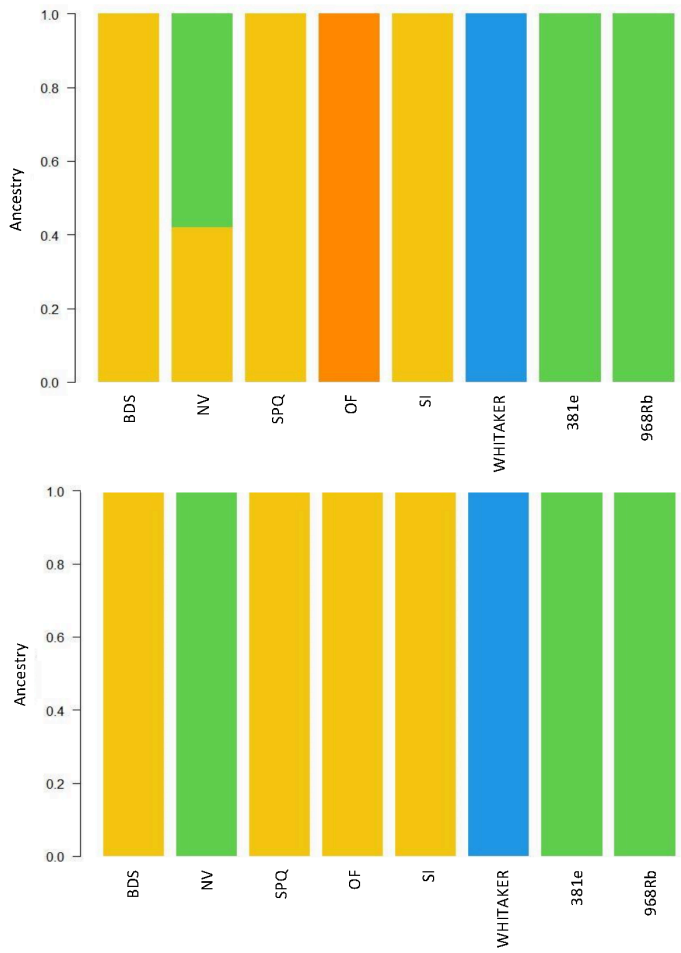


Figure 2

Population structure analysis of eight accessions of Cucurbita pepo: BDS, NV, SPQ, OF, SI, “Whitaker”, 381e, 968Rb. A) Probability of individual assignment to each of genetic groups (K = 4, indicated by colors). B) Probability of individual assignment to each of genetic groups (K = 3, indicated by colors).

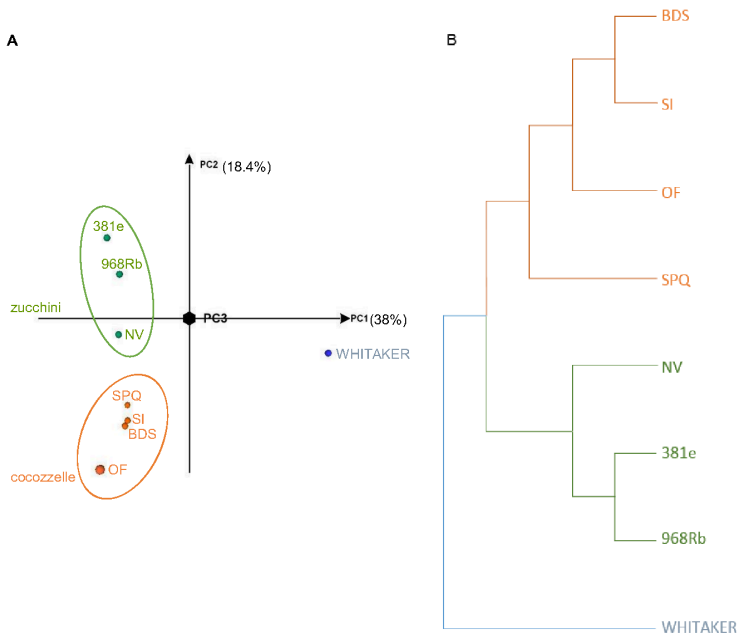


Figure 3

Assessment of genomic relationships of eight accessions of Cucurbita pepo: BDS, NV, SPQ, OF, SI, “Whitaker”, 381e, 968Rb. A) Principal component analysis (PCA) of analyzed varieties. B) UPMGA tree of eight C. pepo varieties. Zucchini group is colored in green and Cocozelle group in orange, while “Whitaker” is colored in blue.

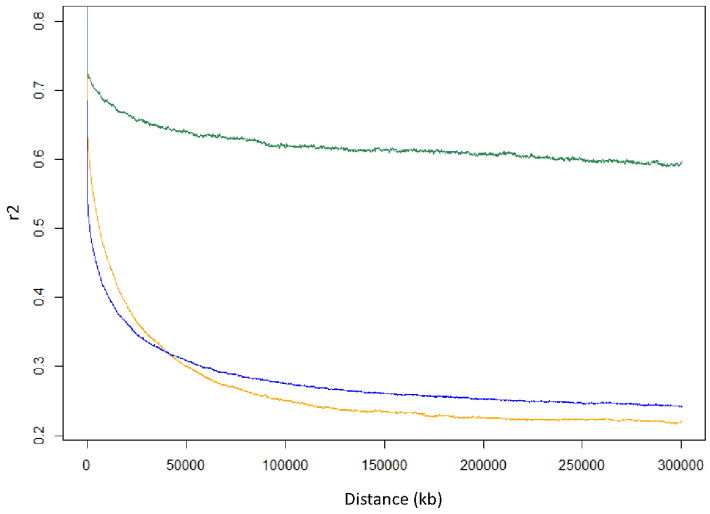
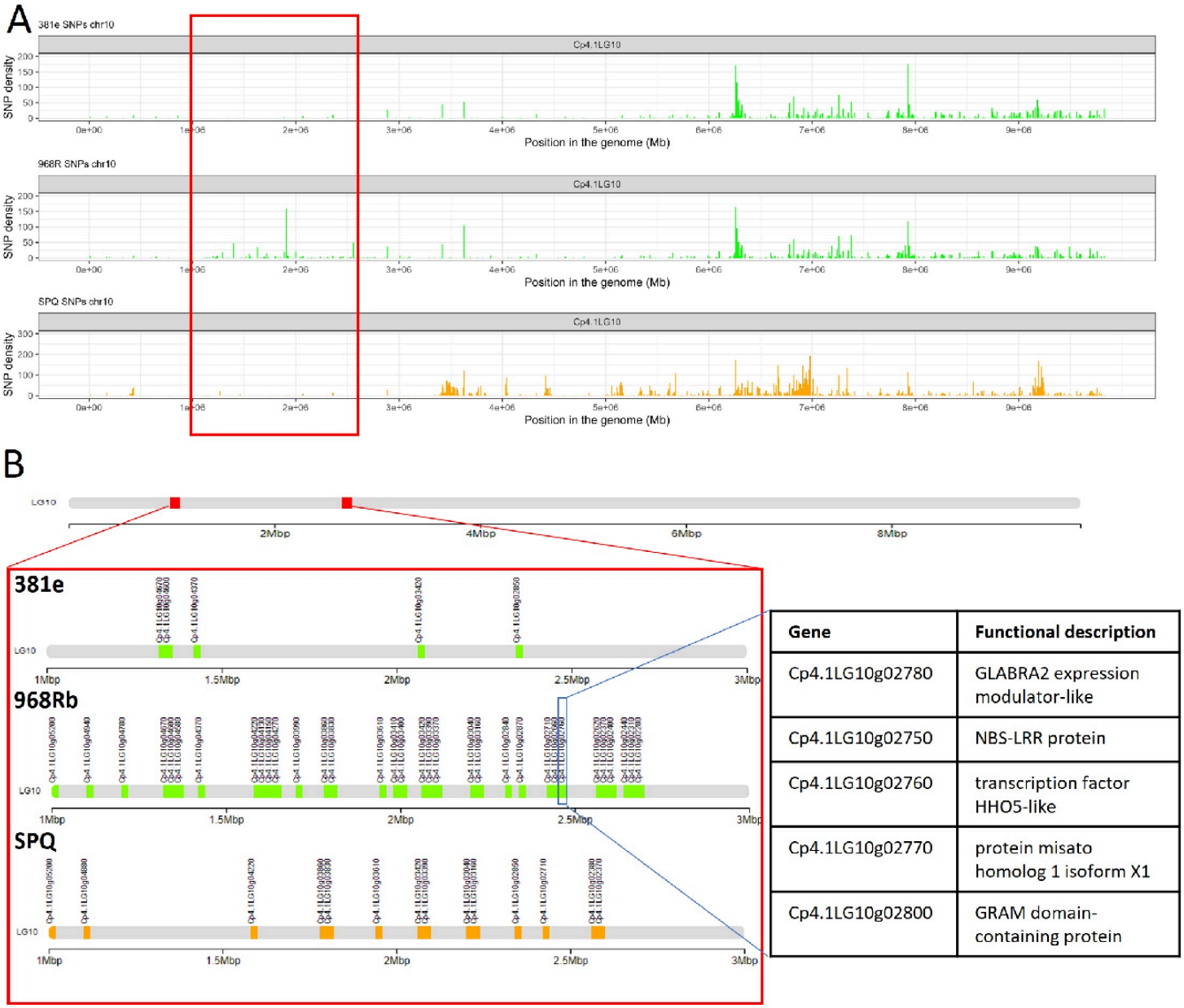


Figure 4

Linkage disequilibrium (r2) for the C. pepo, using all varieties, and for group. Up to down, Zucchini group in green, all varieties in blue and Cocozelle group in orange.



Gene	Functional description
Cp4.1LG10g02780	GLABRA2 expression modulator-like
Cp4.1LG10g02750	NBS-LRR protein
Cp4.1LG10g02760	transcription factor HHO5-like
Cp4.1LG10g02770	protein misato homolog 1 isoform X1
Cp4.1LG10g02800	GRAM domain-containing protein

Figure 5

A) SNP density along the chromosome 10 in 381e, 968Rb and SPQ. B) Genes affected by variants and located on the chromosome 10, in the region of high SNP density for 968Rb, between 1 Mb-2,8 Mb.



Figure 6

A) SNP distribution along the chromosome 8 in 381e, 968Rb and SPQ. B) Genes putatively involved in response to biotic stresses in the region of chromosome 10. For each gene the number and the type of mutation as well the biological processes they are involved are represented.

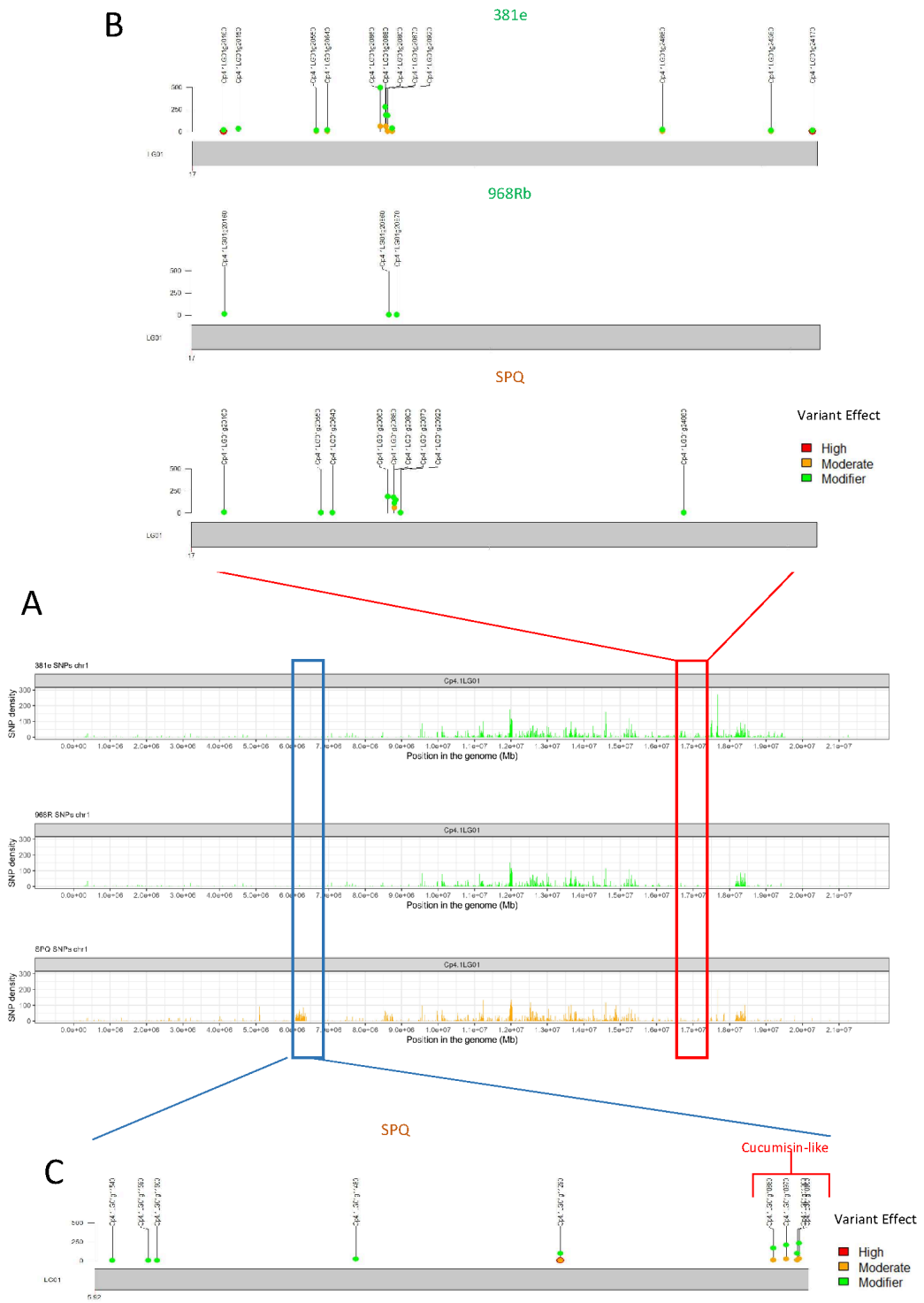


Figure 7

A) SNP distribution along the chromosome 1 in 381e, 968Rb and SPQ. B) Zoom on the region of 17-17.7 Mb on chromosome 1 in 381e, 968Rb and SPQ. C) Zoom on the region of 6 Mb on chromosome 1 in SPQ. The grey horizontal bar indicates the chromosome 1, the vertical black lines indicate the position of the genes along the chromosome. Along the vertical line of each gene is reported the number of

variations affecting the gene and the colors of the circles indicate the effect of variation (high, moderate and modifier).

Supplementary Files

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

- FigureS1.pdf
- FigureS2.pdf
- TableS1.xls
- TableS2.xls
- TableS3.xls
- TableS4.xlsx