the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2022) **110,** 1128–1143 doi: 10.1111/tpj.15728

Chromosome-level pepino genome provides insights into genome evolution and anthocyanin biosynthesis in Solanaceae

Xiaoming Song[1,*,†], Haibin Liu[1,†], Shaoqin Shen[1,†], Zhinan Huang[2,†], Tong Yu[1], Zhuo Liu[1], Qihang Yang[1], Tong Wu[1], Shuyan Feng[1], Yu Zhang[1], Zhiyuan Wang[1] and Weike Duan[2,*]

[1]*School of Life Sciences, North China University of Science and Technology, Tangshan, 063210, Hebei, China, and*
[2]*College of Life Sciences and Food Engineering, Huaiyin Institute of Technology, Huai'an 223003, China*

Received 25 October 2021; accepted 7 March 2022; published online 16 March 2022.
*For correspondence (e-mail songxm@ncst.edu.cn; weikeduan@126.com).
†These authors contributed equally to this work.

SUMMARY

Pepino (*Solanum muricatum*, $2n = 2x = 24$), a member of the Solanaceae family, is an important globally grown fruit. Herein, we report high-quality, chromosome-level pepino genomes. The 91.67% genome sequence is anchored to 12 chromosomes, with a total length of 1.20 Gb and scaffold N50 of 87.03 Mb. More than half the genome comprises repetitive sequences. In addition to the shared ancient whole-genome triplication (WGT) event in eudicots, an additional new WGT event was present in the pepino. Our findings suggest that pepinos experienced chromosome rearrangements, fusions, and gene loss after a WGT event. The large number of gene removals indicated the instability of Solanaceae genomes, providing opportunities for species divergence and natural selection. The paucity of disease-resistance genes (*NBS*) in pepino and eggplant has been explained by extensive loss and limited generation of genes after WGT events in Solanaceae. The outbreak of *NBS* genes was not synchronized in Solanaceae species, which occurred before the Solanaceae WGT event in pepino, tomato, and tobacco, whereas it was almost synchronized with WGT events in the other four Solanaceae species. Transcriptome and comparative genomic analyses revealed several key genes involved in anthocyanin biosynthesis. Although an extra WGT event occurred in Solanaceae, *CHS* genes related to anthocyanin biosynthesis in grapes were still significantly expanded compared with those in Solanaceae species. Proximal and tandem duplications contributed to the expansion of *CHS* genes. In conclusion, the pepino genome and annotation facilitate further research into important gene functions and comparative genomic analysis in Solanaceae.

Keywords: pepino genome, whole-genome triplication, genome evolution, resistance genes, anthocyanin biosynthesis genes, Solanaceae.

INTRODUCTION

Pepino, or sweet cucumber (*Solanum muricatum*, $2n = 2x = 24$), is an herbaceous domesticated crop originating from the Andean region, and is now cultivated worldwide (Anderson & Kim, 1996). The fruit of pepino is popular because of its attractive appearance, with yellow skin covered by purple stripes (Herraiz, Blanca, et al., 2016). Furthermore, it is highly nutritious and is gradually being used as a potential new horticultural plant (Rodríguez-Burruezo et al., 2011; Yalçin, 2012). Recent studies have shown that the pepino is rich in phenols and flavonoids as antioxidants, which can lower total cholesterol and benefit patients with type 2 diabetes (Hsu et al., 2020; Virani et al., 2020). In addition, the aqueous extract of pepino can ameliorate oxidative stress and lipid accumulation in alcoholic fatty liver disease (Hsu et al., 2018).

The pepino belongs to the Solanaceae family, which contains approximately 95 genera and 2300 species (Cao et al., 2021). The genomes of several species in this family have been reported, including tomato (*Solanum pennellii*, *Solanum lycopersicum*, and *Solanum pimpinellifolium*) (Bolger et al., 2014; International Tomato Genome Sequencing Consortium, 2012; Schmidt et al., 2017; Takei et al., 2021; Wang et al., 2020), potato (*Solanum tuberosum*, *Solanum phureja*) (Pham et al., 2020; Xu et al., 2011; Zhou et al., 2020), pepper (*Capsicum baccatum*, *Capsicum*

chinense, and *Capsicum annuum*) (Kim et al., 2014; Kim, Park, et al., 2017; Qin et al., 2014), eggplant (*Solanum melongena*) (Barchi et al., 2019, 2021; Hirakawa et al., 2014; Wei et al., 2020), tobacco (*Nicotiana tabacum*, *Nicotiana attenuate*, *Nicotiana sylvestris*, *Nicotiana tomentosiformis*) (Sierro et al., 2013; Sierro et al., 2014; Xu et al., 2017), wolfberry (*Lycium barbarum*, *Lycium ruthenicum*) (Cao et al., 2021), and *Petunia* species (*Petunia axillaris*, *Petunia inflata*) (Bombarely et al., 2016). However, the pepino genome has not been analyzed until now.

The most important characteristic is that the pepino is closely related to the tomato and potato phylogenetically, and they have the same chromosome number ($x = 12$) (Herraiz, Blanca, et al., 2016; Särkinen et al., 2013). This close relationship may allow the use of the pepino as a genetic source for potato and tomato breeding to improve disease resistance and flavor (Rodríguez-Burruezo et al., 2011; Trognitz & Trognitz, 2005). The introgression of pepino traits into tomatoes has been obtained based on the construction of tomato–pepino somatic hybrids (Sakomoto & Taguchi, 1991). Therefore, pepino is a part of the tertiary gene pool for both tomato and potato breeding.

Research on pepino has primarily focused on pest and disease control (Hu et al., 2016; Ishikawa & Takahata, 2019; Kim, Ishikawa, et al., 2017), fruit composition and processing (Herraiz, Raigón, et al., 2016; Herraiz, Villaño, et al., 2016; Özcan et al., 2020), mosaic virus isolates (Fribourg et al., 2019; Ge et al., 2012; Kim, Ishikawa, et al., 2017), transcriptomes (Herraiz, Blanca, et al., 2016), and development of molecular markers (Anderson & Kim, 1996; Blanca et al., 2007; Herraiz et al., 2015; Nadeem & Muhammad, 2014). Although it is an important crop and a new potential species in many areas, there have been limited molecular and physiological studies. Furthermore, no genomic research has been conducted on pepino. To clarify Solanaceae biology and evolution, we produced a high-quality assembly pepino genome in this study and performed comprehensive comparative genomic analyses of Solanaceae. Here, we report a chromosomal-level pepino genome that integrates PacBio, Illumina, and Hi-C technologies. This study aims to deduce the evolutionary trajectories of Solanaceae genomes and explore important genes regulating disease resistance and anthocyanin biosynthesis.

RESULTS AND DISCUSSION

Pepino genome sequencing, assembly, and assessment

The pepino genome was sequenced using PacBio and Illumina technologies (Figure 1a; Table 1). First, we initially estimated the pepino genome using *k*-mer = 17 by 91.48 Gb Illumina sequencing data (Table 1, Tables S1 and S2). The estimated genome size was 1.195 Gb, and the heterozygosity rate was 0.83% (Figure S1; Table S2). PacBio was used to generate 161.32 Gb data with an average 135.00× coverage depth (Table 1). The PacBio reads were of high quality and long with an N50 length of 38 015 bp and average length of 22 440 bp (Table S3). In total, 252.80 Gb (211.55×) pepino DNA sequences obtained from PacBio and the Illumina platform were used for preliminary *de novo* assembly. The results showed that the cumulative scaffold length was 1.20 Gb and scaffold N50 was 6.41 Mb (Tables S4–S6).

We further performed Hi-C analysis to improve the pepino genome assembly and obtained 73.04 Gb (61.12×) high-quality sequences (Table 1). The Hi-C contact map was used to separate distinct regions of the different chromosomes (Figure 1b). Finally, the revised genome size was 1.20 Gb, with contig N50 length reaching 6.34 Mb, and scaffold N50 of 87.03 Mb (Table S7). In total, 1.10 Gb sequences, accounting for 91.67% of the revised assembled genome, were anchored to 12 chromosomes in the pepino (Figure 1c; Table S8). The mapping rate reads were over 99.35%; therefore, we generated a relatively complete pepino genome (Table S9). BUSCO analysis indicated that 98.1% of 1641 complete genes were detected in pepino (Table S10). CEGMA analysis showed that 97.18% (241) of the core eukaryotic genes were found in pepino (Table S11).

Genome annotation

We found that 61.92% of the estimated pepino genome was composed of repetitive sequences (Figure 1c; Table S12). Most transposable elements were long-terminal repeats, with a total length of over 640.93 Gb, accounting for 53.28% of the pepino genome (Figure S2; Table S12). Long interspersed nuclear elements (LINE), DNA transposons, and simple repeat sequences only accounted for 4.00%, 2.36%, and 1.86% of the pepino genome, respectively (Table S12). The genes were primarily located in the terminal chromosomal regions, which were also confirmed by the expression data (Figure 1c). The distribution of genes was consistent with that of DNA transposable elements, while retrotransposons (primarily copia and gypsy) were nearly inversely distributed on chromosomes. In addition, we found that the distribution of tandem repeats, simple sequence repeats, and long interspersed nuclear elements on chromosomes showed a similar trend (Figure 1c).

Among the 33 734 annotated pepino genes (Figures S3 and S4; Tables S13 and S14), Swiss-Prot, non-redundant protein, InterPro, and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases provided evidence of function for 33 339 (98.83%) genes, with 19 569 annotated by four databases (Figure S5; Table S15). In addition, 370 miRNAs, 2713 rRNAs, 1275 tRNAs, and 696 snRNAs were detected, accounting for 0.18% of the pepino genome (Figure S6; Table S16).

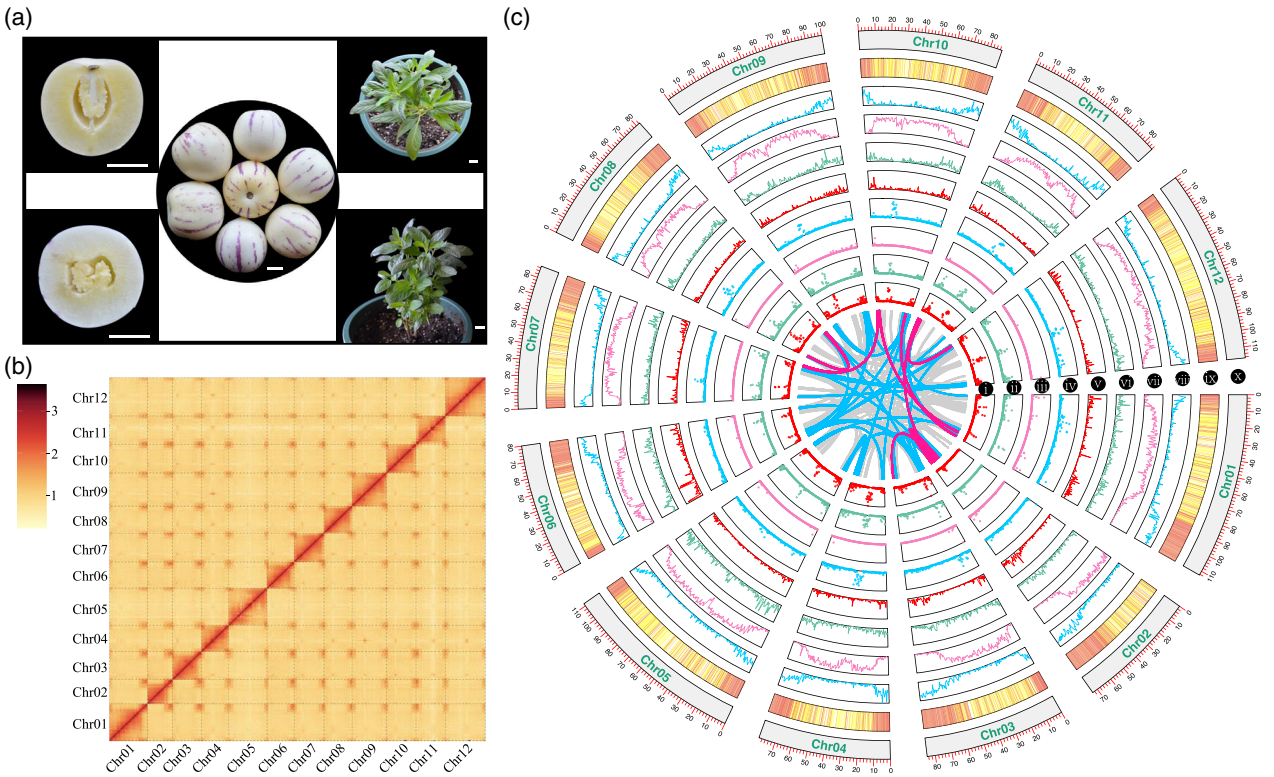

Figure 1. Morphology, Hi-C contact map, and genome assembly of pepino genome.
(a) Morphology of the seedlings and fruit of pepino. Scale bars, 2 cm.
(b) Genome-wide all-by-all interactions among 12 pepino chromosomes obtained by Hi-C.
(c) i, Distribution of TRF genes (non-overlapping, window size, 50 kb); ii, distribution of simple sequence repeats; iii, distribution of SINE; iv, distribution of LINE; v, density of DNA repeats (non-overlapping, window size, 1 Mb); vi, density of copia-type transposons (non-overlapping window size is 1 Mb); vii, density of gypsy-type transposons (non-overlapping, window size, 1 Mb). viii, Gene density. The statistics of various repeat sequences and gene density from i to viii were performed in the non-overlapping regions, and the window size was set as 500 kb. ix, Gene expression levels (Log2FPKM) in pepino, and the colors from white to orange represent the expression level from low to high. x, 12 pepino chromosomes. The inner curve lines indicated collinear gene blocks. The gray, blue, and red colors represent 5–20, 20–50, and >50 gene pairs in collinear blocks, respectively.

Table 1 Summary of pepino genome sequencing data

Paired-end libraries	Insert size (bp)	Total data (Gb)	Read length (bp)	Coverage (×)
Illumina reads	350	91.48	150	76.55
PacBio reads	–	161.32	249 738/ 22 440[a]	135.00
Sub-total	–	252.80	–	211.55
Hi-C	–	73.04	–	61.12
Total	–	325.84	–	272.67

[a]Maximum and average length of the PacBio reads.

Gene family cluster and expansion analysis

First, members of gene families and family size were detected in the pepino and other nine species (Figure 2a). In total, 29 250 gene families were detected in 10 species, including 2309 single-copy gene families and 7469 common gene families. Furthermore, we selected six related Solanaceae species for gene family identification to analyze pepino better. The results showed that pepino has 575 species-specific gene families, fewer than the tobacco (1064), eggplant (872), pepper (743), potato (590), and more than tomato (428) (Figure S7). Notably, pepino, tomato, potato, and eggplant shared 349 Solanum-specific gene families.

Gene family contraction and expansion were detected in pepino and nine other representative species. In total, 29 250 gene families were inferred from their most recent common ancestor (Figure 2b). In pepino, we detected 1003 gene family expansions, fewer than in potato (1219), but more than in all other examined species. However, only 143 gene family contractions were detected in pepino, fewer than in tomato (1863) and potato (1156), whereas more than in other species. Finally, we performed functional enrichment analysis of 4262 genes from 1003 expanded and 348 genes from 143 contracted gene families. The most expanded gene families were related to mismatch repair, DNA replication, and homologous

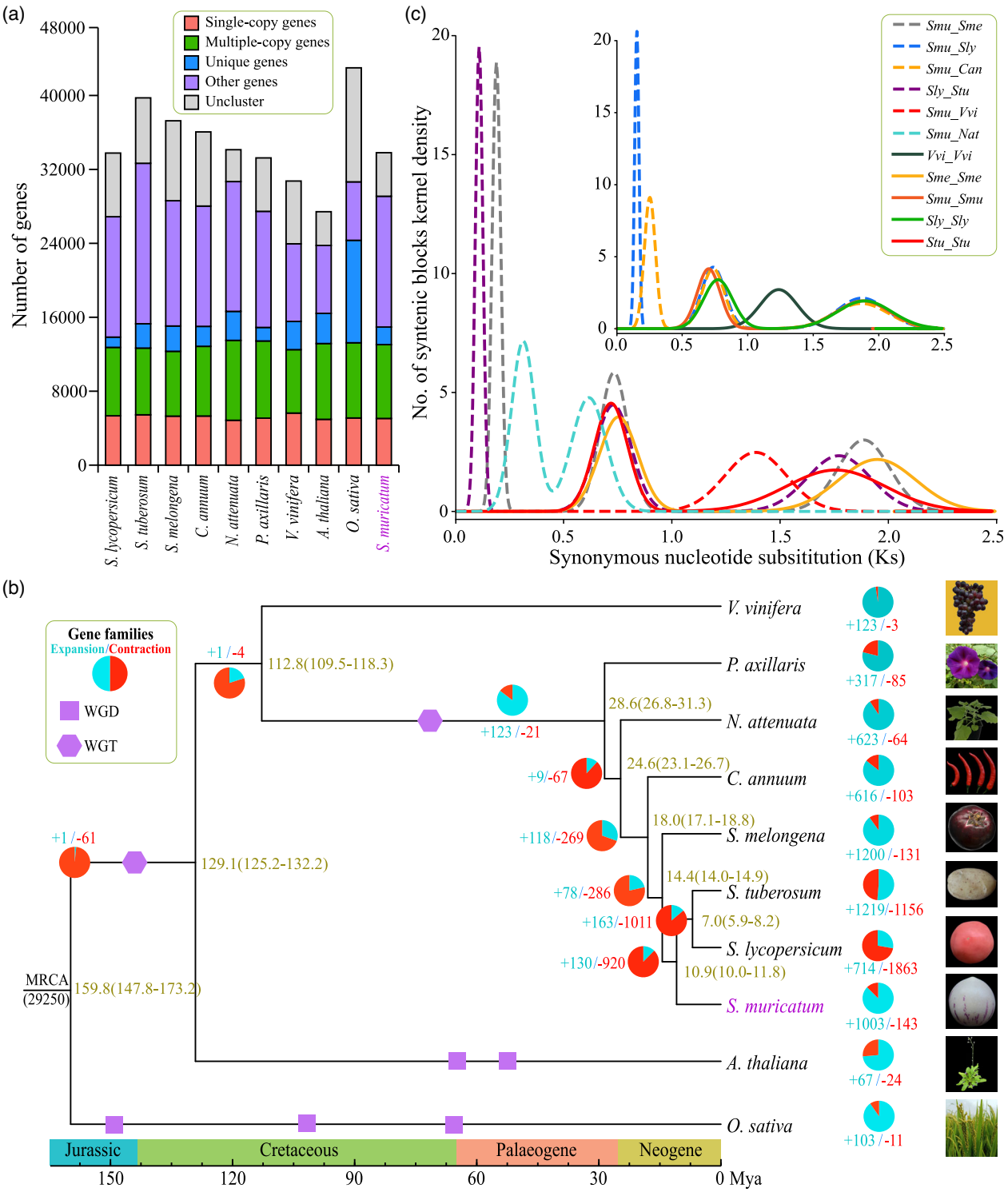

Figure 2. Gene family and evolutionary dating analyses.
(a) Distribution of gene numbers and family sizes in pepino and nine other representative species.
(b) Gene family expansion/contraction analyses and divergence time estimation. The numbers on the nodes indicated the divergence time of the species (Mya, million years ago), with the confidence range in brackets. The blue and red pies indicate the expansion and contraction number of gene families. MRCA, most recent common ancestor; WGD, whole-genome duplication; WGT, whole-genome triplication.
(c) Density of several synonymous substitutions per synonymous site (K_s) among collinear genes. The continuous and dashed lines represent the K_s values of genes within and between species, respectively.

recombination, whereas contracted gene families were primarily related to plant–pathogen interactions, endocytosis, sesquiterpenoid, and triterpenoid biosynthesis (Tables S17 and S18). The plant–pathogen interaction gene family contraction might be related to reduced resistance genes in pepino, which was similar to that of the eggplant (Wei et al., 2020).

Evolution and polyploidization of pepino genome

Based on 2309 single-copy gene families, we conducted a phylogenetic analysis and divergence time estimation (Figure 2b). The results showed that pepino diverged from the common ancestor of tomato and potato from 10.0 to 11.8 Ma. The tomato and potato had a close relationship, and their divergence time was 5.9–8.2 Ma. The group of four *Solanum* species (pepino, tomato, potato, and eggplant) is sister to pepper, diverging 17.1–18.8 Ma. Furthermore, we explored the evolution of pepino according to the rate of synonymous nucleotide substitution (K_s) of syntenic blocks (Figure 2c). Among Solanaceae crops, pepino first diverged from tobacco at approximately 20.32 Ma (K_s = 0.31), followed by pepper at approximately 16.39 Ma (K_s = 0.25), eggplant at 12.45 Ma (K_s = 0.19), and then tomato and potato at approximately 9.83 Ma (K_s = 0.15) (Table S19). K_s peaks of collinear genes indicated divergent evolutionary rates among *Solanum* species, with pepino evolving the slowest, followed by potato and eggplant, and tomato had the fastest evolutionary rate (Figure 2c; Table S19). All these conclusions were approved by the phylogenetically inferred divergence time using Mcmctree (Figure 2b).

The polyploidization events in the pepino were detected based on the K_s density plot. Both pepino and other Solanaceae species had two peaks, indicating that two polyploidization events occurred in Solanaceae. Combining dot plots and K_s density plot analysis, we found that Solanaceae experienced two rounds of WGT events. The ancient WGT event was shared with grapes and most eudicots (Jaillon et al., 2007), and the recent WGT event occurred in Solanaceae species. The recent WGT event had occurred during 45.37–51.28 Ma (K_s = 0.70–0.77) (Figure 2c; Table S19).

Genome organization in Solanaceae plants

The WGT event that occurred in Solanaceae led to the pepino genome organization (Figure 2c). We identified 1003 collinear blocks within the pepino genome, involving 15 732 collinear gene pairs. We mapped the pepino sequences on to other Solanaceae to infer their collinearity (Figures S8–S13). In total, 312 collinear blocks were detected between pepino and grape, and the largest one contained 782 gene pairs. Among Solanaceae species, the most collinear blocks occurred between pepino and pepper (203), followed by eggplant (156), tomato (125), potato (104), and tobacco (90) (Figure 3a, Figure S14). However, there were only five large collinear blocks (collinear gene pairs >500) between pepino and pepper, and no large collinear blocks were found between pepino and tobacco. These results indicated that these species underwent chromosome rearrangements after their divergence, which was consistent with the composition of chromosome fragments between pepino and other examined species (Figures S15–S20).

The ratio of collinear regions between pepino and grape was 1:3, and the ratio between pepino and five Solanaceae species (tobacco, pepper, eggplant, potato, tomato) was 1:1 because of the two rounds of WGT events in pepino (Figure 3b,c; Table S20). For example, chromosome (Chr)9 was aligned to pepino Chr3, 6, and 12, and they were further aligned to the corresponding tomato chromosomes (Figure 4a,b). The microsynteny analysis also showed a 1:1 ratio between pepino and other Solanaceae. For example, the 0.99–1.92 Mb of Chr9 in pepino was perfectly collinear with tobacco (Chr7: 113.58–114.72 Mb), pepper (Chr9: 237.33–238.70 Mb), eggplant (Chr9: 0.71–1.61 Mb), potato (Chr9: 2.42–3.28 Mb), and tomato (Chr9: 2.89–3.56 Mb) (Figure 3c). In collinear regions, we found several chromosomal rearrangements and fusion between the pepino and other Solanaceae species. Using tomato as a reference, chromosome fusion and exchange occurred between Chr4 and Chr11 in pepino. In addition, the end of Chr11 in pepino underwent inversion (Figure 4c,d). A similar phenomenon occurred in pepino and other Solanaceae species, which is consistent with previous reports in other species (Wei et al., 2020).

Randomness of gene loss and gradual genome fractionation in pepino

The pepino collinear regions were divided into three groups using grapes as a reference. The three groups of duplicated regions, containing only 7183, 2362, and 1593 collinear genes in pepino, cover 21.29%, 7.00%, and 4.72% of the total genes, respectively (Table S20). Therefore, the majority of duplicated genes produced by WGTs have been lost.

We performed gene retention analysis in homologous regions by comparing it with other species. The chromosomal regions duplicated by WGTs had divergent gene retention levels (Figure 5a,b; Table S21). Overall, the average retention rates of collinear pepino genes were 57.77%, 50.30%, 34.44%, 36.37%, and 37.17% using tomato, eggplant, potato, pepper, and tobacco, respectively (Table S21). The average retention rates of collinear pepino genes in the three groups were 14.48%, 5.43%, and 3.38% using grape as a reference (Figure 5a; Table S21). The pattern of syntenic depth was 4:2 between grape and pepino due to gene loss or variation in pepinos (Figure 5c). Similarly, the pattern was 2:1 between pepino and pepper or tobacco (Figures S21 and S22). This phenomenon

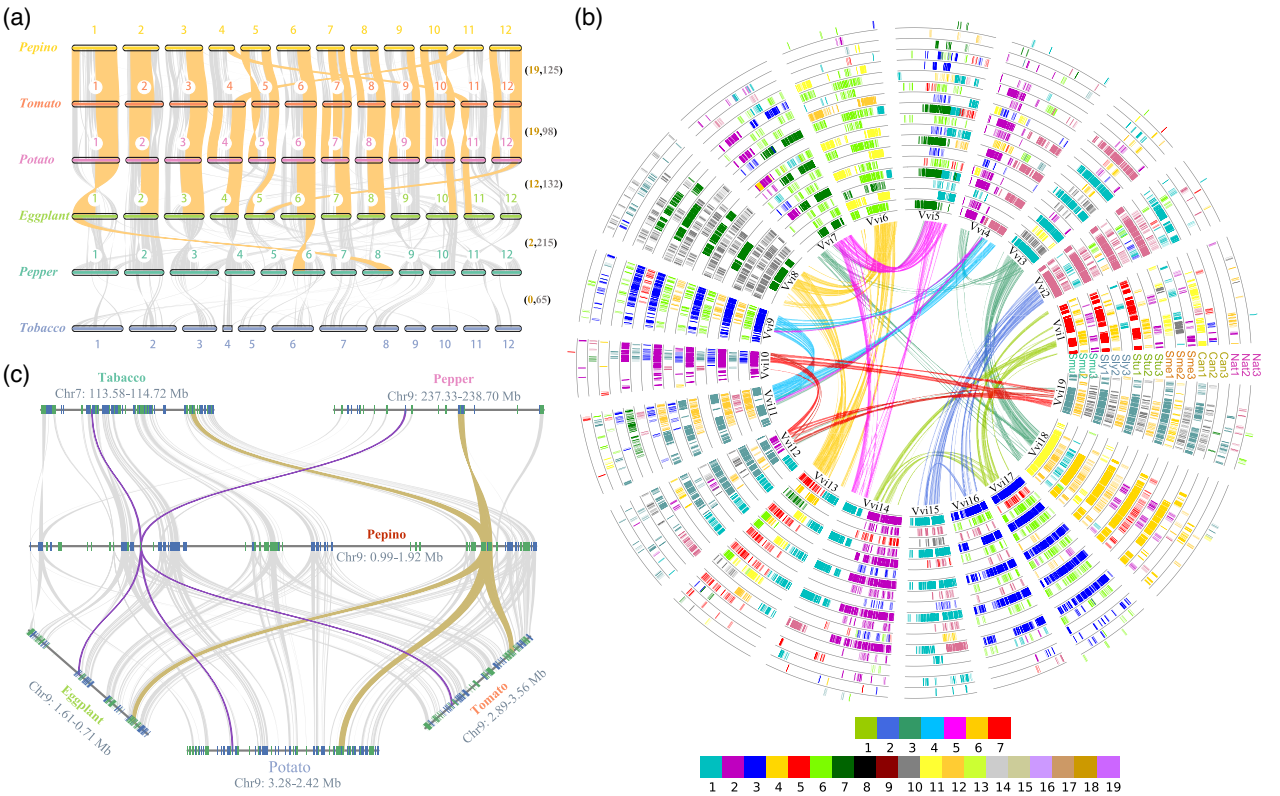

Figure 3. Global and microsynteny alignment of genomes.
(a) Syntenic comparison between pepino and other Solanaceae species, including tomato, potato, eggplant, pepper, and tobacco. Syntenic blocks were linked by gray lines, and the large syntenic blocks (>500 gene pairs) were highlighted in orange.
(b) Global alignment of homologous regions in pepino and five other Solanaceae genomes with the grape as a reference. Each Solanaceae genome was further divided into three subgenomes due to an additional whole-genome triplication event occurred in their genomes after divergence from the grape. Colinear genes between each subgenome of Solanaceae species and grape were shown in each circle, colored as to grape chromosome number according to their respective homologous with grape, as shown in the inset color scheme. The curved lines in the inner circle were formed by 19 grape chromosome colors corresponding to the seven ancestral chromosomes before the ancient core-eudicot common hexaploid ancestor (γ event).
(c) Microsynteny alignment of genes between pepino and other Solanaceae species. The representative synteny relationship indicated that one pepino region matched one region in tobacco, pepper, eggplant, potato, and tomato. Rectangles represent annotated genes with orientation on the reverse strand (green) and same strand (blue). The gray lines connected collinear gene pairs, with two regions highlighted in purple and orange colors.

indicated the large-scale genome fractionation and instability of pepino after its split from these species. However, the syntenic depth was still 1:1 between the pepino and eggplant, tomato, or potato because of their close relationship (Figures S23–S25).

Significant reduction of disease resistance genes after WGT event in pepino

In total, 22 855 genes from 63 families were identified in the pepino and other nine species (Figure 6a; Table S22). The nucleotide-binding site (*NBS*, 2625) gene family was the largest among all examined families. In pepino, 2303 genes from 63 families were detected, with the most being members of the *MYB* (264), *APETALA2*/ethylene response factor (*AP2*/*ERF*, 176), basic helix–loop–helix (*bHLH*, 139), and *NBS* (138) gene families (Table S22). Compared with the other nine species, the number of most (51 of 63, 80.95%) gene families was not notably different in pepino (Figure 6a; Table S23). However, the gene number of six gene families (NBS, Whirly, CPP, NF-X1, STAT, and S1Fa-like) in pepino was less than that in other species, and four gene families (TCP, GeBP, EIL, and Alfin) in pepino were higher than in the other species (Figure 6a; Table S23).

Among the gene families with significant differences in number, the *NBS* family contained more genes in each species. The pepino had 138 *NBS* genes, which were far fewer than potato (414), pepper (283), grape (436), and rice (466) (Table S22). Based on the K_s density plot of NBS family genes, the time of NBS burst in pepino (K_s^{peak} = 1.215) was approximately 79.64 Ma, which was the earliest among all the examined Solanaceae species (Figure 6b). The NBS burst times of pepino were similar to those of tomato (K_s^{peak} = 1.174, T = 76.95 Ma) and tobacco (K_s^{peak} = 1.125, T = 73.74 Ma), but far earlier than that in the other four Solanaceae species, including *P. axillaris* (K_s^{peak} = 0.787, T = 5 1.59), potato (K_s^{peak} = 0.773, T = 50.67), eggplant (K_s^{peak} = 0.734, T = 48.11), and pepper

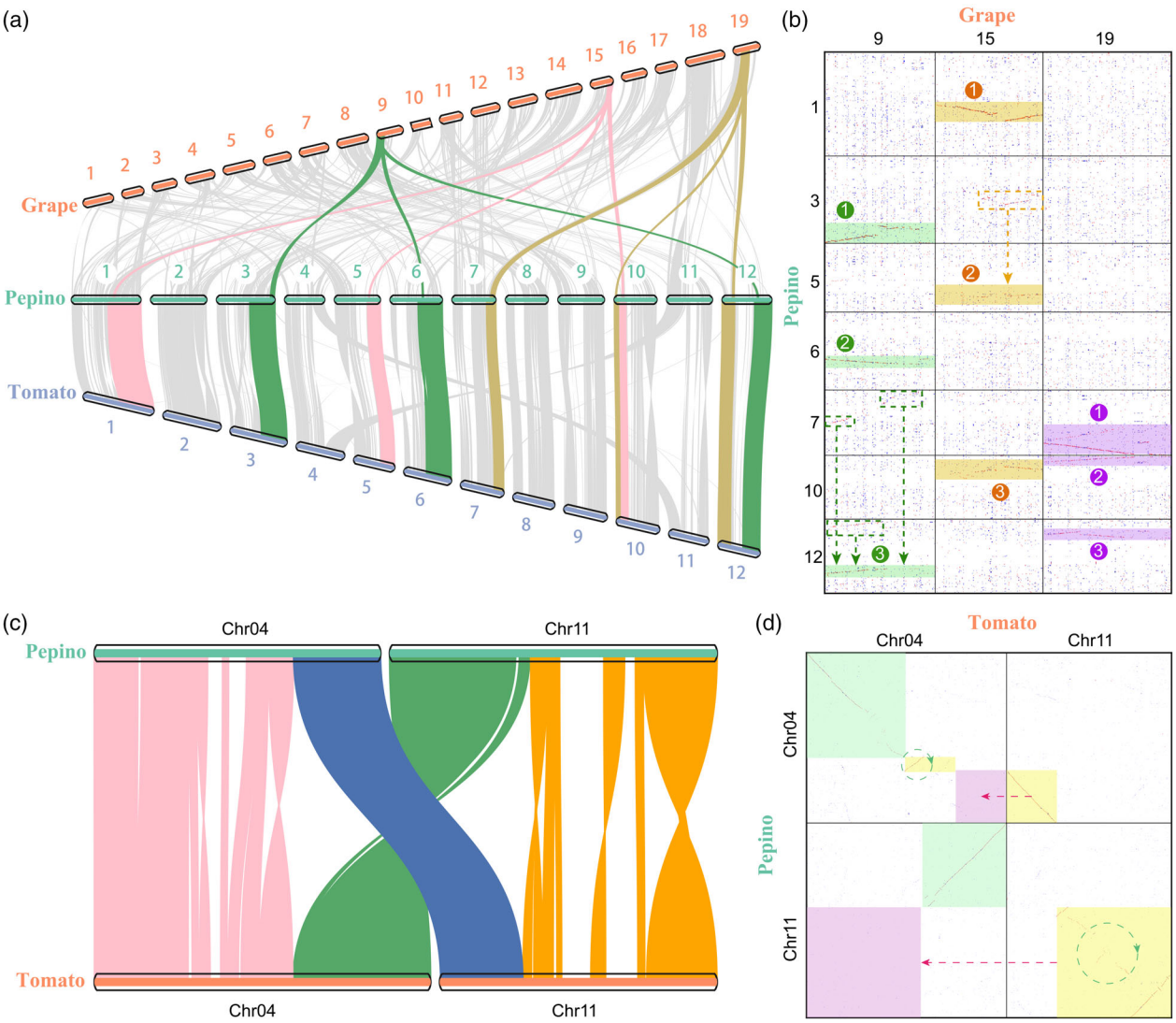

Figure 4. Syntenic comparisons and homologous dot plots between pepino and grape or tomato.
(a) Syntenic comparison revealed that pepino and other Solanaceae species had additional whole-genome triplication after the shared whole-genome triplication in the common ancestor of all angiosperms. Three examples for syntenic relationships of three regions in pepino matching a single genomic region in grape were highlighted in green (Vvi9 versus Smu3, 6, and 12), pink (Vvi15 versus Smu1, 5, and 10), and orange (Vvi19 versus Smu1, 7, and 10), respectively.
(b) Homologous dot plot between selected grape chromosomes (9, 15, and 19) and corresponding pepino chromosomes. Red, blue, and gray dots represent the best, secondary, and other homologous genes, respectively. Three homologous regions were marked out by rectangles numbered by 1, 2, and 3 in circles.
(c) Syntenic comparison revealed that chromosome arrangement and inversion between several chromosomes in pepino and tomato.
(d) Homologous dot plot between selected tomato chromosomes (4 and 11) and corresponding pepino chromosomes. Dashed-line with arrow showed complement correspondence generated by chromosome breakages in the evolution.

(K_s^{peak} = 0.719, T = 47.13). This result indicated that the outbreak of *NBS* genes was not synchronized in different Solanaceae species. Among them, the *NBS* gene outbreak in pepino, tomato, and tobacco occurred before the WGT event in Solanaceae, whereas the NBS outbreak in *P. axillaris*, potato, pepper, and eggplant was almost synchronized with the WGT event of Solanaceae species (Figure 6b). In addition to studying the early outbreak of *NBS* family genes, we also investigated the recent evolution of *NBS* genes in the genomes of pepino and other species (Figure 6c–f, Figure S26). Only 86 pairs of pepino *NBS* genes had K_s < 0.3 (i.e., diverged in the past approximately 20 million years), which was far less than that in potato (1777), grape (1044), and pepper (443) (Figure 6c–f; Table S24). This phenomenon indicated that the *NBS* family genes in pepino had more gene loss than the other Solanaceae species after the WGT event.

Compared with potato and pepper, the *NBS* genes in pepino showed extreme paucity, which is consistent with the trend of eggplant and tobacco (Table S22). Based on

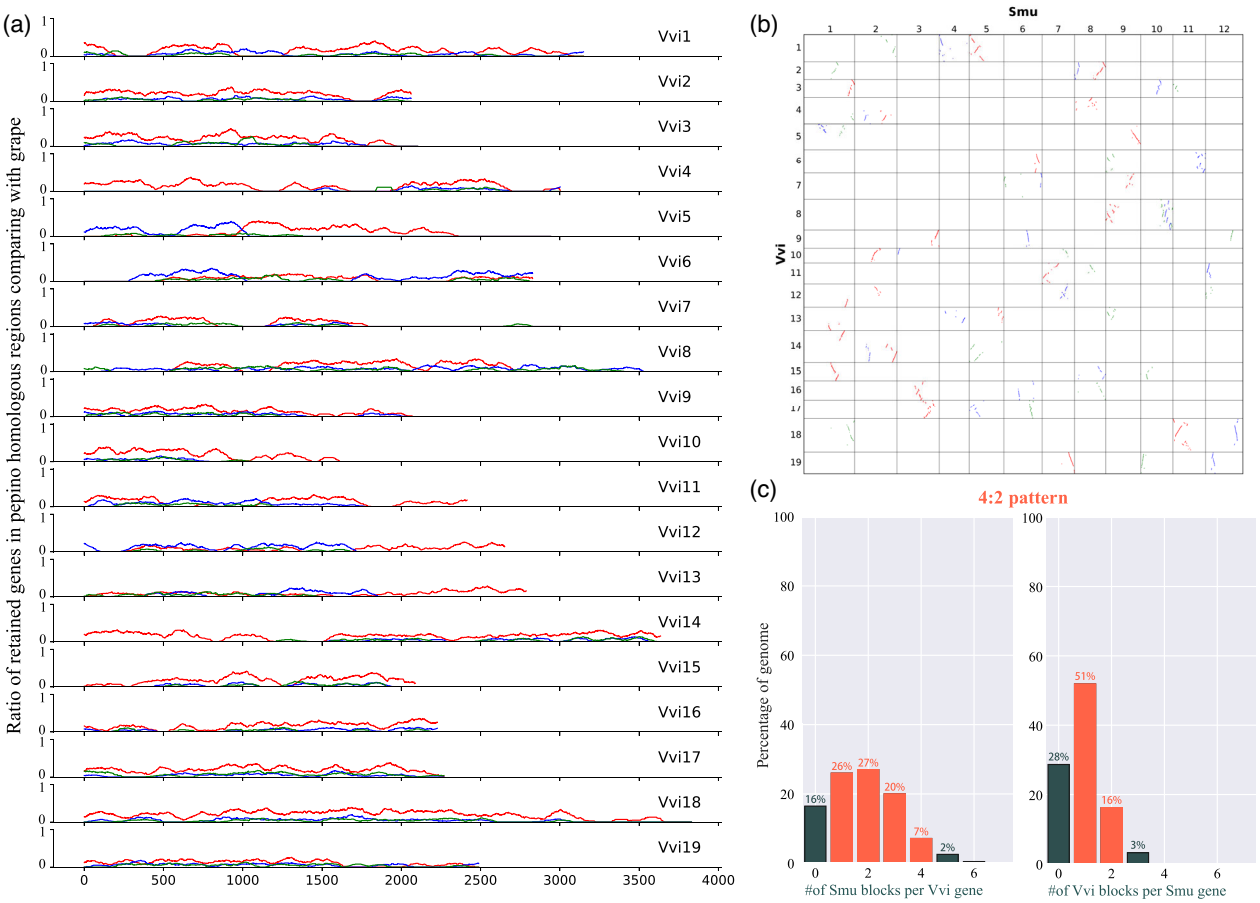

Figure 5. Gene retention analysis of three subgenomes in pepino compared with that of grapes.
(a) Retention of duplicated genes residing in three subgenomes of pepino along with each chromosome of the grape.
(b) Homologous dot plot between pepino and grape genome.
(c) Syntenic depth analysis between pepino and grape.

inferred collinearity, a group of 51 *NBS* genes on potato Chr4 corresponded to only 11 *NBS* genes at the orthologous region on pepino Chr4, implying loss of at least 40 pepino *NBS* genes at this location, with similar losses inferred on several other chromosomes (Figure 6c,d). Phylogenetic analysis of *NBS* genes also showed large-scale gene loss, many branches with only a singleton pepino gene, and some with no comparison with potato, pepper, grape, and rice (Figure 6g).

Exploration of key genes in the anthocyanin biosynthesis pathway

As a wild relative of domesticated pepino, *Solanum caripense* (wild pepino) provides a rich resource of variation to improve the cultivated species. Here, we tried to detect differentially expressed genes between pepino and wild pepino using the RNA-sequencing (RNA-seq) dataset. Compared with wild pepino, a total of 1105 upregulated and 1061 downregulated genes were identified in pepino. Functional enrichment analysis indicated that the upregulated genes in pepino were involved in anthocyanin biosynthesis (Figure S27; Table S25).

The sequences of 29 *Arabidopsis* genes encoding 11 enzymes implicated in anthocyanin biosynthesis were used as seeds to identify homologs in pepino and other species (Figure 7a; Table S26). Most nodes in the pathway had one or more gene copies among the 10 species. There were four chalcone isomerase (*CHI*) genes in pepino, more than in the other Solanaceae species (Figure 7a; Tables S26 and S27). Five genes (*Sm05G02138*, *SmUnG137G00010*, *Sm02G02175*, *Sm08G01977*, and *Sm09G02395*) were upregulated in pepino, which encoded CHS, CHI, DFR, ANS, and UGT enzymes, respectively (Table S28). One gene (*Sm09G00505*) encoding PAL and two genes (*Sm09G00151* and *Sm09G00152*) encoding CHI enzymes were downregulated in pepino. Furthermore, we also detected 17 regulatory genes involved in anthocyanin biosynthesis in pepino using homologous and phylogenetic analyses (Figure 7b; Table S28). Most of these genes belonged to MYB, bHLH, and later organ boundary

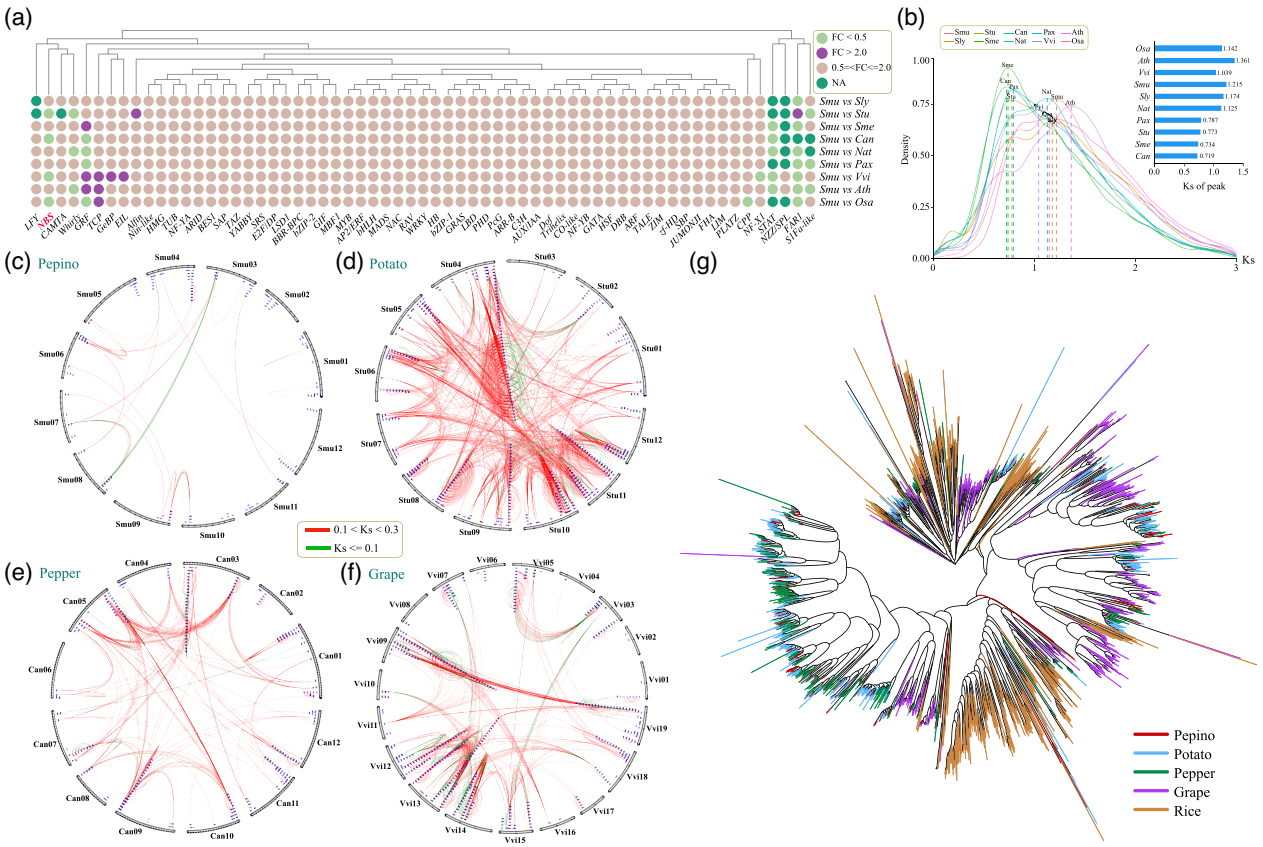

Figure 6. Comparative analysis of the gene families of pepino and nine representative plants.
(a) Heatmap of the fold change (FC) of the gene family number between pepino and other species. The purple and light green circle represent the fold change >2.0 and <0.5, respectively.
(b) K_s density plot of *NBS* family genes in pepino and other nine species.
(c) Distribution of nucleotide-binding site (*NBS*) genes on each chromosome of pepino and three other species, including potato (d), pepper (e), and grape (f). Curve lines represent K_s values of *NBS* gene pairs that were <0.1 (green); or >0.1 but <0.3 (red).
(g) Maximum-likelihood trees of *NBS* family genes that were constructed using the amino acid sequences with 1000 bootstrap repeats in pepino, potato, pepper, grape, and rice.

domain-containing protein (LBD). Among the 17 genes, *SmMYB111*, *SmMYBL2*, *SmTT8*, *SmTT19*, and *SmCPC* were upregulated, whereas *SmLBD39* was downregulated in pepino (Figure 7b, Figures S28 and S29; Table S28).

Furthermore, we conducted a comparative analysis of genes involved in anthocyanin biosynthesis in pepino and nine other species (Figure 8a). Interestingly, we found that 42 genes encoding CHS enzyme were identified in grapes, indicating that it was significantly expanded in grapes compared with other species (Figure 8a,b; Tables S26 and S27). Phylogenetic analysis showed that most of the *CHS* genes in grapes cluster together (Figure 8b). The chromosomal distribution showed that most *CHS* genes (34; 80.95%) in grapes located on Chr16, and were divided into two clusters (Figure 8c,d, Figure S30; Table S29). Further analysis revealed that two duplicated types contributed to the expansion of *CHS* genes in grapes. Twenty-one (50.00%) of *CHS* genes belonged to proximal duplication genes, and 18 (42.86%) genes were tandem duplication genes in grapes (Figure 8d). However, all three *CHS* genes of pepino belonged to the dispersed duplicated type (Figure 8c).

CONCLUSIONS

The first pepino genome sequences, together with comparative genomic analysis, will provide rich resources for both fundamental and applied research in Solanaceae species. The sequence consistency and integrity assessment by CEGMA and BUSCO reflected the high-quality, chromosome level of the assembled pepino genome. In addition to the shared ancient WGT event that occurred in most eudicots (Jaillon et al., 2007), additional WGT events occurred in pepino and other Solanaceae species. Our findings suggested that the pepino genome undergoes chromosome rearrangements, fusions, and gene removal after the WGT event. Several gene removals indicated the instability of the Solanaceae genomes, providing opportunities for species divergence and natural selection. Interestingly, we found large-scale removal of *NBS* family genes in the pepino genome and limited production of new *NBS*

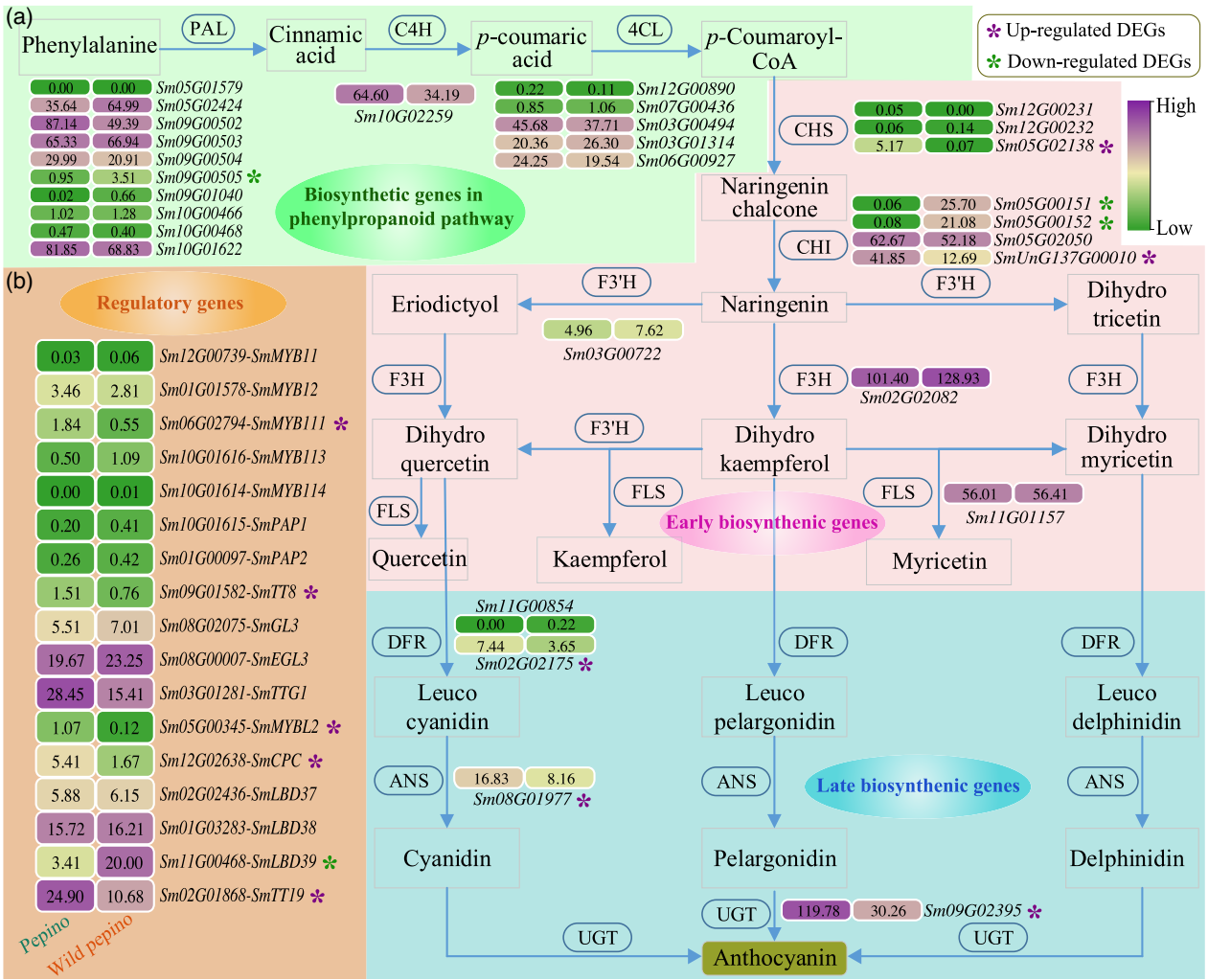

Figure 7. Inferred pepino anthocyanin biosynthesis genes.
(a) Identification of structural anthocyanin biosynthesis genes in pepino, including biosynthetic genes in phenylpropanoid pathway (background with green color), early biosynthetic genes (background with red color), and late biosynthetic genes (background with blue color). Gene expression was detected in the pepino and wild pepino. The green and purple colors indicated low and high expression levels, respectively. Purple and red asterisks represent the up- and downregulated genes in pepino.
(b) Identification and expression analysis of the regulatory genes of the anthocyanin biosynthesis pathway in pepino.

duplicates after the Solanaceae WGT event. Further analysis indicated that the outbreak of *NBS* family genes was not synchronized in Solanaceae species. The *NBS* burst in pepino was the earliest among all the examined Solanaceae species. *NBS* family gene outbreaks in pepino, tomato, and tobacco occurred before the Solanaceae WGT event, whereas it was almost synchronized with the WGT event in *P. axillaris*, potato, pepper, and eggplant.

Comparative genomic studies and transcriptome analysis contributed to understanding the evolutionary and gene functions of pepino. The copy number of genes involved in the anthocyanin biosynthesis pathway was comprehensively studied in pepino and compared with that in other Solanaceae. Several key genes were identified in pepino and their expression levels were explored between different species. In conclusion, this study will lay a solid foundation for studying the gene functions and genome evolution of pepino, and other Solanaceae species.

EXPERIMENTAL PROCEDURES

Genome sequencing and Hi-C technology

Solanum muricatum leaf samples were collected for genomic DNA extraction and library construction. Three sequencing strategies were used in this study: (i) two paired-end libraries were constructed with 350 bp fragments and sequenced using Illumina technology (Illumina Inc., San Diego, CA, USA); (ii) libraries were constructed and sequenced using the PacBio pipeline (Pacific Biosciences, Menlo Park, CA, USA); and (iii) Illumina sequencing data were used to assist assembly according to Hi-C technology. The

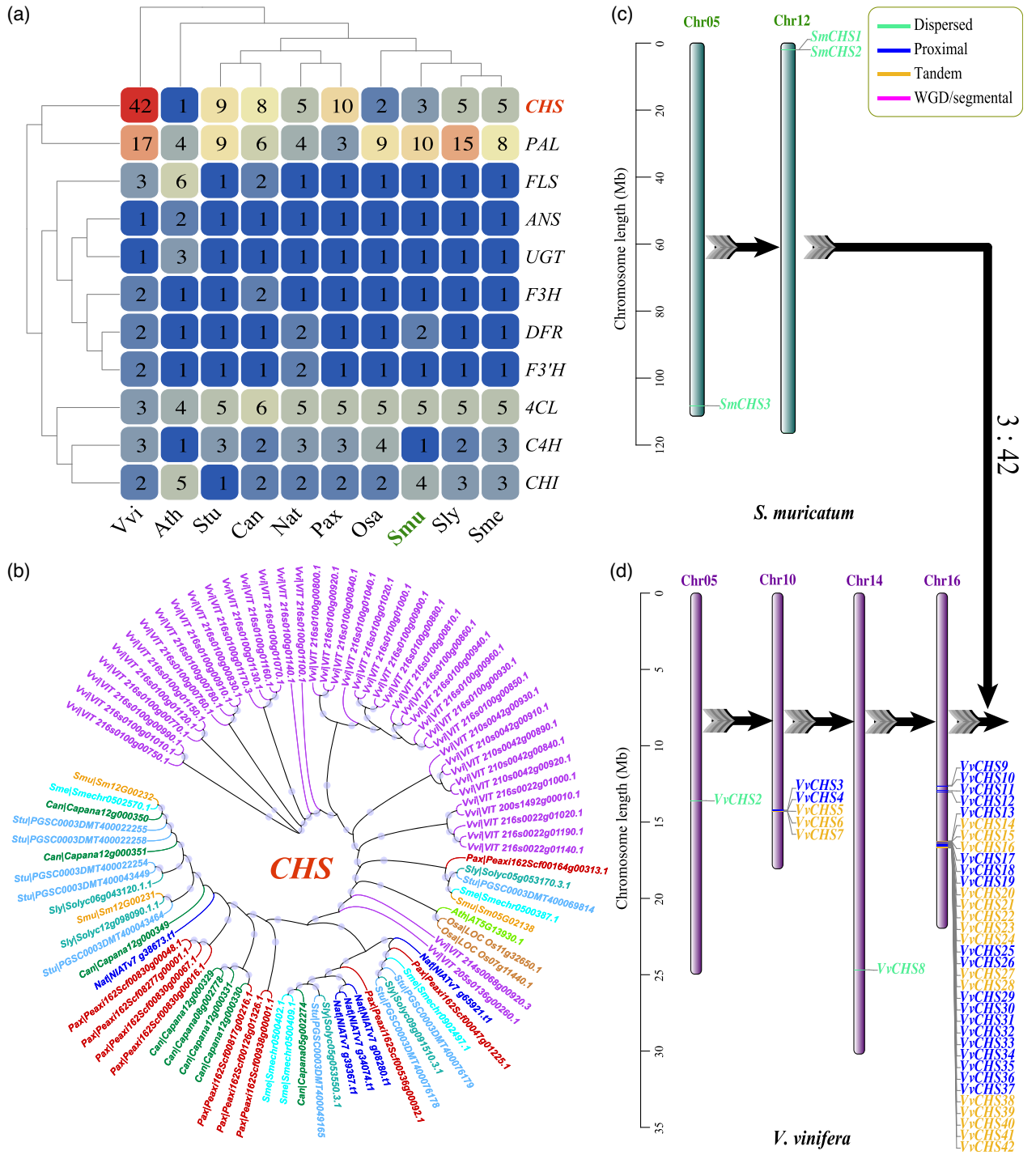

Figure 8. Chromosomal distribution, duplication type, and phylogenetic analyses of the anthocyanin biosynthesis genes.
(a) Heatmap of the structural anthocyanin biosynthesis genes number in pepino and other nine species.
(b) Maximum-likelihood trees of *CHS* family genes that were constructed using the amino acid sequences with 1000 bootstrap repeats in pepino and other nine species.
(c) Distribution of *CHS* genes on each chromosome in pepino. The gene name with different colors represents various duplication types identified by the MCScanX program.
(d) Distribution of *CHS* genes on each chromosome in grape.

17-nt *k*-mers were used to estimate genome size (Marcais & Kingsford, 2011).

Genome assembly and assessment

We conducted genome assembly using different software, such as Canu (Koren et al., 2017), Falcon (https://github.com/PacificBiosciences/FALCON), and Nextdenovo (https://github.com/Nextomics/NextDenovo). After evaluating the effects of different software assemblies, we chose Nextdenovo for genome assembly. Nextdenovo contains three modules: NextCorrect, NextGraph, and NextPolish. Based on the string graph algorithm, we used NextCorrect to conduct error correction of the original sequencing data. Then, NextGraph was used to assemble the genome based on the corrected data. Finally, we corrected the assembled genome using NextPolish to obtain an accurate and high-quality pepino genome. LACHESIS was used to assist genome assembly based on Hi-C technology (Belaghzal et al., 2017; Burton et al., 2013). The CEGMA and BUSCO programs were used to assess the assembled genome further (Manni et al., 2021; Parra et al., 2007).

Genome annotation

We performed pepino genome annotation, which primarily contained repeated sequence annotation, gene annotation, and non-coding RNA annotation.

(i) *De novo* prediction and homologous sequence alignment were used to annotate the repeated sequences. First, we constructed a repeat sequence database using RepeatModeler, LTR_FINDER (Xu & Wang, 2007), Piler (Edgar & Myers, 2005), and RepeatScout programs (Price et al., 2005). Repeatmasker was used to perform repeat sequence prediction. Homologous sequence alignment was performed using repeatproteinmask and Repeatmasker software by searching the RepBase database (Bao et al., 2015; Tarailo-Graovac & Chen, 2009). TRF software was used to predict tandem repeat sequences (Benson, 1999).

(ii) Gene annotation was performed using several protein databases, including SwissProt, TrEMBL, KEGG, and InterPro.

(iii) The non-coding miRNAs and snRNAs were predicted using INFERNAL software (Nawrocki & Eddy, 2013). rRNA and tRNA were predicted using BLAST and tRNAscan-SE software, respectively (Chan & Lowe, 2019). The distribution of repeat sequences, gene density, and non-coding genes on chromosomes was determined using TBtools (Chen et al., 2020).

Gene prediction

We conducted gene prediction using homologous and *de novo* prediction. Homologous prediction was performed using Genewise and BLAST programs (Birney et al., 2004; Camacho et al., 2009). *De novo* gene prediction was primarily performed using GlimmerHMM (Stanke & Morgenstern, 2005), Augustus (http://bioinf.uni-greifswald.de/augustus/), and SNAP software (Korf, 2004). The IntegrationModeler (EVM) pipeline was then used to integrate the above results (Haas et al., 2008). Finally, we used PASA software to correct EVM gene prediction results by combining transcriptome data (Haas et al., 2003).

Gene family detection and expansion analysis

The identification of gene families among pepino and the other nine representative species was conducted using the OrthoFinder program (Emms & Kelly, 2019). First, we filtered the gene sequences with alternative splicing and retained only the longest transcript. Second, we removed the gene that encoded a protein with a length of <50 amino acids. Third, an all-vs-all BLAST was conducted using the protein sequences of all species to obtain similarity relationships (E-value <1e-5). Finally, multi- and single-copy gene families were obtained by conducting cluster analysis using the MCL graph clustering algorithm. The contraction and amplification of gene families was conducted using the CAFE program (De Bie et al., 2006).

Phylogenetic tree construction and divergence time estimation

The genes of each single-copy family were separately used to conduct multiple sequence alignments using MUSCLE (Edgar, 2004). Then, all alignments of each family were combined to form a super alignment matrix, which was further used to construct a phylogenetic tree. Trees of 10 species were constructed using RAxML with the maximum likelihood model (Stamatakis, 2014). The single-copy gene families were used to calculate the divergence time using the MCMC tree program (Yang, 2007). The TimeTree database (http://www.timetree.org) was used to obtain time correction points (Kumar et al., 2017). The parameters of the MCMC tree program were the sample number = 1 000 000, burn-in = 5 000 000, and sample frequency = 50.

Transcriptome analysis

The raw reads of RNA-seq datasets were downloaded from NCBI with accession numbers SRS1052501 for pepino and SRS1054035 for *S. caripense* (Herraiz, Blanca, et al., 2016). The tissues of young leaves, flowers, and mature fruits were sampled for each species. After filtering, the clean reads were mapped to the pepino genome using the HISAT program (Kim et al., 2015). Gene expression was normalized as fragments per kilobase of transcript sequence per million base pairs (FPKM) (Trapnell et al., 2010). The DESeq program was used to perform differentially expressed gene analysis with the following parameters: $|\log2(\text{fold-change})| > 1$ and $P_{adj} < 0.05$ (Anders & Huber, 2010).

Gene collinearity detection and visualization

A collinearity analysis was performed using the whole-genome duplication integrated analysis (WGDI), which contained an improved version of ColinearScan ("-icl" model) (Sun et al., 2021; Wang et al., 2006). Specifically, the Blastp program was used to detect homologous genes within one genome or between two genomes (E-value <1e-5, Score > 100). Then, the "-icl" model was used to run the improved version of ColinearScan for collinearity analysis. The maximal gap length between two neighboring genes in collinearity was set to 50 genes. The gene families with the number of over 30 genes were removed before running "-icl" model. Dot plots were used to identify homologous blocks generated by various polyploidization events. Dot plots of homologous genes were produced using the WGDI toolkit (Sun et al., 2021).

Lastly, we constructed collinear gene alignments for each Solanaceae species using grape as a reference. Generally, every grape gene might have two additional collinear genes due to the WGT event (Jaillon et al., 2007). For each gene in grape, a gene ID was filled in the cell of a related column when a collinear gene was present. Otherwise, the cell marked a dot when a collinear gene was absent due to gene translocation or loss. For pepino and other Solanaceae species, we assigned them four columns because their genomes underwent additional WGT events. Therefore, the alignment had 18 columns for six Solanaceae species, reflecting layers of two tripled homologies due to recursive

polyploidies across Solanaceae genomes. Finally, the alignment was visualized as a circos plot, which was drawn using the -ci module of WGDI (Sun et al., 2021). Synteny and microsynteny among Solanaceae species were illustrated using the Python version of MCscan (Tang et al., 2008). The duplicate_gene_classifier program in MCScanX was adopted to infer duplicated types (Wang et al., 2012).

K_s calculation and distribution fitting

First, MUSCLE was used to perform alignment based on homologous protein sequences (Edgar, 2004). The program PAL2NAL was used to convert protein alignment into codon alignment based on the coding sequences (Suyama et al., 2006). Finally, the yn00 program in PAML was used to calculate K_a and K_s using the Nei-Gojobori approach (Yang, 2007). In each collinear block, the median K_s of homologous genes was used to classify the blocks generated by each duplication event. The K_s values were marked on a collinear block with different colors using the WGDI program (Sun et al., 2021). The density distribution of K_s was determined using three modules in WGDI. First, the K_s density distribution curve was obtained using K_s peaks. Then, the multi-peak fitting of the curves was conducted using PeaksFit. Finally, K_s figures were used to integrate multiple fitted density curves into one graph.

Anthocyanin biosynthetic and resistance genes identification

Arabidopsis genes related to anthocyanin biosynthesis were retrieved from the KEGG database (Mao et al., 2005). These genes were then used to search for homologous genes in pepino and other examined species using Blastp (E-value <1e-5, identify >60%, score > 150) according to previous reports with slight modifications (Duan et al., 2016; Song et al., 2020). The chromosomal distribution of anthocyanin biosynthetic genes was determined using TBtools (Chen et al., 2020). The *NBS* genes were detected using the number "PF00931" with an E-value <1e-5.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (32172583, 31801856, 31902021), Natural Science Foundation of Hebei (C2021209005), and the China Postdoctoral Science Foundation (2020 M673188, 2021 T140097). The genome sequencing and Hi-C were conducted in the Novogene Corporation.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

XS and WD conceived the project and were responsible for the project initiation. XS and WD supervised and managed the project and research. Experiments and analyses were designed by XS, WD, HL, and ZH Data generation and bioinformatic analyses were led by XS, S.S, TY, ZL, QY, TW, SF, YZ, and ZW. The manuscript was organized, written, and revised by X S, WD, HL, and ZH. All authors read and revised the manuscript.

AVAILABILITY OF DATA AND MATERIALS

The genome sequence and RNA-seq datasets of pepino reported in this paper have been deposited in the Genome Sequence Archive (Wang et al., 2017) in BIG Data Center (Members, 2019), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession numbers CRA005032 and CRA005096 that are publicly accessible at http://bigd.big.ac.cn/gsa. All materials and related data in this study are available upon request.

SUPPORTING INFORMATION

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

Figure S1. *k*-mer distribution of the pepino genome. (a) *k*-mer = 17 Depth and *k*-mer number frequency distribution. (b) *k*-mer = 17 depth and *k*-mer type frequency distribution.

Figure S2. The frequency and the length of the major types of repetitive sequences in pepino genome, including DNA, LINE, LTR, and SINE type repeats.

Figure S3. Comparative analysis of CDS length, exon length, exon number, gene length, and intron length among pepino and other representative species.

Figure S4. The statistics of gene set evidence supports in pepino genome. *De novo*, EVM integrates genes supported by *de novo* prediction; homolog, genes supported by homologous prediction when EVM integration; RNA, genes supported by RNA-seq during EVM integration. The gene overlap is <50% as a standard, and the number indicates the number of genes.

Figure S5. The Venn diagram of gene function annotations in pepino obtained using four databases, including InterPro, Swiss-Prot, NR, and KEGG.

Figure S6. The chromosomal distribution of pepino ncRNAs, including miRNA, rRNA, snRNA, and tRNA.

Figure S7. Common and lineage-specific gene families in pepino and other five Solanaceae species.

Figure S8. The homologous dotplot between pepino (smu) and grape (vvi) genome.

Figure S9. The homologous dotplot between pepino (smu) and tobacco (nat) genome.

Figure S10. The homologous dotplot between pepino (smu) and pepper (can) genome.

Figure S11. The homologous dotplot between pepino (smu) and pepper (can) genome.

Figure S12. The homologous dotplot between pepino (smu) and potato (stu) genome.

Figure S13. The homologous dotplot between pepino (smu) and tomato (sly) genome.

Figure S14. Syntenic comparison between pepino and other Solanaceae species, including potato, eggplant, pepper, and tobacco.

Figure S15. The chromosome representation of pepino using 19 grape chromosomes according to their collinear genes.

Figure S16. The chromosome representation of pepino using 12 tobacco chromosomes according to their collinear genes.

Figure S17. The chromosome representation of pepino using 12 tobacco chromosomes according to their collinear genes.

Figure S18. The chromosome representation of pepino using 12 eggplant chromosomes according to their collinear genes.

Figure S19. The chromosome representation of pepino using 12 potato chromosomes according to their collinear genes.

Figure S20. The chromosome representation of pepino using 12 tomato chromosomes according to their collinear genes.

Figure S21. Gene retention analysis of pepino genome comparing with tobacco. (a) The retention of duplicated genes residing in pepino genome along with each chromosome of tobacco. (b) The syntenic depth analysis between pepino and tobacco genome.

Figure S22. Gene retention analysis of pepino genome comparing with pepper. (a) The retention of duplicated genes residing in pepino genome along with each chromosome of pepper. (b) The syntenic depth analysis between pepino and pepper genome.

Figure S23. Gene retention analysis of pepino genome comparing with eggplant. (a) The retention of duplicated genes residing in pepino genome along with each chromosome of eggplant. (b) The syntenic depth analysis between pepino and eggplant genome.

Figure S24. Gene retention analysis of pepino genome comparing with potato. (a) The retention of duplicated genes residing in pepino genome along with each chromosome of potato. (b) The syntenic depth analysis between pepino and potato genome.

Figure S25. Gene retention analysis of pepino genome comparing with tomato. (a) The retention of duplicated genes residing in pepino genome along with each chromosome of tomato. (b) The syntenic depth analysis between pepino and tomato genome.

Figure S26. The distribution of nucleotide-binding site (NBS) genes on each chromosome of tomato (a), eggplant (b), tobacco (c), Arabidopsis (d), and rice (e). Curve lines represent K_s values of NBS gene pairs that were <0.1 (green); or >0.1 but <0.3 (red).

Figure S27. The KEGG functional enrichment analysis of upregulated genes in pepino.

Figure S28. Maximum-likelihood trees of MYB family genes that were constructed using the amino acid sequences with 1000 bootstrap repeats in pepino (orange) and Arabidopsis (green). The red five-pointed star represents genes related to anthocyanin biosynthetic genes.

Figure S29. Maximum-likelihood trees of MYB family genes that were constructed using the amino acid sequences with 1000 bootstrap repeats in pepino. The red five-pointed star represents genes related to anthocyanin biosynthetic genes.

Figure S30. The chromosomal distribution of the anthocyanin biosynthesis genes in pepino and grape. (a) The distribution of anthocyanin biosynthesis genes on each chromosome in pepino. The gene name with different colors represents the various duplication types identified by MCScanX program. Rectangles with red dashed-lines showed the CHS genes on chromosome. (b) The distribution of anthocyanin biosynthesis genes on each chromosome in grape.

Table S1. Statistics of sequencing data obtained by Illumina Hiseq platform for pepino genome survey.

Table S2. *k*-mer statistics of the genomic characteristics of pepino obtained by genome survey analysis.

Table S3. Statistics of sequencing data of pepino by Pacbio sequel platform.

Table S4. The summary of preliminary assembly of the pepino genome using Illumina and Pacbio technology.

Table S5. Statistics on the base content of the pepino genome.

Table S6. The SNP statistics of the pepino genome.

Table S7. Statistics of pepino genome assembly quality.

Table S8. The assembled length and cluster number of each chromosome of pepino genome.

Table S9. The read coverage statistics of the pepino genome.

Table S10. The summary of BUSCO assessment of the pepino genome.

Table S11. The summary of CEGMA assessment of the pepino genome.

Table S12. The statistics of the repeat sequence classification in pepino genome.

Table S13. The statistical of gene structure prediction in pepino genome.

Table S14. Statistics of gene structure of pepino and other related species.

Table S15. Statistics of gene functional annotations in pepino genome.

Table S16. The statistics of non-coding RNA in pepino genome.

Table S17. The KEGG enrichment analysis of genes from expanded gene family in pepino (adjusted *P*-value <0.05).

Table S18. The KEGG enrichment analysis of genes from contracted gene family in pepino (adjusted *P*-value <0.05).

Table S19. Kernel function analysis of K_s distribution related to duplication events within each genome and between two genomes.

Table S20. The table listing the homologous gene sets between grape and other Solanaceae species. A dot (.) is placed where no homolog is identified in the respective genome.

Table S21. Retained genes in Pepino homologous regions comparing with other species.

Table S22. The number of various kinds of gene families in pepino and other nine species.

Table S23. The fold change of each gene family in pepino comparing with other nine species.

Table S24. The number NBS gene pairs of different K_s in pepino and other nine species.

Table S25. The KEGG enrichment analysis of upregulated genes in pepino comparing with wild pepino.

Table S26. The main enzymes and genes involved in anthocyanins biosynthesis in Arabidopsis according to the KEGG. The candidated anthocyanins biosynthesis related genes in pepino and other species were identified by comparing with Arabidopsis.

Table S27. The number of anthocyanins biosynthesis genes in pepino and other nine species.

Table S28. The differential gene expression analyses for the anthocyanin biosynthesis-related genes in pepino and wild pepino using RNA-seq data.

Table S29. The rename of anthocyanins biosynthesis related genes in grape.

REFERENCES

Anders, S. & Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biology*, **11**, R106.

Anderson, G.J. & Kim, J.Y. (1996) The origin and relationships of the Pepino, *Solanum muricatum* (Solanaceae): DNA restriction fragment evidence. *Economic Botany*, **50**, 369–380.

Bao, W., Kojima, K.K. & Kohany, O. (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6**, 11.

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A. *et al.* (2019) A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Scientific Reports*, **9**, 11769.

Barchi, L., Rabanus-Wallace, M.T., Prohens, J., Toppino, L., Padmarasu, S., Portis, E. *et al.* (2021) Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *The Plant Journal*, **107**, 579–596.

Belaghzal, H., Dekker, J. & Gibcus, J.H. (2017) Hi-C 2.0: An optimized hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods*, **123**, 56–65.

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**, 573–580.

Birney, E., Clamp, M. & Durbin, R. (2004) GeneWise and genomewise. *Genome Research*, **14**, 988–995.

Blanca, J.M., Prohens, J., Anderson, G.J., Zuriaga, E., Cañizares, J. & Nuez, F. (2007) AFLP and DNA sequence variation in an Andean domesticate, Pepino (*Solanum muricatum*, Solanaceae): implications for evolution and domestication. *American Journal of Botany*, **94**, 1219–1229.

Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T. *et al.* (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nature Genetics*, **46**, 1034–1038.

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C.S. *et al.* (2016) Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nature Plants*, **2**, 16074.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O. & Shendure, J. (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, **31**, 1119–1125.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Cao, Y.L., Li, Y.L., Fan, Y.F., Li, Z., Yoshida, K., Wang, J.Y. *et al.* (2021) Wolfberry genomes and the evolution of Lycium (Solanaceae). *Commun Biol*, **4**, 671.

Chan, P.P. & Lowe, T.M. (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods in Molecular Biology*, **1962**, 1–14.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y. *et al.* (2020) TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, **13**, 1194–1202.

De Bie, T., Cristianini, N., Demuth, J.P. & Hahn, M.W. (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, **22**, 1269–1271.

Duan, W., Huang, Z., Song, X., Liu, T., Liu, H., Hou, X. *et al.* (2016) Comprehensive analysis of the polygalacturonase and pectin methylesterase genes in *Brassica rapa* shed light on their different evolutionary patterns. *Scientific Reports*, **6**, 25107.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edgar, R.C. & Myers, E.W. (2005) PILER: identification and classification of genomic repeats. *Bioinformatics*, **21**(Suppl 1), i152–i158.

Emms, D.M. & Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**, 238.

Fribourg, C.E., Gibbs, A.J., Adams, I.P., Boonham, N. & Jones, R.A.C. (2019) Biological and molecular properties of wild potato mosaic virus isolates from Pepino (*Solanum muricatum*). *Plant Disease*, **103**, 1746–1756.

Ge, B.B., Liu, G.J. & Wang, H.Q. (2012) First report of tomato mosaic virus infecting Pepino in China. *Plant Disease*, **96**, 1704.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I. *et al.* (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, **31**, 5654–5666.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J. *et al.* (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, **9**, R7.

Herraiz, F.J., Blanca, J., Ziarsolo, P., Gramazio, P., Plazas, M., Anderson, G.J. *et al.* (2016) The first de novo transcriptome of Pepino (*Solanum muricatum*): assembly, comprehensive analysis and comparison with the closely related species *S. caripense*, potato and tomato. *BMC Genomics*, **17**, 321.

Herraiz, F.J., Raigón, M.D., Vilanova, S., García-Martínez, M.D., Gramazio, P., Plazas, M. *et al.* (2016) Fruit composition diversity in land races and modern Pepino (*Solanum muricatum*) varieties and wild related species. *Food Chemistry*, **203**, 49–58.

Herraiz, F.J., Vilanova, S., Andújar, I., Torrent, D., Plazas, M., Gramazio, P. *et al.* (2015) Morphological and molecular characterization of local varieties, modern cultivars and wild relatives of an emerging vegetable crop, the Pepino (*Solanum muricatum*), provides insight into its diversity, relationships and breeding history. *Euphytica*, **206**, 301–318.

Herraiz, F.J., Villaño, D., Plazas, M., Vilanova, S., Ferreres, F., Prohens, J. *et al.* (2016) Phenolic profile and biological activities of the Pepino (*Solanum muricatum*) fruit and its wild Relative *S. caripense*. *International Journal of Molecular Sciences*, **17**, 394.

Hirakawa, H., Shirasawa, K., Miyatake, K., Nunome, T., Negoro, S., Ohyama, A. *et al.* (2014) Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world. *DNA Research*, **21**, 649–660.

Hsu, J.Y., Lin, H.H., Hsu, C.C., Chen, B.C. & Chen, J.H. (2018) Aqueous extract of Pepino (*solanum muriactum* Ait) leaves ameliorate lipid accumulation and oxidative stress in alcoholic fatty liver disease. *Nutrients*, **10**(7), 931.

Hsu, J.Y., Lin, H.H., Wang, Z.H. & Chen, J.H. (2020) Aqueous extract from Pepino (*Solanum muricatum* Ait.) leaves ameliorated insulin resistance, hyperlipidemia, and hyperglycemia in mice with metabolic syndrome. *Journal of Food Biochemistry*, **44**, e13518.

Hu, J., Yang, H., Long, X., Liu, Z. & Rengel, Z. (2016) Pepino (*Solanum muricatum*) planting increased diversity and abundance of bacterial communities in karst area. *Scientific Reports*, **6**, 21938.

International Tomato Genome Sequencing Consortium. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.

Ishikawa, T. & Takahata, K. (2019) Insect and mite pests of Pepino (*Solanum muricatum* Ait.) in Japan. *Biodiversity Data Journal*, **7**, e36453.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

Kim, D., Langmead, B. & Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**, 357–360.

Kim, O.K., Ishikawa, T., Yamada, Y., Sato, T., Shinohara, H. & Takahata, K. (2017) Incidence of pests and viral disease on Pepino (*Solanum muricatum* Ait.) in Kanagawa prefecture, Japan. *Biodiversity Data Journal*, (5): e14879.

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.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. & Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, **27**, 722–736.

Korf, I. (2004) Gene finding in novel genomes. *BMC Bioinformatics*, **5**, 59.

Kumar, S., Stecher, G., Suleski, M. & Hedges, S.B. (2017) TimeTree: a resource for timelines, Timetrees, and divergence times. *Molecular Biology and Evolution*, **34**, 1812–1819.

Manni, M., Berkeley, M.R., Seppey, M., Simao, F.A. & Zdobnov, E.M. (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, **38**(10), 4647–4654.

Mao, X., Cai, T., Olyarchuk, J.G. & Wei, L. (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*, **21**, 3787–3793.

Marcais, G. & Kingsford, C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, **27**, 764–770.

Members, B.I.G.D.C. (2019) Database resources of the BIG data center in 2019. *Nucleic Acids Research*, **47**, D8–D14.

Nadeem, M.S. & Muhammad, K. (2014) Morphogenetic study of pepino and other members of solanaceae family. *American Journal of Plant Sciences*, **5**, 3761–3768.

Nawrocki, E.P. & Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.

Özcan, M.M., Al Juhaimi, F., Ahmed, I.A.M., Uslu, N., Babiker, E.E. & Ghafoor, K. (2020) Effect of microwave and oven drying processes on antioxidant activity, total phenol and phenolic compounds of kiwi and Pepino fruits. *Journal of Food Science and Technology*, **57**, 233–242.

Parra, G., Bradnam, K. & Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**, 1061–1067.

Pham, G.M., Hamilton, J.P., Wood, J.C., Burke, J.T., Zhao, H., Vaillancourt, B. *et al.* (2020) Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience*, **9**(9), giaa100.

Price, A.L., Jones, N.C. & Pevzner, P.A. (2005) De novo identification of repeat families in large genomes. *Bioinformatics*, **21**(Suppl 1), i351–i358.

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J. *et al.* (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *capsicum* domestication and specialization. *Proceedings of the National Academy of Sciences*, **111**, 5135–5140.

Rodríguez-Burruezo, A., Prohens, J. & Fita, A.M. (2011) Breeding strategies for improving the performance and fruit quality of the Pepino (*Solanum muricatum*): a model for the enhancement of underutilized exotic fruits. *Food Research International*, **44**, 1927–1935.

Sakomoto, K. & Taguchi, T. (1991) Regeneration of intergeneric somatic hybrid plants between *Lycopersicon esculentum* and *Solanum muricatum*. *Theoretical and Applied Genetics*, **81**, 509–513.

Särkinen, T., Bohs, L., Olmstead, R.G. & Knapp, S. (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology*, **13**, 214.

Schmidt, M.H.W., Vogel, A., Denton, A.K., Istace, B., Wormit, A., van de Geest, H. *et al.* (2017) De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *The Plant Cell*, **29**, 2336–2348.

Sierro, N., Battey, J.N., Ouadi, S., Bakaher, N., Bovet, L., Willig, A. *et al.* (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nature Communications*, **5**, 3833.

Sierro, N., Battey, J.N.D., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N. *et al.* (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biology*, **14**, R60.

Song, X.M., Wang, J.P., Sun, P.C., Ma, X., Yang, Q.H., Hu, J.J. *et al.* (2020) Preferential gene retention increases the robustness of cold regulation in Brassicaceae and other plants after polyploidization. *Horticultural Research*, **7**, 20.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stanke, M. & Morgenstern, B. (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, **33**, W465–W467.

Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., Xi, Z., Wang, X. and Liu, J. (2021) WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *bioRxiv*, https://doi.org/10.1101/2021.04.29.441969

Suyama, M., Torrents, D. & Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, W609–W612.

Takei, H., Shirasawa, K., Kuwabara, K., Toyoda, A., Matsuzawa, Y., Iioka, S. *et al.* (2021) De novo genome assembly of two tomato ancestors, *Solanum pimpinellifolium* and *Solanum lycopersicum* var. cerasiforme, by long-read sequencing. *DNA Research*, **28**(1), dsaa029.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M. & Paterson, A.H. (2008) Synteny and collinearity in plant genomes. *Science*, **320**, 486–488.

Tarailo-Graovac, M. and Chen, N. (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, **25**(4), 4–14.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J. *et al.* (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, **28**, 511–515.

Trognitz, F.C. & Trognitz, B.R. (2005) Survey of resistance gene analogs in *Solanum caripense*, a relative of potato and tomato, and update on R gene genealogy. *Molecular Genetics & Genomics Mgg*, **274**, 595–605.

Virani, D., Chaerunnisa, N.N., Suarsi, I., Dachlan, D.M. and Thahir, A.I.A. (2020) Pepino extract (*Solanum muricatum* Ait.) on HDL and LDL in type 2 diabetic rats. *Enfermeria Clinica*, **30** (Suppl 4), 163–166.

Wang, X., Gao, L., Jiao, C., Stravoravdis, S., Hosmani, P.S., Saha, S. *et al.* (2020) Genome of Solanum pimpinellifolium provides insights into structural variants during tomato breeding. *Nature Communications*, **11**, 5817.

Wang, X., Shi, X., Li, Z., Zhu, Q., Kong, L., Tang, W. *et al.* (2006) Statistical inference of chromosomal homology based on gene colinearity and applications to Arabidopsis and rice. *BMC Bioinformatics*, **7**, 447.

Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T. *et al.* (2017) GSA: genome sequence archive. *Genomics, Proteomics & Bioinformatics*, **15**, 14–18.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X. *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**, e49.

Wei, Q., Wang, J., Wang, W., Hu, T., Hu, H. & Bao, C. (2020) A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Horticultural Research*, **7**, 153.

Xu, S., Brockmöller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z. *et al.* (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proceedings of the National Academy of Sciences*, **114**, 6133.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P. *et al.* (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, **475**, 189–195.

Xu, Z. & Wang, H. (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, **35**, W265–W268.

Yalçin, H. (2012) Effect of ripening period on composition of Pepino (*Solanum muricatum*) fruit grown in Turkey. *African Journal of Biotechnology*, **9**, 3901.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Zhou, Q., Tang, D., Huang, W., Yang, Z., Zhang, Y., Hamilton, J.P. *et al.* (2020) Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nature Genetics*, **52**, 1018–1023.