

Genetic changes in F1 hybrids of the genetically divergent *Solanum* L., *Solanum melongena* L. and *Solanum aethiopicum* L.

Ning Yu (✉ yuningkjilove@126.com)

Guangxi Academy of Agricultural Science <https://orcid.org/0000-0002-6030-9937>

Guiyun Gan

Guangxi Academy of Agricultural Science

Dandan Li

Guangxi Academy of Agricultural Science

Weiliu Li

Guangxi Academy of Agricultural Science

Yaqin Jiang

Guangxi Academy of Agricultural Science

Qihong Yang

Guangxi Academy of Agricultural Science

Peng Wang

Guangxi Academy of Agricultural Science

Wenjia Li

Guangxi Academy of Agricultural Science

Yikui Wang

Guangxi Academy of Agricultural Science

Research Article

Keywords: *Solanum*, Heterosis, Single-nucleotide polymorphism, Insertion/deletion, Structural variation, Copy number variation

Posted Date: August 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1893350/v1>

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

[Read Full License](#)

Abstract

Distant hybridization constitutes an important process in plant evolution. Outbreeding of *Solanum* plants can produce hybrid progeny with heterobeltiosis in terms of growth, development, and resistance. In this study, the genomic characteristics and genetic variation of reciprocal cross F1 hybrids of strains 177 (*Solanum melongena* L.) and Y11 (*Solanum aethiopicum* L.) were quantified to analyze the relationship between heterosis and genomic characteristics. While no significant phenotypic differences existed between the F1 hybrids, their phenotypes were closer to those of the Y11 parent. Using whole-genome resequencing, large numbers of single-nucleotide polymorphisms, insertion or deletions, structural variations, and copy number variation (CNV) were found. The variation rate of the Y11 parent was the highest, while those of the F1 hybrids were similar to each other but closer to Y11. These results reflect the fact that the genome of the heterozygous progenies was stable at the chromosome level; nevertheless, a lot of variation sites were produced, which may underlie the increased phenotypic diversities compared to that found in the cultivated parent plants. Mining and functional annotation of variant genes with different variant types revealed that the highest number of variant genes were caused by CNVs. Moreover, the variant genes were significantly enriched in pathways associated with glucose metabolism, protein synthesis, and phytohormone signaling, indicating that CNVs may be responsible for the heterosis observed in the progeny of outbred crosses. These data can provide a reference for future heterosis research, molecular marker development, genetic engineering, high-density linkage map construction, and genome-wide association analysis.

1. Introduction

The Solanaceae family comprises more than 90 genera and 3,000–4,000 species, many of which are crop plants (e.g., potato, tomato, pepper, tobacco) (Doganlar et al. 2002a; Knapp et al. 2013). Comparisons of the genome structures of 16 solanaceous species revealed that genome-level doubling and large-scale rearrangement events have not occurred in individual species during their independent evolution and that despite the high consistency exhibited between genomes, the phenotypes vary greatly (Doganlar et al. 2002b; Mueller et al. 2005; Wu et al. 2009; Grandillo et al. 2011).

Eggplants (*Solanum melongena* L.) are the third largest group of solanaceous plants after potatoes and tomatoes (Godfrey et al. 2017; Gramazio et al. 2017). Together with two African eggplants (*Solanum aethiopicum* L. and *Solanum macrocarpon* L.), they constitute an important, economically cultivated group (Jacoby et al. 2003; Shackleton et al. 2009; Vorontsova et al. 2013; Gramazio et al. 2016). *Solanum aethiopicum* L. is mainly cultivated in Brazil (Sunseri et al. 2010) and was domesticated from *Solanum anguivi* Lam. (Lester 1986). This species also constitutes one of the five most important vegetables in tropical Africa. Studies on *S. aethiopicum* have focused on classification (Gramazio et al. 2016; Acquadro et al. 2017; Gramazio et al. 2017), hybridization (Prohens et al. 2012; Rattan et al. 2014), and identification of resistance (Collonnier et al. 2001; Gisbert et al. 2011). Despite the morphological diversity of *S. melongena* (Kaushik et al. 2016), the genetic basis is narrow (Augustinos et al. 2016), particularly in commercial cultivars (Muñoz-Falcón et al. 2009), and has been seriously threatened in recent years by

soil-borne diseases (e.g., *Fusarium* wilt, *Verticillium* wilt, and root-knot nematodes). Crossing of *S. melongena* with *S. aethiopicum* can transfer a solanaceous wild relative resistance gene into the cultivated eggplants (Toppino et al. 2008). However, interspecific hybrid incompatibility and high sterility rates in hybrids and individual plant organs have been reported (Rotino et al. 2014; Plazas et al. 2016). Many hybrid progeny traits are intermediate and may represent non-biparental traits, such as stings and inflorescences. Nevertheless, most F1 hybrids bear more similarities to their wild parents than their cultivated parents (Prohens et al. 2013) and exhibit strong growth and disease resistance (Calvo-Asensio et al. 2014; Kaushik et al. 2016; Afful et al. 2018). Heterosis refers to the advantages of hybrid progeny in terms of yield, plant height, viability, and resistance (Fu et al. 2015; Holá et al. 2017). Although heterosis has been widely utilized and various models have been proposed to explain this phenomenon (Comings and MacMurray 2000), its genetic and molecular bases remain largely unknown (Birchler et al. 2010).

In recent years, whole-genome resequencing (WGRS) has emerged as a high-throughput sequencing method for detecting various effective genetic variants on a genome-wide scale. Deep sequencing can detect single-nucleotide polymorphisms (SNPs), insertions or deletions (InDels), structural variations (SVs), copy number variations (CNVs), and other types of genetic variations (Rasmus et al. 2011; Würschum et al. 2015; Zhang et al. 2015). This technique has been widely used in studies of crop yield, resistance, and quality breeding (Ray and Satya 2014; Barabaschi et al. 2016; Rasheed et al. 2017). Furthermore, heterosis-related gene identification and quantitative trait locus (QTL) mapping can be performed via population resequencing (Yates et al. 2004; Li et al. 2010; Meyer et al. 2010; Yuan et al. 2012; Shang et al. 2016; Pan et al. 2017). Molecular studies of eggplant have enabled the construction of genetic maps of related traits (Fukuoka et al. 2012; Gramazio et al. 2014; Wei et al. 2020b), the development of molecular markers (Gramazio et al. 2016), and transcript (Yang et al. 2014) and genome sequence assembly (Hirakawa et al. 2014; Song et al. 2019; Wei et al. 2020a; Li et al. 2021). Particularly, the acquisition of a high-quality reference genome for eggplant would enable studies of heterosis in this species at the genomic level.

In this study, we created reciprocal hybrids of *S. melongena* and *S. aethiopicum*. The phenotypic descriptions of these hybrids and their parents were analyzed along with their genomic characteristics and genetic variation to provide preliminary indications of the influence of differences in genetic background between the parents on the formation of distant heterosis in eggplant.

2. Materials And Methods

2.1 Plant materials

Plants of the cultivated strain 177 of *S. melongena* (originally collected from Japan) and the wild strain Y11 of *S. aethiopicum* (originally collected from Africa) were obtained from the Vegetable Research Institute, Guangxi Academy of Agricultural Science, Guangxi Province, China (28°N; 118°E). Both strains were purified over 10 generations by selfing and then reciprocally crossed to produce F1 hybrid progeny.

The parents and F1 hybrid progeny were planted at the science research base of Guangxi Academy of Agricultural Science (23°15'8.83"N; 108°03'49.63"E) using deep groove high-grain cultivation.

Traits and growth and development data of the two parental accessions and the reciprocal F1 progeny were analyzed according to the Distinctness, Uniformity, and Stability classification (GB/T 19557.20–2017) and standards used in relation to eggplant germplasms (Li 2006).

2.2 Apical somatic chromosome observation and genome size determination

Seeds were germinated in Petri dishes as described previously (Ranil et al. 2015) and then transferred into seedling trays in a climatic chamber under a photoperiod and temperature regimen of 16-h light (25°C)/8-h dark (18°C). The root tips of the parental and F1 hybrid progeny were dissected, capped, and placed in an ice–water mixture at approximately 0°C for 20–24 h before fixation in Carnoy's solution (absolute ethanol and glacial acetic acid at a ratio of 3:1) for 12–24 h. Next, the samples were dissociated with 1 N hydrochloric acid at 60°C for 10–15 min and stained with modified basic fuchsin for tableting. Under 40× microscopic examination, we counted the number of chromosomes in at least 50 cells from each line. The number of chromosomes in more than 85% of cells was considered the number of chromosomes in that variety.

Fresh, tender, and unfolded eggplant leaves were placed in clean Petri dishes and fragments of approximately 1 cm in length were cut using a sharp blade. Next, 1.5 mL of nuclear extraction buffer was added to each sample and the samples were quickly chopped further. The samples were allowed to stand for 2 min to extract nuclei, and 2 mL PARTEC nuclear staining solution added. After 4 min, the stained filter membrane was filtered into a 3-mL sample tube, and the samples loaded for flow cytometry. During the measurement, the voltage, threshold, and other parameters were optimized, and the coefficient of variation of samples controlled within 5%. Three different individual plants of each species or hybrid were collected, with three technical replicates performed per plant. Using chicken red blood cells as the reference, the external reference method was used. The sample genome value was calculated as follows: $P = (F1/F2) \times H$, where P represents the DNA content of the sample, $F1$ the sample peak fluorescence intensity, $F2$ the chicken red blood cell fluorescence intensity, and H the chicken red blood cell genome value (2.3 pg/N, equivalent to 2249.4 Mb).

2.3 DNA extraction and sequencing

Total genomic DNA (gDNA) was isolated from an individual plantlet of each accession using the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle 1990) with slight modifications. Genomic DNA quantity and quality were using agarose electrophoresis and a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), respectively. The concentration was measured with a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.4 Whole-genome sequencing and quality control

High-quality gDNA was sent to Pfomic Information Technology Co., Ltd. (Nanning, Guangxi, China) for sample processing and sequencing. The DNA samples were randomly fragmented to 350 bp-long fragments using a Covaris ultrasonic disruptor (Woburn, MA, USA), and a complete library was prepared using end repair, A-tailing, sequencing-adaptor ligation, purification, and PCR amplification. Paired-end sequencing libraries with an insert size of approximately 500 bp were built using the Illumina DNA library kit according to the manufacturer's instructions and sequenced on an Illumina HiSeq 2500 platform (San Diego, CA, USA) at 150 bp length.

The original image data generated through sequencing were converted into sequence data via base calling (Illumina pipeline CASAVA v1.8.2). Next, the sequencing data were subjected to quality control to remove unusable reads using the following exclusion criteria: 1) the read contained Illumina library construction adapters; 2) the read contained more than 10% unknown bases; and 3) one end of the read contained more than 50% low-quality bases (sequencing quality value ≤ 5).

2.5 Read alignment and variant calling

A high-quality, chromosome-scale eggplant genome of the cultivar GUIQIE-1, assembled by combining PacBio long reads and Hi-C sequencing data (Li et al. 2021), served as the reference genome sequence for determining variation levels. For the re-sequenced genome, FastQC was used to check the quality of the raw sequencing data (Andrews 2014). Raw reads were trimmed using Trimmomatic (Bolger et al. 2014) read mapping. Sequencing reads were then aligned to the reference genome (GUIQIE-1) using bwa-mem (version 0.7.17) with default parameters (Li and Durbin 2009). Subsequent processing, including duplicate removal, was performed using Samtools (Li et al. 2009) and PICARD (<http://picard.sourceforge.net>). The raw SNP/InDel sets were called by Samtools with the following parameters: -q, 1; -C, 50; -m, 2; -F, 0.002; and -d, 1000. We then filtered these sets using the following criteria: 1) mapping quality > 20; and 2) depth of the variant position > 4. BreakDancer (Chen et al. 2009) and CNVnator (Abyzov et al. 2011) were used to detect SVs and CNVs, respectively. ANNOVAR (Wang et al. 2010) was used for functional annotation of the variants. UCSC known genes were used to generate gene and region annotations.

2.6 Screening and analysis of gene variations

In combination with detecting variations in SNPs, InDels, SVs, and CNVs occurring in the coding sequences (CDS) of the samples, we mined and analyzed the genes with associated functional differences between the progeny from crosses. Then, mutant genes were BLAST compared with functional databases such as NR (Altschul et al. 1997), Swiss-Prot (Boeckmann et al. 2003), GO (Ashburner et al. 2000) and KEGG (Kanehisa et al. 2004), and annotated, screened, and analyzed.

3. Results

3.1 Biparental and hybrid progeny trait statistics

Phenotypic comparisons between hybrids and their parents are the simplest approach for identifying the authenticity of distant hybrids (Jacoby et al. 2003; Prohens et al. 2012). The field characteristics of the plants were investigated from planting-out onwards; the vegetative growth period indices were determined at the beginning of flowering and fruit characteristic indices were determined at commercial maturity. The results revealed a large difference between the 177 and Y11 strains. The characteristics of the reciprocal progenies were inclined toward those of wild eggplants. The two reciprocal progenies did not differ significantly. The corolla color of distant hybrid F1 was intermediate, and its inflorescences were structured, mostly composed of 9–12 flowers, whereas cultivated eggplant generally exhibited a single flower and no inflorescence. Although the related wild strain Y11 had inflorescences, these numbered from 1–4. The color, top shape, and cross-sectional shape of fruits of the distant hybrid F1 were similar to those of wild eggplant when it was the male parent, which is consistent with the results of Prohens et al. (2013). There were no stripes on the fruit surface of either parent but some immature fruits of F1 plants were deformed, with wide, dark green, irregularly distributed stripes. Moreover, the color turned orange after ripening (Table S1, Fig. 1).

3.2 Observation of chromosomes in apical somatic cells and genome size determination

Chromosome number remains relatively stable during plant growth and development and is not easily changed by the external environment; therefore, it can be used as a basis for plant classification and phylogeny and to identify genetic relationships. As shown in Fig. 2, the chromosome number in all parents and reciprocal progenies was $2n = 24$ and the karyotype was normal diploid ($2A$). The chromosome numbers and karyotypes of the parents and their progenies remained unchanged after crossing the cultivated and wild relative eggplants.

Since the relative values were measured using flow cytometry, it is necessary to identify reference plant samples with known genome sizes to determine the genome size of the test samples. In this case, the chicken red blood cell genome is not a suitable internal reference because of its similarity to those of the test samples, hence we used this genome as an external reference in our detection and calibration methods. The results revealed significant differences in the genome sizes between the parents and offspring. Y11 had the largest genome size of 1,312.75 Mbp, while strain 177 had the smallest genome size of 1,013.68 Mbp. The genome sizes of the reciprocal cross progeny were 1,136.14 and 1,185.21 Mb, which were between the parents. Moreover, no significant differences were observed between the parents and progeny based on the analysis of variance (ANOVA; Table 1).

Table 1
Flow cytometry detection results.

Sample	fluorescence intensity	genome size (pg/N)	genome size (M)
Y11	85,417	1.35	1,312.75 ± 12.86 Aa
	83,767	1.33	
	84,732	1.34	
177	64,219	1.02	1,013.68 ± 8.57 Cc
	63,130	1.00	
	68,720	1.09	
Y11×177	74,832	1.19	1,185.21 ± 45.96 Bb
	76,339	1.21	
	78,075	1.24	
177×Y11	72,718	1.15	1,136.14 ± 25.17 Bb
	73,216	1.16	
	73,821	1.17	
Note: Y11: wild African eggplant, <i>S. aethiopicum</i> ; 177: cultivated Japanese eggplant, <i>S. melongena</i> ; 177×Y11: F1 heterozygous progeny of 177×Y11; Y11×177: F1 heterozygous progeny of Y11×177.			

3.3 Sequencing data

The sequencing depth distribution results showed that the sequencing depth was consistent with this study (Fig. S1). A total of 194.753 g of raw data and 194.753 g of clean data were generated (Table 2). The raw data of each sample ranged from 3,2437.028 to 5,6071.642 M. The sequencing quality was high ($Q20 \geq 96.62\%$, $Q30 \geq 90.96\%$), while the GC content ranged from 36.4–37.2%, with an average of 36.81%. Thus, there was no external contamination during WGRS. The clean reads of the four samples were tested for variation, and 3,137,091 InDels, 87,970,955 SNPs, 184,851 SVs, and 128,113 CNVs were obtained (Fig. 3).

Table 2
Data sequencing quality and depth.

Sample	Raw Base (G)	Clean Base (G)	Effective Rate (%)	Q20(%)	Q30(%)	GC Content (%)	Mapping rate (%)	Average depth (X)
Y11	56.00	56.00	99.8	96.76	91.37	37.2	94.21	58.86
177	56.07	56.11	99.94	96.62	90.96	36.4	97.99	53.09
177×Y11	50.24	50.24	100.00	97.84	93.50	36.79	96.82	49.33
Y11×177	32.44	32.44	100.00	97.96	93.81	36.83	96.35	33.32
Note: Sample, name of samples; Raw Base (bp), number of sequencing sequences multiplied by the length of sequences, in bp; Clean Base (G), number of filtered sequences multiplied by the length of sequences, in bp; Effective Rate (%), the ratio of clean data and raw data obtained after filtering; Q20 (%) and Q30 (%), percentages of total bases with Phred values greater than 20 and 30; GC Content (%), percentage of the total number of G and C bases of all the bases; Mapping Rate (%), comparison rate, the number of reads on the reference genome divided by the number of reads of valid sequencing data; Average Depth (X), average sequencing depth, comparing the total number of bases to the reference genome divided by the genome size. Y11: wild African eggplant, <i>S. aethiopicum</i>; 177: cultivated Japanese eggplant, <i>S. melongena</i>; 177×Y11: F1 heterozygous progeny of 177×Y11; Y11×177: F1 heterozygous progeny of Y11×177.								

3.4 SNP and InDel profiling

Based on the whole-genome variation data, the genomes of 177 and Y11 and their hybrid progenies were compared at the DNA level to improve our understanding of the genetic variation characteristics of heterosis in distant hybrid progenies of eggplants. Strain Y11 contained the largest number of SNPs (31,570,174), while strain 177 contained the smallest number of SNPs (6,240,123). The numbers for the reciprocal heterozygote generation were just below that of Y11 (Fig. 4A), with 28,312,016, and 27,879,149 for 177×Y11, and Y11×177, respectively.

The 177×Y11 and Y11×177 heterozygotes showed no significant changes in their phenotypes. The number of shared SNPs was 23,914,108 and the total number of SNP sites found only in one heterozygote was 5,981,994. The number of non-shared SNP sites did not differ significantly between strain Y11 and the heterozygotes, at 9,383,624 and 10,375,794, respectively; the numbers of SNP sites shared between strain Y11 and the hybrids were 25,307,987 (177×Y11) and 24,723,417 (Y11×177). The numbers of shared SNP sites between strain 177 and each heterozygote type did not differ significantly, at 653,734 (177×Y11) and 851,314 (Y11×177). The total numbers of non-shared SNP sites with strain 177 were 268,408,559 (177×Y11) and 26,268,729 (Y11×177). The phenotypic difference between strains Y11 and 177 was large, and the total number of non-shared SNPs was 33,533,229, whereas the number of shared SNPs was 313,758 (Fig. 4A). Regarding mutation types, T:A > C:G was the most frequent mutation type, followed by C:G > T:A (Fig. S2). SNP heterozygosity increased from 0.916 and 9.376% to 23.006 and 23.215%, indicating that the number of SNPs in the hybrid progenies increased (Fig. 5). In

addition, SNP variation was mainly concentrated in the intergenic region, accounting for 44.6–46.3%. The SNP variation in upstream, downstream, intron, and exon regions accounted for 0.07–0.09%, 0.07–0.08%, 23.41–25.69%, and 10.07–12.15%, respectively. The proportion of SNPs in exon regions causing synonymous and nonsynonymous mutations was 0.48–0.49. The number of SNPs causing stop gains was 25,032–98,418, and the number of SNPs causing stop losses was 2,705–12,229 (Fig. 6A; Table S2). The degree of variation of strain 177 was lower than those of strain Y11 and the progeny. Moreover, the progeny variation levels were closer to those of the wild relative strain Y11, which is consistent with the phenotype. Regarding the direction of mutation, transformation was significantly more common than transversion (Fig. 6A; Table S2).

Functional annotation of the InDels showed that they occurred mainly in intergenic regions, accounting for 51–53.56% of the total InDels. The second most common locations were intron regions and either up- or downstream of genes, accounting for 23.48–26% and 19.86–20% of the total InDels, respectively. The number of InDels in the exon regions was relatively small, accounting for only 3–4.18% of the total. The number of variations at splice sites was the smallest, ranging from 0.1–0.12% (Fig. 6B, Table S3). Overall, there was slightly more insertion than deletion mutations across the whole genome. A longer total length of InDels across the whole genome was associated with fewer InDel variations. Single-base mutations (both insertion and deletion) were the most common, accounting for more than 40% of the total number, while InDels with lengths of 6, 9, or 12 were also relatively common (Fig. S3). The results show that the change trend in InDel length and bit number of the four materials was consistent.

There were significant differences in phenotype between strains 177 and Y11, with the number of InDels reflecting this degree of difference, reaching 1,325,898 unshared InDels and only 14,450 shared InDels (the lowest level of commonality) (Fig. 4B). The total number of non-shared InDels between the two heterozygous progenies, 177×Y11 and Y11×177, was lower than that of shared InDels, at 347,475 and 717,409, respectively. The numbers of non-shared InDels between Y11 and each reciprocal heterozygote were 571,582 (177×Y11) and 620,965 (Y11×177). The numbers of shared InDels were 788,757 (177×Y11) and 666,190 (Y11×177). Finally, the numbers of non-shared InDels between strain 177 and each reciprocal heterozygote were 937,230 (177×Y11) and 872,497 (Y11×177), and for shared InDels 53,184 (177×Y11) and 57,749 (Y11×177) (Fig. 4B). Thus, the heterozygous progenies of 177 and Y11, whether from crosses or reciprocal crosses, showed large variations compared with 177, which constitute the fundamental genetic basis of heterosis in distant hybrids.

3.5 SV and CNV profiling

An SV constitutes a large-scale rearrangement of the nucleotide sequence in the genome and may include insertions, deletions, inversions, duplications, and/or heterotopies. A typical SV affects a sequence of around 1 kb to 3 Mb in length, which is larger than a SNP but smaller than a chromosomal abnormality (Feuk et al. 2006). In comparison with the GUIQIE-1 reference genome, the actual insert size and insertion, deletion, inversion, intra-chromosome migration, and inter-chromosome migration numbers in the four samples were determined. A total of 1,844,851 SVs were detected in all samples (excluding those with paired-end reads < 2), and deletions, insertions, and inversions were annotated using

ANNOVAR. The SV length distribution in strain 177 and Y11×177 heterozygotes showed that their SVs were mainly concentrated in two intervals, ranging from 300–400 bp and greater than 1200 bp, whereas those of strain Y11 and 177×Y11 heterozygotes ranged from 200–300 bp, or greater than 1200 bp in length (Fig. S4). Among strains 177, Y11, and their heterozygous progenies, the number of large deletion fragments was highest, accounting for 36.99–47.94%, while intra-chromosome migration and inter-chromosome migration accounted for 25.58–32.36% and 18.81–30.32% of all SVs, respectively (Fig. 7A, Table S4).

CNV refers to a change in the copy number of a genome segment, which can be derived through four primary mechanisms: deletion, insertion, duplication, and recombination. In this study, we detected 128,113 CNVs in the four samples. Analysis of the CNV locations showed that the intergenic region contained the most CNVs, accounting for 36.76–49.63% of the total, followed by the exonic region, accounting for 18.36–49.82% of the total. The fewest CNVs were detected in the upstream and downstream groups, accounting for only 4–8% of the total. With respect to the four plant sources, there was no significant difference in the CNV region distribution between strain Y11, heterozygous progeny of 177×Y11, and the heterozygous progeny of Y11×177. The number of variations in the exonic region of strain 177 was lower than those in the other three sources, whereas the number of variations in the intronic region of strain 177 was significantly higher than those in the other three plants (Fig. 7B, Table S5).

3.6 Mining of variant genes and functional annotation

The progeny of 177×Y11 showed vigorous growth compared with that of their parents. To explore which gene variations were responsible for the emergence of such growth and developmental heterosis, nonsynonymous SNP variants, as well as InDel, CNV, and SV variants in the CDS region were analyzed. A total of 34,870 and 34,868 genes harbored variants in 177×Y11 and Y11×177, respectively. These variants included 33,353 and 33,210 occurrences of nonsynonymous SNP variations, 29,284 and 29,330 occurrences of InDel variations, 22,310 and 25,148 occurrences of CNVs, and 7,343 and 7,182 occurrences of SVs in 177×Y11 and Y11×177, respectively (Fig. S5, Table 3). Moreover, multiple genes were identified as simultaneously harboring multiple types of variants. Using a Venn diagram analysis, we found that the numbers of variant genes exhibiting differential variation between the heterozygous offspring and the two parents, and among the different variant types, the highest number of variant genes caused by CNVs was in the heterozygous offspring, 12,499 (177×Y11) and 15,334 (Y11×177), respectively (Fig. S6, Table S6, Table S7). These results suggest that the generation of heterosis in heterozygous offspring from distant hybridization may be related to genomic CNV.

Table 3
Statistical results of gene variations.

Sample	Genes with nonsynonymous SNPs	Genes with Small InDel	Genes with SV	Genes with CNV	Total gene variations
177×Y11	33,353	29,284	7,343	22,310	34,870
Y11×177	33,120	29,330	7,182	25,148	34,868
Note: non-identical mutations of the same gene were not repeated. 177×Y11: F1 heterozygous progeny of 177×Y11; Y11×177: F1 heterozygous progeny of Y11×177.					

GO classification and KEGG enrichment analysis were further performed on the variant genes. GO classification revealed that genes with different variant types occurring in heterozygous offspring of orthogonal (177×Y11) and backcross (Y11×177) derivation were mainly enriched for biological regulatory processes such as “regulation of RNA metabolic process” (GO: 0051252), “regulation of cellular biosynthetic process” (GO: 0031326), and “carbohydrate metabolic process” (GO: 0005975), and cellular components such as “nucleus” (GO: 0005634), “intrinsic component of membrane” (GO: 0031224), “integral component of membrane” (GO: 0016021), and “cell wall” (GO: 0005618). Moreover, they were enriched for molecular functions such as “ADP binding” (GO: 0043531), “ATPase activity” (GO: 0016887), “transferase activity” (GO: 0016757, GO: 0016758), and “transporter activity” (GO: 0005215) (Table S8). KEGG enrichment analysis results revealed that the variant genes were enriched in a total of 136 metabolic pathways, including “plant-pathogen interaction” (ko04626), “plant hormone signaling transduction” (ko04075), “carbon metabolism” (ko01200), “starch and sucrose metabolism” (ko00500), “ribosome” (ko03010), “MAPK signaling pathway-plant” (ko04016), “circadian rhythms–plant” (ko04712), and “biosynthesis of amino acids” (ko01230) (Table S9).

4. Discussion

Interspecific hybridization occurs frequently and repeatedly and is the source of genetic variation in the resulting F1 hybrid phenotype (Abbott et al. 2013). The novelty of F1 hybrids is attributed to various factors and interactions including physiological responses, changes in gene expression, and inheritance at the molecular level. In this study, deep resequencing was used to identify genomic features and variant loci in representative eggplant types (Asian cultivated eggplant, *S. melongena* 177 and African eggplant, *S. aethiopicum* Y11), and the hybrid progenies of reciprocal crosses between them. This study provides a theoretical basis for further exploring genetic variations in outbred crosses among eggplants, as well as the mechanisms of heterosis.

4.1 Comparison of traits between 177 (*S. melongena*), Y11 (*S. aethiopicum*), and F1 hybrid progenies of reciprocal crosses

More than 2,000 solanaceous plant species are known; however, research regarding their breeding, genetics, and genomics are limited because of their interspecific hybridization incompatibilities (McCammon and Honma 1983; Bletsos et al. 1998; Bletsos et al. 2004; Plazas et al. 2016). The eggplant

strains 177 and Y11 belong to the subgenus *Leptostemonum* but show obvious differences in morphology. In a previous study, fluorescence and cytological studies of hybrid progenies of strains 177 and Y11 revealed good compatibility between them; however, although the F1 hybrid progeny could blossom and bear fruit normally, despite some twisted and deformed examples, the fertility of the fruits was poor, with few or no seeds. This may be related to the post-fertilization obstacle in distant hybridization (Martins et al. 2015). Therefore, these strains provide good material for genetic research into *Solanum* and a good interbreeding platform.

The phenotypes of reciprocal crosses markedly differed from those of their parents, although being closer to that of the 177 wild parent. This similarity arose because most wild-type traits were dominant, which is consistent with previous studies (Costanzo et al. 2005). However, there was no significant difference in phenotype between the reciprocal crosses, indicating that the female parent had no significant effect on the phenotype of the offspring, which is consistent with previous studies of *Arabidopsis* (Hofmann 2012) but is in contrast with the results in maize (Gonzalo et al. 2007).

Distant hybridization leads to genetic instability (Ozkan et al. 2001). Root-tip somatic-cell analysis showed that the chromosome number in all four plant sources was $2n = 24$, with no significant difference in the genomic arrangement among the four plant types. The flow cytometry results differed, which can be explained by the large error in the external parameter calibration methodology used. Although there was little difference among the four plants in chromosomal terms, there were large phenotypic differences among them; this discrepancy requires further study. The hybrid progenies showed strong developmental and resistance advantages, and their appearance was closer to that of perennial shrubs. Thus, they exhibited strong heterosis, partially because of their heterozygosity rate, which was 10-fold higher than that of their parents.

4.2 Genetic variation in strain 177, strain Y11, and hybrid progenies of reciprocal crosses

Deep resequencing can reveal a large number of trait markers, and SNP and InDel markers can be closely linked with specific traits such as disease resistance, quality, phenology, and yield (Ray and Satya 2014); therefore, this approach is used as a conventional tool in plant breeding (Rasheed et al. 2017). Our InDel data analysis confirmed that (excluding InDels with a length of 1 base) the highest proportions of InDels had lengths of 3, 6, 9, 12, 15, and 18 nucleotides, and occurred in exonic regions. This was because deletion of 3 bases would result in deletion of a single amino acid without breaking the overall reading frame. Preferential avoidance of intronic InDels may be a protective mechanism by which organisms limit the extent of changes to coding information owing to alterations in the reading frame, which can reduce mutation-related damage to the organism.

SV and CNV have attracted wide attention in human genetics and are associated with various chronic diseases (Audano et al. 2019; Ho et al. 2020). These variations are similarly abundant in plant genomes and significantly influence phenotype, although the impact of genotypic variation in crop genomics

remains unknown (Saxena et al. 2014). Whole-genome resequencing has identified numerous CNVs in the *Arabidopsis* genome but only a few true CNV polymorphisms are differentially expressed, resulting in genetic variation that may have only a small effect on the phenotype (Xiangchao et al. 2011). Relevant studies showed that SVs and CNVs in the single gene and genome ranges are associated with a variety of plant resistance and other phenotypes (Sutton et al. 2007; Knox et al. 2010; Maron et al. 2013; Würschum et al. 2015; Zhang et al. 2015). Many diseases and phenotypic variations in organisms are related to variations in the genomic structure (Brandler et al. 2016); such variation significantly influences individual characteristics. Additionally, by detecting genome structure variations within a population, biodiversity can be better explored (McHale et al. 2012). Studies of genomic structure variation, along with key gene annotation, can not only clarify the genetic variation mechanism of distant hybridization heterosis but also be used to support modification of the crop genome to obtain better varieties (Abel and Duncavage 2013). In plant genomes, the presence of large-scale CNVs affects plant height, growth and development, flowering time, seed dormancy, metabolic processes, and other characteristics, and is related to plant responses to biotic and abiotic stresses. They are thus likely to be strongly related to plant adaptive evolution, speciation, and heterosis (Yang and Zhang 2015). Here, we detected several SVs and CNVs between the parents and their reciprocal hybrid progenies; however, whether this is related to heterosis requires further analysis.

There are abundant genetic variations among different *Solanum* species at the whole-genome level, which is consistent with the numeric trend of differences among the four variation types, with strains 177 and Y11 showing the largest differences. This is also in accordance with previous research findings showing that these two species cluster into different branches using various markers (Issshiki et al. 2008; Polignano et al. 2010). The number of variants was similar between 177×Y11 and Y11×177 compared to that in strain 177 and similar between 177×Y11 and Y11×177 compared with strain Y11. The phenotypes between 177×Y11 and Y11×177 were not only similar, but also the number of variations was higher than the parents; the specific reasons for this discrepancy should be evaluated further. The variation rate of strain Y11 was highest, that of strain 177 was lowest, and that of the reciprocal hybrid progenies was between those of the parents but tended to be higher in wild relatives, which is consistent with the phenotype.

4.3 Variant gene mining and analysis in strain 177, strain Y11, and hybrid progenies of reciprocal crosses

Genes determine animal and plant biological traits. Nonsynonymous SNPs change nucleotide composition and InDels reflect the insertion or deletion of DNA fragments with different sizes in the genome. Both affect the function of related genes. In addition, SVs and CNVs also constitute important genomic structural variations that can affect genomic stability, and the expression regulation of related genes and, in turn, determine the phenotype of a species (Sebastian et al. 2019; Kumar et al. 2021).

Combining variant gene functional annotation with enrichment analysis results revealed that although heterozygous offspring of orthogonal (177×Y11) and backcross (Y11×177) derivation harbored numerous genes with different variant types, the number of genes with differential variation between the

crosses was not significant, potentially explaining the similarity between the 177×Y11 and Y11×177 heterozygote phenotypes.

Notably, we identified fewer genes with nonsynonymous SNP variations, InDels, or SVs, but more genes with CNVs in the CDS regions of the reciprocal cross heterozygotes compared with those of the two parents, implying that the rearrangements occurred in the genomes of the 177 offspring outbred with Y11, which may have led to gene dosage effects, gene breakage, gene fusion, or position effect, thereby causing heterozygous progeny to develop heterosis for some traits (e.g., plant growth and development, or resistance). In addition, the total number of nucleotides in the genome that are involved in CNV exceeds the total number of SNPs, implying potential effects on gene expression, phenotypic variation, and evolutionary adaptability through changes in gene dose and gene regulatory regions. Accordingly, CNVs represent an important source of genetic variation and phenotypic difference and play an important role in the occurrence and development of complexity (Cesar et al. 2011). We found that more genes were associated with CNVs than other variation types in heterozygous offspring, suggesting an important consequence of distant hybridization.

Variation of the crop genome leads to a change in phenotypic traits, which is fundamentally due to the alteration of gene function (Gaut et al. 2018). For the orthorhombic (177×Y11) and backcross (Y11×177) plants, functional annotation, and enrichment analysis of the variant genes in the heterozygous offspring revealed a predominance of genes with variants involved in “biological regulation”, “cell wall building”, “cytoskeleton synthesis”, “ATPase activity”, and “transferase activity”. Numerous variant genes were also associated with pathways closely related to plant growth and development, such as “plant hormone signal transduction”, “MAPK signaling pathway–plant”, and “circadian rhythm–plant”. Therefore, it is speculated that the growth and development of heterosis exhibited by the heterozygous offspring of distant hybridization between 177 and Y11 may be caused by the functional changes of genes in these metabolic pathways’ by genetic variation. These data provide a basis for further screening of heterosis functional genes in distant hybridization.

In summary, disease resistance is developed by most plant types under long-term natural selection, with distant hybridization introducing resistance genes into cultivars to improve resistance. Distant hybridization can also induce parthenogenesis and sterile male lines, creating new germplasm resources (Peterka et al. 2004). In addition, a series of intermediate types and new species types can appear in the progenies of distant hybridizations, providing a basis for studying their evolutionary history and determining their kinships. Such data will provide insights into the rules governing species or type formation and evolution, which can be used to create new species. Therefore, mapping of the functional genes involved in heterosis should be performed at the DNA level, and the effects of different gene sources on the formation of heterosis should be examined to comprehensively dissect heterosis.

Declarations

Acknowledgments

The authors would like to thank Guangxi Pufei Information Technology Co., LTD for helping to analyze the sequencing data.

Author Contributions

W.L. and Y.W. designed the experiments. W.L. and D.L. performed the observation and analysis of phenotypic traits. Y.J., P.W., and Q.Y. participated in chromosome karyotype experimental analysis. N.Y. and G.G. performed the data analyses and wrote the manuscript. W.L. and Y.W. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Funding

This research was supported by the Guangxi Natural Science Foundation (2018GXNSFDA050015), Guangxi Science and Technology Project (Guike AA22068088), and Basic Special Fund of the Guangxi Academy of Agricultural Sciences (Guinongke 2021YT100, Guinongke 2018YM13).

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Most M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Vainola R, Wolf JB, Zinner D (2013) Hybridization and speciation. *J Evol Biol* 26 (2):229-246. doi:10.1111/j.1420-9101.2012.02599.x
2. Abel HJ, Duncavage EJ (2013) Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. *Cancer Genet* 206 (12):432-440. doi:10.1016/j.cancergen.2013.11.002
3. Abyzov A, Urban AE, Snyder M, Gerstein M (2011) CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 21 (6):974-984. doi:10.1101/gr.114876.110
4. Acquadro A, Barchi L, Gramazio P, Portis E, Vilanova S, Comino C, Plazas M, Prohens J, Lanteri S (2017) Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant

- complexes. PLoS One 12 (7):e0180774. doi:10.1371/journal.pone.0180774
5. Afful NT, Nyadanu D, Akromah R, Amoatey HM, Annor C, Diawouh RG (2018) Evaluation of crossability studies between selected eggplant accessions with wild relatives *S. torvum*, *S. anguivi* and *S. aethopicum* (Shum group). Journal of Plant Breeding and Crop Science 10 (1):1-12. doi:10.5897/jpbcs2017.0695
 6. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25 (17):3389-3402. doi:10.1093/nar/25.17.3389
 7. Andrews S (2014) FastQC A Quality Control tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25 (1):25-29. doi:10.1038/75556
 9. Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, Dougherty ML, Nelson BJ, Shah A, Dutcher SK, Warren WC, Magrini V, McGrath SD, Li YI, Wilson RK, Eichler EE (2019) Characterizing the Major Structural Variant Alleles of the Human Genome. Cell 176 (3):663-675 e619. doi:10.1016/j.cell.2018.12.019
 10. Augustinos AA, Petropoulos C, Karasoulou V, Bletsos F, Papasotiropoulos V (2016) Assessing diversity among traditional Greek and foreign eggplant cultivars using molecular markers and morphometrical descriptors. Span J Agric Res 14 (4). doi:10.5424/sjar/2016144-9020
 11. Barabaschi D, Tondelli A, Desiderio F, Volante A, Vaccino P, Vale G, Cattivelli L (2016) Next generation breeding. Plant Sci 242:3-13. doi:10.1016/j.plantsci.2015.07.010
 12. Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA (2010) Heterosis. Plant Cell 22 (7):2105-2112. doi:10.1105/tpc.110.076133
 13. Bletsos F, Roupakias D, Tsaktsira M, Scaltsoyjanis A (2004) Production and characterization of interspecific hybrids between three eggplant (*Solanum melongena* L.) cultivars and *Solanum macrocarpon* L. Sci Hortic 101 (1-2):11-21. doi:10.1016/j.scienta.2003.09.011
 14. Bletsos FA, Roupakias DG, Thanassouloupoulos CC (1998) Morphological Traits of the Interspecific Eggplant Hybrid *Solanum Melongena* X *Solanum Torvum*. Acta Horticulturae 467 (467):155-162. doi:10.17660/ActaHortic.1998.467.17
 15. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 31 (1):365-370. doi:10.1093/nar/gkg095
 16. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15):2114-2120. doi:10.1093/bioinformatics/btu170
 17. Brandler WM, Antaki D, Gujral M, Noor A, Rosanio G, Chapman TR, Barrera DJ, Lin GN, Malhotra D, Watts AC, Wong LC, Estabillo JA, Gadomski TE, Hong O, Fajardo KV, Bhandari A, Owen R, Baughn M,

- Yuan J, Solomon T, Moyzis AG, Maile MS, Sanders SJ, Reiner GE, Vaux KK, Strom CM, Zhang K, Muotri AR, Akshoomoff N, Leal SM, Pierce K, Courchesne E, Iakoucheva LM, Corsello C, Sebat J (2016) Frequency and Complexity of De Novo Structural Mutation in Autism. *Am J Hum Genet* 98 (4):667-679. doi:10.1016/j.ajhg.2016.02.018
18. Calvo-Asensio I, Prohens J, Gisbert C (2014) Vigor for in vitro culture traits in *S. melongena* x *S. aethiopicum* hybrids with potential as rootstocks for eggplant. *ScientificWorldJournal* 2014:702071. doi:10.1155/2014/702071
 19. Cesar, P, Canales, Katherina, Walz (2011) Copy number variation and susceptibility to complex traits. *EMBO Molecular Medicine* 3 (1):1-4
 20. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, Ley TJ, Wilson RK, Ding L, Mardis ER (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 6 (9):677-681. doi:10.1038/nmeth.1363
 21. Collonnier C, Mulya K, Fock II, Mariska II, Servaes A, Vedel F, Siljak-Yakovlev S, Souvannavong VV, Ducreux G, Sihachakr D (2001) Source of resistance against *Ralstonia solanacearum* in fertile somatic hybrids of eggplant (*Solanum melongena* L.) with *Solanum aethiopicum* L. *Plant Sci* 160 (2):301-313. doi:10.1016/s0168-9452(00)00394-0
 22. Comings DE, MacMurray JP (2000) Molecular heterosis: a review. *Mol Genet Metab* 71 (1-2):19-31. doi:10.1006/mgme.2000.3015
 23. Costanzo S, Simko I, Christ BJ, Haynes KG (2005) QTL analysis of late blight resistance in a diploid potato family of *Solanum phureja* x *S. stenotomum*. *Theor Appl Genet* 111 (3):609-617. doi:10.1007/s00122-005-2053-1
 24. Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002a) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the solanaceae. *Genetics* 161 (4):1697-1711. doi:10.1093/genetics/161.4.1697
 25. Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002b) Conservation of gene function in the solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* 161 (4):1713-1726. doi:10.1093/genetics/161.4.1713
 26. Doyle JJT, Doyle JL (1990) Isolation of Plant DNA from fresh tissue. *Focus* 12
 27. Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7 (2):85-97. doi:10.1038/nrg1767
 28. Fu D, Xiao M, Hayward A, Jiang G, Zhu L, Zhou Q, Li J, Zhang M (2015) What is crop heterosis: new insights into an old topic. *J Appl Genet* 56 (1):1-13. doi:10.1007/s13353-014-0231-z
 29. Fukuoka H, Miyatake K, Nunome T, Negoro S, Shirasawa K, Isobe S, Asamizu E, Yamaguchi H, Ohya A (2012) Development of gene-based markers and construction of an integrated linkage map in eggplant by using *Solanum* orthologous (SOL) gene sets. *Theor Appl Genet* 125 (1):47-56. doi:10.1007/s00122-012-1815-9

30. Gaut BS, Seymour DK, Liu Q, Zhou Y (2018) Demography and its effects on genomic variation in crop domestication. *Nat Plants* 4 (8):512-520. doi:10.1038/s41477-018-0210-1
31. Gisbert C, Prohens J, Raigón MD, Stommel JR, Nuez F (2011) Eggplant relatives as sources of variation for developing new rootstocks: Effects of grafting on eggplant yield and fruit apparent quality and composition. *Sci Hortic* 128 (1):14-22
32. Godfrey S, Nahamya PK, Apolo KK, John NJ, Michael M, Elizabeth BK (2017) Diversity and distribution of African indigenous vegetable species in Uganda. *International Journal of Biodiversity and Conservation* 9 (11):334-341. doi:10.5897/ijbc2017.1120
33. Gonzalo M, Vyn TJ, Holland JB, McIntyre LM (2007) Mapping reciprocal effects and interactions with plant density stress in *Zea mays* L. *Heredity (Edinb)* 99 (1):14-30. doi:10.1038/sj.hdy.6800955
34. Gramazio P, Blanca J, Ziarsolo P, Herraiz FJ, Plazas M, Prohens J, Vilanova S (2016) Transcriptome analysis and molecular marker discovery in *Solanum incanum* and *S. aethiopicum*, two close relatives of the common eggplant (*Solanum melongena*) with interest for breeding. *BMC Genomics* 17 (1):300
35. Gramazio P, Prohens J, Borràs D, Plazas M, Herraiz FJ, Vilanova S (2017) Comparison of transcriptome-derived simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers for genetic fingerprinting, diversity evaluation, and establishment of relationships in eggplants. *Euphytica* 213 (12). doi:10.1007/s10681-017-2057-3
36. Gramazio P, Prohens J, Plazas M, Andujar I, Herraiz FJ, Castillo E, Knapp S, Meyer RS, Vilanova S (2014) Location of chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant. *BMC Plant Biol* 14 (1):350. doi:10.1186/s12870-014-0350-z
37. Grandillo S, Chetelat R, Knapp S, Spooner D, Peralta I, Cammareri M, Perez O, Termolino P, Tripodi P, Chiusano ML, Ercolano MR, Frusciante L, Monti L, Pignone D (2011) *Solanum* sect. *Lycopersicon*. In: Kole C (ed) *Wild Crop Relatives: Genomic and Breeding Resources*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 129-215. doi:10.1007/978-3-642-20450-0_9
38. Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S, Ohyama A, Yamaguchi H, Sato S, Isobe S, Tabata S, Fukuoka H (2014) Draft genome sequence of eggplant (*Solanum melongena* L.): the representative *solanum* species indigenous to the old world. *DNA Res* 21 (6):649-660. doi:10.1093/dnares/dsu027
39. Ho SS, Urban AE, Mills RE (2020) Structural variation in the sequencing era. *Nat Rev Genet* 21 (3):171-189. doi:10.1038/s41576-019-0180-9
40. Hofmann NR (2012) A global view of hybrid vigor: DNA methylation, small RNAs, and gene expression. *Plant Cell* 24 (3):841. doi:10.1105/tpc.112.240312
41. Holá D, Benešová M, Fischer L, Haisel D, Hnilička F, Hniličková H, Jedelský PL, Kočová M, Procházková D, Rothová O (2017) The disadvantages of being a hybrid during drought: A combined analysis of plant morphology, physiology and leaf proteome in maize. *PLoS One* 12 (4):e0176121
42. Isshiki S, Iwata N, Khan MMR (2008) ISSR variations in eggplant (*Solanum melongena* L.) and related *Solanum* species. *Sci Hortic* 117 (3):186-190

43. Jacoby A, Labuschagne MT, Viljoen CD (2003) Genetic relationships between Southern African *Solanum retroflexum* Dun. and other related species measured by morphological and DNA markers. *Euphytica* 132 (1):109-113
44. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32 (Database issue):D277-280. doi:10.1093/nar/gkh063
45. Kaushik P, Prohens J, Vilanova S, Gramazio P, Plazas M (2016) Phenotyping of Eggplant Wild Relatives and Interspecific Hybrids with Conventional and Phenomics Descriptors Provides Insight for Their Potential Utilization in Breeding. *Front Plant Sci* 7:677. doi:10.3389/fpls.2016.00677
46. Knapp S, Vorontsova MS, Prohens J (2013) Wild relatives of the eggplant (*Solanum melongena* L.: Solanaceae): new understanding of species names in a complex group. *PLoS One* 8 (2):e57039. doi:10.1371/journal.pone.0057039
47. Knox AK, Dhillon T, Cheng H, Tondelli A, Pecchioni N, Stockinger EJ (2010) CBF gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. *Theor Appl Genet* 121 (1):21-35. doi:10.1007/s00122-010-1288-7
48. Kumar H, Panigrahi M, Saravanan KA, Rajawat D, Parida S, Bhushan B, Gaur GK, Dutt T, Mishra BP, Singh RK (2021) Genome-wide detection of copy number variations in Tharparkar cattle. *Anim Biotechnol*:1-8. doi:10.1080/10495398.2021.1942027
49. Lester RN (1986) Taxonomy of Scarlet Eggplants, *Solanum Aethiopicum* L. *Acta Horticulturae* 182 (182):125-132. doi:10.17660/ActaHortic.1986.182.15
50. Li D, Qian J, Li W, Yu N, Gan G, Jiang Y, Li W, Liang X, Chen R, Mo Y, Lian J, Niu Y, Wang Y (2021) A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. *Mol Ecol Resour* 21 (4):1274-1286. doi:10.1111/1755-0998.13321
51. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14):1754-1760. doi:10.1093/bioinformatics/btp324
52. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16):2078-2079. doi:10.1093/bioinformatics/btp352
53. Li X (2006) Description specification and data standard for eggplant germplasm resources. China Agriculture Press, Beijing, China
54. Li Z, Peng T, Xie Q, Han S, Tian J (2010) Mapping of QTL for tiller number at different stages of growth in wheat using double haploid and immortalized F2 populations. *J Genet* 89 (4):409-415. doi:10.1007/s12041-010-0059-1
55. Maron LG, Guimaraes CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, Buckler ES, Coluccio AE, Danilova TV, Kudrna D, Magalhaes JV, Pineros MA, Schatz MC, Wing RA, Kochian LV (2013) Aluminum tolerance in maize is associated with higher MATE1 gene copy number. *Proc Natl Acad Sci U S A* 110 (13):5241-5246. doi:10.1073/pnas.1220766110

56. Martins KC, Pereira TNS, Souza SAM, Rodrigues R, Martins KC, Pereira TNS, Souza SAM, Rodrigues R (2015) Crossability and evaluation of incompatibility barriers in crosses between *Capsicum* species. *Crop Breeding & Applied Biotechnology* 15 (3):139-145
57. McCammon KR, Honma S (1983) Morphological and cytogenetic analyses of an interspecific hybrid eggplant, *Solanum melongena* x *Solanum torvum*. *Hortscience* 18:894-895
58. McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddelloh JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol* 159 (4):1295-1308. doi:10.1104/pp.112.194605
59. Meyer RC, Kusterer B, Lisec J, Steinfath M, Becher M, Scharr H, Melchinger AE, Selbig J, Schurr U, Willmitzer L, Altmann T (2010) QTL analysis of early stage heterosis for biomass in *Arabidopsis*. *Theor Appl Genet* 120 (2):227-237. doi:10.1007/s00122-009-1074-6
60. Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005) The SOL Genomics Network: a comparative resource for *Solanaceae* biology and beyond. *Plant Physiol* 138 (3):1310-1317. doi:10.1104/pp.105.060707
61. Muñoz-Falcón JE, Prohens J, Vilanova S, Nuez F (2009) Diversity in commercial varieties and landraces of black eggplants and implications for broadening the breeders' gene pool. *Ann Appl Biol* 154 (3):453-465. doi:10.1111/j.1744-7348.2009.00314.x
62. Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13 (8):1735-1747. doi:10.1105/tpc.010082
63. Pan Y, Qu S, Bo K, Gao M, Haider KR, Weng Y (2017) QTL mapping of domestication and diversifying selection related traits in round-fruited semi-wild Xishuangbanna cucumber (*Cucumis sativus* L. var. xishuangbannanensis). *Theor Appl Genet* 130 (7):1531-1548. doi:10.1007/s00122-017-2908-2
64. Peterka H, Budahn H, Schrader O, Ahne R, Schutze W (2004) Transfer of resistance against the beet cyst nematode from radish (*Raphanus sativus*) to rape (*Brassica napus*) by monosomic chromosome addition. *Theor Appl Genet* 109 (1):30-41. doi:10.1007/s00122-004-1611-2
65. Plazas M, Vilanova S, Gramazio P, Rodriguez-Burruezo A, Fita A, Herraiz FJ, Ranil R, Fonseka R, Nirani L, Fonseka H, Kouassi B, Kouassi A, Kouassi A, Prohens J (2016) Interspecific Hybridization between Eggplant and Wild Relatives from Different Genepools. *J Am Soc Hortic Sci* 141 (1):34-44
66. Polignano G, Ugenti P, Bisignano V, Gatta CD (2010) Genetic divergence analysis in eggplant (*Solanum melongena* L.) and allied species. *Genetic Resources & Crop Evolution* 57 (2):171-181
67. Prohens J, Plazas M, Raigón MD, Seguí-Simarro JM, Stommel JR, Vilanova S (2012) Characterization of interspecific hybrids and first backcross generations from crosses between two cultivated eggplants (*Solanum melongena* and *S. aethiopicum* Kumba group) and implications for eggplant breeding. *Euphytica* 186 (2):517-538. doi:10.1007/s10681-012-0652-x
68. Prohens J, Whitaker BD, Plazas M, Vilanova S, Hurtado M, Blasco M, Gramazio P, Stommel JR (2013) Genetic diversity in morphological characters and phenolic acids content resulting from an

- interspecific cross between eggplant, *Solanum melongena*, and its wild ancestor (*S. incanum*). *Ann Appl Biol* 162 (2):242-257. doi:10.1111/aab.12017
69. Ranil RHG, Niran HML, Plazas M, Fonseka RM, Fonseka HH, Vilanova S, Andújar I, Gramazio P, Fita A, Prohens J (2015) Improving seed germination of the eggplant rootstock *Solanum torvum* by testing multiple factors using an orthogonal array design. *Sci Hortic* 193:174-181. doi:10.1016/j.scienta.2015.07.030
70. Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, He Z (2017) Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. *Mol Plant* 10 (8):1047-1064. doi:10.1016/j.molp.2017.06.008
71. Rasmus N, Paul JS, Anders A, Yun S, Song (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* 12 (x):443
72. Rattan P, Kumar S, Salgotra RK, Samnotra RK, Sharma F (2014) Development of interspecific F1 hybrids (*Solanum melongena* × *Solanum khasianum*) in eggplant through embryo rescue technique. *Plant Cell, Tissue and Organ Culture (PCTOC)* 120 (1):379-386. doi:10.1007/s11240-014-0591-4
73. Ray S, Satya P (2014) Next generation sequencing technologies for next generation plant breeding. *Front Plant Sci* 5:367. doi:10.3389/fpls.2014.00367
74. Rotino GL, Sala T, Toppino L (2014) Eggplant. In: Pratap A, Kumar J (eds) *Alien Gene Transfer in Crop Plants, Volume 2: Achievements and Impacts*. Springer New York, New York, NY, pp 381-409. doi:10.1007/978-1-4614-9572-7_16
75. Saxena RK, Edwards D, Varshney RK (2014) Structural variations in plant genomes. *Brief Funct Genomics* 13 (4):296-307. doi:10.1093/bfpg/elu016
76. Sebastian, Soyk, Zachary, Lemmon, Fritz, Sedlazeck, José (2019) Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato. *Nature Plants*
77. Shackleton CM, Pasquini MW, Drescher AW (2009) *African Indigenous Vegetables in Urban Agriculture*. Routledge, London. doi:10.4324/9781849770019
78. Shang L, Ma L, Wang Y, Su Y, Wang X, Li Y, Abduweli A, Cai S, Liu F, Wang K, Hua J (2016) Main Effect QTL with Dominance Determines Heterosis for Dynamic Plant Height in Upland Cotton. *G3 (Bethesda)* 6 (10):3373-3379. doi:10.1534/g3.116.034355
79. Song B, Song Y, Fu Y, Kizito EB, Kamenya SN, Kabod PN, Liu H, Muthemba S, Kariba R, Njuguna J, Maina S, Stomeo F, Djikeng A, Hendre PS, Chen X, Chen W, Li X, Sun W, Wang S, Cheng S, Muchugi A, Jamnadass R, Shapiro HY, Van Deynze A, Yang H, Wang J, Xu X, Odeny DA, Liu X (2019) Draft genome sequence of *Solanum aethiopicum* provides insights into disease resistance, drought tolerance, and the evolution of the genome. *Gigascience* 8 (10):giz115. doi:10.1093/gigascience/giz115
80. Sunseri F, Polignano GB, Alba V, Lotti C, Bisignano V, Mennella G, D'Alessandro A, Bacchi M, Riccardi P, Fiore MC (2010) Genetic diversity and characterization of African eggplant germplasm collection. *African Journal of Plant Science* 4:231-241

81. Sutton T, Baumann U, Hayes J, Collins NC, Shi BJ, Schnurbusch T, Hay A, Mayo G, Pallotta M, Tester M, Langridge P (2007) Boron-toxicity tolerance in barley arising from efflux transporter amplification. *Science* 318 (5855):1446-1449. doi:10.1126/science.1146853
82. Toppino L, Valè G, Rotino GL (2008) Inheritance of Fusarium wilt resistance introgressed from *Solanum aethiopicum* Gilo and *Aculeatum* groups into cultivated eggplant (*S. melongena*) and development of associated PCR-based markers. *Mol Breed* 22 (2):237-250. doi:10.1007/s11032-008-9170-x
83. Vorontsova MS, Stern S, Bohs L, Knapp S (2013) African spiny *Solanum* (subgenus *Leptostemonum*, Solanaceae): a thorny phylogenetic tangle. *Bot J Linnean Soc* 173 (2):176-193
84. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38 (16):e164. doi:10.1093/nar/gkq603
85. Wei Q, Wang J, Wang W, Hu T, Hu H, Bao C (2020a) A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Hortic Res* 7 (1):153. doi:10.1038/s41438-020-00391-0
86. Wei Q, Wang W, Hu T, Hu H, Wang J, Bao C (2020b) Construction of a SNP-Based Genetic Map Using SLAF-Seq and QTL Analysis of Morphological Traits in Eggplant. *Front Genet* 11:178. doi:10.3389/fgene.2020.00178
87. Wu F, Eannetta NT, Xu Y, Tanksley SD (2009) A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. *Theor Appl Genet* 118 (5):927-935. doi:10.1007/s00122-008-0950-9
88. Würschum T, Boeven PHG, Langer SM, Longin CFH, Leiser WL (2015) Multiply to conquer: Copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. *BMC Genetics*,16,1(2015-07-29) 16 (1):96
89. Xiangchao G, Oliver S, Jonas B, Steffen JG, Philipp D, Hildebrand KL, Rune L, Schultheiss SJ, Osborne EJ, Sreedharan VT (2011) Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature*
90. Yang H, Zhang D (2015) Copy Number Variations in Plant Genomes. *Molecular Plant Breeding*
91. Yang X, Cheng YF, Deng C, Ma Y, Wang ZW, Chen XH, Xue LB (2014) Comparative transcriptome analysis of eggplant (*Solanum melongena* L.) and turkey berry (*Solanum torvum* Sw.): phylogenomics and disease resistance analysis. *BMC Genomics* 15 (1):412. doi:10.1186/1471-2164-15-412
92. Yates HE, Frary A, Doganlar S, Frampton A, Eannetta NT, Uhlig J, Tanksley SD (2004) Comparative fine mapping of fruit quality QTLs on chromosome 4 introgressions derived from two wild tomato species. *Euphytica* 135 (3):283-296. doi:10.1023/b:Euph.0000013314.04488.87
93. Yuan Q, Deng Z, Peng T, Tian J (2012) QTL-based analysis of heterosis for number of grains per spike in wheat using DH and immortalized F2 populations. *Euphytica* 188 (3):387-395. doi:10.1007/s10681-012-0694-0

94. Zhang Z, Mao L, Chen H, Bu F, Li G, Sun J, Li S, Sun H, Jiao C, Blakely R, Pan J, Cai R, Luo R, Van de Peer Y, Jacobsen E, Fei Z, Huang S (2015) Genome-Wide Mapping of Structural Variations Reveals a Copy Number Variant That Determines Reproductive Morphology in Cucumber. *Plant Cell* 27 (6):1595-1604. doi:10.1105/tpc.114.135848

Figures

Figure 1

Stem, leaf, flower, and fruit morphologies of 177, Y11 and their reciprocal plants. Scale bar = 1 cm.

Figure 2

Chromosomes of the four plant materials. Chromosomes of (A) 177 (cultivated Japanese eggplant, *S. melongena*), (B) Y11 (wild African eggplant, *S. aethiopicum*), (C) F1 heterozygous progeny of 177×Y11, and (D) F1 heterozygous progeny of Y11×177. The chromosomes were observed under 40× magnification. Scale bar = 10 μm.

Figure 3

Distribution of structural variations in the genomes of (A) 177 (cultivated Japanese eggplant, *S. melongena*), (B) Y11 (wild African eggplant, *S. aethiopicum*), (C) F1 heterozygous progeny of 177×Y11, and (D) F1 heterozygous progeny of Y11×177. Shown from the outermost to innermost: Chromosome, single-nucleotide polymorphism (SNP), insertion/deletion (InDel), copy number variant (CNV) duplication, CNV deletion, structural variation (SV) insertion, SV deletion, SV inversion, SV intra-chromosomal translocation (ITX), and SV inter-chromosomal translocation (CTX).



Figure 4

Single-nucleotide polymorphism (SNP) and insertion/deletion (InDel) comparisons between the four samples. The numbers of SNP (A) and InDel (B) loci shared and distinct between the samples. Y11: wild African eggplant, *S. aethiopicum*; 177: cultivated Japanese eggplant, *S. melongena*; 177xY11: F1 heterozygous progeny of 177xY11; Y11x177: F1 heterozygous progeny of Y11x177.

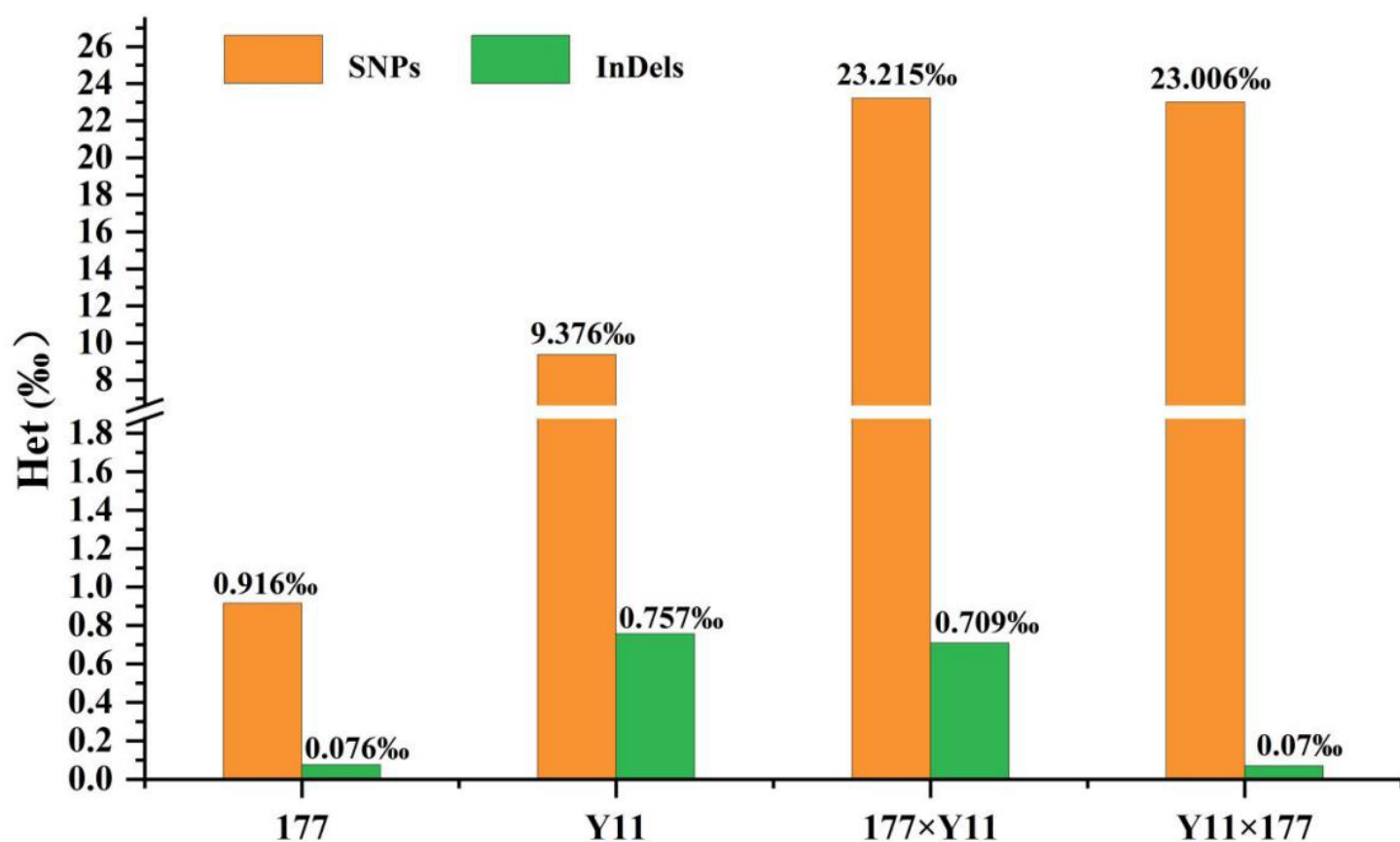


Figure 5

Heterozygosity rates of single-nucleotide polymorphisms (SNPs) and insertion/deletions (InDels). The Het rate indicates the proportion of heterozygosity in the whole genome, calculated as the number of heterozygous SNPs or InDels/genome size. Y11: wild African eggplant, *S. aethiopicum*; 177: cultivated Japanese eggplant, *S. melongena*; 177×Y11: F1 heterozygous progeny of 177×Y11; Y11×177: F1 heterozygous progeny of Y11×177.

Figure 6

Distribution of single-nucleotide polymorphisms (SNPs) and insertion/deletion mutations (InDels) by genomic region in the four samples. (A) SNP number comparison between samples; (B) InDel number comparison between samples. Upstream: 1 kb region upstream of a gene; exonic: within an exon region; intronic: within an intron region; splicing: within 2 bp of an exon/intron boundary; downstream: 1 kb region downstream of a gene; Upstream/Downstream: located within 1 kb upstream of one gene and within 1 kb region downstream of another gene; Intergenic: located in the intergenic region; ts: Transitions; tv: Transversions; Y11: wild African eggplant, *S. aethiopicum*; 177: cultivated Japanese eggplant, *S. melongena*; 177×Y11: F1 heterozygous progeny of 177×Y11; Y11×177: F1 heterozygous progeny of Y11×177.

Figure 7

Structural variations (SVs) (A) and copy number variations (CNVs) (B) in each genome. Upstream: 1 kb region upstream of a gene; Exonic: within an exon region; Intronic: within an intron region; Splicing: within 2 bp of an exon/intron boundary; Downstream: 1 kb region downstream of a gene; Upstream/Downstream: located within 1 kb upstream of one gene and also within 1 kb region downstream of another gene; Intergenic: located in the intergenic region; INS: insertion; DEL: deletion; INV: inversion; ITX: intra-chromosomal translocation; CTX: inter-chromosomal translocation; Y11: wild African eggplant, *S. aethiopicum*; 177: cultivated Japanese eggplant, *S. melongena*; 177×Y11: F1 heterozygous progeny of 177×Y11; Y11×177: F1 heterozygous progeny of Y11×177.

Supplementary Files

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

- [SupplementaryFigures.docx](#)
- [SupplementaryTableS1.xlsx](#)
- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS3.xlsx](#)

- [SupplementaryTableS4.xlsx](#)
- [SupplementaryTableS5.xlsx](#)
- [SupplementaryTableS6.xlsx](#)
- [SupplementaryTableS7.xlsx](#)
- [SupplementaryTableS8.xlsx](#)
- [SupplementaryTableS9.xlsx](#)
- [SupplementaryTableappendix.docx](#)