A High-Quality Melon Genome Assembly Provides Insights into Genetic Basis of Fruit Trait Improvement

Hong Zhang, Xuming Li, Haiyan Yu, ..., Jianshun Miao, Hongkun Zheng, Hongping Yi

zhenghk@biomarker.com.cn (H.Z.)
hpyi1223@163.com (H.Y.)

HIGHLIGHTS
Provides a high-quality assembly for melon genome
Explains a considerable proportion of epidermis thickness
Melons in China are introduced from different routes
Haplotypes of alleles associated with agronomic traits enable efficient breeding

DATA AND CODE AVAILABILITY
SUB4523929

Zhang et al., iScience 22, 16–27
December 20, 2019 © 2019
The Author(s).
https://doi.org/10.1016/j.isci.2019.10.049
A High-Quality Melon Genome Assembly Provides Insights into Genetic Basis of Fruit Trait Improvement

Hong Zhang,1,4 Xuming Li,3,4 Haiyan Yu,3,4 Yongbing Zhang,1 Meihua Li,1 Haojie Wang,1 Dengming Wang,1 Huaisong Wang,2 Qiushi Fu,2 Min Liu,3 Changmian Ji,3 Liming Ma,3 Juan Tang,3 Song Li,3 Jianshun Miao,3 Hongkun Zheng,3,* and Hongping Yi1,5,*

SUMMARY
Accurate reference genomes have become indispensable tools for characterization of genetic and functional variations. Here we generated a high-quality assembly of the melon Payzawat using a combination of short-read sequencing, single-molecule real-time sequencing, Hi-C, and a high-density genetic map. The final 12 chromosome-level scaffolds cover ~94.13% of the estimated genome (398.57 Mb). Compared with the published DHL92 genome, our assembly exhibits a 157-fold increase in contig length and remarkable improvements in the assembly of centromeres and telomeres. Six genes within STHQF12.4 on pseudochromosome 12, identified from whole-genome comparison between Payzawat and DHL92, may explain a considerable proportion of the skin thickness. In addition, our population study showed that melon domesticated at multiple times from whole-genome perspective and melons in China are introduced from different routes. Selective sweeps underlying the genes related to desirable traits, haplotypes of alleles associated with agronomic traits, and the variants from resequencing data enable efficient breeding.

INTRODUCTION
Melon (Cucumis melo L.) is an economically important crop. The annual yield is estimated to be about 30 million tons worldwide, with China contributing to more than half of global melon production (http://www.fao.org). Melon is a eudicot diploid plant species (2n = 2x = 24), belonging to the Cucurbitaceae family. Melon is cultivated worldwide and mainly grows in temperate, subtropical, and tropical regions.

Cucumis melo L. is classified into two sub-species: melo and agrestis. The origin of C. melo remains controversial as wild relatives can be found in Africa and Asia (Diaz et al., 2017). The sub-species agrestis is found mainly in Asia from India to the Far-East and in Africa and Central America. The sub-species melo is found in India, Central and Western Asia, Africa, Europe, and America (Pitrat, 2013). Wild melons are observed in these two sub-species. Until recently, independent domestication of wild melons in Africa and Asia resolved the geographic origin of melon (Endl et al., 2018). The traits in fruit morphology and quality are diverse in different horticultural groups. The wild and cultivated melons show contrasting phenotypic traits. The fruit, flower, and seed of wild melons are small, and the plant is monoecious. The fruit of wild melons is bitter, differing from the sweetness of cultivar melons. These are domestication traits and evidently different in wild and cultivar accessions.

The draft genome of melon was released in 2012 (Garcia-Mas et al., 2012); however, it was sequenced using the short reads, resulting in a highly fragmented genome assembly. Although efforts had been made to anchor the genome into pseudochromosomes using genetic maps (Argyris et al., 2015), their marker coverage was not sufficient to anchor many gene-rich scaffolds or correct mistakes made in the genome assembly. Moreover, short-read sequencing has difficulty in traversing complex repeat structures, leading to incomplete gene models, less accurate representation of repeats, and biases in our understanding of genome biology (Gordon et al., 2016). The lack of an accurate genome assembly also hinders many studies that rely on genome sequences, such as resequencing, transcriptome, and genetic mapping. Until now, the current version of melon genome (v3.6.1) for DHL92 has more than 40,000 sequence gaps (Ruggieri et al., 2018), making it highly fragmented.
The single-molecule real-time (SMRT) sequencing technology allows researchers to obtain long reads with tens of kilobases in size that can span repeat-rich genomic regions, thus significantly reducing sequencing gaps. The completeness of a variety of genome assemblies have been improved by using the SMRT technology, such as quinoa (Jarvis et al., 2017), goat (Bickhart et al., 2017), gorilla (Gordon et al., 2016), and citrus (Wang et al., 2017). In this study, we combined the SMRT sequencing and high-throughput chromosome interaction mapping (Hi-C) technologies to obtain a high-quality genome assembly of a cultivated melon species named Payzawat, a most typical representative of Cucumis melo sp. melo in China. Furthermore, resequencing a number of accessions allows a better understanding of the evolution of melon and provides candidate genes controlling phenotypic traits involved in domestication or diversification.

RESULTS
De Novo Sequencing and Assembly
The SMRT sequencing technologies were used to assemble a high-quality reference genome for melon. A total of 27.80 Gb error-corrected PacBio sub-reads were individually assembled into the contigs with Canu (v1.5) (Koren et al., 2017) assembler and DBG2LOC (Ye et al., 2016), respectively (see Table S2). The resulting two draft assemblies (365 and 389 Mb, see Table S3) then were merged into a consensus assembly. The consensus was polished with PacBio sub-reads and Illumina reads. The estimated genome size for Payzawat was 398.57 Mb using kmer analysis (see Figure S1). As a result, the final genome assembly was 386 Mb with contig N50 and N90 of 2.86 Mb and 511.5 Kb, respectively (see Table S3). This assembly showed substantial improvements compared with the latest assembly of CM3.6.1 (Garcia-Mas et al., 2012), including in terms of assembly completeness, contiguity (gap filling), TE content, etc. With the aid of Hi-C interaction data, 95.53% (363.76 Mb) of the assembly was anchored onto 12 pseudochromosomes (see Table S4 and Figure 1). The Hi-C heatmap of the genome using Hi-C plotter (Akdemir and Chin, 2015) at 200-kb resolution showed the high accuracy of melon assembly (see Figure S2A).

Assembly Assessment
To assess the accuracy of the anchored pseudochromosomes, the genetic markers yielded from an RIL population containing 119 individuals were mapped onto the genome assembly. The results showed that the orders of these markers were highly consistent with the pseudochromosomes (see Figure S2B).

The conserved plant gene sets such as CEG and BUSCO and the specific repeats in melon were also used to evaluate the quality of the genome assembly. A total of 448 (97.82% of 458) CEGs, including 246 (99.19% of 248) highly conserved CEGs, could be found in the melon genome (see Table S7). BUSCO (Simao et al., 2015) assessment showed that 1,336 (92.78%) of gene models were complete, 30 (4.03%) gene models were fragmented, and 76 (5.14%) gene models were missing (Table S8). Melon-specific CentM centromeric repeats (GenBank accession no. 3929695) were found in all 12 chromosomes with a maximum length of 26,811 bp (see Figure S3 and Table S10). In CM3.6.1 genome, centromeric repeats were found only in ten chromosomes with a maximum length of 8,839 bp (see Figure S3 and Table S11). The maximum length of telomeres in Payzawat was 33,998 bp other than 118 bp in CM3.6.1 (see Figure S3 and Tables S12 and S13). In addition, 99.03% of the short reads and 71.82% of the Pacbio sub-reads could be correctly remapped to the assembly with 90.67% and 70% coverage, respectively. All these data verified the high-quality assembly of the Payzawat genome.

Genome Annotation
A total of 192.4 Mb repetitive sequences were identified in Payzawat genome, of which 83.2% (160 Mb) were retrotransposons and 10.02% (38.7 Mb) were DNA transposons (see b in Figure 1 and Table S14).

Protein-coding genes were predicted by integrating evidence from RNA sequencing (RNA-seq)-based prediction, protein-homology-based prediction, and ab initio prediction. A total of high-quality 22,924 gene models were predicted in the Payzawat genome (see Table S15 and Figure S4), of which 22,298 (97.27%) genes were supported by RNA-seq or homolog evidence (see Figure S4 and Table S17) and 22,506 (98.18%) could be annotated by at least one of the public databases (see Table S20). The length distributions of mRNA, CDS, and intron were similar to species within the Cucurbitaceae family (see Figures S5A–S5C). The average length of gene model is 4,491 bp (see Tables 1 and S16), and the median length of mRNA is 3,166 bp in Payzawat longer than 2,326 bp in CM3.6.1. The median number of exons is four in Payzawat but three in CM3.6.1 (see Table 1). In total, 22,037 (96%) of protein-coding genes were allocated to 12
pseudochromosomes. The density of genes tended to be higher and the density of repeats tended to be lower at both ends of the chromosome (see a and b in Figure 1).

**Whole-Genome Comparison between Payzawat and CM3.6.1 Genome**

Genomic collinearity analysis between Payzawat and CM3.6.1 revealed that 76.7% (296.59 Mb) of Payzawat genomes have one-to-one syntenic blocks with 71.1% (296.29 Mb) of CM3.6.1, and these syntenic blocks accommodated 83.36% (19,110) of genes in Payzawat and 89.63% (26,869) in CM3.6.1, respectively (see Figure 2, and Table S23 and Figure S6).
A total of 1,761,822 SNPs and 735,486 indels were identified in syntenic blocks between the two genomes, respectively. Among them, 421,364 SNPs were located in the genic region (9,795 genes) and 33,636 in coding regions (6,971 genes) and 172,760 indels were located in the genic region (9,777 genes) and 4,548 in the coding region (1,842 genes) (Table S24, Figure 3A). For the SNPs within the coding regions, 3,128 SNPs (affecting 796 genes) were non-synonymous or lead to changes at the start or stop codon. These genes were enriched in regulating fruit flavor (diterpenoid biosynthesis and the sulfur relay system) and cell wall biosynthesis (glycosylphosphatidylinositol [GPI]-anchor biosynthesis). The ratio of triple indels on the coding regions is higher than that on the non-coding regions (Figure 3B). The counts of SNPs and indels are higher on chr04, chr07, chr08, chr09, and chr12 (see Figure 3A). Genes with indels causing frameshift were mainly enriched in pathways relating to the cell wall (xylem development, glucuronoxylan metabolic process, xyloglucan biosynthetic process, and D-xylose transport), flavor (organic substance transport and vacuolar transport), sugar metabolism (galactose transport, mannitol transport, sorbitol transport, and galactose transmembrane transporter activity), and fruit color (narigenin 3-dioxigenase activity). Some 49,394 structural variations, including GAP, JMP, INV, and SEQ, were detected on 5,852 genes (see Tables S25 and S26). In addition, 6,955 structural variations were detected in one-to-one syntenic blocks and 1,186 genes were affected. Among the genes located in syntenic blocks, approximately 5,181 genes (23.51%, 22,037 in pseudochromosomes) of Payzawat showed variations in coding regions.

The skin (epidermis) of DHL92 (genome version CM3.6.1) is thin, whereas that of Payzawat is thick (see Figures S7A and S7B). To reveal the candidate genes involved in skin thickness, we analyzed the epidermis thickness-related quantitative trait loci (QTLs) (Diaz et al., 2011) and genes harboring SNPs, indels, and structural variations within these regions. QTL STHQF12.4, controlling the skin thickness, is located at 23,803,761–26,907,220 of chr12, consisting of 455 genes. Six genes (EVM0008156: pectinesterase, EVM0009214: gibberellin 20 oxidase 1-B-like, EVM0014824: pectinesterase, EVM0014071: pectinesterase, and EVM0004080: gibberellin 20 oxidase 1-B-like) with structural variants (SVs) were implicated in cell wall biosynthesis (see Figure 3C). The insertion of 76 bp on exon1 of EVM0014842 on DHL92 leads to the truncated protein, and the variation on a region encompassing the start codon of EVM0014071 and EVM0016862 disables the function of these genes on DHL92 (Figure 3D). The sequence alignment showed the difference between the two genomes (see Figure S9). The other three genes in these two genomes were highly divergent. The variation status of six genes was checked in 49 re-sequencing samples, and the group information is shown in Figure 4B. The ratios of genes with variation in the TC (thick epidermis) group (group three) were higher compared with the TN (thin epidermis) group, such as EVM00008156, EVM0009214, and EVM0014071. Variation of EVM0016862 only can be detected in R13 (African horned melon [Cucumis metuliferus]) (see Figure S10). The variants for EVM0016862 between Payzawat and DHL92 may not be real due to the poor assembly quality of the gene. 15 samples downloaded from NCBI (see Table S29), consisting of thick- and thin-skinned melons, were used to test the expression of

### Table 1. The Statistics of Protein-Coding Genes in Payzawat and DHL92

|                     | Payzawat (C. melo L. cv. inodorus) | DHL92 (C. melo L. cv. DHL92, CM3.6.1) |
|---------------------|-----------------------------------|---------------------------------------|
| Number of genes     | 22,924                            | 29,980                                |
| Number of mRNAs     | 22,924                            | 29,980                                |
| Number of 5’UTR      | 17,419                            | 23,703                                |
| Number of 3’UTR      | 16,078                            | 22,546                                |
| Median mRNA length   | 3,166                             | 2,326                                 |
| Median CDS length    | 452                               | 321                                   |
| Median intron length | 2,227                             | 3,046                                 |
| Median exon number   | 4                                 | 3                                     |
| Percentage of genome covered by genes | 26.64 | 25.85 |

The skin (epidermis) of DHL92 (genome version CM3.6.1) is thin, whereas that of Payzawat is thick (see Figures S7A and S7B). To reveal the candidate genes involved in skin thickness, we analyzed the epidermis thickness-related quantitative trait loci (QTLs) (Diaz et al., 2011) and genes harboring SNPs, indels, and structural variations within these regions. QTL STHQF12.4, controlling the skin thickness, is located at 23,803,761–26,907,220 of chr12, consisting of 455 genes. Six genes (EVM0008156: pectinesterase, EVM0009214: gibberellin 20 oxidase 1-B-like, EVM0014824: pectinesterase, EVM0014071: pectinesterase, and EVM0004080: gibberellin 20 oxidase 1-B-like) with structural variants (SVs) were implicated in cell wall biosynthesis (see Figure 3C). The insertion of 76 bp on exon1 of EVM0014842 on DHL92 leads to the truncated protein, and the variation on a region encompassing the start codon of EVM0014071 and EVM0016862 disables the function of these genes on DHL92 (Figure 3D). The sequence alignment showed the difference between the two genomes (see Figure S9). The other three genes in these two genomes were highly divergent. The variation status of six genes was checked in 49 re-sequencing samples, and the group information is shown in Figure 4B. The ratios of genes with variation in the TC (thick epidermis) group (group three) were higher compared with the TN (thin epidermis) group, such as EVM00008156, EVM0009214, and EVM0014071. Variation of EVM0016862 only can be detected in R13 (African horned melon [Cucumis metuliferus]) (see Figure S10). The variants for EVM0016862 between Payzawat and DHL92 may not be real due to the poor assembly quality of the gene. 15 samples downloaded from NCBI (see Table S29), consisting of thick- and thin-skinned melons, were used to test the expression of
six genes. Except for one gene, which did not express at any sample, other five genes expressed more or
less in the thick-epidermis melon (see Figure S11). All the above-mentioned evidence indicates that the six
genes may function in determining the thickness of the epidermis.

The chromosome-level genomes for Payzawat and CM3.6.1 enable the identification of large structural
variations and complex genome rearrangement. Large structural variations between Payzawat and
CM3.6.1 are intra-chromosomal translocation and inversion. To exclude the assembly error leading to
these differences, PacBio sub-reads were mapped to the inversion breakpoints of five large inversions
on chromosome 6 (see Tables S27 and S28 and Figure 2). All breakpoints were covered by PacBio
sub-reads (see Figure S8), validating the quality of Payzawat assembly. The structure of Payzawat chro-
mosome 6 is also supported by Hi-C contact frequency heatmapging. This discrepancy may be the
true difference between the two lines or the erroneous assembly of CM3.6.1 in this region. Some
QTLs, EayQL6.1 for early yield, FDQN6.1 for fruit diameter, PHYQN6.2 for phytoene content, β-carQN6.1
for α-carotene content, and fwi6.1 for fruit width (Diaz et al., 2011), were found to be co-located within
the inversion regions on chromosome 6.

Figure 2. Whole Genome Alignment between Payzawat and CM3.6.1
The red bar represents the chromosome of Payzawat, and the deep navy-blue bar represents the chromosome of
CM3.6.1. The sky blue line indicates synten between Payzawat and CM3.6.1.
Payzawat is characterized as large fruit size, sweetness, and orange fleshy pulp, whereas DHL92 is medium fruit size and white fleshy pulp (see Figures S7A and S7B). Comparative genomic analysis uncovers numerous line-specific variations, which may serve as the important resource for such trait differences.

**Genome Resequencing and Population Analysis**

Resequencing of the 50 melon accessions (Figure 4A) generated a total of clean 90,443,068 paired-end (PE) reads, with an average of 18X depth per accession (Tables S30 and S31), and the mapping rate ranged from 86.74% to 95.63% when R13 is excluded owing to its distantly phylogenetic relationship with species melo (see Tables S31 and S32).

A total of 240,775 (4.65%; 5,174,489 in total) and 12,794 indels (0.97%; 1,316,073 in total) were located in coding sequences (CDS), among which 98,701 SNPs were non-synonymous and 6,453 indels led to frame-shift mutations.

Neighbor-joining phylogenetic tree recovered from 50 melon accessions showed two clades (see Figure 4B). The sub-species agrestis did not cluster into one clade. Cultivars/landraces and the wild accessions of northern and southern China in sub-species agrestis formed a subclade and then grouped with the
Indian wild subclade, indicating China may be one Asian diversifying center, supporting the idea that some oriental Asian melons originated from India, then spread into northern and southern China before domestication (Akashi et al., 2002; Yashiro et al., 2005). The cultivars/landraces on sub-species agrestis was designated as Group two, with a characteristic thin epidermis (TN group) (see Figure S7D).

The cultivars/landraces on sub-species melo, designated as Group three, have the characteristic of a thick epidermis (TC group) (see Figure S7B), with different geographical origins first grouped together and then grouped with the Indian wild subclade, which supported the opinion that cultivars/landraces were introduced to western China via the Silk Road (Kitamura, 1951) after the domestication occurred in India. The wild melons on these two clades were designated as Group one. The fruit of wild-type (WT) is small compared with that of cultivar (CV) (see Figure S7C).

The evolutionary history of melons in China was further explored with Delta K (individual ancestry coefficients) values. Results revealed that four populations (K = 4) represent the best model (see Figure 4C).
and new sub-populations arose from each clade with K from 2 to 4 (see Figure 4D). When K was set to 4, new sub-populations arose from the wild species, indicating their high diversity and distinct phylogenetic relationships with domesticated melons. The cultivars/landraces were segregated into two sub-populations revealing the geographical distributions thereof in China.

Together, these findings led us to propose the melon evolutionary history in China, revealing that China may be one of the diversifying centers and illustrating the idea that melons in different regions of China originated from distinct Indian regions and spread into China via multiple routes.

Population genetic analysis among groups was performed to test whether, or not, China was a diversification center. Nucleotide diversity ($\pi$) and Tajima’s D were highest in the WT group ($\pi = 0.0020285$ and Tajima’s D = 1.59854), followed by Group three (TC group, $\pi = 0.0011642$ and Tajima’s D = 0.87328), and lowest in Group two (TN group, $\pi = 0.0002681$ and Tajima’s D = −0.50861), indicating the higher diversification of melons in northern and southern China (Group two).

LD analysis further confirmed the nucleotide diversity results. Linkage disequilibrium (LD) analysis for Group two (mainly containing melons from northern and southern China) has a much slower decay of pairwise correlation ($r^2$) values than for Group three (melons from western China) (see Figure 4E). Population analysis indicates that northern and southern China may be one of the diversification centers.

**Selective Sweep Analysis**

The fixation index (Fst) and $\pi$ were used to detect the sweep signals between the wild (Group 1) and cultivated (Groups 2 and 3) populations (see Figure 5A). Some 42,311,973-bp segments were recognized as sweep regions. Since the contrasting genotypes between wild and cultivated populations may contribute to the different phenotypes, e.g., fruit size, leaf size, seed size, and bitterness of the fruit, we focus on those genes with different genotypes between the wild and cultivated populations within the sweep regions. Among 26,919 small mutations (21,010 SNPs and 5,909 indels), 275 mutations (194 genes) change across the coding regions (non-synonymous mutation, start gained, stop gained, coding region change, insertion, deletion, and stop loss), providing the candidate pool for fruit trait improvement (Table S34). These genes were implicated in processes relating to fruit size (expansin-A12, cell division, and cell cycle), fruit development (brassinosteroid biosynthesis, the ethylene-activated signaling pathway, indole-3-acetic acid-induced protein ARG7, etc.), synthesis and transport of sugar (beta-glucosidase and alpha-L-fucosidase 2), fruit flavor (organic substance metabolism, carotenoid cleavage dioxygenase, water transport, solute carrier family, etc.), and biotic or abiotic response (disease resistance protein, etc.) (see Table S34).

*CmACS7* is involved in sex determination, and a causal SNP was reported to be associated with andromonoecy. *CmACS7* affected fruit size through the SNP (C_small fruit-T_large fruit) regulating andromonoecy (Boualem et al., 2008). The gene was co-located with the selective sweep regions. The SNP at A57V (chr01:35853579:C-T) is linked with andromonoecy (see Figure 5B). Interestingly, most wild accessions had the same haplotype as the V57 mutant (andromonoecious), whereas the allele of cultivars/landraces equals *CmACS7* (monoecious) (see Figure 5C and Table S34). The gene *CmACS7* was also implicated in fruit weight, total soluble solids, and ethylene contents (Galpaz et al., 2018). Except for *CmACS7*, the other six genes, including EVM0005618 (3-epi-6-deoxocathasterone 23-monoxygenase), EVM0005054 (mitotic B-type cyclin), EVM0010492 (VAL2), EVM0018337 (cyclin-B3-1), EVM0010058 (CCD1, carotenoid 9,10 (9',10')-cleavage dioxygenase 1), and EVM0009845 (early flowering 3, ELF3) were involved in fruit development, leaf and seed development, and regulation of flower development and flowering. The SNP leading to non-synonymous mutations on these genes may contribute to the difference in fruit size, leaf size, and seed size between WT and CV populations (Figure 5B). The differentiated haplotypes of these SNPs for WT (group one) and CV (group two and group three) could be detected (see Figure 5C). Heterogeneity of the genotype in the wild or CV populations may be due to the complicated taxonomic history of melon, leading to a high number of misclassified germplasm collections. The variations on these genes may be involved in the fruit, seed, or leaf size, and more dense sampling, especially for wild populations, is necessary.

**Identification of Candidate Haplotypes for Reported Genes**

The resequencing data can also be used to identify and validate the candidate causal mutation for previously reported genes. We focus on small mutation on the coding regions (non-synonymous mutation, start gained, stop gained, coding region change, insertion, deletion, and stop loss).
The pH gene has a significant effect on fruit acidity via a 12-base duplication on the coding region, and most sweet melons have this duplication (Cohen et al., 2014). pH in Payzawat matched the allele of sweet melon, with a 12-bp (TTAATTGTTGCA) duplication in the coding region, whereas the allele in DHL92 matched the allele of sour melon. Wild accessions matched the allele of sour melon, whereas cultivars/landraces accessions matched both alleles (see Figure S12 and Table S34), indicating that acidity is the diversifying trait (see Figure S12).

The pH gene has a significant effect on fruit acidity via a 12-base duplication on the coding region, and most sweet melons have this duplication (Cohen et al., 2014). pH in Payzawat matched the allele of sweet melon, with a 12-bp (TTAATTGTTGCA) duplication in the coding region, whereas the allele in DHL92 matched the allele of sour melon. Wild accessions matched the allele of sour melon, whereas cultivars/landraces accessions matched both alleles (see Figure S12 and Table S34), indicating that acidity is the diversifying trait (see Figure S12).

CmOr gene was the causal gene for difference of orange and non-orange colors (Tzuri et al., 2015), and a single SNP (G_green flesh->A_orange flesh) is responsible for the different flesh color pigmentation. The
allele in Payzawat matched the allele of orange melon (“Dulce”), whereas the allele in CM3.6.1 matched the allele of non-orange melon (“Tam Dew”). Sub-species agrestis (Group one and Group two) accessions matched the allele of non-orange melon, whereas sub-species melo (Group three) accessions matched the allele of orange melon (see Table S34 and Figure S12), indicating that the flesh color is the diversifying, or selection, trait.

ETHQB6.3 controls fruit ripening type (climacteric and non-climacteric), and MELO3C016540 was recognized as the causal gene (Ríos et al., 2017). We detected one SNP on EVM0015173, homologous to MELO3C016540. CmKFB (EVM0012228) is implicated in the accumulation of naringenin chalcone, which determined the yellow color of rind (Feder et al., 2015).

Candidate genes EVM0015625 and EVM0019658 were on the QTL SUCQSC5.1, which is responsible for sugar accumulation (Argyris et al., 2017). CmTHAT1 (acyetyl-CoA acetyltransferase, EVM0016460) affects fruit flavor. CmPPR1 (pentatricopeptide repeat-containing protein, EVM0014144) genes may affect carotenoid accumulation and flesh color, which may correspond to the wf locus (Galpaz et al., 2018). RGH10 (EVM001444) confers resistance to Papaya ring-spot virus (PRSV) and RGH9 (EVM0004621) confers resistance to races 0 and 2 (Brotman et al., 2013). ACS11 (EVM0021716) is involved in sex determination (Boualem et al., 2015), and some contrasting haplotypes were detected in TC and TN groups, such as SUCQSC5.1, CmTHAT1, and RGH10; some were detected in sub-species agrestis and melo, such as RGH9, PPR1, and ETHQB6.3. These traits are diversifying and showed taxonomic specificity.

The population analysis provides the basis of genetic control of many traits, e.g., flowering, fruit morphology, ripening behavior, rind characteristics, flesh color, and sugar content. The results will expand our knowledge of candidate genes controlling desirable traits.

DISCUSSION

In summary, by combing the SMRT sequencing technology and high-throughput chromosome interaction mapping technology, we obtained a high-quality melon genome with contig N50 up to 2.8 Mb and more than 98.53% sequences were anchored to 12 chromosomes, which provided a fine genome for future melon studies. Compared with use of NGS technology, SMRT sequencing exhibited superiority in genome assembly. Compared with traditional genetic maps, Hi-C technology exhibits several advantages in chromosome construction, such as encompassing up to one million markers, being convenient when preparing materials and offering high accuracy at lower cost. Both technologies help us obtain more intact genome information, which should be helpful in understanding the genetic changes in long-term natural evolution and artificial domestication of melon.

A high-quality genome assembly with higher contiguity and completeness shows its values in studying evolution, domestication, and breeding, as well as gaining insights into genome-wide structural variants (Sun et al., 2018; Zhang et al., 2019; Xie et al., 2019). The whole-genome sequence comparison between Payzawat and DHL92 enables the identification of the large inversions, e.g., five inversions on the tail of chromosome 6, and translocations, e.g., translocation on chromosome 5. These differences between two genomes may be line-specific or mis-assembly, and further validation of assembly quality is necessary. The comparative genomic analysis between Payzawat and DHL92 indicates that chromosome regions (e.g., Chr3, Chr4, Chr7, Chr8, Chr9, and Chr12) enriched with small variations (SNP and InDels) tend to colocalize with QTLs. The QTL STHQF12.4 (skin thickness) was located on chr12 (23,803,761–26,907,220), where small variations were enriched. Six genes annotated with pectinesterase, gibberellin 20 oxidase 1-B-like, and pectate lyase-like, and these six genes have variations in resequencing data. The homologs of pectinesterase, gibberellin 20 oxidase 1-B-like, and pectate lyase-like in plants are involved in cell wall biosynthesis (Micheli, 2001; Alqsous et al., 2004; Yang et al., 2017; Jeon; et al., 2016; Bai et al., 2014). The correlation between expression profiles of six genes and epidermis thickness must be tested in more samples. There is a shortage in the number of samples in our present analysis, and whether the six genes have variants in samples should be checked. Owing to the unavailable data (transcriptome data and resequencing data for each sample), the conclusion is less strong at present. More samples should be included in future to test the correlation among expression profiles, structural variants, and skin thickness. The identification of variants (large structural variants and small variants) will
allow the exploration of genotypes and phenotypes. However, the mechanism controlling the skin thickness in melon remains to be explored.

Previously, evidence backing up multiple domestications in melon was mostly derived from short DNA sequences (Endl et al., 2018), and whole-genome level evidence is more persuasive. We here present evidence supporting multiple domestications in melon from the whole-genome perspective. Because most samples in our study were collected from China, more samples from wider geographical locations are needed to strengthen the evidence. The higher diversification in northern and southern melons (TC group) compared with the western melons (TN group) supported with LD, π, and Tajima’s D explains the diverse shape, color, etc. in the TC group. The haplotype analysis indicates that some traits in modern melons, like fruit size through evaluating the haplotype of causative SNP in CmACS7, are domesticated from wild melon and some traits are lineage specific. The domestication and diversification processes contribute to the breeding through providing the molecular foundation.

Limitations of the Study
We reported a high-quality assembly of melon. We identified six genes potentially related to skin thickness. More samples are required to test the correlation among the variation, expression profile of six genes, and skin thickness. Furthermore, the samples for population analysis are inadequate and samples from wider geographical locations are needed.

METHODS
All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY
The DNA short reads, RNA-seq data, linked reads, and Pacbio sub-reads used in this study were submitted to NCBI under submission number SUB4523929 and project number PRJNA491307. The accession number for Payzawat genome assembly reported in this paper is SUB6534844 under the project PRJNA491307. The resequencing raw data have been deposited into NCBI Short Read Archive with the accession number SRP081058.

SUPPLEMENTAL INFORMATION
Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.10.049.

ACKNOWLEDGMENTS
This study was supported by The National Natural Science Foundation of China (Grant No. 31860564, http://www.nsfc.gov.cn), the Special Fund for Agroscientific Research in the Public Interest of The Ministry of Agriculture of the People’s Republic of China (Grant No. 201203004), and the earmarked fund for Modern Agro-industry Technology Research System (Grant No. CARS-25). We would like to thank Jordi Garcia-Mas from Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB-UB, Campus UAB, Edifici CRAG for providing the picture and trait information of DHL92 in this paper.

AUTHOR CONTRIBUTIONS
Conceptualization, H. Yi, H. Zhang, and H. Zheng; Methodology, X. Li, H. Yu, M. Liu, C. Ji, L. Ma, J. Tang, and J. Miao; Investigation, H. Zhang, Y. Zhang, M. Li, Ha. Wang, D. Wang, Hu. Wang, and Q. Fu; Visualization, H. Yu, L. Ma, and J. Tang; Writing – Original Draft, H. Zhang, H. Yi, X. Li, and H. Yu; Writing – Review & Editing, H. Zhang, and H. Yu; Supervision, H. Yi.

DECLARATION OF INTERESTS
The authors declare no competing interests.

Received: July 24, 2019
Revised: September 23, 2019
Accepted: October 24, 2019
Published: December 20, 2019
REFERENCES

Akashi, Y., Fukuda, N., Wako, T., Masuda, M., and Kato, K. (2000). Genetic variation and phylogenetic relationships in East and South Asian melons, Cucumis melo L. based on the analysis of five isozymes. Euphytica 125, 385–396.

Akdemir, K.C., and Chin, L. (2015). HiCPlotter integrates genomic data with interaction matrices. Genome Biol. 16, 198.

Altspsou, S., Carpentier, E., Kleineude, D., Burel, C., Mareck, A., Dauchel, H., Gomord, V., and Balangé, A.P. (2004). Identification and isolation of a pectin methylesterase isoform that could be involved in flax cell wall stiffening. Planta 219, 369–378.

Argyris, J.M., Ruiz-Herrera, A., Madriz-Masis, P., Sanseverino, W., Morata, J., Pujol, M., Ramos-Onsins, S.E., and Garcia-Mas, J. (2015). Use of an improved NB-LRR gene reveals how unisexual flowers develop and contribute to the evolution of sweet melons. Nat. Commun. 5, 4026.

Díaz, A., Fergany, M., Formisano, G., Ziarolo, P., Blanca, J., Fei, Z., Staub, J.E., Zalapa, J.E., Cuevas, H.E., Dace, G., et al. (2011). A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (Cucumis melo L.). BMC Plant Biol. 11, 111.

Díaz, A., Martinhernandez, A.M., Dolcetsanjuan, R., Garcesclaver, A., Alvarez, J.M., Gomiscas, J., and Monforte, A.J. (2017). Quantitative trait loci analysis of melon (Cucumis melo L.) domestication-related traits. Theor. Appl. Genet. 130, 1837–1856.

Endl, J., Achigandako, E.G., Pandey, A.K., Monforte, A.J., Pico, B., and Schaefer, H. (2018). Repeated domestication of melon (Cucumis melo) in Africa and Asia and a new close relative from India. Am. J. Bot. 105, 1662–1671.

Feder, A., Burger, J., Gao, S., Levinsohn, E., Kato, K., Schaffer, A.A., Meir, A., Davdovich-Rikanati, R., Portnoy, V., Gal-On, A., et al. (2015). A Kelch domain-containing F-box coding gene negatively regulates flavonoid accumulation in Cucumis melo L. Plant Physiol. 169, 1714–1726.

Galpaz, N., Gong, I., Shem-Tov, D., Barad, O., Izuri, G., Lev, S., Fei, Z., Xu, Y., Mao, L., Jiao, C., Harel-Beja, R., et al. (2018). Deciphering genetic factors that determine melon fruit-quality traits using RNA-seq-based high-resolution QTL and eQTL mapping. Plant J. 94, 169–191.

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., Gonzalez, V.M., Henaff, E., Cámara, F., Cozzuto, L., Lowy, E., et al. (2012). The genome of melon (Cucumis melo L.). Proc. Natl. Acad. Sci. U S A 109, 11872.

Gordon, D., Huddleston, J., Chaisson, M.J., Hillier, L.W., et al. (2016). Long-read sequence assembly of the gorilla genome. Science 352, eaad3644.

Janis, D.E., Ho, Y.S., Lightfoot, D.J., Schmckel, S.M., Li, B., Borm, T.J., Ohyanagi, H., Mineta, K., Michell, C.T., Saber, N., et al. (2017). The genome of Chenopodium quinoa. Nature 542, 307–312.

Jeon, H.W., Cho, J.S., Park, E.J., Han, K.H., Choi, Y.I., and Ko, J.H. (2016). Developing xylem-preferential expression of PDA2O20x1, a gibberellin 20-oxidase 1 from Pinus densiflora, improves woody biomass production in a hybrid poplar. Plant Biotechnol. J. 14, 1161–1170.

Kitamura, S. (1951). The origin of cultivated plants of China. Acta Phytosoy. Geobot. 14, 81–86.

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

Micheli, F. (2001). Pectin methyltransferases: cell wall enzymes with important roles in plant physiology. Trends Plant Sci. 6, 414–419.

Pitrat, M. (2013). Phenotypic diversity in wild and cultivated melons (Cucumis melo) Plant. Biotechnol. 30, 273–278.

Rios, P., Argyris, J., Vegas, J., Leida, C., Kengsawal, M., Tauri, G., Troadec, C., Bendahmane, A., Katzir, N., Picó, B., et al. (2017). ETHQV6.3 is involved in melon climacteric fruit ripening and is encoded by a NAC domain transcription factor. Plant J. 91, 671–683.

Ruggieri, V., Alexiou, K.G., Morata, J., Argyris, J., Pujol, M., Yano, R., Nonaka, S., Ezura, H., Latrasse, D., Boualem, A., et al. (2018). An improved assembly and annotation of the melon (Cucumis melo L.) reference genome. Sci. Rep. 8, 8088.

Smato, F.A., Waterhouse, R.M., Joannides, P., Kruvels, E.V., and Zdobnov, E.M. (2015). 1,440 BUSCO embryophyta subset of genes: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Sun, S., Zhou, Y., Chen, J., Shi, J., Zhao, H., Zhao, H., Song, W., Zhang, M., Cui, Y., Dong, X., et al. (2018). Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nat. Genet. 50, 1299–1295.

Tzuri, G., Zhou, X., Chayut, N., Yuan, H., Portnoy, V., Meir, A., Sa’ar, U., Baumkoler, F., Mazurek, M., Levinsohn, E., et al. (2015). A ‘golden’ SNP in CmGr defines the fruit flesh color of melon (Cucumis melo). Plant J. Cell Mol. Biol. 82, 267–279.

Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., Wu, G., Tian, S., Chen, C., Liu, Y., et al. (2017). Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat. Genet. 49, 765–772.

Xie, M., Chung, C.Y., Li, M., Wong, F.L., Wang, X., Liu, A., Wang, Z., Leung, A.K., Wong, T.H., Tong, S.W., et al. (2019). A reference-grade wild soybean genome. Nat. Commun. 10, 1–12.

Yang, L., Huang, W., Xiong, F., Xian, Z., Su, D., Ren, M., and Li, Z. (2017). Silencing of SLP, which encodes a pectate lyase in tomato, confers enhanced fruit firmness, prolonged shelf-life and reduced susceptibility to grey mould. Plant Biotechnol. J. 15, 1544–1555.

Yashiro, K., Iwata, H., Akashi, Y., Tomita, K., Kuzuya, M., Tsumura, Y., and Kato, K. (2005). Genetic relationship among East and South Asian melon (Cucumis melo) revealed by AFLP analysis. Breed. Sci. 55, 197–206.

Ye, C., Hill, C.M., Wu, S., Ruan, J., and Ma, Z.S. (2016). DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci. Rep. 6, 31900.

Zhang, L., Hu, J., Han, X., Gao, Y., Richards, C.M., Zhang, C., Tian, Y., Liu, G., Guf, H., et al. (2019). A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat. Commun. 10, 1494.
Supplemental Information

A High-Quality Melon Genome Assembly Provides Insights into Genetic Basis of Fruit Trait Improvement

Hong Zhang, Xuming Li, Haiyan Yu, Yongbing Zhang, Meihua Li, Haojie Wang, Dengming Wang, Huaisong Wang, Qiushi Fu, Min Liu, Changmian Ji, Liming Ma, Juan Tang, Song Li, Jianshun Miao, Hongkun Zheng, and Hongping Yi
Figure S1. 17-Kmer frequency distribution using Illumina short reads. Related to Figure 1.
The depth of peak was at 64 and peak at depth of more than 128 was a repetitive peak.
Figure S2. Assessment of genome assembly. Related to Figure 1.

a. The interaction frequency distribution of Hi-C links among Payzawat melon chromosomes. We scanned the genome using a 500-kb non-overlapping window as a bin and calculated valid interaction links of Hi-C data between any pair of bins. The base-2 logarithm of the link number was calculated. The distribution of links among chromosomes was exhibited by heat-map based on Hi-Cplotter. The colour key of the heat-map ranges from light yellow to dark red indicating the frequency of Hi-C interaction links from low to high (0 to 6).

b. The genetic-map-based evaluation of assembly. Genetic markers were mapped to the genome and locations were plotted according to the genetic map and HiC anchored pseudochromosomes.
Figure S3. Identification of centromeric repeats and telomeric motifs on Payzawat assembly. Related to Figure 1. The red bar represents the chromosome of Payzawat and the blue bar represents the chromosome of CM3.6.1. The magenta rectangles indicate the position of centromeric repeats and the black rectangles indicates the position of telomeres.
Figure S4. Veen plot of gene sets. Related to Figure 1.
The predicted genes were supported by *De novo*, homolog or transcriptome evidence.
Figure S5. Length distribution of mRNA, CDS, and intron. Related to Figure 1.

a. Distribution of mRNA length in species within the Cucurbitaceae family and other plants;
b. Distribution of CDS length in species within the Cucurbitaceae family and other plants;
c. Distribution of intron length in species within the Cucurbitaceae family and other plants.
Figure S6. Whole genome comparison between Payzawat and CM3.6.1. Related to Figure 2.

The red line on the diagonal indicates the same orientation and the blue line indicates the reverse orientation.
Figure S7. Fruit pictured: DHL92 (a), Payzawat (b), wild (c), and thin-skin melon (d). Related to Figure 2,3 and 4.
Figure S8. Validation of inversion breakpoints on chr06 through PacBio subreads mapping back to Payzawat assembly. Related to Figure 2.

a. PacBio subreads mapping to the left breakpoint of INV1. b. PacBio subreads mapping to the right breakpoint of INV1. c. PacBio subreads mapping to the right breakpoint of INV2. d. PacBio subreads mapping to the left breakpoint of INV3. e. PacBio subreads mapping to the right breakpoint of INV3. f. PacBio subreads mapping to the left breakpoint of INV5.

The red line indicates PacBio subreads mapping back to the Payzawat assembly on the forward orientation and the blue line indicates PacBio subreads mapping back to the Payzawat assembly on the reverse orientation. The grey line indicates non-mapping PacBio subreads. The green bar represents the segments of Payzawat assembly. The upper panel on the bar represents the high-quality mapping subreads and the lower panel on the bar represents
low-quality mapping subreads.

Figure S9. Sequence alignment for candidate genes implicated in skin thickness. Related to Figure 3.
Figure S10. Variation of six genes in 49 resequencing samples. Related to Figure 3.
Figure S11. Expression profile of genes. Related to Figure 3.
DAP : day after pollination. PI: PI 161375 (PI, conomon group); Ved: cv. Védrantais (Ved; cantalupensis group); Fla:C. melo var ameri, cv ‘Flavor No. 4’.
Figure S12. Candidate alleles for reported genes. Related to Figure 5.
a. candidate causative SNPs for reported genes.  
b. Structure and validated deletions for CmPH.
**Table S1. Summary of Illumina PE reads for melon. Related to Figure 1.**

| Date type                                      | Library Insert Size (bp) | PE Reads Number | Base Number                  | Q30    |
|------------------------------------------------|--------------------------|-----------------|-----------------------------|--------|
| Genome survey and genome base error correction | 270                      | 108,577,515     | 32,528,871,468              | 91.94% |
| HiC data for genome error correction           |                          |                 |                             |        |
| chromosome construction                        | 300~700                  | 164,495,209     | 49,070,802,940              | 90.24% |
Table S2a. Summary of SMRT reads for melon. Related to Figure 1.

| Type       | Filter  | Read Bases (bp) | Reads Num | Read N50 | Mean Reads Length (bp) | Reads Quality |
|------------|---------|-----------------|-----------|----------|------------------------|---------------|
| Polymerase | Pre-Filter | 30,783,578,459  | 4,057,884 | 20,305   | 7,586                  | 0.405         |
|            | Post-Filter | 27,848,664,023  | 1,876,049 | 20,690   | 14,844                 | 0.835         |
| Subreads   | Filter 500 bp | 27,795,975,505 | 2,643,196 | 13,947   | 10,516                 | ---           |

Table S2b. The statistic of Pacbio subreads length distribution. Related to Figure 1.

| Length (bp) | Num   | Total length (bp) | Average length (bp) |
|-------------|-------|-------------------|---------------------|
| 0~2000      | 227,002 | 308,289,603       | 1,358               |
| 2000~4000   | 293,359 | 859,695,674       | 2,931               |
| 4000~6000   | 283,464 | 1,419,329,012     | 5,007               |
| 6000~8000   | 281,390 | 1,967,306,789     | 6,991               |
| 8000~10000  | 273,231 | 2,461,567,552     | 9,009               |
| 10000~12000 | 308,358 | 3,395,984,822     | 11,013              |
| 12000~14000 | 275,809 | 3,573,685,572     | 12,957              |
| 14000~16000 | 203,459 | 3,040,961,048     | 14,946              |
| 16000~18000 | 144,022 | 2,440,252,006     | 16,944              |
| 18000~      | 353,102 | 8,328,903,427     | 23,588              |
| Total       | 2,643,196 | 27,795,975,505 | 10,516              |
Table S3. The assembly results of melon genome using different softwares. Related to Figure 1.

| Software      | Contig number | Contig length (bp) | Contig N50 (bp) | Contig N90 (bp) | Contig Max (bp) | GC%  |
|---------------|---------------|--------------------|-----------------|-----------------|-----------------|------|
| Canu v1.5     | 755           | 365,394,387        | 1,016,639       | 267,218         | 9,266,155       | 33.82|
| DBG2OLC       | 1,328         | 389,219,408        | 821,979         | 180,906         | 4,725,009       | 33.80|
| Quickmerage   | 803           | 386,470,950        | 2,863,989       | 511,464         | 11,158,838      | 33.89|
Table S4. Statistics of pseudochromosomes of melon. Related to Figure 1.

| Group | Sequence Number | Sequence Length (bp) |
|-------|-----------------|----------------------|
| Chr1  | 59              | 37,360,996           |
| Chr2  | 24              | 25,498,186           |
| Chr3  | 36              | 31,151,059           |
| Chr4  | 60              | 34,816,742           |
| Chr5  | 65              | 30,780,632           |
| Chr6  | 55              | 37,833,529           |
| Chr7  | 70              | 31,462,808           |
| Chr8  | 68              | 35,445,645           |
| Chr9  | 28              | 24,937,816           |
| Chr10 | 54              | 28,136,047           |
| Chr11 | 55              | 34,134,534           |
| Chr12 | 63              | 29,233,345           |
| Total Sequences Clustered (Ratio %) | 637(72.22) | 380,791,339(98.53) |
| Total Sequences Ordered and Oriented (Ratio %) | 271(42.54) | 363,764,534(95.53) |
Table S5. Genome completeness evaluation based on short reads mapped to genome. Related to Figure 1.

| Total reads   | Mapped reads | Mapped(%) | Properly mapped reads | Properly mapped(%) |
|---------------|--------------|-----------|-----------------------|-------------------|
| 218,378,255   | 216,286,575  | 99.04     | 210,224,508           | 96.81             |
Table S6. Genome completeness evaluation based on all PacBio subreads mapped back to genome. Related to Figure 1.

| Parameter                        | Value     |
|----------------------------------|-----------|
| Total Number                     | 2,643,196 |
| Total Length                     | 2,026,171,729 |
| Total Mapped Number              | 2,343,297(88.65) |
| Total Mapped_Length              | 24,691,702,398(88.83) |
| Mapping_identity                 | 84.06     |
| Properly_Mapped_Number           | 1,898,351(71.82) |
| Properly_Mapped_length           | 18,225,688,665(65.57) |
| Properly_Mapped_identity         | 84.83     |
Table S7. Genome completeness evaluation based on CEGMA database. Related to Figure 1.

| Species          | Number of 458 CEGs* present in assembly | % of 458 CEGs present in assemblies | Number of 248 highly conserved CEGs present | % of 248 highly conserved CEGs present |
|------------------|----------------------------------------|-----------------------------------|-------------------------------------------|---------------------------------------|
| Payzawat melon   | 448                                    | 97.82%                            | 246                                       | 99.19%                                |
| Complete BUSCOs (C)            | 1,336 |
|--------------------------------|-------|
| Complete and single-copy BUSCOs (S) | 1,278 |
| Complete and duplicated BUSCOs (D) | 58    |
| Fragmented BUSCOs (F)         | 30    |
| Missing BUSCOs (M)            | 74    |
| Total BUSCO groups searched   | 1,440 |
| Contig length (bp) | Correct base number (bp) | Error base number (bp) | Error base percentage (%) |
|-------------------|--------------------------|------------------------|---------------------------|
| 386,470,950       | 386,469,646              | 1,304                  | 0.0003374                 |
| Chr | start   | end     | length |
|-----|---------|---------|--------|
| chr07 | 8,015,331 | 8,018,992 | 3,661  |
| chr07 | 8,206,075  | 8,226,875  | 20,800 |
| chr07 | 20,841,172 | 20,842,193 | 1,021  |
| chr07 | 20,843,257 | 20,847,128 | 3,871  |
| chr07 | 20,848,967 | 20,859,599 | 10,632 |
| chr07 | 20,862,464 | 20,868,572 | 6,108  |
| chr07 | 20,872,481 | 20,875,457 | 2,976  |
| chr07 | 20,878,955 | 20,880,323 | 1,368  |
| chr09 | 12,317,864 | 12,326,398 | 8,534  |
| chr09 | 12,336,629 | 12,343,584 | 6,955  |
| chr09 | 12,419,626 | 12,426,587 | 6,961  |
| chr09 | 12,432,876 | 12,436,582 | 3,706  |
| chr09 | 12,438,202 | 12,439,594 | 1,392  |
| chr09 | 12,483,168 | 12,487,617 | 4,449  |
| chr01 | 27,893,563 | 27,898,799 | 5,236  |
| chr01 | 27,902,574 | 27,904,244 | 1,670  |
| chr01 | 27,914,340 | 27,917,772 | 3,432  |
| chr01 | 27,923,006 | 27,927,215 | 4,209  |
| chr01 | 27,936,171 | 27,940,186 | 4,015  |
| chr01 | 27,950,089 | 27,958,282 | 8,193  |
| chr10 | 21,539,424 | 21,544,931 | 5,507  |
| chr10 | 21,582,330 | 21,588,656 | 6,326  |
| chr10 | 21,592,978 | 21,595,406 | 2,428  |
| chr10 | 23,409,214 | 23,419,616 | 10,402 |
| chr10 | 23,438,756 | 23,440,072 | 1,316  |
| chr10 | 23,441,655 | 23,442,704 | 1,049  |
| chr08 | 24,260,189 | 24,273,071 | 12,882 |
| chr08 | 24,275,260 | 24,280,456 | 5,196  |
| chr08 | 24,295,050 | 24,300,327 | 5,277  |
| chr08 | 24,323,583 | 24,339,816 | 16,233 |
| chr08 | 24,352,904 | 24,357,464 | 4,560  |
| chr08 | 24,358,854 | 24,368,249 | 9,395  |
| chr08 | 24,562,073 | 24,575,419 | 13,346 |
| chr08 | 26,997,779 | 26,999,774 | 1,995  |
| chr08 | 27,003,939 | 27,021,444 | 17,505 |
| chr04 | 22,747,982 | 22,763,666 | 15,684 |
| chr04 | 22,950,178 | 22,960,972 | 10,794 |
| chr04 | 22,966,229 | 22,972,403 | 6,174  |
| Chr  | Start    | End      | Length  |
|------|----------|----------|---------|
| chr04| 22,975,165 | 22,978,123 | 2,958   |
| chr06| 12,727,898  | 12,732,373 | 4,475   |
| chr06| 12,746,562  | 12,766,615 | 20,053  |
| chr11| 19,214,550  | 19,221,839 | 7,289   |
| chr11| 19,225,959  | 19,247,642 | 21,683  |
| chr11| 20,169,026  | 20,177,304 | 8,278   |
| chr11| 20,193,590  | 20,206,634 | 13,044  |
| chr11| 21,838,773  | 21,852,751 | 13,978  |
| chr05| 11,136,461  | 11,141,615 | 5,154   |
| chr05| 11,144,384  | 11,160,121 | 15,737  |
| chr02| 17,684,383  | 17,698,691 | 14,308  |
| chr02| 19,266,330  | 19,268,697 | 2,367   |
| chr02| 19,271,003  | 19,281,783 | 10,780  |
| chr02| 19,294,902  | 19,296,307 | 1,405   |
| chr02| 19,297,779  | 19,300,853 | 3,074   |
| chr02| 19,303,933  | 19,310,266 | 6,333   |
| chr03| 17,115,534  | 17,130,132 | 14,598  |
| chr03| 19,587,074  | 19,603,704 | 16,630  |
| chr03| 19,651,784  | 19,653,131 | 1,347   |
| chr03| 19,658,110  | 19,672,851 | 14,741  |
| chr12| 5,332,493   | 5,337,079  | 4,586   |
| chr12| 5,370,036   | 5,385,820  | 15,784  |
| chr12| 5,431,354   | 5,435,327  | 3,973   |
| chr12| 11,305,901  | 11,317,300 | 11,399  |
| chr12| 11,319,463  | 11,324,210 | 4,747   |
| chr12| 11,325,236  | 11,330,678 | 5,442   |
| chr12| 11,340,154  | 11,341,360 | 1,206   |
| chr12| 11,342,976  | 11,352,077 | 9,101   |
| chr12| 11,361,571  | 11,373,911 | 12,340  |
| chr12| 11,378,487  | 11,405,001 | 26,514  |
| chr12| 11,414,540  | 11,441,351 | 26,811  |
| chr12| 11,443,885  | 11,454,420 | 10,535  |
| chr12| 11,458,460  | 11,463,850 | 5,390   |
| chr12| 12,546,050  | 12,560,234 | 14,184  |
| chr12| 13,262,503  | 13,265,941 | 3,438   |
| chr12| 13,279,579  | 13,284,715 | 5,136   |
| chr12| 15,875,192  | 15,885,071 | 9,879   |
| chr12| 15,894,482  | 15,898,343 | 3,861   |
| chr12| 15,899,582  | 15,903,423 | 3,841   |
| chr12| 15,908,872  | 15,916,292 | 7,420   |
| Chr  | start     | end        | length   |
|------|-----------|------------|----------|
| chr03| 17,742,757| 17,743,802 | 1,045    |
| chr08| 27,672,447| 27,675,819 | 3,372    |
| chr08| 27,712,504| 27,714,162 | 1,658    |
| chr08| 27,717,811| 27,719,191 | 1,380    |
| chr08| 29,909,421| 29,910,698 | 1,277    |
| chr08| 29,933,035| 29,936,649 | 3,614    |
| chr02| 20,259,278| 20,260,332 | 1,054    |
| chr02| 21,870,709| 21,871,714 | 1,005    |
| chr02| 21,874,072| 21,875,685 | 1,613    |
| chr11| 19,614,632| 19,615,904 | 1,272    |
| chr11| 19,620,451| 19,622,053 | 1,602    |
| chr11| 19,623,517| 19,626,128 | 2,611    |
| chr11| 20,612,742| 20,620,499 | 7,757    |
| chr11| 20,640,710| 20,642,256 | 1,546    |
| chr11| 20,649,687| 20,651,263 | 1,576    |
| chr11| 20,667,175| 20,670,878 | 3,703    |
| chr11| 20,674,239| 20,682,558 | 8,319    |
| chr06| 16,344,750| 16,346,423 | 1,673    |
| chr07| 3,841,902 | 3,844,347  | 2,445    |
| chr07| 3,846,355 | 3,847,591  | 1,236    |
| chr07| 6,351,604 | 6,355,252  | 3,648    |
| chr12| 10,596,545| 10,598,804 | 2,259    |
| chr12| 10,604,391| 10,609,560 | 5,169    |
| chr12| 10,615,471| 10,617,219 | 1,748    |
| chr12| 10,630,371| 10,632,002 | 1,631    |
| chr12| 10,633,459| 10,637,214 | 3,755    |
| chr12| 10,638,910| 10,641,712 | 2,802    |
| chr12| 10,654,436| 10,655,923 | 1,487    |
| chr12| 10,659,565| 10,662,153 | 2,588    |
| chr12| 10,663,300| 10,665,685 | 2,385    |
| chr12| 10,675,727| 10,684,566 | 8,839    |
| chr12| 10,703,341| 10,707,125 | 3,784    |
| chr12| 10,709,139| 10,712,579 | 3,440    |
| chr12| 10,718,792| 10,720,687 | 1,895    |
| chr12| 10,727,683| 10,730,544 | 2,861    |
| chr12| 10,734,364| 10,736,114 | 1,750    |
| chr12| 10,743,571| 10,745,328 | 1,757    |
| chr12| 10,746,758| 10,748,320 | 1,562    |
| Chromosome | Start   | End     | Length |
|------------|---------|---------|--------|
| chr12      | 10,759,308 | 10,761,076 | 1,768 |
| chr12      | 10,763,116 | 10,765,444 | 2,328 |
| chr09      | 12,397,946 | 12,399,622 | 1,676 |
| chr09      | 12,401,997 | 12,403,402 | 1,405 |
| chr09      | 12,425,831 | 12,427,816 | 1,985 |
| chr09      | 12,433,047 | 12,435,085 | 2,038 |
| chr09      | 12,530,564 | 12,531,613 | 1,049 |
| chr04      | 22,568,462 | 22,571,832 | 3,370 |
| chr04      | 22,809,439 | 22,811,373 | 1,934 |
| chr04      | 22,818,031 | 22,819,084 | 1,053 |
| chr10      | 26,646,135 | 26,648,001 | 1,866 |
| chr10      | 26,649,714 | 26,653,004 | 3,290 |
| Chr   | start  | end   |
|-------|--------|-------|
| chr01 | 18945  | 19019 |
| chr01 | 18945  | 19098 |
| chr01 | 20909  | 20962 |
| chr01 | 20913  | 20962 |
| chr03 | 247641 | 252634|
| chr04 | 1435   | 4527  |
| chr05 | 3854888| 3861842|
| chr07 | 2790259| 2790909|
| chr08 | 490689 | 490759 |
| chr09 | 225715 | 225784 |
| chr10 | 13359  | 13602 |
| chr12 | 947196 | 947242 |
| chr03 | 30117021| 30117116|
| chr04 | 33368710| 33374109|
| chr07 | 29453786| 29453843|
| chr09 | 24250739| 24262802|
| chr12 | 27486643| 27520640|
Table S13. The position of telomeres in CM3.6.1. Related to Figure 1.

| Chr | start  | end    |
|-----|--------|--------|
| chr01 | 1,364  | 1,417  |
| chr04 | 131,820 | 131,878 |
| chr05 | 108    | 225    |
| chr06 | 36,915,748 | 36,915,775 |
| chr10 | 544,858 | 544,886 |
Table S14. The repeat elements of melon genome. Related to Figure 1.

| Type                        | Number | Length (bp) | Percentage (%) |
|-----------------------------|--------|-------------|-----------------|
| ClassI/?/?                  | 551    | 271,757     | 0.07            |
| ClassI/DIRS/?               | 21,562 | 17,936,642  | 4.64            |
| ClassI/DIRS/DIRS            | 3      | 149         | 0.00            |
| ClassI/LARD/?               | 53,700 | 19,381,169  | 5.01            |
| ClassI/LINE/?               | 2,162  | 955,712     | 0.25            |
| ClassI/LINE/I               | 11     | 738         | 0.00            |
| ClassI/LINE/Jockey          | 9      | 522         | 0.00            |
| ClassI/LINE/L1              | 8,158  | 3,066,023   | 0.79            |
| ClassI/LINE/R2              | 10     | 462         | 0.00            |
| ClassI/LINE/RTE             | 9      | 689         | 0.00            |
| ClassI/LTR/?                | 2,012  | 908,864     | 0.24            |
| ClassI/LTR/Copia            | 76,236 | 48,966,451  | 12.67           |
| ClassI/LTR/Gypsy            | 109,600| 66,674,086  | 17.25           |
| ClassI/LTR/Retrovirus       | 1      | 101         | 0.00            |
| ClassI/PLE/Penelope         | 408    | 40,059      | 0.01            |
| ClassI/PLE/LARD/?           | 1,025  | 238,717     | 0.06            |
| ClassI/SINE/?               | 5,601  | 1,074,120   | 0.28            |
| ClassI/TRIM/?               | 768    | 319,729     | 0.08            |
| ClassII/?/?                 | 1,164  | 193,622     | 0.05            |
| ClassII/?/Academ            | 2      | 83          | 0.00            |
| ClassII/?/Ginger2/TDD       | 2      | 695         | 0.00            |
| ClassII/?/ISL2EU            | 1      | 48          | 0.00            |
| ClassII/?/Kolobok           | 628    | 60,431      | 0.02            |
| ClassII/?/MuDR              | 5,757  | 3,660,728   | 0.95            |
| ClassII/?/Novosib           | 34     | 4391        | 0.00            |
| ClassII/?/Sola              | 7      | 600         | 0.00            |
| ClassII/Crypton/Crypton     | 830    | 352,479     | 0.09            |
| ClassII/Helitron/Helitron   | 7,524  | 1,209,828   | 0.31            |
| ClassII/MITE/?              | 2,227  | 429,934     | 0.11            |
| ClassII/Maverick/?          | 2,878  | 1,140,186   | 0.30            |
| ClassII/TIR/?               | 14,406 | 8,560,416   | 2.22            |
| ClassII/TIR/CACTA           | 23,208 | 15,115,707  | 3.91            |
| ClassII/TIR/P               | 303    | 23,952      | 0.01            |
| ClassII/TIR/PIF-Harbinger   | 7,795  | 3,646,425   | 0.94            |
| ClassII/TIR/PiggyBac        | 10     | 610         | 0.00            |
| Class/TIR/Tc1-Mariner | 25  | 2,060 | 0.00 |
|-----------------------|-----|-------|------|
| Class/TIR/hAT         | 13,809 | 4,331,570 | 1.12 |
| PotentialHostGene     | 13,128 | 2,633,731 | 0.68 |
| SSR                   | 5,317 | 1,040,848 | 0.27 |
| Unknown               | 135,682 | 38,668,955 | 10.01 |
| Total without overlap | 516,563 | 192,457,182 | 49.80 |
Table S15. Summary of the melon genome gene model prediction. Related to Figure 1.

| Method          | Software and Species                     | Gene number |
|-----------------|------------------------------------------|-------------|
| Ab initio       | **Genscan**                              | 21,436      |
|                 | **Augustus**                             | 23,111      |
|                 | **GlimmerHMM**                           | 23,925      |
|                 | **GeneID**                               | 34,013      |
|                 | **SNAP**                                 | 42,865      |
| Homology-based  | **GeMoMa**                               |             |
|                 | *Cucumis sativus (Cultivar: Borszczagowski)* | 16,909      |
|                 | *Cucumis sativus (Cultivar: 9930)*       | 20,171      |
| Integration     | **PASA**                                 | 39,751      |
|                 | **EVM**                                  | 22,924      |
Table S16. Statistics of the melon genome gene model prediction. Related to Figure 1.

| Software | Gene Number | Total length (bp) | Average gene length (bp) | Total exon length (bp) | Average exon length (bp) | Total Intron Length (bp) | Average Intron length (bp) |
|----------|-------------|-------------------|--------------------------|------------------------|--------------------------|--------------------------|---------------------------|
| EVM      | 22,924      | 102,944,993       | 4,491                    | 29,654,362             | 238                      | 59,893,433               | 480                       |
Table S17. The gene-set evaluation by mapping mRNA reads to genome. Related to Figure 1.

| Type    | BaseNum (bp)   | Percentage (%) |
|---------|----------------|----------------|
| Exon    | 6,547,111,128  | 88.97%         |
| Intron  | 366,647,491    | 4.98%          |
| Intergenic | 445,330,510 | 6.05%          |
| Total   | 7,359,089,129  | 100            |
Table S18. Summary of the ncRNAs in melon genome. Related to Figure 1.

| RNA classification | Number | Family |
|--------------------|--------|--------|
| miRNA              | 75     | 19     |
| rRNA               | 443    | 4      |
| tRNA               | 717    | 24     |
Table S19. Summary of the pseudogenes in melon genome. Related to Figure 1.

| RNA classification       | Number | Total length | Average Length |
|--------------------------|--------|--------------|----------------|
| Frameshift               | 740    | 2,256,944    | 3,050          |
| Stop codon termination   | 1,694  | 5,177,533    | 3,056          |
| Total                    | 2,434  | 7,434,477    | 3,054          |
Table S20. The gene functional annotation of melon. Related to Figure 1.

| Database | Annotated number | Percentage (%) |
|----------|------------------|----------------|
| KOG      | 8,596            | 37.5           |
| GO       | 12,436           | 54.25          |
| KEGG     | 7,582            | 33.07          |
| TrEMBL   | 22,413           | 97.77          |
| nr       | 22,492           | 98.12          |
| All Annotated | 22,506       | 98.18          |
Table S21. The valid data evalution of Hi-C clean reads. Related to Figure 1.

| Type                          | Reads Number | Ratio(%) |
|-------------------------------|--------------|----------|
| Total Read Pairs              | 164,495,209  | 100      |
| Mapped Reads Pairs            | 149,968,147  | 91.17    |
| Unique Mapped Read Pairs      | 76202174     | 46.32    |
| Unique Mapped Read Pairs      |              |          |
| Valid Interaction Pairs       | 53,271,510   | 69.91    |
| Dangling End Pairs            | 9,199,190    | 12.07    |
| Re-ligation Pairs             | 1,172,025    | 1.54     |
| Self-cycle Pairs              | 8,710,138    | 11.43    |
| Dumped Pairs                  | 3,849,311    | 5.05     |
Table S22. Summary of the genome correction by Hi-C data. Related to Figure 1.

| Statistics       | Merged Assembly contigs | Corrected contigs |
|------------------|-------------------------|-------------------|
| Contig number    | 803                     | 863               |
| Contig length    | 386,470,950             | 386,470,950       |
| Contig N50 (bp)  | 2,863,989               | 2,316,192         |
| Contig N90 (bp)  | 511,464                 | 393,785           |
| Contig Max (bp)  | 11,158,838              | 6,858,417         |
| Gap number       | 0                       | 0                 |
Table S23. Statistics of aligned sequences, SNPs and Indels in Payzawat from Payzawat and CM3.6.1 whole-genome sequence comparison. Related to Figure 2 and 3.

| Category                  | Syntenic block length | No. of genes |
|---------------------------|-----------------------|--------------|
| One-to-One syntenic blocks| 296,593,814 bp (76.7%) | 19,110       |
| SNPs                      | 1,761,822             | 9,795        |
| Insertion                 | 371,560               | 7,779        |
| Deletion                  | 363,926               | 8,216        |
Table S24. Genome distribution of SNPs and Indels within one-to-one syntenic blocks in Payzawat from Payzawat and CM3.6.1 whole-genome sequence comparison. Related to Figure 2 and 3.

| Region   | SNP No. | SNP Percent | No. of affected genes | SNP No. | SNP Percent | No. of affected genes |
|----------|---------|-------------|-----------------------|---------|-------------|-----------------------|
| 5'UTR    | 109,259 | 6.204       | 6,055                 | 69,714  | 7.32        | 4,386                 |
| CDS      | 33,636  | 1.909       | 6,971                 | 20,591  | 2.28        | 4,771                 |
| Intron   | 168,487 | 9.563       | 7,460                 | 109,455 | 12.14       | 5,467                 |
| 3'UTR    | 109,982 | 6.242       | 6,438                 | 73,847  | 8.19        | 4,701                 |
| Intergenic | 1,340,458 | 76.115   | 627,971               | -       | -           | -                     |
| Total    | 1,761,822 | 100        | 9,795                 | 901,578 | 100         | 7,046                 |

| Region   | Deletion No. | Deletion Percent | No. of affected genes | Deletion No. | Deletion Percent | No. of affected genes |
|----------|--------------|------------------|-----------------------|--------------|------------------|-----------------------|
| 5'UTR    | 23,943       | 6.579            | 4655                  | 14,562       | 8.125            | 3,444                 |
| CDS      | 2,517        | 0.691            | 1239                  | 1,335        | 0.744            | 736                   |
| Intron   | 37,340       | 10.260           | 6099                  | 22,830       | 12.733           | 4,538                 |
| 3'UTR    | 23,125       | 6.354            | 4715                  | 14,789       | 8.251            | 3,469                 |
| Intergenic | 277,001      | 76.114           | -                     | 125,707      | 70.139           | -                     |
| Total    | 363,926      | 100              | 8216                  | 179,223      | 100              | 6,081                 |

| Region   | Insertion No. | Insertion Percent | No. of affected genes | Insertion No. | Insertion Percent | No. of affected genes |
|----------|---------------|-------------------|-----------------------|---------------|-------------------|-----------------------|
| 5'UTR    | 23,050        | 6.204             | 4,397                 | 13,177        | 7.409             | 3,227                 |
| CDS      | 2,031         | 0.546             | 966                   | 924           | 0.519             | 546                   |
| Intron   | 37,318        | 10.044            | 5,761                 | 21,543        | 12.113            | 4,274                 |
| 3'UTR    | 23,436        | 6.307             | 4,574                 | 14,329        | 8.057             | 3,391                 |
| Intergenic | 285,725      | 76.899            | -                     | 127,868       | 71.900            | -                     |
| Total    | 371,560       | 100               | 7,779                 | 177,841       | 100              | 5,749                 |
Table S25. Position distribution of structural variation. Related to Figure 2 and 3.

| Type       | No. | 5'UTR | CDS  | Intron | 3'UTR | Intergenic Total |
|------------|-----|-------|------|--------|-------|------------------|
| **Structural variation** |     |       |      |        |       |                  |
| GAP        | 2,507 | 995   | 3,595| 2,386  | 21,471| 26,067           |
| DUP        | 8     | 9     | 12   | 8      | 68    | 83               |
| BRK        | 114   | 115   | 99   | 107    | 338   | 459              |
| JMP        | 340   | 325   | 403  | 317    | 3,455 | 4,021            |
| INV        | 325   | 331   | 386  | 299    | 2,935 | 3,481            |
| SEQ        | 1,351 | 1,086 | 1,636| 1,277  | 13,004| 15,279           |
| **Structural variation within one-to-one syntenic blocks** |     |       |      |        |       |                  |
| GAP        | 491   | 73    | 670  | 438    | 4,381 | 5,226           |
| DUP        | 0     | 0     | 0    | 0      | 0     | 0                |
| BRK        | 0     | 0     | 0    | 0      | 0     | 0                |
| JMP        | 8     | 8     | 9    | 6      | 61    | 73               |
| INV        | 3     | 0     | 2    | 1      | 19    | 23               |
| SEQ        | 132   | 75    | 155  | 115    | 1,372 | 1,633           |
Table S26. Statistics of genes with large variation. Related to Figure 2 and 3.

| Type  | No. of affected genes | No. | Percent  |
|-------|-----------------------|-----|----------|
| GAP   |                       | 906 | 76.3912  |
| DUP   |                       | 0   | 0        |
| BRK   |                       | 0   | 0        |
| JMP   |                       | 19  | 1.60202  |
| INV   |                       | 4   | 0.337268 |
| SEQ   |                       | 270 | 22.7656  |
| Total |                       | 1186|          |
| chr  | Payzawat start | Payzawat end | Payzawat length | CM3.6.1 chr | CM3.6.1 start | CM3.6.1 end | CM3.6.1 length |
|------|----------------|--------------|-----------------|-------------|---------------|-------------|----------------|
| chr06| 16,635,346     | 18,619,220   | 1,983,875       | chr06       | 17,483,80     | 19,510,21   | 2,026,417     |
| chr06| 18,682,858     | 24,319,858   | 5,637,001       | chr06       | 19,585,83     | 25,280,50   | 5,694,689     |
| chr06| 24,404,623     | 28,895,149   | 4,490,527       | chr06       | 25,796,22     | 30,486,08   | 4,689,788     |
| chr06| 28,975,754     | 30,534,865   | 1,559,112       | chr06       | 30,835,56     | 32,330,17   | 1,494,558     |
| chr06| 30,563,968     | 36,185,168   | 5,621,201       | chr06       | 29,270,98     | 38,297,73   | 9,026,392     |
**Table S28. Validation of large inversions detected in chromosome chr06. Related to Figure 2.**

| Chr | Payzawat start | Payzawat end | Left breakpoint start | Left breakpoint end | Right breakpoint start | Right breakpoint end |
|-----|----------------|--------------|-----------------------|---------------------|------------------------|----------------------|
| chr06 | 16,635,346 | 18,619,220 | 16,557,453 | 16,635,346 | 18,619,220 | 18,682,858 |
| chr06 | 18,682,858 | 24,319,858 | 18,619,220 | 18,682,858 | 24,319,858 | 24,344,072 |
| chr06 | 24,404,623 | 28,895,149 | 24,355,748 | 24,404,623 | 28,895,149 | 28,975,754 |
| chr06 | 28,975,754 | 30,534,865 | 28,895,149 | 28,975,754 | 30,534,865 | 30,545,753 |
| chr06 | 30,563,968 | 36,185,168 | 30,534,865 | 30,563,968 | 30,545,753 | 30,545,753 |
| SRR         | Library strategy | Description                                      | rind thickness |
|-------------|------------------|--------------------------------------------------|----------------|
| SRR6320593  | RNAseq           | fruit flesh and peel at 10 days thin rind post anthesis | PI161375 (non-climacteric) |
| SRR7343307  | RNAseq           | fruit flesh and peel at 20 days thin rind post anthesis | PI161375 (non-climacteric) |
| SRR5405113  | RNAseq           | fruit flesh and peel at 40 days thin rind post anthesis | PI161375 (non-climacteric) |
| SRR5405110  | RNAseq           | flesh and peel at 10 days thick rind post anthesis | Vedrantais (climacteric) fruit |
| SRR5405109  | RNAseq           | flesh and peel at 20 days thick rind post anthesis | Vedrantais (climacteric) fruit |
| SRR5405108  | RNAseq           | flesh and peel at 30 days thick rind post anthesis | Vedrantais (climacteric) fruit |
| SRR5405107  | RNAseq           | flesh and peel at 40 days thick rind post anthesis | Vedrantais (climacteric) fruit |
| SRR1501088  | RNAseq           | No. 4’ fruits at 0 days post anthesis            | C. melo var ameri, cv ‘Flavor’ |
| SRR1501219  | RNAseq           | No. 4’ fruits at 10 days post anthesis           | C. melo var ameri, cv ‘Flavor’ |
| SRR1503342  | RNAseq           | No. 4’ fruits at 20 days post anthesis           | C. melo var ameri, cv ‘Flavor’ |
| SRR1503343  | RNAseq           | No. 4’ fruits at 30 days post anthesis           | C. melo var ameri, cv ‘Flavor’ |
| SRR1503344  | RNAseq           | No. 4’ fruits at 40 days post anthesis           | C. melo var ameri, cv ‘Flavor’ |
| Sample NO | Accession name          | Group            | Orgin     | Botanical variety                  |
|-----------|-------------------------|------------------|-----------|------------------------------------|
| R13       | African Horn Melon      | wild relative species | African   | Cucumis metulifeus                 |
| R22       | PI 614512               | wild             | India     | C. agrestis var. agrestis          |
| R12       | Agretis                 | wild             | India     | C. agrestis var. agrestis          |
| R26       | PI 180283               | wild             | India     | C. agrestis var. agrestis          |
| R32       | PI 614307               | wild             | India     | C. agrestis var. agrestis          |
| R31       | PI 614173               | wild             | India     | C. agrestis var. agrestis          |
| R33       | PI 614309               | wild             | India     | C. agrestis var. agrestis          |
| R30       | PI 536473               | wild             | Maldives  | C. agrestis var. agrestis          |
| R28       | PI 406737               | wild             | Costarica | C. agrestis var. agrestis          |
| R29       | PI 532829               | wild             | Shanxi, China | C. agrestis var. agrestis       |
| R34       | Mapaogua                | wild             | Shandong, China | C. agrestis var. agrestis       |
| R11       | Kangua                  | wild             | China     | C. agrestis var. chinensis        |
| R24       | Zhonghuacaigua          | cultivar/landrace | Henan, China | C. agrestis var. conomon        |
| R01       | Xianggua                | wild             | China     | C. agrestis var. chinensis        |
| R09       | PI 161375               | landrace         | Korea     | C. agrestis var. chinensis        |
| R17       | Yangjiaosu              | cultivar/landrace | Hebei, China | C. agrestis var. chinensis       |
| R15       | Huangpimiangua          | cultivar/landrace | Henan, China | C. agrestis var. chinensis       |
| R20       | Huangjingua             | cultivar/landrace | Japan     | C. agrestis var. chinensis        |
| R14       | Tinglinxuegua           | cultivar/landrace | Shanghai, China | C. agrestis var. chinensis  |
| R38       | Baizhuogua              | cultivar/landrace | Zhejiang, China | C. agrestis var. chinensis       |
| R18       | Zhimasu                 | cultivar/landrace | Shanxi, China | C. agrestis var. chinensis       |
| R39       | Bingtangzi              | cultivar/landrace | Shandong, China | C. agrestis var. chinensis       |
| R08       | Yagua                   | cultivar/landrace | Northeast China | C. agrestis var. chinensis   |
| R16       | Balixiang               | cultivar/landrace | Jilin, China | C. agrestis var. chinensis        |
| R19       | Tiebaqing               | cultivar/landrace | Liaoning, China | C. agrestis var. chinensis       |
| R07       | PI 414723               | wild             | India     | C. agrestis var. momordica        |
| R04       | PI 371795               | wild             | India     | C. agrestis var. momordica        |
| R06       | PI 313970               | wild             | India     | C. agrestis var. momordica        |
| R03       | PI 446928               | cultivar         | Israel    | C. agrestis var. momordica        |
| Code | Name          | Type          | Origin        | Variety                |
|------|---------------|---------------|---------------|------------------------|
| R36  | MR-1          | breeding material | USA         | C. agrestis var. momordica |
| R37  | PI 124111     | wild          | India         | C. agrestis var. momordica |
| R02  | Im-2          | landerce      | Italy         | C. melo var. chandalak   |
| R23  | Edisto 47     | cultivar      | America       | C. melo var. reticulatus |
| R35  | PI 446929     | cultivar      | Israel        | C. melo var. chandalak   |
| R25  | Fcousss       | landerce      | France        | C. melo var. flexuosus   |
| R41  | Tashikehong   | Landrace      | Xinjiang, China | C. melo var. chandalak   |
| R40  | Im41          | cultivar/landrace | Italy       | C. melo var. inodorus   |
| R05  | PI 234607     | cultivar/landrace | South Africa | C. melo var. inodorus   |
| R10  | Pil de sapo   | cultivar/landrace | Spain       | C. melo var. inodorus   |
| R49  | Boxiekexin    | Landrace      | Xinjiang, China | C. melo var. cassaba    |
| R21  | PI 222187     | landerce      | Afghanistan   | C. melo var. flexuosus   |
| R27  | Minoo         | cultivar/landrace | Iran        | C. melo var. ameri      |
| R43  | Bokezade      | Landrace      | Xinjiang, China | C. melo var. ameri      |
| R46  | Mizigua       | Landrace      | Xinjiang, China | C. melo var. ameri      |
| R44  | Baibicui      | Landrace      | Xinjiang, China | C. melo var. ameri      |
| R48  | Pishan        | Landrace      | Xinjiang, China | C. melo var. ameri      |
| R42  | Naxigan       | Landrace      | Xinjiang, China | C. melo var. ameri      |
| R45  | Huapijinbangzi | Landrace      | Xinjiang, China | C. melo var. ameri      |
| R47  | Hongxincui    | Landrace      | Xinjiang, China | C. melo var. ameri      |
| R50  | Huanghuapibair | Landrace      | Xinjiang, China | C. melo var. ameri      |
Table S31. Statistics of reads and mapping for 50 accessions. Related to Figure 4 and 5.

| Accession | Total_reads | Mapped(%) | Properly_mapped(%) | Average depth | Coverage 1X(%) | Coverage 5X(%) | Coverage 10X(%) |
|-----------|-------------|-----------|-------------------|---------------|----------------|----------------|----------------|
| R01       | 90,443,068  | 98.25     | 93.31             | 18            | 87.09          | 80.45          | 69.9           |
| R02       | 109,867,078 | 98.69     | 95.47             | 22            | 88.33          | 84.04          | 77.05          |
| R03       | 96,454,242  | 98.51     | 94.64             | 19            | 87.78          | 82.15          | 73.13          |
| R04       | 92,316,822  | 98.41     | 94.21             | 19            | 87.59          | 82.45          | 73.52          |
| R05       | 88,244,156  | 98.98     | 92.66             | 16            | 89.13          | 81.7           | 68.35          |
| R06       | 101,321,854 | 98.51     | 94.54             | 20            | 87.56          | 81.68          | 72.83          |
| R07       | 88,429,654  | 98.51     | 94.7              | 17            | 87.24          | 79.23          | 67.68          |
| R08       | 106,328,706 | 97.94     | 93.62             | 22            | 86.85          | 82.06          | 75.08          |
| R09       | 90,979,806  | 98.14     | 93.94             | 19            | 86.53          | 80.94          | 71.58          |
| R10       | 101,505,104 | 98.96     | 93.53             | 18            | 89.26          | 81.99          | 69.88          |
| R11       | 95,639,144  | 98.31     | 94.28             | 18            | 86.24          | 76.97          | 65.43          |
| R12       | 95,063,612  | 98.26     | 94.21             | 17            | 86.65          | 78.66          | 67.07          |
| R13       | 97,975,978  | 42.1      | 33.76             | 12            | 34.72          | 24.6           | 19.56          |
| R14       | 107,135,906 | 98.29     | 94.39             | 20            | 86.59          | 79.36          | 69.85          |
| R15       | 101,394,312 | 98.11     | 94.25             | 20            | 86.77          | 80.96          | 72.33          |
| R16       | 93,361,670  | 98.31     | 92.16             | 17            | 86.43          | 76.19          | 64.13          |
| R17       | 92,116,208  | 98.23     | 92.09             | 18            | 86.67          | 78.38          | 67.36          |
| R18       | 87,152,822  | 98.35     | 94.06             | 17            | 85.66          | 75.79          | 63.62          |
| R19       | 101,719,444 | 98.15     | 93.85             | 20            | 86.55          | 78.89          | 69.21          |
| R20       | 102,527,186 | 98.42     | 93.74             | 19            | 86.5           | 77.16          | 66.04          |
| R21       | 91,825,482  | 98.99     | 95.31             | 18            | 88.89          | 82.32          | 71.42          |
| R22       | 105,856,152 | 98.08     | 93.56             | 22            | 88.02          | 83.49          | 76.46          |
| R23       | 100,373,426 | 98.84     | 94.22             | 19            | 88.97          | 82.28          | 72.58          |
| R24       | 92,555,320  | 98.34     | 94.12             | 18            | 86.58          | 78.78          | 68.09          |
| R25       | 91,196,176  | 98.87     | 95.5              | 17            | 88.16          | 79.16          | 66.85          |
| R26       | 106,718,396 | 98.26     | 93.58             | 20            | 86.99          | 80.67          | 71.63          |
| R27       | 90,552,366  | 99.09     | 96.12             | 17            | 88.2           | 77.49          | 65.1           |
Table S32. Statistics of SNPs for 50 accessions. Related to Figure 4 and 5.

| Accession | SNP number | Transition | Transversion |
|-----------|------------|------------|--------------|
| R01       | 1299330    | 894092     | 405238       |
| R02       | 746957     | 512928     | 234029       |
| R03       | 967729     | 665461     | 302268       |
| R04       | 1115180    | 764894     | 350286       |
| R05       | 597754     | 409376     | 188378       |
| R06       | 1100618    | 756192     | 344426       |
| R07       | 1091484    | 750497     | 340987       |
| R08       | 1323638    | 910416     | 413222       |
| R09       | 1296099    | 891730     | 404369       |
| R10       | 599633     | 410889     | 188744       |
| R11       | 1286260    | 887636     | 398624       |
| R12       | 1340595    | 923456     | 417139       |
| R13       | 1397537    | 857300     | 540237       |
| R14       | 1306832    | 901132     | 405700       |
| R15       | 1321611    | 909679     | 411932       |
| R16       | 1292906    | 893231     | 399675       |
| R17       | 1294154    | 892706     | 401448       |
| R18       | 1272472    | 879242     | 393230       |
| R19       | 1299286    | 896103     | 403183       |
| R20       | 1285367    | 886944     | 398423       |
| R21       | 565455     | 387673     | 177782       |
| R22       | 1604550    | 1102928    | 501622       |
| R23       | 773922     | 531617     | 242305       |
| R24       | 1313675    | 905988     | 407687       |
| R25       | 818625     | 564057     | 254568       |
| R26       | 1364308    | 938448     | 425860       |
| R27       | 447215     | 307781     | 139434       |
| R28       | 1303077    | 899170     | 403907       |
| R29       | 1055085    | 726906     | 328179       |
| R30       | 1391303    | 959524     | 431779       |
| R31       | 1330168    | 916129     | 414039       |
| R32       | 1247410    | 859287     | 388123       |
| R33       | 1271175    | 876475     | 394700       |
| R34       | 1295543    | 893604     | 401939       |
| R35       | 677096     | 465094     | 212002       |
| R36       | 878104     | 606031     | 272073       |
| R37       | 828899     | 571299     | 257600       |
| R38       | 1312911    | 904489     | 408422       |
|   |      |      |      |
|---|------|------|------|
| R39 | 1303084 | 897891 | 405193 |
| R40 | 527224  | 360712 | 166512 |
| R41 | 544124  | 373776 | 170348 |
| R42 | 339382  | 231687 | 107695 |
| R43 | 337273  | 231665 | 105608 |
| R44 | 335147  | 229099 | 106048 |
| R45 | 299581  | 205159 | 94422  |
| R46 | 322078  | 220150 | 101928 |
| R47 | 289933  | 198764 | 91169  |
| R48 | 275650  | 188896 | 86754  |
| R49 | 372668  | 254272 | 118396 |
| R50 | 274280  | 186742 | 87538  |
Table S33. Statistics of structure variation for 50 accessions. Related to Figure 4.

| Accession | SV   | INS  | DEL  | INV  | ITX  | CTX  | UN  |
|-----------|------|------|------|------|------|------|-----|
| R01       | 27886| 9714 | 12359| 637  | 637  | 4513 | 26  |
| R02       | 22949| 9549 | 7466 | 780  | 555  | 4572 | 27  |
| R03       | 20644| 5679 | 9071 | 573  | 627  | 4667 | 27  |
| R04       | 22915| 6992 | 10158| 633  | 588  | 4514 | 30  |
| R05       | 23763| 13839| 5065 | 462  | 584  | 4792 | 23  |
| R06       | 22768| 6359 | 10447| 563  | 584  | 4792 | 23  |
| R07       | 19769| 5323 | 8832 | 499  | 597  | 4491 | 27  |
| R08       | 29659| 10378| 12856| 745  | 721  | 4937 | 22  |
| R09       | 30381| 11494| 13095| 665  | 581  | 4518 | 28  |
| R10       | 23462| 13054| 5238 | 457  | 627  | 4059 | 27  |
| R11       | 27914| 10466| 11868| 473  | 626  | 4455 | 26  |
| R12       | 23753| 6012 | 11803| 508  | 676  | 4715 | 39  |
| R13       | 8225 | 1592 | 2745 | 149  | 198  | 3502 | 39  |
| R14       | 29825| 12141| 11718| 506  | 645  | 4789 | 26  |
| R15       | 28206| 9270 | 12270| 537  | 684  | 4870 | 25  |
| R16       | 25673| 10878| 10967| 503  | 616  | 2685 | 24  |
| R17       | 24345| 8829 | 11770| 522  | 593  | 2602 | 29  |
| R18       | 25970| 9205 | 11438| 468  | 564  | 4273 | 22  |
| R19       | 29133| 11246| 12270| 520  | 627  | 4444 | 26  |
| R20       | 28005| 11932| 10495| 487  | 618  | 4451 | 22  |
| R21       | 16805| 6058 | 5911 | 384  | 550  | 3874 | 28  |
| R22       | 28506| 7817 | 14313| 728  | 746  | 4859 | 43  |
| R23       | 25237| 12445| 7425 | 437  | 624  | 4284 | 22  |
| R24       | 26891| 10339| 10915| 505  | 611  | 4494 | 27  |
| R25       | 20514| 7685 | 7695 | 371  | 539  | 4196 | 28  |
| R26       | 29538| 11451| 11963| 543  | 670  | 4880 | 31  |
| R27       | 17668| 8241 | 4409 | 332  | 461  | 4209 | 16  |
| R28       | 22509| 6949 | 10324| 493  | 652  | 4063 | 28  |
| R29       | 25747| 10522| 9753 | 461  | 576  | 4410 | 25  |
| R30       | 25294| 7378 | 12256| 525  | 698  | 4409 | 28  |
| R31       | 27038| 8880 | 12300| 509  | 689  | 4641 | 19  |
| R32       | 26927| 9340 | 11402| 560  | 686  | 4907 | 32  |
| R33       | 24917| 7568 | 11678| 511  | 663  | 4475 | 22  |
| R34       | 27417| 10921| 11002| 514  | 630  | 4323 | 27  |
| R35       | 16822| 4580 | 6890 | 392  | 583  | 4355 | 22  |
| R36       | 17871| 4246 | 7995 | 428  | 582  | 4592 | 28  |
| R37       | 17838| 4491 | 7680 | 434  | 550  | 4659 | 24  |
| R38       | 27272| 9670 | 11834| 515  | 626  | 4599 | 28  |
|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| R39 | 29733 | 11863 | 11902 | 512 | 647 | 4775 | 34 |
| R40 | 24479 | 13335 | 5691 | 388 | 562 | 4476 | 27 |
| R41 | 19942 | 8975 | 5502 | 356 | 552 | 4539 | 18 |
| R42 | 18477 | 9621 | 3904 | 319 | 477 | 4144 | 12 |
| R43 | 15931 | 7310 | 3784 | 302 | 434 | 4078 | 23 |
| R44 | 12833 | 3560 | 4185 | 362 | 477 | 4226 | 23 |
| R45 | 18488 | 9953 | 3672 | 305 | 479 | 4064 | 15 |
| R46 | 13626 | 4427 | 3973 | 330 | 457 | 4424 | 15 |
| R47 | 21010 | 12482 | 3621 | 299 | 490 | 4100 | 18 |
| R48 | 19897 | 11420 | 3356 | 263 | 465 | 4375 | 18 |
| R49 | 22614 | 12704 | 4443 | 324 | 528 | 4584 | 31 |
| R50 | 12859 | 3696 | 3545 | 340 | 503 | 4756 | 19 |
Table S34. 275 mutations implicated in the contrasting phenotypes between wild and cultivated populations. Related to Figure 5.
Transparent Methods

1 Genome sequencing and assembly

1.1 Sample preparation and sequencing

The strain for sequencing was a cultivated one (Cucumis melo L. var. inodorus) from Payzawat county in Kashgar, Sinkiang. Usually the Payzawat natives call it Payzawat melon and the others call it Kalakesai. The DNA of the melon was extracted from leaf following a previous published protocol.

1.2 Genome size estimation and assembly

A total of 32.53 Gb of Illumina PE (2×150 bp) clean bases were generated with about 81-fold coverage (see Tables S1). The genome survey analysis was performed using “kmer freq stat” software (developed by Biomarker Technologies). The highest peak in the k-mer distribution curve was at 64. The peak at depth of more than 128 was a repetitive peak. Finally, the melon genome size was estimated to be 398.57 Mb, with about 0.03% heterozygosity and 43.39% repeat sequences (see Figure S1).

For SMRT sequencing, a total of 4,057,884 polymerase reads (30.78 Gb) were generated on the PacBio RSII platform (see Tables S2a and S2b). In quality control step, only subreads with length more than 500 bp and RQ value higher than 0.75 were retained for future analysis. Finally, A total of 2,643,196 PacBio post-filtered reads which accounting for approximately 27.80 Gb sequencing data were generated. The subreads N50 reached 13,947 and the average subread length was 10.52 kb. Canu (v1.5)(Koren S, et al., 2017) assembler and DBG2LOC(Ye C, Hill CM, Wu S, Ruan J, Ma ZS, 2016) were used for de novo assembly. To improve the assembly result, Quickmerge(Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ, 2016) software was used to merge the two genome assemblies (see Tables S3). The genome was first polished by Quiver (--minLength 3000 --maxHits 1) and followed by three round Pilon (v1.22, --mindepth 20 --fix base)(Walker BJ, et al., 2014) using the Illumina short reads.

1.3 Chromosome construction and evaluation

To anchor the contigs into pseudo-chromosomes, we performed HI-C experiments. A total of 49.07 Gb clean data were sequenced using the Illumina X-Ten platform, with about 127-fold coverage (see Table S21). The clean reads were mapped to the contigs using BWA aln method with default parameters and 46.32% reads were uniquely mapped. 69.91% of the unique mapped reads were identified as the valid
interaction read pairs with paired HindIII restriction site by HiC-Pro (Servant N, et al., 2015) (see Table S21). Before anchoring the contigs, we first performed a contigs correction step. In this step, the contigs were firstly interrupted at 500 kb window, and then the valid linked reads were mapped to the interrupted contigs. Once the broken contigs could not be linked back to the original contigs, the original contigs were interrupted at a low coverage site. After correction, the contig N50 went to 2.31 Mb (see Table S22). Next, the pseudochromosomes were constructed using LACHESIS (Burton JN, et al., 2013) software with: CLUSTER MIN RE SITES = 71, CLUSTER MAX LINK DENSITY = 2, CLUSTER NONINFORMATIVE RATIO = 2, ORDER MIN N RES IN TRUNK = 59, and ORDER MIN N RES IN SHREDS = 59.

1.4 Genome assembly

The assembly completeness was assessed by means of three data sets: Genetic map, PE short reads, Pacbio subreads, CEGMA, and BUSCO. To evaluate the quality of our new assembly, the short reads used for estimation of genome size were realigned to the assembled genome using the bwa mem method with default parameters (see Table S5) (Li H, Durbin R, 2010). PacBio subreads were aligned to the assembly using blasr with parameters “-bestn 1 -minPctIdentity 70 -nproc 4” (see Table S6). Besides, we found that the single-base error rate was 0.0003374% (see Table S9), suggesting that the base of the genome was accurate. Genome completeness was also assessed using the core eukaryotic gene-mapping approach (CEGMA v2.5) (Parra G, 2007). It provides a rapid method of assessing genome completeness because it comprises a set of highly conserved, single-copy genes, present in all eukaryotes. CEGMA contains 458 core eukaryotic genes (CEGs) and among them 248 are highly-conserved CEGs. The BUSCO (Version 2) program was run against the embryophyta_odb9 dataset with the default parameters.

1.5 Centromere and telomere identification

Melon-specific CentM repeats (GenBank accession no. 3929695) were identified through blasting (BLAST, Version 2.2.31, E-value < 1 × 10^{-5}) search CentM sequences against the genome sequences (centromere repositioning in cucurbit species: Implication of the genomic impact from centromere activation and inactivation, Organization of highly repetitive satellite DNA of two Cucurbitaceae species (Cucumis melo and Cucumis sativus)). Repeats shorter than 1,000 bp were discarded. The repeat clusters between 1000 bp or overlapped clustered were merged. Telomeric repeats (TTTAGGG/CCCTAAA) were identified by tandem repeat finder (TRF) (Version 4.07b) (G. Benson, 1999).
2 Genome annotation

2.1 Repeat content identification

Both homolog and de novo strategies were applied to identify the repetitive sequence in melon genome. There were four de novo prediction software suites, including RepeatScout(Price AL, Jones NC, Pevzner PA, 2005), LTR-FINDER(Xu Z, Wang H, 2007), MITE-Hunter(Han Y, Wessler SR, 2010), and PILER-DF(Edgar RC, Myers EW, 2005), that were adopted for ab initio prediction. RepeatScout can find all repeat classes. LTR-FINDER was used to predict locations and structure of full-length LTR retrotransposons. MITE-Hunter was used to discover miniature inverted-repeat transposable elements (MITEs) from genomic sequences. PILER-DF is suitable when finding repeated elements such as satellites and transposons. Results from all ab initio predictions were combined to form a repetitive sequence library. The library was then merged with RepBase(Bao W, Kojima KK, Kohany O, 2015) and classified into different categories by the PASTECclassifier.py script included in REPET. The repetitive sequences in melon genome were identified by homolog searching in that database through RepeatMasker(Tarailo-Graovac M, Chen N, 2009).

2.2 Gene model, ncRNA and pseudogene prediction

Gene model annotation was performed using a combination of three methods: ab initio prediction, homology-based gene prediction, and transcript evidence from RNA-seq data. The ab initio prediction was conducted using five different suites of software: Genscan(Burke C, Karlin S, 1997), Augustus v2.4(Stanke M, Waack S, 2003), GlimmerHMM v3.0.4(Majerus WH, Pertea M, Salzberg SL, 2004), GeneID v1.4(Blanco E, Parra G, Guigo R, 2007), SNAP(version 2006-07-28)(Korf I, 2004), and GeMoMa (v1.3.1)(Keilwagen J, et al., 2016) was employed for gene prediction based on homologous genes, using model training based on coding sequences from Arabidopsis thaliana, Oryza sativa, Cucumis sativus (Cultivar: Borszczagowski), Cucumis sativus (Cultivar: 9930), and the previously sequenced Cucumis melon. The RNA-seq reads generated from various tissues of melon were assembled without a reference genome using Hisat v2.0.4(Kim D, Langmead B, Salzberg SL, 2015) and Stringtie v1.2.3(Pertea M, et al., 2015), then the unigenes were used to predict genes using PASA v2.0.2(Campbell MA, et al., 2006). Finally, we used EVM v1.1.1(Haas BJ, et al., 2008) to combine the results of the three methods. The ncRNAs were also predicted including miRNAs, tRNAs and rRNAs. tRNAs was detected by using tRNA-scan-SEM (version 1.23)(Lowe TM, Eddy SR, 1997). The miRNAs and rRNAs were found based on sequence alignment using BLASTN v2.26 with an e-value of less than 1 × 10^{-10} and Infernal v1.1(Nawrocki EP, Eddy SR, 2013) against Rfam v12.0(Nawrocki EP, 2015) database. A total of 1,235 non-coding RNAs were also identified in our melon assembly, including 75 microRNAs of 19 families, 443 rRNAs
of four families, and 717 tRNAs of 24 families (Table S18). The GeneWise(Birney E, Clamp M, Durbin R, 2004) software was used to identify pseudogenes, and a total of 2,434 pseudogenes were identified in melon genome (Tables S19).

2.3 Gene functional annotation analysis

Gene function was annotated by comparing their protein sequences against a number of protein sequence databases, including: Nr (NCBI non-redundant protein sequences), Nt (NCBI non-redundant nucleotide sequences), Pfam (Protein family); KOG (Clusters of Orthologous Groups of proteins), Swiss-Prot (a manually annotated and reviewed protein sequence database), KO (KEGG Ortholog database), and GO (Gene Ontology).

3 Whole-genome sequencing comparison

A whole-genome comparison between Payzawat and CM3.6.1 was performed using nucmer module of MUMmer package (version 4.0)(Marcais, G. et al., 2018). Structural variations were identified from the one-to-one syntenic blocks.

4 Resequencing and diversity analysis

4.1 Library construction and sequencing

A total of 49 accessions were collected (Table S30). Fresh young leaves were collected from each individual and immediately frozen in liquid nitrogen. DNA was extracted using the CTAB method(Murray MG, Thompson WF, 1980). Some 5 ug DNA for each individual could be used for library construction with c. 350 bp insert size. Libraries were sequenced on an Illunima Hiseq X Ten (2 x 150 bp). Raw reads were subjected to removal of adaptors and trimmed of low-quality bases and then 599.39-Gb clean data were obtained (approximately 12X coverage for each individual).

4.2 Reads mapping and variants calling

High-quality PE reads for each accession were mapped to the genome using the BWA program (Version 0.7.10, mapping method: MEM) with default parameters. Duplicated reads were discarded using the MarkDuplicates utility in Picard(http://sourceforge.net/projects/picard/). Variants were called using the HaplotypeCaller module in GATK (version 3.1.1)(McKenna A, et al., 2010). To obtain high-quality variants, stringent filtering was performed as follows: 1) clusterSize 2 clusterWindowSize 5; 2) QUAL < 30; 3) QD < 2.0; 4) MQ < 40; 5) FS > 60.6. 2,702,589 high-quality SNPs and 370,207 indels were retained for further analysis.
The identified SNPs and indels were annotated using SnpEff software (Version 3.6c)(Cingolani P, et al., 2012). BreakDancer-1.1(Chen K, et al., 2009) was used to identify large structural variants(see Table 33).

4.3 Population diversity

Population diversity was analysed using three methods, including phylogenetic tree construction, population stratification using ADMIXTURE(Alexander D H, Novembre J, Lange K, 2009), and high-quality SNPs (MAF = 0.05, missing rate = 0.5) were used to construct the phylogenetic relationships among 49 accessions using MEGA X(Kumar, Sudhir, et al., 2018) with a neighbour-joining method with Kimura 2-parameter model and 1,000 bootstrap replicates. Population stratification among 49 accessions was inferred using ADMIXTURE software(Alexander D H, Novembre J, Lange K, 2009). Cross-validation error was tested for K (the most likely number of clusters) from 2 to 5 and the population with the smallest cross-validation error was chosen and results was plotted for K = 2 to K = 5.

4.4 Linkage disequilibrium (LD)

A total of 2,248,795 high-quality SNPs (MAF = 0.05, missing rate = 0.2) in 49 accessions were used to perform LD using Software plink(Purcell S, et al., 2007) through calculating the squared correlation coefficient ($r^2$) between any two SNPs within a 100 kb window.

4.5 Selective sweep

SNPs with minor allele frequency below 5% and missing rate above 0.2 were excluded. Software PopGenome(Pfeifer B, et al., 2014) was used to calculate Fst and pai with a 100 kb sliding window and step size of 10 kb.
Supplemental references

Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ: Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res 2016, 44:e147.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM: Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014, 9:e112963.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E: HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 2015, 16:259.

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J: Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol 2013, 31:1119-1125.

Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010, 26:589-595.

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

G. Benson,"Tandem repeats finder: a program to analyze DNA sequences"Nucleic Acids Research (1999) Vol. 27, No. 2, pp. 573-580.

Price AL, Jones NC, Pevzner PA: De novo identification of repeat families in large genomes. Bioinformatics 2005, 21 Suppl 1:i351-358.

Xu Z, Wang H: LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 2007, 35:W265-268.

Han Y, Wessler SR: MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res 2010, 38:e199.

Edgar RC, Myers EW: PILER: identification and classification of genomic repeats. Bioinformatics 2005, 21 Suppl 1:i152-158.

Bao W, Kojima KK, Kohany O: Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 2015, 6:11.

Tarailo-Graovac M, Chen N: Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics 2009, Chapter 4:Unit 4 10.

Burge C, Karlin S: Prediction of complete gene structures in human genomic DNA. J Mol Biol 1997, 268:78-94.

Stanke M, Waack S: Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 2003, 19 Suppl 2:i215-225.

Majoros WH, Pertea M, Salzberg SL: TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 2004, 20:2878-2879.

Blanco E, Parra G, Guigo R: Using geneid to identify genes. Curr Protoc Bioinformatics 2007, Chapter 4:Unit 4 3.

Korf I: Gene finding in novel genomes. BMC Bioinformatics 2004, 5:59.

Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F: Using intron position conservation for homology-based gene prediction. Nucleic Acids Res 2016, 44:e89.
Kim D, Langmead B, Salzberg SL: HISAT: A fast spliced aligner with low memory requirements. Nat Methods 2015, 12:357-360.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL: StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 2015, 33:290-295.

Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR: Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. BMC Genomics 2006, 7:327.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR: Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol 2008, 9:R7.

Lowe TM, Eddy SR: tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997, 25:955-964.

Nawrocki EP, Eddy SR: Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 2013, 29:2933-2935.

Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD: Rfam 12.0: updates to the RNA families database. Nucleic Acids Res 2015, 43:D130-137.

Binney E, Clamp M, Durbin R: GeneWise and Genomewise. Genome Res 2004, 14:988-995.

Marcais, G. et al. MUMmer4: a fast and versatile genome alignment system. PLoS Comput. Biol. 14, e1005944 (2018).

Murray MG, Thompson WF: Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res,1980. 8:4321-5.

Picard: http://sourceforge.net/projects/picard/ (Picard).

McKenna A, Hanna M, Banks E, Sivachenko A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 20:1297-303.

Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3., Fly (Austin). 2012 Apr-Jun;6(2):80-92.

Chen K, Wallis J W, Mcellan M D, et al. BreakDancer: An algorithm for high resolution mapping of genomic structural variation[J]. Nature Methods, 2009, 6(9): 677-681.

Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome research, 2009, 19(9): 1655-1664.

Kumar, Sudhir, et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular biology and evolution 35.6 (2018): 1547-1549.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

Pfeifer B, Wittelsburger U, Ramosonsins S E, et al. PopGenome: an efficient Swiss army knife for population genomic analyses in R,[J]. Molecular Biology and Evolution, 2014, 31(7): 1929-1936.