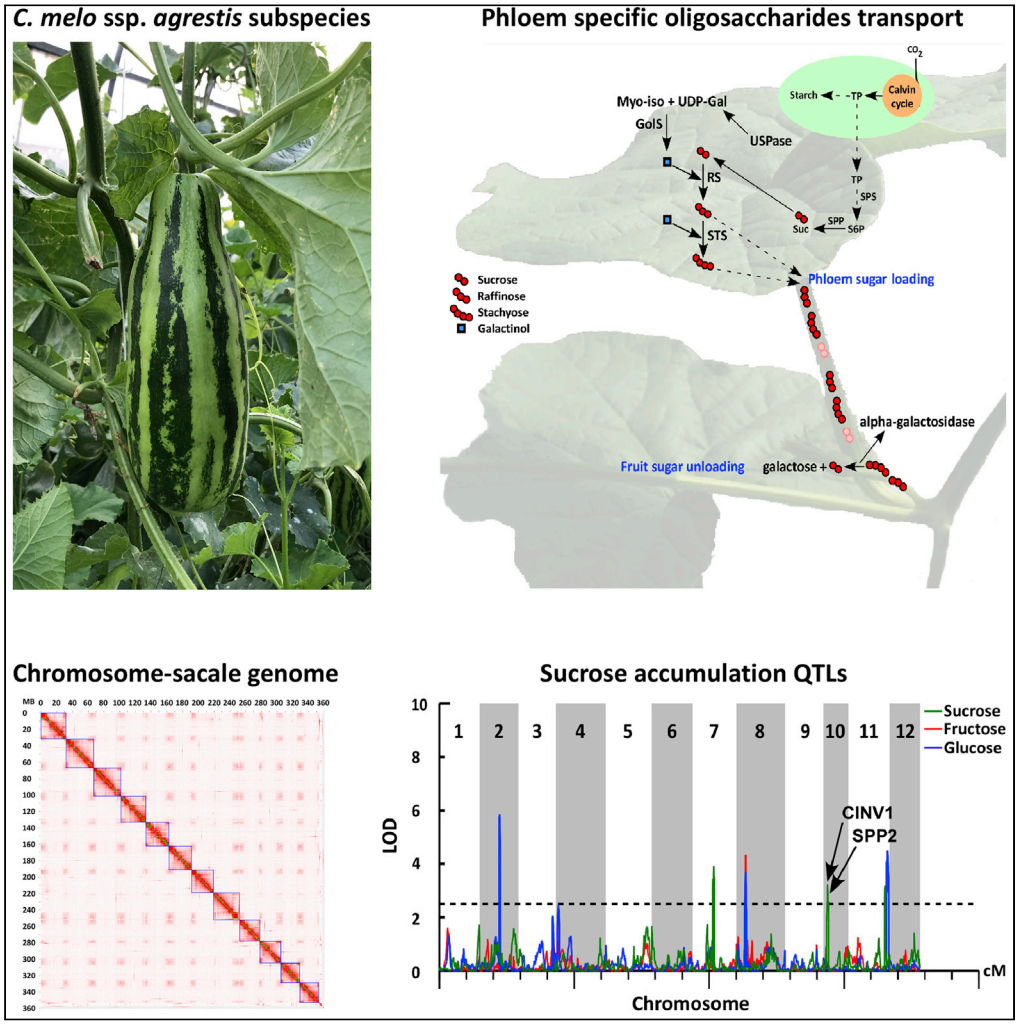
iScience

CellPress
OPEN ACCESS

Article

The Chromosome-Scale Genome of Melon Dissects Genetic Architecture of Important Agronomic Traits



Jinghua Yang, Guancong Deng, Jinmin Lian, ..., Zhongyuan Hu, Jingquan Yu, Mingfang Zhang

jqyu@zju.edu.cn (J.Y.)
mfzhang@zju.edu.cn (M.Z.)

HIGHLIGHTS

A chromosome-scale genome of *Cucumis melo* ssp. *agrestis* subspecies

The work provides insight into oligosaccharides transport in *Cucurbitaceae*

Findings provide insight into the epigenetic regulation of sucrose accumulation in developing fruits

Genetic architectures of sucrose accumulation and gummy stem blight resistance

Yang et al., iScience *23*, 101422
August 21, 2020
https://doi.org/10.1016/j.isci.2020.101422

iScience

CellPress
OPEN ACCESS

Article

The Chromosome-Scale Genome of Melon Dissects Genetic Architecture of Important Agronomic Traits

Jinghua Yang,[1,2] Guancong Deng,[1] Jinmin Lian,[3] Jenella Garraway,[1] Yongchao Niu,[3] Zhongyuan Hu,[1] Jingquan Yu,[1,2,*] and Mingfang Zhang[1,2,4,*]

SUMMARY

Comparative and evolutionary genomics analyses are the powerful tools to provide mechanistic insights into important agronomic traits. Here, we completed a chromosome-scale assembly of the "neglected" but vital melon subspecies *Cucumis melo* ssp. *agrestis* using single-molecule real-time sequencing, Hi-C, and an ultra-dense genetic map. Comparative genomics analyses identified two targeted genes, *UDP-sugar pyrophosphorylase* and *α-galactosidase*, that were selected during evolution for specific phloem transport of oligosaccharides in *Cucurbitaceae*. Association analysis of transcriptome and the DNA methylation patterns revealed the epigenetic regulation of sucrose accumulation in developing fruits. We constructed the melon recombinant inbred lines to uncover *Alkaline/Neutral Invertase* (*CINV*), *Sucrose-Phosphatase 2* (*SPP2*), *α-galactosidase*, and *β-galactosidase* loci related to sucrose accumulation and an LRR receptor-like serine/threonine-protein kinase associated with gummy stem blight resistance. This study provides essential genomic resources enabling functional genomics studies and the genomics-informed breeding pipelines for improving the fruit quality and disease resistance traits.

INTRODUCTION

Melon (*Cucumis melo* L., 2n = 24) is one of the most important fruit crops in the world belonging to *Cucurbitaceae*, with 32 million tons of fresh melons and 1 million tons of melon seeds produced in 2017 (http://www.fao.org/faostat/). After an intense process of diversification during domestication that has been proposed to have originated from both Africa and Asia (Sebastian et al., 2010), *C. melo* is considered to be divided into two subspecies, *melo* and *agrestis* (Jeffrey, 1980), with each recently separated into several cultivar groups (Pitrat, 2008). The draft genome sequence of melon, a double-haploid line derived from the cross between *inodorus* and *agrestis* subspecies, has been used in genetic studies, using short-reads sequencing strategy (Garcia-Mas et al., 2012), and the genome has recently been improved with single-molecule sequencing method (Castanera et al., 2020). In addition, another melon subspecies (*C. melo* spp. *melo*) genome has been just released (Zhang et al., 2019). Importantly, a genomics-enabled breeding strategy requires a thorough and robust understanding of genomic organization in melon subspecies.

Cucurbit crops specifically transport oligosaccharides raffinose and stachyose via the phloem from source to sink (Rennie and Turgeon, 2009). However, the evolutionary mechanism of the oligosaccharides transport, known as polymer trapping, remains poorly understood. The *conomon* melon (*C. melo* ssp. *agrestis*), also named oriental melon, has particular aromatic flavor and taste, environmental adaptation, and disease resistance (Zhang and Li, 2005; Wang et al., 2011; Hu et al., 2018). Sucrose metabolism has been well documented because of its pivotal roles in fruit quality (Ren et al., 2018; Guo et al., 2019), development, stress response, and yield formation (Wan et al., 2018; Ruan, 2014). Gummy stem blight (GSB) is a fatal fungus disease affecting most *Cucurbit* species, causing severe yield losses, especially in humid tropics and sub-tropics (Hu et al., 2018). Some GSB-resistance genetic loci have been reported in different resistant germplasm (Hu et al., 2018); however, a functional GSB-resistant gene has not been identified so far.

In this study, we employed the PacBio SMRT long reads and high-throughput chromosome conformation capture (Hi-C) technologies to assemble for the genome of melon subspecies of *C. melo* spp. *agrestis* (HS, a few-sucrose accumulator and GSB-resistant landrace). We anchored 359 Mb of the assembly onto 12

[1]Laboratory of Germplasm Innovation and Molecular Breeding, Institute of Vegetable Science, Zhejiang University, Hangzhou 310058, China

[2]Key Laboratory of Horticultural Plant Growth and Development, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, China

[3]Biozeron Shenzhen, Inc., Shenzhen 518081, China

[4]Lead Contact

*Correspondence: jqyu@zju.edu.cn (J.Y.), mfzhang@zju.edu.cn (M.Z.)

https://doi.org/10.1016/j.isci.2020.101422

pseudo-chromosomes of HS, estimating to cover above 98% of the assembly with significant completeness than extant published melon genomes. We pinpointed the candidate genes involved in specific phloem transport of oligosaccharides in *Cucurbitaceae*. In addition, we presented the relevant evidence of fruit quality regulated by epigenetic factors, particularly regulating sugar accumulation. We mapped the genetic loci associated with sucrose accumulation and gummy stem blight resistance using a recombinant inbred line population. The genomic and genetic resources developed in the present study will further empower the implementation and acceleration of genomic breeding in melon crops.

RESULTS AND DISCUSSION

De Novo Genome Sequencing and Assembly

We generated 35.6 Gb data of the SMRT sequencing long reads for a melon subspecies (HS, *Cucumis melo* ssp. *agrestis*), with an estimated depth of genome coverage of 100X (Table S1 and Figure S1). Illumina paired-end reads were used to polish the PacBio sequencing data (Table S2). This allowed us to assemble the genome into contigs using the Falcon pipeline and, subsequently, to construct the genome into super-scaffolds using the Hi-C interaction data with more than 85X genome coverage (31.0 Gb) (Figure 1A, S2, and S3 and Table S3). We then mapped genetic markers from previously constructed maps of two populations HS x XH (Hu et al., 2018) and PS x SC (Argyris et al., 2015) linkage maps to the contigs of HS genome and constructed a collinear genetic map and pseudo-chromosomes (Tables S4 and S5 and Figure S4). We recognized and split four breakpoints in two contigs after comparison between the HS genome assembly with the collinear genetic map (Table S6). As a result, the final assembly was 366 Mb with contigs N50 and N90 of 3.45 Mb and 926.7 Kb, respectively, and super-scaffold N50 of 29.76 Mb (Tables 1, S7, and S8). This assembly displayed considerable improvements in contiguity (gap filling) and completeness compared with the two published genome assemblies of melon varieties (*Cucumis melo* ssp. *melo*) DHL92_CM3.6.1 (Garcia-Mas et al., 2012) and Payzawat (Zhang et al., 2019) (Tables 1 and S7–S9). We finally anchored 98.15% (359.4 Mb) of the assembly onto 12 pseudo-chromosomes (Tables 1 and S8).

Genome Annotation and Chromosome-Scale Variations

A total of 157.6 Mb (43.05%) of repetitive sequences were annotated in the HS genome, of which only 4.48% was tandem repeats (Table S10). About 33.93% and 8.27% were class I and class II transposons, respectively (Tables S11 and S12 and Figure S5). The protein-coding genes were predicted through a combination of the prediction strategies (homology based, *de novo*, and transcriptome based). In total, 28,898 gene models with high fidelity were identified in the HS genome (Table S13), of which 86.29% were supported by RNA sequencing (RNA-seq) and 93.55% were annotated using at least one of the public databases (Table S14). The BUSCO assessment indicated that 91.0% of the conserved gene models were complete, with only 3.0% fragmented and 6.0% missing, indicating a high-quality assembly of the HS genome (Table S13). In addition, we simultaneously annotated the non-coding RNAs including 91 miRNAs, 778 tRNAs, 433 rRNAs, and 327 snRNAs (Table S15). Finally, we demonstrated the genomic annotation of the HS genome (Figure 1B).

The chromosome-scale genome enabled us to identify the large structural variations (SVs) and the complicated genome rearrangement. We observed some large SVs between the HS and the DHL92 genomes in Chr_1, Chr_5, Chr_6, Chr_10, and Chr_11, with the SVs validated by mapping of the PacBio sub-reads (Figures 1C and S6). The pseudo-chromosomes of HS were constructed after the collinear comparison of two genetic maps used for the HS and the DHL92 genomes assembly. We aligned Chr_06 of HS and DHL92 with the collinear genetic map to confirm that the conspicuous differences between the genetic map and DHL92 assembly resulted from the inverted assembly of the DHL92 genome (Figure 1D). Consequently, we supposed that the discrepancies in Chr_05 and Chr_10 (similar to those in Chr_06) originated from the inaccurate assembly of the DHL92 genome. The variations found in Chr_01, Chr_04, Chr_08, and Chr_11 among the genomes of HS, DHL92, and Payzawat varieties were suggested to be intra-chromosomal translocations and inversions (Figure 1B).

Evolutionary Analysis of Phloem Oligosaccharide Transport in Cucurbitaceae

The great majority of plants transport photo-assimilates in the form of sucrose from source to sink. However, plants in *Cucurbitaceae* and *Scrophulariaceae* families transport mainly other forms of sugar, such as raffinose and stachyose oligosaccharides (RFOs) (Turgeon and Wolf, 2009; Zhang and Turgeon, 2009). The transport of RFOs using polymer-trapping is an active phloem loading strategy, although it is the energy-consuming initial step of sucrose synthesis (Turgeon and Wolf, 2009). We constructed a phylogenetic tree using several major sequenced crops in *Cucurbitaceae*, with the single copy orthologous genes of

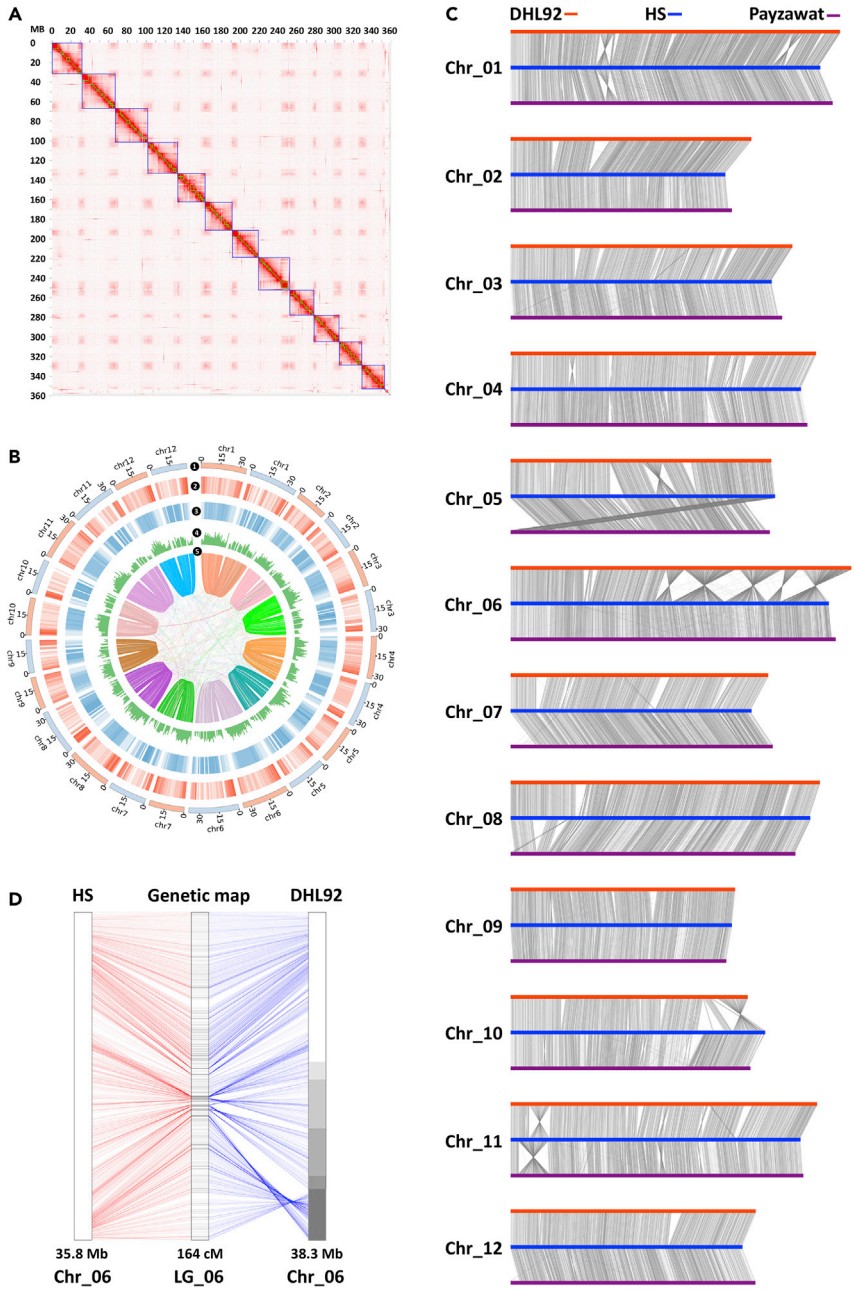


Figure 1. Characterization of Chromosome-Scale *C. melo* ssp. *agrestis* Genome

(A) The interaction frequency distribution of Hi-C links among HS chromosomes.

(B) Genomic landscape of the HS and DHL92 genomes. (1) Length of pseudo-chromosomes of HS (light red) and DHL92 (light blue); (2) gene density calculated on the basis of the number of genes in non-overlapping 1-Mb windows; (3) percent coverage of TEs in non-overlapping 1-Mb windows; (4) GC content in non-overlapping 1-Mb windows; (5) syntenic alignments between the HS and DHL92 genomes based on one-to-one orthologous genes.

(C) Genome alignment of pseudo-chromosomes in HS, DHL92, and Payzawat genomes displaying inverted assembly in DHL92.

(D) Presentation with lines connecting the physical positions on the pseudo-chromosome 6 and the map positions showing some inverted assembly regions relative to the genetic map in the assembly of DHL92.

Assembly Feature	HS	DHL92 (v3.6.1)[4]	DHL92 (v4.0)[5]	Payzawat[6]
Size of assembly (Mb)	366	417	358	386
Contig N50 (Mb)	3.45	0.023	0.714	2.86
Anchored pseudo-chromosomes (%)	98.15	90.01	96	95.53
Repeat content (%)	43.05	44	45.2	49.8
Annotated protein-coding genes	28,898	29,980	28,299	22,924
Complete BUSCOs (%)	91.8	91.9	94.8	92.78

Table 1. Comparison of Melon Assemblies
Complete BUSCOs are the BUSCO matches that scored within the expected range of scores and within the expected range of length alignments to the BUSCO profile.

Arabidopsis thaliana, *Solanum lycopersicum*, *Malus domestica*, and *Oryza sativa* as controls (Figures 2A and S7–S9 and Table S16). In contrast to *A. thaliana*, *S. lycopersicum*, *M. domestica*, and *O. sativa*, the *Cucurbitaceae* crops showed many gene family expansions significantly associated with β-galactosidase activity, carbohydrate binding, and carbohydrate metabolic process (Tables S17 and S18). The functional annotation of expanded gene families strongly suggested that carbohydrate transport is designedly presented in *Cucurbitaceae* crops.

The adaptive or selective advantage of active phloem loading allow plants to maintain low photo-assimilate concentrations in leaves (Turgeon, 2010). We proposed a schematic pathway of how the photo-assimilate products are transported from source to sink (Figure 2B). Using positive selection gene analysis, we discovered two candidate genes, *α-galactosidase* (*MEL O 06810*) and *UDP-sugar pyrophosphorylase* (*MEL O 05057*), with parallel shifts of amino acids in *Cucurbitaceae* (Figure 2C and Table S19). *α-Galactosidase* is involved in the first step of RFOs catabolism generating sucrose and galactose during phloem unloading (Ohkawa et al., 2010). The RFOs transported to fruits are catabolized in the peduncle before eventually entering the fruits. *UDP-sugar pyrophosphorylase* (*USPase*) catalyzes a reversible transfer of the uridyl group from UTP to sugar-1-phosphate, producing UDP sugar and pyrophosphate, which is involved in RFOs synthesis (Kleczkowski et al., 2011). The initial and final steps of RFOs synthesis and hydrolysis before entry into fruit are shown to have been selected during evolution of *Cucurbitaceae* crops, suggesting their roles in the carbohydrates phloem transport evolution. The biological function of these two selected genes in *Cucurbitaceae* crops is quite worthy of being investigated to reveal their important roles in the phloem transport.

Association Study of Transcriptome and DNA Methylation during Fruit Development

Fruits are essential for human nutrition, providing fundamental sugars, vitamins, and numerous other compounds, such as carotenoids, polyphenols, and fatty acids (Seymour et al., 2013). Fruit development and ripening need various internal and environmental cues, including developmental genes, light, and phytohormones (Matas et al., 2009). Recently, in tomato, epigenetic remodeling is reported to be associated with fruit ripening (Zhong et al., 2013) and DNA methylation can regulate mRNA m^6A methylation in a feedback loop affecting fruit ripening (Zhou et al., 2019). Sugar such as sucrose, fructose, and glucose contribute the sweetness and influence flavor in fruits, in addition to other dynamic and diverse signaling roles in growth and development (Ruan, 2014). In the present study, we compared sucrose, fructose, and glucose accumulations in developing fruit in varieties HS and DHL92. The fructose and glucose accumulations did not differ significantly between HS and DHL92, but sucrose accumulation displayed a distinct difference, whereby HS accumulated less sucrose than DHL92 20 days after pollination (Figure 3A). We then identified the co-differentially expressed and the methylated genes in developing fruits using association analysis of the transcriptome and DNA methylation data (Figures 3B and Table S20). These co-differentially expressed and methylated genes in fruits 20 days after pollination that were associated mainly with metabolic processes involved in fruit quality (Figure S10).

Using homologs from DHL92 via blast analysis, we distinguished 164 genes associated with sugar metabolism (Table S21). We further retrieved the patterns of the 127 genes from the co-differentially expressed and methylated genes data, showing 50 genes of CG type, 40 genes of CHG type, and 37 genes of CHH

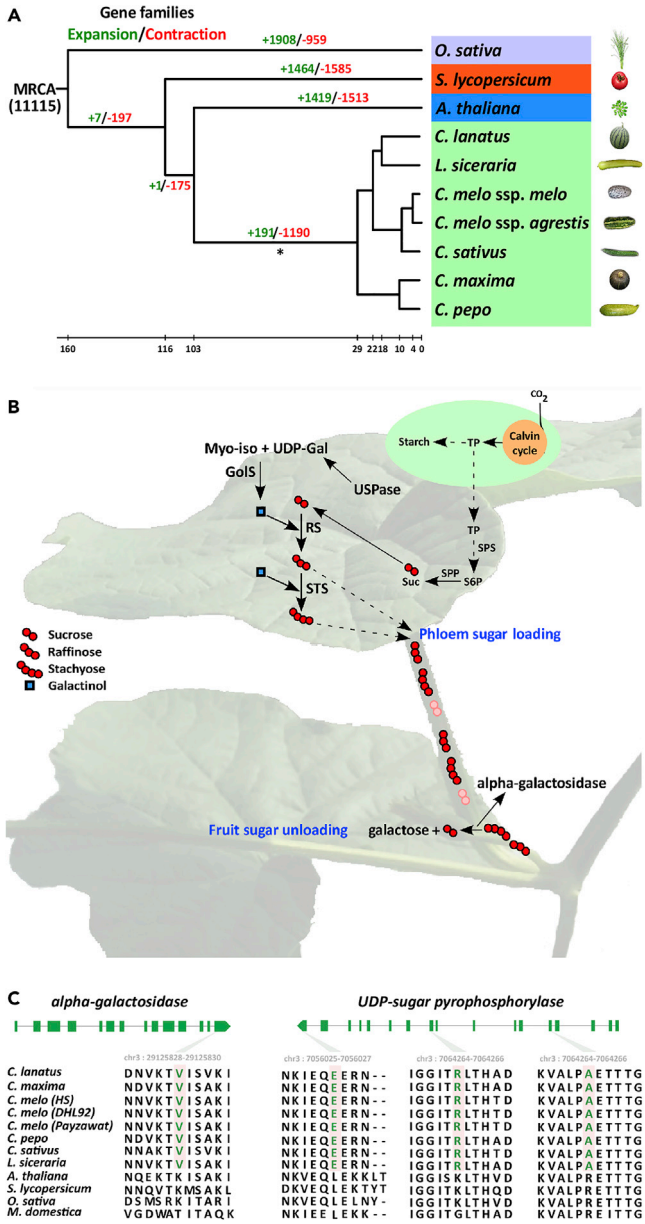


Figure 2. Positive Selection and Evolutionary Analysis of *Cucurbitaceae*

(A) Phylogenetic tree constructed with orthologous genes on phase 1 sites by Bayesian-based method. The star showed gene family expansion and contraction in *Cucurbitaceae*. MRCA, most recent common ancestor.

(B) Oligosaccharides transport (loading and unloading) via phloem in *Cucurbitaceae*.

(C) Amino acid characteristics of specific genes (*α-galactosidase* and *UDP-sugar pyrophosphorylase*) by positive selection analysis in *Cucurbitaceae* compared with *A thaliana*, *S. lycopersicum*, *M. domestica*, and *O. sativa*. USPase, UDP-sugar pyrophosphorylase. The green colored amino acids denote the sites under positive selection.

type (Figures 3C and Tables S22, S23, and S24). These results revealed epigenetic regulation of fruit quality in melon, particularly regarding sugar metabolism, providing us with quality-influencing targets and metabolic processes mediated by epigenetics.

Genetic Architecture of Important Traits in Melon

Fruit quality is of strong interest, and several QTLs have been mapped using genetically segregated or natural populations in melon (Galpaz et al., 2018; Zhao et al., 2019; Zhang et al., 2019). Sugar, particularly

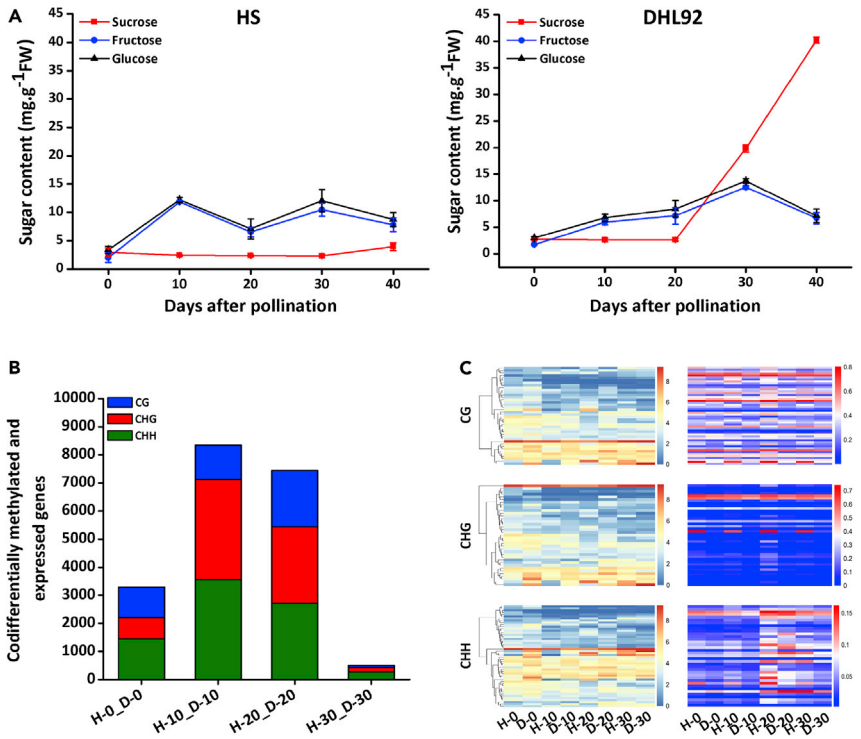


Figure 3. Transcription and DNA Methylation Patterns during Fruit Development of HS (Few-Sucrose Accumulator) and DHL (High-Sucrose Accumulator)

(A) Sugar accumulation in developing fruits after pollination. Data are represented as mean ± SEM.

(B) Co-differentially expressed and methylated genes in developing fruits after pollination (0–30 days).

(C) Co-differentially methylated and expressed genes involved in sugar metabolism of developing fruits after pollination. Left: DNA methylated patterns of genes involved in sugar metabolism; right: transcriptional expression patterns of the same genes as shown on the left. H, HS; D, DHL92.

sucrose, contributes to sweetness of fruit, which is the result of selection and improvement for desirable quality appealing for human consumption. However, causative genes associated with sucrose accumulation have not been identified to date. Using the *C. melo* spp. *agrestis* subspecies variety (HS) that accumulates few sucrose and a high-sucrose accumulator, *C. melo* spp. *melo* subspecies (XH), we constructed and re-sequenced a F_8 recombinant inbred population (RILs) (Figure 4A). After filtering the sequencing reads, we collected 465.64 Gb of clean reads, subsequently mapping these reads to the *de novo* assembly of HS (Table S25). Using 1,350,040 high-quality SNPs called from the resequencing of the RILs population (Table S26), we figured out a physical recombination map of each line in the RILs population (Figure S11). Finally, we retrieved 88 RILs after eliminating questionable lines, defined as one bin of recombined sites in 100 Kb region, and constructed the bin map using 1,110 bins for this RILs population (Figure 4B). We evaluated sucrose, fructose, and glucose accumulation in fruits of the RILs population in both spring and fall (Figures 4C and 4E).

We mapped two QTLs regions for sucrose accumulation in spring (Chr_10 and Chr_7) and fall (Chr_10 and Chr_5) (Figures 4D and 4F and Table S27). Totally, we identified five genes that were associated with sugar metabolism based on annotation (Table S28). Two genes, alkaline/neutral invertase 1 (*CINV1*, *MELO21653*) and sucrose-phosphatase 2 (*SPP2*, *MELO21692*), were located in the QTLs regions in spring (Figure 4D), and other three candidate genes, α-galactosidase (*MELO09061*), β-galactosidase (*MEL O 16006*), and putative sugar phosphate/phosphate translocator (*MELO16058*), were found in the QTLs regions in fall (Figure 4F). The genotype distributions of these two genes (*MELO21653* and *MELO21692*) in the RILs populations indicated their relevance to sucrose accumulation (Figure S12). We further confirmed two non-synonymous SNPs (A472T and F486S) in the *CINV1* (*MELO21653*) in several few-sucrose accumulators and high-sucrose accumulators melon varieties (Figures 4G and S13). In addition, we showed the same non-synonymous SNPs in two sequenced cucumber genomes (Figure S14), in which cucumber is, in parallel,

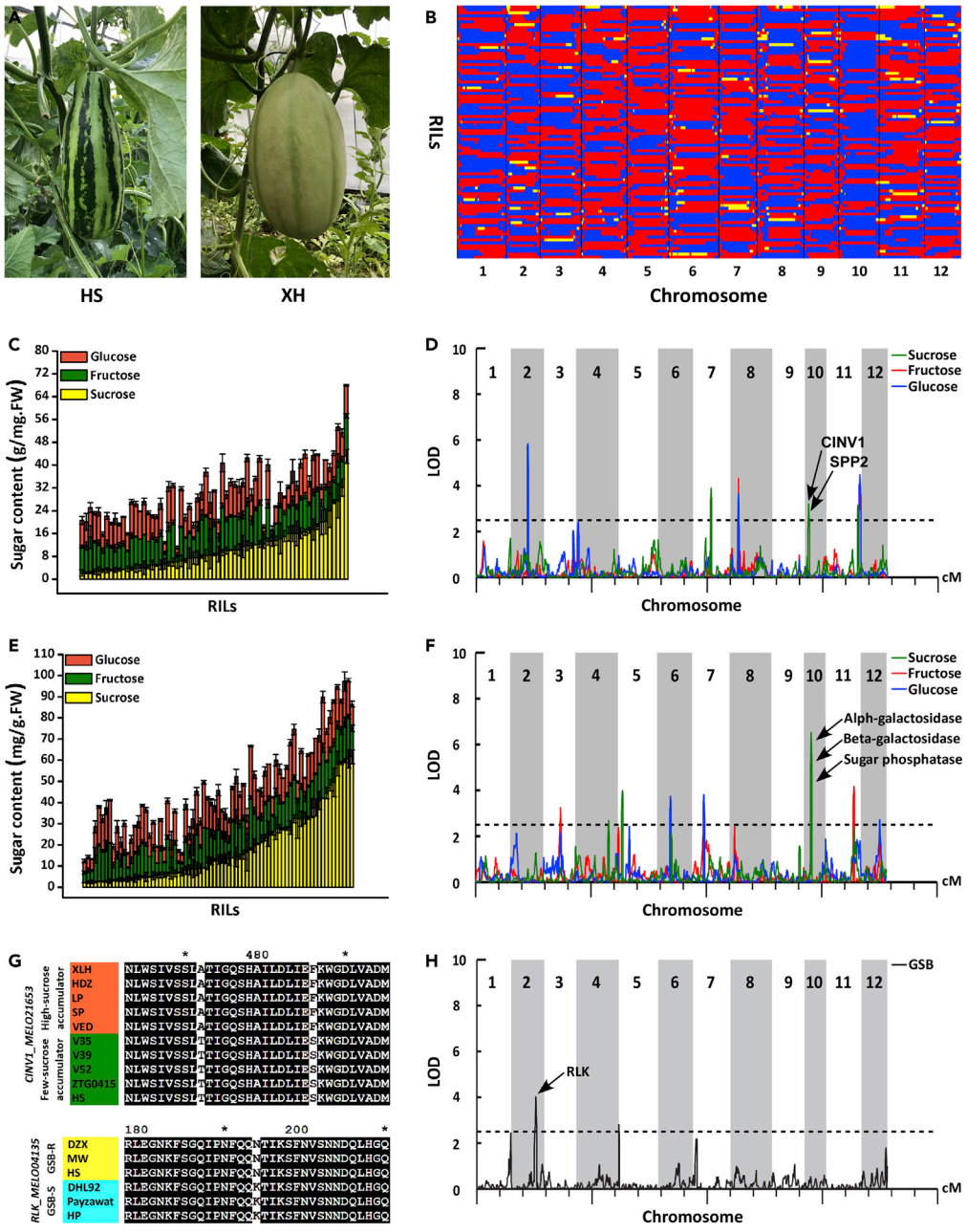


Figure 4. Identification of Genetic Loci Controlling Sucrose Accumulation and Gummy Stem Blight Resistance Using Recombinant Inbred Lines

(A) Fruit phenotypes of HS (*agrestis*-type, few-sucrose accumulator, and GSB resistance) and XH (*melo*-type, high-sucrose accumulator, and GSB susceptibility) used for the recombinant inbred lines construction.

(B) Bin map of the recombinant inbred lines. Blue bands represent markers from the HS parent; red bands show markers from the XH parent; yellow ones denote heterozygous genotype.

(C) Sugar accumulation of the recombinant inbred lines in spring. Data are represented as mean ± SEM.

(D) QTLs analysis of sugar accumulation in spring. CINV, alkaline/neutral invertase; SPP2, sucrose-phosphatase 2.

(E) Sugar accumulation of the recombination inbred lines in fall. Data are represented as mean ± SEM.

(F) QTLs analysis of sugar accumulation in fall.

(G) Genotyping analysis of candidate genes in natural varieties.

(H) QTLs analysis of gummy stem blight resistance. RLK, LRR receptor-like serine/threonine-protein kinase.

few-sucrose accumulator suggesting the *CINV1* might be a major player in sucrose accumulation. In higher plants, the CINVs, located in cytoplasm, are maintaining cytosolic sugar homeostasis for the cellular function (Wan et al., 2018), in addition to regulating reproductive and root development (Barratt et al., 2009). The CINVs with other two families, acid invertase (cell wall invertase, CWIN, and vacuole invertase, VIN), specifically hydrolyze sucrose into glucose and fructose. The sucrose-phosphatase (SPP) catalyzes the final step in the sucrose biosynthesis to convert sucrose-6-phosphate to sucrose, in which SPP can be involved in sucrose phosphate synthesis (SPS) to form SPS-SPP complex to control sucrose biosynthesis (Lunn and Macrae, 2003). The significant different expressions of these candidate genes between fruits of few-sucrose and high-sucrose accumulation varieties (HS and XH) indicate their association with sucrose accumulation (Figure S15). Of these five candidate genes associated with sucrose accumulation, we found that expressions of four genes (*MEL O 16006*, *MELO16058*, *MELO21692* and *MELO21653*) were regulated by DNA methylation (Table S29). These results indicated an epigenetics mechanism of fruit quality formation in melon fruit.

The *C. melo* spp. *agrestis* subspecies variety (HS) used for RILs population construction has gummy stem blight resistance (Hu et al., 2018). Using this RILs population, we mapped two QTLs (Chr_2 and Chr_5) for GSB resistance in melon (Figure 4H). We identified a typical R gene, LRR receptor-like serine/threonine-protein kinase (RLK, *MELO04135*), was associated with GSB resistance (Figure 4H and Table S30). A non-synonymous SNP (N194K) was identified in CDS of the RLK gene between GSB-resistant and -susceptible varieties, and we further confirmed the genotypes of the SNP in some GSB-resistant and -susceptible varieties (Figure 4G and Figure S16). The LRR receptor-like serine/threonine-protein kinase (LRR-RLKs), as transmembrane cell surface receptors, can recognize microbe-associated molecular patterns (MAMPs) and are required for MAMP-triggered immunity (Couto and Zipfel, 2016). A number of RLKs were identified to regulate plant innate immunity (Wu et al., 2016). Furthermore, the LRR-RLK genes commonly play an essential role in innate immunity to hemibiotrophic and biotrophic pathogens (Roux et al., 2011). The GSB-associated *Didymella bryoniae* is a hemibiotrophic pathogen. The RLK mapped in the present study might be a functional gene of GSB resistance based on the genotyping and expression analysis in melon.

Recently, evolutionary genomic analyses indicated that *C. melo* spp. *agrestis* subspecies was independently domesticated compared with *C. melo* spp. *melo* subspecies in melon (Zhao et al., 2019). In the present study, the release of the chromosome-scale genome of *agrestis* subspecies is closing the gap of two melon subspecies genomes, enabling us to use these genomes in the genetics and evolutionary and gene mapping studies for academic and industrial communities. In melon, *agrestis* subspecies possess many desirable traits for breeding improvement, particularly regarding disease resistance, fruit flavor, and environmental adaptation. Inter-subspecies hybridization is routinely employed to improve various agronomic traits in creating novel resistant and wide adaptable elite cultivars. The selected genes identified from *Cucurbitaceae* allow us to further explore their conserved function and reconsider the role of phloem loading in plants. Moreover, the sucrose accumulation-associated genes provide us targets for seeking natural variations in sugar accumulation and opportunities for fruit quality enhancement using genetic techniques. In addition, the SNP identified in the RLK gene associated with GSB resistance will be developed to design a reproducible marker, facilitating marker-assisted selection for GSB resistance in melon. This study provides indispensable genomic resources accelerating of functional genomics studies and the genomics-informed breeding pipeline for improving the fruit quality and disease resistance traits in melons.

Limitations of the Study

We reported a high-quality assembly of *C. melo* ssp. *agrestis* providing comparative genomic analysis and functional genes mining in melon species. We identified candidate genes associated with oligosaccharide transport in the phloem, sucrose accumulation, and GSB-resistance in *Cucurbitaceae*. These candidate genes involving in specific phloem transport of oligosaccharides, sucrose accumulation, and gummy stem blight resistance need be further investigated using gene editing tool and facilitate their applications in genomic breeding pipelines of melon crops.

Resource Availability

Materials Availability

All materials should be requested from M.Z (mfzhang@zju.edu.cn).

Data and Code Availability

All data, including genome sequencing raw data, genome assembly and annotation data, RNA-seq, DNA methylation, and re-sequencing data of RILs generated in the present study have been deposited in CNSA (https://db.cngb.org/cnsa/) under accession number CNP0000863 and in NCBI (https://www.ncbi.nlm.nih.gov/geo/) under accession number PRJNA648029.

METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101422.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (Grant 2018YFD1000800) and China Agriculture Research System (CARS-25-17).

AUTHOR CONTRIBUTIONS

J.Yang, M.Zhang and J.Yu conceived and designed the study. J.Lian, Y.Niu and J.Yang performed the genome sequencing, assembly and bioinformatics analyses. G.Deng, J.Garraway and Z.Hu performed phenotyping, genotyping and contributed to data analyses. J.Yang wrote the manuscript. M.Zhang and J.Yu discussed and revised the manuscript.

DECLARATION OF INTERESTS

All authors declared no interests.

Received: February 22, 2020
Revised: June 5, 2020
Accepted: July 26, 2020
Published: August 21, 2020

REFERENCES

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 targeted SNP selection for an improved anchoring of the melon (*Cucumis melo* L.) scaffold genome assembly. BMC Genomics *16*, 4.

Barratt, D.H.P., Derbyshire, P., Findlay, K., Pike, M., Wellner, N., Lunn, J., Feil, R., Simpson, C., Maule, A.J., and Smith, A.M. (2009). Normal growth of Arabidopsis requires cytosolic invertase but not sucrose synthase. Proc. Natl. Acad. Sci. U S A *106*, 13124–13129.

Castanera, R., Ruggieri, V., Pujol, M., Garcia-Mas, J., and Casacuberta, J.M. (2020). An improved melon reference genome with single-molecule sequencing uncovers a recent burst of transposable elements with potential impact on genes. Front. Plant Sci. *10*, 1815.

Couto, D., and Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. Nat. Rev. Immunol. *16*, 537–552.

Galpaz, N., Gonda, I., Shem-Tov, D., Barad, O., Tzuri, G., Lev, S., Fei, Z.J., Xu, Y.M., Mao, L.Y., Jiao, C., 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., Camara, F., Cozzuto, L., Lowy, E., et al. (2012). The genome of melon (*Cucumis melo* L.). Proc. Natl. Acad. Sci. U S A *109*, 11872–11877.

Guo, S.G., Zhao, S.J., Sun, H.H., Wang, X., Wu, S., Lin, T., Ren, Y., Gao, L., Deng, Y., Zhang, J., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. Nat. Genet. *51*, 1616–1623.

Hu, Z.Y., Deng, G.C., Mou, H.P., Xu, Y.H., Chen, L., Yang, J.H., and Zhang, M.F. (2018). A re-sequencing-based ultra-dense genetic map reveals a gummy stem blight resistance-associated gene in *Cucumis melo*. DNA Res. *25*, 1–10.

Jeffrey, C. (1980). A review of the Cucurbitaceae. Bot. J. Linn. Soc. *81*, 233–247.

Kleczkowski, L.A., Decker, D., and Wilczynska, M. (2011). UDP-Sugar pyrophosphorylase: a new old mechanism for sugar activation. Plant Physiol. *156*, 3–10.

Lunn, J.E., and Macrae, E. (2003). New complexities in the synthesis of sucrose. Curr. Opin. Plant Biol. *6*, 208–214.

Matas, A.J., Gapper, N.E., Chung, M.Y., Giovannoni, J.J., and Rose, J.K.C. (2009). Biology and genetic engineering of fruit maturation for enhanced quality and shelf-life. Curr. Opin. Biotechnol. *20*, 197–203.

Ohkawa, W., Kanayama, Y., Daibo, N., Sato, T., Nishiyama, M., and Kanahama, K. (2010). Metabolic process of the C^{-14}-sugars on the translocation pathways of cucumber plants. Scientia Horticulturae *124*, 46–50.

Pitrat, M. (2008). Melon (*Cucumis melo* L.). In Handbook of Crop Breeding Vol I: Vegetables, J. Prohens and F. Nuez, eds. (Springer), pp. 282–315.

Ren, Y., Guo, S.G., Zhang, J., He, H.J., Sun, H.H., Tian, S.W., Gong, G.Y., Zhang, H.Y., Levi, A., Tadmor, Y., et al. (2018). A tonoplast sugar transporter underlies a sugar accumulation QTL in watermelon. Plant Physiol. *176*, 836–850.

Rennie, E.A., and Turgeon, R. (2009). A comprehensive picture of phloem loading strategies. Proc. Natl. Acad. Sci. U S A *106*, 14162–14167.

Roux, M., Schwessinger, B., Albrecht, C., Chinchilla, D., Jones, A., Holton, N., Malinovsky, F.G., Tor, M., De Vries, S., and Zipfel, C. (2011). The Arabidopsis Leucine-rich repeat receptor-

like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. Plant Cell *23*, 2440–2455.

Ruan, Y.L. (2014). Sucrose metabolism: gateway to diverse carbon use and sugar signaling. Annu. Rev. Plant Biol. *65*, 33–67.

Sebastian, P., Schaefer, H., Telford, I.R.H., and Renner, S.S. (2010). Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. Proc. Natl. Acad. Sci. U S A *107*, 14269–14273.

Seymour, G.B., Ostergaard, L., Chapman, N.H., Knapp, S., and Martin, C. (2013). Fruit development and ripening. Annu. Rev. Plant Biol. *64*, 219–241.

Turgeon, R. (2010). The role of phloem loading reconsidered. Plant Physiol. *152*, 1817–1823.

Turgeon, R., and Wolf, S. (2009). Phloem transport: cellular pathways and molecular trafficking. Annu. Rev. Plant Biol. *60*, 207–221.

Wan, H.J., Wu, L.M., Yang, Y.J., Zhou, G.Z., and Ruan, Y.L. (2018). Evolution of sucrose metabolism: the dichotomy of invertases and beyond. Trends Plant Sci. *23*, 163–177.

Wang, S.W., Yang, J.H., and Zhang, M.F. (2011). Developments of functional markers for Fom-2-mediated fusarium wilt resistance based on single nucleotide polymorphism in melon (*Cucumis melo* L.). Mol. Breed. *27*, 385–393.

Wu, Y.Z., Xun, Q.Q., Guo, Y., Zhang, J.H., Cheng, K.L., Shi, T., He, K., Hou, S.W., Gou, X.P., and Li, J. (2016). Genome-wide expression pattern analyses of the Arabidopsis leucine-rich repeat receptor-like kinases. Mol. Plant *9*, 289–300.

Zhang, C.K., and Turgeon, R. (2009). Downregulating the sucrose transporter VpSUT1 in *Verbascum phoeniceum* does not inhibit phloem loading. Proc. Natl. Acad. Sci. U S A *106*, 18849–18854.

Zhang, H., Li, X., Yu, H., Zhang, Y., Li, M., Wang, H., Wang, D., Fu, Q., Liu, M., Ji, C., et al. (2019). A high-quality melon genome assembly provides insights into genetic basis of fruit trait improvement. iScience *22*, 16–27.

Zhang, M.F., and Li, Z.L. (2005). A comparison of sugar-accumulating patterns and relative compositions in developing fruits of two oriental melon varieties as determined by HPLC. Food Chem. *90*, 785–790.

Zhao, G.W., Lian, Q., Zhang, Z.H., Fu, Q.S., He, Y.H., Ma, S., Ruggieri, V., Monforte, A.J., Wang, P.Y., Julca, I., et al. (2019). A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. Nat. Genet. *51*, 1607–1615.

Zhong, S.L., Fei, Z.J., Chen, Y.R., Zheng, Y., Huang, M.Y., Vrebalov, J., Mcquinn, R., Gapper, N., Liu, B., Xiang, J., et al. (2013). Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat. Biotechnol. *31*, 154–159.

Zhou, L.L., Tian, S.P., and Qin, G.Z. (2019). RNA methylomes reveal the m^6A-mediated regulation of DNA demethylase gene SlDML2 in tomato fruit ripening. Genome Biol. *20*, 156.

iScience, Volume *23*

Supplemental Information

The Chromosome-Scale Genome of Melon Dissects Genetic Architecture of Important Agronomic Traits

Jinghua Yang, Guancong Deng, Jinmin Lian, Jenella Garraway, Yongchao Niu, Zhongyuan Hu, Jingquan Yu, and Mingfang Zhang

Supplemental Information

The Chromosome-scale Genome of Melon Dissects Genetic Architecture of Important Agronomic Traits

Jinghua Yang, Guancong Deng, Jinmin Lian, Jennella Garraway, Yongchao Niu, Zhongyuan Hu, Jingquan Yu, Mingfang Zhang

Supplementary Figures

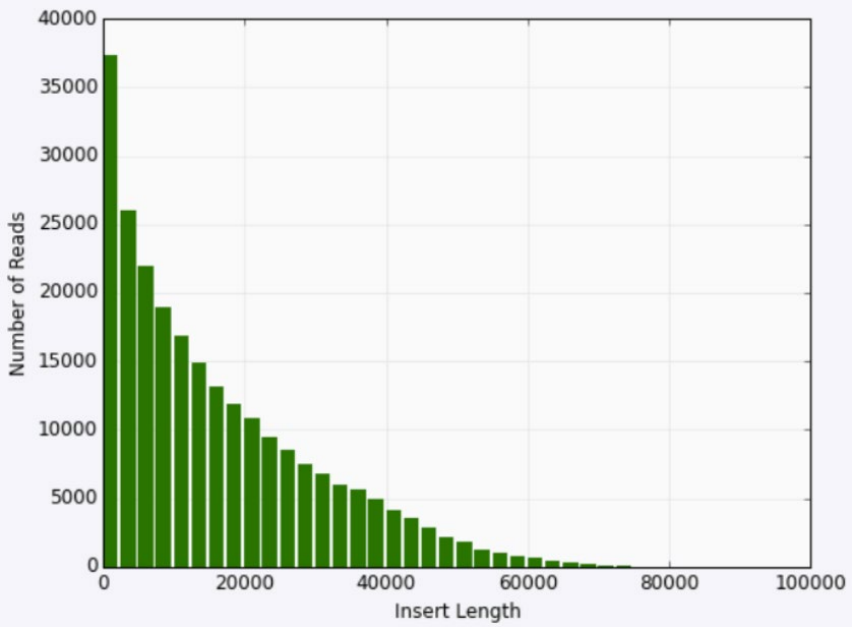


Figure S1. Distribution of estimated insert length of PacBio Flow Cell. Related to Figure 1.

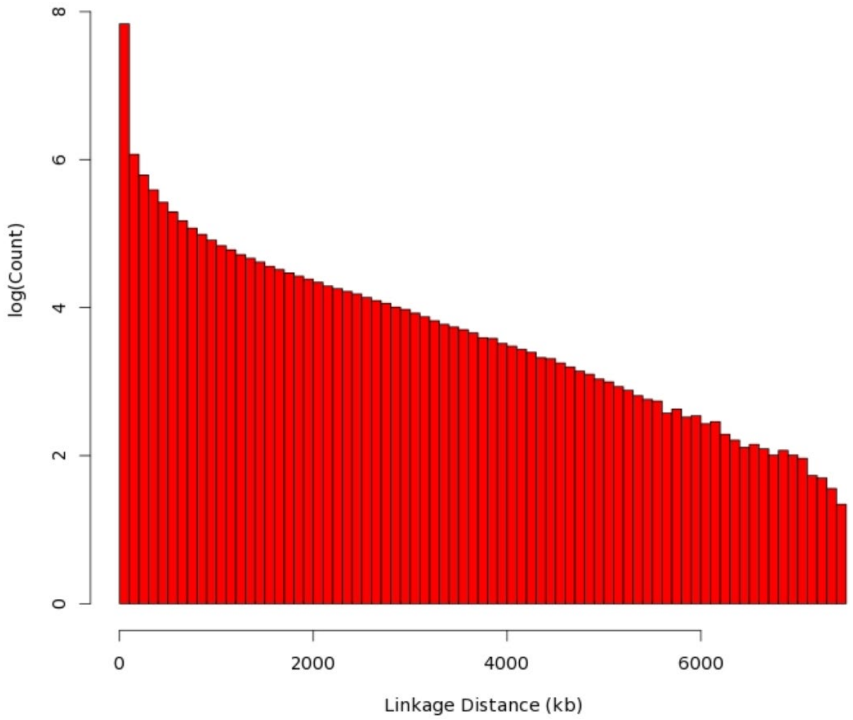


Figure S2. Intra-contig Hi-C linkage distance histogram. Related to Figure 1.

Hi-C linkages follow a power law as expected, and linkages are observed out to near 8 Mbp, the size of the largest contig in the alternate assembly.

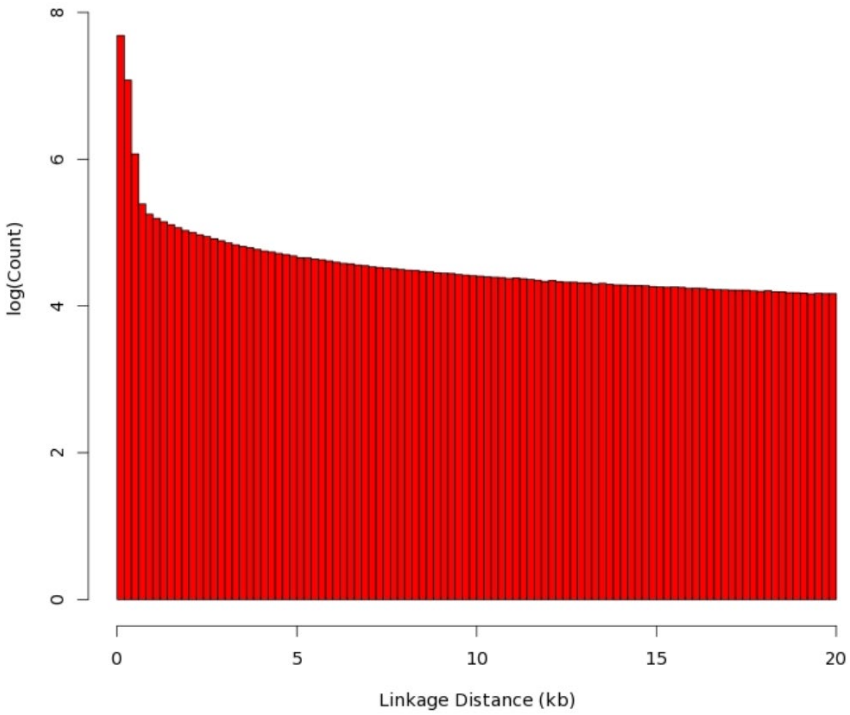


Figure S3. Intra-contig Hi-C linkage histogram, zoomed in to first 20kbp. Related to Figure 1.

The zoomed in figure is useful to check that the spike in the shortest spanned distances is not too large (<2 orders of magnitude is ideal) and that the power law pattern is smooth.

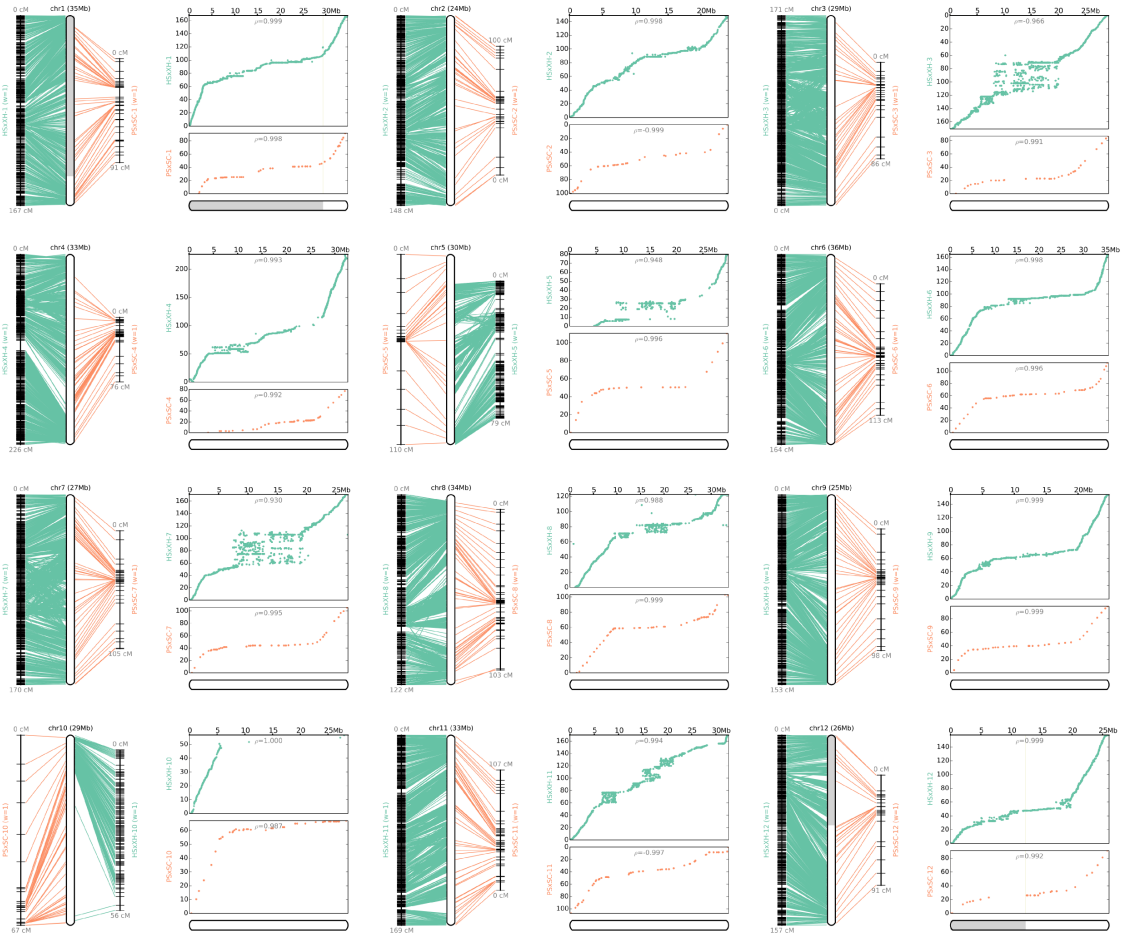


Figure S4. The 12 pseudo-chromosomes of melon genome, reconstructed from HS x XH and PS x SC genetic maps. Related to Figure 1.

Left panel of each picture: CMAP-style presentation with lines connecting the physical positions on the reconstructed chromosome and the map positions.

Right panel of each picture: Scatter plots, with dots representing the physical position on the chromosome (*x*-axis) versus the map location (*y*-axis).

Adjacent scaffolds within the reconstructed chromosome are shown as boxes with alternating shades, marking the boundaries of the component scaffolds. The ρ-value on each scatter plot measures the Pearson correlation coefficient, with values in the range of -1 to 1 (values closer to -1 and 1 indicate near-perfect collinearity).

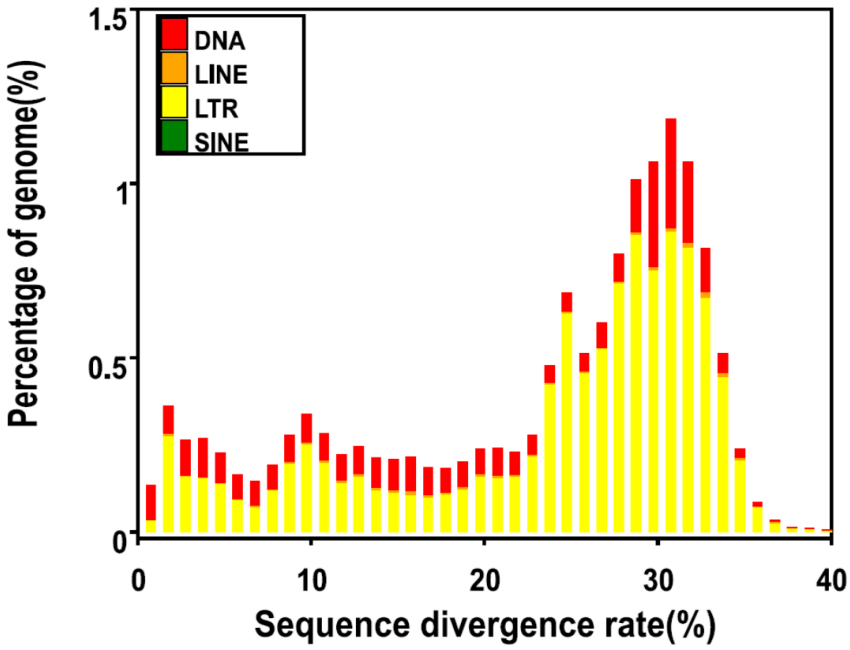


Figure S5. Distribution of divergence rate of each type of TE. Related to Figure 1.

The divergence rate was calculated between the identified TE elements in the genome by homology-based method and the consensus sequence in the Repbase.

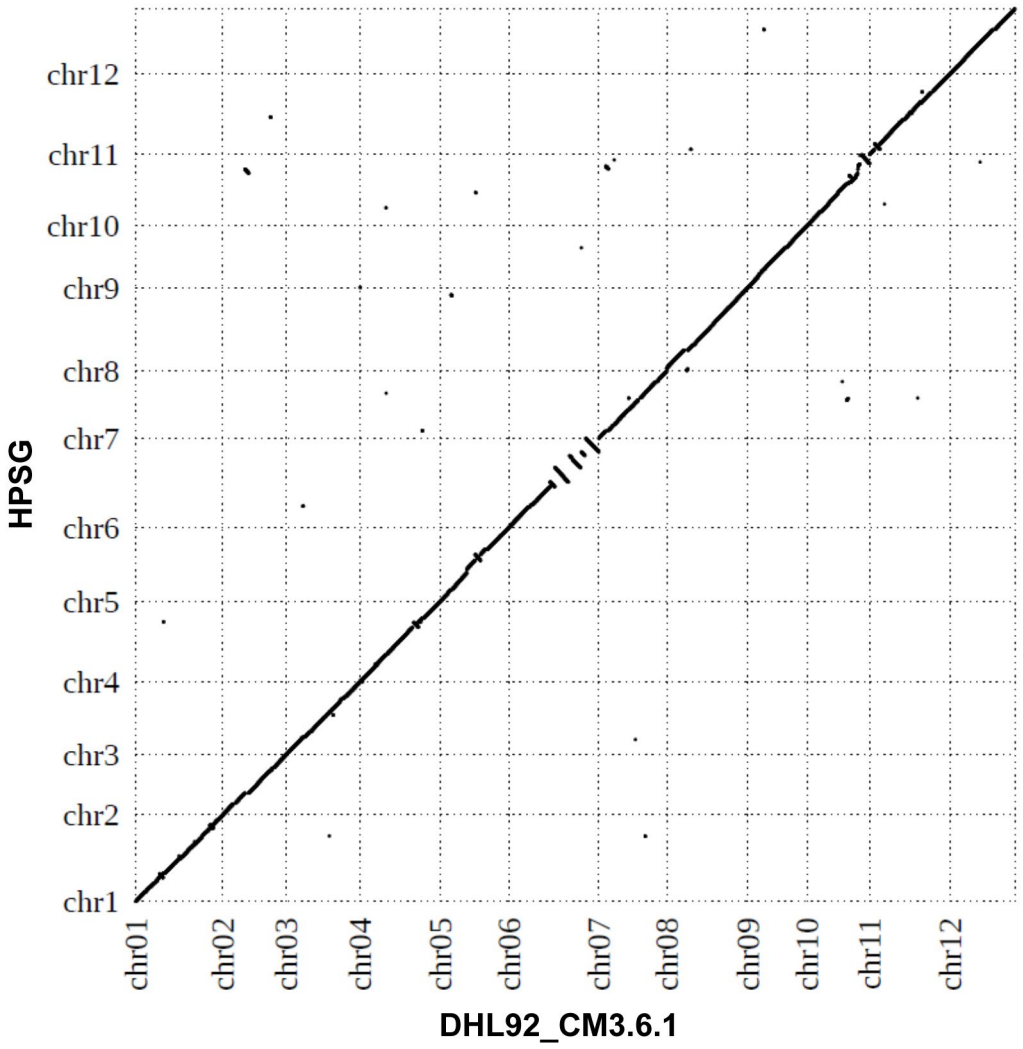


Figure S6. The synteny between DHL9.2_3.6.1 and HS assembly. Related to Figure 1.

HPSG assembly in this study was aligned to the DHL9.2_3.6.1 genome with MUMmer (version 3.23) with default parameters and the genomic alignment results were extracted with the delta-filter -i 95 -l 10000 -u 50 -q -r parameters.

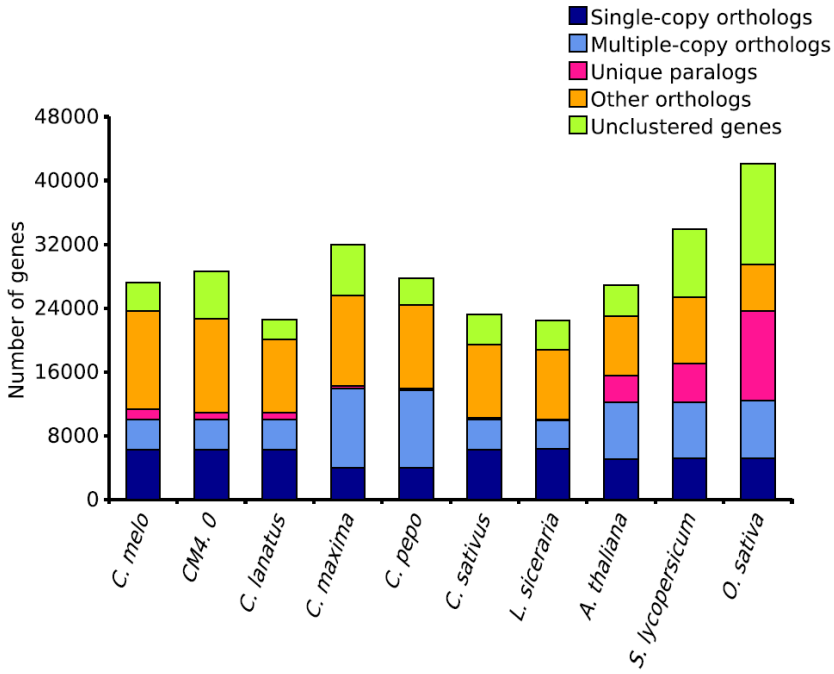


Figure S7. Protein orthology comparison among genomes of *Cucumis melo (HS), Cucumis melo (CM4.0), Cucurbita pepo, Cucumis sativus, Citrullus lanatus, Cucurbita maxima, Lagenaria Siceraria, Solanum lycopersicum, Arabidopsis thaliana and Oryza sativa*. Related to Figure 2.

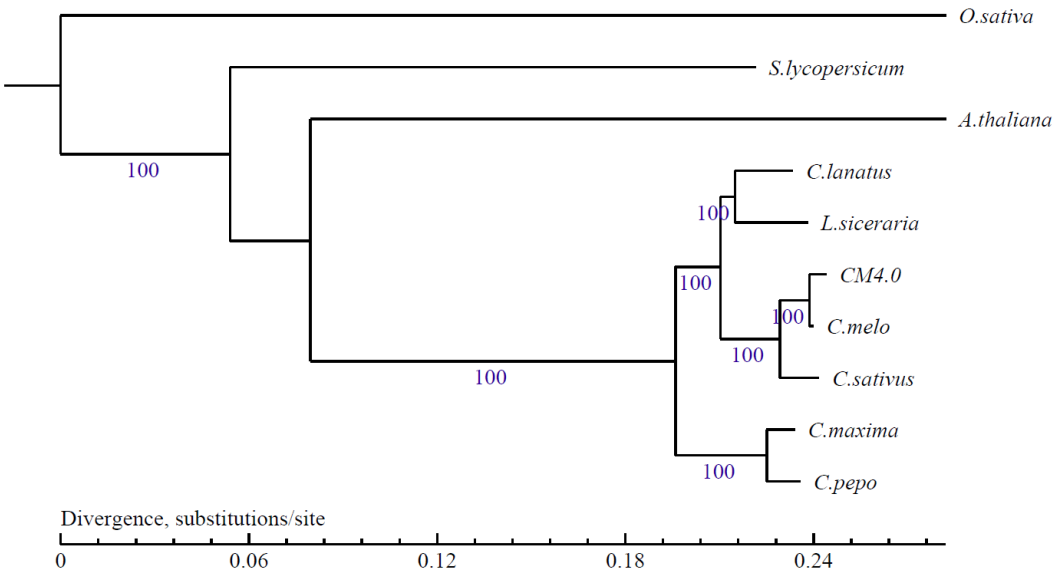


Figure S8. Phylogenetic tree constructed with orthologous genes on phase 1 sites by Bayesian-based method. The branch length represents the neutral divergence rate. Related to Figure 2.

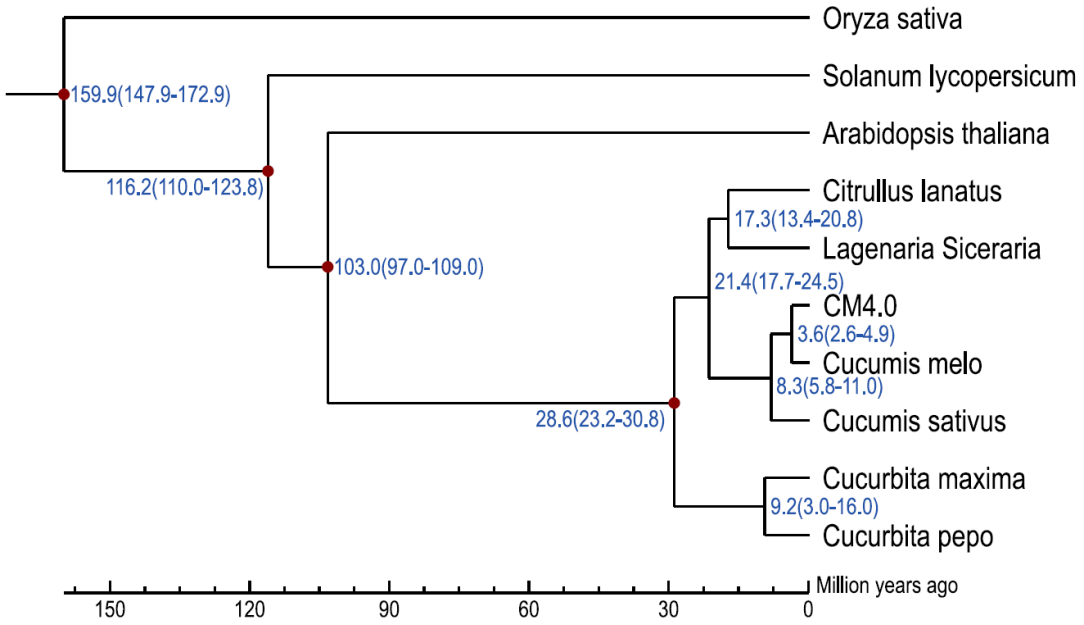


Figure S9. Phylogenetic tree and divergence time. The blue numbers on the nodes are the divergence time from present (million years ago, Mya). Related to Figure 2.

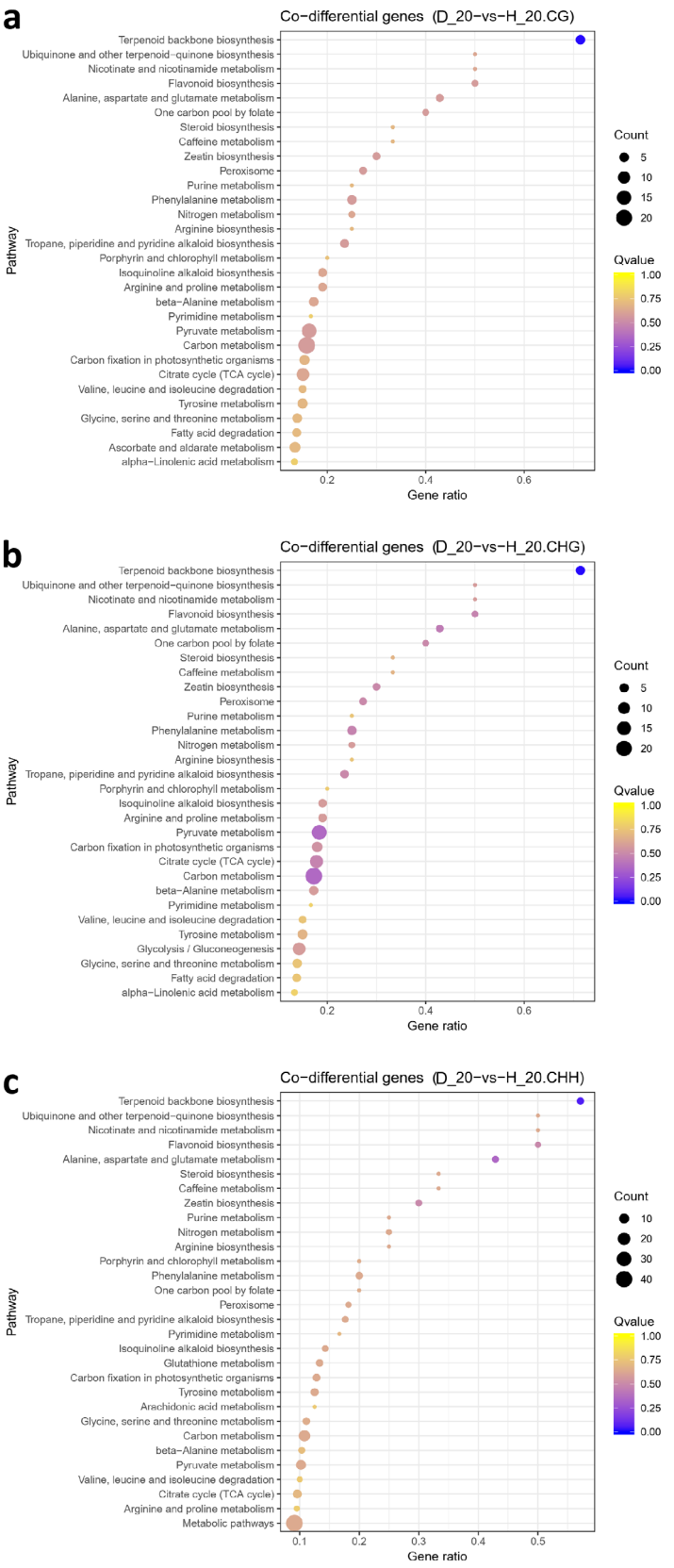


Figure S10. KEGG enrichment annotation of co-differential expressed and methylated genes in 20 days melon fruit after pollination. Related to Figure 3.

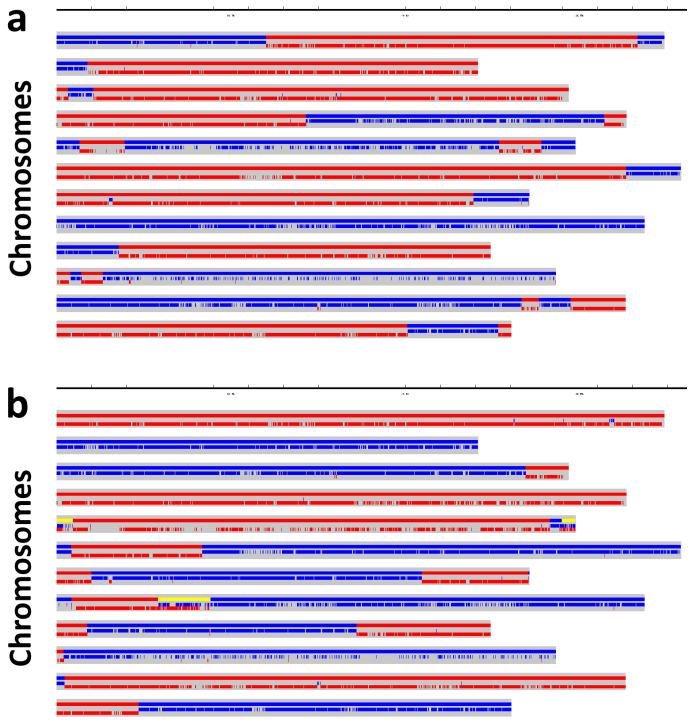


Figure S11. Physical recombination map in representative R9 and R97 lines of the RILs population. Related to Figure 4.

a, R9; b, R97. Each chromosome (ordinate) consisted three lines, in which the first line is deduced genotypes, the second and third lines are parents' genotypes with blue and red colors, and yellow one means heterozygous genotype.

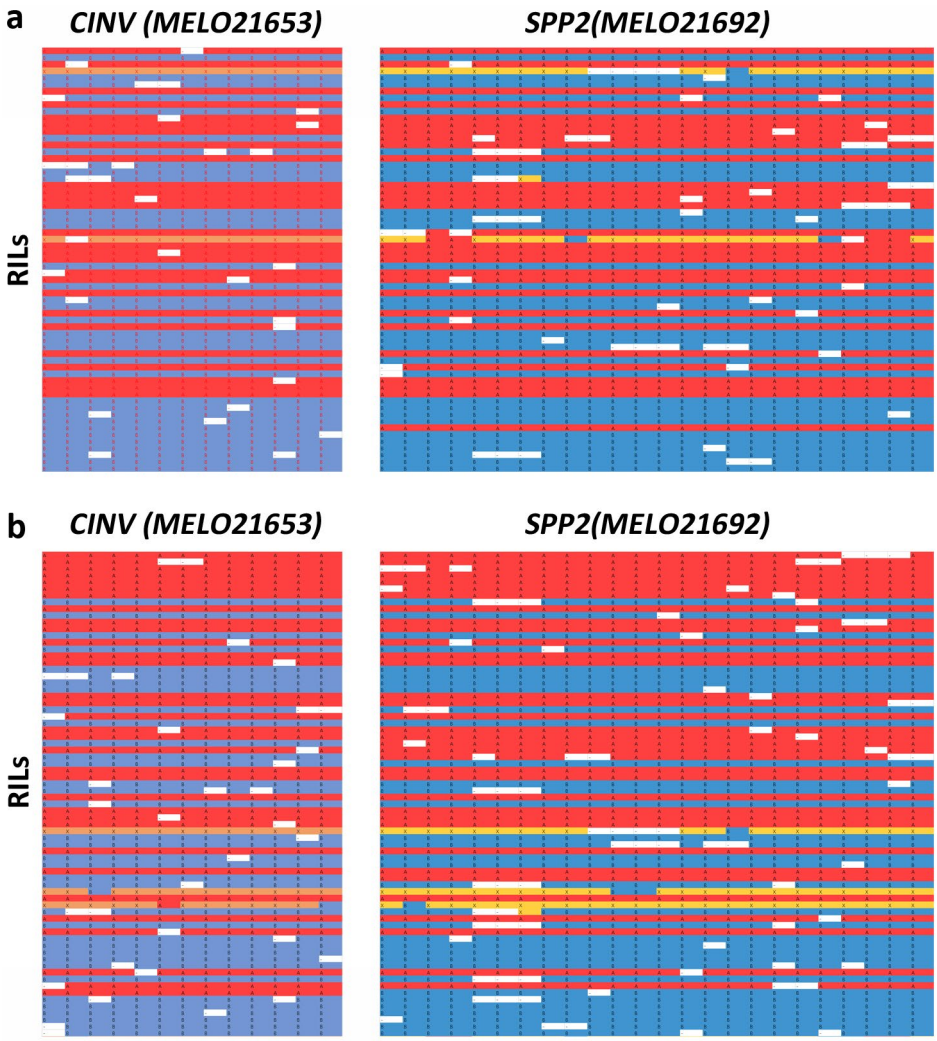


Figure S12. SNPs genotyping analysis of representative genes in RILs. Related to Figure 4.

The Y-axis represented RILs lines from high- to few-sucrose accumulation, and the X-axis showed SNPs of the genes. a, Spring; b, Fall.

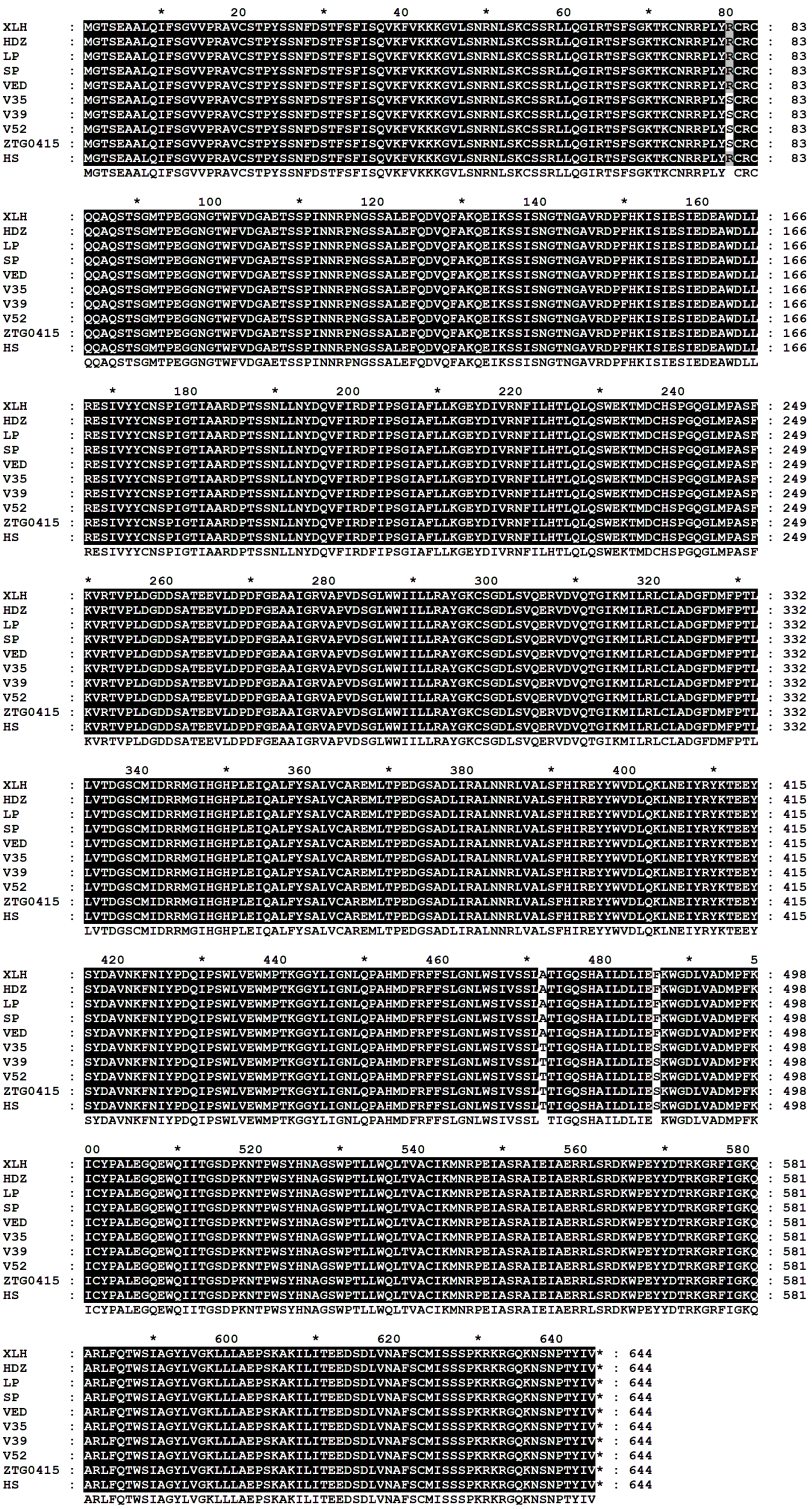


Figure S13. Alignment of *CINV1* (*MELO21653*) amino acids in few-sucrose accumulator (V35, V39, V52, ZTG0415 and HS) and high-sucrose accumulator (XLH, HDZ, LP, SP and VED) melon varieties. Related to Figure 4.

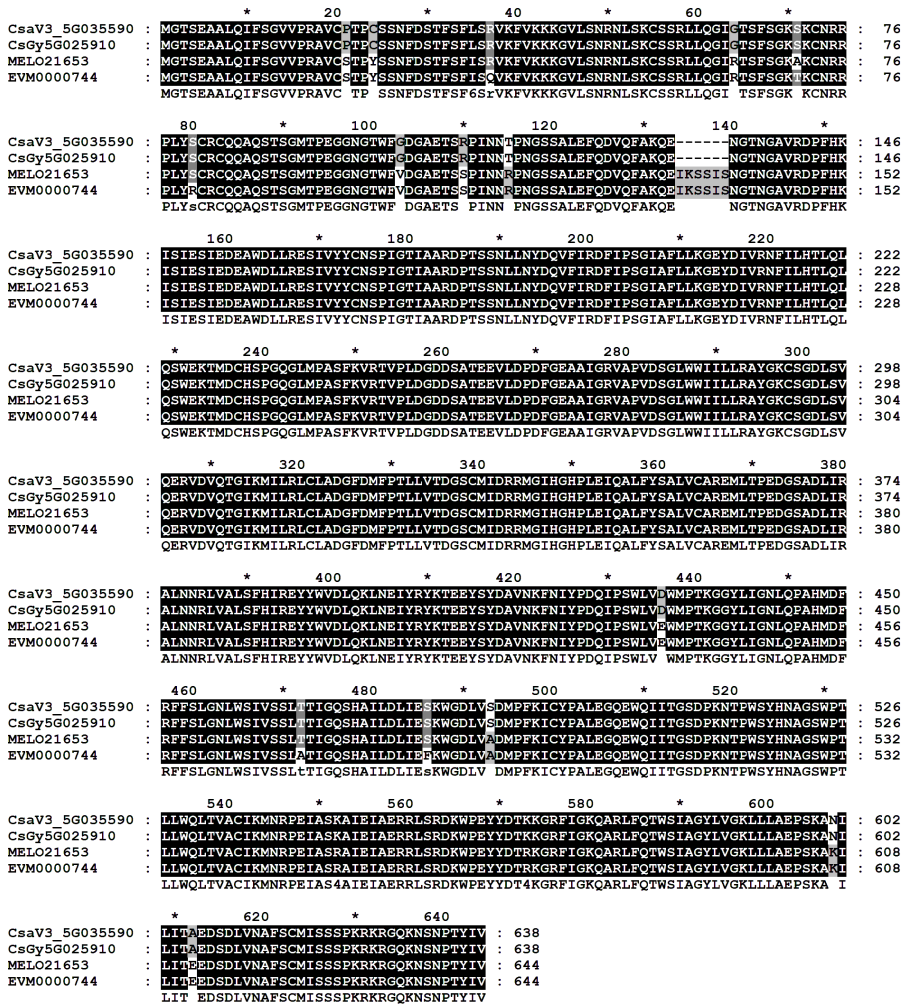


Figure S14. Alignment of *CINV1* (*MELO21653*) amino acids in cucumber (CsaV3_5G035590 and CsGy5G025910) and few-sucrose accumulator (MELO21653) and high-sucrose accumulator (EVM0000744) melon varieties. Related to Figure 4.

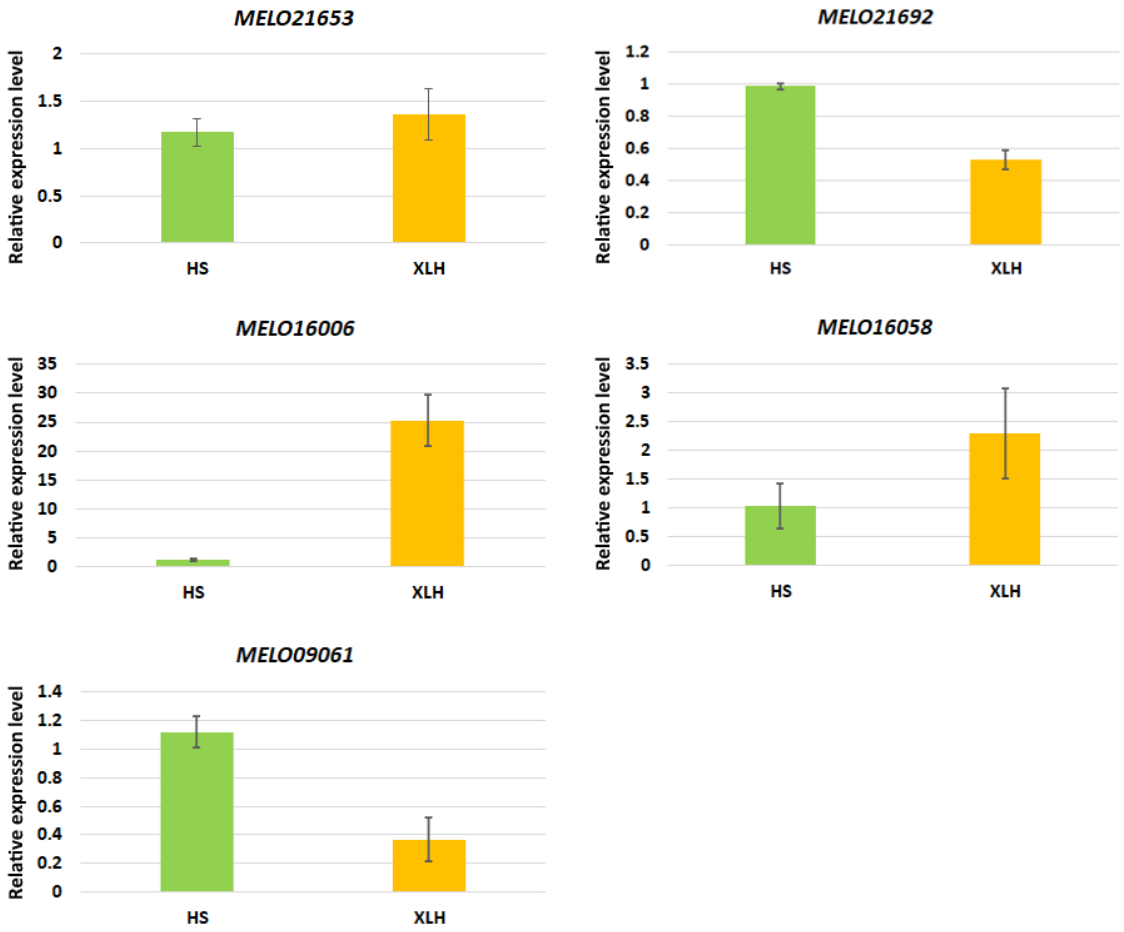


Figure S15. Expressions of mapped candidate genes involving sucrose accumulation between fruits of few-sucrose and high-sucrose accumulation varieties (HS and XLH). Data are represented as mean ± SEM. Related to Figure 4.

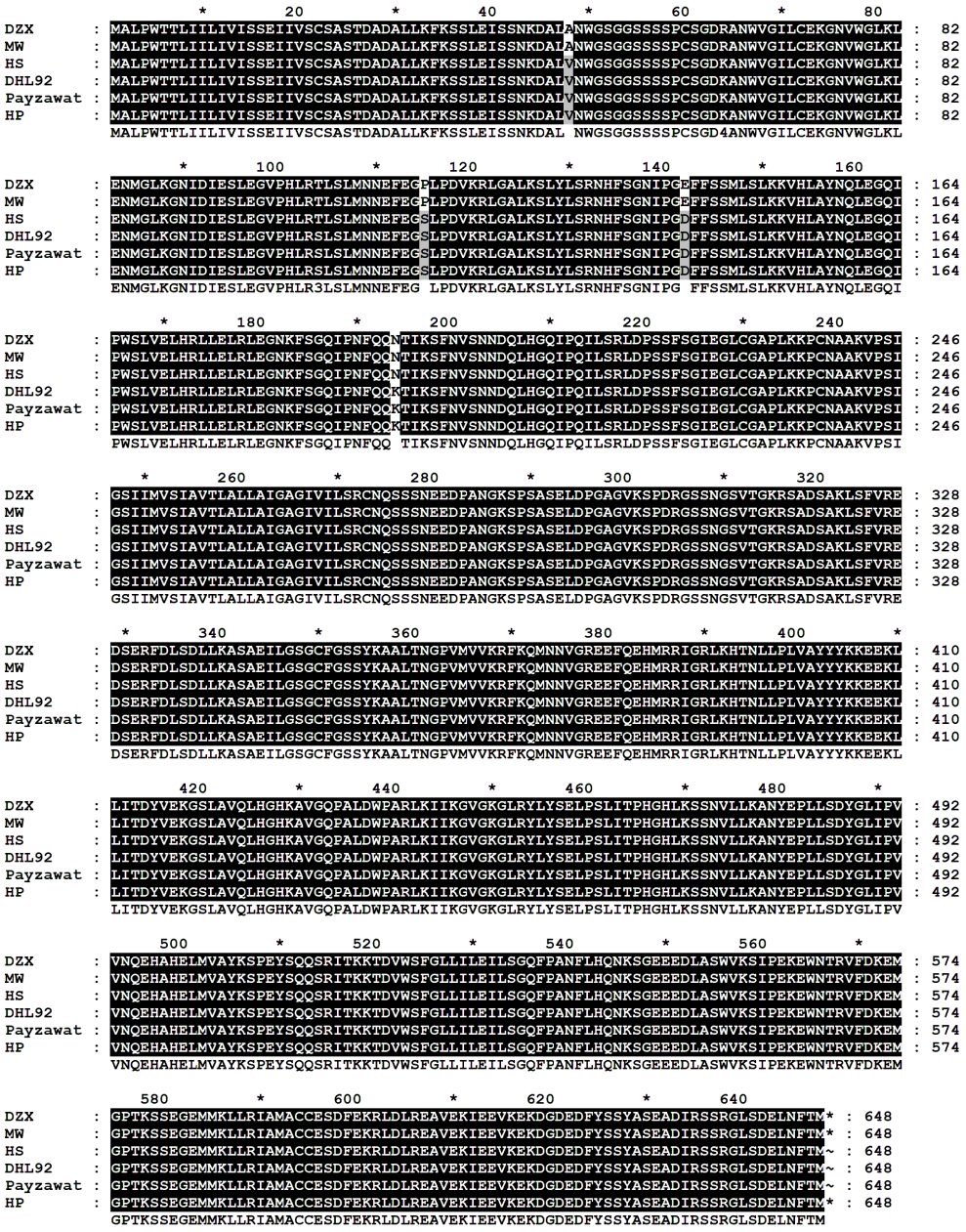


Figure S16. Alignment of *RLK* (*MELO04135*) amino acids in GSB resistance (DZX, MW and HS) and GSB susceptibility (DHL92, Payzawat and HP) melon varieties. Related to Figure 4.

Supplementary Tables

Table S1. Summary of PacBio clean subreads from HS. Related to Figure 1.

Library ID	Reads number	Total Reads Bases	Mean Reads Length	Max Reads Length
m54160_180612_093057	82420	797317983	9674	77680
m54269_180614_171529	201292	1883915643	9359	70625
m54050R1_180317_001655	335799	4361776490	12989	83403
m54050R1_180317_102228	233921	3625672419	15500	89634
m54050R1_180317_203211	252533	3928798436	15558	82516
m54050R1_180318_064146	264928	3970142421	14986	80390
m54050R1_180313_201421	286885	3894105847	13574	78880
m54050R1_180327_000730	612750	6577709279	10735	108190
m54050R1_180327_101313	615783	6601409443	10720	79327
Total	2886311	35640847961	12348	108190

Table S2. Summary of Illumina short reads for polishing from HS. Related to Figure 1.

	Reads number	Reads length (bp)	Total reads bases (bp)	Sequencing coverage
HS	108,035,006	151	16,313,285,906	45

Table S3. Summary of Hi-C clean reads from HS. Related to Figure 1.

	Reads number	Reads length (bp)	Total reads bases (bp)	Sequencing coverage
HS	310,797,204	101	31,390,517,604	86

Table S4. Statistic of HSxXH population genetic markers mapped to HS assembly. Related to Figure 1.

Contig_ID	Chr_ID	HS x XH Linkage groups											
		1	2	3	4	5	6	7	8	9	10	11	12
Contig_8	Chr_1	963	1	3	17	0	1	3	0	1	0	2	0
Contig_12	Chr_2	1	1020	0	0	0	0	1	1	0	0	1	1
Contig_4	Chr_3	2	0	1550	0	2	2	2	0	1	0	0	0
Contig_7	Chr_4	0	1	0	1048	1	1	1	6	0	0	1	0
Contig_3	Chr_5	0	0	0	0	626	0	0	4	1	1	2	1
Contig_2	Chr_6	0	1	2	0	0	1020	0	2	0	0	0	1
Contig_6	Chr_7	0	2	1	0	1	0	1392	3	1	0	2	1
Contig_5	Chr_8	1	5	1	0	1	1	0	1441	2	0	3	1
Contig_10	Chr_9	0	4	0	0	0	2	1	1	942	0	0	1
Contig_1	Chr_10	1	1	2	2	1	1	0	0	0	168	1	0
Contig_0	Chr_11	1	4	2	0	0	1	0	3	1	0	1499	0
Contig_9	Chr_12	0	1	0	0	0	0	0	1	1	0	9	1014

Table S5. Statistic of PSxSC population genetic markers mapped to HS assembly. Related to Figure 1.

Contig_ID	Chr_ID	PS x SC Linkage groups											
		1	2	3	4	5	6	7	8	9	10	11	12
Contig_8	Chr_1	51	0	0	1	0	0	3	0	0	0	0	0
Contig_12	Chr_2	0	38	0	0	0	0	0	0	0	0	0	0
Contig_4	Chr_3	0	0	35	0	0	0	0	0	0	0	0	0
Contig_7	Chr_4	0	0	0	53	0	0	0	0	0	0	0	0
Contig_3	Chr_5	0	0	0	0	28	0	0	0	0	0	0	0
Contig_2	Chr_6	0	0	0	0	0	54	0	0	0	0	0	0
Contig_6	Chr_7	0	0	0	2	0	0	50	0	0	0	0	0
Contig_5	Chr_8	0	0	1	0	0	0	0	55	0	0	0	0
Contig_10	Chr_9	0	0	0	0	0	0	0	0	37	0	0	0
Contig_1	Chr_10	0	0	0	0	0	0	0	0	0	51	0	0
Contig_0	Chr_11	0	1	0	0	0	0	0	0	0	0	55	0
Contig_9	Chr_12	2	0	0	0	0	0	0	0	0	0	0	28

Table S6. Information of 4 breakpoints by ALLMAPS software. Related to Figure 1.

Scaffold_ID	Start	End
Contig_8	5345204	5506305
Contig_8	5708103	5840368
Contig_9	13692024	13739639
Contig_9	13895771	13904364

Table S7. The pseudo-chromosome of HS. Related Figure 1 and Table 1.

HS	Contigs		Super-scaffolds (Pseudo-chromosomes)	
	Size (bp)	Number	Size (bp)	Number
N90	928,690	110	24,873,067	11
N80	1,565,629	81	26,064,071	10
N70	2,257,174	61	28,609,348	8
N60	2,831,336	47	29,341,949	7
N50	3,451,447	35	29,735,421	6
Longest	12,575,093	-	35,776,948	-
Total Size	366,161,843	-	366,172,327	-
Gap Size	0	-	10,484	-
Non-Gap Size	366,161,843	-	366,161,843	-

* The contigs are generated from pseudo-chromosomes by breaking the gaps (Ns).

Table S8. The length of HS pseudo-chromosomes compared with DHL92.Related to Figure 1 and Table 1.

Pseudo-chromosome ID	Length (bp)	
	HS	DHL92_CM3.6.1
Chr_1	34,827,257	37,037532
Chr_2	24,147,995	27,064,691
Chr_3	29,341,949	31,666,927
Chr_4	32,650,583	34,318,044
Chr_5	29,735,421	29,324,717
Chr_6	35,776,948	38,297,372
Chr_7	27,089,338	28,958,359
Chr_8	33,089,338	34,765,488
Chr_9	24,873,067	25,243,276
Chr_10	28,609,348	26,663,822
Chr_11	32,603,692	34,457,057
Chr_12	26,603,692	27,563,660
Anchored	359,414,177 (98.15%)	375,360,399 (90.01%)
Unanchored (Chr00)	6,758,150 (1.85 %)	41,641,883 (9.99%)
Total	366,172,327 (100%)	417,002,282 (100%)

Table S9. Statistic of gap proportion of melon assemblies. Related to Figure 1 and Table 1.

Assembly	Total size(bp)	Effective size (bp)	Number of Gaps	Gap size (bp)	Gap proportion (%)
HS	366,172,327	366,161,843	212	10,484	0.003
DHL92_V3.6.1	417,002,282	337,325,314	44,652	79,676,968	19.11
DHL92_V3.5.1	406,928,820	336,097,046	60,625	70,831,774	17.41

Table S10. General statistics of repeats in HS genome. Related to Figure 1.

Type	Repeat Size	% of genome
Tandem Repeats	16,419,912	4.484203
Interspersed repeat	152,105,439	41.539305
Total	157,621,021	43.04559

Note: Some elements may partly include another element domain.

Table S11. TEs content in the HS genome. Related to Figure 1.

Type	Repbase TEs		TE proteins		De novo		Combined TEs	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
ClassI	40,122,070	10.957155	40,253,856	10.993145	115,267,394	31.479002	124,235,646	33.928191
LTR	39,491,521	10.784955	37,846,127	10.335605	113,646,804	31.036426	120,648,182	32.948471
Gypsy	23,841,789	6.511084	19,072,264	5.208549	67,912,864	18.546695	71,616,293	19.558084
Copia	14,821,974	4.047814	18,172,933	4.962945	39,511,260	10.790346	45,054,137	12.30408
LINE	718,304	0.196166	2,413,009	0.658982	1,804,333	0.492755	4,259,347	1.163208
SINE	7,977	0.002178	0	0	30,053	0.008207	38,030	0.010386
ClassII-DNA	11,370,095	3.105121	10,300,802	2.813102	22,111,026	6.038421	30,282,747	8.270081
Unclassified	4,857	0.001326	0	0	6,278,176	1.714541	6,283,033	1.715868
Total	51,120,945	13.960898	50,513,363	13.79497	139,744,406	38.163563	152,105,439	41.539305

Note: Some elements may partly include another element domain.

Table S12. Comparison of TE content. Related to Figure 1.

variety	Total TE coverage (%)	Coverage of Class I transposons (%)	Coverage of Class II transposons (%)
HS	41.54	33.93	8.27
DHL92_CM3.6.1	44	33.2	7.9

Table S13. Comparison of gene annotation between the HS and DHL92_CM4.0. Related to Figure 1.

Gene set	HPSG	DHL92_CM4.0
Gene number	28,898	29,980
Average gene length (bp)	3,917	3,596
Average mRNA length (bp)	1,052	974
Average exons per gene	4.67	4.61
Average exon length (bp)	225	211
Average intron length (bp)	630	564
Complete BUSCOs (%)	91.0	87.3
Fragmented (%)	3.0	4.7
Missing (%)	6.0	8.0

Note: BUSCOs analysis included 1440 embryophyta genes, BUSCO were run with the "--mode proteins --limit 20 --long" parameters.

Table S14. Functional annotation of the predicted genes in the assembly of HS. Related to Figure 1.

Type	Gene number	%
Total	28,898	100
Nr	26,960	93.29
Swissprot	17,272	59.77
KEGG	17,333	59.98
TrEMBL	26,342	91.16
Interpro	22,277	77.09
GO	13,025	45.07
Annotated	27,035	93.55
Unannotated	1,863	6.45

Table S15. Non-coding RNAs in the HS assembly. Related to Figure 1.

Type		Copy(w)	Average length(bp)	Total length(bp)	% of genome
miRNA		91	131	11,926	0.003257
tRNA		778	75	58,582	0.015999
rRNA	rRNA	433	373	161,645	0.044146
	18S	100	1,166	116,625	0.031851
	28S	232	136	31,448	0.008589
	5.8S	68	149	10,098	0.002758
	5S	33	105	3,474	0.000949
snRNA	snRNA	327	113	36,871	0.01007
	CD-box	174	96	16,644	0.004546
	HACA-box	45	126	5,678	0.001551
	splicing	108	135	14,549	0.003973

Note: '% of genome' was calculated by the non-gap genome size 366,161,843 bp

Table S16. Statistic analysis of gene families. Related to Figure 2.

Species	Genes number	Genes in families	Unclustered genes	Family number	Unique families	Average genes per family
C. melo_HS	27,154	24,015	3,139	17,603	271	1.36
C. melo_DHL92_CM4.0	28,590	23,169	5,421	17,934	360	1.29
C. lanatus	22,509	20,302	2,207	15,147	189	1.34
C. maxima	31,989	26,049	5,940	15,560	110	1.67
C. pepo	27,787	24,951	2,836	15,144	88	1.65
C. sativus	23,246	19,791	3,455	15,382	62	1.29
L. siceraria	22,408	19,090	3,318	14,631	46	1.3
A. thaliana	26,925	23,420	3,505	12,864	775	1.82
S. lycopersicum	33,866	26,050	7,816	13,759	1,059	1.89
O. sativa	42,135	30,189	11,946	13,176	2,233	2.29

Note: Genes encoding more than 50 proteins were used for this analysis. Unclustered genes refer to special gene of corresponding species; Unique families refer to special gene families of corresponding species.

Table S25. Clean and mapped reads of resequencing of the RILs population. Related to Figure 4.

Sample_ID	Reads num (M)	Base num (M)	GC (%)	Q20(%)	Total reads	Mapped reads	Mapping rate(%)
M	81.8	12154.85	36.22	96.75	81803478	80043021	97.85
R100	26.07	3869.38	36.5	96.9	26069840	25690702	98.55
R101	24.47	3633.15	37.04	96.7	24469314	24057222	98.32
R102	32.13	4777.83	37.72	97.58	32132326	31706652	98.68
R103	26.85	3986.44	36.56	97.16	26850250	26390276	98.29
R104	30.01	4453.92	36.68	97.11	30007110	29503458	98.32
R10	30.53	4539.35	36.38	97.27	30528588	30032304	98.37
R11	30.85	4583.63	36.55	96.8	30848892	30126494	97.66
R12	28.96	4301.85	36.3	96.64	28958780	28365463	97.95
R13	31.76	4724.9	36.38	97.13	31762828	31292240	98.52
R14	30.26	4496.12	36.21	97.14	30258202	29660678	98.03
R15	28.71	4261.9	36.27	97.03	28710804	28168540	98.11
R16	30.58	4542.44	36.09	97.15	30575520	30051341	98.29
R17	26.96	4008.49	36.47	97.04	26964424	26457356	98.12
R18	29.11	4326.31	36.55	97.25	29105258	28479635	97.85
R19	35.85	5330.59	36.43	97.29	35854450	35288087	98.42
R1	32.12	4777.16	36.7	97.09	32123122	31222426	97.20
R20	29.32	4357.63	36.54	97.21	29319046	28748829	98.06
R21	28.9	4293.36	36.3	97.13	28896626	28462418	98.50
R22	30.6	4553.65	36.41	97.24	30599700	30123391	98.44
R23	29.67	4407.92	36.15	97.02	29671526	29179508	98.34
R24	27.61	4101.58	36.32	97.13	27606578	27156101	98.37
R25	30.43	4524.62	36.42	97.3	30433554	29886299	98.20
R26	26.16	3889.19	36.37	96.98	26156254	25610577	97.91
R27	33.13	4928.24	36.68	97.2	33133924	32669328	98.60
R28	28.5	4240.26	36.15	97.1	28500702	28000458	98.24
R29	26.56	3943.71	36.93	97.08	26561218	25767528	97.01
R2	26.5	3935.58	36.39	97.15	26498564	26103692	98.51
R30	30.9	4594.56	36.35	97.07	30897862	30456762	98.57
R31	30.24	4493.76	36.57	96.93	30242484	29829732	98.64
R32	33.66	4998.18	36.41	97.07	33658678	32923860	97.82
R33	25.78	3830.4	36.19	96.74	25779872	25396577	98.51
R34	29.43	4367.58	36.34	96.95	29426808	28964476	98.43
R35	35.9	5332.23	36.44	97.09	35903396	35361168	98.49
R36	30	4453.85	36.28	96.92	30001858	29510188	98.36
R37	29.16	4335.29	36.47	97.35	29157662	28740560	98.57
R38	31.64	4705.61	36.61	97.2	31636986	31179944	98.56
R39	29.56	4388.03	36.16	96.95	29557122	29134651	98.57
R3	29.2	4338.48	36.31	96.74	29204338	28780983	98.55
R41	26.88	3993.39	36.29	96.79	26880458	26435090	98.34

R42	31.61	4694.61	36.26	97.09	31605544	31161179	98.59
R43	31.97	4746.14	36.22	97.02	31971768	31386607	98.17
R44	27.46	4083.54	36.52	96.85	27463370	27006923	98.34
R45	27.03	4016.3	36.57	96.78	27033164	26573270	98.30
R46	29.66	4412.39	36.34	97.17	29664294	29149058	98.26
R47	26.03	3863.07	36.42	96.83	26033866	25562435	98.19
R48	25.4	3772.91	36.71	97.15	25395052	24968212	98.32
R49	30.62	4549.91	36.21	96.83	30620314	30166310	98.52
R4	26.64	3956.86	36.28	97	26639308	26208559	98.38
R50	26.94	4005.24	36.47	96.86	26944382	26313310	97.66
R51	31.08	4627.36	36.27	97.32	31084910	30567911	98.34
R52	29.2	4336.08	36.28	96.91	29197690	28748359	98.46
R53	29.51	4384.9	36.15	97.11	29514600	28993442	98.23
R54	28.19	4187.41	36.33	96.73	28190376	27346415	97.01
R55	26.51	3935.25	36.36	96.77	26513992	26095046	98.42
R56	28.13	4180.08	36.55	97.15	28130190	27705886	98.49
R57	28.84	4286.25	36.41	96.85	28843172	28434370	98.58
R58	30.54	4541.78	36.3	97.41	30536824	30075348	98.49
R59	31.05	4618.55	36.35	97.14	31045428	30518535	98.30
R5	26.5	3938.54	36.21	96.86	26503952	26085060	98.42
R60	33.24	4937.26	36.21	96.98	33236314	32750858	98.54
R61	28.95	4299.66	36.64	97.18	28946954	28470160	98.35
R62	30.93	4598.37	36.81	97.19	30934768	30381034	98.21
R63	34.83	5186.08	36.58	97.63	34830998	34271439	98.39
R64	25.49	3791.52	36.77	97.27	25490266	25088812	98.43
R65	29.47	4379.32	36.4	97.27	29469762	29026878	98.50
R66	31.63	4691.47	36.57	96.76	31626970	31183979	98.60
R67	26.04	3871.52	36.63	97.46	26040436	25635626	98.45
R68	31.94	4749.22	36.56	97.06	31939046	31492436	98.60
R69	30.36	4502.86	36.45	96.94	30355240	29945314	98.65
R6	27.19	4035.33	36.43	96.95	27186476	26656381	98.05
R70	28.53	4238.79	36.5	97.18	28531480	28087581	98.44
R71	33.18	4931.48	35.88	96.96	33177184	32635411	98.37
R72	31.48	4680.15	36.3	97.58	31479024	30937398	98.28
R73	25.45	3777.68	36.51	97.02	25446240	25027257	98.35
R74	31.77	4721.11	36.39	97.44	31767446	31358298	98.71
R75	29.84	4430.85	36.23	96.75	29836060	29422899	98.62
R76	34.5	5128.12	36.31	97.38	34504316	34007431	98.56
R77	26.43	3923.07	36.34	97.04	26427058	25985909	98.33
R78	30.1	4475.74	36.58	97.42	30103164	29666941	98.55
R79	31.65	4701.26	36.63	97.2	31646546	31145577	98.42
R7	32.72	4860.3	36.23	97.1	32717012	32135509	98.22
R80	31.3	4653.24	36.62	97.3	31296126	30843972	98.56
R81	29.77	4427.33	36.1	97.34	29773092	28810300	96.77

R82	31.54	4688.27	36.24	97.46	31538728	31074304	98.53
R83	34.92	5195.58	36.48	97.24	34920060	34476527	98.73
R84	31.86	4729.6	36.24	96.99	31863806	31368025	98.44
R85	27.36	4061.37	36.21	97.02	27355464	26983490	98.64
R86	31.46	4671.3	36.44	97.03	31462810	30708558	97.60
R87	29.6	4394.57	36.29	97.05	29598076	29173454	98.57
R88	29.67	4404.1	36.29	96.89	29672648	29219082	98.47
R89	27.57	4081.27	36.55	96.23	27569508	27134626	98.42
R8	25.34	3766.3	36.45	96.78	25338794	24911638	98.31
R90	28.13	4180.09	36.64	96.83	28129144	27699071	98.47
R91	30.45	4519.79	36.32	96.91	30452574	29912219	98.23
R92	39.98	5929.1	36.09	96.75	39978042	39371594	98.48
R93	30.91	4594.1	36.11	97.33	30907370	30423363	98.43
R94	27.27	4049.9	36.33	96.7	27274484	26743362	98.05
R95	24.52	3635.58	36.26	96.33	24519014	24114276	98.35
R96	31.32	4649.98	36.61	97.04	31323428	30760530	98.20
R97	28.22	4187.02	36.35	96.99	28217772	27788250	98.48
R98	28.73	4267.89	36.31	96.71	28734452	28307124	98.51
R99	25.61	3806.74	36.73	96.91	25611866	25202131	98.40
R9	33.12	4927.84	36.46	97.33	33117192	32542676	98.27

Table S26. SNPs called from resequencing of the RILs population. Related to Figure 4.

Type	Number
UTR3	34186
UTR5	19108
UTR5\x3bUTR3	250
downstream	206259
exonic	137508
exonic\x3bsplicing	27
intergenic	2595546
intronic	494387
splicing	1099
upstream	246142
upstream\x3bdownstream	27151
Total	3761663

Table S28. Candidate genes involving in sucrose accumulation. Related to Figure 4.

Gene ID	gene model
MELO09061	Alpha-galactosidase
MELO16006	Beta-galactosidase
MELO16058	Putative sugar phosphate/phosphate translocator
MELO21692	Sucrose phosphatase
MELO21653	Neutral/alkaline invertase 3

Transparent Method

Plant materials

An inbred line of *C. melo* ssp. *agrestis* (HS) was selected for genome sequencing due to its key traits, including few-sucrose accumulation and resistance to a range of diseases, such as gummy stem blight (GSB) and *Fusarium* wilt. The true leaves from a single 20-day-old plant were harvested for PacBio and Illumina sequencing. Then 10-day-old etiolated seedlings grew on the MS medium were sampled for Hi-C sequencing. Two species were specifically selected to investigate sugar accumulation, transcriptomics and DNA methylation: HS and a double-haploid line DHL92 (a high-sucrose accumulator) which was previously derived from a cross between *inodorus* and *agrestis* subspecies. For the genetic mapping of sugar metabolism and GSb resistance , a recombinant inbred line (RIL) population by crossing HS with another high-sucrose accumulation line, *C. melo* ssp. *melo* (XH) for eight times was constructed. Finally, a total of 88 RILs were used in this study after eliminating questionable lines.

Genome sequencing

The genomic DNA were extracted from tissue samples by using an improved CTAB approach. A Phase Genomics kit (https://www.phasegenomics.com/hi-c-kits/) was used to prepare genomic DNA from tissues for Hi-C sequencing. Three different genome libraries were constructed and sequenced according to the manufacturer's instructions to generate a chromosome-scale assembly: (i) whole genome

sequencing (WGS) using a PacBio Sequel platform (20-kb library), (ii) Hi-C chromosome conformation captured reads sequencing by Phase Genomics, and (iii) short reads paired-end sequencing (150 bp in length) using an Illumina NovaSeq platform.

***De novo* assembly**

The PacBio sequencing data were assembled into contigs using FALCON (V2.0.5) and the data polishment was finished by FALCON-Unzip and Arrow (V2.2.2). Then we employed the Phase Genomics Proximo Hi-C Genome Scaffolding Pipeline to create chromosome scale scaffolds from the draft assembly (Bickhart et al., 2017) once we had assembled a draft set of contigs. Next, errors that had been introduced into the assembly in the long reads was rectified by Pilon (V1.22) program. In order to collect the best mapped reads for each marker, the markers from the published genetic maps of HS x XH (HS was used for genome sequencing in this study) (Hu et al., 2018) and PS x SC (described in the published DHL92 genome) (Argyris et al., 2015) were used to map to the new melon assembly with BWA-MEM. Following cutting the scaffolds where there are significant matches (at least 4 markers) to create multiple linkage groups by ALLMAPS program (Tang et al., 2015). Finally, we combined the HS x XH and PS x SC linkage groups, and then ordered and orientated the split scaffolds to re-construct chromosomes using ALLMAPS (with default parameters).

Genome annotation

We first used Tandem Repeats Finder (TRF, version 4.07) to identify the tandem repetitive sequences. Interspersed repeats were identified using *de novo* repeat identification and known repeat searches against existing databases. Then the *de novo* software packages including PILER (v1.0), Repeatscout (v1.0.5) and LTR_FINDER (v1.0.6) were used to predict repeat sequences in the assembly to generate the initial repeat library. We searched the genome against the library by RepeatMasker (version 4.0.7) and subsequently a homology-based approach was applied for common databases of known repetitive sequences. Finally, RepeatMasker (version 4.0.7) and Repbase database (version 21) were used in conjunction to identify TE repeats in the assembled genome. All programs used in this study were identical to those used to generate the published melon genome (Ruggieri et al., 2018); this practice maintained consistency for subsequent comparative analysis between the two genomes using the same criteria.

Protein-coding genes were predicted using a variety of *de novo*, protein homology and transcriptome-based approaches. Five *ab initio* gene prediction programs, Augustus (version 3.2.1), GlimmerHMM (version 3.0.4), Genscan (version 1.0), Geneid (version 1.4.4) and SNAP (version 2006-07-28), were used for the *de novo* prediction of coding regions in the repeat-masked genome. Next, the protein sequences downloaded from Phytozome (Release 11) and Cucurbit Genomics Database (CuGenDB) (*Arabidopsis thaliana*, *Oryza sativa*, *Solanum lycopersicum*, *Malus domestica*, *Cucumis sativus*, *Cucumis melo* (V4.0), *Citrullus lanatus*, *Cucurbita maxima*, *Lagenaria Siceraria* and *Cucurbita pepo*) were then aligned to the assembly

using genblasta (version 1.0.4). Next, we predicted the exact gene structure of the corresponding genomic regions on each genblasta hit by GeneWise (version 2.4.1) . Finally, RNA-seq data were mapped to the assembly using hisat2 (version 2.0.1), stringtie (version 1.2.2) and TransDecoder (version 3.0.1). The mapped RNA-seq data were then used to assemble the transcripts and identify candidate coding regions in the gene models. By using EvidenceModeler (EVM), the gene models predicted from these three different approaches were then combined into a non-redundant set of gene structures. The generated gene models were finally refined with the Program to Assemble Spliced Alignments (PASA v2.3.3). We then used BLASTP (E-value 1e-05) against two integrated protein sequence databases (SwissProt and TrEMBL) to analyze the functional annotations of protein-coding genes. Protein domains were then annotated by InterProScan (V5.19). The Gene Ontology (GO) terms for each gene were subsequently identified application of Blast2GO on the nr protein database. We blasted the potential pathways for the identified genes in the KEGG database (release 59.3), with an E-value cutoff of 1e-05.

Next, we used tRNAscan-SE (version 1.3.1) to identify tRNAs and predict rRNA fragments by alignment to the Arabidopsis and rice template rRNA sequences using BlastN (version 2.2.26) at an E-value of 1e-10. Then using INFERNAL (version 1.1.1) to identify miRNAs and snRNAs by searching the Rfam database (release 12.0).

Identification of chromosomal structure variation

The genome sequences of DHL92 and Payzawat were aligned to the HS assembly

using MUMmer (version 3.23) with default parameters. Next, the genomic alignment results were extracted with the delta-filter -1 -l 1000 parameters. We then identified the collinearity relationships in the genomes as grid lines. The collinear genetic marker sequences were mapped to the HS and DHL92 assemblies using BWA (version 0.7.5a) with default parameters. The physical positions on the reconstructed pseudo-chromosome was connected by the grid lines mentioned before .

Evolutionary and positive selection gene analysis

A number of single copy gene families from *C. lanatus*, *L. siceraria*, *C. melo* ssp. *melo*, *C. melo* ssp. *agrestis*, *C. sativus*, *C. maxima*, *C. pepo*, *O. sativa*, *S. lycopersicum* and *A. thaliana* were identified by OrthoMCL program (Li et al., 2003). Using these single copy orthologous genes, we then constructed a phylogenetic tree and estimated the divergence time among species. The Computational Analysis of Gene Family Evolution (CAFE, version 2.1) was used to detect gene family expansion and contraction (De Bie et al., 2006) once the constructed phylogenetic tree being annotated with the estimated divergence times . Expanded and contracted gene families in the *Cucurbitaceae* family were subjected to GO enrichment analysis. P values were determined using Fisher's exact test and were adjusted by the Benjamini-Hochberg (BH) method.

Next, based on the created phylogenetic tree for *Cucurbitaceae* species, we incorporated a branch-site model into the PAML package and screened for Positive Selection Genes (PSGs). The branch of *Cucurbitaceae* species was used as the

foreground branch, while the branches created for *O. sativa*, *S. lycopersicum*, *A. thaliana* branches were used as background branches. The null model used in the branch-site test assumed that the Ka/Ks values for the codons in all branches must be ≤1, whereas the alternative model assumed that the foreground branch included codons evolving with Ka/Ks >1. A maximum likelihood ratio test (LRT) was used to compare the null model with the alternative model. The P value was calculated using the Chi-squared distribution with 1 degree of freedom (df = 1). Then the P values were adjusted for multiple testing using the false discovery rate (FDR) method. Genes were identified as being positively selected when the FDR < 0.05. At least one amino acid site possessed needed to possess a high probability of being positively selected (Bayes probability >95%). If none of the amino acids passed this cutoff threshold in the PSGs, then the gene under investigation was identified as a false positive and was excluded.

Transcriptome and DNA methylation analysis

Paired-end RNA-seq libraries were prepared from with 1 μg of total RNA using aTruSeqTM RNA Kit (Illumina, San Diego, CA) to prepare the Paired-end RNA-seq. The libraries were then sequenced by the Illumina HiSeq PE 2X151 bp cycles (read length). Clean reads were then aligned to the HS genome using HIAST2 (version 2.0.5) (Kim et al., 2015). Next, we calculated the expression level of each transcript using the fragments per kilobase of exon per million mapped reads (FPKM) method. SAM files were then sorted to allow differential gene expression analysis using functions

within the R package DESeq2 (version 1.16.1). The GOSeq (version 1.10.0) and KOBAS (version v2.0.12) functions in the R package were employed to perform GO and KEGG annotation and enrichment analyses.

For sequencing, 25 ng of lamda-DNA was first mixed with 5 μg genomic DNA. This mixture of DNAs was then fragmented to sizes of approximately 450 bp with a Sonicator and then supplemented by adapters. Next, by application of the ZYMO EZ DNA Methylation-Gold Kit (ZYMO), the DNAs were treated with bisulfite. Ultra-high-throughput pair-end sequencing was then carried out using the Illumina Hiseq system. Clean BS-seq reads were then mapped to the reference genome by the BSMAP aligner (version v2.90) (Li and Li, 2009). Uniquely mapped reads were then used to determine the levels of cytosine methylation (Shao et al., 2014). We employed the DSS package in the R package to detect differential methylation regions (DMRs) in smoothing mode and based on CpG sites (Feng et al., 2014). Differentially methylated loci (DML) were first identified with default parameters (P value<1e-5). Then, DMRs were called on the basis that the DML required a minimum read of 50 bp and a minimum of 3 CpG sites per region, neighboring DMRs were combined if the distance was less than 100 bp.

The construction of RILs and the QTLs mapping

HS plants were crossed with XH (a high-sucrose accumulator of *melo* subspecies line) and continuously self-crossed for eight times generation to create population of recombination inbred lines (RILs). 103 RILs true leaves tissues were sampled to

extract the genomic DNAs for resequencing on the Illumina Hiseq platform. On average, the use of these sequences allowed us to collate a 10-fold coverage of the genome for each line. Clean reads were mapped to the HS genome using BWA and SAMtools and then SNPs were identified with GATK. After filtration and eliminating questionable RILs, high-quality SNPs were used to construct a physical recombination map and Bin-map (Huang et al., 2009). We calculated the genetic distance between adjacent markers using the Kosambi program and QTLs mapping using WinQTLcart. The genotyping analysis were further carried out by cloning the candidate genes CDS that were mapped by QTLs analysis from natural varieties exhibiting few/high-sucrose accumulation, and GSB susceptibility/resistance.

Sugar content and GSB resistance assay

Flesh from the equator of each fruit was uniformly collected with a sampler from developing HG and DHL92 melon fruits (0, 10, 20, 30 and 40 days after pollination) for transcriptomic and DNA methylation analyses. RILs melons were sampled 35 days after pollination and weighted 1 mg fresh flesh. Soluble sugar (sucrose,fructose and glucose) were extracted by 80% ethanol then digested sufficiently in 45℃ incubator, three times repeat to get 10 ml complete extraction. 1 ml upper layer extraction liquid was dried up by rotary evaporator. The sugar content was then determined using high performance liquid chromatography (HPLC) after suction filtration the resuspended sugar solution by 1 ml ultrapure water.

Didymella bryoniae fungi were cultured on PDA medium plates in dark for one

week, and then under both dark and ultra-violet light (12 h UV and 12 h darkness). Next, we collected the pathogenic conidial suspension and calculated the concentration of conidia with a hemocytometer. We then diluted the conidial suspension to 500,000 spores per mL. 20 drops of Tween 20 were then added to the diluted suspension and adjusted the pH to 4 with lactic acid. The adjusted conidial suspensions were then sprayed on 4-6 true leaves while maintaining the air humidity above 90% and the temperature approximately 25℃ for the next 3 days. Two to three weeks after inoculation, we determined the grades of disease resistance exhibited by the different melon plants according to established criteria.

Supplemental References

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 targeted SNP selection for an improved anchoring of the melon (*Cucumis melo* L.) scaffold genome assembly. *BMC Genomics,* 16**,** 4.

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., Lee, J., Lam, E. T., Liachko, I., Sullivan, S. T*., et al.* 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics,* 49**,** 643-650.

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

Feng, H., Conneely, K. N. and Wu, H. 2014. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Research,* 42**,** e69-e69.

Hu, Z. Y., Deng, G. C., Mou, H. P., Xu, Y. H., Chen, L., Yang, J. H. and Zhang, M. F. 2018. A re-sequencing-based ultra-dense genetic map reveals a gummy stem blight resistance-associated gene in Cucumis melo. *DNA Research,* 25**,** 1-10.

Huang, X. H., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A. H., Guan, J. P., Fan, D. L., Weng, Q. J., Huang, T*., et al.* 2009. High-throughput genotyping by whole-genome resequencing. *Genome Research,* 19**,** 1068-1076.

Kim, D., Landmead, B. and Salzberg, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods,* 12**,** 357-360.

Li, L., Stoeckert, C. J. and Roos, D. S. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research,* 13**,** 2178-2189.

Li, Y. X. and Li, W. 2009. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics,* 10**,** 232.

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. *Scientific Reports,* 8**,** 8088.

Shao, C. W., Li, Q. Y., Chen, S. L., Zhang, P., Lian, J. M., Hu, Q. M., Sun, B., Jin, L. J., Liu, S. S., Wang, Z. J.*, et al.* 2014. Epigenetic modification and inheritance in sexual reversal of fish. *Genome Research,* 24**,** 604-615.

Tang, H. B., Zhang, X. T., Miao, C. Y., Zhang, J. S., Ming, R., Schnable, J. C., Schnable, P. S., Lyons, E. and Lu, J. G. 2015. ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biology,* 16**,** 3.