nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-019-0518-4

In the format provided by the authors and unedited.

# Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits

**Shaogui Guo[1,9], Shengjie Zhao[2,9], Honghe Sun[1,3,9], Xin Wang[3,9], Shan Wu[3,9], Tao Lin[4,9], Yi Ren[1], Lei Gao[3], Yun Deng[2], Jie Zhang[1], Xuqiang Lu[2], Haiying Zhang[1], Jianli Shang[2], Guoyi Gong[1], Changlong Wen[1], Nan He[2], Shouwei Tian[1], Maoying Li[1], Junpu Liu[2], Yanping Wang[1], Yingchun Zhu[2], Robert Jarret[5], Amnon Levi[6], Xingping Zhang[1], Sanwen Huang[4,7*], Zhangjun Fei[3,8*], Wenge Liu[2*] and Yong Xu[1*]**

[1]National Watermelon and Melon Improvement Center, Beijing Academy of Agricultural and Forestry Sciences, Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (North China), Beijing Key Laboratory of Vegetable Germplasm Improvement, Beijing, China. [2]Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, China. [3]Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY, USA. [4]Lingnan Guangdong Laboratory of Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [5]U.S. Department of Agriculture-Agricultural Research Service, Plant Genetic Resources Conservation Unit, Griffin, GA, USA. [6]U.S. Department of Agriculture-Agricultural Research Service, U.S. Vegetable Lab, Charleston, SC, USA. [7]Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of the Ministry of Agriculture, Sino-Dutch Joint Laboratory of Horticultural Genomics, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China. [8]U.S. Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA. [9]These authors contributed equally: Shaogui Guo, Shengjie Zhao, Honghe Sun, Xin Wang, Shan Wu, Tao Lin. *e-mail: huangsanwen@caas.cn; zf25@cornell.edu; liuwenge@caas.cn; xuyong@nercv.org

**Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits**

Shaogui Guo[1,2,9], Shengjie Zhao[3,9], Honghe Sun[1,2,9], Xin Wang[2,9], Shan Wu[2,9], Tao Lin[4,9], Yi Ren[1], Lei Gao[2], Yun Deng[3], Jie Zhang[1], Xuqiang Lu[3], Haiying Zhang[1], Jianli Shang[3], Guoyi Gong[1], Changlong Wen[1], Nan He[3], Shouwei Tian[1], Maoying Li[1], Junpu Liu[3], Yanping Wang[1], Yingchun Zhu[3], Robert Jarret[5], Amnon Levi[6], Xingping Zhang[1], Sanwen Huang[4,7], Zhangjun Fei[2,8], Wenge Liu[3] & Yong Xu[1]

[1]National Watermelon and Melon Improvement Center, Beijing Academy of Agricultural and Forestry Sciences, Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (North China), Beijing Key Laboratory of Vegetable Germplasm Improvement, Beijing, China.
[2]Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY, USA.
[3]Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, China.
[4]Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China.
[5]U.S. Department of Agriculture-Agricultural Research Service, Plant Genetic Resources Conservation Unit, Griffin, GA, USA.
[6]U.S. Department of Agriculture-Agricultural Research Service, U.S. Vegetable Lab, Charleston, SC, USA.
[7]Lingnan Guangdong Laboratory of Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China.
[8]U.S. Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA.
[9]These authors contributed equally to this work.

Correspondence should be addressed to Yong Xu (xuyong@nercv.org), Wenge Liu (liuwenge@caas.cn), Zhangjun Fei (zf25@cornell.edu) and Sanwen Huang (huangsanwen@caas.cn).

## Contents

**Supplementary Note**

## 1. Evaluation of the watermelon '97103' genome assembly

### *1.1 Coverage by Illumina reads*

A total of 45,504,289 '97103' cleaned Illumina read pairs with read lengths of 150 bp, which corresponded to ~13.5 Gb sequences covering ~31.4× of the watermelon genome, were aligned to the newly assembled '97103' genome (v2) using the BWA aln algorithm[1] with parameters "-n 0.03 -o 1 -e 2". About 94.30% of the reads were mapped to the v2 assembly, whereas 88.01% could be mapped to the previous version of '97103' genome assembly[2] (v1). In the 359,800,263 bp non-gapped region of the v2 assembly, 359,433,964 bp (~99.9%) were covered by at least 1 read, and 358,962,274 bp (~99.7%) were covered by at least 5 reads. Further check of the 5.7% unmapped reads showed that 83.9% of them could be mapped to the plastid (GenBank acc#: KY430692) or mitochondrion (GenBank acc#: NC_014043) genomes, indicating their organelle origin.

To confirm that there were no potential contaminated sequences of microorganism or organelle origin in the '97103' v2 assembly, we calculated the GC content of each of the non-overlapping 500-bp sliding windows and the average per-base sequencing coverage of the region, which indicated no signature of these contaminations (**Extended Data Fig. 5**).

Potential collapsed repeat regions were identified based on the Illumina read coverage. Genomic regions with an average coverage higher than three times of the genome-wide average coverage were considered as potential repeats. Our result indicated that these potential collapsed repeat regions had an accumulative length of ~891 kb (0.25% of the '97103' v2 assembly).

### *1.2 Evaluation of gene space coverage*

The completeness of the '97103' v2 assembly in terms of gene content was first evaluated by BUSCO[3] (v2.0.1) with the "embryophyta_odb10" dataset. About 96.80% of the core conserved plant genes (1,331 out of 1,375 BUSCOs) were found complete in the v2 genome assembly, and 94.69% (1,302 out of 1,375) were found complete in the v2 predicted genes (**Supplementary Table 3**).

We then evaluated the gene space coverage by mapping RNA-Seq reads to the '97103' v2 assembly using HISAT2 (ref. [4]) allowing up to 3% mismatches. RNA-Seq reads from a total of 72 libraries, which were constructed from six different tissues of '97103' (apical point, carpopodium,

fruit flesh, stem, leaf and root) sampled at four different time points (10, 18, 26 and 34 days after pollination), were used for mapping. Up to 96.5% of the RNA-Seq reads could be aligned to the v2 assembly, and the mean mapping rate of RNA-Seq data from all libraries was 92.55%. The high BUSCO coverage and RNA-Seq read mapping rates indicated the watermelon genes were well captured by the v2 assembly.

Furthermore, we evaluated the quality of genes predicted in the '97103' v2 assembly. Of the 22,596 predicted genes, 21,048 (93.15%) were supported by the RNA-Seq data, with an RPKM (reads per kilobase exon model per million mapped reads) of at least one, 20,008 (88.55%) were supported by homologous proteins from public databases, and 18,927 (83.76%) were supported by both. Only 467 (2.01%) genes were purely derived from *ab initio* predictions without support from the RNA-Seq data or homologous proteins. These results together with the high BUSCO coverage by the predicted genes (above) suggested the high quality of the predicted genes in the '97103' v2 assembly.

### *1.3 Assessment of assembly contiguity with Long terminal repeat (LTR) Assembly Index*

LTR Assembly Index[5] (LAI) of the '97103' v2 assembly was calculated to evaluate the contiguity and compared to those of other genomes. The intact LTR retrotransposon content of '97103' v1 assembly was too low for performing this analysis. Therefore, the recently published Illumina short-read assembly of watermelon 'Charleston Gray'[6] and the PacBio long-read assembly of cucumber 'Chinese long'[7] were used here for comparison. The LAI score of '97103' v2 assembly was 7.8±3.4, which was much better than that of the 'Charleston Gray' assembly (4.7±2.6), and comparable to that of the 'Chinese long' long-read assembly (8.3±6.1), supporting the high contiguity of the '97103' v2 assembly.

### *1.4 BAC sequence alignment*

Four fully sequenced watermelon '97103' BAC sequences (GenBank accessions: JN402338, JN402339, JX027061 and JX027062) were aligned to the '97103' v1 and v2 assemblies using BLAST[8] with no filtering of low complexity regions. For all BACs, their complete sequences were covered by the v2 assembly except the first 32 bases of the BAC JX027061, compared to 97.8% (JN402338), 97.6% (JN402339), 90.2% (JX027061) and 64.2% (JX027062) covered by the v1 assembly[2] (**Extended Data Fig. 6**). Further comparison of the first 32 bases of the BAC JX027061

to the GenBank nt database indicated this 32-bp sequence was from a cloning vector. Therefore, the v2 assembly correctly captured the entire sequences of the four fully sequenced BACs. In addition to the increased coverage, the v2 assembly also had higher sequence identity to the BAC sequences (**Extended Data Fig. 6**).

### *1.5 Collinearity with genetic maps*

The '97103' v2 assembly had high collinearity with the three published watermelon genetic maps[9-11] for most chromosomes (**Extended Data Fig. 3**). An inversion on chromosome 11 was suggested by two genetic maps. However, the structure of chromosome 11 was consistent with the BioNano map and the regions near the potential "breakpoints" were well supported by '97103' mate-pair reads (**Extended Data Fig. 7**), indicating the accuracy of the assembly. The inconsistency between the assembly and the genetic maps could be due to naturally occurring structural variations in the populations used for map construction: *C. amarus* USVL246 × USVL114 (ref. [9]) and elite watermelon HMw017 (*Fon* race 1 resistant) × HMw013 (susceptible)[10], or errors in the genetic map[11].

## 2. Repeat contents in the watermelon '97103' genome assemblies

A total of 202.8 Mb of repeat sequences were identified in the '97103' v2 assembly, accounting for 55.55% of the assembly (**Supplementary Table 4**). On the other hand, 164.7 Mb of repeats (46.6%) were found in the '97103' v1 assembly. Among the repeats in the v2 assembly, 156.1 Mb (77.15% of identified repeats) were classified into known repeat families, with the Gypsy- (72.5 Mb; 35.83% of the identified repeats) and Copia-type (30.6 Mb; 15.13% of the identified repeats) LTR retrotransposons (LTR-RTs) being the predominant types. The repeat composition of '97103' v2 assembly was largely consistent with that in the v1 assembly (**Supplementary Table 4**). A total of 519 intact (fully-assembled) LTR-RTs with an accumulative length of 4,432,097 bp were identified in the '97103' v2 assembly, whereas only 65 (330,102 bp) full-length LTR-RTs could be identified in the v1 assembly. Together, the '97103' v2 assembly captured more repeat sequences and had a higher abundance of intact LTR-RTs in comparison to v1 owing to the improved assembly using PacBio long reads.

## 3. Candidate fruit quality-related genes in the selective sweeps

Selective sweeps overlapping with fruit weight QTLs *Qfwt2-1* and *Qfwt3* contained a gene (*Cla97C02G038570*) homologous to *SLOW GROWTH 1* (ref. [12]), and an auxin transporter (*Cla97C03G060620*) and cell wall metabolism (*Cla97C02G036910* and *Cla97C03G060980*) genes, all of which could to be involved in regulating organ size (**Supplementary Table 15**). *Cla97C02G036910*, encoding a xyloglucan endotransglucosylase/hydrolase, was highly expressed in the flesh at the early stages of fruit development of cultivated watermelon while expressed at very low levels in the flesh of wild watermelon. *Cla97C03G060620* was mainly expressed in the rind of cultivated fruit while *Cla97C02G038570* and *Cla97C03G060980* were either not expressed or expressed at a very low level in the flesh and rind (**Supplementary Table 15**).

The candidate gene underlying QTL *QBRX2-1*, which overlaps with both domestication improvement sweeps, is *ClTST2*, which encodes a sugar transporter and has been functionally proved to be a key transporter responsible for accumulating sucrose, fructose, and glucose in the vacuole of watermelon fruit cells[13]. QTL *QBrix6*, only overlapping with domestication sweeps, contains another sugar transporter gene, *Cla97C06G118310*, which is homologous to Arabidopsis *AtPMT5*/*AtPLT5* that transports sugars into the specific sink tissue[14] and was expressed in the flesh at the early stage of fruit development (**Supplementary Table 15**). In addition, a cluster of three sugar transporter genes, *Cla97C10G190640*, *Cla97C10G190650* and *Cla97C10G190690*, homologous to tomato *SlSTP1* (ref. [15]), were found in the improvement sweeps overlapping with QTL *FCE10.1* (**Fig. 3c**). However, all three genes were expressed at very low levels in the flesh during fruit development (**Supplementary Table 15**). Further experiments are required to investigate whether they play roles in sugar accumulation in watermelon flesh cells.

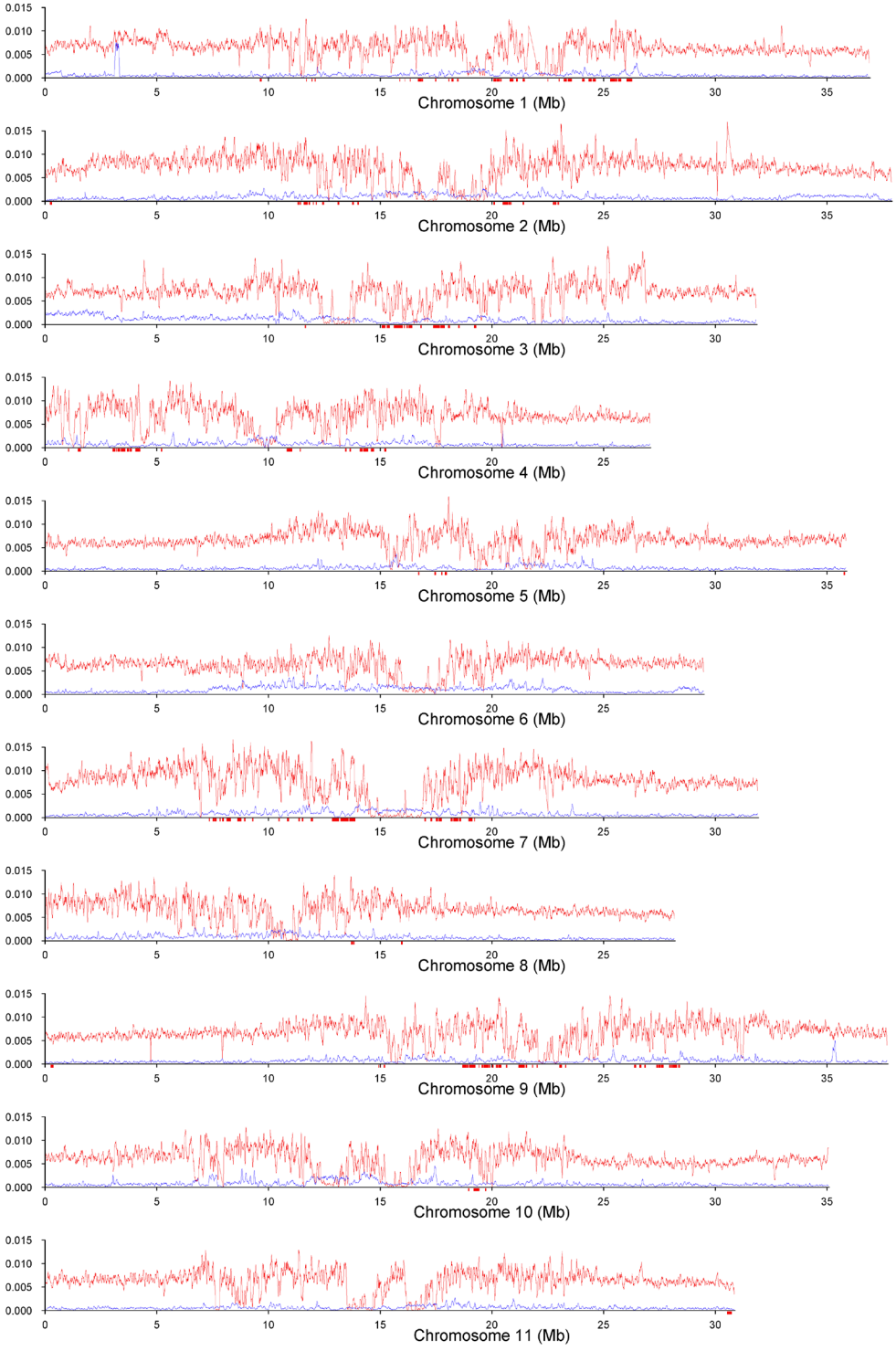
A cluster of genes on chromosome 2, including five benzyl alcohol O-benzoyltransferase (BEBT) genes (*Cla97C02G034580*, *Cla97C02G034590*, *Cla97C02G034600*, *Cla97C02G034610* and *Cla97C02G034620*) in the volatile ester biosynthesis pathway, were found in the improvement sweeps, but were not under selection during the domestication of sweet watermelon or the speciation process of *C. mucosospermus* (**Fig. 3c**). Among these BEBT genes, only *Cla97C02G034580* was expressed in fruit flesh and its expression was substantially higher in cultivated watermelon than in wild (**Supplementary Table 15)**, suggesting its potential contribution to the fruit aromatic flavor of sweet watermelon. Several other fruit quality-related genes were found in the improvement sweeps, including those encoding β-galactosidase (*Cla97C01G005260*, *Cla97C01G005270* and *Cla97C07G140540*), pectinesterase

(*Cla97C07G141130* and *Cla97C09G179380*) and carotenoid isomerase (*Cla97C10G190930*), among which the β-galactosidase gene *Cla97C07G140540*, the two pectinesterase genes and the carotenoid isomerase gene were expressed in fruit flesh and rind (**Supplementary Table 15**), indicating that these genes may contribute to the improved texture and nutritional profiles of cultivated watermelon fruit.

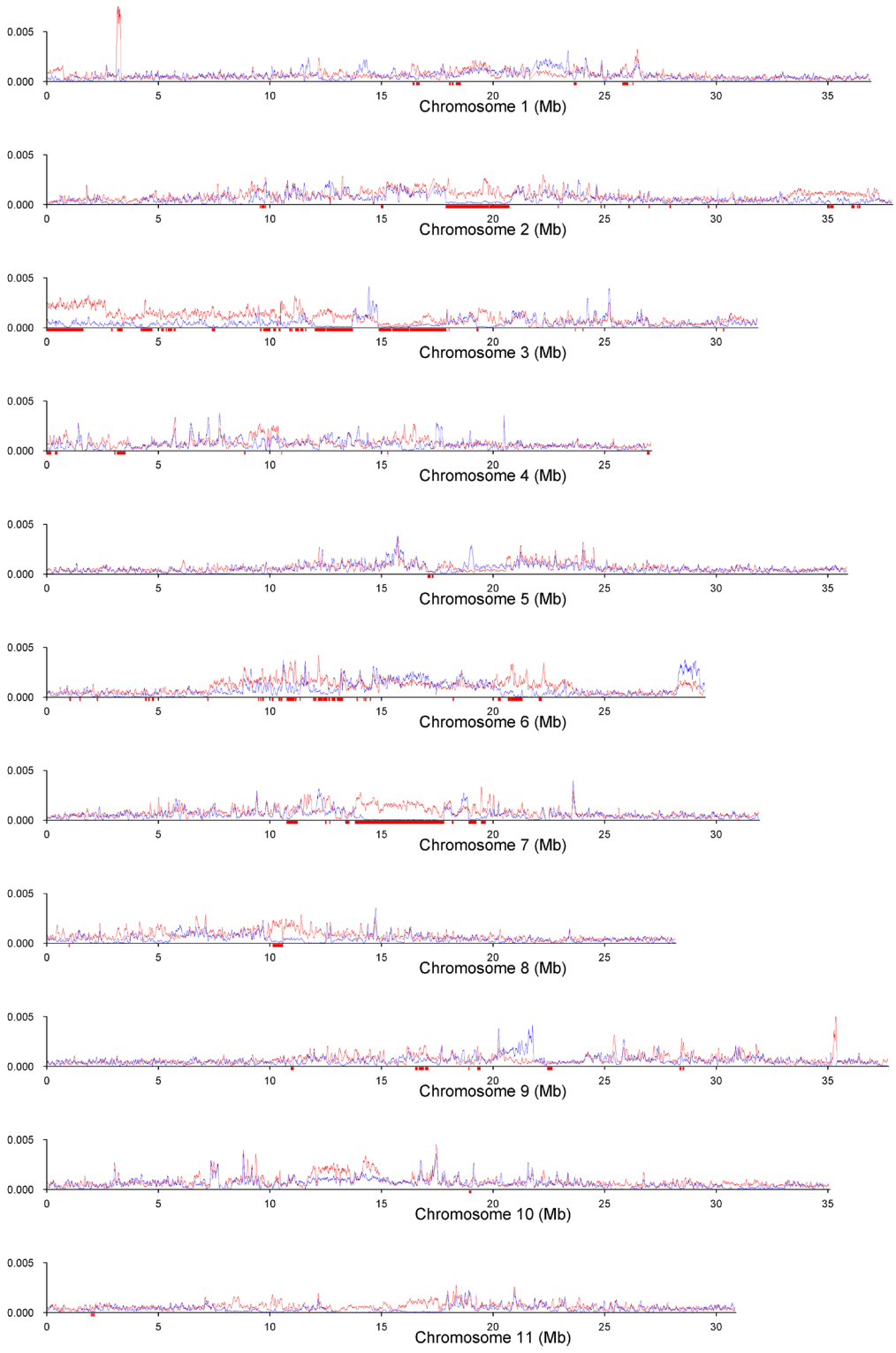
## References


1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
2. Guo, S. G. et al. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* **45**, 51–58 (2013).
3. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
4. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
5. Ou, S. J., Chen, J. F. & Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126 (2018).
6. Wu, S. et al. Genome of 'Charleston Gray', the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the U.S. National Plant Germplasm System watermelon collection. *Plant Biotechnol. J.* doi: 10.1111/pbi.13136 (2019).
7. Li, Q. et al. A chromosome-scale genome assembly of cucumber (*Cucumis sativus* L.). *Gigascience* **8**, giz072 (2019).
8. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
9. Branham, S. E., Levi, A., Farnham, M. W. & Wechter, W. P. A GBS-SNP-based linkage map and quantitative trait loci (QTL) associated with resistance to *Fusarium oxysporum* f. sp. *niveum* race 2 identified in *Citrullus lanatus* var. *citroides*. *Theor. Appl. Genet.* **130**, 319–330 (2017).
10. Lambel, S. et al. A major QTL associated with *Fusarium oxysporum* race 1 resistance identified in genetic populations derived from closely related watermelon lines using selective genotyping and genotyping-by-sequencing for SNP discovery. *Theor. Appl. Genet.* **127**, 2105–2115 (2014).
11. Ren, Y. et al. An integrated genetic map based on four mapping populations and quantitative trait loci associated with economically important traits in watermelon (*Citrullus lanatus*). *BMC Plant Biol.* **14**, 33 (2014).
12. Sung, T. Y., Tseng, C. C. & Hsieh, M. H. The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in Arabidopsis mitochondria. *Plant J.* **63**, 499–511 (2010).
13. Ren, Y. et al. A tonoplast sugar transporter underlies a sugar accumulation QTL in watermelon. *Plant Physiol.* **176**, 836–850 (2018).
14. Reinders, A., Panshyshyn, J. A. & Ward, J. M. Analysis of transport activity of *Arabidopsis* sugar alcohol permease homolog AtPLT5. *J. Biol. Chem.* **280**, 1594–1602 (2005).
15. Reuscher, S. et al. The sugar transporter inventory of tomato: genome-wide identification and expression analysis. *Plant and Cell Physiol.* **55**, 1123–1141 (2014).

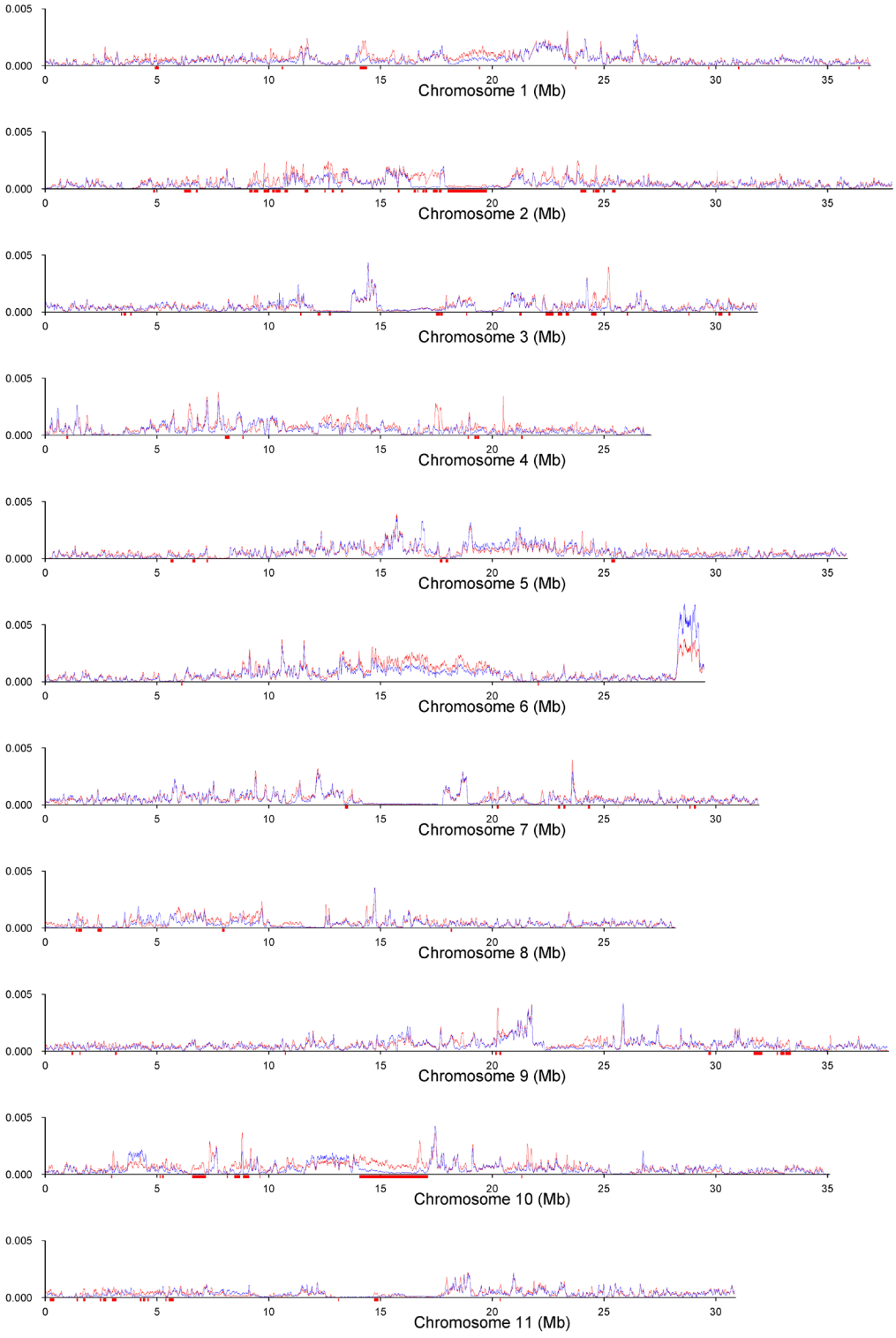
## Supplementary Figures



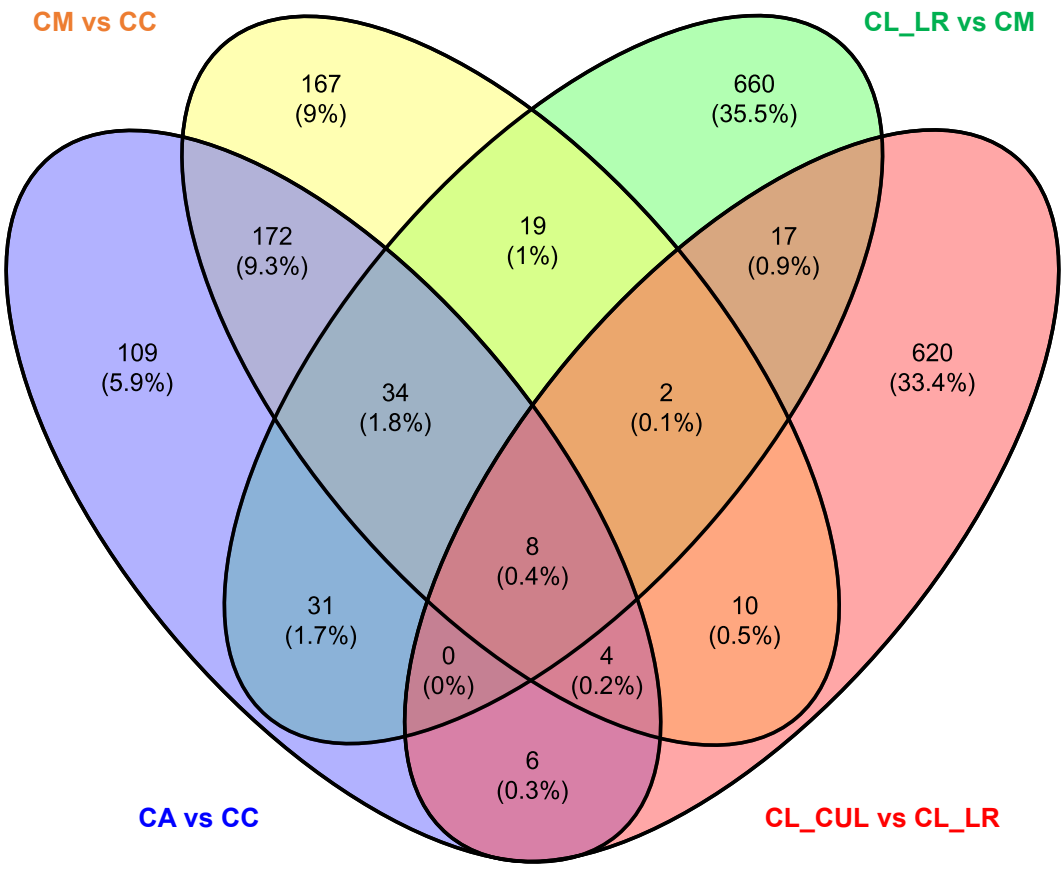

**Supplementary Fig. 1** Distribution of nucleotide diversity ($\pi$) for *C. mucosospermus* (blue) and *C. colocynthis* (red) across the 11 watermelon chromosomes. Red bars below the x-axis indicated candidate selective sweeps.



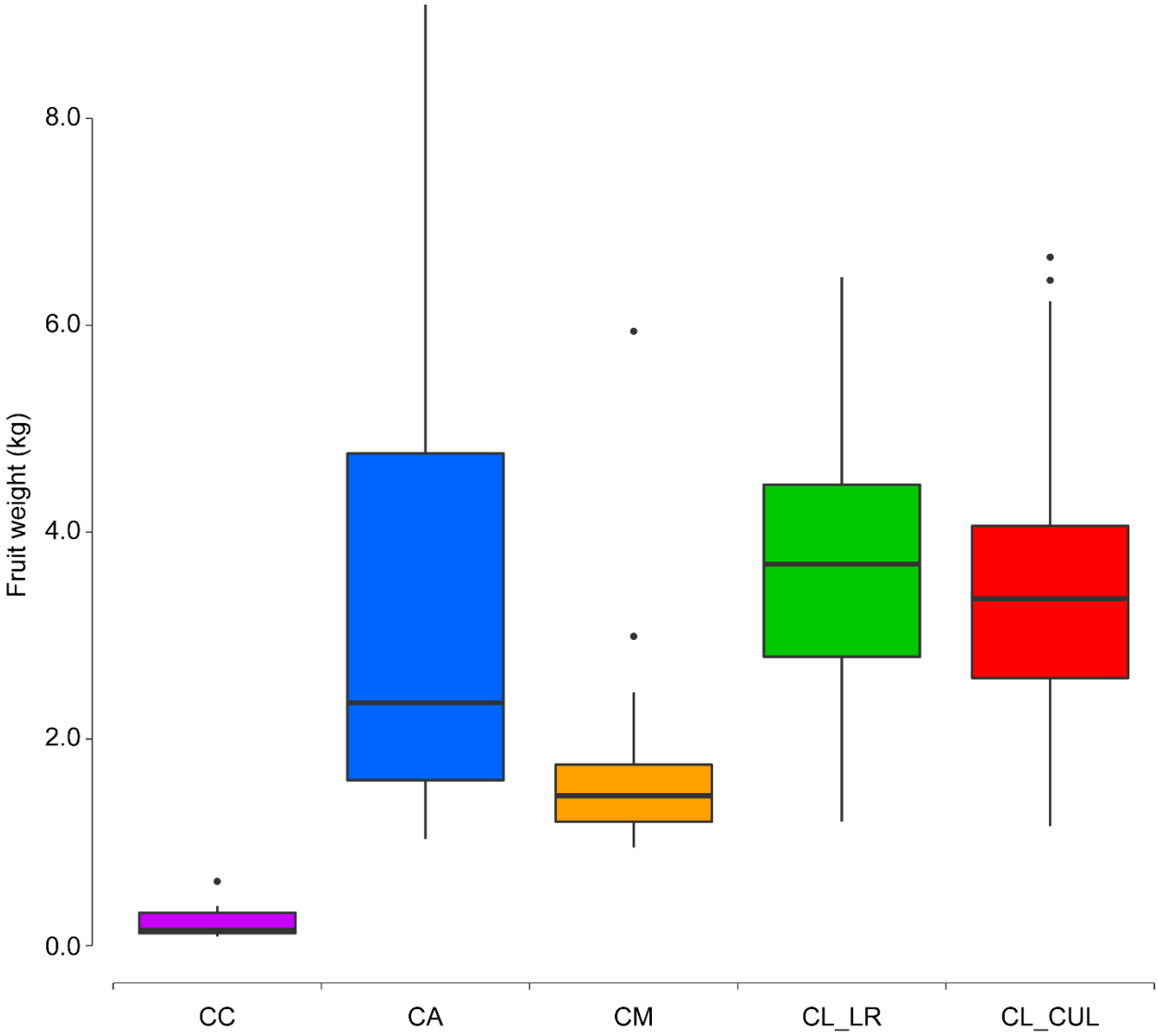

**Supplementary Fig. 2** Distribution of nucleotide diversity (π) for *C. lanatus* landrace (blue) and *C. mucosospermus* (red) across the 11 watermelon chromosomes. Red bars below the x-axis indicated candidate selective sweeps.



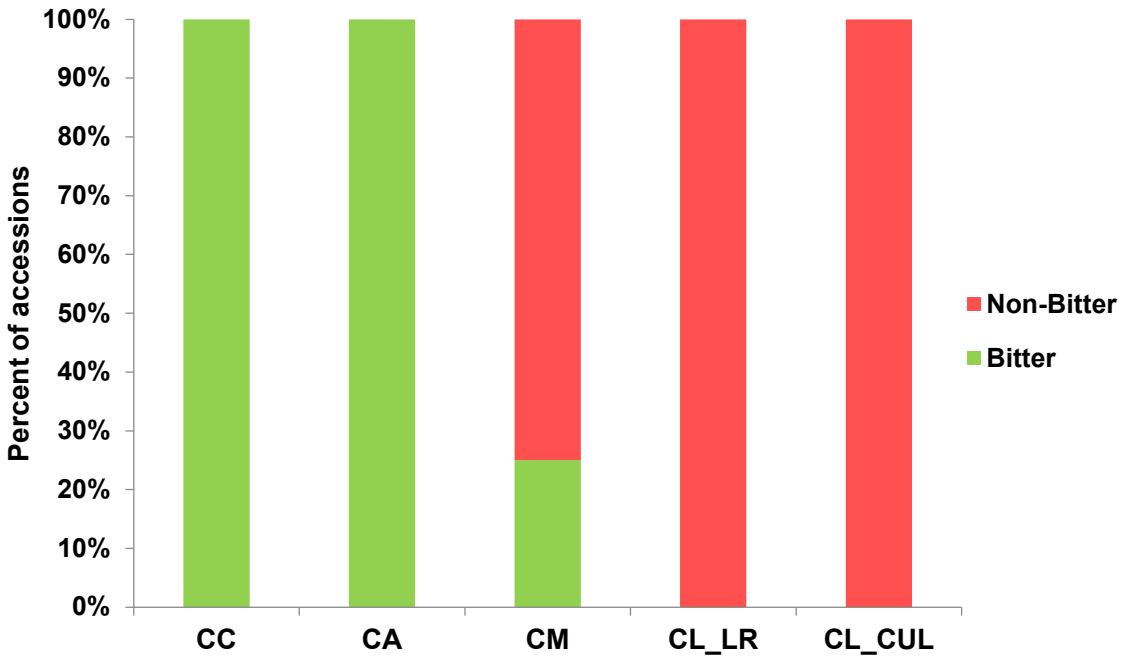

**Supplementary Fig. 3** Distribution of nucleotide diversity (π) for *C. lanatus* cultivar (blue) and *C. lanatus* landrace (red) across the 11 watermelon chromosomes. Red bars below the x-axis indicated candidate selective sweeps.



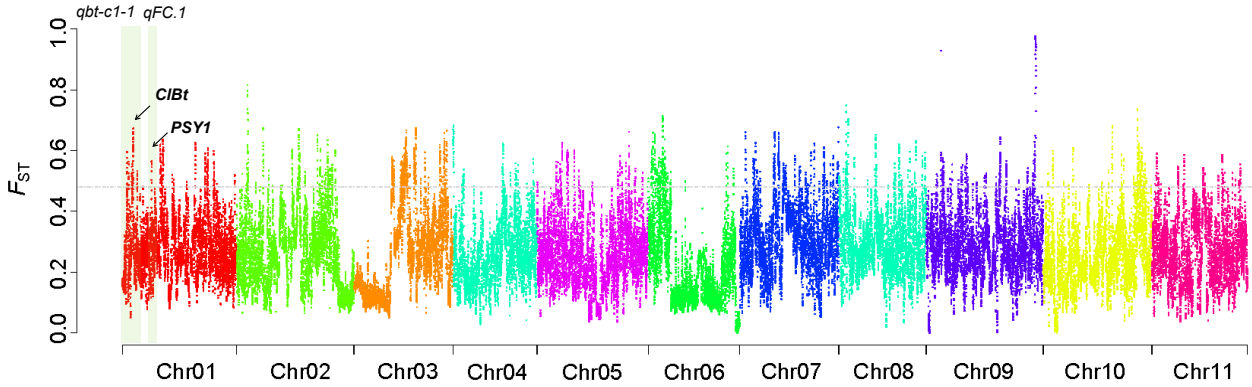

**Supplementary Fig. 4** Venn diagram of genes in selective sweeps during watermelon speciation, domestication and improvement. CM vs CC, selective sweeps in *C. mucosospermus* (CM) compared to *C. colocynthis* (CC) (speciation sweeps); CL_LR vs CM, selective sweeps in *C. lanatus* landraces (CL_LR) compared to CM (domestication sweeps); CA vs CC, selective sweeps in *C. amarus* (CA) compared to CC (speciation sweeps); CL_CUL vs CL_LR, selective sweeps in *C. lanatus* cultivar (CL_CUL) compared to CL_LR (improvement sweeps).



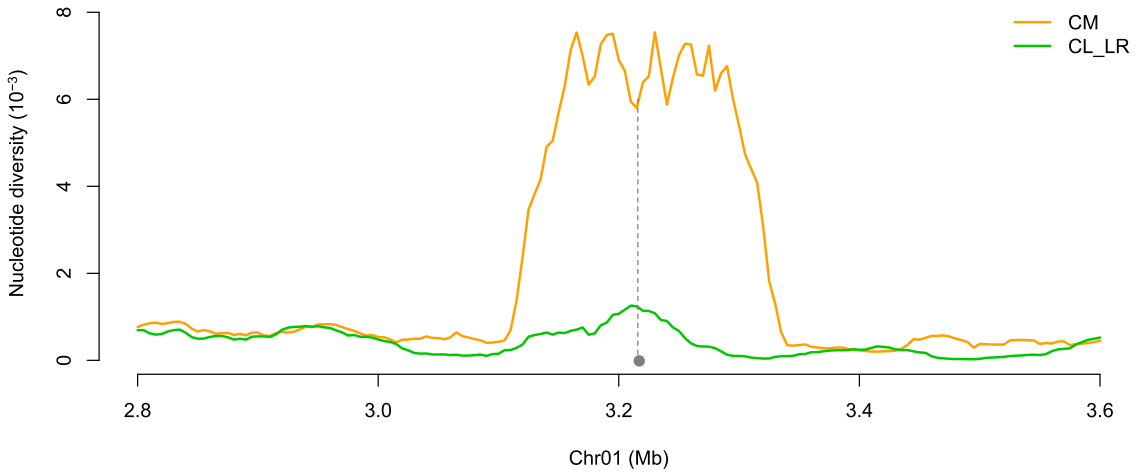

**Supplementary Fig. 5** Fruit weight in different *Citrullus* species. CC, *C. colocynthis* (n = 7 accessions); CA, *C. amarus* (n = 21 accessions); CM, *C. mucosospermus* (n = 16 accessions); CL_LR, *C. lanatus* landrace (n = 85 accessions); CL_CUL, *C. lanatus* cultivar (n = 221 accessions). For each boxplot, the lower and upper bounds of the box indicate the first and third quartiles, respectively, and the center line indicates the median.



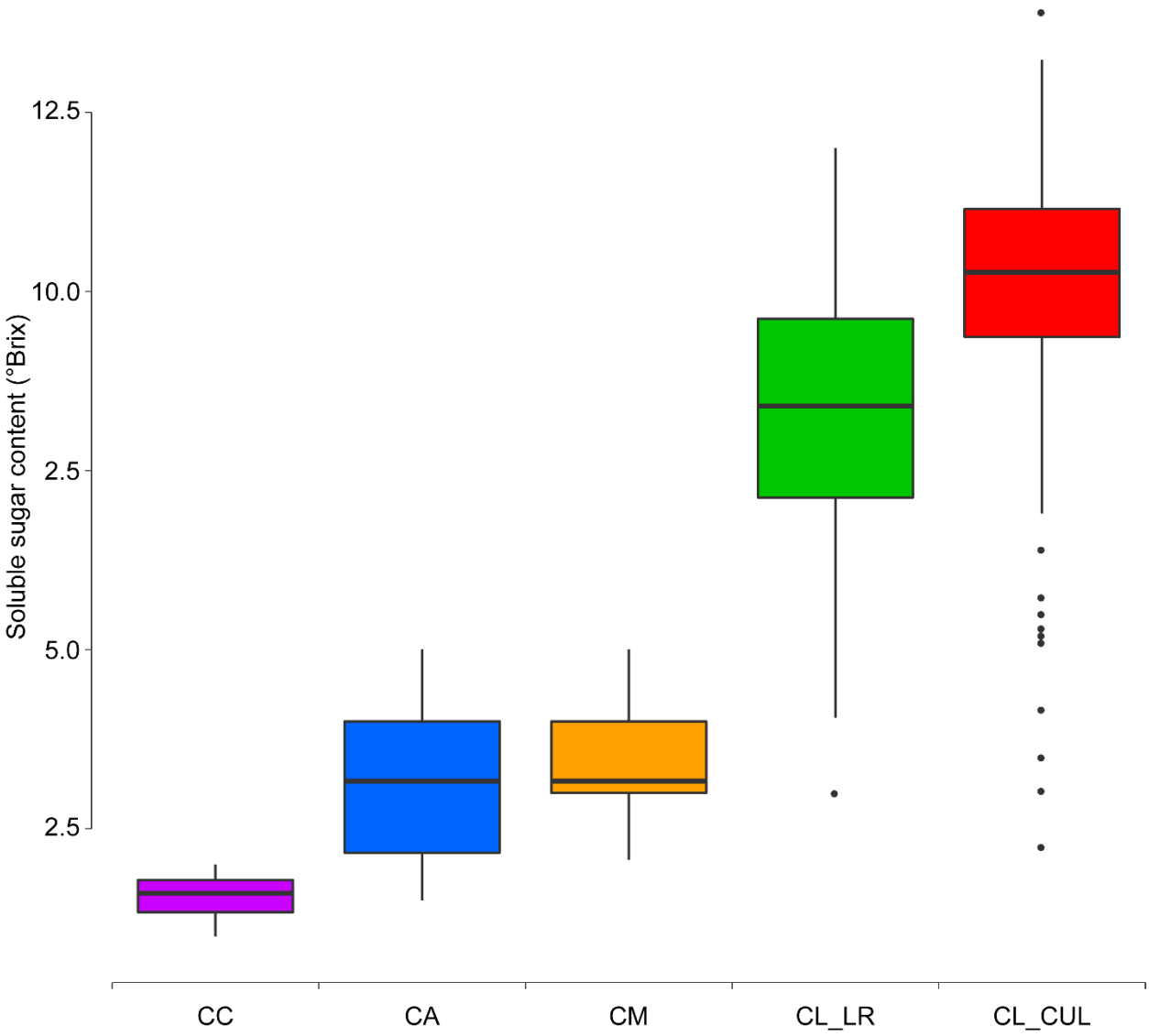

**Supplementary Fig. 6** Fruit flesh bitterness in different *Citrullus* species**.** CC, *C. colocynthis* (n = 9 accessions); CA, *C. amarus* (n = 25 accessions); CM, *C. mucosospermus* (n = 16 accessions); CL_LR, *C. lanatus* landrace (n = 84 accessions); CL_CUL, *C. lanatus* cultivar (n = 240 accessions).



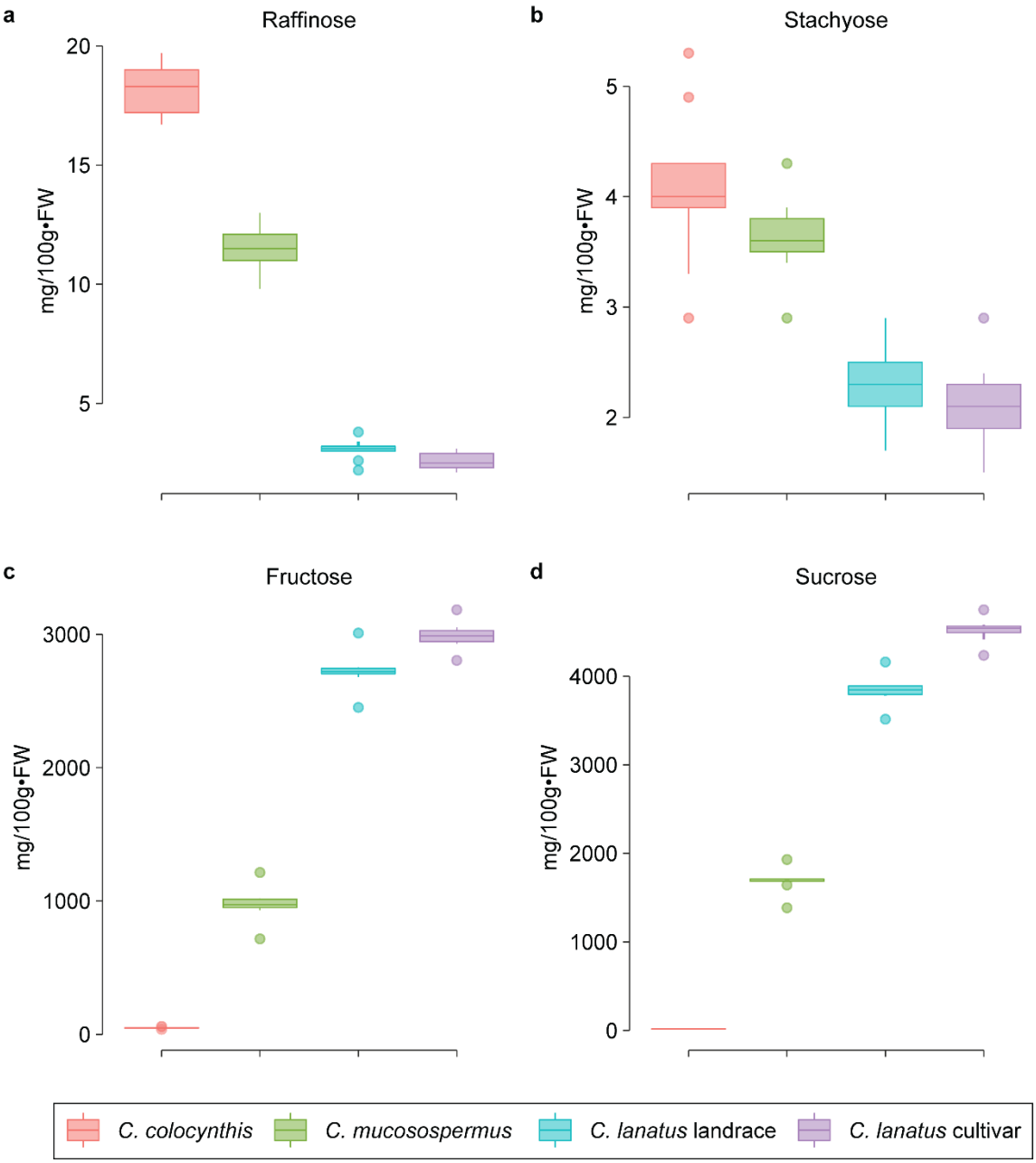

**Supplementary Fig. 7** Population differentiation between *C. lanatus* landraces and *C. mucosospermus* measured by $F_{ST}$. The loci underlying fruit flesh bitterness and coloration are highlighted. Candidate genes within the loci are pointed by arrows.



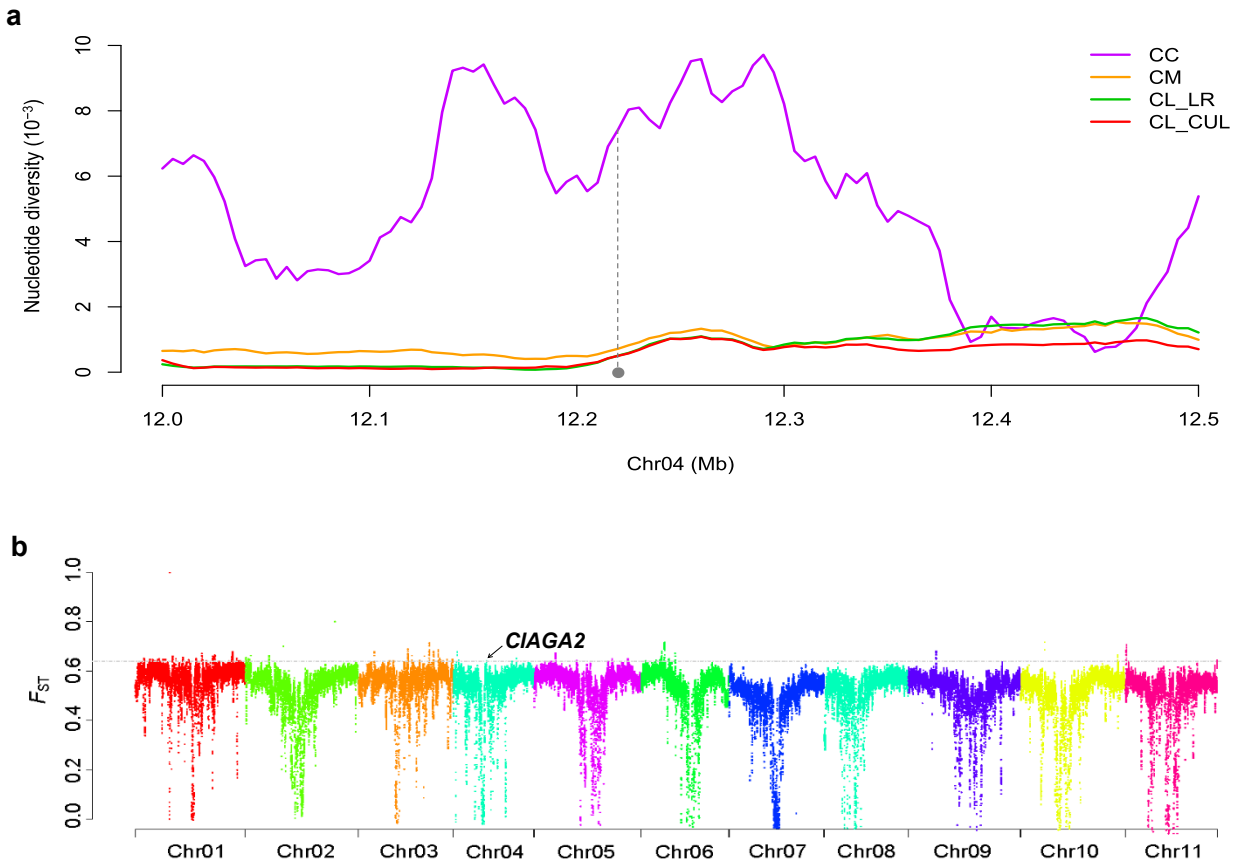

**Supplementary Fig. 8** Genetic diversities in *C. mucosospermus* (CM) and *C. lanatus* landraces (CL_LR) in genome regions surrounding the *ClBt* gene. The location of *ClBt* is indicated by the grey dot.



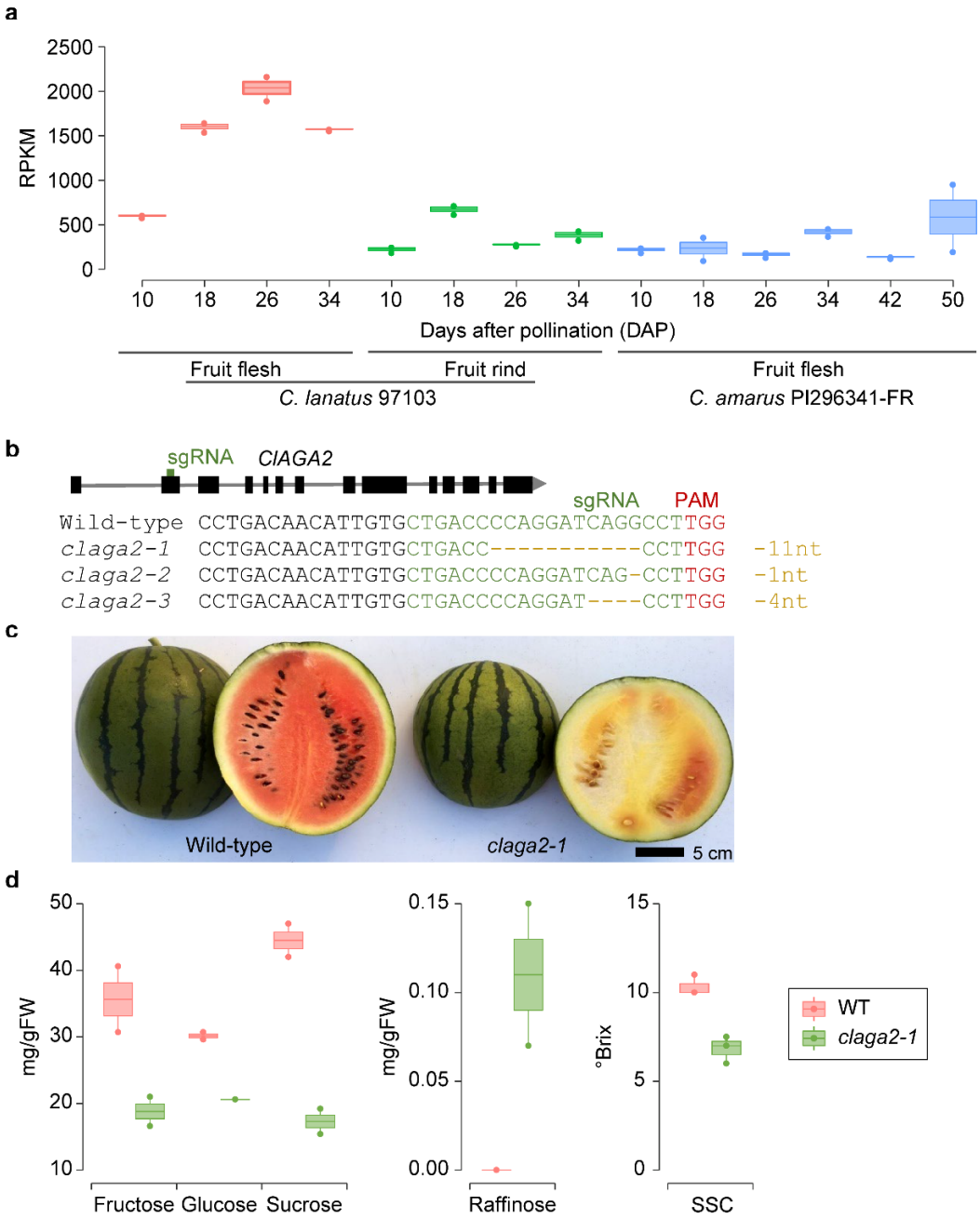

**Supplementary Fig. 9** Flesh sweetness (soluble solid content; ºBrix) in different *Citrullus* species. CC, *C. colocynthis* (n = 7 accessions); CA, *C. amarus* (n = 21 accessions); CM, *C. mucosospermus* (n = 16 accessions); CL_LR, *C. lanatus* landrace (n = 84 accessions); CL_CUL, *C. lanatus* cultivar (n = 218 accessions). For each boxplot, the lower and upper bounds of the box indicate the first and third quartiles, respectively, and the center line indicates the median.



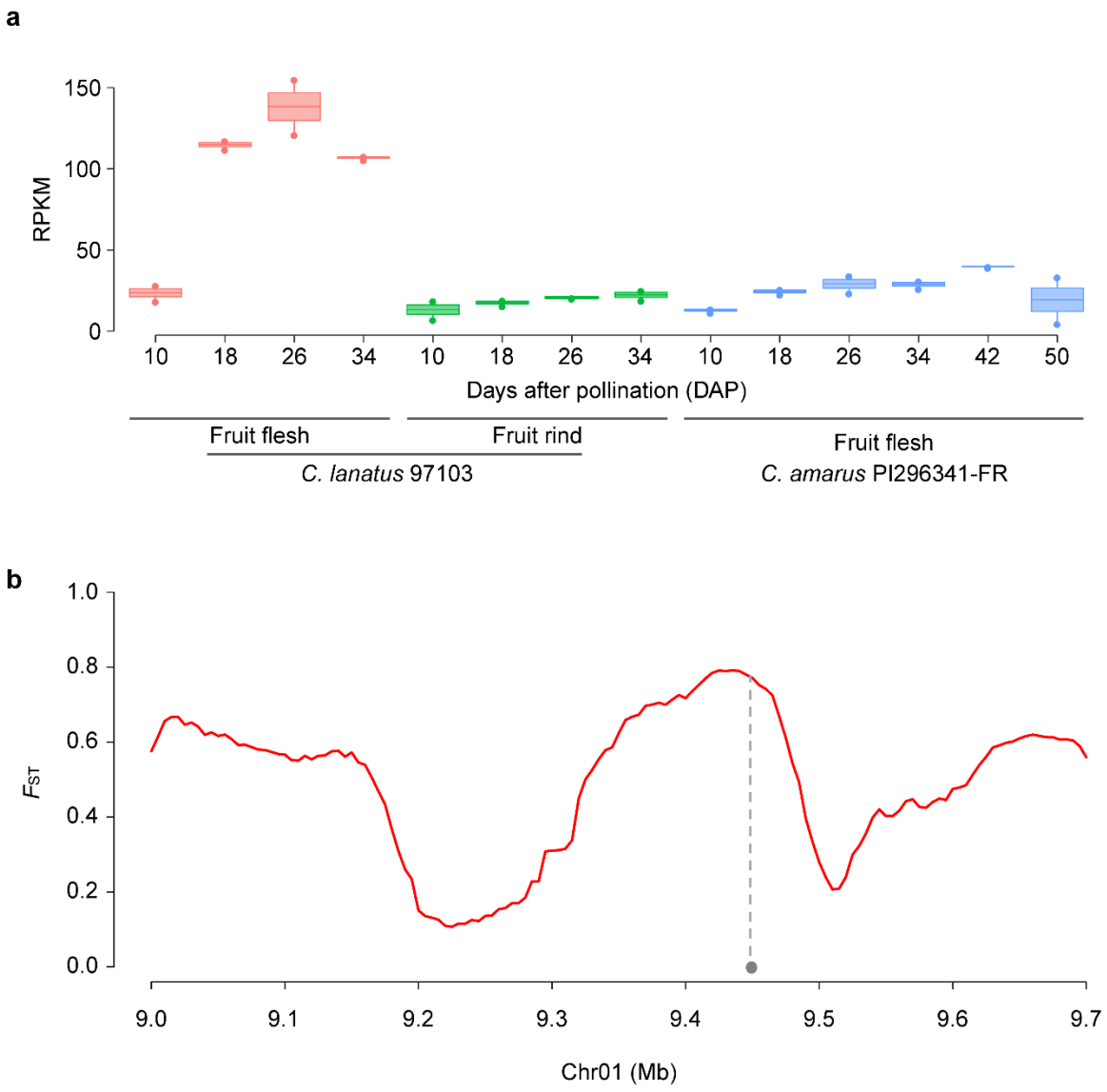

**Supplementary Fig. 10** Levels of raffinose (**a**), stachyose (**b**), fructose (**c**) and sucrose (**d**) in *C. colocynthis, C. mucosospermus* and *C. lanatus* landraces and cultivars. For each group, three accessions (*C. colocynthis*: PI 386019, PI 549161, PI 388770; *C. mucosospermus*: PI 595203, PI248178, PI 254740; *C. lanatus* landraces: PI 525084, PI 512404, HBJ; *C. lanatus* cultivars: 97103, Sugarlee, ZZJM), each with three independent experiments, were used to measure the sugar contents. For each boxplot, the lower and upper bounds of the box indicate the first and third quartiles, respectively, and the center line indicates the median.



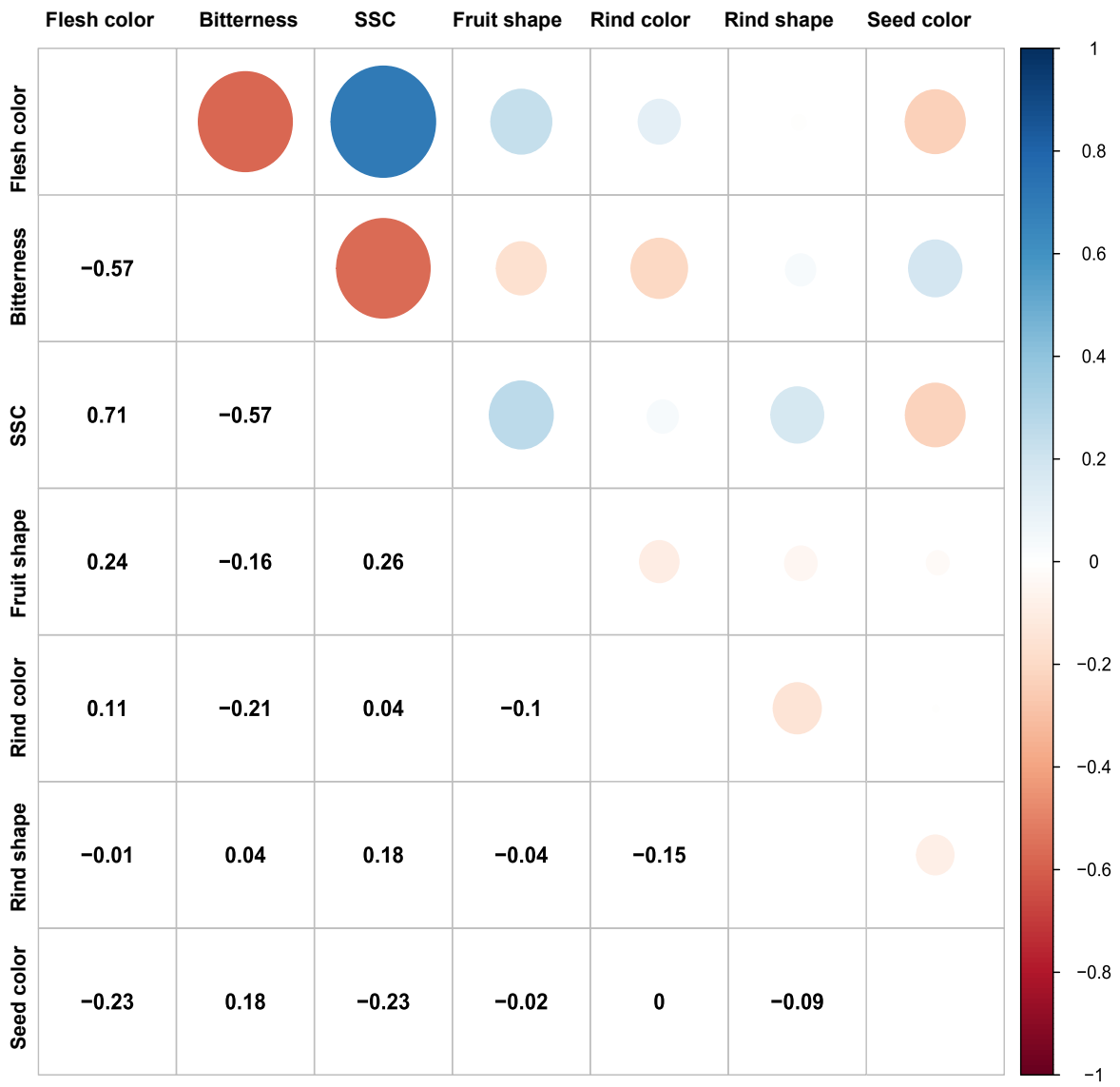

**Supplementary Fig. 11** Nucleotide diversities and population differentiation in the genome region of *ClAGA2*. **a,** Nucleotide diversities in the genome region surrounding *ClAGA2* in *C. colocynthis* (CC), *C. mucosospermus* (CM), *C. lanatus* landrace (CL_LR), and *C. lanatus* cultivar (CL_CUL). The location of *ClAGA2* is indicated by the grey dot. **b**, Population differentiation between *C. colocynthis* and *C. mucosospermus* measured by $F_{ST}$. The location of *ClAGA2* is indicated by the arrow.



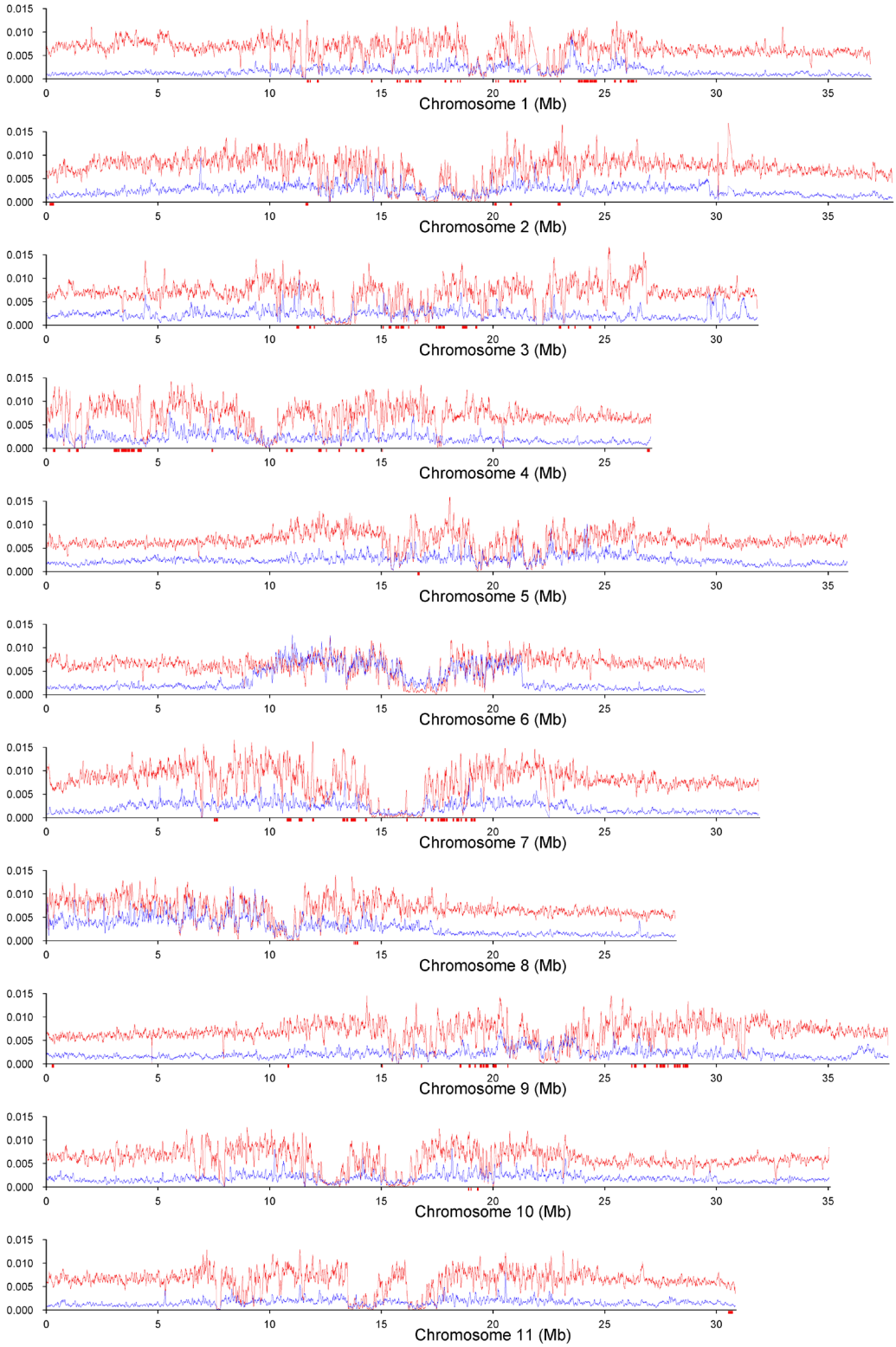

**Supplementary Fig. 12** Expression pattern of *ClAGA2* and characterization of the *ClAGA2*-knockout watermelon fruit. **a**, Expression levels of *ClAGA2* in *C. lanatus* 97103 and *C. amarus* PI296341-FR fruit tissues. n = 2 independent experiments. DAP, days after pollination. **b**, CRISPR/Cas9 induced indels at the *ClAGA2* guide RNA target sites in *claga2* mutant lines. sgRNA, single guide RNA. PAM, protospacer adjacent motif. **c**, Representative mature fruits of sweet watermelon 'ZZJM' plants in the *Cas9*-free T2 generation carrying the wild-type and *claga2-1* mutant allele (-11 nt) at *ClAGA2*. **d**, Glucose, fructose, sucrose, raffinose and soluble solid contents in the fruit of wild-type and *claga2-1* mutants. SSC, soluble solid contents. n = 2 independent experiments for glucose, fructose, sucrose, raffinose and 3 for SSC. For each boxplot, the lower and upper bounds of the box indicate the first and third quartiles, respectively, and the center line indicates the median.



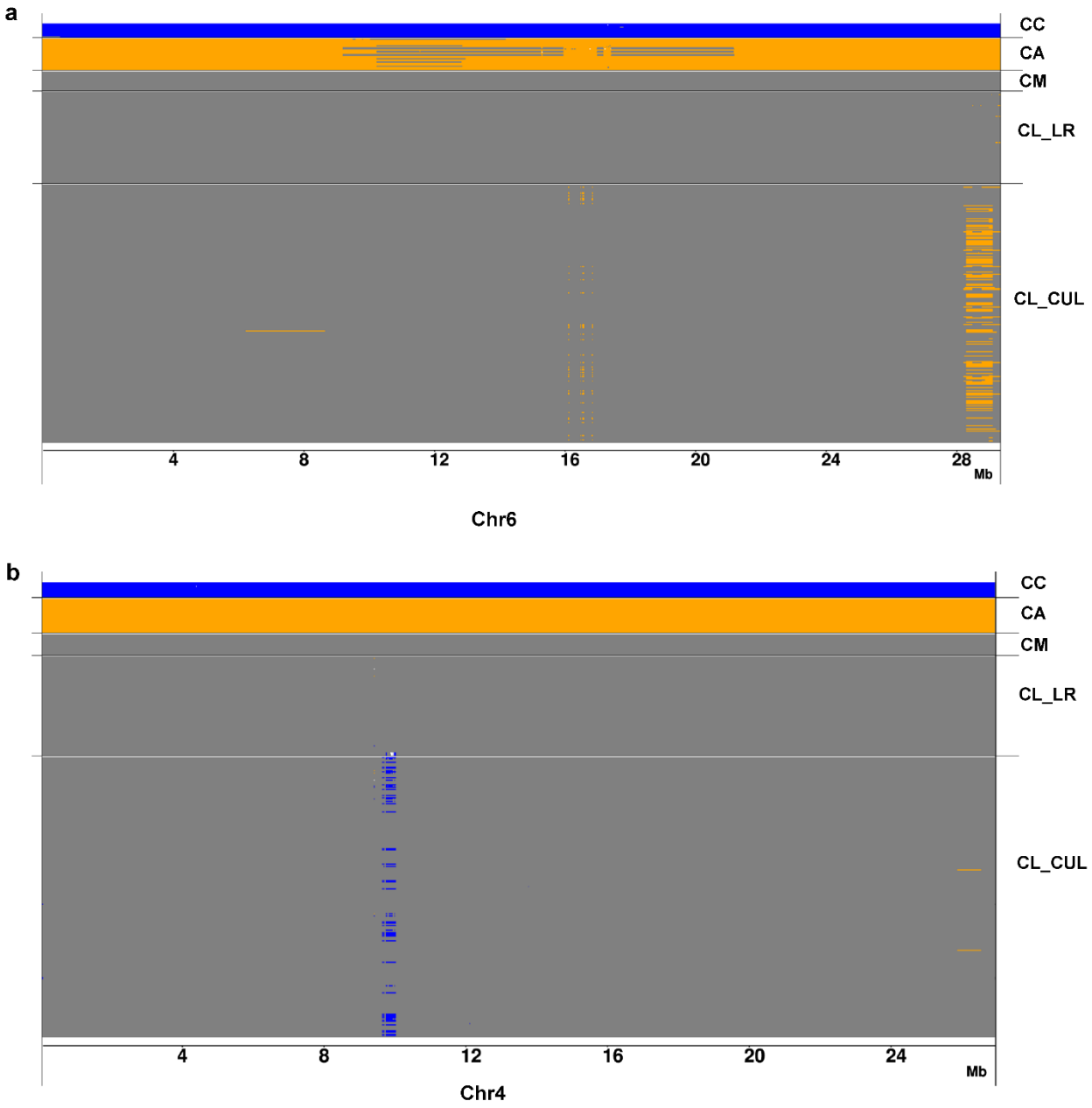

**Supplementary Fig. 13** Expression and population differentiation of the *PSY1* gene, *Cla97C01G008760*. **a**, Expression patterns of *PSY1* during fruit development. n = 2 independent experiments. For each boxplot, the lower and upper bounds of the box indicate the first and third quartiles, respectively, and the center line indicates the median. Values are mean of two biological replicates ± SE. DAP, days after pollination. **b**, Population differentiation measured by $F_{ST}$ in the genome region surrounding the *PSY1* gene (grey dot) between *C. mucosospermus* (CM) and *C. lanatus* landrace (CL_LR).




**Supplementary Fig. 14** Correlations among different watermelon fruit quality traits. Circle colors and sizes represent Pearson correlation coefficients, ranging from -1 (red) to +1 (blue).




**Supplementary Fig. 15** Distribution of nucleotide diversity (π) for *C. amarus* (blue) and *C. colocynthis* (red) across the 11 watermelon chromosomes. Red bars below the x-axis indicated candidate selective sweeps.




**Supplementary Fig. 16** Introgression among different *Citrullus* species. **a**, Introgression of a region on chromosome 6 from *C. amarus* (CA) to *C. lanatus* cultivar (CL_CUL). **b**, Introgression of a region on chromosome 4 from *C. colocynthis* (CC) to CL_CUL. Species-specific genome regions are shown in different colors. CM, *C. mucosospermus*; CL_LR, *C. lanatus* landrace.