nature genetics

ARTICLES

https://doi.org/10.1038/s41588-019-0522-8

A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits

Guangwei Zhao[1,20], Qun Lian[2,20], Zhonghua Zhang[3,4,20], Qiushi Fu[3,20], Yuhua He[1], Shuangwu Ma[1], Valentino Ruggieri[5,6], Antonio J. Monforte[7], Pingyong Wang[1], Irene Julca[8,9,10], Huaisong Wang[3], Junpu Liu[1], Yong Xu[11], Runze Wang[12], Jiabing Ji[2], Zhihong Xu[1], Weihu Kong[1], Yang Zhong[2], Jianli Shang[1], Lara Pereira[5,6], Jason Argyris[5,6], Jian Zhang[1], Carlos Mayobre[5,6], Marta Pujol[5,6], Elad Oren[13], Diandian Ou[1], Jiming Wang[1], Dexi Sun[1], Shengjie Zhao[1], Yingchun Zhu[1], Na Li[1], Nurit Katzir[13], Amit Gur[13], Catherine Dogimont[14], Hanno Schaefer[15], Wei Fan[2], Abdelhafid Bendahmane[14], Zhangjun Fei[16,17], Michel Pitrat[14], Toni Gabaldón[9,10,18], Tao Lin[2,19], Jordi Garcia-Mas[5,6]*, Yongyang Xu[1]* and Sanwen Huang[2]*

Melon is an economically important fruit crop that has been cultivated for thousands of years; however, the genetic basis and history of its domestication still remain largely unknown. Here we report a comprehensive map of the genomic variation in melon derived from the resequencing of 1,175 accessions, which represent the global diversity of the species. Our results suggest that three independent domestication events occurred in melon, two in India and one in Africa. We detected two independent sets of domestication sweeps, resulting in diverse characteristics of the two subspecies *melo* and *agrestis* during melon breeding. Genome-wide association studies for 16 agronomic traits identified 208 loci significantly associated with fruit mass, quality and morphological characters. This study sheds light on the domestication history of melon and provides a valuable resource for genomics-assisted breeding of this important crop.

Melon (*Cucumis melo* L.) is an important crop of the Cucurbitaceae family that is cultivated worldwide and of which more than 32 million tons was produced in 2017 (FAO; http://faostat.fao.org/). Melon was domesticated four thousand years ago[1]. As most wild *Cucumis* species emerged in Africa and have the same number of chromosomes as *C. melo*, it has been proposed that the center of origin of cultivated melon is Africa[2,3]. However, recent studies revealed that the closest wild relatives of melon are found in India and Australia[4,5]. Melon has been classified into two subspecies, *C. melo* ssp. *melo* (hereafter *melo*) and *C. melo* ssp. *agrestis* (hereafter *agrestis*), based on ovary pubescence[6]. Both domesticated subspecies have increased sizes of fruit, leaf and seed, and show loss of fruit bitterness and acidity compared with wild species. However, the fruit sizes of the two domesticated subspecies are highly different and whereas bitterness was fully lost in *melo*, it was only partially lost in *agrestis*. Therefore, the history and genetic basis of melon domestication still remain poorly understood. Current knowledge of melon domestication is largely derived from molecular marker analyses[7,8], and limited archeological[9] and historical data[10]. The availability of the melon genome sequence (454 Mb)[11] and a collection of melon germplasm resources enabled the rapid

[1]Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, China. [2]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. [3]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. [4]College of Horticulture, Qingdao Agricultural University, Qingdao, China. [5]Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Barcelona, Spain. [6]Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Barcelona, Spain. [7]Instituto de Biologia Molecular y Celular de Plantas, Universitat Politecnica de Valencia-Consejo Superior de Investigaciones Cientificas (CSIC-UPV), Valencia, Spain. [8]Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. [9]Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain. [10]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [11]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. [12]Centre of Pear Engineering Technology Research, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China. [13]Plant Science Institute, Israeli Agricultural Research Organization, NeweYa'ar Research Center, Ramat Yishay, Israel. [14]INRA, Génétique et Amélioration des Fruits et Légumes, Montfavet, France. [15]Department of Ecology and Ecosystem Management, Plant Biodiversity Research, Technical University of Munich, Freising, Germany. [16]Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY, USA. [17]US Department of Agriculture–Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA. [18]Institució Catalana de Recercai Estudis Avançats (ICREA), Barcelona, Spain. [19]College of Horticulture, China Agricultural University, Beijing, China. [20]These authors contributed equally: Guangwei Zhao, Qun Lian, Zhonghua Zhang, Qiushi Fu. *e-mail: Jordi.Garcia@irta.cat; xuyongyang@caas.cn; huangsanwen@caas.cn

detection genomic variations and insights into the trajectory of melon domestication on a genome-wide scale.

C. melo is a highly diversified species and a model system for studying several important biological processes[11], yet only limited number of genes related to agronomic traits like fruit monoecy[12], flesh color[13] and peel color[14] have been identified. Genome-wide association studies (GWAS) are a powerful approach for the identification of genes or quantitative trait loci (QTLs) that underlie complex traits, as has been demonstrated in rice[15], maize[16], foxtail millet[17], soybean[18], cotton[19], cucumber[20] and tomato[21,22]. Here we present the genome resequencing of 1,175 diverse accessions to characterize the population structure and domestication history of melon, and we provide genomic evidence for elucidating melon taxonomy. We also performed GWAS to identify a number of candidate genes and loci that underlie several important agricultural traits.

Results

Genome variation map of melon. We used 1,175 diverse accessions of *C. melo*, which included 667 from *C. melo* ssp. *melo* and 508 from *C. melo* ssp. *agrestis*, and an additional 9 accessions from closely related species (Fig. 1a and Supplementary Table 1). The *C. melo* collection consisted of 134 wild and 1,041 cultivated accessions spanning most of the native range of the species. We generated a total of 4.29 trillion base pairs (bp) of sequence using next-generation sequencing technology, with a median depth of 4.98-fold and coverage of 80.73% of the assembled melon genome[23] (release 3.5.1). After aligning the reads to the melon reference genome[23], we identified a total of 5,678,165 SNPs and 957,421 small insertions or deletions (indels; ≤5 bp), with an average of 13.99 SNPs and 2.36 indels per kb (Extended Data Fig. 1, Supplementary Tables 2 and 3) that is similar to cucumber (17.22 SNPs and 1.75 indels per kb)[24]. The accuracy of the identified SNPs was estimated to be 99.07% when comparing 10 pairs of accessions sequenced with low depth (4.71×) and high depth (18.92×; Supplementary Table 4). A total of 197,113 SNPs (3.47%) and 10,114 (1.06%) indels were located in the coding regions, among which 13,022 showed potentially large effects: 7,030 SNPs (0.12%) affected 5,598 genes by causing start codon changes, premature stop codons or elongated transcripts, and 5,992 (0.63%) indels led to a frameshift in 4,587 annotated genes (Supplementary Tables 2 and 3). Collectively, this comprehensive dataset of the genome variation of melon provides a resource for melon biology and breeding.

Population structure. The phylogenetic relationships for these melon accessions were inferred using a subset of 17,055 SNPs at fourfold degenerate sites. To better deduce the relationships of melon accessions from different areas, 207 with uncertain origin were excluded from the next analysis. The phylogenetic tree based on the nuclear and chloroplast genome SNPs (Fig. 1b and Extended Data Fig. 2) supported three distinct clades, which exhibited strong geographic separation and distinctive botanical features. We found that only the primitive African domesticated types were clustered with wild African accessions in clade I, suggesting the marginal impact of wild African accessions during melon domestication outside of Africa, consistent with a previous study[5]. The remaining accessions in clade II and clade III corresponded to *melo* and *agrestis* ssp. according to passport information and morphological characteristics. A similar result was obtained by DAPC analysis (Extended Data Fig. 3).

Model-based clustering and principal component analyses further classified clade II and clade III each into two main subclades (Fig. 1c and Extended Data Fig. 4). The majority of *C. melo* var. *momordica* (*momordica*) accessions, which are native to India and traditionally considered to be a cultivated *agrestis* (CA), formed a single subclade (clade II-1) with an obvious admixture in genetic composition (Fig. 1c), and clustered closely together with cultivated *melo* (CM) accessions (clade II-2). Furthermore, clade II-1 generally showed similar characteristics to wild melon, such as monoecy, low sugar content, acid flesh and high resistance to pests and disease[25]. Thus, our data suggested that cultivated *melo* was domesticated directly from *momordica* (Fig. 1b). Clade III-1 consisted of wild (*C. melo* L. var. *agrestis*; WA) and cultivated *agrestis* accessions. In this clade, the cultivated *agrestis* accessions from southern Africa unexpectedly clustered with wild *agrestis* melon derived from India. These accessions share similarities to wild melon from India with respect to bearing small fruits, monoecy and having a gelatinous sheath around the seeds. Based on both genetic and trait similarities, South African *agrestis* accessions could represent recent migrants from India. Both morphological and genomic data largely support that the cultivated *melo* (clade II-2) and *agrestis* (clade III-2) accessions were domesticated from clade II-1 (wild *melo*; WM) and clade III-1 (wild *agrestis*), respectively. Finally, we can speculate that there were three independent domestication events leading to the three main clusters: two of which occurred in India and another in Africa, consistent with a recent study[5]. Owing to the relatively few wild African melon accessions in our study and their low influence during melon domestication, we used the melon accessions from clade II and clade III for further analyses.

Nucleotide diversity measured by the π value[26] for the wild *melo* ($\pi = 0.00347$) and wild *agrestis* ($\pi = 0.0031$) groups was substantially higher than that for the cultivated *melo* ($\pi = 0.00256$) and cultivated *agrestis* ($\pi = 0.00074$) groups, which is consistent with the result of Watterson estimator analysis (θ_W(WM) = 0.00256, θ_W(WA) = 0.00218, θ_W(CM) = 0.00151 and θ_W(CA) = 0.00069). Furthermore, different groups exhibited different degrees of heterozygosity (Extended Data Fig. 5) and we also observed gene flow between these groups, which may be due to the overlapping distribution of accessions, open pollination and modern breeding programs (Extended Data Fig. 6). The decay of linkage disequilibrium (LD) with physical distance between SNPs to half of the maximum values occurred at 22.0 kb and 20.6 kb in the wild *melo* ($r^2 = 0.17$) and wild *agrestis* ($r^2 = 0.14$) groups, respectively, which were considerably smaller than the values in the cultivated *melo* (58.0 kb, $r^2 = 0.31$) and cultivated *agrestis* (610.2 kb, $r^2 = 0.43$) groups (Fig. 1d). Taken together, these results are strongly suggestive of a significant reduction in genetic diversity in cultivated melons because of domestication. Notably, the cultivated *agrestis* group has much lower nucleotide diversity and higher LD decay than the cultivated *melo* group, suggesting that the cultivated *agrestis* group has undergone a more-severe bottleneck during domestication.

Independent domestication of *melo* and *agrestis* subspecies. Since the Neolithic revolution and the development of agriculture, people have preferably kept and propagated seeds from wild plants with larger and more delicious fruits. To identify potential selective signals during melon domestication, we scanned genomic regions that showed marked reductions in nucleotide diversity by comparing each cultivated group with its corresponding wild group (π_{WM}/π_{CM} and π_{WA}/π_{CA}) over 50-kb windows. We identified 148 and 185 putative selection sweeps associated with domestication in *melo* ($\pi_{WM}/\pi_{CM} \geq 3.49$) and *agrestis* ($\pi_{WA}/\pi_{CA} \geq 26.18$), respectively, covering 6.28% (25.52 Mb) and 7.23% (29.39 Mb) of the assembled genome and comprising 1,481 and 1,710 genes, respectively (Supplementary Tables 5–8). Notably, only 143 out of 27,427 genes (2.67 Mb; 0.66% of the assembled genome) were shared between the *melo* and *agrestis* sweeps (Supplementary Table 9). By contrast, in rice, most well-characterized domestication genes were shared between the *indica* and *japonica* types[27]. Jointly, the *melo* and *agrestis* sweeps cover 52.24 Mb (12.86% of the assembled genome), which encompasses 3,048 genes.

To discover potential domestication loci, we developed two F_2 segregating populations derived from a cross between a wild *melo*

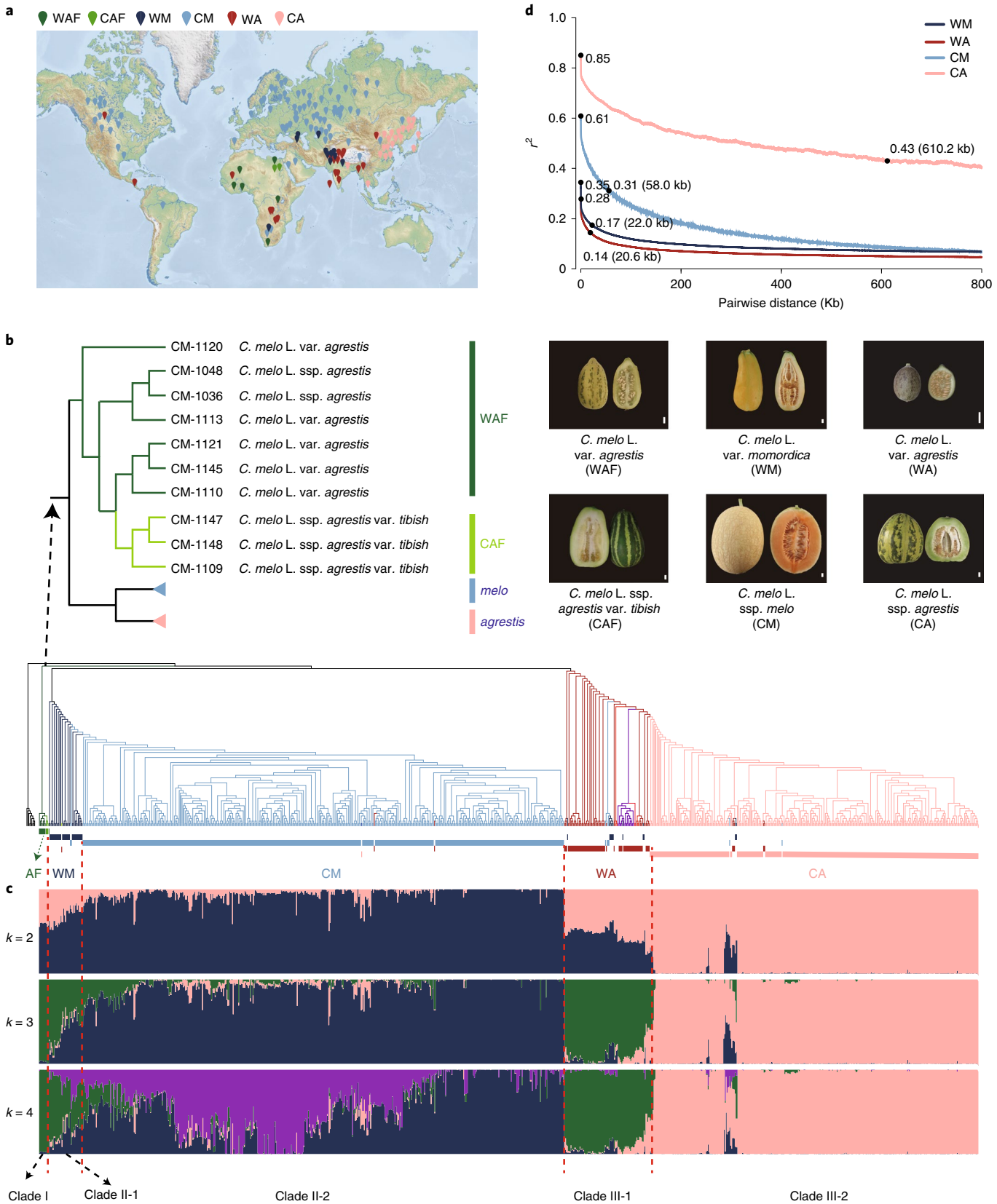


Fig. 1 | Geographic distribution and population structure of melon accessions. a, Geographic distribution of melon accessions, which are represented by dots on the world map (the map was downloaded from Mapswire, https://mapswire.com/world/physical-maps/). **b**, Phylogenetic tree of the population (531 *melo*, 437 *agrestis* accessions and 9 wild relatives as the outgroup) constructed using 17,055 SNPs at fourfold degenerate sites. AF, African; CAF, cultivated African; WAF, wild African. Representative fruits of the three clades studied are shown. Scale bars, 1.0 cm. **c**, Model-based clustering analysis with different numbers of clusters ($K = 2$, 3 and 4). The *y* axis quantifies clusters membership, and the *x* axis lists the different accessions. The orders and positions of these accessions on the *x* axis are consistent with those in the phylogenetic tree. **d**, Genome-wide average LD decay estimated from different melon groups.

(MS-542) and a cultivated *melo* (additional accession B460), and a cross between a wild *agrestis* (additional accession yesheng) and a cultivated *agrestis* (MS-79), and identified 19 new QTLs that overlap with sweep regions, including 10 in *melo* and 9 in *agrestis* (Fig. 2a,b and Supplementary Table 10). Among the QTLs, one genomic region (around 2.95–4.77 Mb on chromosome 8) related to fruit mass (*fwqaz8.1*, *fdqaz8.1* and *ftqaz8.1* for fruit weight, fruit diameter and flesh thickness, respectively; Fig. 2c), a trait that has been under human selection was detected using the above F_2 segregating population from a cross between a wild *agrestis* and a cultivated *agrestis* accession. This region is consistent with a previously reported QTL (*fwqc8.1*) that contributes negatively to the increase in fruit weight in melon[28]. The nucleotide diversity of this interval was markedly reduced in the cultivated *agrestis* group compared to wild *agrestis* ($\pi_{WA}/\pi_{CA}=7.45$), but the reduction was only minor in this region in cultivated *melo* ($\pi_{WM}/\pi_{CM}=1.45$; Fig. 2d). This region included two genes, *MELO3C007596* and *MELO3C007597*, both of which encode auxin-responsive GH3-like proteins. Auxin-responsive GH3-like proteins are reported to be involved in fruit growth and development in longan[29] and tomato[30], suggesting both of the two genes are logical candidates to be associated with fruit mass during *agrestis* domestication. Moreover, we found that *Cm-HMGR* (*MELO3C026512*) on chromosome 3, which encodes a hydroxymethylglutaryl coenzyme A reductase (HMGR) in the mevalonate pathway that is involved in controlling fruit size in melon[31], was located in a *melo* domestication sweep ($\pi_{WM}/\pi_{CM}=4.07$; Fig. 2e).

Bitterness is an essential domestication trait, and has been (partially) lost during the domestication of melon. We observed that 95.18% of cultivated *melo* accessions carried non-bitter young fruits, whereas 25.87% of young fruits from the cultivated *agrestis* accessions were bitter, especially the fruits exposed to stress conditions. This suggests that the two melon groups possess different domestication mechanisms conferring the loss of bitterness. We detected a notable decrease in nucleotide diversity in the *Bi* (encoding a cucurbitadienol synthase) cluster ($\pi_{WM}/\pi_{CM}=3.61$ and $\pi_{WA}/\pi_{CA}=1.75$) in *melo* and in the *CmBt* locus (encoding a transcription factor that activates *CmBi* transcription)[32] in *agrestis* ($\pi_{WA}/\pi_{CA}=28.21$ and $\pi_{WM}/\pi_{CM}=3.70$; Fig. 2a,b). Additionally, there is an obvious population differentiation in *CmBi* and *CmBt* between the cultivated *melo* and cultivated *agrestis* groups (Fig. 2f,g).

To further verify potential sweeps related to bitterness during melon domestication, we performed QTL mapping using the above two F_2 populations (WM × CM and WA × CA). Two QTLs (around 29.04–29.93 Mb and about 21.07–22.05 Mb on chromosome 11 and 9, respectively) associated with fruit bitterness were identified in *melo* and *agrestis*, harboring the *CmBi*[32] and the *CmBt* gene[32], respectively (Fig. 2f,g). In general, alleles from different bitterness-related genes have been domesticated in the two melon subspecies. This was further validated by assessment of gene expression (Fig. 2h,i and Extended Data Fig. 7) and cucurbitacin B content (Supplementary Table 11). Furthermore, we found that hybrids (F_1) from a cross between non-bitter cultivated *melo* (accession MS-251) and cultivated *agrestis* (accession MS-42) lines always had bitter young fruits, suggesting that complementary genes exist conferring bitterness, as previously hypothesized[33,34].

Acidity as a major component of taste is selected by farmers and breeders during melon breeding. The acidic genotype is presumably the ancestral form[35], and the acidity in domesticated *melo* and *agrestis* also seems to be related to different genes. For example, we found that a domestication sweep ($\pi_{WA}/\pi_{CA}=1.94$ and $\pi_{WM}/\pi_{CM}=9.99$) containing the *CmPH* gene (*MELO3C025264*)[35] on chromosome 8 occurred in the *melo* but not in the *agrestis* group (Fig. 2a,b). The *CmPH* gene, encoding a transmembrane transporter, determines fruit acidity based on the presence of an insertion of a four amino acid duplication in non-acidic melon accessions, and contributes to the evolution of sweet melons[35]. We further verified the causative variation and found that the insertion occurred in non-acidic melon accessions of the cultivated *melo* group, but not all accessions were consistent in the cultivated *agrestis* group, suggesting that other genes might contribute to the acidity in the cultivated *agrestis* group. The *CmPH* gene was also detected within the association signals in a GWAS analysis for flesh acidity in *melo* accessions (Fig. 2j), and is located in a *melo* sweep (Fig. 2a). Intriguingly, *MELO3C011482* on chromosome 3, which encodes ATP-citrate synthase subunit 1, was located in an association signal (Fig. 2k) and a sweep of *agrestis* (Fig. 2b).

In summary, our results suggest that distinct domestication mechanisms for fruit mass, flesh bitterness and acidity occurred in *melo* and *agrestis* accessions, which further supports the hypothesis that the cultivated *melo* and cultivated *agrestis* groups were domesticated independently from the wild *melo* and wild *agrestis* groups, respectively.

Divergence between the *melo* and *agrestis* groups. Melon is a diverse species cultivated by local farmers and used by breeders in many countries. Geographically, *melo* is cultivated worldwide, whereas *agrestis* is concentrated in east Asia. In general, cultivated *agrestis* melons possess greener leaves, edible epicarp, thinner flesh, lower sugar content and ecological differences that result from their distinct geographic distributions. To elucidate the genomic regions that underlie these differences, we measured the pairwise genome-wide fixation index (F_{ST}) values based on SNPs between different melon groups. The average F_{ST} value between the cultivated *melo* and cultivated *agrestis* groups was estimated to be 0.46, which was similar to that of *indica* and *japonica* rice (0.55)[15], indicating that there is strong population differentiation in the two subspecies. Based on F_{ST}, 289 divergent genomic regions between *melo* and *agrestis* were identified, which covered 56.9 Mb (14.01%) of the assembled genome and comprised 3,535 predicted genes (Supplementary Tables 12 and 13).

Seventeen previously reported QTLs for flesh thickness and sugar content overlapped with these divergent genomic regions (Extended Data Fig. 8a and Supplementary Table 14). We also identified two new QTLs (Extended Data Fig. 8b,c) for flesh thickness on chromosome 4 and 5 using an F_2 population from a cross between a cultivated *melo* (additional accession Y14) and a cultivated *agrestis* line (MS-1006), both of which were close to divergence regions. Notably, one GWAS signal on chromosome 4 for ovary pubescence, a trait used to distinguish *melo* and *agrestis*[6], was also located in a divergence region (Extended Data Fig. 8d). Furthermore, five xyloglucan endo-transglycosylase genes involved in plant growth[36,37] were located in a divergence region (around 2.15–2.28 Mb) of chromosome 5. This dataset constitutes a relevant resource for the exploitation of genes conferring genetic differentiation between *melo* and *agrestis*.

Identification of genes or loci related to important agronomic traits. Melon has several diverse characteristics of agronomic importance, such as sex determination, peel and flesh color, and fruit shape. However, few genes or loci underlying agronomic traits have been identified in melon so far. To explore the potential of GWAS to identify causal genes for complex traits, we performed an association study using a panel composed of 1,067 diverse accessions for 16 agronomic traits (Supplementary Table 15). A total of 208 significant association signals were identified in the melon genome. Four previously dissected genes were found in these association signals, including *CmACS-7* for monoecy[12], *CmOr* for flesh color[13], *CmKFB* for peel color[14] and *CmPH* for acidity[35] (Fig. 2 and Extended Data Fig. 9). The remaining 204 signals were associated with yield (76), fruit quality (29) and morphological (99) traits (Extended Data Fig. 10 and Supplementary Figs. 1–10). We further validated major GWAS signals for rind sutures, peel color and flesh color

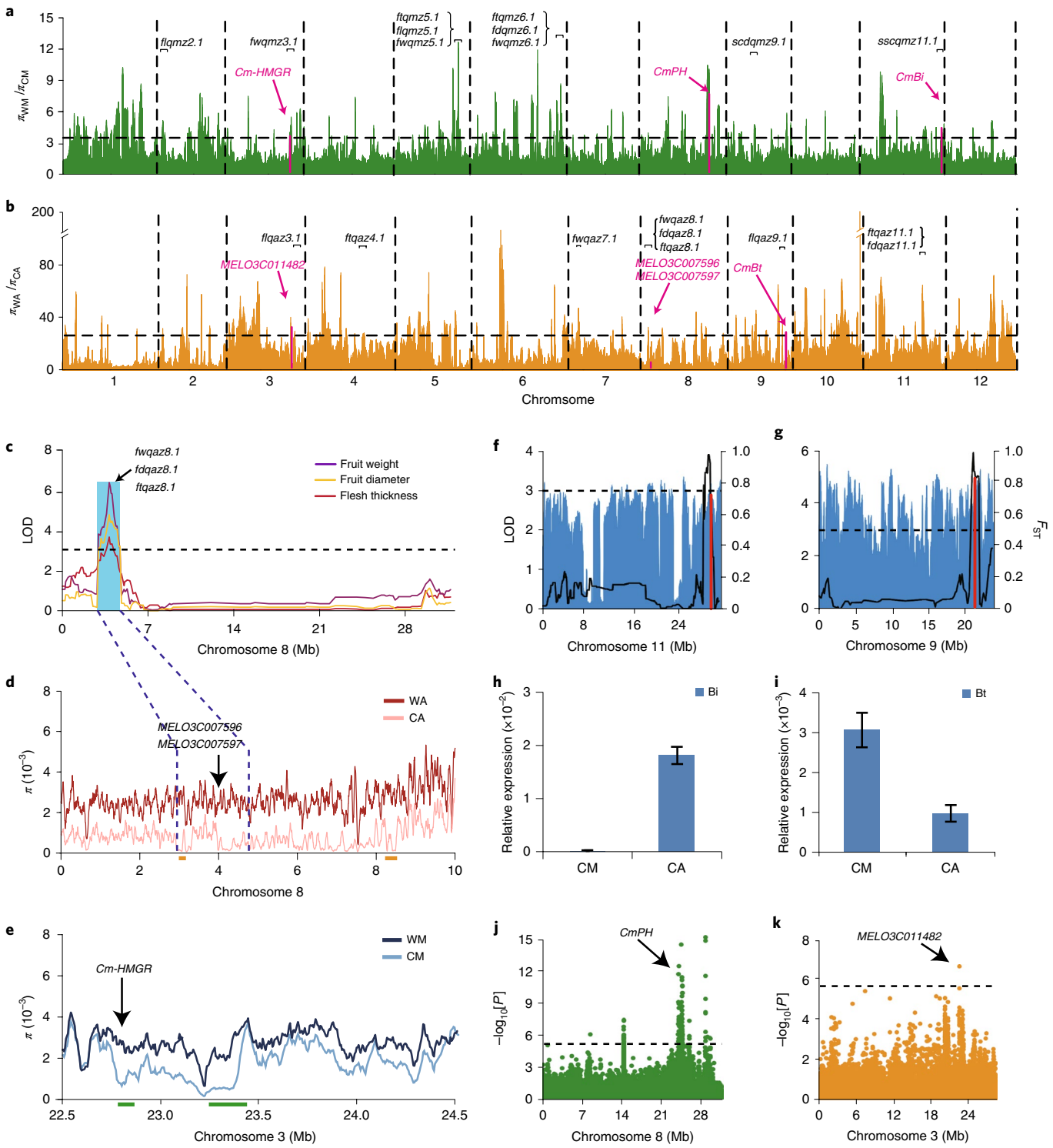


Fig. 2 | Independent selection in domesticated traits between *C. melo*. ssp. *melo* and *agrestis*. **a**,**b**, Selection signals in domestication of *melo* (**a**) and *agrestis* (**b**) populations on the 12 melon chromosomes. Horizontal dashed lines indicate the genome-wide threshold of selection signals. Candidate genes that have previously been reported or identified in this study (red) and QTLs (black) that overlapped with selective sweeps are marked. **c**, Overlapping genomic regions of QTLs for fruit weight, fruit diameter and flesh thickness mapped by genetic analysis of an F_2 population from the cross of a wild and a cultivated (MS-79) *agrestis* accession. **d**,**e**, Distribution of nucleotide diversity (π) of wild *agrestis* and cultivated *agrestis* (**d**), and wild *melo* and cultivated *melo* (**e**) accessions. *MELO3C007596*, *MELO3C007597* and *Cm-HMGR*[31] were located within the *agrestis* and *melo* sweep regions, respectively. **f**,**g**, QTL mapping for young fruit bitterness using two F_2 populations from the cross between wild *melo* (**f**) and *agrestis* (**g**) with their corresponding cultivated accessions. F_{ST} values of *CmBi*[32] and *CmBt*[32] genes between cultivated *melo* and cultivated *agrestis* are shown in the red vertical line. **h**,**i**, Quantitative PCR with reverse transcription (RT-qPCR) of *CmBi* (**h**) and *CmBt* (**i**) in young fruits of cultivated *melo* and *agrestis* accessions. Data are mean ± s.d. ($n = 3$ independent experiments). **j**,**k**, Manhattan plots of GWAS for flesh acidity in *melo* (**j**) and *agrestis* (**k**) populations. *CmPH*[35] and *MELO3C011482* were identified residing within the association signals on chromosomes 8 and 3, respectively.

using segregating populations and molecular biology approaches (Figs. 3, 4 and Supplementary Fig. 11).

Rind sutures (also called vein tracks) are an important trait commonly found in commercial melons, which is controlled by a single gene[38] on chromosome 11. A strong GWAS signal for rind sutures was identified ($P = 2.14 \times 10^{-68}$; around 20.6–24.8 Mb) on chromosome 11 (Fig. 3a). To further validate this signal, we constructed a recombinant inbred line (RIL) population[39] obtained by crossing a sutured line (MS-1152, Vedrantais, VED) with a non-sutured line (Piel de sapo T111, PS) and narrowed down the QTL to a 1.7-Mb (about 22.8–24.5 Mb) interval (Fig. 3b). We further delimited this interval to an approximately 86-kb (around 23.17–23.25 Mb) region using additional F_2 and RIL populations. The region contains four putative protein-coding genes, of which two are expressed in flower and fruit tissues (Fig. 3c). We found that *MELO3C019694*, encoding an AGAMOUS MADS-box transcription factor, resides 16.8 kb upstream of the strongest association signal. The orthologs of this gene include *SHP1* (*AT3G58780*)/*SHP2* (*AT2G42830*), which regulate pod dehiscence in *Arabidopsis*[40], and *TAGL1* (*Solyc07g055920.2.1*), which is required for pericarp expansion and climacteric ripening in tomato[41] (Fig. 3d).

We further analyzed the upstream and downstream sequences of *MELO3C019694* to identify structural variations and discovered a 1,070-bp deletion at 23.85 kb upstream of *MELO3C019694* that was present in most sutured accessions (83.66% of 257 accessions; Fig. 3e,f). The expression level of *MELO3C019694* is delayed until 7 days after pollination in the sutured line compared to the non-sutured line, which is the initial stage of suture development (Fig. 3g). These results indicate that *MELO3C019694* could be a candidate gene for sutures, and the 1,070-bp deletion might impair the transcriptional regulation of *MELO3C019694*. However, the mechanism and causal variation of *MELO3C019694* need to be further validated functionally.

Peel and flesh color are important fruit quality traits that influence the choice of the consumer and acceptability of the melon. Peel colors of commercial melons are green, white or yellow (Fig. 4a), which are conferred by distinct pigment accumulation[42]. We identified a 12:3:1 segregation ratio for green, white and yellow traits by analyzing an F_2 segregating population from a cross between green peel (MS-723) and yellow peel (B432, an additional accession) lines (Fig. 4b), indicating that green peel is a dominant epistatic to non-green (white and yellow) peel. We selected 254 green and 381 non-green accessions (145 white and 236 yellow accessions) in a GWAS analysis (Supplementary Dataset 1) and identified two strong association signals on chromosome 4 ($-\log_{10}[P] = 14.17$) and chromosome 8 ($-\log_{10}[P] = 10.78$; Fig. 4c). Moreover, we performed a GWAS analysis using the non-green accessions (145 white and 236 yellow accessions). One significant peak ($-\log_{10}[P] = 20.80$) was detected on chromosome 10 (Fig. 4d).

To further verify these signals related to peel color, we sequenced three bulk populations with green, white and yellow peel from the above F_2 population (MS-723 × B432). We computed the differences in SNP indices (ΔSNP index)[43] between the green and non-green peel bulk populations, and between the white and yellow peel bulk populations, respectively, and identified a single overlapping genomic region with those of GWAS on chromosome 4 (Fig. 4c) and chromosome 10 (Fig. 4d), respectively. The overlapping genomic regions were also identified in another F_2 segregating population obtained by crossing a green peel line with a yellow peel line (Fig. 4e–g). Within these genomic intervals (around 0.30–0.80 Mb on chromosome 4 and about 3.41–3.50 Mb on chromosome 10), we detected two candidate genes related to peel color, *MELO3C003375* on chromosome 4, which encodes a two-component response regulator-like protein APRR2, and the already identified *CmKFB* gene on chromosome 10, which negatively regulates the accumulation of naringenin chalcone[14]. The orthologous genes of *MELO3C003375* in cucumber (*w*)[44], watermelon (*ClCG09G012330*)[45], pepper (GeneBank accession no. KC175445)[46] and tomato (*SolyC08g077230*)[46] have been demonstrated to control chlorophyll metabolism and pigment accumulation in fruit peel.

MELO3C003375 exhibited much higher transcript levels in green peel lines than white and yellow peel lines (Fig. 4h). There is hardly any expression of *CmKFB* in the yellow peel accession (Fig. 4i), consistent with its function of negatively regulating flavonoid accumulation. Additionally, we detected a gene (*MELO3C003097*) in the genomic interval (around 29.74–29.77 Mb) of the association signal on chromosome 8 that is an ortholog of the *Arabidopsis SG1*, which encodes the protein SLOW GREEN 1, is required for chloroplast development[47] and is expressed in accessions of every peel color (Fig. 4j). We speculate that *MELO3C003375* and *CmKFB* are associated with the green and yellow peel trait, respectively; *MELO3C003097* could be a minor gene involved in the formation of peel color.

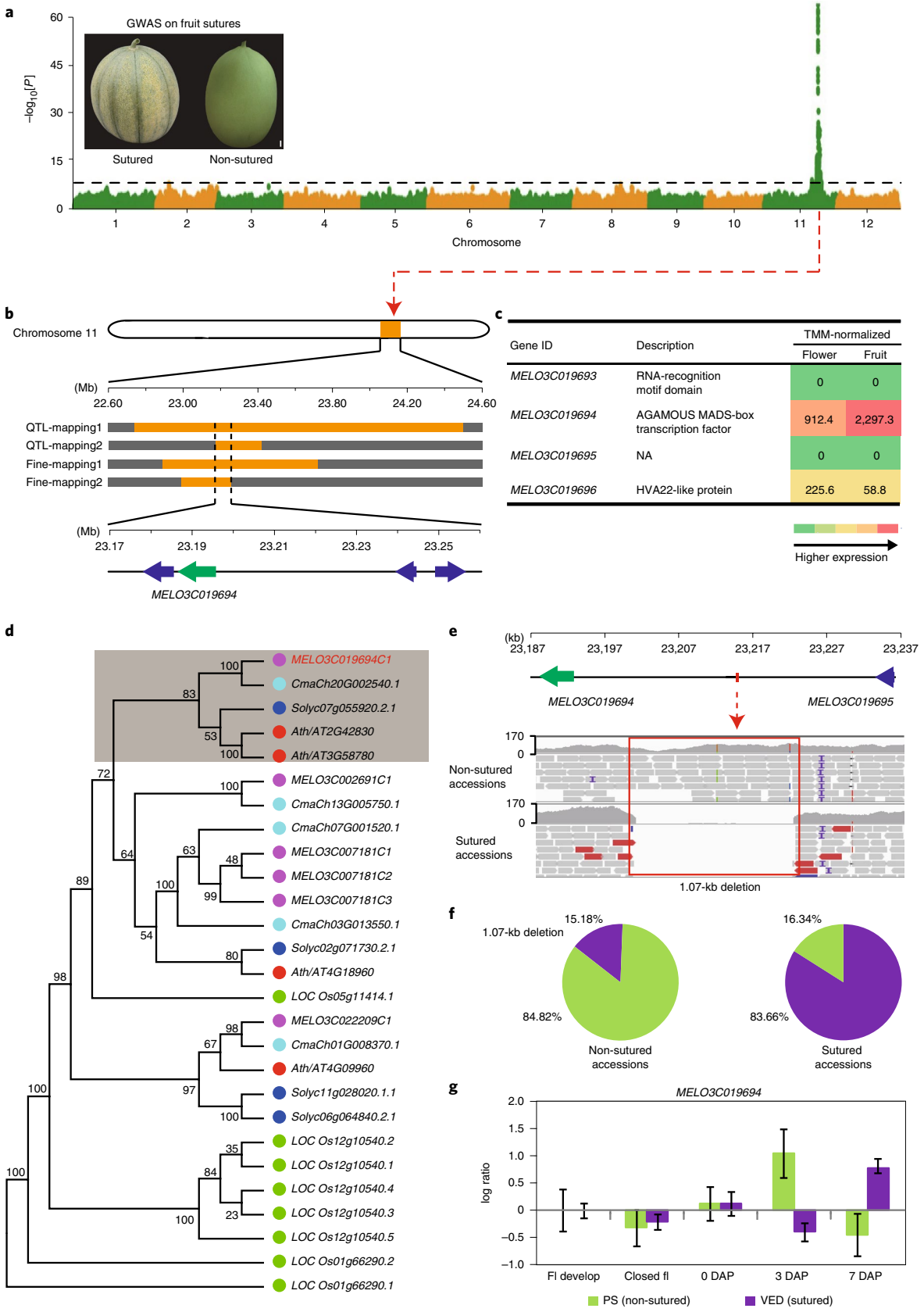
Moreover, we performed GWAS on flesh color using 688 melon accessions. In addition to the identified *Gf* gene (*CmOr*)[13] controlling orange flesh, we detected a strong association signal on chromosome 8 with a highest $-\log[P]$ value of 22.66 (Supplementary Fig. 11a). The association signal overlapped with the reported *Wf* locus[38], which controls white and green flesh. To identify the candidate gene, we constructed a RIL population[39] from a cross between an orange flesh line (MS-1152, Vedrantais, VED) and a white flesh line (Piel de sapo T111, PS) for QTL mapping, and detected a significant QTL (*LUMQU8.1*; around 29.63–29.87 Mb; LOD (likelihood of odds) score = 10.23) corresponding to the *Wf* locus on chromosome 8 (Supplementary Fig. 11b). Combining the QTL with GWAS results, a 96-kb overlapping interval containing 11 protein-coding genes was detected (Supplementary Fig. 11c). A previously reported candidate gene *MELO3C003069*[48] for *Wf*, encoding a pentatricopeptide protein, is 202-kb away from our mapping interval. Among the 11 genes, we found one gene, *MELO3C003097*, for which the ortholog (*SG1*) in *Arabidopsis* has been reported to be essential for chloroplast development and chlorophyll biosynthesis[45]. The expression level of *MELO3C003097* during fruit development was significantly higher in the green flesh accession (MS-982) than in the white flesh accession (MS-531; Supplementary Fig. 11d). These results suggest that *MELO3C003097* may be a strong candidate for

Fig. 3 | Identification of a candidate gene for the melon sutures trait. a, Manhattan plots of GWAS for fruit sutures in melon accessions. Fruits of representative sutured and non-sutured melon accessions are shown. Scale bar, 1.0 cm. **b**, Fine mapping of melon fruit sutures using diverse segregating populations. An 86-kb interval harboring four genes was identified (represented by arrows, of which the green arrow indicates the candidate gene *MELO3C019694*). **c**, Expression of the four genes in flower and developed fruit of different genotypes. NA, no annotated information; TMM, trimmed mean of M values. **d**, Phylogenetic tree of *MELO3C019694* and its homologues in rice (green points), *Arabidopsis* (red points), tomato (blue points), pumpkin (light blue points) and melon (purple points). The closest homologues of *MELO3C019694* indicated in a gray-shadowed box include those from *Arabidopsis* (*AT3G58780* and *AT2G42830*)[40] and tomato (*Solyc07g055920.2.1*)[41] that have been reported to be associated with pod dehiscence and pericarp expansion, respectively. **e**, Identification of a 1.07-kb deletion upstream of the *MELO3C019694* gene in accessions with sutures compared with non-sutured accessions. **f**, Proportion of the 1.07-kb deletion (purple) existing in sutured and non-sutured melon accessions. **g**, RT-qPCR analysis of *MELO3C019694* in the female flowers and young fruits in PS (a non-sutured accession) and VED (a sutured accession). Fl develop, flower in development; closed fl, flower before anthesis; 0 DAP, 3 DAP and 7 DAP represent fruits at 0, 3 and 7 days after pollination. Data are mean ± s.d. ($n = 3$ independent experiments).

the *Wf* locus. Notably, the same peak on chromosome 8 was found in GWAS for both peel and flesh color, indicating that this peak may play an important part in the color formation of both peel and flesh tissues in melon.

Discussion

In summary, our analysis based on large-scale genome resequencing suggests that there were three independent domestications of melon, one in Africa and two in India. Although the African clade



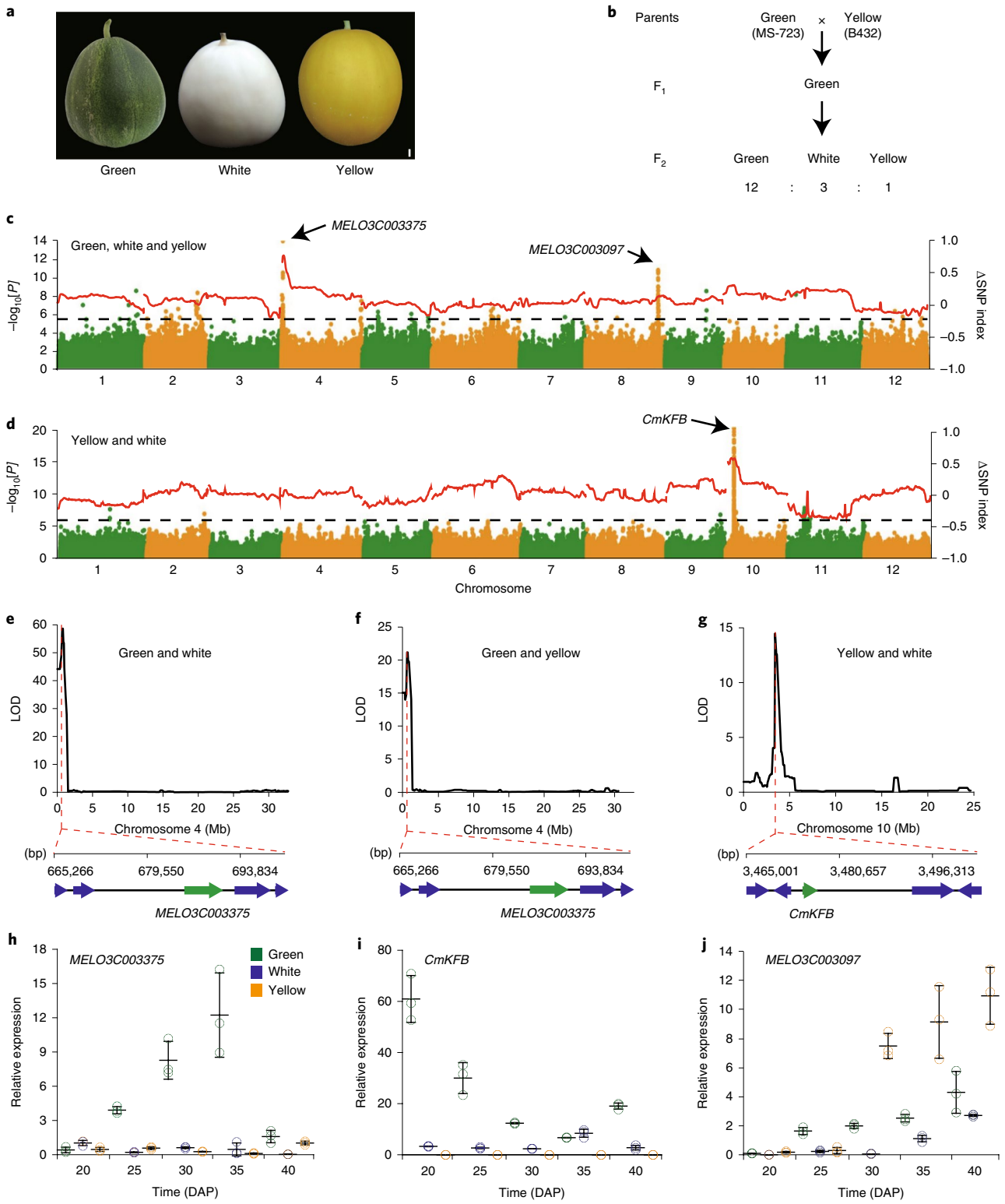


Fig. 4 | GWAS, bulked segregation analysis and QTL analysis identified the same region as being potentially important for peel color. a, Phenotypes of green, white and yellow peel melon accessions. Scale bar, 1.0 cm. **b**, Segregation of peel color in an F_2 population derived from crossing a green peel accession (MS-723) with a yellow peel accession (B432). **c**,**d**, Identification of overlapping intervals using GWAS and bulk segregation analyses for the peel color trait in melon. GWAS analyses (Manhattan plots) were performed using green, white and yellow peel melon accessions (**c**), and white and yellow peel melon accessions (**d**), respectively. Bulk segregation analyses (red lines) were conducted with the green and non-green (white peel and yellow peel) bulk populations (**c**), and the white peel and yellow peel bulk populations (**d**) from the above F_2 population. The candidate gene for each signal is provided. The horizontal dashed lines indicate the genome-wide threshold of GWAS signals ($P = 2.51 \times 10^{-6}$). **e**–**g**, QTL mapping using another F_2 population derived from crossing a green peel accession and a yellow peel accession. Candidate genes (in green), *MELO3C003375* and *CmKFB*[14], were located in the intervals of the identified QTLs. **h**–**j**, RT-qPCR analysis of *MELO3C003375*, *CmKFB*[14] and *MELO3C003097* during fruit development in melon accessions with different peel colors. Data are mean ± s.d. ($n = 3$ independent experiments).

(wild African and cultivated African) is clearly a different gene pool from *melo* (wild *melo* and cultivated *melo*) and *agrestis* (wild *agrestis* and cultivated *agrestis*) groups[5], there is a limited number of African accessions captured in our collection. Therefore, it would be worthwhile to explore a wider African diversity panel in future studies. The Indian domestication events were derived from distinct wild populations, and the small number of common selective sweeps suggests that domestication was achieved through diverse genetic pathways that ultimately resulted in similar phenotypes. The strong differentiation between *melo* and *agrestis* may be useful in breeding, because inter-subspecies crosses have the potential to generate heterosis and high diversity (Supplementary Note). Furthermore, our identification of candidate genes related to domestication and important agronomic traits (Supplementary Table 16) will be useful for melon breeding.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0522-8.

Received: 30 April 2019; Accepted: 24 September 2019;
Published online: 1 November 2019

References

1. Pitrat, M., Hanelt, P. & Hammer, K. Some comments on infraspecific classification of melon. *Acta Hortic.* **510**, 29–36 (2000).
2. Luan, F., Delannay, I. & Staub, J. E. Chinese melon (*Cucumis melo* L.) diversity analyses provide strategies for germplasm curation, genetic improvement, and evidentiary support of domestication patterns. *Euphytica* **164**, 445–461 (2008).
3. Kerje, T. & Grum, M. The origin of melon, *Cucumis melo*: a review of the literature. *Acta Hortic.* **510**, 37–44 (2000).
4. Sebastian, P., Schaefer, H., Telford, I. R. & Renner, S. S. 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. USA* **107**, 14269–14273 (2010).
5. Endl, J. et al. Repeated domestication of melon (*Cucumis melo*) in Africa and Asia and a new close relative from India. *Am. J. Bot.* **105**, 1662–1671 (2018).
6. Jeffrey, C. A review of the Cucurbitaceae. *Bot. J. Linn. Soc.* **81**, 233–247 (1980).
7. Serres-Giardi, L. & Dogimont, C. How microsatellite diversity helps to understand the domestication. In *Proc. Xth EUCARPIA Meeting on Genetics and Breeding of Cucurbitaceae* (eds Sari, N. et al.) 254–263 (Antalya, 2012).
8. Staub, J. E., Lopez-Sese, A. I. & Fanourakis, N. Diversity among melon landraces (*Cucumis melo* L.) from Greece and their genetic relationships with other melon germplasm of diverse origins. *Euphytica* **136**, 151–166 (2004).
9. Tanaka, K. et al. Seed size and chloroplast DNA of modern and ancient seeds explain the establishment of Japanese cultivated melon (*Cucumis melo* L.) by introduction and selection. *Genet. Resour. Crop Evol.* **63**, 1237–1254 (2016).
10. Paris, H. S., Amar, Z. & Lev, E. Medieval emergence of sweet melons, *Cucumis melo* (Cucurbitaceae). *Ann. Bot.* **110**, 23–33 (2012).
11. Garcia-Mas, J. et al. The genome of melon (*Cucumis melo* L.). *Proc. Natl Acad. Sci. USA* **109**, 11872–11877 (2012).
12. Boualem, A. et al. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* **321**, 836–838 (2008).
13. Tzuri, G. et al. A 'golden' SNP in *CmOr* governs the fruit flesh color of melon (*Cucumis melo*). *Plant J.* **82**, 267–279 (2015).
14. Feder, A. et al. A kelch domain-containing F-box coding gene negatively regulates flavonoid accumulation in muskmelon. *Plant Physiol.* **169**, 1714–1726 (2015).
15. Huang, X. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967 (2010).
16. Tian, F. et al. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* **43**, 159–162 (2011).
17. Jia, G. et al. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* **45**, 957–961 (2013).
18. Zhou, Z. et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **33**, 408–414 (2015).
19. Du, X. et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* **50**, 796–802 (2018).
20. Shang, Y. et al. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* **346**, 1084–1088 (2014).
21. Tieman, D. et al. A chemical genetic roadmap to improved tomato flavor. *Science* **355**, 391–394 (2017).
22. Zhu, G. et al. Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**, 249–261 (2018).
23. Argyris, J. M. et al. Use of targeted SNP selection for an improved anchoring of the melon (*Cucumis melo* L.) scaffold genome assembly. *BMC Genomics* **16**, 4 (2015).
24. Qi, J. et al. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* **45**, 1510–1515 (2013).
25. Dhillon, N. P. S. et al. Diversity among landraces of Indian snapmelon (*Cucumis melo* var. *momordica*). *Genet. Resour. Crop Evol.* **54**, 1267–1283 (2007).
26. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460 (1983).
27. Huang, X. et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
28. Diaz, A. et al. Quantitative trait loci analysis of melon (*Cucumis melo* L.) domestication-related traits. *Theor. Appl. Genet.* **130**, 1837–1856 (2017).
29. Kuang, J. F. et al. Two *GH3* genes from longan are differentially regulated during fruit growth and development. *Gene* **485**, 1–6 (2011).
30. Lin, T. et al. Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* **46**, 1220–1226 (2014).
31. Kato-Emori, S., Higashi, K., Hosoya, K., Kobayashi, T. & Ezura, H. Cloning and characterization of the gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase in melon (*Cucumis melo* L. *reticulatus*). *Mol. Genet. Genomics* **265**, 135–142 (2001).
32. Zhou, Y. et al. Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat. Plants* **2**, 16183 (2016).
33. Ma, D., Sun, L., Gao, S., Hu, R. & Liu, M. Studies on the genetic pattern of bitter taste in yong fruit of melon (*Cucumis melo* L.) [in Chinese]. *Acta Hortic. Sin.* **23**, 255–258 (1996).
34. Fujishita, N., Furukawa, H. & Morii, S. Distribution of three genotypes for bitterness of F_1 immature fruit in *Cucumis melo* [in Japanese]. *Jpn. J. Breed.* **43** (Suppl. 2), 206 (1993).
35. Cohen, S. et al. The *PH* gene determines fruit acidity and contributes to the evolution of sweet melons. *Nat. Commun.* **5**, 4026 (2014).
36. Nardi, C. F. et al. Expression of *FaXTH1* and *FaXTH2* genes in strawberry fruit. Cloning of promoter regions and effect of plant growth regulators. *Sci. Hortic.* **165**, 111–122 (2014).
37. Dogra, V., Sharma, R. & Yelam, S. Xyloglucan endo-transglycosylase/hydrolase (XET/H) gene is expressed during the seed germination in *Podophyllum hexandrum*: a high altitude Himalayan plant. *Planta* **244**, 505–515 (2016).
38. Périn, C. et al. A reference map of *Cucumis melo* based on two recombinant inbred line populations. *Theor. Appl. Genet.* **104**, 1017–1034 (2002).
39. Pereira, L. et al. QTL mapping of melon fruit quality traits using a high-density GBS-based genetic map. *BMC Plant Biol.* **18**, 324 (2018).
40. Liljegren, S. J. et al. *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**, 766–770 (2000).
41. Vrebalov, J. et al. Fleshy fruit expansion and ripening are regulated by the tomato *SHATTERPROOF* gene *TAGL1*. *Plant Cell* **21**, 3041–3062 (2009).
42. Tadmor, Y. et al. Genetics of flavonoid, carotenoid, and chlorophyll pigments in melon fruit rinds. *J. Agric. Food Chem.* **58**, 10722–10728 (2010).
43. Abe, A. et al. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* **30**, 174–178 (2012).
44. Liu, H. et al. Map-based cloning, identification and characterization of the *w* gene controlling white immature fruit color in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* **129**, 1247–1256 (2016).
45. Oren, E. et al. The multi-allelic *APRR2* gene is associated with fruit pigment accumulation in melon and watermelon. *J. Exp. Bot.* **70**, 3781–3794 (2019).
46. Pan, Y. et al. Network inference analysis identifies an *APRR2-like* gene linked to pigment accumulation in tomato and pepper fruits. *Plant Physiol.* **161**, 1476–1485 (2013).
47. Hu, Z. et al. The tetratricopeptide repeat-containing protein slow green1 is required for chloroplast development in *Arabidopsis*. *J. Exp. Bot.* **65**, 1111–1123 (2014).
48. Galpaz, N. et al. Deciphering genetic factors that determine melon fruit-quality traits using RNA-Seq-based high-resolution QTL and eQTL mapping. *Plant J.* **94**, 169–191 (2018).

Methods

Plant materials and sequencing. A diverse worldwide collection of 1,175 melon accessions and of 9 accessions from related species of the *Cucumis* genus was obtained from the National Mid-term Genebank for Watermelon and Melon, Zhengzhou, China, the Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, the US Department of Agriculture and the National Institute for Agricultural Research. Information about the accessions, including the names of the individual accessions, country of origin, group, identity of the variety and resequencing data summary, is provided in Supplementary Table 1. Genomic DNA was extracted from fresh young leaves using the cetyltriethylammnonium bromide method[49]. At least 5 μg of genomic DNA was used for each accession to construct sequencing libraries according to the manufacturer's instructions (Illumina). The libraries were sequenced on the Illumina HiSeq 2500 or HiSeq 3000 platform, generating 150-bp or 125-bp paired-end reads. Five F_2 populations were used in our study, of which the genome of individuals of three F_2 populations were sequenced with 5× depth and three bulk populations developed from an F_2 population were sequenced with 15× depth.

Sequence alignment and variation calling. To call SNPs, reads of all accessions were mapped to the melon reference genome[23] (version 3.5.1) using SOAP2[50] with the following parameters: -m 100 -x 888 -s 35 -l 32 -v 3. Mapped reads were filtered to remove PCR duplicates, assigned to chromosomes and sorted according to the mapping coordinates. Both pair-end and single-end mapped reads were used for SNP detection throughout the entire collection of melon accessions.

We identified possible SNPs for each accession relative to the reference using SOAPsnp[51] with the following parameters: -L 150 -u -F 1. The likelihood of the genotype of each individual accession in GLF format was then generated for each chromosome with SNP quality ≥ 40 and base quality ≥ 40.

To integrate SNPs across the entire collection, we called each SNP using GLFmulti[52] according to the maximum-likelihood estimation of site frequency. The core set of SNPs was obtained by filtering on the base of allele frequency and the quality score given by GLFmulti[52]. SNPs were further filtered using the following criteria: (i) one position with more than two alleles was considered to be a polymorphic site in the population and was excluded from the next analyses; (ii) the total sequencing depth should be >500× and <6,800× and the SNP quality value should be greater than 40; (iii) position with an average mapping rate of reads of <1.5 were retained to rule out the effect of duplications; and (iv) the nearest SNPs should be more than 1 bp away.

To obtain the final set of SNPs, we further performed filtering using segregation tests, which can distinguish any segregation pattern from random sequencing errors on the base of the sequencing depth of the two putative alleles in different individuals. Permutations were used to determine the significance of allele depth in the population, and only sites with $P < 0.01$ were retained. To detect small indels (≤5 bp in length), we mapped all of the sequence reads from each accession with a gap of ≤5 bp allowed (parameter -g 5) using SOAP2[50]. Indels (1–5 bp) were called by SOAPindel pipeline.

Planting and phenotyping. In total, 1,175 melon accessions and 9 accessions from related species were grown in Zhengzhou (Henan province), Sanya (Hainan province) and Changji (Xinjiang province) in 2015 and 2016. Because of poor adaptation of some accessions, several traits were evaluated in only one or two locations. Three replicates were performed at each location.

Three RIL populations[39,53,54] and five F_2 populations were used to identify candidates for sutures, peel color, flesh color, fruit bitterness, flesh thickness and sugar content (Supplementary Table 17). We phenotyped melons for the traits of fruit suture (a qualitative trait identified as presence or absence) and flesh color (a qualitative trait identified as yellow, green or white) at harvest. For flesh color, ripe fruit was cut in two longitudinal sections; one section was then used to evaluate flesh color visually, and the other section was scanned to perform the flesh color analysis in color spaces RGB and CIElab using the Tomato Analyzer 3.0 software[55,56]. Total chlorophyll and carotenoid contents were determined using ultraviolet–visible light spectroscopy. The other accessions and populations were grown in Xinxiang, Sanya or Beijing and phenotyped following a Chinese technical specification for evaluating melon[57].

Phylogenetic and population structure analyses. A subset of 17,055 SNPs with a minor allele frequency (MAF) ≥ 0.05 and missing rate ≤ 0.4 at fourfold degenerate sites representing neutral or near-neutral variants were used for phylogenetic and population structure analyses. The alignment was trimmed using trimAl[58] (v.1.4. rev22v) to remove positions that are non-variable and include more than 90% of gaps. The remaining sites were used to construct the phylogenetic tree with RAxML[59] (v.8.1.17) using the evolutionary model GTR. The branch length and rate parameters were optimized, and the aLRT SH-like branch support was calculated using PhyML[60] v.3.0 with the options -b -4, -o lr. The same method was used to construct the phylogenetic tree of the chloroplast genome. Population structure was investigated using the program STRUCTURE[61] (v.2.3.1), with the same data used in the phylogenetic tree construction. Furthermore, to confirm the result of STRUCTURE, we also performed the analysis using DAPC in the R package adegenet[62] v.2.1.0 with the parameter max.n.clust = 40, PCs to retain = 900, discriminant functions to retain = 5. In addition, principal component analysis (PCA) using the whole-genome SNPs with missing value ≤ 40% was performed with EIGENSOFT[63] v.6.0.1. Combining the phylogenetic tree analysis and PCA, we classified these accessions into three distinct clades (African, *melo* and *agrestis*). Considering the passport information, the *melo* and *agrestis* groups could each be further divided into two sub-clades (wild and cultivated *agrestis*; wild and cultivated *melo*), respectively.

Identification of domestication sweeps. To detect genomic regions affected by domestication, we measured the level of genetic diversity (π) using a 50-kb window with a step size of 5 kb in wild *melo*, cultivated *melo*, wild *agrestis* and cultivated *agrestis*. Genome regions affected by domestication should have substantially lower diversity in cultivated *melo* (π_{CM}) and cultivated *agrestis* (π_{CA}) than the diversity in wild *melo* (π_{WM}) and wild *agrestis* (π_{WA}), respectively. Windows with π_{WM} or π_{WA} lower than 0.002 were excluded from the analysis. Windows with the top 5% highest ratios of π_{WM}/π_{CM} (≥3.49) or π_{WA}/π_{CA} (≥26.18) were selected as candidate domestication sweeps. We also performed QTL mapping for fruit mass and bitterness to analyze QTLs or genes that segregated between the wild and cultivated parents by resequencing the individuals of two F_2 segregating populations derived from crossing between the wild and cultivated accessions. If genetic intervals of these QTLs and reported genes (loci) were close to or located in domestication sweeps, we considered them to be candidate domesticated QTLs or genes (Supplementary Table 10).

Identification of differentiated regions. The population fixation statistics (F_{ST}) were estimated for 50-kb sliding windows with a step size of 5 kb and for each SNP using a variance component approach implemented in the HIERFSTAT R package[64]. The average F_{ST} of all sliding windows was regarded as the value at the whole-genome level across different groups. Sliding windows with the top 10% highest F_{ST} values were selected initially. Neighboring windows were then merged into one fragment. If the distance between two fragments was <50 kb, fragments were merged into one region. The final merged regions were considered as highly diverged between different groups.

Watterson estimator analysis. The Watterson estimator of θ_W was evaluated using the software VariScan[65] v.2.0.3 for four main sub-populations. SNPs with a MAF ≥ 1% within each sub-population were used as the input data. A sliding window of 50 kb in length was used to scan the whole genome. The average θ_W value of all windows in the genome was then calculated to present the polymorphism.

LD analysis. Haploview software[66] was used to calculate LD values for each of the groups (wild *melo*, cultivated *melo*, wild *agrestis* and cultivated *agrestis*) using SNPs with MAF ≥ 0.05 with the following parameters: -n -pedfile -info -log -maxdistance 1000 -minMAF 0.05 -hwcutoff 0 -dprime -memory 10480. LD decay was measured on the basis of the r^2 value and the corresponding distance between two given SNPs.

HPLC analysis of cucurbitacin B. Fruit flesh and leaf samples were frozen in liquid nitrogen and ground using a mortar and pestle. The fine powder (0.5 g) was added to methanol (2 ml) and homogenized for 15 min, followed by centrifugation at 10,000*g* at 4 °C for 10 min. The solution was filtered through a 0.22-μm membrane before injection and then analyzed on an HPLC system (Agilent 1200) equipped with an XDB-C18 column (5 μm, 150 mm × 4.6 mm) and eluted with 55% methanol at 1 ml min^{-1} under a wavelength of 230 nm.

GWAS. Only SNPs with MAF ≥ 0.05 and missing rate ≤ 0.4 in a population were used to carry out GWAS. This resulted in 1,599,428, 872,244 and 2,028,259 SNPs that were used in GWAS for subspecies of *melo*, *agrestis* and the entire population (*melo* and *agrestis*), respectively. We performed GWAS using the efficient mixed-model association expedited (EMMAX) program[67]. Population stratification and hidden relatedness were modeled with a kinship (K) matrix in the emmax-kin-intel package of EMMAX. The P-value thresholds for significance were approximately 2.51×10^{-6}.

Bulked segregant analysis of F_2 population by whole-genome resequencing. We planted 450 individuals of an F_2 population derived from a cross between MS-723 (green peel accession) and B432 (yellow peel accession) in the winter of 2017 in Sanya, China. The fruit peel color of each individual was recorded. Genomic DNA was isolated from fresh leaves using the cetyltriethylammnonium bromide method. For bulked segregant analysis, bulked DNA samples were constructed by mixing equal amounts of DNA from 30, 29 and 9 individuals with representative green, white and yellow peel colors, respectively. Roughly 13× genome sequences for each of the two parents (B432 and MS-723) and 15× data for each of the three bulked samples (green peel, white peel and yellow peel) were generated. Short reads were aligned against the reference genome[23] (released 3.5.1) using the Burrows–Wheeler aligner (BWA)[68], and SNPs were identified using SAMtools[69]. The average SNP index for the green peel bulk and non-green peel bulk (white peel bulk and yellow

peel bulk), and white peel bulk and yellow peel bulk were calculated using a 1,000-kb sliding window with a step size of 100 kb.

Expression analysis of candidate genes for peel color, flesh color and suture. Fruits of MS-348 (yellow peel, orange flesh), MS-531 (white peel, white flesh) and MS-982 (green peel, green flesh) were sampled at 20, 25, 30, 35 and 40 days after pollination, respectively. Total RNA was extracted using RNAprep Pure plant kit (TIANGEN Biotech). First-strand cDNA synthesis was conducted using SuperScript RT Mix (Bio-Connect Biotech) according to the manufacturer's instructions. Then, 2 µl cDNA was used to perform RT–qPCR in a 10-µl reaction mixture. We analyzed the expression of *MELO3C003375*, *CmKFB* and *MELO3C003097* in the peel of MS-348, MS-531 and MS-982, and the expression of *MELO3C003097* in the flesh of MS-982 and MS-531. Expression analyses of *CmBi* and *CmBt* were performed in fruits at 7 days after anthesis of cultivated *melo* and *agrestis* accessions. For suture, we harvested MS-1152 (sutured line) and Piel de sapo T111 (non-sutured line) at Fl-develop (7-mm female flower), closed-Fl (10-mm female flower) and 0, 3, 7 days after pollination and calculated the expression of the candidate *MELO3C019694*. Three replicates were performed for each experiment. Relative expression levels were calculated using the $2^{-\Delta\Delta C_t}$ or $2^{-\Delta C_t}$ method.

Statistical analysis. Calculations of the χ^2 test statistic and s.d. were performed with the SPSS software. The significance was determined by two-tailed Student's *t*-tests.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw sequencing data reported in this paper have been deposited in the Sequence Read Archive (SRA) under a NCBI BioProject accession (PRJNA565104) and NCBI BioSample accessions (SAMN12791768–SAMN12792667, SAMN12791484–SAMN12791767). The sequencing data are also accessible from the BIG Data Center (http://bigd.big.ac.cn/gsa) under the accession number CRA001513. In addition, the data are also available from the corresponding authors on reasonable request.

Code availability

All codes are available from the corresponding authors upon request.

References

49. Gawel, N. J. & Jarret, R. L. A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. *Plant Mol. Biol.* **9**, 262–266 (1991).
50. Li, R. et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966–1967 (2009).
51. Li, R. et al. SNP detection for massively parallel whole-genome resequencing. *Genome Res.* **19**, 1124–1132 (2009).
52. He, W. et al. ReSeqTools: an integrated toolkit for large-scale next-generation sequencing based resequencing analysis. *Genet. Mol. Res.* **12**, 6275–6283 (2013).
53. Harel-Beja, R. et al. A genetic map of melon highly enriched with fruit quality QTLs and EST markers, including sugar and carotenoid metabolism genes. *Theor. Appl. Genet.* **121**, 511–533 (2010).
54. Gur, A. et al. Genome-wide linkage-disequilibrium mapping to the candidate gene level in melon (*Cucumis melo*). *Sci. Rep.* **7**, 9770 (2017).
55. Brewer, M. T. et al. Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. *Plant Physiol.* **141**, 15–25 (2006).
56. Darrigues, A., Hall, J., Knaap, E. & Francis, D. M. Tomato analyzer-color test: a new tool for efficient digital phenotyping. *J. Am. Soc. Hort. Sci.* **133**, 579–586 (2008).
57. Ma, S. et al. *Descriptors and Data Standard for Melon (*Cucumis melo *L.)* (China Agriculture Press, 2006).
58. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
59. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
60. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
61. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
62. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
63. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
64. Goudet, J. Hierfstat, a package for R to compute and test hierarchical *F*-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
65. Hutter, S., Vilella, A. J. & Rozas, J. Genome-wide DNA polymorphism analysesusing VariScan. *BMC Bioinformatics* **7**, 409 (2006).
66. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
67. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
68. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
69. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

Acknowledgements

We thank B. S. Gaut (Department of Ecology and Evolutionary Biology, University of California Irvine), W. Lucas (University of California, Davis), J. Ruan (Agricultural Genome Institute at Shenzhen, Chinese Academy of Agricultural Sciences) and D. Wu (Kunming Institute of Zoology, Chinese Academy of Sciences) for critical comments. This work was supported by funding from the Agricultural Science and Technology Innovation Program (to Yongyang Xu, S.H., Z.Z. and H.W.), the China Agriculture Research System (CARS-25 to Yongyang Xu and H.W.), the Leading Talents of Guangdong Province Program (00201515 to S.H.), the Shenzhen Municipal (The Peacock Plan KQTD2016113010482651 to S.H.), the Dapeng district government, National Natural Science Foundation of China (31772304 to Z.Z.), the Science and Technology Program of Guangdong (2018B020202007 to S.H.), the National Natural Science Foundation of China (31530066 to S.H.), the National Key R&D Program of China (2016YFD0101007 to S.H.), USDA National Institute of Food and Agriculture Specialty Crop Research Initiative (2015-51181-24285 to Z.F.), the European Research Council (ERC-SEXYPARTH to A.B.), the Spanish Ministry of Economy and Competitiveness (AGL2015–64625-C2-1-R to J.G.-M.), Severo Ochoa Programme for Centres of Excellence in R&D 2016–2010 (SEV-2015–0533 to J.G.-M.), the CERCA Programme/Generalitat de Catalunya to J.G.-M. and the German Science Foundation (SPP1991 Taxon-OMICS to H.S.).

Author contributions

S.H., Yongyang Xu and J.G.-M. designed studies and contributed to the original concept of the project, S.H., G.Z., T.L., Z.Z. and Q.F. managed the project, T.G., I.J., R.W., V.R. and W.F. performed the bioinformatics, S.M., J.S., Yongyang Xu, M. Pitrat, C.D., J.W., J.L. and A.J.M. contributed to the collection of the melon accessions, Y.H., G.Z., W.K., H.W., J.Z., Z.X., A.G., N.K., E.O., D.S., S.Z., Y.Z. and N.L. planted accessions, prepared the samples and performed phenotyping, P.W., Y.H., Y.Z., J.A., C.M., L.P., M. Pujol and D.O. designed and performed the molecular experiments, G.Z., Q.L. and T.L. prepared the figures and tables, S.H., T.L., J.G.-M., Z.F., T.G., A.J.M., V.R., A.G., Yong Xu, A.B., H.S. and J.J. revised the manuscript, G.Z., Q.L. T.L., Z.Z. and Q.F. analyzed data and wrote the paper.

Competing interests

The authors declare no competing interests.

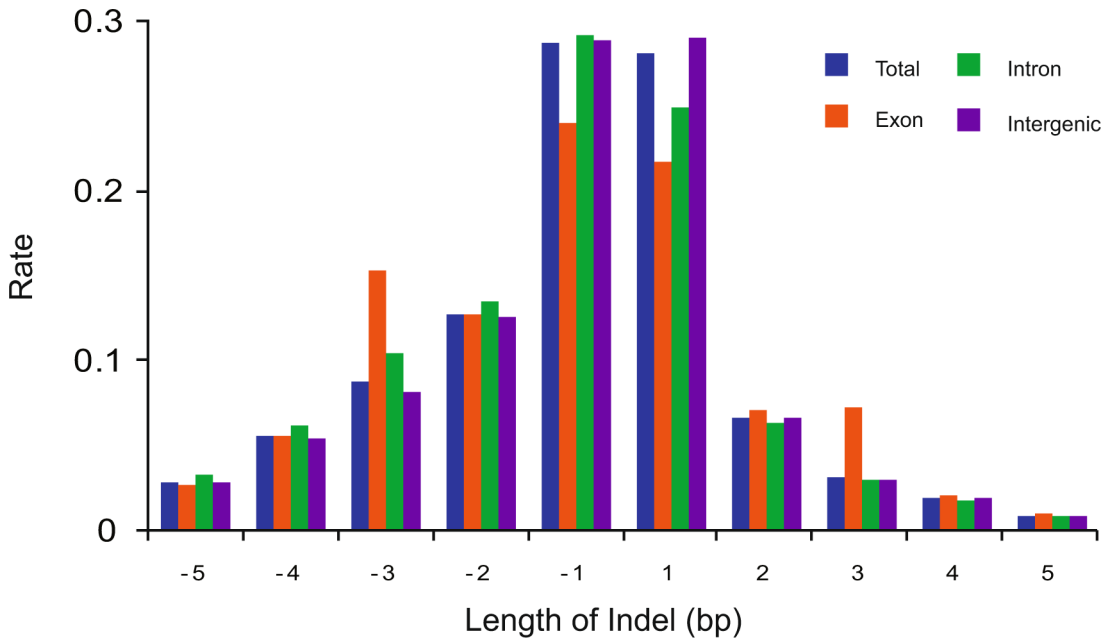
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-019-0522-8.

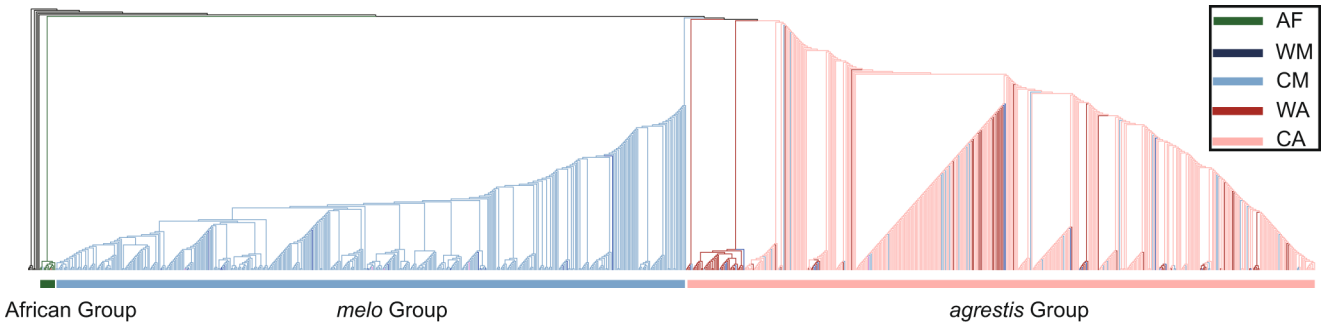
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0522-8.

Correspondence and requests for materials should be addressed to J.G.-M., Y.X. or S.H.

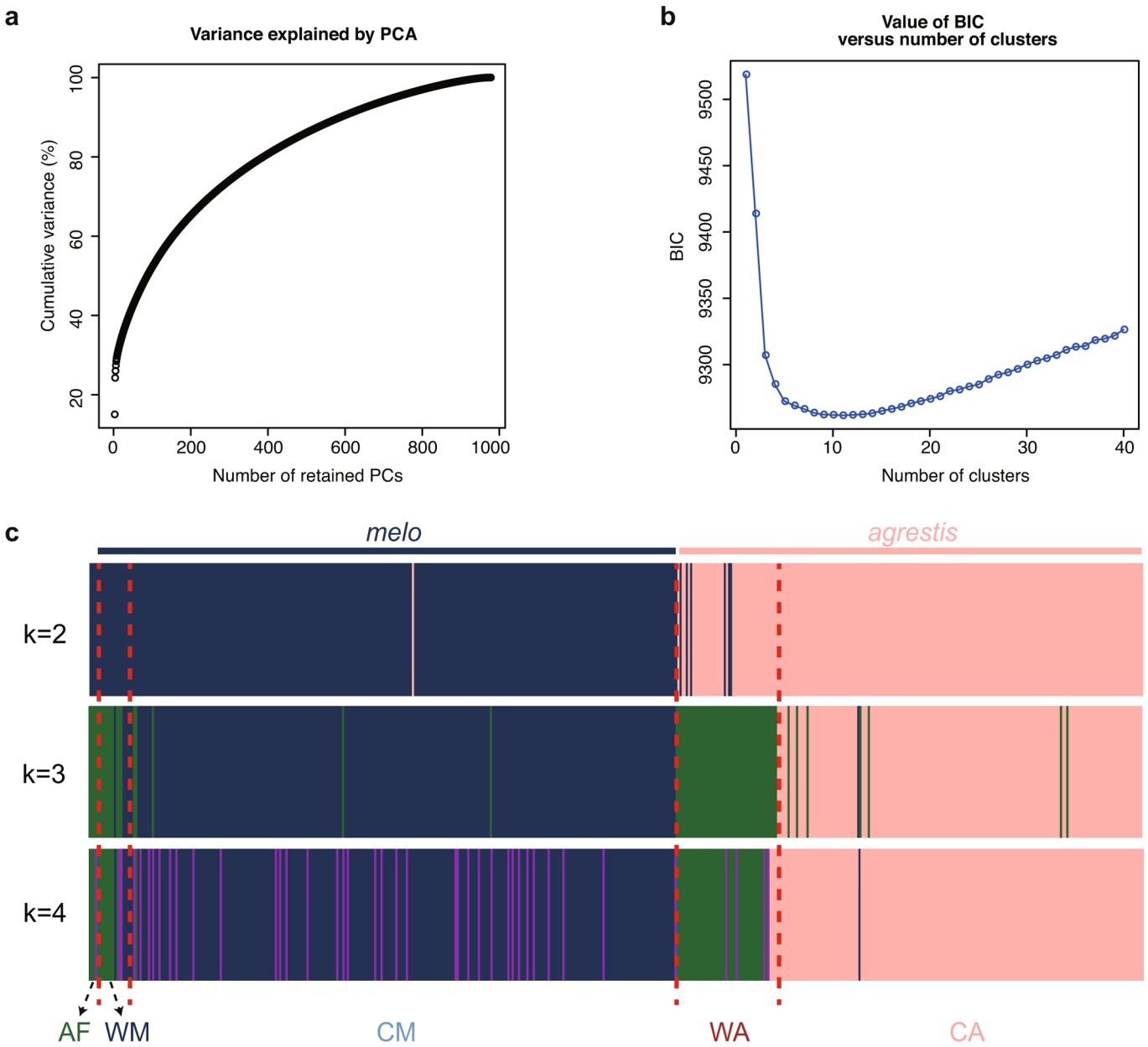
Reprints and permissions information is available at www.nature.com/reprints.



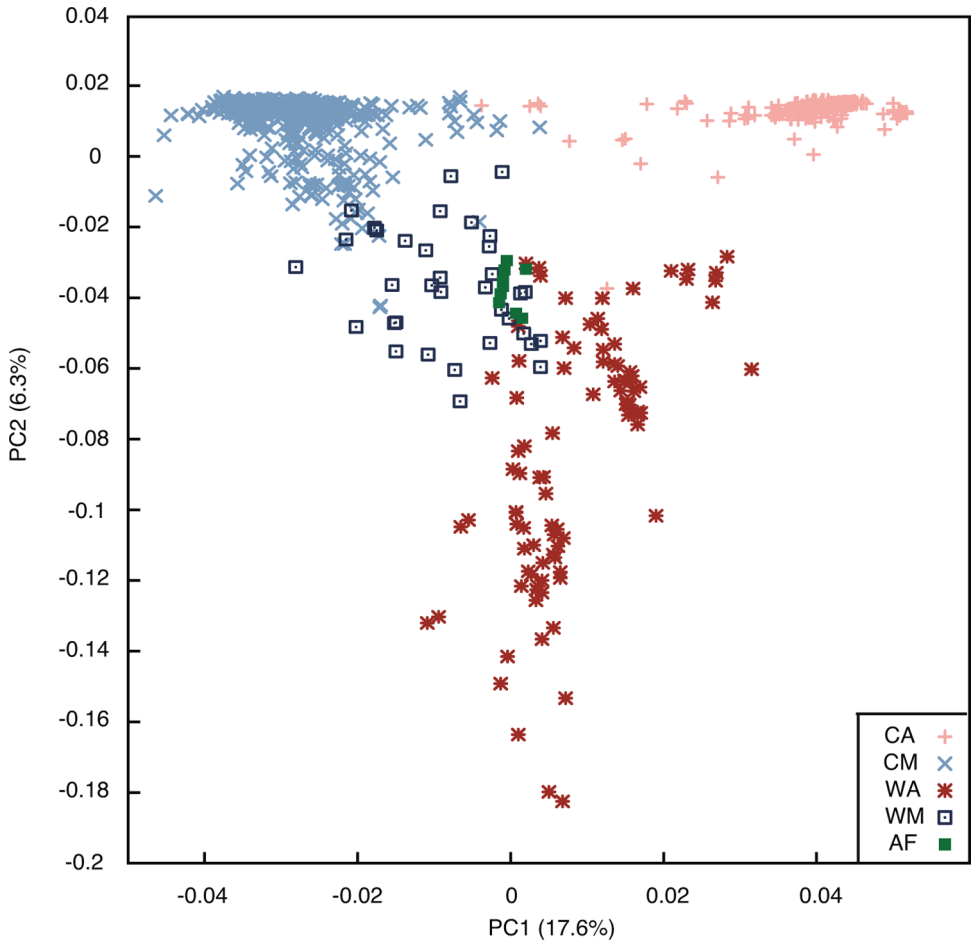
Extended Data Fig. 1 | Distribution of small indels. Distribution of small indels (≤ 5 bp) in different genomic regions. Indels in all regions were shown in blue. Those in intergenic regions were shown in purple. Indels in introns (green) are predominantly short (1 or 2 bp), whereas those in exons (orange) are often 3-bp long as this length does not cause a frame shift.



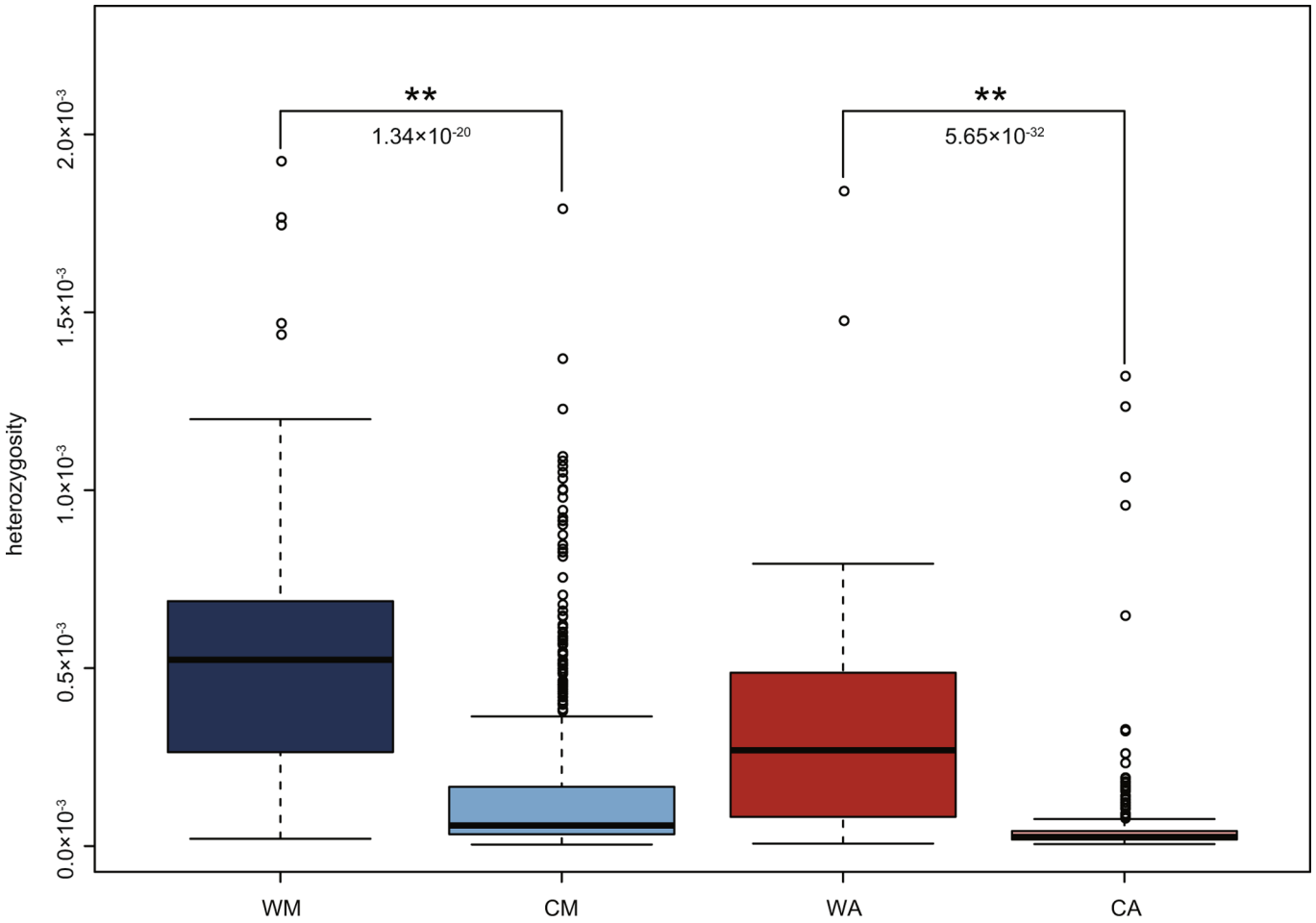
Extended Data Fig. 2 | Chloroplast Phylogenetic tree. Phylogenetic tree constructed with the chloroplast SNPs for 977 melon accessions.



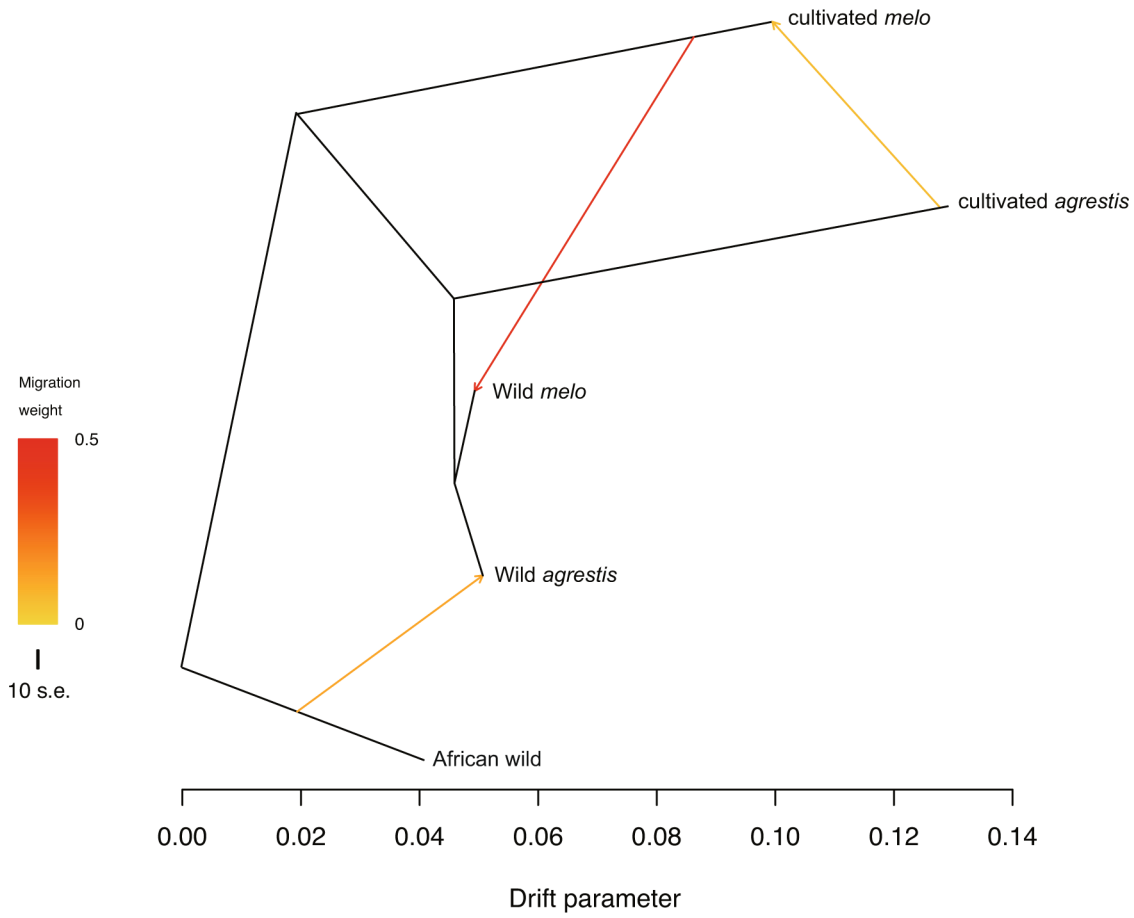
Extended Data Fig. 3 | The result of DAPC. Population structure analysis of melon accessions with DAPC. a, Cumulated variance explained by the eigenvalues of the PCA. b, Variation curve of BIC value. c, Model-based clustering analysis with different numbers of clusters (K = 2, 3 and 4).



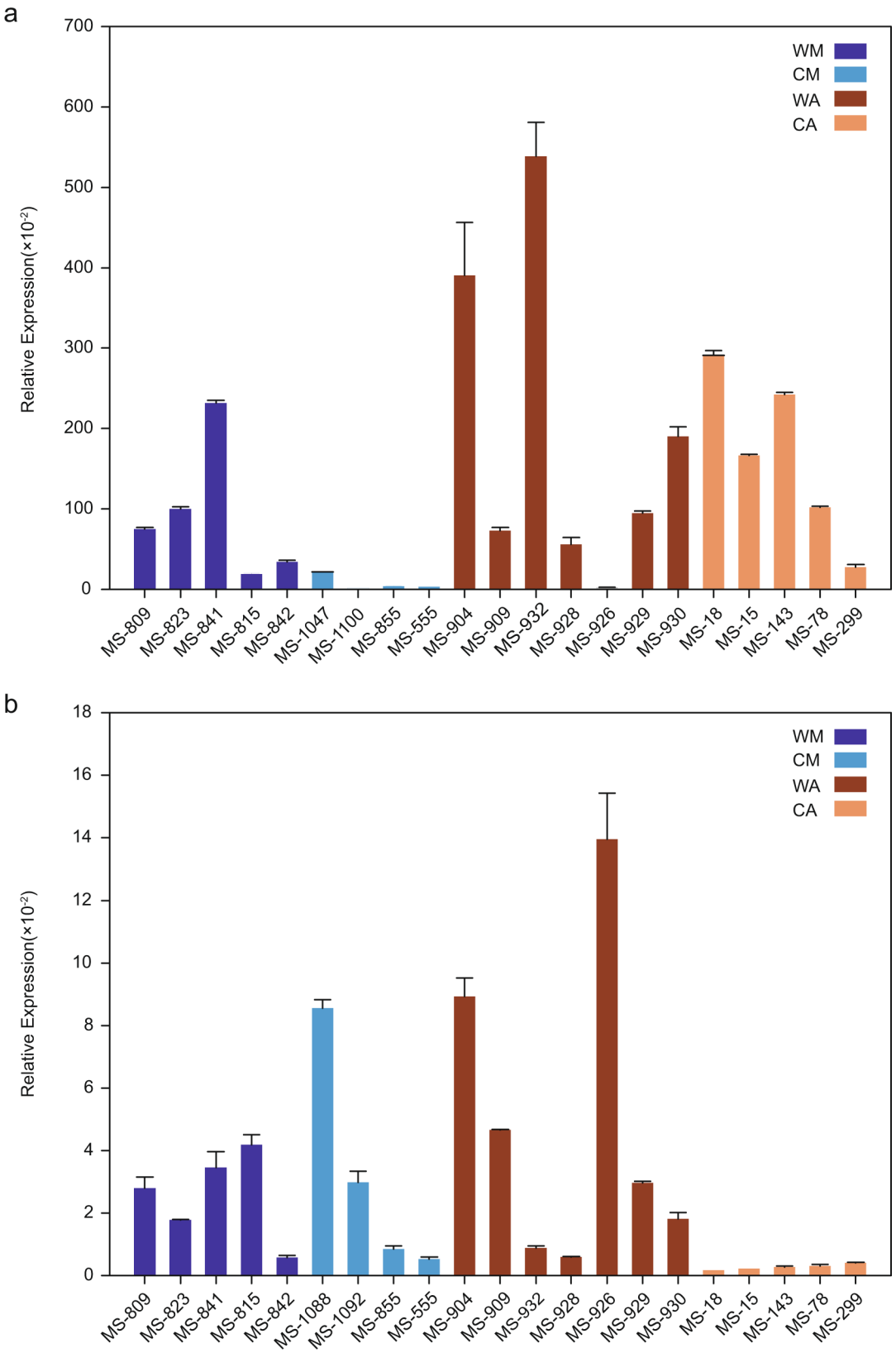
Extended Data Fig. 4 | Principal component analysis (PCA). Principal component analysis (PCA) of 968 melon accessions. SNPs with missing data rate ≤ 40% were used for PCA. Two-dimension coordinates were plotted for the 968 melon accessions. The African group (green) and *melo* group (blue) have a discrete distribution; the *agrestis* group (red) has an obviously centralized distribution.



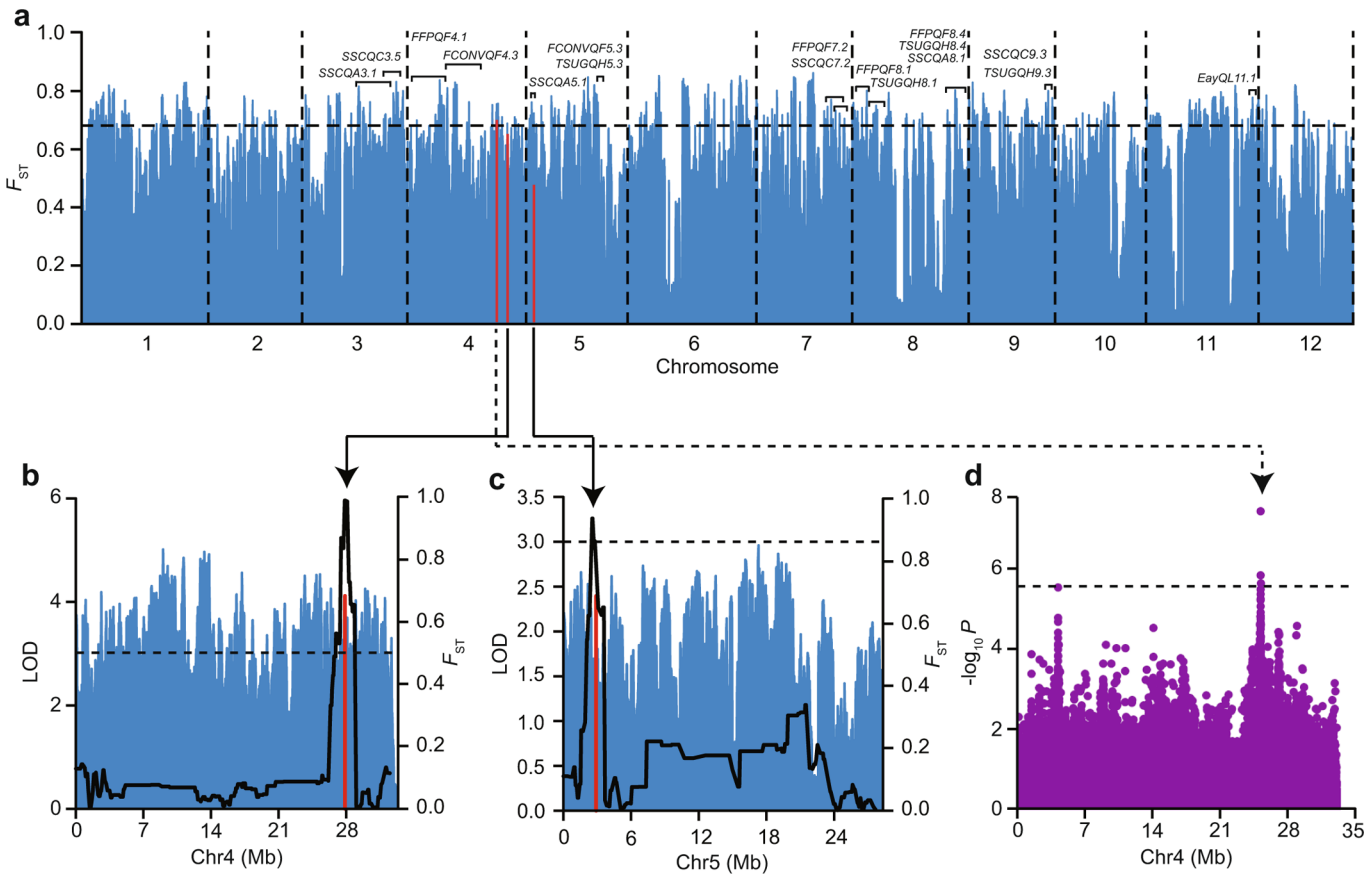
Extended Data Fig. 5 | The differentiation of Heterozygosity. The heterozygosity of different groups. Each box represents the mean and interquartile range. The top whisker denotes the maximum value and the bottom whisker denotes the minimum value. The significance was determined by two-tailed Student's t tests.



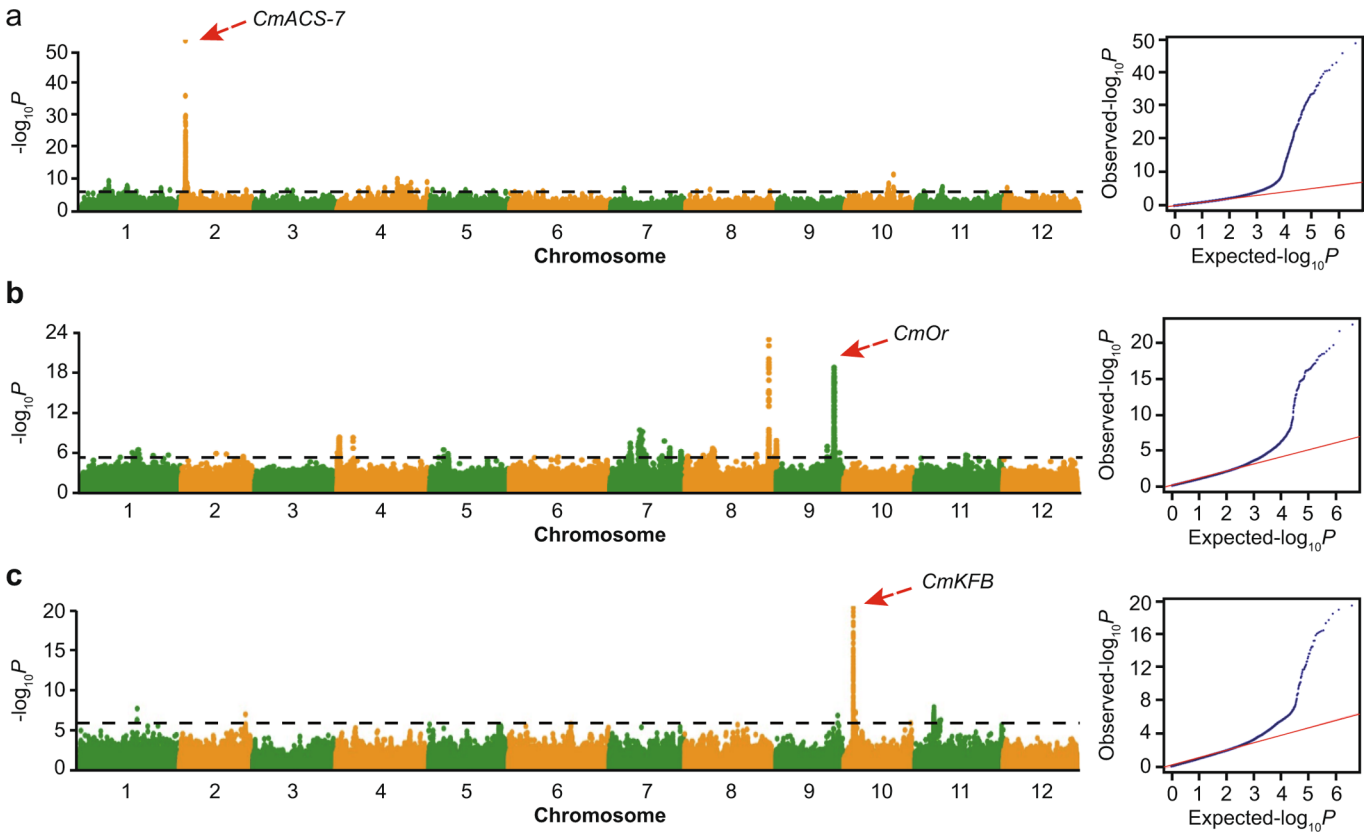
Extended Data Fig. 6 | The analysis of introgression. Treemix analysis of the main genetic clusters. Arrows represent the direction of migrations.



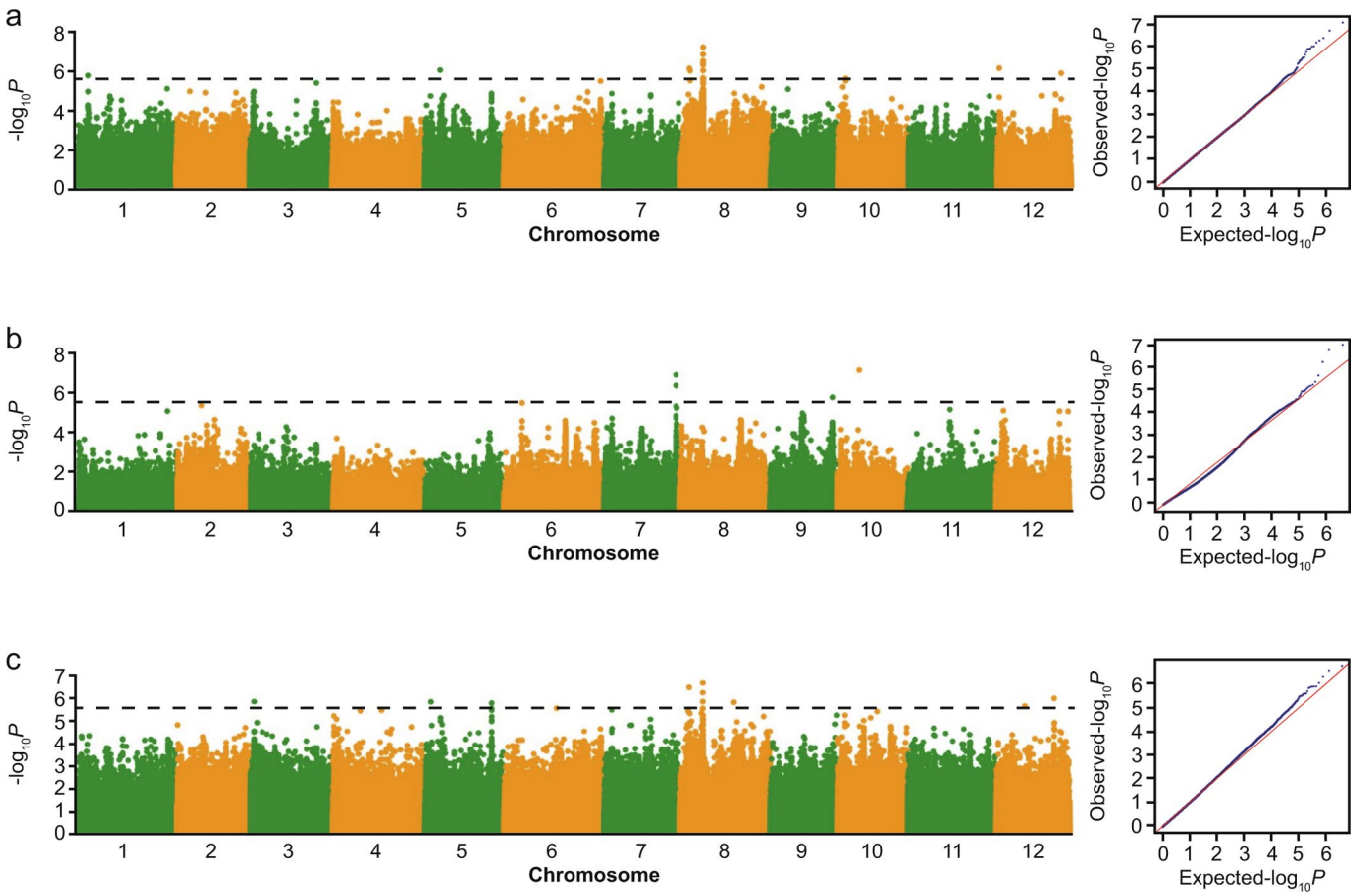
Extended Data Fig. 7 | Expression of *CmBi* and *CmBt*. RT-qPCR of *CmBi* (a) and *CmBt* (b) in young fruits of WM, CM, WA and CA accessions. Data are presented as mean ± s.d.(n = 3 independent measurements).



Extended Data Fig. 8 | Population differentiation in cultivated melon. Population differentiation between CM (cultivated *melo*) and CA (cultivated *agrestis*) groups. a, Distribution of F_{ST} across the melon genome. Highly divergent genomic regions overlapping previously reported QTL signals are indicated. The horizontal dashed line indicates the top 10% threshold. b,c, QTLs for flesh thickness identified from an F2 population from the cross of a cultivated *melo* accession and a cultivated *agrestis* accession. Both QTLs are located in regions with higher divergence levels ($F_{ST}=0.69$) and ($F_{ST}=0.48$), respectively. The black horizontal dashed lines indicate the threshold (LOD > 3.0) of QTL-mapping. d, Association signals identified by GWAS on ovary pubescence using the whole population. The significant threshold of -log10*P* value was set at 5.6.



Extended Data Fig. 9 | The verification of known genes in GWAS analysis. Previously reported genes identified in the GWAS analysis. a-c, Manhattan plots (left) and quantile-quantile plots (right) of GWAS for sex determination (a), orange flesh color (b) and yellow and white peel color (c) using the MLM model. The significant threshold of -log10P value was set at 5.6. Genes CmACS-7 (ref. [1]), CmOr2 and CmKFB3 are marked by red arrows.



Extended Data Fig. 10 | GWAS analysis of flesh aroma. GWAS analysis of flesh aroma in three different populations. a-c, Manhattan plots (left) and quantile-quantile plots (right) for GWAS on flesh aroma in the melo population (a), in the agrestis population (b), and in the whole population (c). The significant threshold of -log10P value was set at 5.6.

nature research

Corresponding author(s): Sanwen Huang
Yongyang Xu
Jordi Garcia-Mas

Last updated by author(s): 16/9/19

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect data.
Data analysis	Detailed description for all the softwares used for analysis have been provided in the Methods as well as Supplementary Note. The tools and software used in this study: SOAP2, SOAPsnp, GLFmulti, SOAPindel, BWA, Samtools, Vcftools, trimAl, RAxML, PhyML v3.0, STRUCTURE, EIGENSOFT 6.0.1, HIERFASAT, Haploview, EMMAX, SPSS, MSTMap, R/qtl

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The raw sequence data reported in this paper has been deposited in the Sequence Read Archive (SRA) under the NCBI BioProject (PRJNA565104) and BioSample (SAMN12791768 - SAMN12792667, SAMN12791484 - SAMN12791767) that are publicly accessible. The sequence data is also accessible in BIG Data Center (http://bigd.big.ac.cn/gsa) under the accession number CRA001513. In addition, the data is also available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples were selected to have enough representation from cultivated and wild species to derive meaningful and decisive conclusion. The germplasm collected in this research cover majority of the growing regions across world.
Data exclusions	No data was excluded.
Replication	Phenotyping experiments were performed in at least one or two locations and three replicates were performed at each location and confirmed the reproducibility of data. Assessment of gene expression and cucurbitacin B content were randomly replicated and confirm that results are reproducible.
Randomization	The samples were grouped base on the result of population structure analysis (phylogenetic tree, structure and PCA) and the geographical distribution as well as the botanical characteristics.
Blinding	The investigators were blinded to the group allocations.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging