

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

3 Guangwei Zhao^{1,20}, Qun Lian^{2,20}, Zhonghua Zhang^{3,4,20}, Qiushi Fu^{3,20}, Yuhua He¹,
4 Shuangwu Ma¹, Valentino Ruggieri^{5,6}, Antonio J. Monforte⁷, Pingyong Wang¹, Irene
5 Julca^{8,9,10}, Huaisong Wang³, Junpu Liu¹, Yong Xu¹¹, Runze Wang¹², Jiabing Ji², Zhihong
6 Xu¹, Weihu Kong¹, Yang Zhong², Jianli Shang¹, Lara Pereira^{5,6}, Jason Argyris^{5,6}, Jian
7 Zhang¹, Carlos Mayobre^{5,6}, Marta Pujol^{5,6}, Elad Oren¹³, Diandian Ou¹, Jiming Wang¹,
8 Dexi Sun¹, Shengjie Zhao¹, Yingchun Zhu¹, Na Li¹, Nurit Katzir¹³, Amit Gur¹³, Catherine
9 Dogimont¹⁴, Hanno Schaefer¹⁵, Wei Fan², Abdelhafid Bendahmane¹⁴, Zhangjun Fei^{16,17},
10 Michel Pitrat¹⁴, Toni Gabaldón^{9,10,18}, Tao Lin^{2,19}, Jordi Garcia-Mas^{5,6*}, Yongyang Xu^{1*},
11 Sanwen Huang^{2*}

12
13 ¹Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences,
14 Zhengzhou, China.

15 ²Lingnan Guangdong Laboratory of Modern Agriculture, Genome Analysis Laboratory of
16 the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese
17 Academy of Agricultural Sciences, Shenzhen, China.

18 ³Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of the
19 Ministry of Agriculture, Sino-Dutch Joint Laboratory of Horticultural Genomics, Institute
20 of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China.

21 ⁴College of Horticulture, Qingdao Agricultural University, Qingdao, China.

22 ⁵Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Barcelona, Spain.

23 ⁶Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Barcelona, Spain.

24 ⁷Instituto de Biología Molecular y Celular de Plantas, Universitat Politècnica de
25 Valencia-Consejo Superior de Investigaciones Científicas (CSIC-UPV), Valencia, Spain.

26 ⁸Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.

27 ⁹Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology,

28 Barcelona, Spain.

29 ¹⁰Universitat Pompeu Fabra (UPF), Barcelona, Spain.

30 ¹¹National Watermelon and Melon Improvement Center, Beijing Academy of Agricultural
31 and Forestry Sciences, Key Laboratory of Biology and Genetic Improvement of
32 Horticultural Crops (North China), Beijing Key Laboratory of Vegetable Germplasm
33 Improvement, Beijing, China.

34 ¹²Centre of Pear Engineering Technology Research, State Key Laboratory of Crop
35 Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China.

36 ¹³Plant Science Institute, Israeli Agricultural Research Organization, NeweYa'ar Research
37 Center, Ramat Yishay, Israel.

38 ¹⁴INRA, Génétique et Amélioration des Fruits et Légumes, Montfavet, France.

39 ¹⁵Department of Ecology and Ecosystem Management, Plant Biodiversity Research,
40 Technical University of Munich, Freising, Germany.

41 ¹⁶Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY, USA.

42 ¹⁷US Department of Agriculture–Agricultural Research Service, Robert W. Holley Center
43 for Agriculture and Health, Ithaca, NY, USA.

44 ¹⁸Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys,
45 Barcelona, Spain.

46 ¹⁹China Agricultural University, College of Horticulture, Beijing, China.

47 ²⁰These authors contributed equally to this work.

48 *For correspondence: Sanwen Huang, huangsanwen@caas.cn; Yongyang Xu,
49 xuyongyang@caas.cn; Jordi Garcia-Mas, Jordi.Garcia@irta.cat.

50 **ABSTRACT**

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

63

64 **INTRODUCTION**

65 Melon (*Cucumis melo* L.), an important crop in the Cucurbitaceae family, is cultivated
66 worldwide, with more than 32 million tons produced in 2017 (United Nations Food and
67 Agriculture Organization (FAO) statistics), and was domesticated four thousand years
68 ago¹. Since most wild *Cucumis* species emerged in Africa and have the same
69 chromosome number as *C. melo*, it has been proposed that the center of origin of
70 cultivated melon is Africa^{2,3}. However, recent studies revealed that the closest wild
71 relatives of melon are found in India and Australia^{4,5}. Melon has been classified into two
72 subspecies, *C. melo* subsp. *melo* (*melo*) and *C. melo* subsp. *agrestis* (*agrestis*), based on
73 ovary pubescence⁶. Both domesticated subspecies exhibit increased size of fruit, leaf and
74 seed, and loss of fruit bitterness and acidity. However, the fruit sizes of the two
75 domesticated subspecies are highly different, and bitterness was fully lost in *melo*, but
76 partially in *agrestis*. Therefore, the history and genetic basis of melon domestication still

77 remain poorly understood. Current knowledge of melon domestication is largely derived
78 from molecular marker analyses^{7,8}, and limited archaeological⁹ and historical data¹⁰. The
79 availability of the melon genome sequence (454 Mb)¹¹ and a collection of melon
80 germplasm resources made it possible to rapidly detect genomic variations and to offer
81 new and powerful insights into the trajectory of melon domestication on a genome-wide
82 scale.

83

84 *Cucumis melo* is a highly diversified species and a model system for studying several
85 important biological processes¹¹, but only limited number of genes related to agronomic
86 traits like fruit monoecy¹², flesh color¹³ and peel color¹⁴ have been identified.
87 Genome-wide association studies (GWAS) are a powerful approach for identifying genes
88 or quantitative trait loci (QTLs) underlying complex traits as has been demonstrated in
89 rice¹⁵, maize¹⁶, foxtail millet¹⁷, soybean¹⁸, cotton¹⁹, cucumber²⁰, and tomato^{21,22}. Here we
90 present the genome resequencing of 1,175 diverse accessions to characterize the
91 population structure and domestication history of melon, and we provide genomic
92 evidence for elucidating melon taxonomy. We also performed GWAS to identify a
93 number of candidate genes and loci underlying several important agricultural traits.

94

95 **RESULTS**

96 **Melon genome variation map**

97 We used 1,175 diverse accessions of *C. melo*, which included 667 from subspecies *melo*
98 and 508 from *agrestis*, and an additional 9 from closely related species (Fig. 1a and
99 Supplementary Table 1). The *C. melo* collection consisted of 134 wild and 1,041 cultivated
100 accessions spanning most of the species' native range. We generated a total of 4.29 trillion
101 base pairs of sequence using next-generation sequencing technology, with a median depth
102 of 4.98-fold and coverage of 80.73% of the assembled melon genome²³ (released 3.5.1).
103 After aligning the reads against the melon reference genome²³, we identified a total of

104 5,678,165 single-nucleotide polymorphisms (SNPs) and 957,421 small indels (≤ 5 bp),
105 with an average of 13.99 SNPs and 2.36 indels per kilobase (Supplementary Fig. 1,
106 Supplementary Tables 2 and 3) that is similar to cucumber (17.22 SNPs and 1.75 indels
107 per kilobase)²⁴. The accuracy of the identified SNPs was estimated to be 99.07% when
108 comparing 10 pairs of accessions sequenced with low (4.71 \times) and high depth (18.92 \times)
109 (Supplementary Table 4). A total of 197,113 SNPs (3.47%) and 10,114 (1.06%) indels
110 were located in the coding regions, among which 13,022 showed potentially large effects:
111 7,030 SNPs (0.12%) affected 5,598 genes by causing start codon changes, premature stop
112 codons or elongated transcripts, and 5,992 (0.63%) indels led to frame-shift in 4,587
113 annotated genes (Supplementary Tables 2 and 3). Collectively, this comprehensive melon
114 genome variation dataset provides a new resource for melon biology and breeding.

115

116 **Population structure**

117 The phylogenetic relationships for these melon accessions were inferred using a subset of
118 17,055 SNPs at four-fold degenerate sites. To better deduce the relationships of melon
119 accessions from different areas, 207 with uncertain origin were excluded in the next
120 analysis. The phylogenetic tree based on the nuclear and chloroplast genome SNPs (Fig.1b
121 and Supplementary Fig. 2) supported three distinct clades, which exhibited strong
122 geographic separation and distinctive botanical features. We found that only the primitive
123 African domesticated types (CAF) were clustered with wild African accessions (WAF) in
124 Clade I (AF), suggesting the marginal impact of WAF during melon domestication outside
125 of Africa, consistent with a previous study⁵. The remaining accessions in Clade II and
126 Clade III corresponded to *melo* and *agrestis* subspecies according to passport information
127 and morphological characteristics. A similar result was obtained by DAPC analysis
128 (Supplementary Fig. 3).

129

130 Model-based clustering and principal component (PCA) analyses further classified

131 each of Clade II and Clade III into two main subclades (Fig. 1c and Supplementary Fig. 4).
132 The majority of *C. melo* var. *momordica* (*momordica*) accessions native to India, and
133 traditionally considered as cultivated *agrestis*, formed a single subclade (Clade II-1) with
134 an obvious admixture in genetic composition (Fig. 1c), and closely clustered together with
135 cultivated *melo* accessions (Clade II-2). In addition, Clade II-1 generally showed similar
136 characteristics to wild melon, such as monoecy, low sugar content, acid flesh and high
137 resistance to pests and disease²⁵. Thus, our data suggested that cultivated *melo* was
138 domesticated directly from *momordica* (Fig. 1b). Clade III-1 consisted of wild (*C. melo* L.
139 var. *agrestis*) and cultivated *agrestis* accessions. In this clade, the cultivated *agrestis*
140 accessions from southern Africa unexpectedly clustered with wild *agrestis* melon derived
141 from India. These accessions share similarities to wild melon from India with respect to
142 bearing small fruits, monoecy and having gelatinous sheath around the seeds. Based on
143 both genetic and trait similarities, South African *agrestis* accessions could represent
144 recent migrants from India. Both morphological and genomic data largely support that the
145 cultivated *melo* (Clade II-2; CM) and *agrestis* (Clade III-2; CA) accessions were
146 domesticated from Clade II-1 (wild *melo*; WM) and Clade III-1 (wild *agrestis*; WA),
147 respectively. Finally, we can speculate that there were three independent domestication
148 events leading to the three main clusters: two occurring in India and another in Africa,
149 consistent with a recent study⁵. Due to the relatively few wild African melon accessions in
150 our study and their low influence during melon domestication, we used the melon
151 accessions from Clade II and Clade III for further analyses.

152

153 Nucleotide diversity measured by the π value²⁶ for the WM (0.00347) and WA
154 (0.0031) groups was substantially higher than that for the CM (0.00256) and CA (0.00074)
155 groups, which is consistent with the result of Watterson estimator analysis (θ_w -WM:
156 0.00256, θ_w -WA: 0.00218, θ_w -CM: 0.00151, and θ_w -CA: 0.00069). Furthermore, different
157 groups exhibited different degrees of heterozygosity (Supplementary Fig. 5) and we also

158 observed gene-flow between these groups, which may be due to the overlapping
159 distribution of accessions, open pollination and modern breeding program (Supplementary
160 Fig. 6). The decay of linkage disequilibrium (LD) with physical distance between SNPs to
161 half of the maximum values occurred at 22.0 kb and 20.6 kb in the WM ($r^2 = 0.17$) and WA
162 ($r^2 = 0.14$) groups, respectively, which were considerably smaller than that in the CM (58.0
163 kb, $r^2 = 0.31$) and CA (610.2 kb, $r^2 = 0.43$) groups (Fig.1d). These results together are
164 strongly suggestive of a significant genetic diversity reduction in cultivated melon because
165 of domestication. Notably, the CA group has much lower nucleotide diversity and higher
166 LD decay than the CM group, suggesting that the CA group has undergone a more severe
167 bottleneck during domestication.

168

169 **Independent domestication of *melo* and *agrestis* melon**

170 Since the Neolithic revolution and the development of agriculture, human preferably kept
171 and propagated seeds from wild plants with larger and more delicious fruits. To identify
172 potential selective signals during melon domestication, we scanned genomic regions
173 showing drastic reduction in nucleotide diversity by comparing each cultivated group with
174 its corresponding wild group (π_{WM}/π_{CM} and π_{WA}/π_{CA}) over 50 kb windows. We identified
175 148 and 185 putative selection sweeps associated to domestication in *melo* ($\pi_{WM}/\pi_{CM} \geq$
176 3.49) and *agrestis* ($\pi_{WA}/\pi_{CA} \geq 26.18$), respectively, covering 6.28% (25.52 Mb) and 7.23%
177 (29.39 Mb) of the assembled genome and harboring 1,481 and 1,710 genes
178 (Supplementary Tables 5-8). Notably, only 143 of 27,427 genes (2.67 Mb; 0.66% of the
179 assembled genome) were shared between the *melo* and *agrestis* sweeps (Supplementary
180 Table 9). By contrast, in rice, most well-characterized domestication genes were shared
181 between the *indica* and *japonica* types²⁷. Jointly, the *melo* and *agrestis* sweeps cover 52.24
182 Mb (12.86% of the assembled genome), which encompasses 3,048 genes.

183

184 To discover potential domestication loci, we developed two F₂ segregating
185 populations derived from a cross between a WM (MS-542) and a CM (additional
186 accession B460), and a cross between a WA (additional accession yesheng) and a CA
187 (MS-79), and identified 19 new QTLs that overlap with sweep regions, including 10 in
188 *melo* and 9 in *agrestis* (Fig. 2a,b and Supplementary Table 10). Among the QTLs, one
189 genomic region (~2.95-4.77 Mb on chromosome 8) related to fruit mass (*fwqaz8.1*,
190 *fdqaz8.1*, *ftqaz8.1* for fruit weight, fruit diameter and flesh thickness, respectively) (Fig.
191 2c), a trait that has been under human selection was detected using the above F₂
192 segregating population from a cross between a WA and a CA accession. This region is
193 consistent with a previously reported QTL (*fwqc8.1*) that contributes negatively to the
194 increase of fruit weight in melon²⁸. The nucleotide diversity of this interval was
195 drastically reduced in the CA group compared to WA ($\pi_{WA}/\pi_{CA} = 7.45$), but the reduction
196 was only minor in this region in CM ($\pi_{WM}/\pi_{CM} = 1.45$) (Fig. 2d). This region included two
197 genes, *MELO3C007596* and *MELO3C007597*, both encoding auxin-responsive GH3-like
198 proteins. Auxin-responsive GH3-like proteins are reported to be involved in fruit growth
199 and development in longan²⁹ and tomato³⁰, suggesting both of the two genes are logical
200 candidates to be associated with fruit mass during *agrestis* domestication. Moreover, we
201 found that *Cm-HMGR* (*MELO3C026512*) on chromosome 3, which encodes a
202 hydroxy-methylglutaryl coenzyme A reductase (HMGR) in the mevalonate (MVA)
203 pathway that is involved in controlling fruit size in melon³¹, was located in a *melo*
204 domestication sweep ($\pi_{WM}/\pi_{CM} = 4.07$) (Fig. 2e).

205

206 Bitterness is an essential domestication trait, and has been (partially) lost during the
207 domestication of melon. We observed that 95.18% of CM accessions carried non-bitter
208 young fruits, whereas 25.87% of young fruits from the CA accessions were bitter,
209 especially the fruits exposed to stress conditions. This suggests that the two melon groups

210 possess different domestication mechanisms conferring the loss of bitterness. We detected
211 a notable decrease in nucleotide diversity in the *Bi* (encoding a cucurbitadienol synthase)
212 cluster ($\pi_{WM}/\pi_{CM} = 3.61$ and $\pi_{WA}/\pi_{CA} = 1.75$) in *melo* and in the *CmBt* locus (encoding a
213 transcription factor that activates *CmBi* transcription)³² in *agrestis* ($\pi_{WA}/\pi_{CA} = 28.21$ and
214 $\pi_{WM}/\pi_{CM} = 3.70$) (Fig. 2a,b). Additionally, there is an obvious population differentiation in
215 *CmBi* and *CmBt* between the CM and CA groups (Fig. 2f,g).

216

217 To further verify potential sweeps related to bitterness during melon domestication,
218 we performed QTL mapping using the above two F₂ populations (WM × CM and WA ×
219 CA). Two QTLs (~29.04-29.93 and ~21.07-22.05 Mb on chromosome 11 and 9,
220 respectively) associated with fruit bitterness were identified in *melo* and *agrestis*,
221 harboring the *CmBi*³² and the *CmBt* gene³², respectively (Fig. 2f,g). In general, alleles
222 from different bitterness-related genes have been domesticated in the two melon
223 subspecies. This was further validated by assessment of gene expression (Fig. 2h,i and
224 Supplementary Fig. 7) and cucurbitacin B content (Supplementary Table 11). Furthermore,
225 we found that hybrids (F₁) from a cross between non-bitter CM (accession MS-251) and
226 CA (accession MS-42) lines always had bitter young fruits, suggesting that
227 complementary genes exist conferring bitterness, as previously hypothesized^{33,34}.

228

229 Acidity as one major component of taste is selected by farmers and breeders during
230 melon breeding. The acidic genotype is presumably the ancestral form³⁵, and the acidity
231 in domesticated *melo* and *agrestis* too seems to be related to different genes. For example,
232 we found that one domestication sweep ($\pi_{WA}/\pi_{CA} = 1.94$ and $\pi_{WM}/\pi_{CM} = 9.99$) containing
233 the *CmPH* gene (*MELO3C025264*)³⁵ on chromosome 8 occurred in the *melo* but not in
234 the *agrestis* group (Fig. 2a,b). The *CmPH* gene, encoding a transmembrane transporter,
235 determines fruit acidity based on the presence of an insertion of a four amino-acid

236 duplication in non-acidic melon accessions, and contributes to the evolution of sweet
237 melons³⁵. We further verified the causative variation and found that the insertion
238 occurred in non-acidic melon accessions of the CM group, but not all accessions were
239 consistent in the CA group, suggesting that other genes might contribute to the acidity in
240 the CA group. The *CmPH* gene was also detected within the association signals in a
241 GWAS analysis for flesh acidity in *melo* accessions (Fig. 2j), and located in a *melo* sweep
242 (Fig. 2a). Intriguingly, *MELO3C011482* on chromosome 3 encoding a ATP-citrate
243 synthase subunit 1 was located in an association signal (Fig. 2k) and a sweep of *agrestis*
244 (Fig. 2b).

245

246 In summary, our results suggest that distinct domestication mechanisms for fruit mass,
247 flesh bitterness and acidity occurred in *melo* and *agrestis* accessions, which further
248 supports the hypothesis that the CM and CA groups were domesticated independently from
249 the WM and WA groups, respectively.

250

251 **Divergence between the *melo* and *agrestis* groups**

252 Melon is a diverse species cultivated by local farmers and used by breeders in many
253 countries. Geographically, *melo* is cultivated worldwide, whereas *agrestis* is concentrated
254 in East Asia. In general, cultivated *agrestis* melons possess greener leaves, edible epicarp,
255 thinner flesh, lower sugar content and ecological differences resulting from their distinct
256 geographical distributions. To dissect genomic regions underlying these differences, we
257 measured the pairwise genome-wide fixation index (F_{ST}) values based on SNPs between
258 different melon groups. The average F_{ST} value between the CM and CA groups was
259 estimated at 0.46, which was similar to that of *indica* and *japonica* rice (0.55)¹⁵, indicating
260 strong population differentiation in the two subspecies. Based on F_{ST} , 289 divergent
261 genomic regions between *melo* and *agrestis* were identified, which covered 56.9 Mb
262 (14.01%) of the assembled genome and harboring 3,535 predicted genes (Supplementary

263 Tables 12 and 13).

264

265 Seventeen previously reported QTLs for flesh thickness and sugar content
266 overlapped with these divergent genomic regions (Supplementary Fig. 8a and
267 Supplementary Table 14). We also identified two new QTLs (Supplementary Fig. 8b,c)
268 for flesh thickness on chromosome 4 and 5 using an F₂ population from a cross between a
269 CM (additional accession Y14) and a CA line (MS-1006), both of which were close to
270 divergence regions. Intriguingly, one GWAS signal on chromosome 4 for ovary
271 pubescence, a trait used as to distinguish *melo* and *agrestis*⁶, was also located in a
272 divergence region (Supplementary Fig. 8d). Furthermore, five xyloglucan
273 endo-transglycosylase genes involved in plant growth^{36,37} were located in a divergence
274 region (~2.15-2.28 Mb) of chromosome 5. This dataset constitutes a relevant resource for
275 the exploitation of genes conferring genetic differentiation between *melo* and *agrestis*.

276

277 **Identification of genes or loci related to important agronomic traits**

278 Melon has several diverse characteristics of agronomic importance, such as sex
279 determination, peel and flesh color, and fruit shape. However, few genes or loci underlying
280 agronomic traits have been identified in melon so far. To explore the potential of GWAS to
281 identify causal genes for complex traits, we performed an association study using a panel
282 composed of 1,067 diverse accessions for 16 agronomic traits (Supplementary Table 15).
283 A total of 208 significant association signals were identified in the melon genome. Four
284 previously dissected genes were found in these association signals, including *CmACS-7*
285 for monoecy¹², *CmOr* for flesh color¹³, *CmKFB* for peel color¹⁴ and *CmPH* for acidity³⁵
286 (Fig. 2 and Supplementary Fig. 9). The remaining 204 signals were associated with yield
287 (76), fruit quality (29) and morphological (99) traits (Supplementary Figs. 10-20). We
288 further validated major GWAS signals for rind sutures, peel color and flesh color using
289 segregating populations and molecular biology approaches (Fig. 3,4 and Supplementary

290 Fig. 21).

291

292 Rind sutures (also called vein tracks) is an important trait commonly found in
293 commercial melons, which is controlled by a single gene³⁸ on chromosome 11. A strong
294 GWAS signal for rind sutures was identified ($P = 2.14 \times 10^{-68}$; ~20.6-24.8 Mb) on
295 chromosome 11 (Fig. 3a). To further validate this signal, we constructed a RIL population³⁹
296 obtained by crossing a sutured line (MS-1152, Vedrantaïs) with a non-sutured line (Piel de
297 sapo T111) and narrowed down the QTL to a 1.7-Mb (~22.8-24.5 Mb) interval (Fig. 3b).
298 We further delimited this interval to an approximately 86-kb (~23.17-23.25 Mb) region
299 using additional F₂ and recombinant inbred line (RIL) populations. The region contains
300 four putative protein-encoding genes, of which two are expressed in flower and fruit
301 tissues (Fig. 3c). We found that *MELO3C019694*, encoding an AGAMOUS MADS-box
302 transcription factor, resides 16.8 kb upstream of the strongest association signal. The
303 orthologues of this gene include *SHP1* (*AT3G58780*) / *SHP2* (*AT2G42830*), which
304 regulate pod dehiscence in *Arabidopsis*⁴⁰, and *TAGL1* (*Solyc07g055920.2.1*), which is
305 required for pericarp expansion and climacteric ripening in tomato⁴¹ (Fig. 3d).

306

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

316

317 Peel and flesh color are important fruit quality traits influencing consumers' choice
318 and acceptability. Peel colors of commercial melons are green, white or yellow (Fig. 4a),
319 which are conferred by distinct pigment accumulation⁴². We identified a 12:3:1
320 segregation ratio for green, white and yellow traits by analyzing an F₂ segregating
321 population from a cross between a green-peel (MS-723) and a yellow-peel (B432, an
322 additional accession) lines (Fig. 4b), indicating that green peel is dominant epistatic to
323 non-green (white and yellow) peel. We selected 254 green and 381 non-green accessions
324 (145 white and 236 yellow accessions) in a GWAS analysis and identified two strong
325 association signals on chromosome 4 ($-\log_{10}P$ value = 14.17) and chromosome 8 ($-\log_{10}P$
326 value = 10.78) (Fig. 4c). Moreover, we performed a GWAS analysis using the non-green
327 accessions (145 white and 236 yellow accessions). One significant peak ($-\log_{10}P$ value =
328 20.80) was detected on chromosome 10 (Fig. 4d).

329

330 To further verify these signals related to peel color, we sequenced three bulk
331 populations with green, white and yellow peel from the above F₂ population (MS-723 ×
332 B432). We computed the differences of SNP indices (Δ SNP index)⁴³ between the green
333 and non-green peel bulk populations, and between the white and yellow peel bulk
334 populations, respectively, and identified a single overlapping genomic region with those
335 of GWAS on chromosome 4 (Fig. 4c) and chromosome 10 (Fig. 4d), respectively. The
336 overlapping genomic regions were also identified in another F₂ segregating population
337 obtained by crossing a green peel line with a yellow peel line (Fig. 4e-g). Within these
338 genomic intervals (~0.30-0.80 Mb on chromosome 4 and ~3.41-3.50 Mb on chromosome
339 10), we detected two candidate genes related to peel color, *MELO3C003375* on
340 chromosome 4 encoding a two-component response regulator-like protein APRR2 and the
341 already identified *CmKFB* gene on chromosome 10 negatively regulating naringenin
342 chalcone accumulation¹⁴. The orthologous genes of *MELO3C003375* in cucumber (*w*)⁴⁴,
343 watermelon (*CICG09G012330*)⁴⁵, pepper (GeneBank No. KC175445)⁴⁶ and tomato

344 (*SolyC08g077230*)⁴⁶ have been demonstrated to control chlorophyll metabolism and
345 pigment accumulation in fruit peel.

346

347 *MELO3C003375* exhibited much higher transcript levels in green-peel lines than
348 white and yellow-peel lines (Fig. 4h). There is hardly any expression of *CmKFB* in
349 yellow-peel accession (Fig. 4i), consistent with its function of negatively regulating
350 flavonoid accumulation. Additionally, we detected a gene (*MELO3C003097*) in the
351 genomic interval (~29.74-29.77 Mb) of the association signal on chromosome 8, an
352 ortholog of the Arabidopsis *SG1*, which encodes a Protein Slow Green 1, required for
353 chloroplast development⁴⁷, and expressed in every peel color accessions (Fig. 4j). We
354 speculate that *MELO3C003375* and *CmKFB* are associated with the green and yellow
355 peel trait, respectively; *MELO3C003097* could be a minor gene involved in peel color
356 formation.

357

358 Moreover, we performed GWAS on flesh color using 688 melon accessions. Besides
359 the identified *Gf* gene (*CmOr*)¹³ controlling orange flesh, we detected a strong association
360 signal on chromosome 8 with a highest $-\log P$ value of 22.66 (Supplementary Fig. 21a).
361 The association signal overlapped with the reported *Wf* locus³⁸ controlling white and green
362 flesh. To identify the candidate gene, we constructed a RIL population³⁹ from a cross
363 between an orange-flesh (MS-1152, Vedrantaïs) and a white-flesh (Piel de sapo T111) lines
364 for QTL mapping, and detected a significant QTL (*LUMQU8.1*; ~29.63-29.87 Mb; LOD
365 score = 10.23) corresponding to the *Wf* locus on chromosome 8 (Supplementary Fig. 21b).
366 Combining the QTL with GWAS results, a 96-kb overlapping interval containing 11
367 protein-coding genes was detected (Supplementary Fig. 21c). A previously reported
368 candidate gene *MELO3C003069* (ref. 48) for *Wf*, encoding a pentatricopeptide protein, is
369 202-kb away from our mapping interval. Among the 11 genes, we found that one gene,
370 *MELO3C003097*, whose orthologue (*SG1*) in Arabidopsis was reported to be essential for

371 chloroplast development and chlorophyll biosynthesis⁴⁵. The expression level of
372 *MELO3C003097* during fruit development was significantly higher in green-flesh
373 accession (MS-982) than in white-flesh accession (MS-531) (Supplementary Fig. 21d).
374 These results suggest that *MELO3C003097* may be a strong candidate for the *Wf* locus.
375 Interestingly, the same peak on chromosome 8 was found in GWAS for both peel and flesh
376 color, indicating that this peak may play an important role in the color formation of both
377 peel and flesh tissues in melon.

378

379 **DISCUSSION**

380 In summary, our analysis based on large-scale genome resequencing suggests three
381 independent domestications of melon, one in Africa and two in India. Though the African
382 clade (WAF and CAF) is clearly a different gene pool from *melo* (WM and CM) and
383 *agrestis* (WA and CA) groups⁵, there is a limited number of African accessions captured
384 in our collection. Therefore, it would be worthwhile to explore a wider African diversity
385 panel in future studies. The Indian domestication events were derived from distinct wild
386 populations, and the small number of common selective sweeps suggests that
387 domestication was achieved via diverse genetic pathways that ultimately resulted in
388 similar phenotypes. The strong differentiation between *melo* and *agrestis* may be useful in
389 breeding, because inter-subspecies crosses have the potential to generate heterosis and
390 high diversity (Supplementary Note). Furthermore, our identification of candidate genes
391 related to domestication and important agronomic traits (Supplementary Table 16) will be
392 useful for melon breeding.

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

412

413 **Author contributions** S.H., Yongyang. Xu. and J.G.-M. designed studies and contributed
414 to the original concept of the project, S.H., G.Z., T.L., Z.Z. and Q.F. managed the project,
415 T.G., I.J., R.W., V.R. and W.F performed the bioinformatics, S.M., J.S., Yongyang. Xu., M.
416 Pitrat., C.D., J.W., J.L. and A.J.M. contributed to the collection the melon accessions, Y.H.,
417 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,
418 prepared the samples and performed phenotyping, P.W., Y.H., Y.Z., J.A., C.M., L.P., M.
419 Pujol. and D.O. designed and performed the molecular experiments, G.Z., Q.L. and T.L.

420 prepared the figures and tables, S.H., T.L., J.G.-M., Z.F., T.G., A.J.M., V.R., A.G., Yong.
421 Xu., A.B., H.S. and J.J. revised the manuscript, G.Z., Q.L. T.L., Z.Z. and Q.F. analyzed
422 whole data and wrote the paper.

423

424 **Competing interests** The authors declare no competing interests.

425

References

- 426 1 Pitrat, M., Hanelt, P. & Hammer, K. Some comments on infraspecific classification
427 of melon. *Acta Horticulturae* **510**, 29-36 (2000).
- 428 2 Luan, F., Delannay, I. & Staub, J. E. Chinese melon (*Cucumis melo* L.) diversity
429 analyses provide strategies for germplasm curation, genetic improvement, and
430 evidentiary support of domestication patterns. *Euphytica* **164**, 445-461 (2008).
- 431 3 Kerje, T. & Grum, M. The origin of melon, *Cucumis melo*: a review of the literature.
432 *Acta Horticulturae* **510**, 37-44 (2000).
- 433 4 Sebastian, P., Schaefer, H., Telford, I. R. & Renner, S. S. Cucumber (*Cucumis sativus*)
434 and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the
435 sister species of melon is from Australia. *Proc. Natl. Acad. Sci. USA* **107**,
436 14269-14273 (2010).
- 437 5 Endl, J. et al. Repeated domestication of melon (*Cucumis melo*) in Africa and Asia
438 and a new close relative from India. *Am. J. Bot.* **105**, 1662-1671 (2018).
- 439 6 Jeffrey, C. A review of the Cucurbitaceae. *Botanical Journal of the Linnean Society*
440 **81**, 233-247 (1980).
- 441 7 Serres-Giardi, L. & Dogimont, C. *How microsatellite diversity helps to understand*
442 *the domestication*. in Proceedings of Xth EUCARPIA meeting on genetics and
443 breeding of Cucurbitaceae (eds. Sari, N., Solmaz, I. & Aras, V.) 254-263 (Antalya,
444 2012).
- 445 8 Staub, J. E., Lopez-Sese, A. I. & Fanourakis, N. Diversity among melon landraces
446 (*Cucumis melo* L.) from Greece and their genetic relationships with other melon
447 germplasm of diverse origins. *Euphytica* **136**, 151-166 (2004).
- 448 9 Tanaka, K. et al. Seed size and chloroplast DNA of modern and ancient seeds explain
449 the establishment of Japanese cultivated melon (*Cucumis melo* L.) by introduction
450 and selection. *Genetic Resources and Crop Evolution* **63**, 1237-1254 (2015).
- 451 10 Paris, H. S., Amar, Z. & Lev, E. Medieval emergence of sweet melons, *Cucumis melo*
452 (*Cucurbitaceae*). *Ann. Bot.* **110**, 23-33 (2012).
- 453 11 Garcia-Mas, J. et al. The genome of melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci.*
454 *USA* **109**, 11872-11877 (2012).
- 455 12 Boualem, A. et al. A conserved mutation in an ethylene biosynthesis enzyme leads to
456 andromonoecy in melons. *Science* **321**, 836-838 (2008).
- 457 13 Tzuri, G. et al. A 'golden' SNP in *CmOr* governs the fruit flesh color of melon
458 (*Cucumis melo*). *Plant J.* **82**, 267-279 (2015).
- 459 14 Feder, A. et al. A kelch domain-containing F-Box coding gene negatively regulates
460 flavonoid accumulation in muskmelon. *Plant Physiol.* **169**, 1714-1726 (2015).
- 461 15 Huang, X. et al. Genome-wide association studies of 14 agronomic traits in rice
462 landraces. *Nat. Genet.* **42**, 961-967 (2010).
- 463 16 Tian, F. et al. Genome-wide association study of leaf architecture in the maize nested
464 association mapping population. *Nat. Genet.* **43**, 159-162 (2011).
- 465 17 Jia, G. et al. A haplotype map of genomic variations and genome-wide association

- 466 studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* **45**, 957-961
467 (2013).
- 468 18 Zhou, Z. et al. Resequencing 302 wild and cultivated accessions identifies genes
469 related to domestication and improvement in soybean. *Nat. Biotechnol.* **33**, 408-414
470 (2015).
- 471 19 Du, X. et al. Resequencing of 243 diploid cotton accessions based on an updated A
472 genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* **50**, 796-802
473 (2018).
- 474 20 Shang, Y. et al. Biosynthesis, regulation, and domestication of bitterness in cucumber.
475 *Science* **346**, 1084-1088 (2014).
- 476 21 Tieman, D. et al. A chemical genetic roadmap to improved tomato flavor. *Science*
477 **355**, 391-394 (2017).
- 478 22 Zhu, G. et al. Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**,
479 249-261 (2018).
- 480 23 Argyris, J. M. et al. Use of targeted SNP selection for an improved anchoring of the
481 melon (*Cucumis melo* L.) scaffold genome assembly. *BMC Genomics* **16**, 4 (2015).
- 482 24 Qi, J. et al. A genomic variation map provides insights into the genetic basis of
483 cucumber domestication and diversity. *Nat. Genet.* **45**, 1510-1515 (2013).
- 484 25 Dhillon, N. P. S. et al. Diversity among landraces of Indian snapmelon (*Cucumis melo*
485 var. *momordica*). *Genetic Resources and Crop Evolution* **54**, 1267-1283 (2007).
- 486 26 Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics*
487 **105**, 437-460 (1983).
- 488 27 Huang, X. et al. A map of rice genome variation reveals the origin of cultivated rice.
489 *Nature* **490**, 497-501 (2012).
- 490 28 Diaz, A. et al. Quantitative trait loci analysis of melon (*Cucumis melo* L.)
491 domestication-related traits. *Theor. Appl. Genet.* **130**, 1837-1856 (2017).
- 492 29 Kuang, J. F. et al. Two *GH3* genes from longan are differentially regulated during fruit
493 growth and development. *Gene* **485**, 1-6 (2011).
- 494 30 Lin, T. et al. Genomic analyses provide insights into the history of tomato breeding.
495 *Nat. Genet.* **46**, 1220-1226 (2014).
- 496 31 Kato-Emori, S., Higashi, K., Hosoya, K., Kobayashi, T. & Ezura, H. Cloning and
497 characterization of the gene encoding 3-hydroxy-3-methylglutaryl coenzyme A
498 reductase in melon (*Cucumis melo* L. *reticulatus*). *Mol. Genet. Genomics* **265**,
499 135-142 (2001).
- 500 32 Zhou, Y. et al. Convergence and divergence of bitterness biosynthesis and regulation
501 in Cucurbitaceae. *Nat. Plants* **2**, 16183 (2016).
- 502 33 Ma, D., Sun, L., Gao, S., Hu, R. & Liu, M. Studies on the genetic pattern of bitter taste
503 in yong fruit of melon (*Cucumis melo* L.). *Acta Horticulturae Sinica* **23**, 255-258
504 (1996).
- 505 34 Fujishita, N., Furukawa, H. & Morii, S. Distribution of three genotypes for bitterness
506 of F₁ immature fruit in *Cucumis melo*. *Jpn. J. Breed* **43** (Suppl. 2), 206 (1993).

- 507 35 Cohen, S. et al. The *PH* gene determines fruit acidity and contributes to the evolution
508 of sweet melons. *Nat. Commun.* **5**, 4026 (2014).
- 509 36 Nardi, C. F. et al. Expression of *FaXTH1* and *FaXTH2* genes in strawberry fruit.
510 Cloning of promoter regions and effect of plant growth regulators. *Scientia*
511 *Horticulturae* **165**, 111-122 (2014).
- 512 37 Dogra, V., Sharma, R. & Yelam, S. Xyloglucan endo-transglycosylase/hydrolase
513 (XET/H) gene is expressed during the seed germination in *Podophyllum hexandrum*:
514 a high altitude Himalayan plant. *Planta* **244**, 505-515 (2016).
- 515 38 Perin, C. et al. A reference map of *Cucumis melo* based on two recombinant inbred
516 line populations. *Theor. Appl. Genet.* **104**, 1017-1034 (2002).
- 517 39 Pereira, L. et al. QTL mapping of melon fruit quality traits using a high-density
518 GBS-based genetic map. *BMC Plant Biol.* **18**, 324 (2018).
- 519 40 Liljegren, S. J. et al. *SHATTERPROOF* MADS-box genes control seed dispersal in
520 *Arabidopsis*. *Nature* **404**, 766-770 (2000).
- 521 41 Vrebalov, J. et al. Fleshy fruit expansion and ripening are regulated by the tomato
522 *SHATTERPROOF* gene *TAGL1*. *Plant Cell* **21**, 3041-3062 (2009).
- 523 42 Tadmor, Y. et al. Genetics of flavonoid, carotenoid, and chlorophyll pigments in
524 melon fruit rinds. *J. Agric. Food Chem.* **58**, 10722-10728 (2010).
- 525 43 Abe, A. et al. Genome sequencing reveals agronomically important loci in rice using
526 MutMap. *Nat. Biotechnol.* **30**, 174-178 (2012).
- 527 44 Liu, H. et al. Map-based cloning, identification and characterization of the *w* gene
528 controlling white immature fruit color in cucumber (*Cucumis sativus* L.). *Theor. Appl.*
529 *Genet.* **129**, 1247-1256 (2016).
- 530 45 Oren, E. et al. Multi-allelic *APRR2* gene is associated with fruit pigment
531 accumulation in melon and watermelon. *J. Exp. Bot.* **70**, 3781-3794 (2019).
- 532 46 Pan, Y. et al. Network inference analysis identifies an *APRR2-like* gene linked to
533 pigment accumulation in tomato and pepper fruits. *Plant Physiol.* **161**, 1476-1485
534 (2013).
- 535 47 Hu, Z. et al. The tetratricopeptide repeat-containing protein slow green1 is required
536 for chloroplast development in *Arabidopsis*. *J. Exp. Bot.* **65**, 1111-1123 (2014).
- 537 48 Galpaz, N. et al. Deciphering genetic factors that determine melon fruit-quality traits
538 using RNA-Seq-based high-resolution QTL and eQTL mapping. *Plant J.* **94**, 169-191
539 (2018).

541 **Figure legends**

542

543 **Fig.1** Geographic distribution and population structure of melon accessions. **a**,
544 Geographic distribution of melon accessions, which are represented by dots on the world
545 map. **b**, Phylogenetic tree of the population (531 *C. melo* subsp. *melo*, 437 *C. melo* subsp.
546 *agrestis* accessions and 9 wild relatives as the outgroup) constructed using 17,055 SNPs at
547 four-fold degenerate sites. AF, African; WAF, wild African; CAF, cultivated African; WM,
548 wild *melo*; CM, cultivated *melo*; WA, wild *agrestis*; CA, cultivated *agrestis*.
549 Representative fruits of the three clades studied are shown. Scale bars represent 1.0 cm. **c**,
550 Model-based clustering analysis with different numbers of clusters ($K = 2, 3$ and 4). The y
551 axis quantifies clusters membership, and the x axis lists the different accessions. The orders
552 and positions of these accessions on the x axis are consistent with those in the phylogenetic
553 tree. **d**, Genome-wide average LD decay estimated from different melon group.

554

555 **Fig. 2** Independent selection in domesticated traits between *C. melo* subsp. *melo* and
556 *agrestis*. **a,b**, Selection signals in domestication of *C. melo* subsp. *melo* (**a**) and *agrestis* (**b**)
557 populations on the twelve melon chromosomes. Horizontal dashed lines indicate the
558 genome-wide threshold of selection signals. Candidate genes previously reported or
559 identified in this study (red) and QTLs (black) that overlapped with selective sweeps are
560 marked. **c**, Overlapped genomic regions of QTLs for fruit weight, fruit diameter and flesh
561 thickness mapped by genetic analysis of an F_2 population from the cross of a wild and a
562 cultivated (MS-79) *agrestis* accession. **d,e**, Distribution of nucleotide diversity (π) of WA
563 and CA (**d**), and WM and CM (**e**) accessions. *MELO3C007596*, *MELO3C007597* and
564 *Cm-HMGR*³¹ were located within the *agrestis* and *melo* sweep regions, respectively. **f,g**,
565 QTL mapping for young fruit bitterness using two F_2 populations from the cross between
566 wild *melo* (**f**) and *agrestis* (**g**) with their corresponding cultivated accessions. F_{ST} values
567 of *CmBi*³² and *CmBt*³² genes between CM and CA are shown in the red vertical line. **h,i**,
568 qRT-PCR of *CmBi* (**h**) and *CmBt* (**i**) in young fruits of cultivated *melo* and *agrestis*
569 accessions. Data are presented as mean \pm s.d. ($n = 3$ independent experiments).

570 **j,k**, Manhattan plots of GWAS for flesh acidity in *melo* (**j**) and *agrestis* (**k**) populations.
571 *CmPH*³⁵ and *MELO3C011482* were identified residing within the association signals on
572 chromosomes 8 and 3, respectively.

573

574 **Fig. 3** Identification of a candidate gene for the melon sutures trait. **a**, Manhattan plots of
575 GWAS for fruit sutures in melon accessions. Fruits of representative sutured and
576 non-sutured melon accessions are shown. Scale bars represent 1.0 cm. **b**, Fine mapping of
577 melon fruit sutures using diverse segregating populations. An 86-kb interval harboring
578 four genes was identified (represented by arrows, of which the green one is the candidate
579 gene *MELO3C019694*). **c**, Expression of the four genes in flower and developed fruit of
580 different genotypes. **d**, Phylogenetic tree of *MELO3C019694* and its homologues in rice
581 (green points), Arabidopsis (red points), tomato (blue points), pumpkin (light blue points)
582 and melon (purple points). The closest homologues of *MELO3C019694* indicated in a
583 shadow box include those from Arabidopsis (*AT3G58780*, *AT2G42830*)⁴⁰ and tomato
584 (*Solyc07g055920.2.1*)⁴¹ that have been reported associated with pod dehiscence and
585 pericarp expansion, respectively. **e**, Identification of a 1.07-kb deletion upstream of the
586 *MELO3C019694* gene in accessions with sutures compared with non-sutured accessions.
587 **f**, Proportion of the 1.07-kb deletion (purple) existing in sutured and non-sutured melon
588 accessions. **g**, qRT-PCR analysis of *MELO3C019694* in the female flowers and young
589 fruits in PS (a non-sutured accession) and VED (a sutured accession). fl develop: flower
590 in development; closed fl: flower before anthesis; 0 DAP, 3 DAP and 7 DAP represent
591 fruits at 0, 3, 7 days after pollination. Data are presented as mean \pm s.d. (n = 3
592 independent experiments).

593

594

595 **Fig. 4** GWAS, BSA and QTL analyses identified the same region as being potentially
596 important for peel color. **a**, Phenotypes of green, white and yellow-peel melon accessions.
597 Scale bars represent 1.0 cm. **b**, Segregation of peel color in an F₂ population derived from
598 crossing a green-peel accession (MS-723) with a yellow-peel accession (B432). **c,d**,
599 Identification of overlapping intervals using GWAS and BSA analyses for the peel color
600 trait in melon. GWAS analyses (Manhattan plots) were performed using green, white and
601 yellow-peel melon accessions (**c**), and white and yellow-peel melon accessions (**d**),
602 respectively. BSA analyses (red lines) were conducted with the green and non-green
603 (white and yellow) bulks (**c**), and the white and yellow bulks (**d**) from the above F₂
604 population. Candidate gene in each signal is provided. The horizontal dashed lines
605 indicate the genome-wide threshold of GWAS signals (P -value = 2.51×10^{-6}). **e-g**, QTL
606 mapping using another F₂ population derived from crossing a green-peel accession and a
607 yellow-peel accession. Candidate genes (in green), *MELO3C003375* and *CmKFB*¹⁴, were
608 located in the intervals of the identified QTLs. **h-j**, qRT-PCR analysis of
609 *MELO3C003375*, *CmKFB*¹⁴ and *MELO3C003097* during fruit development in melon
610 accessions with different peel colors. Data are presented as mean \pm s.d. (n = 3
611 independent experiments)

612 **Materials and Methods**

613 **Plant materials and sequencing**

614 A diverse worldwide collection of 1,175 melon accessions and 9 from related species of
615 the *Cucumis* genus was obtained from NMGWM (National Mid-term Genebank for
616 Watermelon and Melon, Zhengzhou, China), ZFRI-CAAS (Zhengzhou Fruit Research
617 Institute, Chinese Academy of Agricultural Sciences), USDA (US Department of
618 Agriculture) and INRA (National Institute for Agricultural Research). Information about
619 the accessions, including individual name, country of origin, group, varieties identity and
620 resequencing data summary, is provided in Supplementary Table 1. Genomic DNA was
621 extracted from fresh young leaves using the cetyltriethylammonium bromide (CTAB)
622 method⁴⁹. At least 5 µg of genomic DNA was used for each accession to construct
623 sequencing libraries according to the manufacturer's instructions (Illumina Inc). The
624 libraries were sequenced on the Illumina HiSeq 2500 or HiSeq 3000 platform, generating
625 150-bp or 125-bp paired-end reads. Five F₂ populations were used in our study, of which
626 the genome of individuals of three F₂ populations were sequenced with 5 × depth, three
627 bulks developed from an F₂ population were sequenced with 15 × depth.

628

629 **Sequence alignment and variation calling**

630 To call SNPs, reads of all accessions were mapped to the melon reference genome²³
631 (version 3.5.1) using SOAP2 (ref. ⁵⁰) with the following parameters: -m 100 -x 888 -s 35 -l
632 32 -v 3. Mapped reads were filtered to remove PCR duplicates, assigned to chromosomes
633 and sorted according to the mapping coordinates. Both pair-end and single-end mapped
634 reads were used for SNP detection throughout the entire collection of melon accessions.

635

636 We identified possible SNPs for each accession relative to the reference using
637 SOApsnp⁵¹ with the following parameters: -L 150 -u -F 1. The likelihood of each
638 individual's genotype in glf format was then generated for each chromosome with SNP
639 quality ≥ 40 and base quality ≥ 40 .

640

641 To integrate SNPs across the entire collection, we called each SNP using GLFmulti⁵²
642 according to the maximum-likelihood estimation of site frequency. The core set of SNPs
643 was obtained by filtering on the base of allele frequency and the quality score given by
644 GLFmulti⁵². SNPs were further filtered using the following criteria: (i) one position with
645 more than two alleles was considered to be a polymorphic site in the population and was
646 excluded in the next analyses; (ii) the total sequencing depth should be $> 500 \times$ and $<$
647 $6,800 \times$ and the SNP quality value should be greater than 40; (iii) position with an average
648 mapping rate of reads of less than 1.5 were retained to rule out the effect of duplications;
649 and (iv) the nearest SNPs should be more than 1 bp away.

650

651 To obtain the final set of SNPs, we further performed filtering using segregation tests,

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

658

659 **Planting and phenotyping**

660 A total of 1175 melon accessions and 9 from related species were grown in Zhengzhou
661 (Henan province), Sanya (Hainan province) and Changji (Xinjiang province) in 2015 and
662 2016. Because of poor adaptation for some accessions, several traits were evaluated in only
663 one or two locations. Three replicates were performed at each location.

664

665 Three RIL populations^{39,53,54} and five F_2 populations were used to identify
666 candidates for sutures, peel color, flesh color, fruit bitterness, flesh thickness and sugar
667 content (Supplementary Table 17). We phenotyped for traits of fruit suture (as a
668 qualitative trait for presence/absence) or flesh color as a qualitative trait for yellow, green
669 or white at harvest. For flesh color, ripe fruit was cut in two longitudinal sections, and one
670 of them was used to evaluate flesh color visually and scanned to perform the flesh color
671 analysis in color spaces RGB and CIElab using the Tomato Analyzer 3.0 software^{55,56}.
672 Total chlorophyll and carotenoid contents were determined using UV-VIS Spectroscopy.
673 The other accessions and populations were grown in Xinxiang, Sanya or Beijing and
674 phenotyped following a Chinese technical specification for evaluating melon⁵⁷.

675

676 **Phylogenetic and population structure analyses**

677 A subset of 17,055 SNPs with a minor allele frequency (MAF) ≥ 0.05 and missing rate \leq
678 0.4 at four-fold degenerate sites representing neutral or near-neutral variants, were used for
679 phylogenetic and population structure analyses. The alignment was trimmed using
680 trimAl⁵⁸ (version 1.4. rev22v) in order to remove positions that are non-variable and
681 include more than 90% of gaps. The remaining sites were employed to construct the
682 phylogenetic tree with RAxML⁵⁹ (version 8.1.17) using the evolutionary model GTR.
683 The branch length and rate parameters were optimized, and the aLRT SH-like branch
684 support was calculated using PhyML v3.0 (ref. ⁶⁰) with the options '-b -4, -o lr'. The
685 same method was used to construct the phylogenetic tree of chloroplast genome.
686 Population structure was investigated using the program STRUCTURE⁶¹ (version 2.3.1),
687 with the same data used in the phylogenetic tree construction. Furthermore, to confirm
688 the result of STRUCTURE, we also performed the analysis using DAPC in R package
689 *adegenet* 2.1.0 (ref. ⁶²) with the parameter "max.n.clust = 40, PCs to retain = 900,
690 discriminant functions to retain = 5". In addition, principal component analysis (PCA)
691 using the whole-genome SNPs with missing value $\leq 40\%$ was performed with the
692 EIGENSOFT 6.0.1 (ref. ⁶³). Combining with the phylogenetic tree and principal

693 component analyses, we classified these accessions into three distinct clades (African
694 group, *melo* group and *agrestis* group). Considering the passport information, the *melo*
695 group and *agrestis* groups could be further divided into two sub-clades (wild *agrestis*,
696 cultivated *agrestis*; wild *melo* and cultivated *melo* groups), respectively.

697

698 **Identification of domestication sweeps**

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

711

712 **Identification of differentiated regions**

713 The population fixation statistics (F_{ST}) were estimated for 50-kb sliding windows with a
714 step size of 5 kb and for each SNP using a variance component approach implemented in
715 the HIERFSTAT R package⁶⁴. The average F_{ST} of all sliding windows was regarded as
716 the value at the whole-genome level across different groups. Sliding windows with the
717 top 10% highest F_{ST} values were selected initially. Neighboring windows were then
718 merged into one fragment. If the distance between two fragments was < 50 kb, fragments
719 were merged into one region. The final merged regions were considered as highly
720 diverged between different groups.

721

722 **Watterson estimator analysis**

723 The Watterson estimator of θ_w was evaluated using the software VariScan 2.0.3 (ref. ⁶⁵)
724 for four main sub-populations. SNPs with a minor allele frequency (MAF) $\geq 1\%$ within
725 each sub-population were used as the input data. A sliding window of 50 kb in length was
726 used to scan the whole genome. The average θ_w value of all windows in the genome was
727 then calculated to present the polymorphism.

728

729 **Linkage disequilibrium analysis**

730 Haploview software⁶⁶ was used to calculate LD values for each of the groups (WM, CM,
731 WA, CA) using SNPs with MAF ≥ 0.05 with the following parameters: -n -pedfile -info
732 -log -maxdistance 1000 -minMAF 0.05 -hwcutoff 0 -dprime -memory 10480. LD decay
733 was measured on the basis of the r^2 value and the corresponding distance between two

734 given SNPs.

735

736 **HPLC analysis of cucurbitacin B**

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

743

744 **Genome-wide association studies**

745 Only SNPs with minor allele frequency ≥ 0.05 and missing rate ≤ 0.4 in a population
746 were used to carry out GWAS. This resulted in 1,599,428, 872,244 and 2,028,259 SNPs
747 that were used in GWAS for subspecies of *melo*, *agrestis* and the entire population (*melo*
748 and *agrestis*), respectively. We performed GWAS using Efficient Mixed-Model
749 Association eXpedited (EMMAX) program⁶⁷. Population stratification and hidden
750 relatedness were modeled with a kinship (*K*) matrix in the emmax-kin-intel package of
751 EMMAX. The *P*-value thresholds for significance were approximately 2.51×10^{-6} .

752

753 **Bulked segregant analysis of F₂ population by whole-genome resequencing**

754 We planted 450 individuals of an F₂ population derived from a cross between MS-723
755 (green-peel accession) and B432 (yellow-peel accession) in the winter of 2017 in Sanya,
756 China. The fruit peel color of each individual was recorded. Genomic DNA was isolated
757 from fresh leaves using the CTAB method. For bulked segregant analysis, bulked DNA
758 samples were constructed by mixing equal amounts of DNA from 30, 29 and 9
759 individuals showing representative green, white and yellow peel color, respectively.
760 Roughly 13 × genome sequences for each of the two parents (B432 and MS-723) and 15
761 × data for each of the three bulked samples (green peel, white peel and yellow peel) were
762 generated. Short reads were aligned against the reference genome²³ (released 3.5.1) using
763 the Burrows-Wheeler Aligner (BWA)⁶⁸, and SNPs were identified using SAMtools⁶⁹. The
764 average SNP index for the green-peel bulk and non-green-peel bulk (white-peel bulk and
765 yellow-peel bulk), and white-peel bulk and yellow-peel bulk were calculated using a
766 1,000-kb sliding window with a step size of 100 kb.

767

768 **Expression analysis of candidate genes for peel color, flesh color and suture**

769 Fruits of MS-348 (yellow peel, orange flesh), MS-531 (white peel, white flesh) and
770 MS-982 (green peel, green flesh) were sampled at 20, 25, 30, 35 and 40 days after
771 pollination, respectively. Total RNA was extracted using RNAPrep Pure plant kit
772 (TIANGEN Biotech). The first-strand cDNA synthesis was conducted following
773 SuperScript RT Mix (Bio-Connect Biotech). Then 2-µl cDNA was used to preform
774 qRT-PCR in a 10-µl reaction mixture. We conducted the expression of *MELO3C003375*,

775 *CmKFB* and *MELO3C003097* in the peel of MS-348, MS-531 and MS-982, and the
776 expression of *MELO3C003097* in flesh of MS-982 and MS-531. Expression of *CmBi* and
777 *CmBt* were performed in fruits at 7 days after anthesis of cultivated *melo* and *agrestis*
778 accessions. For suture, we harvested MS-1152 (sutured line) and Piel de sapo T111
779 (non-sutured line) at Fl-develop (7 mm-female flower), closed-Fl (10 mm-female flower)
780 and 0, 3, 7 days after pollination, respectively, and calculated the expression of the
781 candidate *MELO3C019694*. Three replicates were performed for each experiment.
782 Relative expression levels were calculated by the $2^{-\Delta\Delta Ct}$ or $2^{-\Delta Ct}$ method.

783

784 **Reporting Summary.**

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

787

788 **Statistical analysis**

789 Chi-square test statistic and standard deviation (stdev) were performed with the SPSS
790 software. The significance was determined by two-tailed Student's t tests.

791

792 **Code availability**

793 All codes are available from the corresponding author upon request.

794

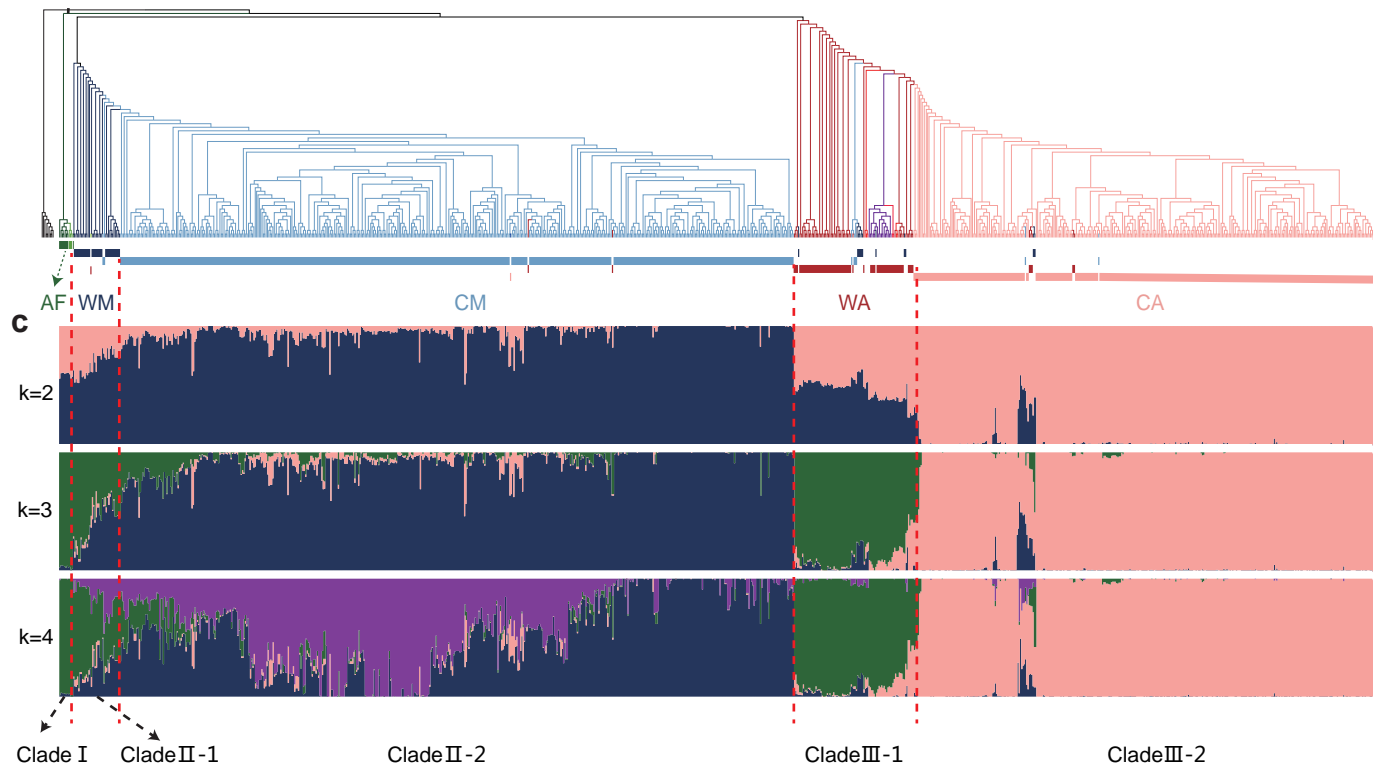
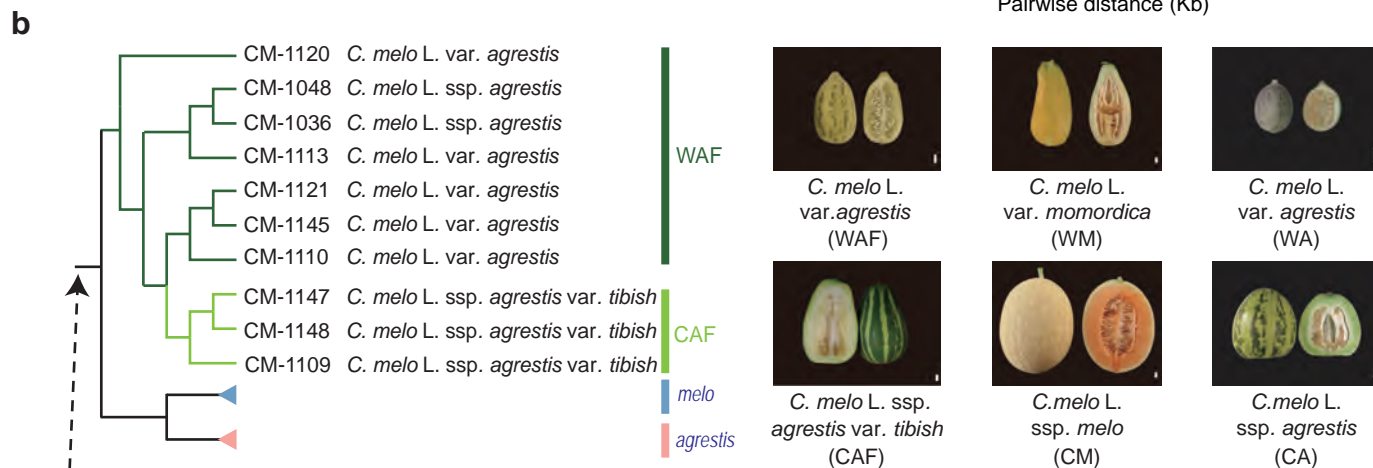
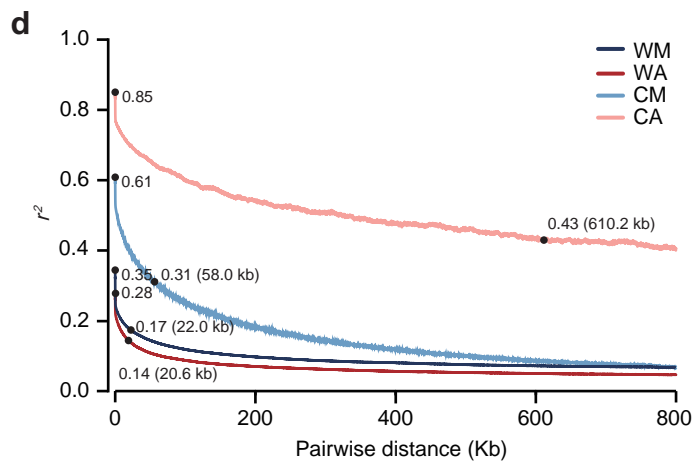
795 **Data Availability**

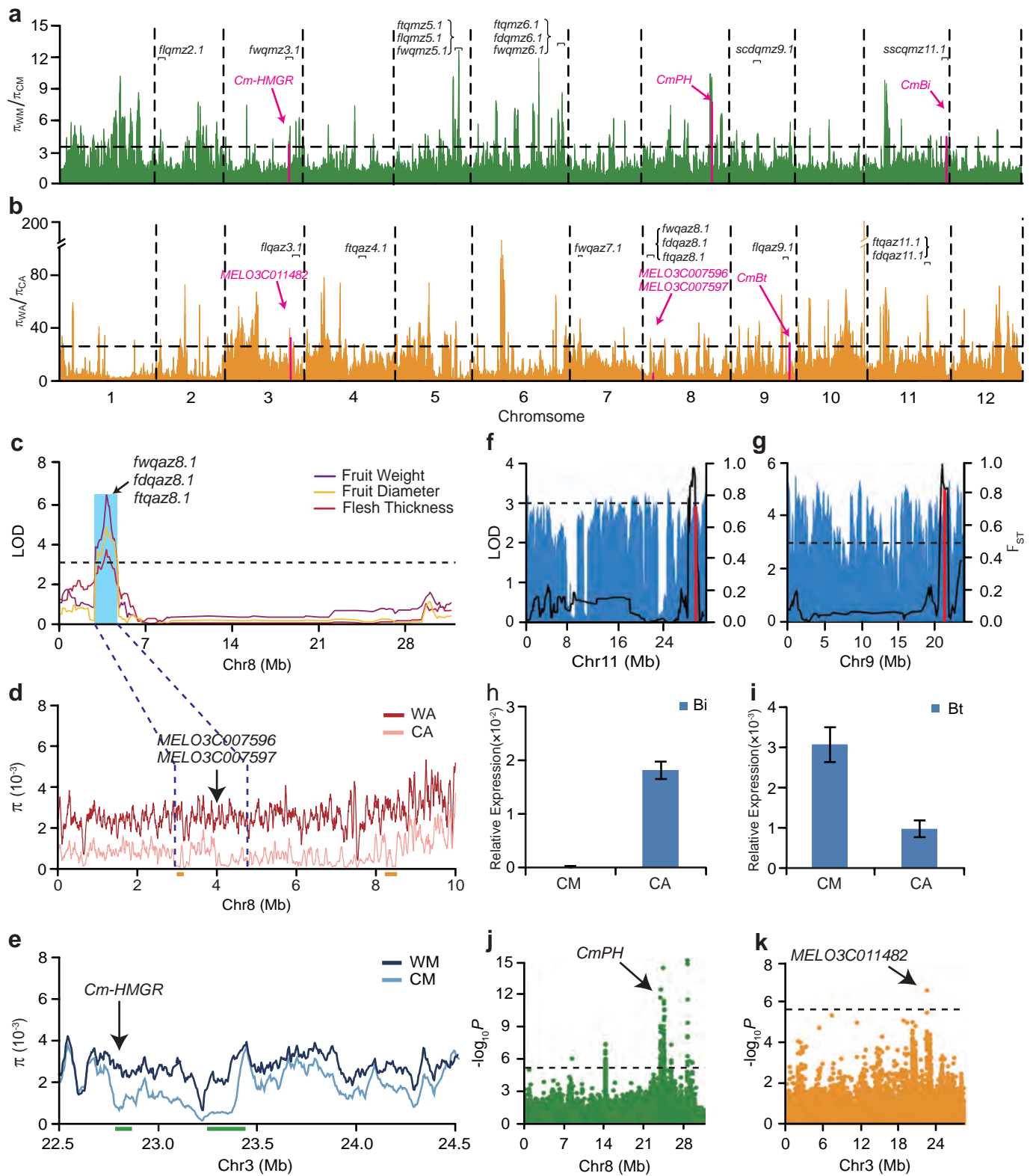
796 The raw sequence data reported in this paper has been deposited in the Sequence Read
797 Archive (SRA) under accession PRJNA565104 that are publicly accessible.

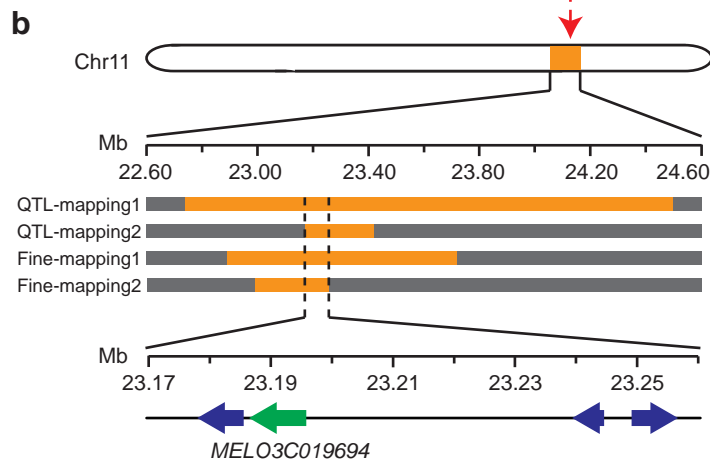
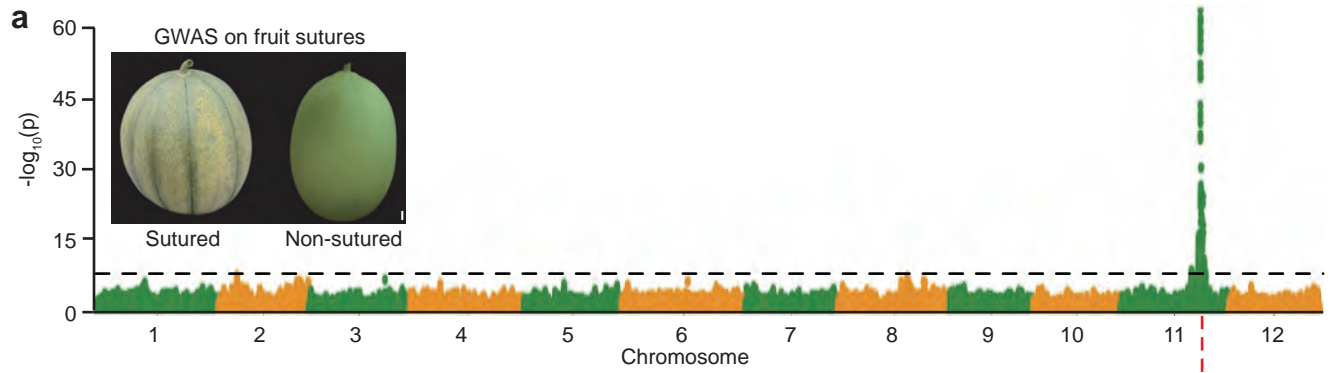
798 **References**

- 799 49 Gawel, N. J. & Jarret, R. L. A modified CTAB DNA extraction procedure for *Musa*
800 and *Ipomoea*. *Plant molecular biology* **9**, 262-266 (1991).
- 801 50 Li, R. et al. SOAP2: an improved ultrafast tool for short read alignment.
802 *Bioinformatics* **25**, 1966-1967 (2009).
- 803 51 Li, R. et al. SNP detection for massively parallel whole-genome resequencing.
804 *Genome Res.* **19**, 1124-1132 (2009).
- 805 52 He, W. et al. ReSeqTools: an integrated toolkit for large-scale next-generation
806 sequencing based resequencing analysis. *Genet. Mol. Res.* **12**, 6275-6283 (2013).
- 807 53 Harel-Beja, R. et al. A genetic map of melon highly enriched with fruit quality QTLs
808 and EST markers, including sugar and carotenoid metabolism genes. *Theor. Appl.*
809 *Genet.* **121**, 511-533 (2010).
- 810 54 Gur, A. et al. Genome-wide linkage-disequilibrium mapping to the candidate gene
811 level in melon (*Cucumis melo*). *Sci. Rep.* **7**, 9770 (2017)
- 812 55 Brewer, M. T. et al. Development of a controlled vocabulary and software application
813 to analyze fruit shape variation in tomato and other plant species. *Plant Physiol.* **141**,
814 15-25 (2006).
- 815 56 Darrigues, A., Hall, J., Knaap, E. & Francis, D. M. Tomato analyzer-color test: a new
816 tool for efficient digital phenotyping. *J. Amer. Soc. Hort. Sci.* **133**, 579-586 (2008).
- 817 57 Ma, S. et al. *Descriptors and data standard for melon (Cucumis melo L.)* (China
818 Agriculture press, Beijing, 2006).
- 819 58 Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. TrimAl: a tool for
820 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*
821 **25**, 1972-1973 (2009).
- 822 59 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis
823 of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
- 824 60 Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood
825 phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307-321 (2010).
- 826 61 Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using
827 multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**,
828 1567-1587 (2003).
- 829 62 Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers.
830 *Bioinformatics* **24**, 1403-1405 (2008).
- 831 63 Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS*
832 *Genet.* **2**, e190 (2006).
- 833 63 Goudet J. Hierfstat, a package for R to compute and test hierarchical *F*-statistics.
834 *Molecular Ecology Notes* **5**, 184-186 (2005).
- 835 64 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization
836 of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).
- 837 65 Hutter, S., Vilella, A. J. & Rozas, J. Genome-wide DNA polymorphism
838 analyses using VariScan. *BMC Bioinformatics* **7**, 409 (2006).

- 839 66 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization
840 of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).
- 841 67 Kang, H. M. et al. Variance component model to account for sample structure in
842 genome-wide association studies. *Nat. Genet.* **42**, 348-354 (2010).
- 843 68 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
844 transform. *Bioinformatics* **25**, 1754-1760 (2009).
- 845 69 Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
846 2078-2079 (2009)







c

Gene ID	Description	TMM-normalized	
		Flower	Fruit
<i>MELO3C019693</i>	RNA recognition motif domain	0	0
<i>MELO3C019694</i>	AGAMOUS MADS box transcription factor	912.4	2297.3
<i>MELO3C019695</i>	NA	0	0
<i>MELO3C019696</i>	HVA22-like protein	225.6	58.8

more expression

