

Watermelon domestication was shaped by stepwise selection and regulation of the metabolome

Pingli Yuan^{1†}, Congping Xu^{2†}, Nan He^{1†}, Xuqiang Lu¹, Xingping Zhang³, Jianli Shang¹, Hongju Zhu¹, Chengsheng Gong¹, Hanhui Kuang⁴, Tang Tang⁵, Yong Xu⁶, Shuangwu Ma¹, Dexi Sun¹, Weiqin Zhang⁵, Muhammad J. Umer¹, Jian Shi⁵, Alisdair R. Fernie⁷, Wenge Liu^{1*} & Jie Luo^{2,5,8*}

¹Henan Joint International Research Laboratory of South Asian Fruits and Cucurbits, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450009, China;

²Hainan Yazhou Bay Seed Laboratory, Sanya Nanfan Research Institute of Hainan University, Sanya 572025, China;

³Peking University Institute of Advanced Agricultural Sciences, Weifang 261325, China;

⁴College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan 430070, China;

⁵Wuhan Metware Biotechnology Co., Ltd., Wuhan 430070, China;

⁶National Watermelon and Melon Improvement Center, Beijing Academy of Agricultural and Forestry Sciences, Beijing 100097, China;

⁷Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm 144776, Germany;

⁸College of Tropical Crops, Hainan University, Haikou 572208, China

Received July 9, 2022; accepted September 16, 2022; published online November 4, 2022

Although crop domestication has greatly aided human civilization, the sequential domestication and regulation of most quality traits remain poorly understood. Here, we report the stepwise selection and regulation of major fruit quality traits that occurred during watermelon evolution. The levels of fruit cucurbitacins and flavonoids were negatively selected during speciation, whereas sugar and carotenoid contents were positively selected during domestication. Interestingly, fruit malic acid and citric acid showed the opposite selection trends during the improvement. We identified a novel gene cluster (*CGC1*, cucurbitacin gene cluster on chromosome 1) containing both regulatory and structural genes involved in cucurbitacin biosynthesis, which revealed a cascade of transcriptional regulation operating mechanisms. In the *CGC1*, an allele caused a single nucleotide change in *CIERF1* binding sites (GCC-box) in the promoter of *CIBH1*, which resulted in reduced expression of *CIBH1* and inhibition of cucurbitacin synthesis in cultivated watermelon. Functional analysis revealed that a rare insertion of 244 amino acids, which arose in *C. amarus* and became fixed in sweet watermelon, in *CIOSC* (oxidosqualene cyclase) was critical for the negative selection of cucurbitacins during watermelon evolution. This research provides an important resource for metabolomics-assisted breeding in watermelon and for exploring metabolic pathway regulation mechanisms.

metabolome, mGWAS, gene cluster, domestication, cucurbitacin biosynthesis

Citation: Yuan P, Xu C, He N, Lu X, Zhang X, Shang J, Zhu H, Gong C, Kuang H, Tang T, et al. (2023). Watermelon domestication was shaped by stepwise selection and regulation of the metabolome. *Sci China Life Sci* 66, 579–594. <https://doi.org/10.1007/s11427-022-2198-5>

†Contributed equally to this work

*Corresponding authors (Jie Luo, email: jie.luo@hainanu.edu.cn; Wenge Liu, email: liuwenge@caas.cn)

INTRODUCTION

Watermelon (*Citrullus lanatus*; 2n=2×=22) is native to Africa (Robinson and Decker-Walters, 1997) and is among the top five most consumed flesh fruits worldwide, with a global production of 101.62 million tonnes in 2020 (FAO-STAT; <http://faostat.fao.org>). Watermelon fruits are rich in health-promoting compounds and are considered a valuable nutritional source for the human diet (Saminathan et al., 2018; Wechter et al., 2008). For watermelon and many fleshy fruits, fruit quality traits such as flesh taste and appearance are very important, which are both ultimately determined by metabolite composition.

The plant metabolome can be regarded as both a readout of the physiological status and a bridge spanning the genome-phenome divide (Luo, 2015). High-throughput metabolic profiling and its integration with other omics tools, especially in association with deep genome sequencing, has recently become a powerful forward genetics strategy to reveal the genetic and biochemical basis of plant metabolism in a range of species (Chen et al., 2014; Chen et al., 2022; Richter et al., 2016; Tieman et al., 2017; Wen et al., 2014; Zeng et al., 2020; Zhou et al., 2019; Zhu et al., 2018). These metabolomic and multi-omics studies have provided valuable insights into the metabolic breeding history of crop plants (Wang et al., 2022; Xu et al., 2019; Zhu et al., 2018) and have led to metabolomic-based classification (Moing et al., 2020). It is anticipated that they could also aid in designing breeding strategies aimed at selecting valuable compounds and agronomic traits (Chen et al., 2016; Chen et al., 2020; Fernie and Schauer, 2009; Gong et al., 2020; Jin et al., 2021; Riedel-sheimer et al., 2012).

Metabolites are very meaningful in fleshy fruits because they determine taste, quality and nutritional properties. A few studies have been performed regarding the functional interpretation of metabolic pathway genes in watermelon. The *CIAGA2*, *CISWEET3*, and *CITST2* were identified to be related to sugar accumulation in watermelon fruits (Ren et al., 2018; Ren et al., 2021). The elevated chloroplast-localized phosphate transporter *CIPHT4;2* expression level is necessary for carotenoid accumulation and flesh color formation (Zhang et al., 2017). The decreased level of lycopene β -cyclase protein results in the appearance of red flesh (Zhang et al., 2020a). The flesh color and sweetness may have co-developed during watermelon development and domestication (Guo et al., 2019). In addition, the cucurbitacin pathway has been analyzed using a comparative genomic strategy in cucumber, melon and watermelon, revealing an important transcription factor (Bt/BI) and a gene cluster (Ma et al., 2022; Xu et al., 2022; Zhou et al., 2016). However, the global watermelon metabolome and the mechanisms that regulate it have yet to be identified in watermelon.

Here, we performed comprehensive broadly targeted metabolic profiling and metabolite-GWAS (mGWAS) using a worldwide collection of 204 watermelon accessions. We revealed a stepwise selection of flesh taste/appearance-related metabolites associated with watermelon fruit quality breeding. This study provides novel insights into the domestication and regulation of major fruit quality-related metabolic traits for shaping crop improvement and provides considerable resources for both watermelon genetic improvement and plant metabolic biology.

RESULTS

Metabolic profiling of watermelon fruits

We selected 204 representative watermelon accessions from various geographic regions around the world (Table S1 in Supporting Information) for metabolite quantification using a widely targeted liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based metabolomics approach (Chen et al., 2013). A total of 3,730 metabolic features were detected in watermelon fruit, among which 346 were annotated, including cucurbitacins, carbohydrates, amino acids and their derivatives, organic acids, flavonoids, nucleic acids and their derivatives, vitamins, and lipids (Table S2 and Figure S1A in Supporting Information). We found that 50.3% of these metabolic features displayed broad-sense heritability (H^2) values greater than 0.5 (Figure S1B in Supporting Information). Additionally, the coefficients of variation (CVs) of all metabolites were less than 30% in 12 quality control samples (QC); in contrast, CVs for 87.1% of the metabolites were greater than 50% (Figure S1C and Table S2 in Supporting Information), suggesting significant variation of metabolites in fruits of different watermelon accessions. Among them, the cucurbitacins in fruit showed the highest CVs, with an average of 507.7%, ranging from 69.3% to 929.5%, while alcohols and polyols showed the lowest CVs, with an average of 33.4% (Table S2 in Supporting Information).

Principal component analysis (PCA) based on the levels of all metabolic features in fruit grouped the watermelon accessions into six distinct groups, represented by *C. colocynthis* (CC), *C. amarus* (CA), *C. mucospermus* (CM), *C. lanatus* edible seed watermelon (CL_ES), *C. lanatus* land-race watermelon (CL_LR) and *C. lanatus* improved watermelon (CL_IM) (Figure 1 A). Further clustering analyses revealed that CC and CA formed independent clades, while CL_ES, which has the characteristics of soft, juicy flesh and giant seeds, formed a subclade with CM (Figure S2 in Supporting Information). CL_ES is commercially grown in northwest China for seed consumption (Levi et al., 2017; Zhang and Jiang, 1990). Clustering analyses using metabolic data displayed a trend similar to that based on genome-wide

SNPs but with a unique grouping of the different watermelon types, wherein CL_ES could be discriminated from CL_LR (Figure 1B; Figure S2 in Supporting Information). The fruit metabolite accumulation pattern of this genotype lies between that of CM and CL_LR (Figure 1A; Figure S2 in Supporting Information). The results of clustering by metabolic features are in close accordance with the three major evolutionary stages of watermelon: speciation, domestication, and improvement (Guo et al., 2019).

To study the possible reshaping of the *Citrullus* metabolome during evolution, we searched for differential metabolites during the processes of watermelon speciation, domestication, and improvement (Table S3–S10 in Supporting Information), focusing on the identified/annotated metabolites. We observed stage-related differentiation of specific metabolites during these processes: first, the significant decrease and even “loss” during speciation and domestication, respectively, of the triterpenes cucurbitacin, flavonoids, and polysaccharides, followed by the substantial increase in mono- and di-saccharides and carotenoids during domestication and finally by the significant decrease in malic

acid and increase in citric acid during the improvement stage (Figure 1C). Taken together, we suggest a consequent selection of flesh taste- and appearance-related metabolites during watermelon evolution and the genetic improvement process.

Genetic basic of natural variation in metabolic traits

To uncover the genetic basis of important metabolic traits in fruit, mGWAS was performed using a linear mixed model (LMM). The LMM resulted in fewer false positives by taking into account the genome-wide patterns of genetic relatedness, and we used these results for further analysis. We used $P_{LMM}=1.16\times 10^{-7}$ as the genome-wide significance threshold for the watermelon population after Bonferroni correction (Table S11 in Supporting Information). mGWAS hot spots were located mainly on chromosome 1, chromosome 2, chromosome 4, and chromosome 9 (Table S11 and Figure S3 in Supporting Information). Interestingly, we observed a wide range of co-mapping of metabolites both from within the same and different metabolite classes (Table S11

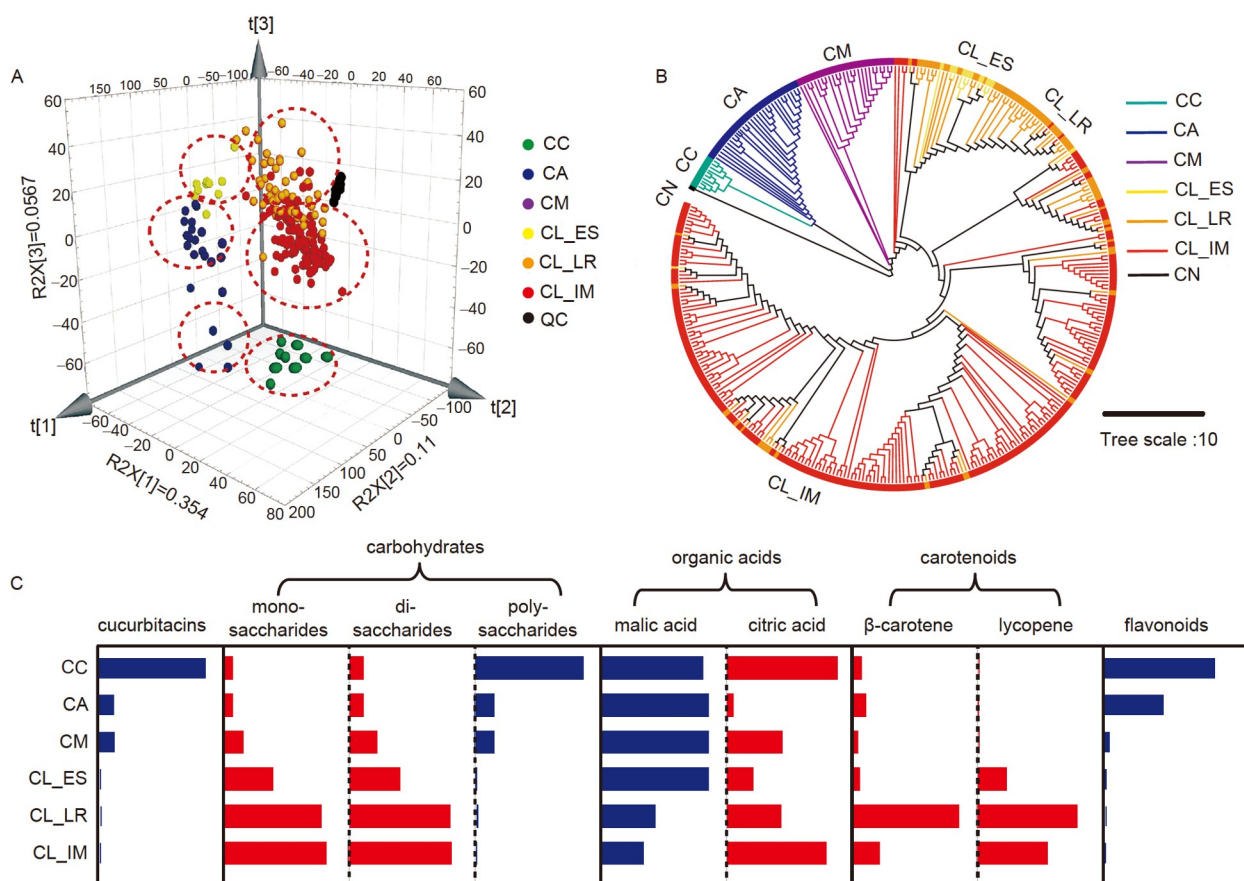


Figure 1 Phylogeny and metabolic profiling of watermelon accessions. A, Principle component analysis of watermelon accessions based on metabolome profiles. Different groups were circled with dash dot lines accordingly. B, Neighbor-joining phylogenetic tree of watermelon accessions constructed using genome-wide SNPs. Colors of branches in the tree indicate different species. C, Distribution of the contents of important metabolites in the six types of watermelon. Each column represents the average content within each watermelon type. The blue and red columns represent decreasing or increasing metabolite contents during watermelon evolutionary process (from wild to cultivated). CN, *C. naudinianus*; CC, *Citrullus colocynthis*; CA, *C. amarus*; CM, *C. mucospermus*; CL_ES, *C. lanatus* edible seed watermelon; CL_LR, *C. lanatus* landrace; CL_IM, *C. lanatus* improved.

in Supporting Information). For example, flavonoids in different subgroups, including flavone, flavonol and flavone C-glycosides, co-mapped to 12.2 Mb on chromosome 6, 27.6 Mb on chromosome 7, 32.5 Mb on chromosome 9, and 25.4 Mb on chromosome 10. Metabolites in different classes, such as cucurbitacin, co-mapped with organic acids, carbohydrates, amino acid derivatives and flavanone at 32.4 Mb on chromosome 9. These observations implied that these regions can control the synthesis of various metabolites.

The important metabolic traits were reanalyzed by multi-locus GWAS using FarmCPU software to reduce false positive and false negative signals and obtain credible results (Liu et al., 2016) (Table S12 in Supporting Information). These reliable signals mapped by two analysis methods were used for the screening of candidate genes. Candidate genes were mined by a combination of bioinformatics and bio-

chemical approaches: (i) searching for a protein or protein cluster that was biochemically related to the associated metabolic trait encoded at these loci; (ii) performing cluster analysis of candidate genes relative to homologous genes with known functions; and (iii) cross-referencing with statistical results of causal genes from multi-locus GWAS using FarmCPU. Finally, 29 candidate genes were identified among 20 significant association loci underlying the variation in metabolites that are regulators of fruit taste, resistance quality or human nutritional importance (Table 1).

Selection and regulation of cucurbitacins during watermelon evolution

Cucurbitacins and flavonoids were negatively selected during watermelon evolution. Flavonoids (including flavonoid

Table 1 The full list of 29 candidate genes from mGWAS results

Locus	Metabolites	Causative SNPs	Effect	<i>P</i> value	Candidate gene	Description
L1	Nicotinic acid-hexoside	10:25998301	UTR_3_PRIME	1.2×10^{-10}	Cla017767	UDP-glycosyl transferase
L1	Methyl nicotinate	10:25931062	Non-Synonymous SNP	3.1×10^{-7}	Cla017759	O-methyl transferase
L2	D(+)-Melezitose O-rhamnoside	7:10308825	Non-Synonymous SNP	5.0×10^{-8}	Cla011774	Membrane transporter
L2	Trehalose	7:10438076	Non-Synonymous SNP	4.0×10^{-8}	Cla011768	Membrane transporter
L3	Luteolin-O-hexoside	4:346052	Non-Synonymous SNP	2.6×10^{-8}	Cla000744	Cytochrome P450
L4	8-C-p-hydroxybenzoylLuteolin-β-D-glucoside	6:3090248	Non-Synonymous SNP	6.1×10^{-7}	Cla006682	Flavanone 3-hydroxylase
L5	Luteolin 6-C-glucoside	1:2358146	Synonymous	4.2×10^{-9}	Cla008060	Zinc finger protein
L6	Citric acid	11:9068215	UTR_3_PRIME	5.2×10^{-9}	Cla007935	Auxin efflux carrier
L7	2-Hydroxyisocaproic acid	8:21448038	Intron	9.2×10^{-7}	Cla022199	NAD-malic enzyme
L8	Cucurbitacin B/D/I-Hex	1:2693392	Synonymous	0	Cla011464	Glycosyl transferase
L8	Cucurbitacin B/D/E/I/T	1:2925085	Non-Synonymous SNP	0	Cla011487	ERF
L8	Cucurbitacin B/D/E/I/T	1:2952640	Non-Synonymous SNP	0	Cla011488	ERF
L8	Cucurbitacin B/D/E/I/T	1:3168674	Non-Synonymous SNP	4.0×10^{-8}	Cla011508	bHLH
L8	Cucurbitacin B/D/E/I/T	1:3271685	Synonymous	3.4×10^{-12}	Cla011510	bHLH
L8	Cucurbitacin B/D/E/I/T	1:3300039	Non-Synonymous SNP	5.1×10^{-14}	Cla011514	Cytochrome P450
L8	Cucurbitacin B/D/E/I/T	1:3311197	Non-Synonymous SNP	1.0×10^{-8}	Cla011515	Cytochrome P450
L9	Lycopene	4:8886977	Non-Synonymous SNP	5.40×10^{-10}	Cla005011	Lycopene beta-cyclase
L10	Cyanidin 3-O-glucoside	2:20007972	Non-Synonymous SNP	2.1×10^{-10}	Cla000767	MYB
L11	5-O-Caffeoylquinic acid	2:8454940	Intron	6.9×10^{-11}	Cla006745	Cytochrome P450
L11	3-O-Caffeoylquinic acid	2:8533020	Introduce stop codon	3.8×10^{-14}	Cla006748	Cytochrome P450
L12	Phytoxanthin	5:11475678	Non-Synonymous SNP	1.4×10^{-8}	Cla013139	Zinc finger protein
L13	Apo-beta-8'-carotenal	2:352440	Non-Synonymous SNP	1.7×10^{-19}	Cla007686	Zinc finger protein
L14	Chrysoeriol O-malonylhexoside	5:13501387	Non-Synonymous SNP	1.1×10^{-7}	Cla013238	O-malonyl transferase
L15	L-Isoleucine	11:54975899	Synonymous	1.1×10^{-8}	Cla011987	Lysine transporter
L16	Quercetin 3-O-glucoside	7:937585	Non-Synonymous SNP	6.9×10^{-29}	Cla002228	UDP-glycosyl transferase
L17	4-Hydroxy-L-glutamic acid	5:4149324	Splicing	2.3×10^{-21}	Cla021561	N-carbamoyl-L amino acid hydrolase
L18	Cucurbitacin E/I	6:12688028	Non-Synonymous SNP	4.2×10^{-16}	Cla004098	bHLH
L19	Cucurbitacin D	6:26980503	Non-Synonymous SNP	1.4×10^{-8}	Cla019330	DS synthase
L20	Cucurbitacin B/D/E/I/T	9:32676489	Synonymous	3.12×10^{-14}	Cla004749	DS synthase

C-glycosides and flavone/flavonol) were selected during the watermelon speciation and domestication stages, and low-level flavonoid alleles were found to be significantly enriched in cultivated watermelon (Figure S4 in Supporting Information). More interestingly, we observed the predominantly negative selection of cucurbitacins during speciation from CC to CA or CC to CM, and a relatively weaker selection during domestication from CM to CL (Figure 2A; Table S13 in Supporting Information). Cucurbitacins, as bitter compounds, are highly oxygenated tetracyclic triterpenes produced in the Cucurbitaceae family. A total of 42

cucurbitacins were found in the identified/annotated metabolites, including cucurbitacin-B (CUB), cucurbitacin-D (CUD), cucurbitacin-E (CUE), cucurbitacin-I (CUI), and cucurbitacin-T (CUT) (Table S2 in Supporting Information).

The mGWAS and genomic region selection analysis were combined to understand cucurbitacin selection at different evolutionary stages of watermelon. Eight major loci (one each on chromosomes 1, 2, 3, 4, 6, and 9) for 42 cucurbitacins were identified by mGWAS (Figure 2B; Figures S5A, S6A and D in Supporting Information). The frequencies of the most highly associated SNPs were located in *cuc3*,

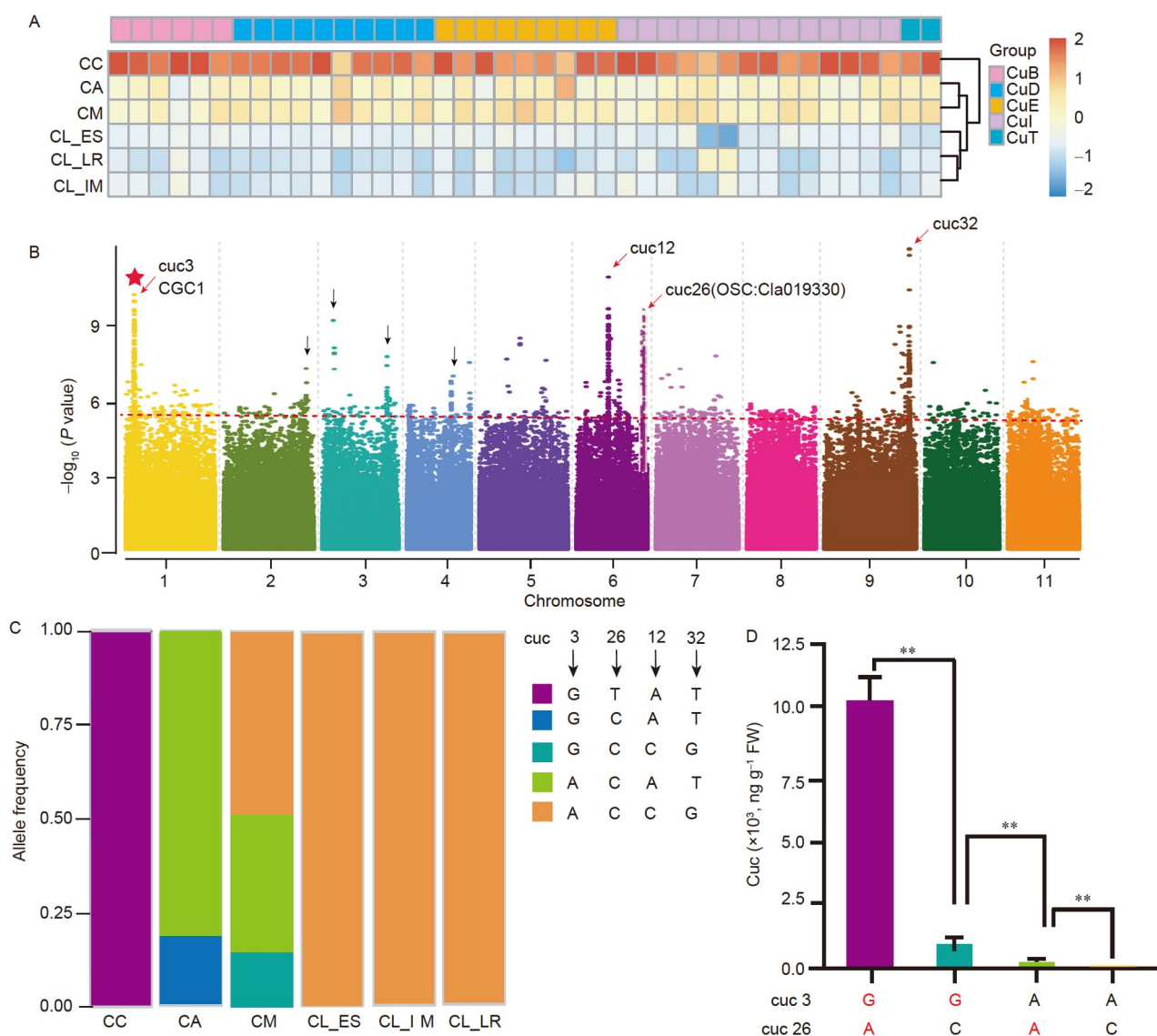


Figure 2 Evolution and domestication of cucurbitacins in watermelon. A, Heatmap of contents of cucurbitacins detected in this study. The relative values of cucurbitacin contents were scaled from CC to cultivars for each metabolite. B, Genomic distribution of major signals associated with cucurbitacin contents. Red arrows denote signals located within sweep regions. Black arrows indicate those outside sweep regions. *cuc3*, the cucurbitacin-related locus at ~3 Mb on chromosome 1; the rest abbreviations are deduced from this. C, Allele frequencies of the four SNP loci (the most significantly associated SNPs in *cuc3*, *cuc12*, *cuc26*, and *cuc32*, respectively) in the six watermelon groups. D, Cucurbitacin contents in accessions with different allele combinations of *cuc3* and *cuc26*, those two loci mainly control the accumulation of cucurbitacins according to the results of Figure S7 in Supporting Information. Allele combinations of less than 5 in watermelon population are not shown. ** indicates a significant difference in cucurbitacin contents as determined by Student's *t*-test ($P < 0.01$).

cuc12, cuc26, and cuc32 of the six watermelon groups (Figure 2C), suggesting that these four major signals were selected during watermelon evolution. Interestingly, we found that cuc3 and cuc12 underwent strong selection (primitive genotypes were almost completely replaced) from CC to CM, whereas there was weaker selection (primitive genotypes were only partially replaced) from CC to CA or from CM to CL (Table S13 in Supporting Information). In addition, the pairwise epistatic interactions between the four significant loci were calculated based on the average content of cucurbitacins. The results indicated that these four loci may act sequentially in controlling cucurbitacin content, and cuc3 and cuc26 were identified as two major effective loci (Figure 2D; Figure S7 in Supporting Information).

The strong selective sweep cuc3 overlapped with the previously reported locus for fruit bitterness identified from the watermelon recombinant inbred lines (RIL) population (Li et al., 2018). In this overlapping region, *CIBh1* (Cla011508) is a transcription factor that regulates cucurbitacin accumulation in watermelon (Li et al., 2018; Zhou et al., 2016). In addition, within close proximity to *CIBh1*, we observed two genes annotated as cytochrome P450 (CYP76b1: Cla011514 and CYP76b2: Cla011515). Phylogenetic analysis revealed that the two P450s have high sequence similarity with the cucurbitacin synthesis genes CYP712D8/CYP81Q58 in cucumber (Shang et al., 2014) (Figure S8 in Supporting Information). The spatiotemporal expression patterns of the related and colocalized genes in different tissues from colocynth and cultivated watermelon were measured, and the results showed that seven genes, including two *CIERF* genes (*CIERF1*: Cla011487 and *CIERF2*: Cla011488), one glycosyltransferase gene (*Clgt1*: Cla011464), two *bHLH* genes (*CIBh1*: Cla011508 and *CIBh2*: Cla011510), and two P450 genes (CYP76b1: Cla011514 and CYP76b2: Cla011515), were highly co-expressed in tissues that accumulated high levels of cucurbitacins (Figure 3A and B). In addition, these seven co-expressed genes were significantly upregulated in watermelon under ABA, drought, or cold treatment (Figure 3C). Thus, we suspected that these genes, including both transcription factors and structural genes within a 70-kilobase region in cuc3 (Figure 3A), may form a cluster involved in cucurbitacin biosynthesis, hereafter referred to as the cucurbitacin gene cluster on chromosome 1 (*CGC1*).

In this putative gene cluster, *CIERF1* and *CIERF2* were shown to belong to the APETALA2 (*AP2*) subfamily proteins according to phylogenetic and gene conserved domain analysis (Figures S9 and S10 in Supporting Information). These two proteins all localized to the nucleus in *N. benthamiana* leaf epidermal cells, as expected for *CIERF1* and *CIERF2* as transcriptional regulators (Figure S11 in Supporting Information). Transient overexpression of either of the *CIERFs* activated transcription of the two *CIBhs* within

the cluster and resulted in an overaccumulation of cucurbitacins in bitter watermelon (Figure S12A in Supporting Information). VIGS results showed that the expression of the two *CIBhs* and the accumulation of cucurbitacins were significantly decreased in silenced-PV190_{*CIERFs*} plants (Figure S13 in Supporting Information). These results suggested that *CIERFs* may regulate the expression of *bHLHs* and the biosynthesis of cucurbitacins. Further investigation showed that *CIERFs* indeed could directly bind to the promoter of *CIBhs* in Y1H assays and luciferase transactivation assays (LUC) (Figure S12B and C in Supporting Information).

The nuclear localization of *CIBh1* and *CIBh2* suggests that they function as transcription factors (Figure S11 in Supporting Information). Phylogenetic analysis indicated that the two *bHLHs* are members of clade Ib and are orthologs of the *CsBt* (*Csa5G157230*) and *CsBl* (*Csa5G156220*) genes, which are responsible for the regulation of cucurbitacin synthesis in cucumber (Figure S14A in Supporting Information) (Shang et al., 2014). Interestingly, the *CIERF1* expression level in colocynth was only ~1.4-fold higher than that in cultivated watermelon, and there was no significant difference between colocynth and cultivated watermelon (*P*-value=0.065, *t*-test). The expression level of *CIBh1* showed an ~100-fold difference in the two types of watermelon (Figure 3D). Overexpression of *CIERF1* increased the RNA level of *CIBh1* and promoted the biosynthesis of cucurbitacins in the fruit of colocynth but not in cultivated watermelon (Figure 3E and F). Further co-overexpression of *CIERF1* and *CIBh1* cloned from colocynth watermelon improved the content of cucurbitacins by ~42-fold in cultivated watermelon (Figure 3G). Based on these data, we hypothesized that *CIERF1* might bind the *CIBh1* promoter differentially between colocynth and cultivated watermelon due to the different transcript levels of *CIBh1* in the two types of watermelon, resulting in a reduction in the accumulation of cucurbitacins in cultivated watermelon but not in colocynth. Analysis of the *CIBh1* promoter region revealed a single nucleotide mutation in the GCC-box (GCCGCC to GTCGCC, abbreviated as *CIBh1mut*), a *cis*-element to which *ERFs* can directly bind, in cultivated watermelon (Figure 3H). We then performed dual-luciferase assays and yeast one-hybrid (Y1H) assays and demonstrated that the C→T change in the promoter significantly reduced *CIBh1mut* inducibility in comparison to the wild-type *CIBh1* (Figure 3I–K). The EMSAs demonstrated that *CIERF1* binds to GCCGCC fragments within the promoter of wild-type *CIBh1*, but mutation in the GCC-box resulted in significantly reduced promoter activity of *CIBh1mut* (Figure 3L). Together, these results clearly show that the *CIERF1* protein binds to the GCC-box in *CIBh1* promoter and has a much higher affinity for colocynth *CIBh1* than the cultivated watermelon *CIBh1mut*.

Y1H assays and luciferase transactivation assays revealed

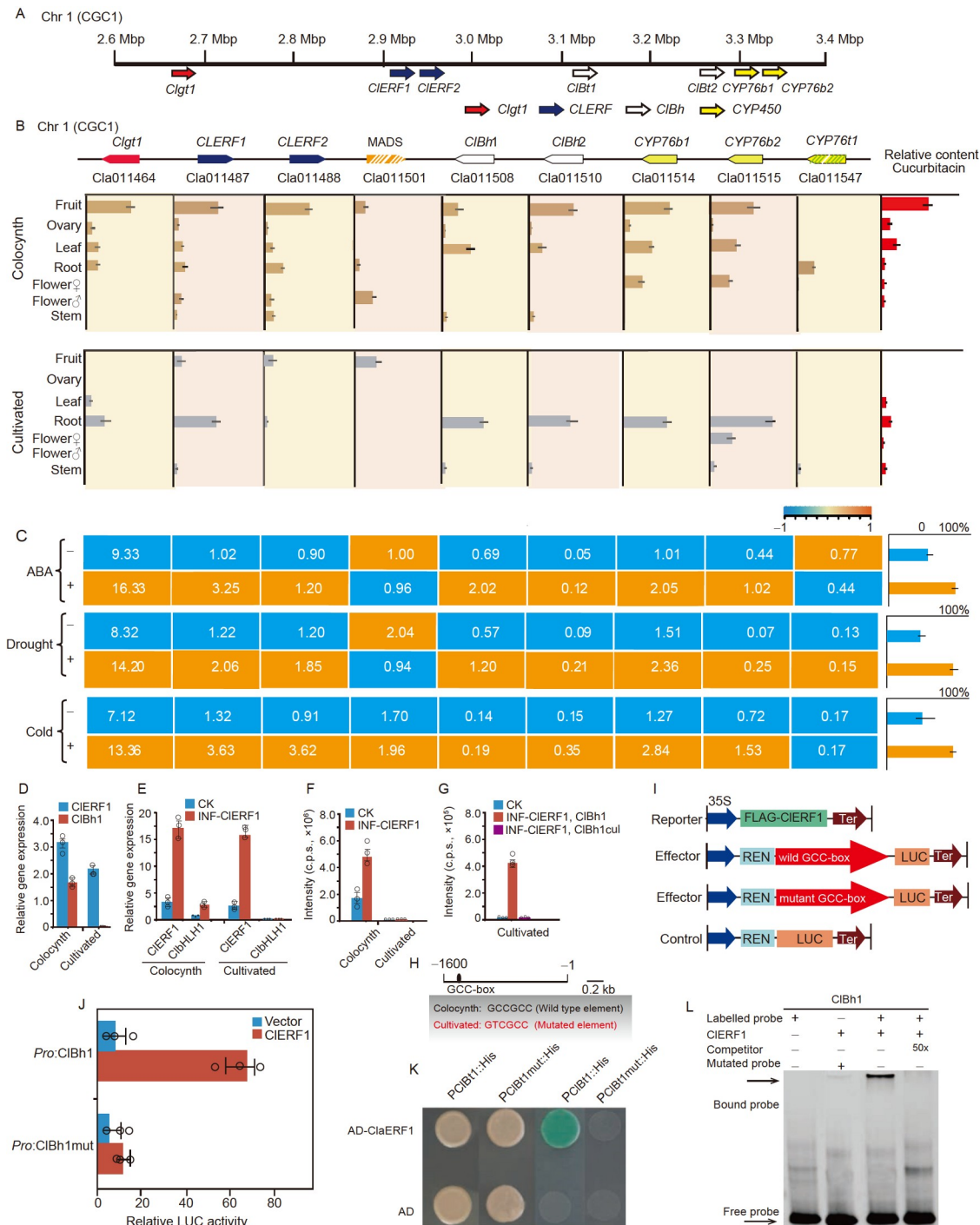


Figure 3 Identification of a cucurbitacin gene cluster and functional elucidations of the regulation of *CIERF1* and *CIBh1*. **A**, Schematic map of cucurbitacin gene cluster identified in the genomic regions on chromosome 1 (*CGC1*). Specific genes are indicated by arrows, and the other irrelevant genes are not shown here. **B**, Expression profiles of co-expressed candidate genes in the *CGC1* and cucurbitacin contents in different tissues of CC (colocynth) and cultivated line. Gene expression profiles were determined by qRT-PCR. The values shown are means±standard errors, *n*=three biological replicates. **C**, Up-regulation of the seven genes in watermelon plants under ABA treatment, drought stress, or cold condition. **D**, The relative expression of *CIERF1* and *CIBh1* in colocynth and cultivated watermelon. **E** and **F**, *CIERF1* was transiently expressed in the fruits of colocynth and cultivated watermelon. c.p.s., counts per second. **G**, Co-expression of *CIERF1* and *CIBh1*, cloned from colocynth, in fruits improved the accumulation of cucurbitacins in cultivated watermelon. **H**, Diagram of the *CIBh1* promoter regions showing a single nucleotide mutation of GCC-box (GCCGCC to GTCGCC) in cultivated watermelon. **I**, Schematic presentation of the effector and reporter used in the dual-luciferase (LUC) assay in leaf epidermal cells of *N.benthamiana*. Ren, Renilla luciferase. **J**, The transactivation activity assay of transcriptional regulation of *CIBh1* and *CIBh1mut* mediated by *CIERF1* in leaves of *N. benthamiana*. The relative LUC activities were normalized to REN activity as internal control. **K**, Yeast one-hybrid activity in *CIERF1* and the promoter of *CIBh1* and *CIBh1mut*. **L**, The *CIERF1* binding to the wild allele type promoters of *CIBh1* by electrophoretic mobility shift assay (EMSA). 50-fold excess of unlabelled probes were used for competition. *Clgt1* (Cla011464); *CIERF1* (Cla011487); *CIERF2* (Cla011488); *CIBh1* (Cla011508); *CIBh2* (Cla011510); *CYP76b1* (Cla011514); *CYP76b2* (Cla011515); *CYP76t1* (Cla011547).

that *ClBhs* could bind to the promoters of *CYP76b1*, *CYP76b2*, and *Clgt1* (Figure S12D–G in Supporting Information). The transcripts of the two *CYP450s* and *Clgt1* were upregulated, and cucurbitacin content was increased in the *ClBhs* transient overexpression plants. However, in the silenced plants with pV190-*ClBhs* constructs, the expression of *Clgt1* and two *CYP450s* showed a significant reduction, and the accumulation of cucurbitacins was reduced by ~2.1-fold when compared to the empty pV190-transformed plants (Figures S12A and S13 in Supporting Information). To investigate the functions of the two *P450s*, *CYP76b1* and *CYP76b2* were transiently overexpressed in bitter watermelon fruits, and the accumulation of cucurbitacins (CuI, CuB, and CuE) was significantly upregulated (Figure S15A and B in Supporting Information). Furthermore, the down-regulation of *CYP450s* (*CYP76b1* and *CYP76b2*) resulted in a decrease in cucurbitacin concentration in the silenced lines compared to the pV190 control lines (Figure S15C and D in Supporting Information). The results suggested that the *CYP76b1* and *CYP76b2* genes indeed possess regulatory functions involved in cucurbitacin biosynthesis.

To demonstrate whether the function of the candidate glycosyltransferase (*Clgt1*, *Cla011464*) in the *CGCI* possessed enzymatic activities involved in cucurbitacin biosynthesis. Significantly elevated accumulation of CuI and CuD glucoside was observed in bitter watermelon fruit that transiently overexpressed *Clgt1* (Figures S16A and B in Supporting Information). The CuI and CuD glucoside contents showed a significant reduction in *Clgt1*-silenced plants compared to pV190 control plants (Figures S16C and D in Supporting Information). The *in vitro* enzymatic assays showed that *Clgt1* displayed higher glycosyl transfer activities for CuI than CuD, with an ~6-fold difference in their *Kcat/Km* values (Figures S16E–G in Supporting Information). These results indicated that *Clgt1* has the capacity to glycosylate cucurbitacins.

In addition to characterizing the *CGCI* gene cluster, we also identified *Cla019330* (*CIOSC*) in cuc26 encoding a putative oxidosqualene cyclase (OSC, responsible for cyclization of 2,3-oxidosqualene into cucurbitadienol) with a high sequence identity to isomultiflorenol synthase 1 (a triterpene synthase) in loofah (Li et al., 2018) (Figure 2B; Figure S17 in Supporting Information). *CIOSC* is co-expressed with the *CGCI* cluster and activated by *CIERFs* and *ClBhs* in fruits (Figure S12A in Supporting Information). To identify genetic variants of *CIOSC*, we cloned and sequenced the coding region of *CIOSC* from bitter and nonbitter watermelon accessions. Interestingly, compared to *CIOSC* in colocynth, the coding region of *Closc* in nonbitter watermelon has a 244 amino acid insertion that results in the addition of five exons in the resulting protein (Figures 4A; Figure S18 in Supporting Information). After expressing these two alleles (*CIOSC* and *Closc*) in yeast, we found that

the production of cucurbitadienol occurred only in the yeast harboring *CIOSC* (Figure 4B), with a *Km* of 36.56 $\mu\text{mol L}^{-1}$ and *Kcat* of 0.0353 s^{-1} (Figure 4C). In-depth analysis of the *CIOSC* gene in all watermelon groups showed that functional *CIOSC* was highly enriched in CC and CA, while a higher proportion of nonfunctional *Closc* was observed in CM and CL watermelons, suggesting that *CIOSC* was mainly selected at the speciation stage (Figure 4D). The selection process of cucurbitacins was further demonstrated by investigating the combination types of functional mutation sites in *ClBh1* and *CIOSC* in watermelon domestication (Figure 4E and F). Analysis of the two well-characterized selection genes revealed that the sweeps for cucurbitacin between CC-CA and CC-CM are nonidentical (Figure 4E), consistent with the idea that CA and CM may have evolved independently after their divergence (Guo et al., 2019).

In summary, a regulatory model for cucurbitacin biosynthesis in watermelon was built based on all the above results from the cucurbitacin research in this study. Abiotic stress induced the expression of *CIERFs*, *CIERFs* regulated the transcription of two *bHLHs*, and *bHLHs* activated the expression of the structural genes *CYP76b1*, *CYP76b2*, *Clgt1*, and *CIOSC*, thereby regulating the accumulation level of cucurbitacins in watermelon plants (Figure S12H in Supporting Information).

Selection of fruit sugar, acid, and carotenoid during watermelon domestication and improvement

The composition of sugars, acids, and carotenoids ultimately determines fruit taste and flesh color, thereby determining watermelon quality and contributing significantly to consumer preferences. We identified 12 carbohydrates, including monosaccharides (fructose, glucose and their derivatives, and threose), disaccharides (sucrose and trehalose), and polysaccharides (raffinose and raffinose *O*-rhamnoside) (Table S2 in Supporting Information), which were positively selected for sweetness during domestication (Figure 5A). Additionally, raffinose showed a decreasing pattern during domestication and improvement, which is consistent with that reported in a previous study (Guo et al., 2019). Combining the results of Tassel v5.2.43 and FarmCPU, eleven major reliable signals (one each on chromosomes 1, 3, 4, 5, 6, 7, 9, and 10 and three on chromosome 11) were identified (Figure 5B; Figures S5B, S6B and E in Supporting Information). Among them, raf3-5, raf5-10, and suc7-10 were under the selection region during domestication (from CL_CM to CL_ES), raf10-11, mon11-9, and mon11-24 were located in the overlapping selective sweeps for improvement I (from CL_ES to CL_LR), and raf1-18, raf9-29, and mon11-24 were under the selection region during the improvement II stage (from CL_ES to CL_LR) (Figure S19 in Supporting Information). These sweep signals showed genomic regions

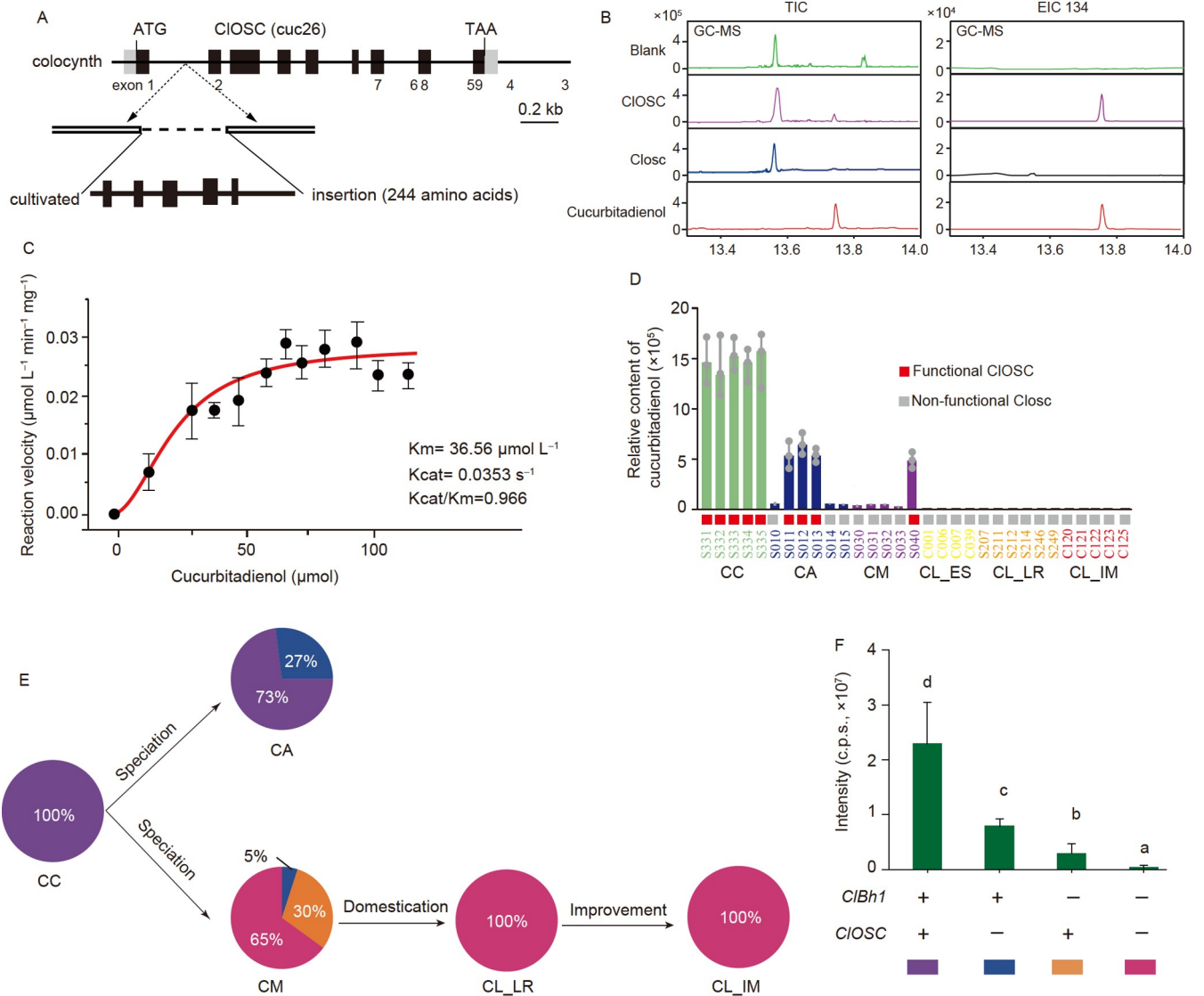


Figure 4 The selection of *ClBh1* and *CIOSC*. **A**, The *CIOSC* (Cucurbita inodora Cucurbitadienol Synthase) variant from cultivated watermelon has a sequence insertion, compared with the sequence of *CIOSC* from colocynthis watermelon. The other portions of the coding region of this gene are identical. The insertion was 244 amino acids. **B**, Gas chromatography analysis of products prepared from yeast harboring *CIOSC* and *Closc*. Empty vector was used as a negative control. **C**, Enzyme kinetic parameter of *CIOSC*. The data are represented as the mean ± SE of two replicates. **D**, The relative content of cucurbitadienol in the randomly selected six types of watermelon. Normal (functional) or insertion (nonfunctional) of *CIOSC* was indicated by red or gray box. Data are presented as the mean ± SD (n=3). **E** and **F**, The selection of *ClBh1* and *CIOSC* responsible for cucurbitacins accumulation during watermelon evolution. Function or unfunction of *ClBh1* and *CIOSC* represented by +/- . The data are presented as mean ± SD, n=3 biologically independent replicates.

with a drastic reduction in nucleotide diversity and high F_{ST} values (Figure S19 in Supporting Information), indicating successive selection for fruit sweetness during different evolutionary transitions. Two membrane transporter genes located in the *suc7-10* locus are homologous to the grape berry *VvHT* gene, which was reported to be involved in hexose transport (Davies et al., 2012) (Figure S20 in Supporting Information). Further in-depth analysis is needed to explore the possible roles of these two genes in sugar transport and sweetness regulation in watermelon.

Malic acid and citric acid are the two main organic acids in watermelon fruits (Chisholm and Picha, 1986). We observed decreased malic acid levels and increased citric acid levels

during watermelon improvement (Figure 5C and D). Five association signals were assessed for citric acid levels (one each on chromosomes 7, 9, and 10 and two on chromosome 11, marked as *cit7-10*, *cit9-32*, *cit10-16*, *cit11-9*, and *cit11-24*, respectively) (Figure 5E; Figures S5C, S6C and F in Supporting Information). *Cit9-32* and *cit11-9* were under the selective sweep regions during improvement stage I (from CL_ES to CL_LR), and *cit11-24* was within the selection regions from CL_ES to CL_IM (Figure 5F–M). We also revealed one GWAS signal underlying malic acid contents (*mal8-5*) that was located in a selective sweep during the improvement stage (from CL_ES to CL_IM) (Figure S21 in Supporting Information), suggesting a successive selection

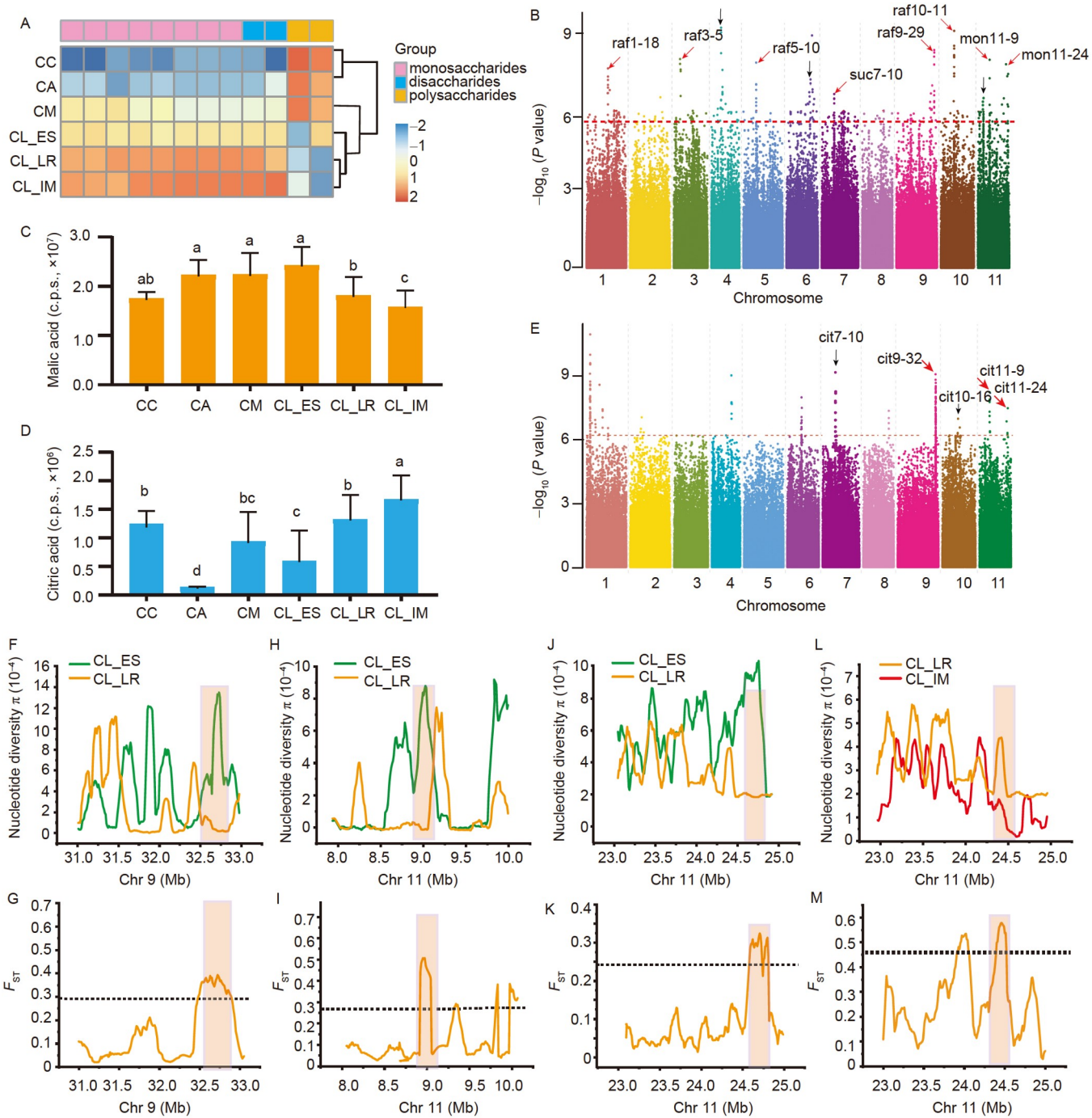


Figure 5 Fruit taste-related metabolites selected during domestication and improvement. A, Heatmap of the contents of carbohydrates detected in this study. The relative values of carbohydrate contents were scaled for each metabolite in the six watermelon groups. B, Genomic distribution of major signals associated with carbohydrate contents. Sweeps are denoted by red arrows. Black arrows indicate that the regions are not located in the sweep regions. C, Relative contents of malic acid in six watermelon types. D, Relative contents of citric acid in six watermelon types. Different letters in (C) and (D) indicate significant differences as determined by Student's *t*-test ($P<0.05$). Data in (C) and (D) represent mean standard error. E, Genomic distribution of major association signals of organic acid contents. Arrows indicate signals for citric acid contents. Red arrows denote signals located within sweep regions. Black arrows indicate those located outside the sweep regions. F–M, Selective sweeps during improvement I and II overlapping with association signals of citric acid and carbohydrate contents on chromosomes 9 and 11. The sweep regions were supported by a sharp reduction of nucleotide diversity (π) and high population divergence (F_{ST}). raf, raffinose O-rhamnoside; suc, sucrose; mon, monosaccharides; cit, citric acid.

for malic acid during improvement. Notably, a series of fruit sugar- and citric acid-related loci were found to overlap with domestication and improvement sweeps (Figure 5B and E; Table S11 in Supporting Information), suggesting that fruit

sugar and acid accumulation coevolved through shared genetic regulation under human selection.

Watermelon flesh color is a vital visual quality trait and is determined by carotenoid composition. We observed that

lycopene and β -carotene levels were positively selected during domestication (Figure 1C). It is worth noting that in genetic mapping of the flesh color trait, β -carotene, β -apocarotenal, and lycopene gave significant signals in the same region on chromosome 4 (Figures S22A–F in Supporting Information) that contains the previously reported lycopene β -cyclase (*LCYB*; *Cla005011*) (Bang et al., 2007; Zhang et al., 2020a). Furthermore, this locus was under the selection region during the domestication stage (Figures S22G in Supporting Information). Two nonsense mutations (SNP1 and SNP2) were present in the coding region of *LCYB* (Table S14 in Supporting Information). Compared with varieties harboring the C allele at the SNP1 (C/G) locus, varieties that harbor the G allele produced 6.5- and 7.5-folds increases in β -carotene and β -apocarotenal, respectively, leading to yellow- or orange-fleshed watermelon. SNP2 (A/C) has been identified in previous studies (Bang et al., 2007). Here, varieties harboring the C allele had 8.6-fold more lycopene than varieties with the A allele, resulting in red-fleshed watermelon (Tables S14 and S15 in Supporting Information). We further analyzed the allele distribution of SNP1 and SNP2 in the six types of watermelon (Figure S22H in Supporting Information). All CC, CA, and CM accessions had pale-colored flesh and harbored the allele combination of C_1A_2 (SNP1:C and SNP2:A), whereas all CL_LR and CL_IM accessions showed colored flesh and possessed allelic combinations of G_1A_2 (orange- or canary yellow-fleshed) or C_1C_2 (red- or pink-fleshed), corresponding to the overaccumulation of β -carotene/ β -apocarotenal and lycopene, respectively (Figure S22I in Supporting Information).

Three allelic combinations were demonstrated in CL-ES (C_1A_2 , G_1A_2 , and C_1C_2) (Table S15 and Figure S22H in Supporting Information), suggesting that the alleles responsible for colorful flesh already existed in the CL_ES genotype and were subsequently further selected and largely fixed in landrace and modern sweet watermelons. Similarly, the alleles related to higher sugar content (two nonsynonymous SNP sites in two membrane transporter genes on chromosome 7 shown in Figure S20 in Supporting Information) were selected and became the predominant alleles in CL_ES (71%) and were further fixed in CL_LR (~100%) (Figure S23 in Supporting Information). These results suggested that watermelon fruit quality is gradually being selected and fixed in breeding history. In addition, we also provided the potentially causal gene lists on the above-mentioned loci for these fruit quality related metabolic traits in Table S16 in Supporting Information.

Evolution and genetic improvement of watermelon fruit quality

The above results have demonstrated a stepwise selection of fruit quality-related metabolomes during watermelon evo-

lution. We next analyzed the selective sweeps (domestication loci) on each chromosome related to watermelon fruit quality. Cucurbitacins were mainly selected at the speciation stage, and four cucurbitacin-related speciation sweeps were identified in the speciation process (Table S17 in Supporting Information). In addition, SNPs from 18 major loci that control fruit taste and quality-related metabolites located within watermelon domestication and improvement sweeps (Table S18 in Supporting Information) were detected in the domestication and improvement processes. Detailed examination of 11 domestication loci underlying major selected metabolites from CM to CL_LR revealed that six loci were selected from CM to CL_ES, while the remaining five loci and one extra (*cit11-9*) were subsequently selected from CL_ES to CL_LR (Figure S24 in Supporting Information).

DISCUSSION

Although a large number of evolution/domestication-related traits have been explored (Abbo et al., 2014; Beleggia et al., 2016; Chen et al., 2015; Doebley et al., 2006; Lenser and Theißen, 2013; Zhang et al., 2020b; Zhu et al., 2018), their order of selection remains mysterious. Using watermelon as a case study, we here demonstrated the stepwise selection of taste/appearance-related metabolites during evolution and revealed the underlying genetic and molecular bases. It seems that the later selected traits are masked until the earlier ones have been selected (Li et al., 2006). Given that marker-assisted/whole genome selection breeding will accelerate the selection process and allow the breeding of varieties with novel combinations of traits, our current study provides an intriguing stepwise model of changes in the metabolome associated with watermelon domestication.

Operon-like clusters in higher plants, simply called gene clusters, are important tools for pathway discovery through genome mining, metabolic engineering, and synthetic biology (Boycheva et al., 2014; Field et al., 2011). To date, all the gene clusters identified consist solely of structural genes encoding enzymes that catalyze different and often consecutive steps in the same pathway and are mainly associated with specialized metabolites (Boycheva et al., 2014; Chen et al., 2019; Shang et al., 2014; Zheng et al., 2015; Zhu et al., 2018). These new results were different from the gene clusters on chromosome 6 (1.53–1.56 Mb) in watermelon reported by Zhou et al. (2016) using comparative genomics analyses, and the possible reason could be the absence of a large natural mutation near the site of chromosome 6 during watermelon domestication and improvement. However, here we discovered a considerably more complicated gene cluster for cucurbitacin biosynthesis, consisting of a cascade of regulation machinery comprised of consecutive transcription

factors and a number of downstream cucurbitacin biosynthetic genes, thereby elucidating a novel regulatory structure in eukaryotic genomes.

MATERIALS AND METHODS

Plant materials

The 354 representative watermelon accessions were collected in different locations throughout the world, including 10 *C. colocynthis*, 31 *C. amarus*, 22 *C. mucosospermus* (Egusi), one *C. naudinianus*, and 290 *C. lanatus* (10 *C. lanatus* edible seed, 64 *C. lanatus* landrace, 216 *C. lanatus* improved). Information about the accessions is provided in Table S1 in Supporting Information. The accessions were maintained at the National Mid-term Genebank for Watermelon and Melon (Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, China).

Metabolite profiling

For metabolite analysis, two biologically replicated sample sets of 204 watermelon accessions (Table S1 in Supporting Information) were grown at Xixiang Experimental Station. At the mature stage, the fruit flesh tissues were harvested from three different plants per line as one biological replicate, snap frozen in liquid nitrogen, and freeze-dried for metabolite profiling with two biological replicates for each accession.

The freeze-dried tissues were crushed using a mix mill (MM 400, Retsch, Germany) with zirconia beads for 1.5 min at 30 Hz. 100 mg of powder was weighed and extracted overnight at 4°C with 1.0 mL of 70% aqueous methanol. Following centrifugation at 10,000g for 15 min, the extracts were filtered (SCAA-104, 0.22 µm pore size; ANPEL, Shanghai, China, <http://www.anpel.com.cn/>) before LC-MS analysis. A previously described relative quantification method of widely targeted metabolites was used to analyze the samples (Chen et al., 2013).

Extraction and measurement of carotenoids

Carotenoid extraction and determinations were based largely on a previously reported protocol (Inbaraj et al., 2008; Petry and Mercadante, 2017) with minor modifications. The weighed frozen flesh powder (100 mg fresh weight) was extracted with *n*-hexane-acetone-ethanol (2:1:1, V/V/V) containing 0.1% butyl hydroxytoluene (BHT). The sample extracts were analyzed using a UPLC-APCI-MS/MS system (Ultra Performance Liquid Chromatography, UPLC, Shim-pack UFLC SHIMADZU CBM30A system, <http://www.shimadzu.com.cn/>, Agilent Technologies, USA; Tandem

mass spectrometry, MS/MS, Applied Biosystems 6500 Quadrupole Triple, <http://www.appliedbiosystems.com.cn/>, AB Sciex, USA). The quantification of the individual carotenoids was determined using dose-response curves constructed with authentic standards, β -carotene, lycopene, and β -apocarotenal.

Phylogenetic and population structure analysis

A subset of 2,108,752 SNPs (bi-allelic with a missing data rate <20% and a minor allele frequency >5%) were used to build a neighbor-joining tree for 354 watermelon accessions using PHYLIP (version 3.695) with 100 bootstrap replicates (Retief, 2000). A data matrix representing the contents of 3730 metabolites from 204 watermelon accessions was used to construct a neighbor-joining tree using PHYLIP with default settings. iTOL (<https://itol.embl.de/tree>) was used for visualizing the phylogenetic trees.

Principal component analysis (PCA) was performed with log₂ transformed metabolite data for 204 accessions to demonstrate the structure of the watermelon population using SIMCA-P version 14.1 (Wu et al., 2010).

Metabolite-based genome-wide association studies

We used 429,811 SNPs (minor allele frequency ≥ 0.05 and missing data rate <10%) for 204 accessions to perform the mGWAS. A compressed mixed linear model was used to detect a trait-marker association (Yu et al., 2006; Zhang et al., 2010) with the *Q* matrix and kinship matrix as covariates by TASSEL v5.2.43 (Bradbury et al., 2007). Population structure (*Q*) was determined using ADMIXTURE version 1.3 (Alexander et al., 2009) with *K*=8. The kinship matrix (*K*) was estimated using Tassel V 5.2.43. Furthermore, to eliminate the influence of the subpopulation on the signals for metabolite traits, we added subpopulations as cofactors with *k*=7. The signals were stable with adjustment on the subpopulation effects.

To further reduce false positive and false negative signals and obtain credible results, 52 important metabolic traits, including cucurbitacins, carbohydrates, and organic acids, were analyzed by multi-locus GWAS using FarmCPU software (Liu et al., 2016). The two commonly used multiple comparison methods to select for the significant threshold level in association mapping studies are Bonferroni correction (Holm, 1979) and false discovery rate (Benjamini and Hochberg, 1995). In the present study, based on the calculated Bonferroni correction, the threshold was set to 3.11×10^{-11} . However, using this very strict threshold can lead to high false negative error rates. Therefore, we used the FDR method to determine the threshold *P*-value for GWAS (Kaler and Purcell, 2019), which can identify signals that are truly associated with metabolite traits.

Identification of selective sweeps

The nucleotide diversity (π) was measured using a 100-kb window and a step size of 10 kb. The ratio of π was calculated in the three comparisons representing different watermelon domestication (from *C. mucospermus* to *C. lanatus* edible seed) and improvement stages (from *C. lanatus* edible seed to *C. lanatus* landrace and from *C. lanatus* landrace to *C. lanatus* improved watermelon), and genome regions with the top 5% highest ratios (6.25, 3.12, and 2.71 for the three comparisons, respectively) were selected as candidate selective sweeps. To reduce the false-positive rate and retain true selection signals, we further calculated the population differentiation statistics (F_{ST}) using a 100-kb window and a step size of 10 kb in each of the three comparisons. Candidate selective sweeps identified using the π ratio that also had the top 5% highest F_{ST} values were defined as final selective sweeps.

Statistical analyses

The coefficient of variation values were computed independently for each metabolite in the entire population. The calculation formula is as follows: σ/μ , where σ and μ represent the standard deviation and the mean of the levels of each metabolite in the population, respectively. The broad-sense heritability (H^2) was calculated using the following formula by treating accessions as a random effect and the biological replication as a replication effect using one-way analysis of variance: $H^2 = \text{Var}_{(G)} / [\text{Var}_{(G)} + \text{Var}_{(E)}]$, where $\text{Var}_{(G)}$ and $\text{Var}_{(E)}$ are the genetic and environmental variances, respectively.

Yeast one-hybrid assay

Yeast cells were cotransformed with the pHIS2 bait vector harboring the promoter of the target gene and the pGADT7 prey vector harboring the CDS of bHLHs or ERFs. As negative controls, yeast cells were transformed with the empty pGADT7 vector and pHIS2 harboring the corresponding promoter. Transformed yeast cells were grown in SD-Leu-Trp medium and SD-Leu-Trp-His medium plates supplemented with 3-AT (Sigma, USA), respectively. The plates were then incubated for three days at 30°C.

Transient expression of heterologous protein in *N. benthamiana* and enzyme activity assay

The recombinant *Agrobacterium tumefaciens* EHA105 harboring the *Cla019330* transient expression construct at a concentration of $A_{600}=0.8$ was infiltrated into the young leaves of six *N. benthamiana* plants, which were sampled 4 days post infiltration. The *Agrobacterium*-mediated infiltration of *N. benthamiana* leaves was performed as previously

described (Chen et al., 2011). These experiments were repeated with three times, independently. For gene expression analysis, the *Nicotiana benthamiana* GAPDH (glyceraldehyde 3-phosphate dehydrogenase) gene was used as a reference gene (Rhee et al., 2016). The empty PEAQ-HT-DEST2 vector was processed with the same treatments as a negative control.

To determine the kinetic constants of OSC/osc for 2,3-oxidosqualene, a series of concentration of the substrate 2,3-oxidosqualene at 0.0, 1.0, 2.0, 3.0, 4.0, 5.0, 10.0, 15.0, 20.0, 25.0 and 30.0 $\mu\text{mol L}^{-1}$ was used. The enzyme was purified from tobacco leaves by the affinity purification with His tag (Valdez-Ortiz et al., 2005; Pagliano et al., 2017). Reactions were initiated by different concentrations of substrate, 30 μg purified OSC/osc in a total volume of 200 μL (250 Mm MOPS, pH=7.2), and 15 mmol L^{-1} MgCl_2 and 2 mmol L^{-1} DTT as co-factors, and were incubated with shaking at 35°C for 60 min. All kinetic parameters were calculated using Michaelis-Menten model (Sigma Plot, version 12.5). All the reactions were run with 3 replicates, and each experiment was repeated 3 times.

Virus-induced gene silencing

For the virus-induced gene silencing (VIGS) experiment, a cucumber green mottle mosaic virus-based pV190 VIGS vector was obtained from Professor Qincheng Gu at Zhengzhou Fruit Research Institute, CAAS. The gene-specific fragments were amplified using the primers listed in Table S19 in Supporting Information. The pV190 VIGS vectors harboring different gene fragments were constructed via homologous recombination with *Bam*HI in the sense orientation, respectively, using a ClonExpress® Ultra One Step Cloning Kit (Vazyme Biotech, Nanjing, China) (Liu et al., 2020). All resulting error-free constructs were separately introduced into the *Agrobacterium tumefaciens* strain GV3101. The single clones were picked, transferred into 500 μL of LB liquid media containing 50 $\mu\text{g mL}^{-1}$ kanamycin and 50 $\mu\text{g mL}^{-1}$ rifampicin and cultured in a shaker at 28°C for 16 h. The bacterium was inoculated into 50 mL of LB medium containing antibiotics and cultured for 16 h. The bacteria were collected and resuspended in inducing buffer solution (10 mmol L^{-1} MgCl_2 , 10 mmol L^{-1} MES, and 100 $\mu\text{mol L}^{-1}$ Acetosyringone) at an $A_{600}=0.8$. The *Agrobacterium* suspension was maintained for 3 h at 25°C and infiltrated into cotyledons of 7-day-old watermelon seedlings from the adaxial side using a needleless syringe.

The pV190 vectors carrying PDS fragments (pV190-PDS) were injected simultaneously as a positive control to check the silencing efficiency, and the empty pV190 vector was used as the negative control. The pV190-PDS infected plants first exhibited a photobleaching phenotype at 20 days post inoculation (dpi). The upper younger leaves and fruits were

sampled, the cucurbitacin content was determined by LC-MS, and the target gene expression level was examined by qRT-PCR. Each experiment was repeated at least six times with the same results. The bitter watermelon accessions PI632751 and PI386015 were used as the inoculated plants for the VIGS experiment in this study.

Electrophoretic mobility shift assay (EMSA)

The full-length CDS of ERF-487 was ligated into the pET30a vector with a His-tag. The recombinant vectors were transformed into *Escherichia coli* strain BL21 (PG-KJE8) competent cells. Expression of the recombinant His-tagged-ERF487 proteins was induced by isopropyl- β -D-thiogalactoside and purified using a HisPur™ Ni-NTA Purification Kit (Thermo Scientific, USA). A Light Shift Chemiluminescent Electrophoretic Mobility Shift Assay Kit (Thermo Fisher Scientific, USA) was used in this experiment. Biotin was labeled at the 5' end of the *cis*-element (Table S19 in Supporting Information). The biotin-labeled DNA probes were synthesized by Sangon Biotech Co., Ltd (Shanghai, China). The detailed procedure for EMSA followed the manufacturer's instructions and methods reported in the literature (Li et al., 2020).

Availability of data and materials

Data supporting the findings of this work are available within the paper and its Supplementary Information files. The genetic materials generated and analyzed during the current study are available from the corresponding author upon request. The raw genome resequencing data are available at NCBI sequence read archive (SRA) under accessions PRJNA641178.

Compliance and ethics The author(s) declare that they have no conflict of interest.

Acknowledgements This work was supported by the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-ZFRI-07), the National Key R&D Program of China (2018YFD0100704), the China Agriculture Research System (CARS-25-03), the National Natural Science Fund for Distinguished Young Scholars (31625021), the National Natural Science Foundation of China (31672178, 31471893), the Hainan University Startup Fund KYQD (ZR) 1866, Project supported by Hainan Yazhou Bay Seed Laboratory (B21Y10901), the Natural Science Foundation of Hainan Province (322RC574). We thank all the participants of this study for their contributions.

References

Abbo, S., Van-Oss, R.P., Gopher, A., Saranga, Y., Ofner, I., and Peleg, Z. (2014). Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci* 19, 351–360.

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based

estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655–1664.

Bang, H., Kim, S., Leskovar, D., and King, S. (2007). Development of a codominant CAPS marker for allelic selection between canary yellow and red watermelon based on SNP in lycopene β -cyclase (*LCYB*) gene. *Mol Breeding* 20, 63–72.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc-Ser B (Methodological)* 57, 289–300.

Beleggia, R., Rau, D., Laidò, G., Platani, C., Nigro, F., Fragasso, M., De Vita, P., Scossa, F., Fernie, A.R., Nikoloski, Z., et al. (2016). Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol Biol Evol* 33, 1740–1753.

Boycheva, S., Daviet, L., Wolfender, J.L., and Fitzpatrick, T.B. (2014). The rise of operon-like gene clusters in plants. *Trends Plant Sci* 19, 447–459.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635.

Chen, J., Hu, X., Shi, T., Yin, H., Sun, D., Hao, Y., Xia, X., Luo, J., Fernie, A.R., He, Z., et al. (2020). Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. *Plant Biotechnol J* 18, 1722–1735.

Chen, Q., Sun, J., Zhai, Q., Zhou, W., Qi, L., Xu, L., Wang, B., Chen, R., Jiang, H., Qi, J., et al. (2011). The basic helix-loop-helix transcription factor MYC2 directly represses *PLETHORA* expression during Jasmonate-mediated modulation of the root stem cell niche in *Arabidopsis*. *Plant Cell* 23, 3335–3352.

Chen, R., Deng, Y., Ding, Y., Guo, J., Qiu, J., Wang, B., Wang, C., Xie, Y., Zhang, Z., Chen, J., et al. (2022). Rice functional genomics: decades' efforts and roads ahead. *Sci China Life Sci* 65, 33–92.

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., Li, Y., Liu, X., Zhang, H., Dong, H., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet* 46, 714–721.

Chen, W., Gong, L., Guo, Z., Wang, W., Zhang, H., Liu, X., Yu, S., Xiong, L., and Luo, J. (2013). A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol Plant* 6, 1769–1780.

Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J., Wang, S., Shi, L., Zhou, B., Li, Z., et al. (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat Commun* 7, 12767.

Chen, X., Liu, F., Liu, L., Qiu, J., Fang, D., Wang, W., Zhang, X., Ye, C., Timko, M.P., Zhu, Q.H., et al. (2019). Characterization and evolution of gene clusters for terpenoid phytoalexin biosynthesis in tobacco. *Planta* 250, 1687–1702.

Chen, Y.H., Gols, R., and Benrey, B. (2015). Crop domestication and its impact on naturally selected trophic interactions. *Annu Rev Entomol* 60, 35–58.

Chisholm, D.N., and Picha, D.H. (1986). Distribution of sugars and organic acids within ripe watermelon fruit. *HortScience* 21, 501–503.

Davies, C., Boss, P., Geros, H., Lecourieux, F., and Delrot, S. (2012). Source/sink relationships and molecular biology of sugar accumulation in grape berries. In: Gerós, H., and Delrot, S., ed. *The Biochemistry of the Grape Berry*. Sharjah: Bentham Science Publishers Ltd. 44–66.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321.

Fernie, A.R., and Schauer, N. (2009). Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet* 25, 39–48.

Field, B., Fiston-Lavier, A.S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A.E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci USA* 108, 16116–16121.

Guo, S., Zhao, S., Sun, H., Wang, X., Wu, S., Lin, T., Ren, Y., Gao, L., Deng, Y., Zhang, J., et al. (2019). Resequencing of 414 cultivated and

- wild watermelon accessions identifies selection for fruit quality traits. *Nat Genet* 51, 1616–1623.
- Gong, Z., Xiong, L., Shi, H., Yang, S., Herrera-Estrella, L.R., Xu, G., Chao, D.Y., Li, J., Wang, P.Y., Qin, F., et al. (2020). Plant abiotic stress response and nutrient use efficiency. *Sci China Life Sci* 63, 635–674.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand J Statist* 6, 65–70.
- Inbaraj, B.S., Lu, H., Hung, C.F., Wu, W.B., Lin, C.L., and Chen, B.H. (2008). Determination of carotenoids and their esters in fruits of *Lycium barbarum* Linnaeus by HPLC-DAD-APCI-MS. *J Pharmaceutical Biomed Anal* 47, 812–818.
- Jin, C., Fang, C., Zhang, Y., Fernie, A.R., and Luo, J. (2021). Plant metabolism paves the way for breeding crops with high nutritional value and stable yield. *Sci China Life Sci* 64, 2202–2205.
- Kaler, A.S., and Purcell, L.C. (2019). Estimation of a significance threshold for genome-wide association studies. *BMC Genomics* 20, 618.
- Lenser, T., and Theißen, G. (2013). Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci* 18, 704–714.
- Levi, A., Jarret, R., Kousik, S., Wechter, W.P., Nimmakayala, P., and Reddy, U.K. (2017). Genetic resources of watermelon. In: Grumet, R., Katzir, N., and Garcia-Mas, J., ed. *Genetics and Genomics of Cucurbitaceae*. Plant Genetics and Genomics: Crops and Models. Cham: Springer. 87–110.
- Li, B., Lu, X., Dou, J., Aslam, A., Gao, L., Zhao, S., He, N., and Liu, W. (2018). Construction of a high-density genetic map and mapping of fruit traits in watermelon (*Citrullus Lanatus* L.) based on whole-genome resequencing. *Int J Mol Sci* 19, 3268.
- Li, C., Zhou, A., and Sang, T. (2006). Rice domestication by reducing shattering. *Science* 311, 1936–1939.
- Li, W., Wang, K., Chern, M., Liu, Y., Zhu, Z., Liu, J., Zhu, X., Yin, J., Ran, L., Xiong, J., et al. (2020). Sclerenchyma cell thickening through enhanced lignification induced by *OsMYB30* prevents fungal penetration of rice leaves. *New Phytol* 226, 1850–1863.
- Liu, M., Liang, Z., Aranda, M.A., Hong, N., Liu, L., Kang, B., and Gu, Q. (2020). A cucumber green mottle mosaic virus vector for virus-induced gene silencing in cucurbit plants. *Plant Methods* 16, 9.
- Liu, X., Huang, M., Fan, B., Buckler, E.S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet* 12, e1005767.
- Luo, J. (2015). Metabolite-based genome-wide association studies in plants. *Curr Opin Plant Biol* 24, 31–38.
- Ma, L., Wang, Q., Zheng, Y., Guo, J., Yuan, S., Fu, A., Bai, C., Zhao, X., Zheng, S., Wen, C., et al. (2022). Cucurbitaceae genome evolution, gene function, and molecular breeding. *Horticulture Res* 9, uh057.
- Moing, A., Allwood, J.W., Aharoni, A., Baker, J., Beale, M.H., Ben-Dor, S., Biais, B., Brigante, F., Burger, Y., Deborde, C., et al. (2020). Comparative metabolomics and molecular phylogenetics of melon (*Cucumis melo*, Cucurbitaceae) biodiversity. *Metabolites* 10, 121.
- Pagliano, C., Bersanini, L., Cella, R., Longoni, P., Pantaleoni, L., Dass, A., Leelavathi, S., and Reddy, V.S. (2017). Use of nicotiana tabacum transplastomic plants engineered to express a His-tagged CP47 for the isolation of functional photosystem II core complexes. *Plant Physiol Biochem* 111, 266–273.
- Petry, F., and Mercadante, A. (2017). New method for carotenoid extraction and analysis by HPLC-DAD-MS/MS in freeze-dried citrus and mango pulps. *J Braz Chem Soc* 29, 205–215.
- Ren, Y., Guo, S., Zhang, J., He, H., Sun, H., Tian, S., Gong, G., Zhang, H., Levi, A., Tadmor, Y., et al. (2018). A tonoplast sugar transporter underlies a sugar accumulation QTL in watermelon. *Plant Physiol* 176, 836–850.
- Ren, Y., Li, M., Guo, S., Sun, H., Zhao, J., Zhang, J., Liu, G., He, H., Tian, S., Yu, Y., et al. (2021). Evolutionary gain of oligosaccharide hydrolysis and sugar transport enhanced carbohydrate partitioning in sweet watermelon fruits. *Plant Cell* 33, 1554–1573.
- Retief, J.D. (2000). Phylogenetic analysis using PHYLP. *Methods Mol Biol* 132, 243–258.
- Rhee, S.J., Jang, Y.J., and Lee, G.P. (2016). Identification of the subgenomic promoter of the coat protein gene of cucumber fruit mottle mosaic virus and development of a heterologous expression vector. *Arch Virol* 161, 1527–1538.
- Richter, A., Schaff, C., Zhang, Z., Lipka, A.E., Tian, F., Köllner, T.G., Schnee, C., Preiß, S., Irmisch, S., Jander, G., et al. (2016). Characterization of biosynthetic pathways for the production of the volatile homoterpenes DMNT and TMTT in *Zea mays*. *Plant Cell* 28, 2651–2665.
- Riedelsheimer, C., Lisec, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., Altmann, T., Stitt, M., Willmitzer, L., and Melchinger, A.E. (2012). Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc Natl Acad Sci USA* 109, 8872–8877.
- Robinson, R.W., and Decker-Walters, D.S. (1997). Cucurbits. New York: CAB International. 65–97.
- Saminathan, T., García, M., Ghimire, B., Lopez, C., Bodunrin, A., Nimmakayala, P., Abburi, V.L., Levi, A., Balagurusamy, N., and Reddy, U.K. (2018). Metagenomic and metatranscriptomic analyses of diverse watermelon cultivars reveal the role of fruit associated microbiome in carbohydrate metabolism and ripening of mature fruits. *Front Plant Sci* 9, 4.
- Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., Zeng, J., Zhou, Q., Wang, S., Gu, W., et al. (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088.
- Tieman, D., Zhu, G., Resende Jr., M.F.R., Lin, T., Nguyen, C., Bies, D., Rambla, J.L., Beltran, K.S.O., Taylor, M., Zhang, B., et al. (2017). A chemical genetic roadmap to improved tomato flavor. *Science* 355, 391–394.
- Valdez-Ortiz, A., Rascón-Cruz, Q., Medina-Godoy, S., Sinagawa-García, S. R., Valverde-González, M.E., and Paredes-López, O. (2005). One-step purification and structural characterization of a recombinant His-tag 11S globulin expressed in transgenic tobacco. *J Biotechnol* 115, 413–423.
- Wang, M., Chen, J., Zhou, F., Yuan, J., Chen, L., Wu, R., Liu, Y., and Zhang, Q. (2022). The ties of brotherhood between japonica and indica rice for regional adaptation. *Sci China Life Sci* 65, 1369–1379.
- Wechter, W.P., Levi, A., Harris, K.R., Davis, A.R., Fei, Z., Katzir, N., Giovannoni, J.J., Salman-Minkov, A., Hernandez, A., Thimmapuram, J., et al. (2008). Gene expression in developing watermelon fruit. *BMC Genomics* 9, 275.
- Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., et al. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat Commun* 5, 3438.
- Wu, Z., Li, D., Meng, J., and Wang, H. (2010). Introduction to SIMCA-P and its application. In: Esposito Vinzi, V., Chin, W., Henseler, J., and Wang, H., ed. *Handbook of partial least squares*. Springer Handbooks of Computational Statistics. Berlin, Heidelberg: Springer. 757–774.
- Xu, G., Cao, J., Wang, X., Chen, Q., Jin, W., Li, Z., and Tian, F. (2019). Evolutionary metabolomics identifies substantial metabolic divergence between maize and its wild ancestor, Teosinte. *Plant Cell* 31, 1990–2009.
- Xu, Y., Zhang, H., Zhong, Y., Jiang, N., Zhong, X., Zhang, Q., Chai, S., Li, H., and Zhang, Z. (2022). Comparative genomics analysis of *bHLH* genes in cucurbits identifies a novel gene regulating cucurbitacin biosynthesis. *Horticulture Res* 9, uhac038.
- Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38, 203–208.
- Zeng, X., Yuan, H., Dong, X., Peng, M., Jing, X., Xu, Q., Tang, T., Wang, Y., Zha, S., Gao, M., et al. (2020). Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of Qingke. *Mol Plant* 13, 112–127.
- Zhang, F., Guo, H., Huang, J., Yang, C., Li, Y., Wang, X., Qu, L., Liu, X., and Luo, J. (2020b). A UV-B-responsive glycosyltransferase, OsUGT706C2, modulates flavonoid metabolism in rice. *Sci China*

- Life Sci* 63, 1037–1052.
- Zhang, J., Guo, S., Ren, Y., Zhang, H., Gong, G., Zhou, M., Wang, G., Zong, M., He, H., Liu, F., et al. (2017). High-level expression of a novel chromoplast phosphate transporter CIPHT4;2 is required for flesh color development in watermelon. *New Phytol* 213, 1208–1221.
- Zhang, J., Sun, H., Guo, S., Ren, Y., Li, M., Wang, J., Zhang, H., Gong, G., and Xu, Y. (2020a). Decreased protein abundance of lycopene β -cyclase contributes to red flesh in domesticated watermelon. *Plant Physiol* 183, 1171–1183.
- Zhang, X.P., and Jiang, Y. (1990). Edible seed watermelons (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) in northwest China. *Cucurbit Genetics Cooperative Report* 13, 40–42.
- Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 42, 355–360.
- Zheng, L., McMullen, M.D., Bauer, E., Schön, C.C., Gierl, A., and Frey, M. (2015). Prolonged expression of the BX1 signature enzyme is associated with a recombination hotspot in the benzoxazinoid gene cluster in *Zea mays*. *J Exp Bot* 66, 3917–3930.
- Zhou, S., Kremling, K.A., Bandillo, N., Richter, A., Zhang, Y.K., Ahern, K. R., Artyukhin, A.B., Hui, J.X., Younkin, G.C., Schroeder, F.C., et al. (2019). Metabolome-scale genome-wide association studies reveal chemical diversity and genetic control of maize specialized metabolites. *Plant Cell* 31, 937–955.
- Zhou, Y., Ma, Y., Zeng, J., Duan, L., Xue, X., Wang, H., Lin, T., Liu, Z., Zeng, K., Zhong, Y., et al. (2016). Convergence and divergence of bitterness biosynthesis and regulation in cucurbitaceae. *Nat Plants* 2, 16183.
- Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., Lin, T., Qin, M., Peng, M., Yang, C., et al. (2018). Rewiring of the fruit metabolome in tomato breeding. *Cell* 172, 249–261.e12.

SUPPORTING INFORMATION

The supporting information is available online at <https://doi.org/10.1007/s11427-022-2198-5>. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.