

Origin and de novo domestication of sweet orange

Received: 24 November 2023

Accepted: 6 February 2025

Published online: 5 March 2025

 Check for updates

Shengjun Liu^{1,2,7}, Yuantao Xu^{1,2,7}, Kun Yang^{1,2}, Yue Huang¹, Zhihao Lu^{1,2}, Shulin Chen¹, Xiang Gao¹, Gongao Xiao¹, Peng Chen³, Xiuli Zeng⁴, Lun Wang¹, Weikang Zheng^{1,2}, Zishuang Liu¹, Guanglian Liao¹, Fa He¹, Junjie Liu¹, Pengfei Wan¹, Fang Ding⁵, Junli Ye¹, Wenbiao Jiao¹, Lijun Chai¹, Zhiyong Pan¹, Fei Zhang¹, Zongcheng Lin¹, Yanjun Zan⁶, Wenwu Guo¹, Robert M. Larkin¹, Zongzhou Xie¹, Xia Wang^{1,2}, Xiuxin Deng¹ & Qiang Xu^{1,2}✉

Sweet orange is cultivated worldwide but suffers from various devastating diseases because of its monogenetic background. The elucidation of the origin of a crop facilitates the domestication of new crops that may better cope with new challenges. Here we collected and sequenced 226 citrus accessions and assembled telomere-to-telomere phased diploid genomes of sweet orange and sour orange. On the basis of a high-resolution haplotype-resolved genome analysis, we inferred that sweet orange originated from a sour orange × mandarin cross and confirmed this model using artificial hybridization experiments. We identified defense-related metabolites that potentially inhibited the growth of multiple industrially important pathogenic bacteria. We introduced diversity to sweet orange, which showed wide segregation in fruit flavor and disease resistance and produced canker-resistant sweet orange by selecting defense-related metabolites. Our findings elucidate the origin of sweet orange and de novo domesticated disease-resistant sweet oranges, illuminating a strategy for the rapid domestication of perennial crops.

Plant diseases increasingly challenge modern agriculture and food security. Long-term selective breeding reduces the genetic diversity of crops, thus decreasing their ability to adapt to changes in the environment and establishing effective disease resistance^{1–3}. De novo domestication of wild relatives has great potential for resolving this problem⁴ and has recently been implemented in several annual crops, such as in rice (*Oryza sativa*)⁵, potato (*Solanum tuberosum*)⁶ and tomato (*Solanum lycopersicum*)⁷. Similar strategies have not yet been deployed in woody perennial crops due to the long juvenile period of woody plants and the unknown genetic background of their wild relatives. Citrus is one of the most important fruit crops worldwide⁸. However, the citrus industry

has neglected the disease-resistance traits available from wild citrus because fruit flavor has been the pre-eminent breeding goal during the past decades.

Sweet orange (*Citrus sinensis*) is one of the most important fruit crops worldwide⁹. Most sweet orange cultivars are derived from somatic mutants and thus their background is monogenetic. Consequently, sweet orange is sensitive to disease threats. For example, all sweet oranges are sensitive to citrus canker, which is one of the most devastating diseases in the citrus industry, caused by the bacterium *Xanthomonas citri* subsp. *citri* (Xcc)^{10,11}. Large amounts of bactericide and copper-related chemicals are used annually to control this disease,

¹National Key Laboratory for Germplasm Innovation and Utilization of Horticultural Crops, Huazhong Agricultural University, Wuhan, P. R. China. ²Hubei Hongshan Laboratory, Wuhan, P. R. China. ³Horticulture Institute, Hunan Academy of Agricultural Sciences, Changsha, P. R. China. ⁴Qinghai-Tibet Plateau Fruit Trees Scientific Observation Test Station, Ministry of Agriculture and Rural Affairs, Lhasa, P. R. China. ⁵Hubei Key Laboratory of Plant Pathology, Huazhong Agricultural University, Wuhan, P. R. China. ⁶Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao, P. R. China.

⁷These authors contributed equally: Shengjun Liu, Yuantao Xu. ✉e-mail: xuqiang@mail.hzau.edu.cn

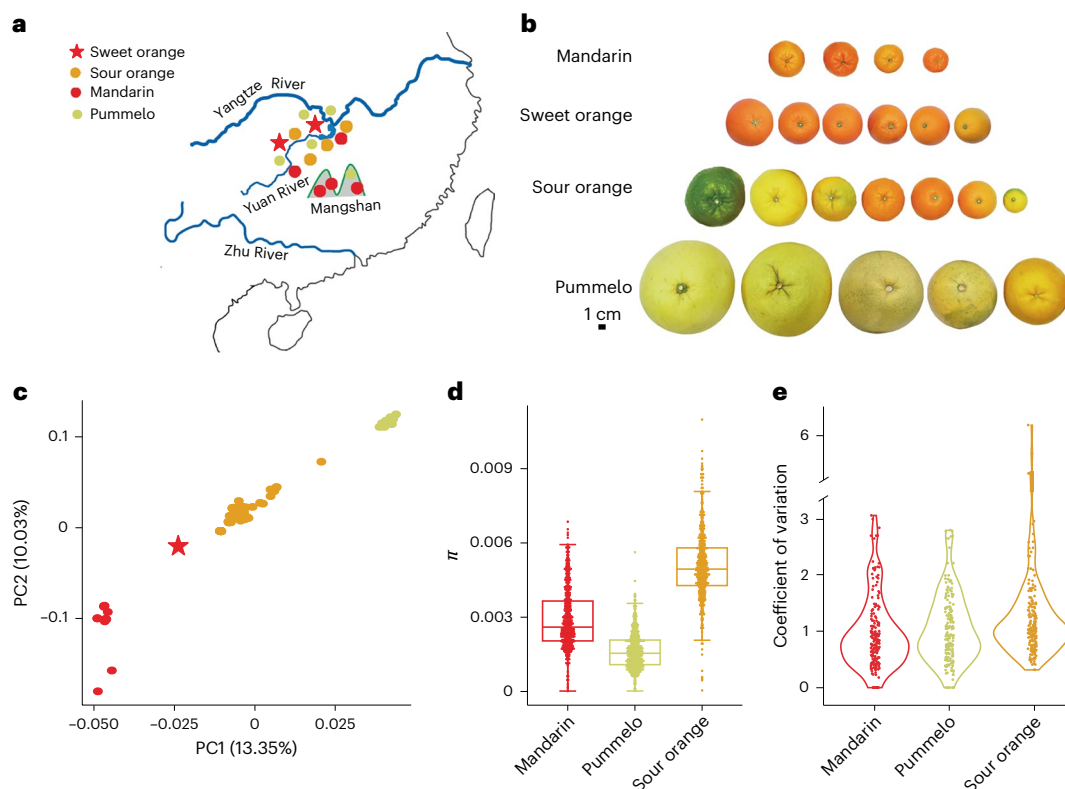


Fig. 1 | Sympatric region and genetic diversity of *Citrus* species from southern China. **a**, The geographical distribution of sweet oranges, sour oranges, mandarins and pummelos. The geographical map was generated manually using the freeform shape function of Microsoft PowerPoint software. **b**, Phenotypic variation among different sweet oranges, sour oranges, mandarins and pummelos in the sympatric region. **c**, PCA of 35 sweet oranges, 99 sour oranges, 11 mandarins and 16 pummelos. The red five-pointed star represents sweet

oranges, orange circles represent sour oranges, red circles represent mandarins and yellowish green circles represent pummelos. **d**, The π values were calculated with 1 Mb sliding windows with 500 kb steps for sour oranges, mandarins and pummelos. The number of genomic statistical windows is 545. Each box represents the median and interquartile range. The error bars represent s.e.m. **e**, Coefficients of metabolite variation for sour oranges, mandarins and pummelos.

but these solutions increase costs, become less efficacious over time and lead to ecological problems. We therefore have a critical need for the de novo domestication of canker-resistant cultivars to counteract these problems.

Plants produce a large number of specialized metabolites to defend themselves against pathogens¹². Specialized metabolites associated with defense play a variety of physiological and ecological roles¹³. Many specialized metabolites have been reported to have antimicrobial activity and are associated with plant disease resistance^{14–18}. Citrus produce a large variety of specialized metabolites such as flavonoids and coumarins^{19–21} that contribute to defense responses^{14,22}. Our previous work indicated that both decreases in disease resistance and specialized metabolite content occurred during the domestication of citrus²³. A fast and promising de novo domestication strategy for breeding new disease-resistant cultivars of citrus involves focusing on specialized metabolites from wild citrus that contribute to disease resistance.

Elucidating the origin of sweet orange is the first step toward the de novo domestication of disease-resistant sweet orange. However, the origin of sweet orange is highly debated. The first genome analysis of sweet orange proposed that it originated from two rounds of interspecific hybridizations between mandarin and pummelo²⁴. Another study that utilized more *Citrus* species indicated that the above model cannot explain some genomic regions and proposed a more complex model with several rounds of hybridization²⁵. Importantly, both models are limited because they rely on only one haplotype genome to infer evolutionary history, although sweet orange is highly heterozygous.

Results

Diversity of sour orange in southern China

The region in southern China containing the Nanling Mountains, Yuan River and Wuling Mountains (NYW) is rich in wild and cultivated citrus resources, including sweet orange (*Citrus sinensis*), sour orange (*Citrus aurantium*), mandarin (*Citrus reticulata*) and pummelo (*Citrus maxima*) (Fig. 1a,b and Extended Data Fig. 1). Diverse sweet oranges, including some varieties with extremely high acidity, are found in the NYW region²⁶. Sour oranges were even more diverse relative to other citrus species, including variations in fruit acidity, seed numbers, fruit size and color (Fig. 1b and Extended Data Figs. 1 and 2).

We collected and sequenced 118 accessions of sour oranges and landrace citrus resources and another 76 cultivated citrus varieties using an Illumina platform (Supplementary Table 1), and also included 16 accessions of pummelo, 11 accessions of mandarin and 35 accessions of sweet oranges published previously (Supplementary Table 2). We performed a principal component analysis (PCA) using the single-nucleotide polymorphism (SNP) data. This analysis revealed that the largest variation exists between pummelo and mandarin, which are two founder *Citrus* species, as indicated by their opposing positions in the PCA plot. Both the PCA and phylogenetic analysis divided sour oranges into three groups that are largely consistent with their geographical distribution (Fig. 1c and Supplementary Fig. 1). We observed the highest level of genetic polymorphisms in sour oranges collected from the NYW region, which formed a separated group. Nucleotide diversity (π) analysis yielded a π value for sour oranges (~ 0.005) that indicated a higher genetic diversity relative to mandarins (~ 0.003) and pummelos (~ 0.002) (Fig. 1d and Supplementary Fig. 2). Consistent

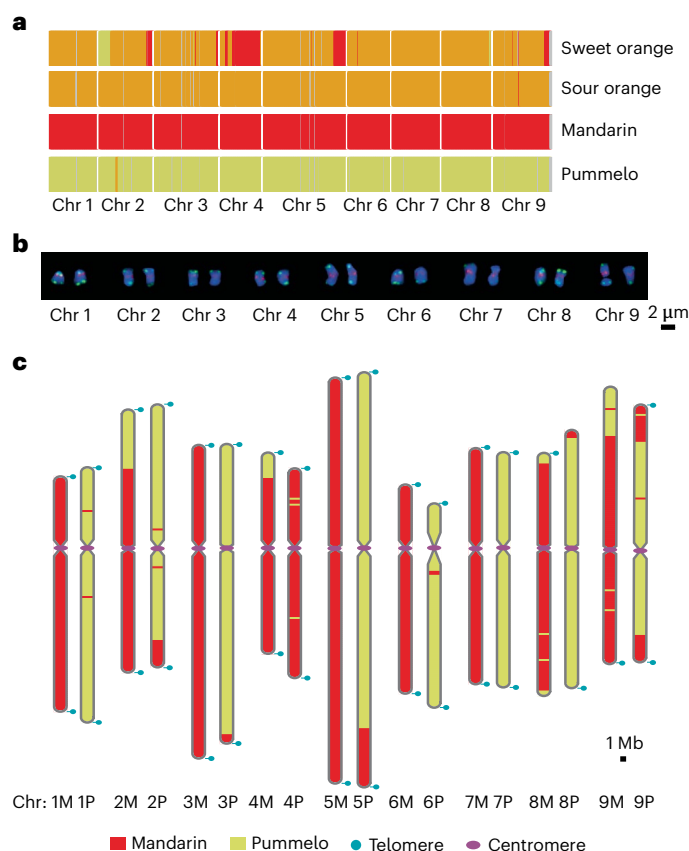


Fig. 2 | A haplotype-resolved telomere-to-telomere genome assembly of sweet orange. **a**, Genomic patterns of sweet orange ('Valencia' sweet orange), sour orange ('Anjiang' sour orange), wild mandarin ('Mangshan' mandarin) and pummelo ('Wanbai' pummelo). These segments were calculated with 250 kb windows. Red, homozygous segment with a mandarin origin; yellowish green, homozygous segment with a pummelo origin; orange, heterozygous segment with a mandarin and pummelo origin; gray, unknown region. Chr, chromosome. **b**, Dual-color FISH of centromeres (pink, using CL34contig88 as centromeric probe³⁴) labeled with biotin-16-dUTP (Roche) and telomeres (green, using synthesized telomeric probes (5'-TTTAGGG-3'), attached to FAM). **c**, A diagram illustrating the contributions of the mandarin (M) and pummelo (P) derived genome segments to the two haplotypes of sweet orange. Red, mandarin origin; yellowish green, pummelo origin; blue circles, telomeres; magenta ellipses, centromeres.

with these data, although all sweet oranges were susceptible to *Xcc*, we observed varying degrees of susceptibility to *Xcc* among the different accessions of sour orange (Supplementary Fig. 3 and Supplementary Table 3). Moreover, the coefficient of variation for metabolites in sour oranges is higher than those in mandarins and pummelos (Fig. 1e and Supplementary Table 4).

Notably, all sweet orange accessions coalesced into a single point on the PCA plot, closer to the sour orange group collected from the NYW region than to the other groups (Fig. 1c). These data provide evidence that sour oranges and sweet oranges are closely related. To explore the genetic constituents of sour oranges and sweet oranges, we analyzed the whole genome with species-specific SNPs²⁷ to evaluate the genomic patterns of sweet oranges, sour oranges, mandarins and pummelos. We confirmed that sweet orange accessions shared a nearly identical genetic background differing²⁶ (Supplementary Fig. 4), which is consistent with most sweet orange accessions arising from somatic mutations. We found that the genome of sour oranges is almost exclusively heterozygous with one allele from mandarin and the other allele from pummelo, which is consistent with the idea that sour oranges are F_1 hybrids derived from a cross between mandarin

and pummelo (Fig. 2a and Supplementary Fig. 5). Notably, most of the genomes from sour oranges and sweet oranges shared the same large heterozygous segments derived from mandarin and pummelo. However, sweet oranges differed from sour oranges in that they contain more homozygous segments from mandarin (Fig. 2a and Supplementary Figs. 4 and 5). We have identified one unknown orange accession from Hunan Province ('TJH2') that produces fruit that is similar to the fruit produced by sweet oranges in terms of fruit size, color and qualities (Extended Data Fig. 3). The genomic landscape of 'TJH2' was similar to that of sweet oranges, although not identical (Extended Data Fig. 3). These results indicate that sweet oranges are abundant and that similar accessions of sweet orange are present in the NYW region.

Telomere-to-telomere genomes of sweet orange and sour orange

To elucidate the genetic origin of sweet orange, we independently obtained haplotype-resolved telomere-to-telomere genomes of sweet orange (*Citrus sinensis*) and sour orange (*Citrus aurantium*) and assembled four chromosome-level genomes of representative *Citrus* accessions, including *C. reticulata* 'Mangshan', *C. maxima* 'Zipi', *C. reticulata* 'Ponkan' and the above mentioned natural hybrid sweet orange 'TJH2' found in southern China. For the sweet orange genome, we used a combination of PacBio high-fidelity (HiFi) long reads (N_{50} of read lengths ~17 kb, 189× coverage) and Nanopore ultralong reads (N_{50} of read lengths ~55 kb, 82× coverage) to de novo assemble the genome of 'Valencia' sweet orange (Supplementary Table 5). We used 97× Hi-C data to scaffold the assembled contigs to pseudochromosomes (Supplementary Table 6). Our assembly filled all 27 gaps in the previous genome assembly and generated a phased telomere-to-telomere genome with high heterozygosity (~2%; Fig. 2c and Supplementary Fig. 6). We divided the two sets of homologous chromosomes from sweet orange into the two haplotypes 'SWO-M' and 'SWO-P' based on their percent sequence identity to the mandarin and pummelo genomes, respectively (Supplementary Fig. 7 and Supplementary Table 7). We reconstructed a maximum likelihood tree using single-copy genes, which showed that the two haplotypes are clustered together with the mandarin and pummelo genomes, respectively, which is consistent with each haplotype being almost entirely derived from either the mandarin or pummelo genome (Supplementary Fig. 8). The sequences of the two haplotypes, 'SWO-M' and 'SWO-P', and the previously published haploid sweet orange genome sequence were highly colinear²⁶. These data indicate that the two haplotype sequences are complete (Supplementary Fig. 9). We estimated the completeness of the new genome assembly by calculating its Benchmarking Universal Single-Copy Orthologs (BUSCO) score, which reached 98% for complete and single-copy BUSCOs. Long-range sequencing combined with HiFi reads helped resolve all 18 centromeres and achieved the complete assembly of most chromosomal telomeres except for chromosome 8 from the pummelo haplotype and chromosomes 8 and 9 from the mandarin haplotype. Importantly, the new genome assembly had no gaps along all chromosomes (Fig. 2c). Fluorescence in situ hybridization (FISH) experiments supported the integrity of the assembly for telomeres and centromeres (Fig. 2b and Supplementary Fig. 10).

The level of presence and absence variation for each chromosome between 'SWO-M' and 'SWO-P' was between 2.09% and 6.77% (Supplementary Table 8). We also identified large inversions and translocations between the 'SWO-M' and 'SWO-P' haplotypes, such as a translocation between chromosomes 4 and 9. We found a ~10 Mb region without paired gene collinearity in the 'SWO-M' and 'SWO-P' haplotypes on chromosomes 4 and 9, which provides evidence for a large structural variation in this region (Supplementary Fig. 11). A comparative genomic analysis using the assembled Hi-C heat map and the Nanopore data confirmed the translocated regions (Supplementary Fig. 12). To test whether these structural variants influence the relative expression of the two haplotypes, we identified 17,850 allelic polymorphic genes in

Table 1 | Statistics for the haplotype-resolved telomere-to-telomere genome of sweet orange and sour orange

	Sweet orange M haplotype	Sweet orange P haplotype	Sour orange M haplotype	Sour orange P haplotype
Assembled size (Mb)	299.35	301.08	297.30	308.63
BUSCO completeness of assembly (%)	98.0	97.20	98.30	98.40
Number of gaps	0	0	0	0
Largest contig (Mb)	50.95	52.10	50.87	49.08
Contig N ₅₀ (Mb)	33.02	32.42	30.48	34.61
Percentage of transposable elements (%)	53.78	53.75	50.48	46.28
Number of gene models	29,777	29,812	26,078	27,914

the sweet orange heterozygous genome (Table 1 and Supplementary Table 9) and detected greater expression differences for each allele in fruit flesh relative to peels and leaves used RNA-sequencing data (Supplementary Table 10).

For the haplotype-resolved telomere-to-telomere genome of sour orange, we assembled the genome of ‘ZGSC’ sour orange (a canker-resistant genotype) using a combination of PacBio HiFi long reads (N₅₀ of read lengths ~20 kb, 102× coverage) and Nanopore ultra-long reads (N₅₀ of read lengths ~92 kb, 157× coverage) and phased the two haplotypes (Supplementary Fig. 13). We used 115.07× coverage Hi-C data for chromosome extension (Supplementary Table 6). The two sets of homologous chromosomes from sour orange were categorized as ‘SSO-M’ chromosomes (that is, chromosomes sharing high sequence identity with the mandarin genome) and ‘SSO-P’ chromosomes (that is, chromosomes sharing high sequence identity with the pummelo genome). These data provide more support for the idea that sour orange was derived from a cross between mandarin and pummelo. In addition, the genome assembly of sour orange resolved all 18 centromeres and achieved the assembly of most telomeres (Extended Data Fig. 4). Importantly, the new genome assembly had no gaps along all chromosomes.

For the genome assembly of the four chromosome-level genomes of the representative citrus accessions, including *C. reticulata* ‘Mangshan’, *C. maxima* ‘Zipi’, *C. reticulata* ‘Ponkan’ and the natural hybrid sweet orange ‘TJH2’, we generated 70.17–98.05× coverage PacBio or Nanopore long reads for the four accessions (Supplementary Table 5). We then used 68–105× coverage Hi-C data to extend the chromosomes (Supplementary Table 6). These data produced chromosome-level assemblies covering 333–365.5 Mb per genome, with contig N₅₀ values of up to 20.7 Mb (Supplementary Table 11).

Elucidating the origin of sweet orange

We used the haplotype variations identified from the whole-genome resequencing data for 305 citrus accessions to assess the extent of their relatedness to sweet orange. We determined that sour oranges are divided into three groups (‘SSO1’, ‘SSO2’ and ‘SSO3’) based on clustering and kinship (Supplementary Table 12). The kinship matrix indicates that Ponkan/Jiankan mandarins are the most closely associated with sweet oranges with the exception of sweet orange-derived cultivars, as indicated by the number of genomic fragments shared between these accessions (Fig. 3a and Supplementary Fig. 14). In addition, a group of sour orange accessions belonging to the ‘SSO2’ cluster was closely related to sweet orange (Fig. 3a and Supplementary Tables 12 and 13).

Jiankan mandarin is a landrace that is currently rarely cultivated and is genetically similar to the widely cultivated Ponkan mandarin (Supplementary Figs. 15 and 16). Both Ponkan and Jiankan mandarins are distributed in the NYW region. Identity by state (IBS) and identity by descent (IBD) analyses based on whole-genome SNP data indicated that sour oranges, Ponkan mandarin and cultivars derived from sweet orange were more closely related to sweet orange than pummelos and other mandarins. Indeed, the IBS values among sweet oranges were the highest (>0.9). The IBS values of sour oranges, Ponkan mandarin and sweet orange-derived cultivars (>0.8) were greater than those of other mandarins and pummelos (<0.8; Fig. 3b). IBD analysis estimates the likelihood that two alleles from two different species were inherited from the same parent. The probability of several sour oranges originating from the NYW region sharing a set of chromosomes with sweet oranges is close to 100%. These data are consistent with a parental relationship between sour oranges and sweet oranges (Fig. 3c). We also calculated allele sharing among accessions, using proportions from SNP data. A pair of accessions that shares 0, 1 or 2 alleles based on an IBS analysis were classified as P0, P1 and P2, respectively. The P0 value of sweet orange and ‘TJH1’ (a sour orange from the NYW region) was only 0.02%; the P0 value of sweet orange and Ponkan mandarins was 0.18% (Supplementary Fig. 17). These data indicate that sour oranges, Ponkan mandarin, and sweet oranges are highly related. For comparison, we used clementine mandarin, a hybrid between sweet orange and willow-leaf mandarin, with a P0 value of 0.04% as a reference for the parentage analysis. On the basis of these data, we hypothesize that sweet orange originated from hybridization between an ancient sour orange accession and a Ponkan-like mandarin in the NYW region. An analysis of the chloroplast genomes of the *Citrus* species revealed that sweet orange is more closely related to sour oranges and provides evidence that sour oranges are probably the maternal parent of sweet orange (Supplementary Fig. 18).

On the basis of the above results, we propose that sweet orange originated from a cross between sour orange (maternal parent) and a Ponkan-like mandarin (paternal parent) and that the former was derived from a more ancient cross between wild mandarin and pummelo (Fig. 2a). The coincidence probability refers to the probability of the genome landscape calculated by 1,014,134 species-specific SNP²⁷ matching the genetic relationship between parents and hybrids and also supports the above inference. The probability that sweet orange is derived from sour orange and Ponkan mandarin is 82.97 ± 2.77% (n = 40; Fig. 3d and Supplementary Table 14) and the probability that sour orange is the hybrid of a cross between pummelo and mandarin is 84.55 ± 3.01% (n = 75; Fig. 3d and Supplementary Table 15).

Introduction of diversity to sweet orange

To test the origin model and introduce diversity into sweet orange, we designed the maternal and paternal parents for a cross (Fig. 4a and Supplementary Fig. 19). Xcc inoculation experiments revealed that different sour oranges varied quantitatively in their resistance to citrus canker. One sour orange accession ‘ZGSC’ was confirmed resistant to citrus canker based on the evaluation of pathogen titer, H₂O₂ production and defense gene expression (Extended Data Fig. 5 and Supplementary Fig. 20). We therefore performed hybridization experiments using the resistant genotype ‘ZGSC’ as the maternal parent and Ponkan mandarin as the paternal parent. We used embryo rescue to obtain 892 hybrids and the genotypes were validated by whole-genome sequencing data and liquid gene chip indicated that 215 of these rescued embryos were indeed hybrids. Hybridization of sour orange and Ponkan mandarin generated vigorous hybrids (Fig. 4a). We conducted a genetic background analysis of parents and hybrids using the 20K liquid gene chip we developed and found a high proportion of heterozygous sites in most of these hybrids (Supplementary Figs. 21–23). We observed considerable trait variations in both fruit quality and disease resistance in the hybrid population. The total soluble solid (TSS) to titratable acidity

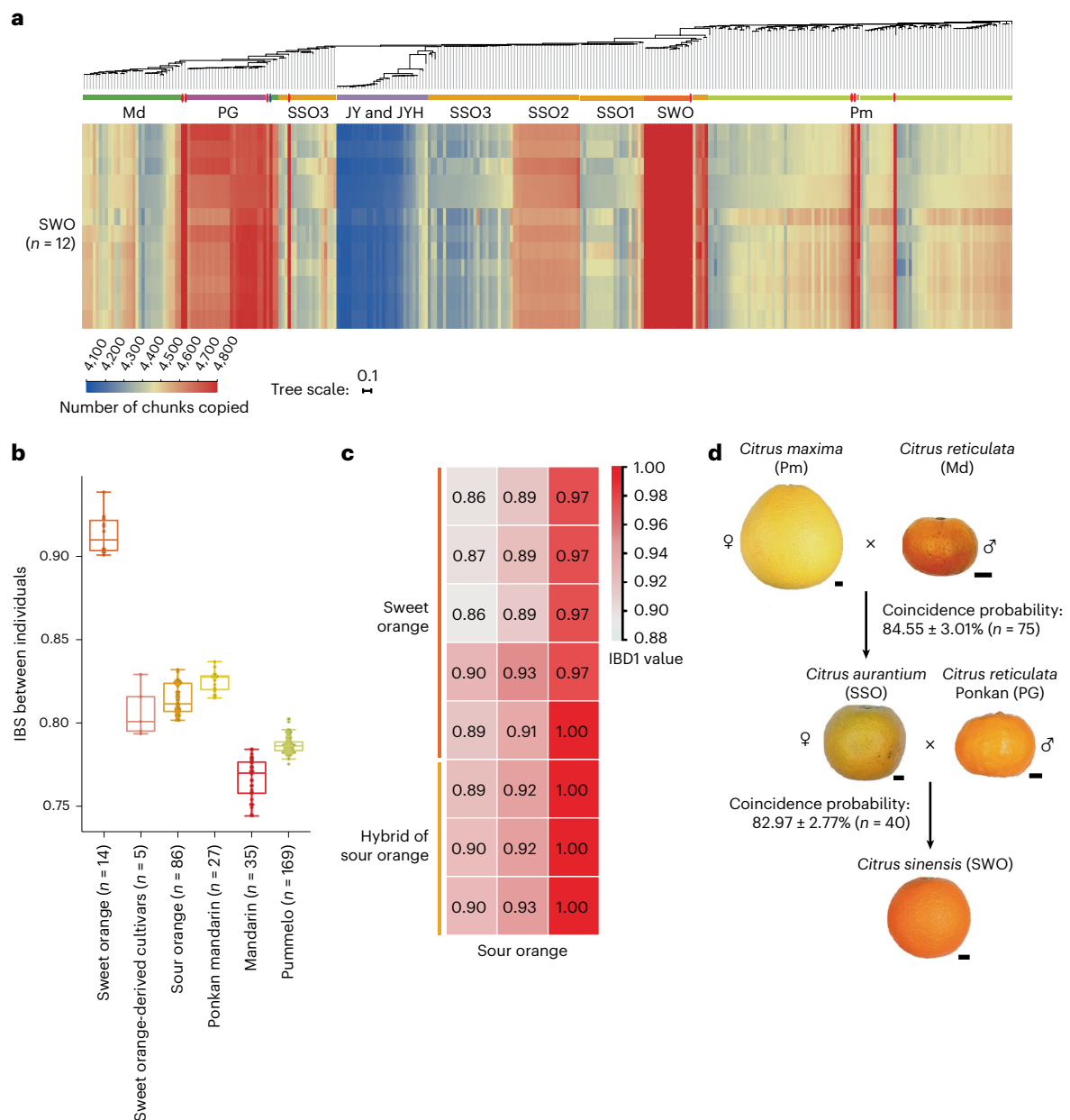


Fig. 3 | Origin of sweet orange. a, fineSTRUCTURE analysis using copying vectors generated from chromosome painting. Each row of the heat map is a recipient copying vector showing the number of fragments shared between the recipient and each individual as a donor (columns), such that block-like patterns are observed on the heat map. Blue represents less haplotype sharing and red indicates more haplotype sharing. SWO, sweet orange; red ellipses, cultivars derived from sweet orange; SSO1, sour orange group 1; SSO2, sour orange group 2; SSO3, sour orange group 3; PG, Ponkan mandarin; blue ellipses, Jiankan mandarin; Md, other mandarins; Pm, pummelo. JY and JYH, citron and the hybrids of citron, such as lemon. **b**, IBS between sweet oranges (n = 14), cultivars

derived from sweet orange (n = 5), sour oranges (n = 86), Ponkan mandarin (n = 27), other mandarins (n = 35) and pummelos (n = 169). Each box represents the median and interquartile range. The error bars represent s.e.m. **c**, IBD analysis among sweet orange, sour orange and hybrids of sour orange. Hybrids of sour orange refer to the hybrids obtained from a cross between sour orange and Ponkan mandarin. Red represents a close relationship. **d**, Model for the origin of sweet orange. The coincidence probability refers to the probability of a genome landscape that matches the genetic relationships between the parents and hybrids. SSO, sour orange. Scale bars, 1 cm.

(TA) ratio varied from 2.96 to 16.98 among hybrids, which makes a compelling case for widespread variation in fruit flavor (Supplementary Fig. 24 and Supplementary Table 16). The disease lesion areas produced by our *Xcc* inoculation assays varied widely among the hybrids (Supplementary Fig. 25). A PCA analysis provided more evidence that we obtained a highly diverse group of hybrids (Fig. 4b).

Among these hybrids, three nearly overlapped with the commercial sweet oranges in a PCA analysis (Fig. 4b). The oval shape of the fruits produced by these three hybrids was very similar to that

of sweet orange. Moreover, the orange–yellow to orange–red pericarp, tight-skinned fruit and the reduced numbers of seeds of these three hybrids was also similar to the commercial sweet oranges. The fruit flavor, as indicated by the ratio of TSS to TA, was also similar to sweet orange (Supplementary Table 16). Moreover, we sequenced the genomes of these three hybrids and used 1,014,134 species-specific SNPs to evaluate the genetic similarity, and found more than 96% regions are similar to that of sweet orange (Fig. 4c,d and Supplementary Table 17). In addition, we observed that hybrids with a specific

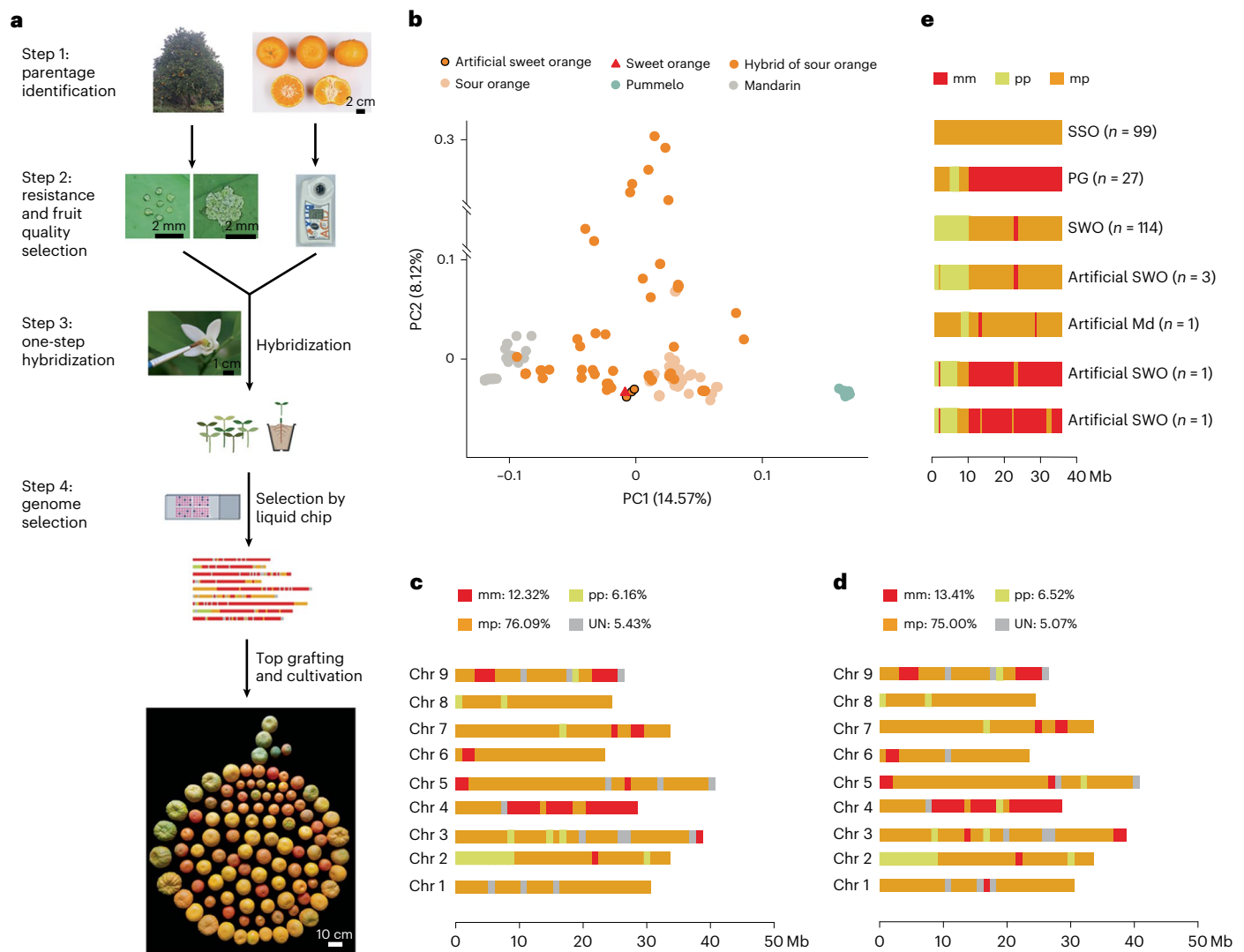


Fig. 4 | De novo domestication of sweet orange. **a**, The process of de novo domestication of sweet orange. **b**, PCA of the hybrids derived from a cross between sour orange and Ponkan mandarin. Orange circles with black borders represent artificial sweet orange: hybrid 228, hybrid 24 and hybrid 55. **c,d**, Genomic pattern of artificial sweet orange (**c**) and cultivated sweet orange (**d**). mm, homozygous segment with a mandarin origin only; mp, heterozygous

segment with a mandarin/pummelo origin; pp, homozygous segment with a pummelo origin only; UN, unknown region. The numbers represent the percentage of each genomic region in the genome. Genomic patterns were evaluated by 1,014,134 species-specific SNPs²⁷. **e**, Genomic variations for chromosome 2 in SWO, SSO, PG, artificial SWO and artificial mandarin (Md). The number of accessions for each variety is indicated.

region of homozygous pummelo haplotypes over the 1.8–5.8 Mb region on chromosome 2 showed sweet orange-like phenotypes (Fig. 4e and Supplementary Fig. 21). RNA-sequencing analysis indicated 36 genes were specifically expressed in sweet orange fruit peels and 27 genes were specifically expressed in sweet orange fruit flesh (Supplementary Table 18). Therefore, our hybridization experiments support the model that sweet orange was derived from a sour orange × mandarin cross.

Metabolites selection for canker-resistant sweet orange

To explore the molecular basis of *Xcc* resistance in sour orange, we examined the expression and sequence variation of the *Xcc* susceptibility gene LATERAL ORGAN BOUNDARIES 1 (*CsLOB1*)²⁸. There was no sequence polymorphism or obvious difference in gene expression levels between resistant and susceptible genotypes (Supplementary Fig. 26). On the basis of these data, we suggest that resistance to *Xcc* in these hybrids was not due to a *CsLOB1*-mediated mechanism. Sour orange is widely used as a type of Chinese folk medicine and accumulates high levels of metabolites with antibacterial activity. Therefore, we

hypothesized that the *Xcc* resistance of sour orange may be attributed to the accumulation of antibacterial metabolites. To test this idea, we performed a comparative metabolomic analysis using leaves and fruits from the *Xcc*-resistant and *Xcc*-susceptible hybrids. We identified and characterized 2,027 metabolites that included 690 flavonoids, 244 phenolic acids, 198 alkaloids, 189 amino acids and their derivatives, 154 lignans and coumarins, 122 lipids, 84 terpenoids, 83 organic acids, 68 nucleotides and their derivatives, and 195 other compounds (Supplementary Tables 19 and 20). Among these metabolites, we detected 215 metabolites with significantly higher levels in the leaves of all three *Xcc*-resistant accessions relative to the three *Xcc*-susceptible accessions (*P* value < 0.05; Fig. 5a and Supplementary Table 21). Additionally, most of these 215 metabolites accumulated at low levels in the fruits of all tested samples (Fig. 5a and Supplementary Table 21).

Most of the 215 metabolites that accumulated to high levels only in *Xcc*-resistant accessions are derivatives of different types of metabolites and some of these metabolites are not commercially available. To identify defense-related metabolites with application potential, we chose 16 typical compounds that are commercially

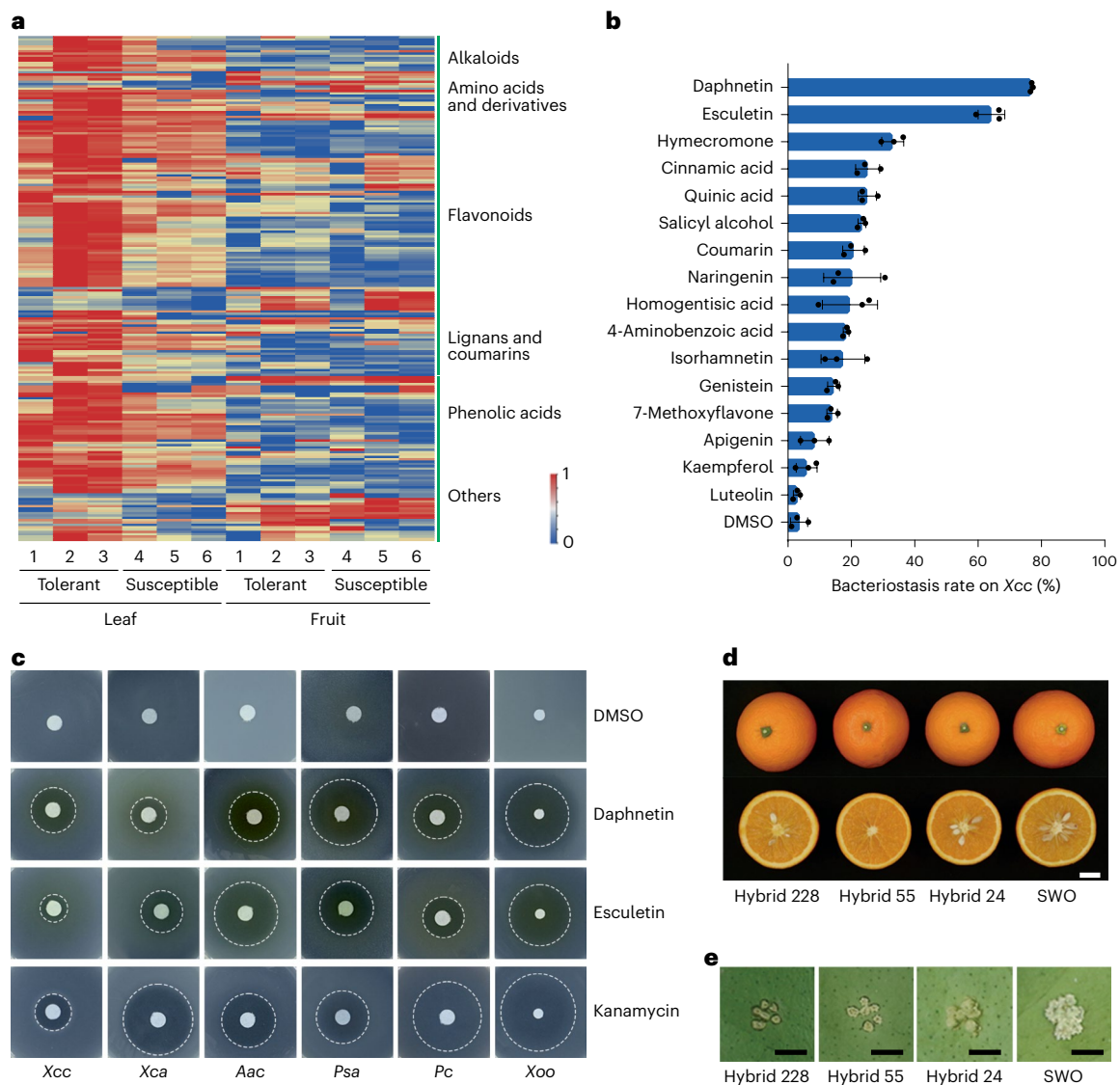


Fig. 5 | Identification of defense-related metabolites and selection of canker-resistant sweet orange. a, Heat map for the levels of 215 metabolites that accumulate to higher levels in the leaves of *Xcc*-resistant accessions compared with *Xcc*-susceptible accessions ($P < 0.05$): 1, sour orange; 2, hybrid 239; 3, hybrid 331; 4, Ponkan mandarin; 5, hybrid 374; 6, hybrid 115. Blue represents low levels and red represents high levels. The content of metabolites were transformed by \log_2 and the normalized output of each row ranges from 0 to 1. **b**, Inhibitory effect of disease resistance-related metabolites (100 $\mu\text{g ml}^{-1}$) on the growth of *Xcc*. Data are presented as mean values \pm s.d.; $n = 3$ independent replicates. **c**, Disk diffusion assay showing the antibacterial activity of daphnetin and esculetin (50 mg ml⁻¹

dissolved in DMSO) on the growth of 6 species of pathogenic bacteria. Kanamycin (25 mg ml⁻¹ dissolved in pure water) and the solvent DMSO were used as positive and negative controls, respectively. Larger diameters for the inhibition zones around the paper disks indicate more antibacterial activity. The inhibition zones are indicated with dashed circles. The diameter of the round white filter paper was 6 mm. **d**, The appearance of fruits from sweet orange and artificial sweet orange. Scale bar, 2 cm. **e**, Leaves from sweet orange and artificial sweet orange were inoculated with *Xcc* (10⁸ CFU ml⁻¹). The photographs were taken 12 days after inoculation. Scale bars, 2 mm.

available and evaluated their antibacterial activity on *Xcc* (Fig. 5b). We observed that 13 of the 16 metabolites had bacteriostatic effects on *Xcc*. Among these metabolites, we observed a 10% bacteriostasis rate for four flavonoid compounds—naringenin, isorhamnetin, genistein and 7-methoxyflavone. In particular, two coumarin metabolites, daphnetin and esculetin, which have been reported to have antibacterial activity toward both *Ralstonia solanacearum*²⁹ and *Ralstonia pseudosolanacearum*³⁰, strongly inhibited the growth of *Xcc*, with a bacteriostasis rate of more than 60% (Fig. 5b). To further explore the broad-spectrum antibacterial effect of daphnetin and esculetin, we used a disk diffusion method to examine the bacteriostasis of the two metabolites on *Xcc* and five other pathogenic bacteria, including *Xanthomonas campestris* pv. *campestris* (*Xca*, a causal agent of the black rot diseases of crucifers), *Acidovorax avenae* subsp. *citrulli* (*Aac*, a causal

agent of fruit blotch disease in watermelon), *Pseudomonas syringae* pv. *actinidiae* (*Psa*, a causal agent of bacterial canker in kiwifruit), *Pseudomonas cannabina* pv. *alisalensis* (*Pc*, a causal agent of bacterial blight in crucifers) and *Xanthomonas oryzae* pv. *oryzae* (*Xoo*, a causal agent of bacterial leaf streak disease in rice). We found that both daphnetin and esculetin can inhibit the growth of all six kinds of pathogenic bacteria to different degrees, as demonstrated by the diameters of inhibition zones that ranged from 10.9 to 42.0 mm (Fig. 5c and Supplementary Fig. 27). Notably, the inhibitory effect of daphnetin on the growth of *Xcc* and *Psa*, and esculetin on the growth of *Aac* was similar to or even larger than the inhibitory effects of kanamycin (Fig. 5c).

To select canker-resistant hybrids, we used metabolic assays to evaluate 29 hybrids. The results indicated that five hybrids accumulated higher levels of daphnetin and *O*-feruloyl 2-hydroxycoumarin

(a derivative of esculetin) than ‘ZGSC’ sour orange (Supplementary Table 22). *Xcc* inoculation experiments revealed that three of the hybrids were resistant to citrus canker and the other two were nearly resistant to citrus canker (Supplementary Figs. 25 and 28). Notably, among the three artificial sweet orange hybrids mentioned above (that is, hybrid 228, hybrid 24 and hybrid 55), hybrid 228 accumulated higher levels of daphnetin and *O*-feruloyl 2-hydroxycoumarin than ‘ZGSC’ (Supplementary Table 22) and was resistant to citrus canker (Fig. 5d,e).

To explore genes related to disease resistance, we generated 1.68 Tb of RNA-sequencing data from the same samples used for the metabolome analysis described above (Supplementary Table 23). From a differential gene expression analysis, we identified 248 significantly upregulated genes in the leaves of the *Xcc*-resistant accessions relative to *Xcc*-susceptible accessions (Supplementary Table 24). We noticed that some genes encoding enzymes involved in coumarin and flavonoid biosynthesis were among the upregulated genes (Supplementary Figs. 29 and 30 and Supplementary Table 25), which is consistent with the high-level accumulation of defense-related metabolites in the resistant accessions (Fig. 5b).

Discussion

Our genomic data, together with evidence from extant native populations and ancient vernaculars, support the conclusion that sweet orange was domesticated in southern China. Prototypic forms with high fruit acidity and regionally distinct sweet oranges were discovered in this region¹⁵. Citrus has been cultivated for at least 4,000 years in this region, which has highly diverse citrus resources including abundant wild forms, semi-domesticated and cultivated sour oranges, mandarins and pummelos. Records indicate that citrus were initially commercially produced in this region, called the Chu state or the Jingzhou region, which corresponds approximately to the regions encompassed by present-day Hunan and Hubei provinces³¹. More archeological evidence comes from herbarium specimens of citrus seeds from the ‘Ma Wang Dui tomb of the Han dynasty’ (first century BC) in this region^{31,32}.

Compared with previous studies on the origin of sweet orange, our study inferred a reasonable model with the extant germplasm of both parents of sweet orange in the NYW region of southern China. A previous study proposed that sweet orange originated from (pummelo × mandarin) × mandarin²⁴. Another previous study proposed that sweet orange had undergone multiple pummelo introgression events³³ and proposed that sour orange was not related to sweet orange²⁵. Our study inferred a simple model for the origin of sweet orange. We propose that sweet orange is hybrid between sour orange and Ponkan-like mandarin based on a high-resolution haplotype-resolved genome analysis that we confirmed with artificial hybridization experiments. As sweet orange and sour orange both are interspecies hybrids with high levels of heterozygosity, a previous genomic analysis that utilized only one haploid genome may lack sufficient allele information. In this study, we clarified the genetic composition of phased haplotypes for both sweet orange and sour orange, allowing for a thorough consideration of the genetic characteristics of sweet oranges. The debate on the different models for the origin of sweet orange has focused on the 1.8–5.8 Mb homozygous pummelo/pummelo (P/P) segment on chromosome 2 of sweet orange. We defined the distribution of the P/P segment in all accessions and cross-referenced them with the phenotypes of the hybrids of sour orange and Ponkan mandarin. Our data revealed that both the sour orange and Ponkan mandarin possess one allele of the pummelo segment and thus, provide a reasonable explanation for the debated P/P segment of sweet oranges (Fig. 4e and Supplementary Fig. 22). We also observed sweet orange-like phenotype among the hybrids with homozygous pummelo haplotypes in this region (Fig. 4e and Supplementary Fig. 23). We further designed hybridization experiment by using sour orange and Ponkan mandarin as parents and created artificial sweet oranges that are very close to commercial sweet

oranges, which provides strong evidence for our origin model of sweet orange. The existence of all parental species in the same sympatric region in the NYW region of southern China provides more support for our model for the origin of sweet orange.

Defense-related metabolites are promising candidates for environmentally friendly bactericides. Using resistant accessions of sour orange, we identified 13 citrus metabolites that can inhibit the growth of *Xcc* with different bacteriostasis rates. We also found broad-spectrum antibacterial activities for daphnetin and esculetin against six kinds of industrially important pathogenic bacteria in *in vitro* experiments, and the inhibitory effects of three combinations were even larger than the kanamycin. A previous study reported that daphnetin and esculetin have antibacterial activity toward both *R. solanacearum*²⁹ and *R. pseudosolanacearum*³⁰. These metabolites are promising candidates and should be tested with field experiments that aim to identify green pesticides with antibactericidal activities. In the future, breeding programs that aim to develop disease-resistant hybrids can use bactericidal metabolites as effective markers for disease-resistant hybrids.

This study developed a strategy that uses sour orange (a kind of wild citrus) for *de novo* domestication of sweet oranges. By choosing highly heterozygous parents and designing a liquid gene chip with high selection efficiency, we obtained diverse sweet oranges with broad variations in their genome sequences, metabolite levels and disease resistance. This study also provides a proof-of-concept for metabolite selection in perennials.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-025-02122-4>.

References

- Lam, H. M. et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* **42**, 1053–1059 (2010).
- Lin, T. et al. Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* **46**, 1220–1226 (2014).
- Wright, S. I. et al. The effects of artificial selection on the maize genome. *Science* **308**, 1310–1314 (2005).
- Xie, X. & Liu, Y. G. *De novo* domestication towards new crops. *Natl Sci. Rev.* **8**, 033 (2021).
- Yu, H. et al. A route to *de novo* domestication of wild allotetraploid rice. *Cell* **184**, 1156–1170 (2021).
- Zhang, C. et al. Genome design of hybrid potato. *Cell* **184**, 3873–3883 (2021).
- Xie, Y., Zhang, T., Huang, X. & Xu, C. A two-in-one breeding strategy boosts rapid utilization of wild species and elite cultivars. *Plant Biotechnol. J.* **20**, 800–802 (2022).
- Webber, H. J. in *The Citrus Industry* (eds Reuther, W. et al.) Ch. 1 (Univ. California Press, 1967).
- Talon, M. & Gmitter, F. G. Citrus genomics. *Int. J. Plant Genomics* **2008**, 528361 (2008).
- Li, Z. et al. A potential disease susceptibility gene *CsLOB* of citrus is targeted by a major virulence effector *PthA* of *Xanthomonas citri* subsp. *citri*. *Mol. Plant* **7**, 912–915 (2014).
- Stover, E. et al. Incidence and severity of Asiatic citrus canker on diverse citrus and citrus-related germplasm in a Florida field planting. *HortScience* **49**, 4–9 (2014).
- Erb, M. & Kliebenstein, D. J. Plant secondary metabolites as defenses, regulators, and primary metabolites: the blurred functional trichotomy. *Plant Physiol.* **184**, 39–52 (2020).
- Wang, W. et al. Rice secondary metabolites: structures, roles, biosynthesis, and metabolic regulation. *Molecules* **23**, 3098 (2018).

14. Chen, S., Sun, B., Shi, Z., Miao, X. & Li, H. Identification of the rice genes and metabolites involved in dual resistance against brown planthopper and rice blast fungus. *Plant Cell Environ.* **45**, 1914–1929 (2022).
15. Bednarek, P. et al. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* **323**, 101–106 (2009).
16. Beyer, S. F. et al. The *Arabidopsis* non-host defence-associated coumarin scopoletin protects soybean from Asian soybean rust. *Plant J.* **99**, 397–413 (2019).
17. Jeon, J. E. et al. A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* **180**, 176–187 (2020).
18. Norvienyeku, J. et al. Bayogenin 3-O-cellobioside confers non-cultivar-specific defence against the rice blast fungus *Pyricularia oryzae*. *Plant Biotechnol. J.* **19**, 589–601 (2021).
19. Dugrand, A. et al. Coumarin and furanocoumarin quantitation in citrus peel via ultraperformance liquid chromatography coupled with mass spectrometry (UPLC–MS). *J. Agric. Food Chem.* **61**, 10677–10684 (2013).
20. Rao, M. J., Ding, F., Wang, N., Deng, X. & Xu, Q. Metabolic mechanisms of host species against citrus Huanglongbing (greening disease). *Crit. Rev. Plant Sci.* **37**, 496–511 (2019).
21. Roowi, S. & Crozier, A. Flavonoids in tropical citrus species. *J. Agric. Food Chem.* **59**, 12217–12225 (2011).
22. Dixon, R. A. Natural products and plant disease resistance. *Nature* **411**, 843–847 (2001).
23. Rao, M. J., Zuo, H. & Xu, Q. Genomic insights into citrus domestication and its important agronomic traits. *Plant Commun.* **2**, 100138 (2021).
24. Xu, Q. et al. The draft genome of sweet orange (*Citrus sinensis*). *Nat. Genet.* **45**, 59–66 (2013).
25. Wu, G. A. et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat. Biotechnol.* **32**, 656–662 (2014).
26. Wang, L. et al. Somatic variations led to the selection of acidic and acidless orange cultivars. *Nat. Plants* **7**, 954–965 (2021).
27. Wang, L. et al. Genome of wild mandarin and domestication history of mandarin. *Mol. Plant* **11**, 1024–1037 (2018).
28. Hu, Y. et al. Lateral organ boundaries 1 is a disease susceptibility gene for citrus bacterial canker disease. *Proc. Natl Acad. Sci. USA* **111**, 521–529 (2014).
29. Yang, L. et al. New insights into the antibacterial activity of hydroxycoumarins against *Ralstonia solanacearum*. *Molecules* **21**, 468 (2016).
30. Yang, L. et al. Hydroxycoumarins: new, effective plant-derived compounds reduce *Ralstonia pseudosolanacearum* populations and control tobacco bacterial wilt. *Microbiol. Res.* **215**, 15–21 (2018).
31. Chinese Citrus Society *Chinese Citrus Industry* (China Agriculture Press, 2008).
32. Dugo, G. & Giacomo, A. D. *Citrus: The Genus of Citrus* (Taylor & Francis, 2002).
33. Wu, G. A. et al. Genomics of the origin and evolution of *Citrus*. *Nature* **554**, 311–316 (2018).
34. Xia, Q. M. et al. Localization and characterization of citrus centromeres by combining half-tetrad analysis and CenH3-associated sequence profiling. *Plant Cell Rep.* **39**, 1609–1622 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025

Methods

Plant materials and whole-genome sequencing

A total of 226 citrus cultivars and landraces (99 sour oranges, 13 sweet oranges, 15 mandarins, 43 pummelos, 19 citrus hybrids, 32 hybrids of sour orange and 5 others citrus accessions) were sequenced in this study (Supplementary Table 1). At least 10 µg of genomic DNA from each accession was used to construct a sequencing library. Paired-end sequencing libraries with an insert size of approximately 200–500 bp were constructed and sequenced on the Illumina HiSeq 2500 platform.

Genome assembly

We assembled six genomes including: *Citrus sinensis* ‘Valencia’, *Citrus aurantium* ‘ZGSC’, *Citrus reticulata* ‘Mangshan’, *Citrus maxima* ‘Zipi’, *Citrus reticulata* ‘Ponkan’ and the natural hybrid sweet orange ‘TJH2’ found in southern China. PacBio HiFi reads and Nanopore ultralong reads were generated for the haplotype-resolved telomere-to-telomere genomes assemblies of *Citrus sinensis* ‘Valencia’ and *Citrus aurantium* ‘ZGSC’. Nanopore ultralong reads were generated for the chromosome-level assemblies of *Citrus reticulata* ‘Mangshan’, *Citrus maxima* ‘Zipi’, *Citrus reticulata* ‘Ponkan’ and the natural hybrid sweet orange ‘TJH2’.

HiFi reads were assembled using Hifiasm (v0.16.1)³⁵ and canu (v2.2)³⁶ with default parameters. Nanopore ultralong reads were assembled using Nextdenovo (v2.5.2)³⁷, Necat (v1.0)³⁸ and SMARTdenovo (v1.0)³⁹. Draft contigs generated using Hifiasm and Nextdenovo were chosen for the following analysis if they passed the quality assessment. Contigs were then polished using Nextpolish (v1.4.0)⁴⁰ with Illumina reads. The polished contigs were then clustered, ordered and oriented to 18 chromosomes using the 3d-dna pipeline (v180922)⁴¹, which contained two sets of genomes. The genomes assembled using Nanopore reads were used to fill gaps in the genomes assembled using Hifiasm and TGS-Gapcloser (v1.1.1)⁴². Another round of gap filling was performed using corrected Nanopore reads, generating a sweet orange assembly without gaps. The plant telomeric sequences (CCCTAAA) were used to identify telomeres in the assembled genome.

Nanopore reads (~10 kb) were generated for three species, including *Citrus reticulata* ‘Ponkan’, *Citrus maxima* ‘ZPY’ and a natural hybrid sweet orange named ‘TJH2’. In addition, we generated Nanopore ultralong reads (~20 kb) for *Citrus reticulata* ‘Mangshan’. These Nanopore reads were assembled using wtdbg (v2.3)⁴³, Nextdenovo (v2.5.2), Necat (v1.0), SMARTdenovo (v1.0) and Flye (v2.8.3)⁴⁴. Then, three rounds of genome polishing were performed to correct assembly errors. The first polishing was performed using racon (v1.5.0)⁴⁵ with Nanopore raw reads. The next two rounds of polishing were performed using Nextpolish (v1.4.0) with Illumina reads. For *Citrus reticulata* ‘Mangshan’, *Citrus reticulata* ‘Ponkan’ and a natural hybrid sweet orange named ‘TJH2’ found in southern China, we generated Hi-C reads to anchor contigs using the 3d-dna pipeline (v180922). For *Citrus maxima* ‘ZPY’, the contigs were anchored based on collinearity with the chromosome-level genome of *Citrus maxima* ‘Wanbai’.

Genome annotation

De novo transposable element libraries were predicted using RepeatModeler2 (v1.0.11)⁴⁶. The transposable element library generated above was used to identify repetitive sequences in particular genomes using RepeatMasker (v4.0.9)⁴⁷. Gene models were annotated by integrating evidence from ab initio gene predictions, homology searches and RNA-sequencing. For ab initio gene predictions, AUGUSTUS (v3.3.3)⁴⁸, SNAP⁴⁹ and GlimmerHMM (v3.0.4)⁵⁰ were used. Homology searching was then conducted using the GenomeThreader software⁵¹. For RNA-sequencing reads, Hisat2 (v2.2.1)⁵² was used for sequence alignments and Cufflinks (v2.2.1)⁵³ was used to construct transcripts. Transcripts were assembled both de novo and genome guided using Trinity (v2.4.0)⁵⁴. The PASA (v2.4.1)⁵⁵ software was used to build a comprehensive transcriptome database for de novo transcriptome assemblies. Correcting exon boundaries, adding untranslated regions

and modeling alternative splicing were based on the PASA alignment assemblies. All the above gene prediction results were integrated using EVM (v1.1.1)⁵⁶ to generate a final structural annotation. Functions of the predicted genes were assigned using eggNOG-mapper⁵⁷.

The approaches of FISH

FISH was performed as previously described by Xia et al.³⁴ with minor modifications. Root tips were harvested from germinated *C. sinensis* ‘Valencia’. After pretreatment with 0.075 M KCl in saturated *p*-dichlorobenzene, the root tips were fixed in fresh Carnot (ethanol:glacial acetic acid, 3:1, v:v) at room temperature for 24 h, and then replaced with 70% ethanol and stored at –20 °C. The CL34contig88 served as a centromeric probe³⁴, was labeled with biotin-16-dUTP (Roche) and was subsequently detected with Cy3-conjugated streptavidin (Sigma). Telomeric probes were synthesized (Sango Biotech) and attached to carboxyfluorescein (FAM). Dual-color FISH was conducted with these two probes. The slides were counterstained with 4,6-diamidino-2-phenylindole (DAPI, Invitrogen) and observed with an Axio imager and an M2 fluorescence microscope (Zeiss). The images were processed and adjusted using ZEN 2 (blue edition) and the Adobe Photoshop 2023 software.

Comparative genomic analyses

Seven representative citrus species were selected for comparative genomic analyses, including *C. sinensis*, *C. reticulata* ‘Mangshan’, *C. maxima* ‘ZPY’, *C. reticulata* ‘Ponkan’, *C. aurantium* ‘ZGSC’, *C. medica* and *Atalantia buxifolia*. Pairwise comparisons (‘SSO-M’ versus *C. reticulata* ‘Mangshan’, ‘SSO-P’ versus *C. maxima* ‘ZPY’, ‘SWO-M’ versus *C. reticulata* ‘Mangshan’, ‘SWO-P’ versus *C. maxima* ‘ZPY’) of genomic sequences were made using LASTZ (v1.02.00)⁵⁸ and the nucmer program from the MUMmer (v4.0.0)⁵⁹ software package using ‘-maxgap = 500 -mincluster = 100 -q -r’. The alignment results were further filtered using the delta-filter program in the MUMmer package. SNPs and structural variants were obtained using the show-diff and show-snps program in the MUMmer package. Genes were clustered using OrthoFinder2 (v2.2.7)⁶⁰ with default parameters. A specific gene family means that this gene family only exists in one species. The Gene Ontology enrichment analyses were performed using the agriGO (v2.0)⁶¹ program.

Genetic diversity and nucleotide diversity (π) analysis

Genomic data from seven citrus species were collected and single-copy genes were identified using OrthoFinder2 (v2.2.7). For each pair of these genes, we used the MUSCLE (v3.8.31)⁶² software for sequence alignment. Aligned amino acid sequences were converted into the corresponding CDS sequences and concatenated. A maximum likelihood tree was produced using *A. buxifolia* as outgroup and using the substitution model GTRGAMMA from the RAxML (v7.7.8) software⁶³, with a total of 100 rapid bootstrap values. High-quality paired-end reads from each accession were mapped to the sweet orange genome using BWA (v0.7.17)⁶⁴ and SAMtools (v1.9)⁶⁵. PCR duplicates were marked using the MarkDuplicates command from GATK (v4.1.1)⁶⁶. SNPs were called using HaplotypeCaller command from GATK (v4.1.1). Raw VCF files were filtered using Vcftools (v0.1.13)⁶⁷. Then, we extracted all 4dTV sites in the sweet orange genome to construct a maximum likelihood phylogenetic tree using RAxML (v7.7.8) with a total of 200 rapid bootstrap values. PCA based on all the SNPs was conducted with default settings using GCTA (v1.26.0)⁶⁸. We used PLINK (v1.90)⁶⁹ for format conversion and used the R package prcomp to calculate PC1 and PC2. For nucleotide diversity (π) analysis, SNPs from the whole genome were used, and a sliding-window approach (1 Mb windows sliding in 500 kb steps) was employed to quantify nucleotide diversity (π) for each pair of citrus populations using the VCFtools software.

Evaluation of genetic relationships

We inferred genetic relationships based on the haplotype polymorphisms at the population level^{70,71}. Before kinship calculations, we first

pruned the SNPs called from previous sections ($r^2 > 0.9$) using plink (v1.90) and phased the filtered SNPs using the beagle 4.0 software (vr1399)⁷². Then, we used ChromoPainter and the FineSTRUCTURE pipeline⁷¹ to analyze 305 citrus resources. We constructed the genetic relationship at the haplotype level, and then calculated the genetic relationship matrix to investigate what proportion of the haplotype each individual contributed to the remaining individuals. In addition, we also calculated IBD and IBS to assess the extent of relatedness among different citrus accessions to sweet orange using PLINK (v1.90). For the evaluation of genetic similarity between hybrids and sweet oranges, the genomic pattern were evaluated by 1,014,134 species-specific SNPs with 100 kb windows refer to the method of Wang et al.²⁷.

Assay of pathogenicity

Fully expanded leaves were inoculated with *Xcc* (10^8 colony-forming units (CFU) per milliliter) with an inoculating needle (0.5 mm in diameter). Each inoculation spot consisted of seven pricks with the inoculation needle as recommended by previous reports with minor modifications. A bacterial suspension was dropped into each inoculation spot. The disease lesion area was measured (36 punctures on average) with ImageJ 2.0.

Hybrid fruit quality determinations

Total soluble solids (TSS) and titratable acid (TA) was determined using a PAL-BX/ACID sugar and acid integrated machine. The data collection was repeated three times. A single fruit weight was determined using an electronic balance with a sensitivity of 0.01 to measure the weight of the fruit. The masses of three fruits were quantified for each genotype. Mean values are reported. The fruit shape index is the longitudinal diameter value/transverse diameter value. A vernier caliper was used to measure the distance from the top of the fruit to the pedicle as the value for the longitudinal diameter (mm), and the cross-sectional diameter of the fruit was used as the value of the transverse meridian (mm). Three fruits were measured from each fruit tree. Mean values are reported.

Metabolomics profiling and analyses

The freeze-dried leaf and fruit samples were powdered using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. First, 100 mg powder was weighed and extracted overnight at 4 °C with 1.0 ml 70% aqueous methanol. Following centrifugation at 10,000 g for 10 min, the extracts were absorbed (CNWBOND Carbon-GCB SPE Cartridge, 250 mg, 3 ml; ANPEL) and filtered (SCAA-104, 0.22 µm pore size; ANPEL) before liquid chromatography–mass spectrometry (LC–MS) analysis.

The sample extracts were analyzed using an LC–ESI–MS/MS system (HPLC, Shim-pack UFLC SHIMADZU CBM30A system; MS, Applied Biosystems 4500 Q TRAP). The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (QTRAP)–MS. LIT and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (Q TRAP), API 4500 Q TRAP LC–MS/MS System, equipped with an ESI Turbo Ion-Spray interface, operating in a positive ion mode and controlled by Analyst 1.6 software (AB Sciex). Orthogonal projections to latent structures–discriminant analysis was applied to identify altered metabolites.

The coefficient of variation was calculated for each metabolite on the basis of the following formula: δ/μ , where δ and μ are the standard deviation and mean of each metabolite in the population, respectively.

Measurement of the antimicrobial activity of metabolites

Six kinds of pathogenic bacteria (*Xcc*, *Xca*, *Aac*, *Psa*, *Pc* and *Xoo*) were plated on Luria-Bertani (LB) solid medium separately and incubated at 28 °C for 14 h. Individual colonies were cultured separately in LB liquid medium at 28 °C on an orbital shaker at 200 rpm for 18 h.

To test the 16 metabolites for antibacterial activity, the *Xcc* bacterial cells were resuspended in sterile water at a concentration of 10^6 CFU ml^{−1} and 0.1 ml of the suspension was added to 5 ml of LB

liquid medium containing different metabolites at a concentration of 100 µg ml^{−1}. Dimethylsulfoxide (DMSO) without metabolites (0 mg ml^{−1}) served as a control. All cultures were shaken (200 rpm) at 28 °C in the dark, and the OD₆₀₀ was measured every 6 h until the bacterial culture reached the stationary phase. The OD₆₀₀ values of all the cultures at 36 h were used to calculate the bacteriostasis rate which is expressed as (OD₆₀₀ of DMSO culture – OD₆₀₀ of metabolite culture)/OD₆₀₀ of DMSO culture × 100%.

To explore the antibacterial effect of daphnetin and esculetin on the six species of pathogenic bacteria, the disk diffusion method for assessing antibacterial activity was performed⁷³. In brief, melted LB solid medium cooled down to about 45 °C was mixed with different bacteria and spread evenly on Petri dishes in sterile conditions. The bacterial suspension concentration of LB solid medium was about 10^6 CFU ml^{−1}. Filter paper disks (6 mm diameter) were soaked with different metabolites at a concentration of 50 mg ml^{−1}. The impregnated disks were symmetrically placed onto the medium using sterile tweezers. Filter paper disks separately soaked with DMSO and 25 mg ml^{−1} kanamycin were used as negative and positive controls, respectively. The Petri dishes were incubated at 28 °C for 36 h. Antibacterial activity was evaluated by measuring the diameter of the inhibition zone, the area around the disks with no bacterial growth. Each experiment was performed three times, with three replicates per experiment.

Geographical map generation

The geographical map was generated manually using the freeform shape function of Microsoft PowerPoint software.

Statistics and reproducibility

A two-tailed Student's *t*-test was used to compare the relative differences in gene expression, metabolite content, the ratio of TSS to TA and the antimicrobial activity of metabolites between two groups of samples. A two-tailed adjusted Fisher's exact test was used to conduct the Gene Ontology enrichment analysis of the target genes relative to the background. All experiments were replicated with at least three biological replicates.

Reporting summary

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

Data availability

Genome assemblies have been deposited in the National Center for Biotechnology Information under accessions [PRJNA1003804](#), [PRJNA1199310](#), [PRJNA1199315](#), [PRJNA1199309](#) and [PRJNA1199314](#) with nos. JBJYGS000000000 (*Citrus reticulata* 'Mangshan'), JBJYGR000000000 (*Citrus maxima* 'Zipi'), JBJYGQ000000000 (*Citrus reticulata* 'Ponkan'), JBKACN000000000 (natural hybrid sweet orange 'TJH2'), JBKABO000000000 (mandarin haplotype of *Citrus sinensis* 'Valencia'), JBKABP000000000 (pummelo haplotype of *Citrus sinensis* 'Valencia'), JBKFCH000000000 (mandarin haplotype of *Citrus aurantium* 'ZGSC') and JBKFCI000000000 (pummelo haplotype of *Citrus aurantium* 'ZGSC'). We deposited all the raw sequencing data generated in this article in the National Center for Biotechnology Information database under accessions [PRJNA1003804](#), [PRJNA1003792](#) and [PRJNA1003676](#). Also, we deposited all the raw sequencing data and assembled genomes used in this article in the National Genomics Data Center (<https://ngdc.cnbc.ac.cn/>) under accession [PRJCA019821](#). The assembled genomes and annotations are also available at <http://citrus.hzau.edu.cn/download.php>. The accession numbers for the whole-genome sequencing data are listed in Supplementary Table 1. The accession numbers for the RNA-sequencing data are listed in Supplementary Table 23. The datasets are available via figshare at <https://figshare.com/s/8d244c4dflafab94cc04> (ref. 74). Source data are provided with this paper.

Code availability

The code used in this article is available via figshare at <https://doi.org/10.6084/m9.figshare.26053999.v2> (ref. 75).

References

35. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
36. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
37. Hu, J. et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* **25**, 107 (2024).
38. Chen, Y. et al. Fast and accurate assembly of Nanopore reads via progressive error correction and adaptive read selection. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.02.01.930107> (2020).
39. Liu, H., Wu, S., Li, A. & Ruan, J. SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte* <https://doi.org/10.46471/gigabyte.15> (2021).
40. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
41. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
42. Xu, M. et al. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience* **9**, giaa094 (2020).
43. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
44. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
45. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
46. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
47. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **4**, 4.10.1–4.10.14 (2009).
48. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
49. Johnson, A. D. et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939 (2008).
50. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
51. Gremme, G., Brendel, V., Sparks, M. E. & Kurtz, S. Engineering a software tool for gene structure prediction in higher organisms. *Inf. Softw. Technol.* **47**, 965–978 (2005).
52. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
53. Trapnell, C. et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
54. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
55. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
56. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
57. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
58. Harris, R. S. *Improved Pairwise Alignment of Genomic DNA*. PhD thesis, Pennsylvania State Univ. (2007).
59. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
60. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
61. Tian, T. et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **45**, W122–W129 (2017).
62. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
63. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
64. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
65. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
66. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
67. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
68. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
69. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
70. Busby, G. B. et al. Admixture into and within sub-Saharan Africa. *eLife* **5**, e15266 (2016).
71. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
72. Browning, B. L. & Browning, S. R. J. G. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
73. Bauer, A. W., Kirby, W. M., Sherris, J. C. & Turck, M. Antibiotic susceptibility testing by a standardized single disk method. *Am. J. Clin. Pathol.* **45**, 493–496 (1966).
74. Liu, S. Origin and de novo domestication of sweet orange. *figshare* <https://figshare.com/s/8d244c4df1afab94cc04> (2024).
75. Liu, S. Code for the study of origin and de novo domestication of sweet orange. *figshare* <https://doi.org/10.6084/m9.figshare.26053999.v2> (2024).

Acknowledgements

We are grateful to E. S. Buckler from the Institute for Genomic Diversity, Cornell University and W. B. Xie from the National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University for their advice on our bioinformatic analysis. We are grateful to X. F. Ma of the College of Horticulture, Hunan Agricultural University for providing us with Xcc. We appreciate G. Lu, D. Y. Guo and L. L. Zhong for the support of plant growth and metabolic platform from National Key Laboratory for Germplasm Innovation and Utilization of Horticultural Crops, Huazhong Agricultural University. This project was financially supported by grants from the National Natural Science Foundation

of China to Q.X. (grant numbers 31925034 and U23A20198), the National Key Research and Development Program of China to Q.X. (grant number 2022YFF1003100), the Key Research and Development Program of Hubei to X.W. (grant number 2022BBA155), the Foundation of Hubei Hongshan Laboratory to Q.X. (grant number 2021hszd016), Key Project of Hubei Provincial Natural Science Foundation to Q.X. (grant number 2021CFA017), Postdoctoral Innovation Research Post in Hubei Province to Y.X. and the National Postdoctoral Program for Innovative Talents (grant number BX20200146) to Y.X.

Author contributions

Q.X. conceived and designed the project and the strategy. S.L. performed the genome and origin analysis, and transcriptome and metabolic analysis. Y.X. collected and created the samples with contributions from L.C., J.Y., Z.X., F.H., J.L., P.W., P.C., X.Z., Z. Liu and Z. Lu. Y.X. identified the phenotype of disease resistance and antibacterial metabolites with help from K.Y., S.L., S.C., X.G. and W.Z. Q.X. coordinated the project with help from X.D., X.W., Y.Z., W.G., Z. Lin, F.Z., Z.P., W.J. and F.D. S.L., Y.X. and Q.X. wrote the paper with contributions from Y.Z., Y.H., L.W., G.X., G.L. and R.M.L.

Competing interests

The authors declare no competing interests.

Additional information

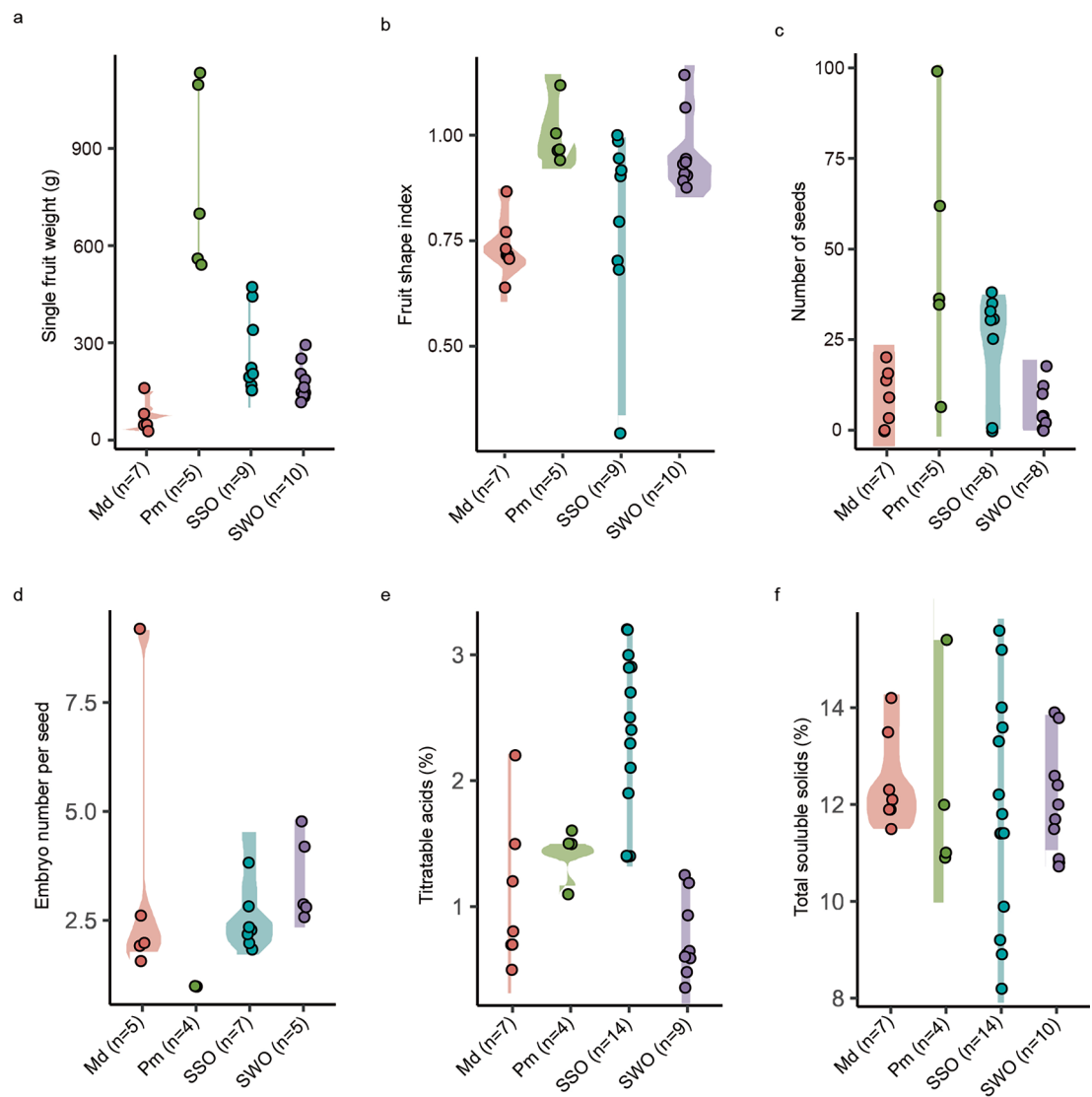
Extended data is available for this paper at <https://doi.org/10.1038/s41588-025-02122-4>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-025-02122-4>.

Correspondence and requests for materials should be addressed to Qiang Xu.

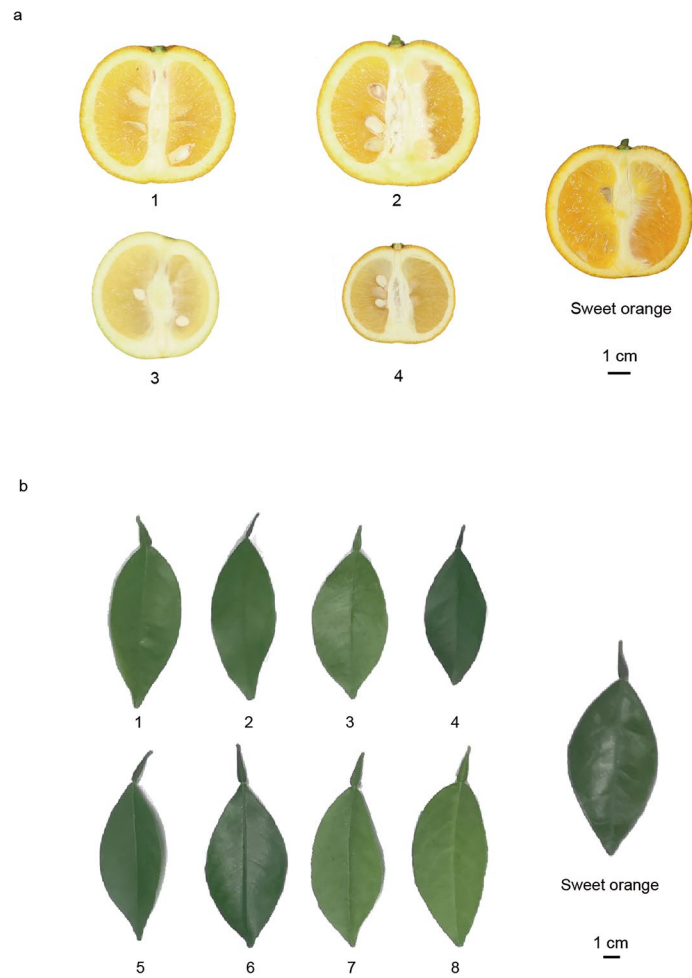
Peer review information *Nature Genetics* thanks Massimo Iorizzo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

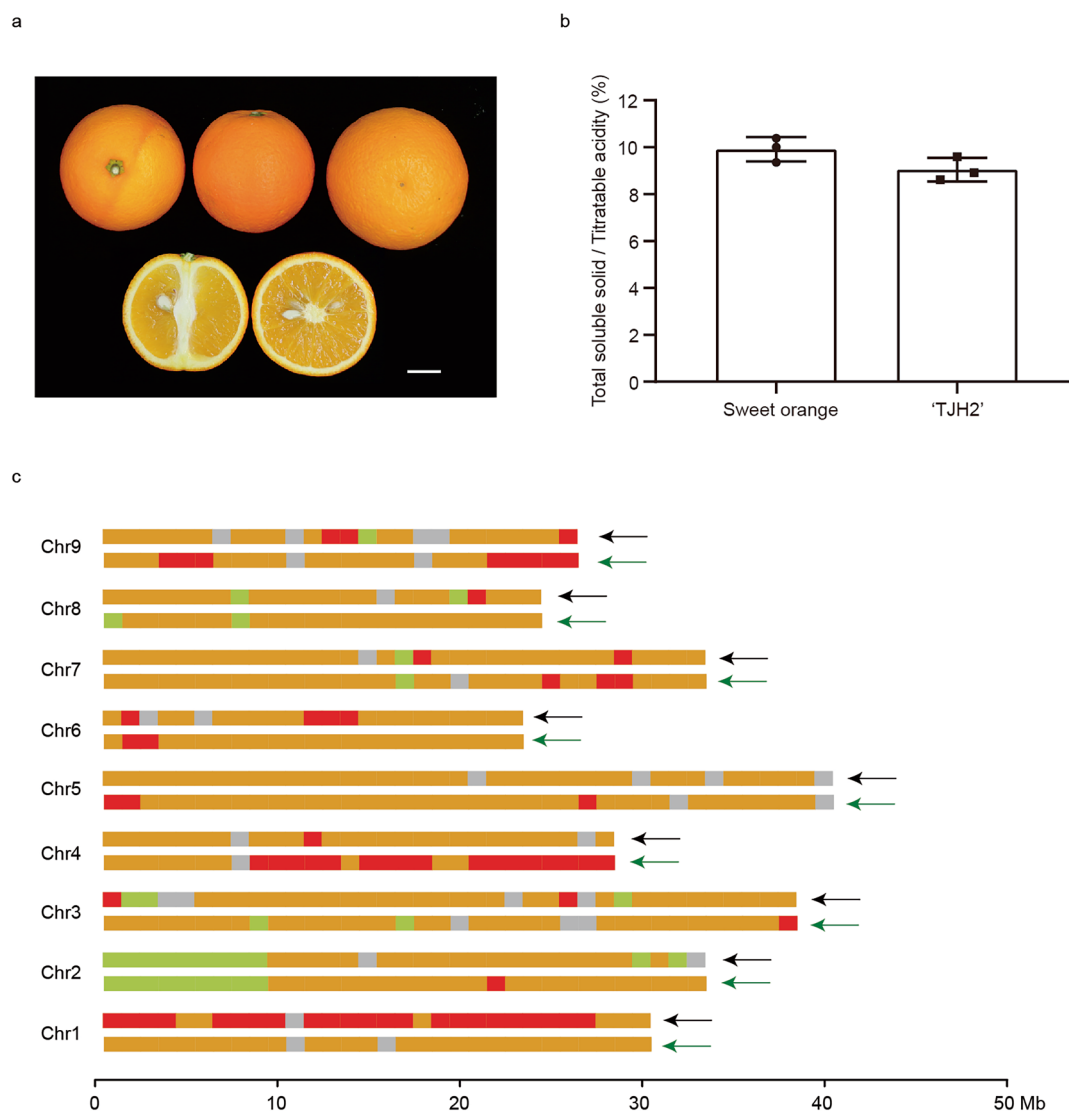


Extended Data Fig. 1 | Phenotypes of citrus accessions in the Nanling Mountains, Yuan River and Wuling Mountains (NYW) region in South China. (a) Single fruit weight. (b) Fruit shape index. (c) Number of seeds. (d) Number of embryos in individual seeds. (e) Titratable acids and (f) total soluble solids in

mandarins, pummelos, sour oranges and sweet oranges from the NYW region. Md, mandarin; Pm, pummelo; SSO, sour orange; SWO, sweet orange. Each point in the figure represents one accession, and the value is the average of three biological replicates.

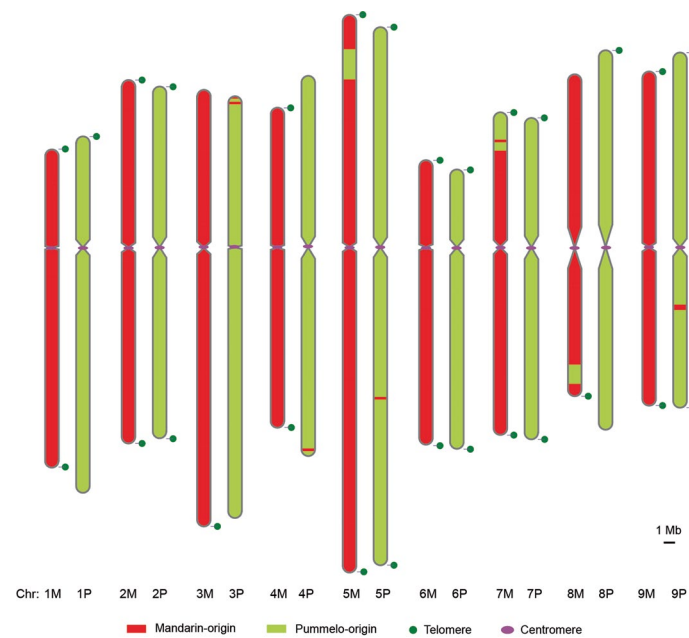


Extended Data Fig. 2 | Fruits and leaves of different sour orange accessions and a sweet orange. (a) Fruit shapes and (b) leaf shapes from sour oranges and a sweet orange provide evidence that sour oranges are similar to sweet orange. The numbers indicate different accessions of sour orange. Scale bar, 1 cm.



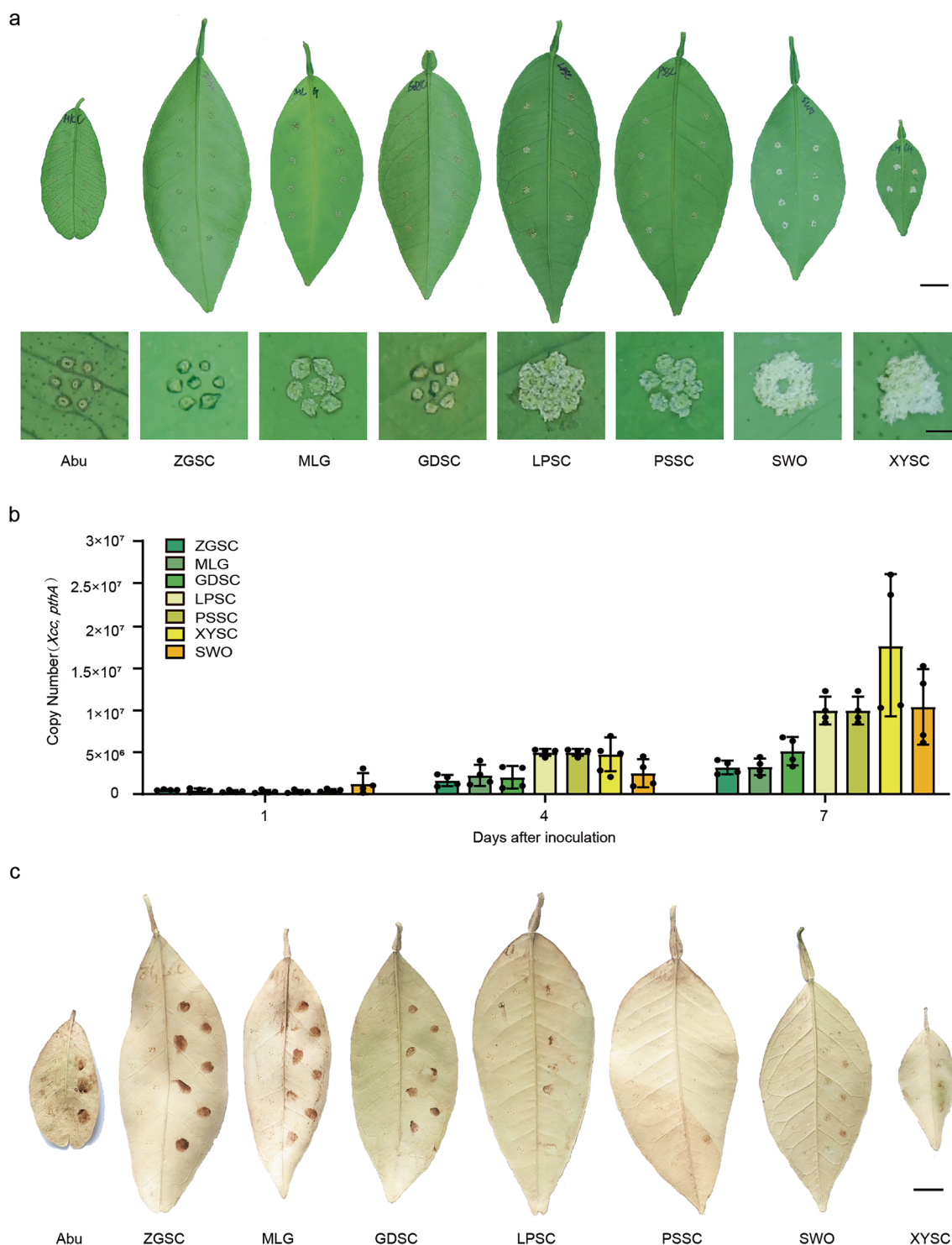
Extended Data Fig. 3 | Fruit phenotype, quality and comparative analysis of the genomic patterns between sweet orange and 'TJH2', a natural sweet orange hybrid found in southern China. (a) Fruit phenotype. Scale bar, 2 cm. **(b)** Ratio of total soluble solid (TSS) to titratable acidity (TA) for 'TJH2' (n=3) and sweet orange (n=3). Each box represents the median and inter quartile range (IQR). The

error bars represent standard error of the mean. **(c)** Genomic composition of sweet orange (black arrow) and TJH2 (green arrow). Red, homozygous segments with a mandarin origin; yellowish green, homozygous segments with a pummelo origin; orange, heterozygous segments with a mandarin/pummelo origin; gray, unknown region.



Extended Data Fig. 4 | Haplotype-resolved telomere-to-telomere genome assembly of sour orange ('ZGSC'). Diagram illustrating the contributions of the mandarin and pummelo genomes to the two haplotypes of sour orange ('ZGSC').

Red, mandarin origin; yellowish green, pummelo origin; green circles, telomeres; magenta ellipses, centromeres. Scale bar, 1 Mb. M represents the mandarin haplotypes, P represents the pummelo haplotypes.



Extended Data Fig. 5 | Phenotypes of different sour orange accessions after inoculation with *Xanthomonas citri* subsp. *citri* (*Xcc*). (a) Canker lesions on the leaves of different sour orange accessions, *Atalantia buxifolia* and sweet orange. The photographs were acquired 7 days after inoculation with *Xcc*. Whole leaves are shown (top). Scale bar, 1 cm. Single inoculation sites are shown (bottom). Scale bar, 1 mm. (b) Pathogen titers in different sour orange accessions and sweet orange after inoculation with *Xcc*. The pathogen titers were measured at 1, 4 and 7 days after inoculation with 5 μ L of 10^8 CFU ml^{-1} *Xcc* as described previously (PLoS Genet. 2021 17(1):e1009316). Three independent samples were

used for the analysis. Each box represents the median and IQR. The error bars represent standard error of the mean. (c) DAB (3,3'-diaminobenzidine) staining of reactive oxygen species in infected leaves. The leaves were stained at 4 days after infection. For each leaf, the left side was inoculated with 5 μ L of pure water as a control. The right side was inoculated with 5 μ L *Xcc* bacteria (10^8 CFU ml^{-1}). Reddish-brown staining indicates the accumulation of H_2O_2 . Scale bar, 1 cm. Abu, *Atalantia buxifolia*; ZGSC, MLG, GDSC, LPSC, PSSC and XYSC are different sour orange accessions; SWO, sweet orange.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

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

n/a Confirmed

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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No data was used for data collection. Data were sequenced from PacBio, Nanopore and Illumina platform.

Data analysis Hifiasm (v0.16.1), canu (v2.2), Nextdenovo (v2.5.2), Necat (v.1.0), SMARTdenovo (v.1.0), Nextpolish (v.1.4.0), 3d-dna pipeline (v180922), TGS-Gapcloser (v1.1.1), wtdbg (v.2.3), Flye (v.2.8.3), racon (v.1.5.0), RepeatModeler2 (v1.0.11), RepeatMasker (v4.0.9), AUGUSTUS (v.3.3), Snap (version 2006-07-28), BUSCO (v 3.0.2), GlimmerHMM (v3.0.4), Hisat2 (v2.2.1), cufflinks (v2.2.1), Trinity (v.2.4.0), PASA (v2.4.1), EVM (v1.1.1), LASTZ (v1.02.00), MUMmer (v4.0.0), OrthoFinder2 (v2.2.7), agriGO (v2.0), MUSCLE (v.3.8.31), RAXML (v7.7.8), BWA (v0.7.17), SAMtools (v1.9), GATK (v4.1.1), Vcftools (v0.1.13), RAXML (v7.7.8), GCTA (v1.26.0), PLINK (v1.90), beagle (v r1399), Adobe Photoshop 2020 software, Adobe Illustrator 2020 software.
Custom scripts are available at <https://doi.org/10.6084/m9.figshare.26053999.v1>

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

We deposited all the raw sequencing data generated in this paper in the National Center for Biotechnology Information database under accession PRJNA1003792 and PRJNA1003676. Also, we deposited all the raw sequencing data used in this paper in the National Genomics Data Center (<https://ngdc.cncb.ac.cn/>) under accession PRJCA019821. We deposited the code used in this paper in Figshare DOI (<https://doi.org/10.6084/m9.figshare.25904626>). The assembled genomes and annotations are also available at <http://citrus.hzau.edu.cn/download.php>. The accession numbers for the whole-genome sequencing data are listed in Supplementary Table 1. The accession numbers for the RNA-seq data are listed in Supplementary Table 23.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	226 accessions were collected to have representative samples from wild, cultivar and hybrids citrus accessions. For the sequencing of wild and cultivated citrus accessions, we sequenced all the accessions that were collected and not previously sequenced. For the hybrids, in order to facilitate the analysis of sequencing data combined with phenotype, we randomly selected individual plants that had already got fruit for sequencing and analysis.
Data exclusions	No data were excluded in this study.
Replication	We confirm all experiments can be replicated. For RNA-seq analysis and metabolite profiling, at least three biological replicates were used. For measurement of the antimicrobial activity of metabolites, at least three biological replicates were used.
Randomization	For each accessions, the sampling process for genome DNA/RNA sequencing was randomly conducted. For the citrus canker inoculation experiment, the hybrids of the resistant phenotype to be identified was the experimental group, and the varieties that had been clearly reported as the susceptible phenotype were used as the control group. For the metabolite antibacterial experiment, the metabolites to be identified as resistant metabolites were the experimental group, and the control group was clearly reported to have recognized antibacterial effect.
Blinding	For genome sequencing, assembly and phylogenetic analysis, we sequenced all the accessions we collected, so there was no need for blinding. For citrus resources, the investigators were blinded to group allocation during data analysis.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Plants

Seed stocks	The seeds, leaves and budsticks of plant materials were collected from Hunan Province, Tibet region, citrus germplasm conservation nursery center of Huazhong Agricultural university of China during the years 2013 to 2023.
Novel plant genotypes	The novel genotypes were either from the hybridization or collection of natural germplasm. The hybrids between sour orange and mandarin were obtained by artificial cross pollination and embryo rescue.
Authentication	The genotypes were validated by whole genome sequencing data and liquid gene chip.