Received: 7 August 2025 | Accepted: 27 February 2026

DOI: 10.1002/tpg2.70232

The Plant Genome

ORIGINAL ARTICLE

Highly heterozygous *Citrus changshan-huyou* Y. B. Chang originated from ancient hybridization between mandarin and pummelo and displayed distinct tissue-specific allelic imbalance

Zhanghui Zeng[1] | **Yingjie Luo**[1] | **Haifei Hu**[2] | **Lan Lan**[3,4] | **Baojin Guo**[3,4] | **Ping Zhou**[5] | **Cong Tan**[5] | **Xiaoping Huang**[1] | **Tuo Qi**[6] | **Zhehao Chen**[1] | **Zhiming Yu**[1] | **Lilin Wang**[1] | **Taihe Xiang**[1] | **Chengdao Li**[3,4,7,8] | **Yong Jia**[3,4]

[1]College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, China

[2]Rice Research Institute, Guangdong Academy of Agricultural Science, Guangzhou, China

[3]Western Crop Genetic Alliance, Murdoch University, Murdoch, Western Australia, Australia

[4]State Agricultural Biotechnology Centre (SABC), College of Science, Health, Engineering and Education, Murdoch University, Murdoch, Western Australia, Australia

[5]BGI Research, Shenzhen, China

[6]Ecological Security and Protection Key Laboratory of Sichuan Province, Mianyang Normal University, Mianyang, China

[7]Department of Primary Industry and Regional Development, Government of Western Australia, South Perth, Western Australia, Australia

[8]College of Agriculture, Shandong Agricultural University, TaiAn, China

Correspondence
Zhanghui Zeng, College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, 311121, China.
Email: zhzeng@hznu.edu.cn

Chengdao Li and Yong Jia, Western Crop Genetic Alliance, Murdoch University, Murdoch, WA 6150, Australia. Email: c.li@murdoch.edu.au and y.jia@murdoch.edu.au

Assigned to Associate Editor Awais Khan.

Funding information
National Natural Science Foundation of China, Grant/Award Number: 32301872; Natural Science Foundation of Zhejiang Province, Grant/Award Number: LQ23C130003; Interdisciplinary Research Project of Hangzhou Normal University, Grant/Award Number: 2025JCXK01

Abstract

The genus Citrus is characterized by a reticulate evolutionary history with frequent hybridization, making it an intriguing subject for genome evolution investigation. *Citrus changshan-huyou* Y. B. Chang (Huyou) is a unique landrace first discovered in Zhejiang Province, China, with premium fruit quality. The evolutionary origin of Huyou has puzzled local botanists and growers. Here, we sequenced a 120-year-old "ancestral tree" of Huyou using PacBio long read and Hi-C sequencing and assembled two high-quality, haplotype-resolved genomes, HY1 and HY2. Huyou displayed a genome heterozygosity level at 3.07%, among the highest in published citrus genomes. Using a k-mer-based tracing approach, we explicitly resolved that HY1 genome contained 87.8% mandarin, 7.3% pummelo, and 0.2% citron origin, whereas HY2 had 85.0% pummelo, 2.9% mandarin, and 0.3% citron, implying a hybridization event between mandarin and pummelo. Phylogeny dating showed that HY1 (2.0 Mya) and HY2 (2.18 Mya) had diverged earlier than the split of *Citrus clementina* and *Citrus reticulata* and the split of *Citrus grandis* and *Citrus maxima*, respectively.

Abbreviations: AI, allelic imbalance; CDS, coding sequence; CNV, copy number variation; DAPI, diamidino-2-phenylindole; FISH, fluorescence in situ hybridization; KEGG, Kyoto encyclopedia of genes and genomes; SNP, single-nucleotide polymorphism; TE, transposable element.

We observed clear chromosomal recombination on chr8 and chr9 in HY1, which may have occurred after the ancestral hybridization. Further transcriptome analyses in six tissues revealed a strong allelic dominance of HY2 over HY1 in root tissue and moderately in stem, leaf, flower, and fruit. Kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses revealed that genes related to antioxidant biosynthesis and lipid metabolism were most significantly affected by allelic imbalance. This pioneering report of allelic imbalance in citrus species support Huyou as an interesting model to investigate genome evolution following distant hybridization.

Plain Language Summary

Citrus changshan-huyou (Huyou) is a unique landrace in Zhejiang, China, with an unknown evolutionary origin. This study sequenced and assembled two complete haplotype genomes from a 120-year-old Huyou tree, revealing it has one of the most genetically diverse citrus genomes known (3.07% heterozygosity). Huyou originated from an ancient cross between mandarin and pummelo. The two haplotypes diverged millions of years ago and show clear evidence of recombination. Gene expression data from six tissues revealed that one haplotype's alleles are more active than the other in specific organs, especially roots. Genes linked to antioxidants and lipid metabolism were most affected by this imbalance, suggesting potential functional consequences of hybrid origin and allele dominance in Huyou.

1 | INTRODUCTION

The genus Citrus belongs to the *Aurantioideae* subfamily of Rutaceae and contains a diverse array of species such as sweet orange (*Citrus sinensis*), mandarin (*Citrus reticulata*), pummelo (*Citrus grandis*), citron (*Citrus medica*), and lemon (*C. limon*). Their diverse flavors, nutritional benefits, and industrial applications have made them beloved fruits worldwide. Citrus ranks as the third largest fruit in terms of production volume (https://www.statista.com/) and holds substantial agricultural, economic, and cultural value globally (Gmitter et al., 2012). Beyond their economic and cultural importance, citrus fruits present a fascinating subject for genomic and evolutionary studies due to their intricate origins and diversification. The genus Citrus is characterized by a reticulate evolutionary history, marked by frequent hybridization and polyploidization events (G. A. Wu et al., 2018). This complex evolutionary past has led to various hybrid accessions with relatively high heterozygosity levels at 1.5%–2.4%, compared to 0.1%–0.6% for intraspecies (G. A. Wu et al., 2018). This makes hybrid citrus an intriguing subject for genome evolution investigation.

Citrus changshan-huyou Y. B. Chang (Huyou) is a unique *Citrus* species that originated in the Changshan County of the Zhejiang Province in China (Y. Li et al., 2019; Yin-bin, 1991). The earliest cultivation of Huyou was recorded over 600 years ago, well before its massive planting and production over 100 years as a prominent landrace. Its fruit has a golden color and contains abundant components, such as amino acids, vitamins, naringin, limonin, citrate, and other nutrients (Guo et al., 2018; Sheng et al., 2017; J. Zhang et al., 2012). These components not only invest the fruit's unique aroma and taste that combines sweet, sour, and a bit of bitterness but also health benefits, including lowering blood sugar and hepatoprotective and anti-inflammatory effects (Guo et al., 2018; Jiang et al., 2019; J. Zhang et al., 2012). These benefits have led to the widespread consumption of Huyou in Zhejiang and neighboring province. Although the economic and scientific importance of Huyou is gaining increasing attention, its evolutionary history and hybrid origin remain subjects of debate among researchers and local growers. Several pioneering studies have attempted to use traditional molecular markers to deduce the origin of Huyou but only yielded limited genetic insights due to data constraint (L. Chen et al., 2002; S. Chen et al., 2006; Y. Li et al., 2019). A recent scientific dataset reported the haplotype-resolved genome assembly of a Huyou accession and transcriptome sequencing in six tissues (Miao et al., 2024). Despite the high quality Huyou genomic data, the study focused on dataset and methodological descriptions. The genomic origin, domestication, and evolution analyses for this important natural hybrid citrus landrace remain to be characterized.

Over the past decade, numerous citrus species have been sequenced, generating genomic data that has significantly advanced our knowledge of citrus biology and genome evolution. Key milestones include the sequencing and assembly of reference genomes for important citrus species such as sweet orange (*C. sinensis*) (Q. Xu et al., 2013), clementine mandarin (*Citrus clementina*) (G. A. Wu et al., 2014), and pummelo (*Citrus maxima*) (G. A. Wu et al., 2014). These were followed by whole genome resequencing for large citrus germplasm collections that encompass diverse citrus, including wild, landrace, and cultivar lines (X. Wang et al., 2017; G. A. Wu et al., 2018). All these studies have highlighted a complex history of admixture and hybridization during citrus domestication. The sequencing projects indicate that modern citrus varieties are often the result of crosses between a few ancestral species such as mandarin, pummelo, and citron. Beyond the genomic insights into genome origin and evolution, citrus genome sequencing studies have also led to the identification of genes associated with important agronomic traits, including fruit quality, disease resistance, and stress tolerance. For example, studies have identified genes involved in flavor synthesis in lemon (Bao et al., 2023), disease resistance in trifoliate orange (Peng et al., 2020), and citric acid accumulation in oranges (Huang et al., 2023). Recently, citrus pangenome data for various citrus and citrus-related species have been made available (Huang et al., 2023; Liu et al., 2022). These genomic resources allow researchers to delve deeper into the genetic basis of key traits, paving the way for more targeted and efficient breeding strategies (Liu et al., 2022).

A haplotype-phased genome refers to a genome sequence where the two sets of chromosomes inherited from each parent are separately distinguished and assembled. The assembly of highly accurate haplotype-phased genomes is achieved through advanced sequencing technologies such as long-read sequencing (e.g., PacBio HiFi, Nanopore) and Hi-C sequencing (chromosome conformation capture). It allows for the identification of allele-specific variations, structural variations, and gene expression patterns that would be obscured in a standard genome assembly. This is particularly crucial for genomic study on heterozygous organisms such as Citrus species, which are marked with frequent hybridization events. Recently, several haplotype-resolved genome assemblies have been generated, including *C. limon* (Mario et al., 2021), *C. australis* (Nakandala et al., 2023), *C. changshanensis* (Miao et al., 2024), *Papeda* (F. Wang et al., 2025), and a haploid genome assembly for pummelo (X. Wang et al., 2017), offering a more detailed view of their genetic makeup.

In this study, we performed genome sequencing and assembly of a 120-year-old "ancestor tree" of Huyou currently still growing in Changshan, Zhejiang Province, where Huyou was first discovered. We integrate high-coverage Illumina short-read and PacBio long-read sequencing with Hi-C sequencing technology for the assembly of two haplotype-resolved chromosome-scale genomes. This is followed by in-depth genome tracing, structural variation, and comparative genomic analysis with previously established citrus ancestor species to fully characterize the unique genomic features of Huyou and to resolve its evolutionary origin. Furthermore, assisted by previously published transcriptome data and the haplotype-resolved genome assemblies in this study, we comprehensively characterized the allelic imbalance (AI) across different tissues. Gene family expansion/contraction and transcriptome profiles for key genes from important metabolic pathways related to citrus fruit quality were also analyzed. We highlight the exceptional genome heterozygosity (3.07%) of Huyou compared to other known hybrid *Citrus* genomes (1.5%–2.4%) (G. A. Wu et al., 2018) and the unique tissue-specific AI in the root tissue. Our study supports Huyou as an interesting model to investigate genome evolution following distant hybridization.

Core Ideas

- De novo assembled two haplotype-phased genomes for a 120-year-old ancestral tree of *Citrus changshan-huyou* (Huyou).
- Observed exceptional genome heterozygosity (3.07%) in Huyou, among the highest across published citrus genomes.
- Explicitly resolved the origin of Huyou as a hybridization of mandarin and pummelo using k-mer-based tracing approach.
- Revealed a strong tissue-specific allelic imbalance pattern from haplotype-based transcriptome analysis in six tissues.

2 | MATERIALS AND METHODS

2.1 | Plant materials and nucleic acid extraction

Fresh leaves, shoots, and fruits were collected from a 120-year-old *C. changshan-huyou* Y. B. Chang tree in Chengtan Village, Qingshi Town, Changshan County (118°30′E, 28°51′N), Zhejiang Province, China, in the spring of 2022 (Figure 1A). These tissues were quickly frozen in liquid nitrogen and then stored at −80°C before DNA and RNA extraction. Total genomic DNA was extracted from the leaf tissue, and its quality was assessed by the Qubit 3.0 Fluorometer (Thermo Fisher), while the total RNA from leaves, shoots, and fruits was extracted using TRIzol Reagent (Invitrogen). The quantity and quality of the RNA were examined using NanoDrop One spectrophotometer (Thermo Fisher)

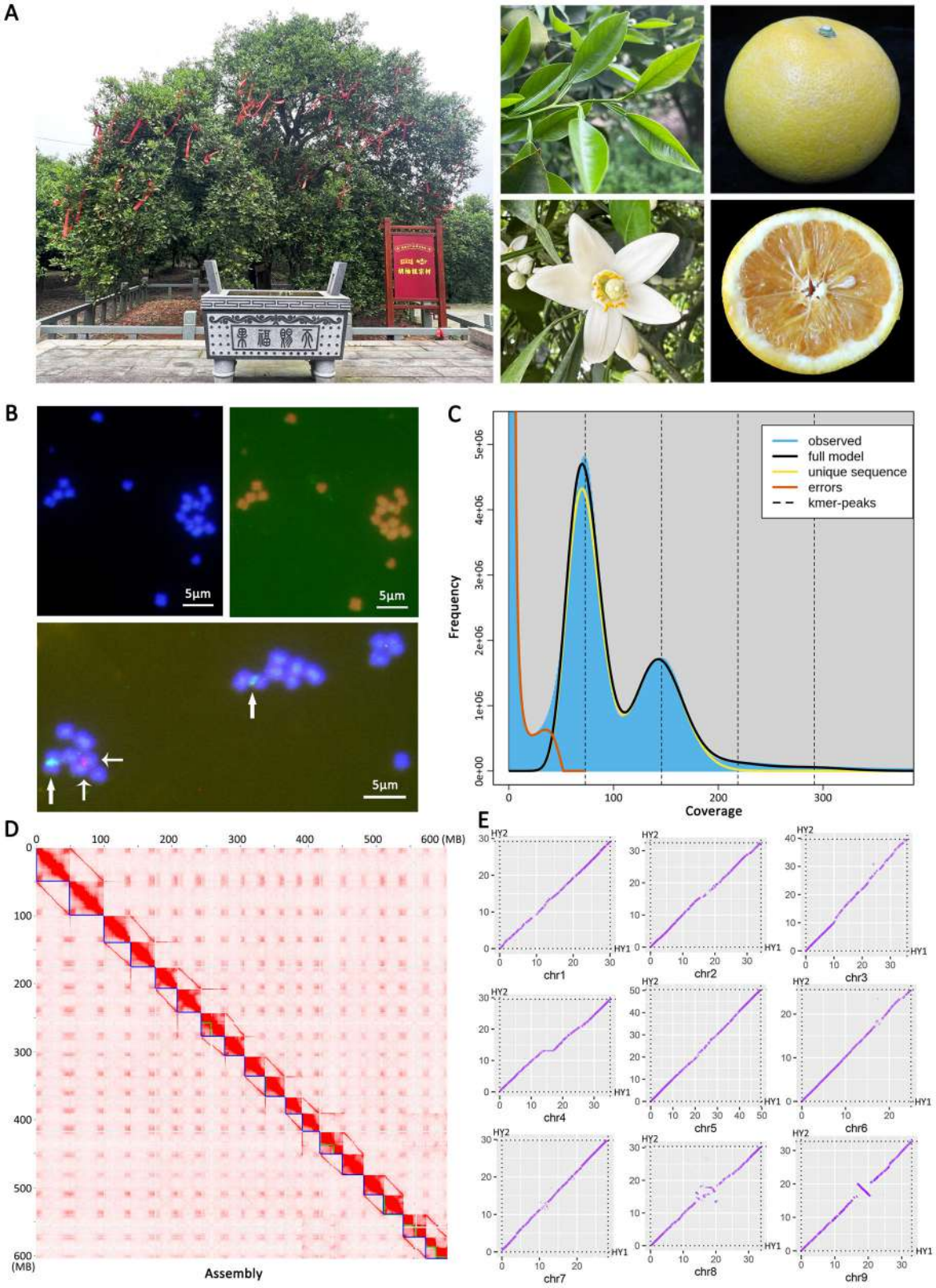


FIGURE 1 Huyou tree characteristics, karyotype analysis, and genome survey. (A) The 120-year-old "ancestral" Huyou tree sequenced in this study (whole tree, stem, leaves, flowers, and fruits). (B) Karyotype analysis on chromosomes of Huyou root tips. Karyotype pictures of Huyou using diamidino-2-phenylindole (DAPI) staining (top left), fluorescence in situ hybridization (FISH) analysis with a telomere specific repeat probe (top right), and 5S rDNA (red) and 18S rDNA (green) repeat sequence probes (bottom). Bar = 5 μm. (C) Genome size and heterozygosity estimation

(Continues)

FIGURE 1 (Continued)
based on K-mer frequency analysis ($K = 21$). The first peak reveals heterozygous k-mers, while the second peak indicates the homozygous k-mers. (D) Hi-C interaction heatmap for the assembled HY1 and HY2 genomes. Blue lines indicate chromosomal boundary. The 18 chromosomes were sorted based on length in decreasing order. The bottom-right cluster represented unanchored contigs. (E) Dot plots displaying the synteny between HY1 and HY2 for each chromosome.

and an Agilent 2100 Bioanalyzer (Agilent Technologies), respectively.

2.2 | Genome karyotype identification

Karyotype analysis was conducted with fluorescence in situ hybridization (FISH) method as previously described (X. Wang et al., 2017; H. Zhang et al., 2022) and was performed by OMIX Technologies Corporation (Chengdu). Briefly, excised young root tips (2–3 cm) of Huyou via tissue culture (Y. Li et al., 2017) were treated with nitrous oxide gas for 2 h under 1 MPa, then fixed in ice-cold 90% acetic acid for 10 min. After washing in water, the root tips were diced and digested in 1% pectolyase Y23 and 2% cellulase Onozuka R-10 (Yakult Pharmaceutical) for 1 h at 37°C. After digestion, the root sections were washed in 70% ethanol three times briefly, then broken by using a needle and vortexed at maximum speed in 100% ethanol for 30 s at room temperature to separate cells from one another. The cells were collected at the bottom of the tube by centrifugation (4000 r/min) and resuspended in glacial acetic acid. The cell suspension was dropped onto glass slides in a box lined with wet paper. A telomere-specific repeat probe 5′-$(TTTAGGG)_6$-3′) was used to detect the number of intact chromosomes, and 5s rDNA and 18s rDNA repeat sequence probes were used to identify multiple copies of chromosomes. The fluorescence staining of the chromosomes was performed using 4′,6-diamidino-2-phenylindole (DAPI). After DAPI staining, the dispersed metaphase chromosome cells were counted under a fluorescence microscope (Zeiss LSM880).

2.3 | Genome size and heterozygosity estimation

For genome profiling, the DNA library was constructed with insert sizes of ~300 bp using Nextera DNA Flex Library Prep Kit (Illumina) and subsequently sequenced on the Illumina NovaSeq 6000 platform (Illumina), which generated 150 bp paired-end reads. The software FastQC (v 0.20.1) (Andrews, 2012) was used to filter the original reads and discard the low-quality reads. A total of 57.05 Gb clean data were generated by Illumina sequencing and were used to perform the genome survey for Huyou. Reference-free estimation of genome size and heterozygosity was performed using a k-mer frequency approach implemented in GenomeScope 2.0 (Ranallo-Benavidez et al., 2020).

2.4 | PacBio, Hi-C, and RNA sequencing

To generate long-read sequencing reads for Huyou, DNA libraries were prepared using the SMRTbell Express Template Preparation kit 1.0 following the PacBio 20-kb protocol (https://www.pacb.com/) and sequenced on the PacBio Sequel II platform (Pacific Biosciences of California, Inc.). After the data quality control, by removing sequencing adaptors and filtering low-quality short read, and correcting errors via SMRT Link (v8.0), 26.17 Gb of HiFi reads were obtained (Table 1).

The generation of a Hi-C library was performed as described previously (Strijk et al., 2021). Briefly, the genomic DNA was randomly broken into 300–700 bp fragments via DNA cross-linking, restriction enzyme digestion, cohesive end repair, DNA cyclization, and DNA purification. Then the sample was sequenced on an Illumina NovaSeq 6000 platform (PE 150). After the removal of sequencing adaptors and low-quality reads using fastp software (v 0.21.0) (S. Chen et al., 2018), 31.77 Gb Hi-C clean reads were obtained.

For gene annotation, a pooled RNA-seq sample was generated by mixing equal amounts of RNA extracted from leaves, shoots, and fruits. For RNA sequencing, the cDNA library was constructed in this study following manufacturer instructions for the Illumina NovaSeq 6000 platform. Approximately 10 Gb of sequencing data were obtained in total.

2.5 | De novo genome assembly and annotation

Haplotype-resolved de novo genome assembling was performed using hifiasm software (Cheng et al., 2021) with HiFi reads and Hi-C phasing. The default k-mer size was changed to 19 to obtain correct estimation of heterozygous site coverage. The resulting haplotype-resolved genome contigs were further purged for duplicates using Purge_Dups (Guan et al., 2020) by aligning the HiFi reads back to the assembled genomes using minimap2 tool (Li, 2018). The purged haplotype genomes were scaffolded into chromosome groups using the pipeline implemented at https://github.com/WarrenLab/hic-scaffolding-nf, which employs Chromap (H. Zhang et al., 2021) and YaHS (Zhou et al., 2023) for sequence alignment and Hi-C scaffolding, respectively. Genome statistics were calculated using the assembly-stats tool (https://github.com/sanger-pathogens/assembly-stats). Then, the juicebox software (v1.11.08) (Durand et al., 2016) was used to draw a Hi-C heatmap based on the interaction intensity and relative position between contigs. Following manual correction, the

TABLE 1 The statistics for genome assembly and gene annotation of Huyou.

Sequencing platform	Illumina, PacBio, Hi-C	
Assembly	HY1	HY2
Chromosome number	9	9
Total genome size	326.65 Mb	322.79 Mb
Size anchored to chromosome	299.73 Mb/91.8%	304.81 Mb/94.4%
Number of contigs	324	235
N50 contigs	21.15 Mb	18.77 Mb
Largest contig	46.45 Mb	34.26 Mb
Complete BUSCOs	98.1%	97.7%
GC content	56.60%	55.02%
Genome annotation		
Number of protein-coding genes	42,091	42,050
Mean mRNA length	2570 bp	2554 bp
Mean CDS length	1245 bp	1245 bp
Mean exons per gene	5.75	5.77
Mean exon length	216.40	215.87
Number of TEs		
DNA	14,097	15,545
LINE	6762	7428
SINE	248	269
LTR	50,487	56,301
Total size of TEs	181.5 Mb/55.56%	179.4Mb/55.58%
Number of non-coding RNA		
miRNA	161	150
tRNA	1003	790
rRNA	1590	2686
snoRNA	1393	1463

Note: HY1 and HY2 refers the two haplotype genomes.
Abbreviation: CDS, coding sequence.

uniquely mapped and valid interaction paired-end reads were used to build the pseudochromosome sequences. The pseudochromosomes were sorted and oriented using *C. sinensis* as the reference by RagTag (Alonge et al., 2022).

The quality of the assembled genomes was evaluated by benchmarking universal single-copy orthologs against the eukaryota_odb10 database (BUSCO v.4.1.4) (Simão et al., 2015). Repeat sequences of the haplotype genomes were searched using an automated pipeline implemented by EarlGrey (Baril et al., 2024). The soft-masked genomes were used for gene model prediction using GeMoMa v1.8 (Keilwagen et al., 2019) based on homology and transcriptome evidence. For homology references, the gene models from three ancestor genomes (*C. sinensis*, *Citrus ichangensis*, and *C. medica*) were used. For transcriptome evidence, the RNA sequencing reads were mapped to the masked genomes using STAR (Dobin et al., 2013), and the resulting alignment files were used as input for GeMoMa. For noncoding RNA annotation, the transfer RNAs (tRNAs) were annotated by tRNAscan-SE (v1.23) (Chan et al., 2021). Other non-coding ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), and microRNAs (miRNAs) were identified using the Infernal toolkits (Nawrocki et al., 2009) based on information from the Rfam database. Gene function annotations were performed using Interproscan for GO annotation and kofam_scan for Kyoto encyclopedia of genes and genomes (KEGG) annotation. GO and KEGG enrichment analyses for Huyou genome-specific genes were performed using TBtools (C. Chen et al., 2020). The program circos (v0.69) (Krzywinski et al., 2009) was used to visualize genomic characteristic information, including gene density, GC density, transposable element (TE) density, non-coding RNA density, and synteny block.

2.6 | Variants, synteny, and genome origin analyses

The genomic variations and synteny across the haplotype genomes of Huyou and their ancestor genomes, *C. reticulata* (mandarin, NCBI accession ID: ASM325862v1) (L. Wang et al., 2018) and *C. maxima* (pummelo, NCBI accession ID: ASM2964120v1) (Zheng et al., 2023), were analyzed using SyRI software (Goel et al., 2019). Genome alignment was performed using minimap2 (Li, 2018). The ancestral origin of Huyou haplotype genomes was determined using the k-part tool at https://github.com/sc-zhang/kPart. Variations between HY1 and HY2 were counted per 1 Mb window based on the output from SyRI. For ancestor tracing, genomic data for three ancestor genomes (*C. reticulata*, *C. maxima*, and *C. medica*) were downloaded and used as references. The ancestor of Huyou genomes were assigned for each 100 kb window using kPart with −k 21 parameter (https://github.com/sc-zhang/kPart). The percentages of each ancestor origin were calculated. The haplotype origin for each chromosome was also phased and classified using SubPhaser program with a similar k-mer-based method (Jia et al., 2022).

2.7 | Ortholog inferences, phylogeny development, and gene family analyses

Protein sequences of Huyou genomes and 19 other species, including *Aquilegia coerulea*, *Mangifera indica*, *Vitis vinifera*, *Malus domestica* Golden, *Solanum lycopersicum*, *C. sinensis*, *C. reticulata*, *C. maxima*, *C. medica*, *C. clementina*, *Poncirus trifoliata*, *Fortunella hindsii*, *C. ichangensis*, *Atalantia buxfoliata*, and *Litchi*

chinensis Sonn (downloaded from Citrus Genome Database at https://www.citrusgenomedb.org/) (Liu et al., 2022), were used as input for OrthoFinder v2.3.12 (https://github.com/davidemms/OrthoFinder) (Emms & Kelly, 2019) for ortholog inferences. For alternative transcripts, only the sequences for the longest transcripts were used. The obtained species tree was used as input for downstream Bayesian estimation of divergence time using mcmctree from PAML v4.9 toolkits (Yang, 2007). The coding sequences (CDS) for the identified single copy genes from OrthoFinder were extracted. Codon-based sequence alignments were performed using a combination of MUSCLE (Edgar, 2004) and SeqMagick (v0.8.6) backtrans-align functions. The concatenated CDS alignment was trimmed using Gblocks (v0.91b) (Castresana, 2000) before input into mcmctree (Yang & Rannala, 2006). Five divergence time points derived from https://timetree.org/ were used to calibrate the species tree. The final obtained species tree from MCMCTree and orthologous gene count results from OrthoFinder were input into CAFE (De Bie et al., 2006) for gene family expansion/contraction analyses. The gene family changes along the species tree were visualized using tvBOT (Xie et al., 2023) online tool at https://www.chiplot.online/tvbot.html. For gene copy number analysis, the gene counts for corresponding hierarchical orthogroups from OrthoFinder outputs were used.

2.8 | Genome divergence time analyses

The genome divergence time was assessed by *Ks* profile analyses using WGDI program (H. Sun et al., 2022). Each citrus genome was compared to itself following the WGDI instruction. Protein sequences for the primary transcripts of each species were used for BLASTP using diamond program (Buchfink et al., 2021). Syntenic blocks were identified using the build-in function in WGDI. Codon-based sequence alignments were performed using muscle (Edgar, 2004) and pal2nal (Suyama et al., 2006) programs. Ks values calculation for collinear genes using YN00 from PAML v4.9 program (Yang, 2007). Ks peaks at p-value ≤ 0.05 were identified and fitted using WGDI. The whole genome duplication event in *V. vinifera* (Jaillon et al., 2007) was used as a reference.

2.9 | Transcriptome and AI analyses

Transcriptome RNA sequencing data in six Huyou tissues (root: SRR28430799, stem: SRR28430800, leaf: SRR28430798, flower: SRR28430801, unripe fruit: SRR28430803, ripe fruit: SRR28430802) were downloaded from a previous study (Miao et al., 2024). Raw reads were trimmed using fastp program (S. Chen et al., 2018) with –qualified_quality_phred 20 and –length_required 30 parameters. The resulting clean reads were mapped to merged haplotype genomes using STAR program (Dobin et al., 2013) and then were quantified using RSEM (Li & Dewey, 2011) based on the annotated gene models in this study.

To perform AI analyses, first the syntenic gene pairs between HY1 and HY2 were identified using MCScanX program (Y. Wang et al., 2012). Then, for the identified syntenic gene pairs, the $\log_2$ values of the ratios of HY1 gene expression to HY2 gene expression were calculated. Allelic genes with fold changes ≥ 2 ($\log_2$ ratio value ≥ 1 or ≤ -1) were identified. GO and KEGG enrichment analyses for Huyou genome-specific genes were performed using TBtools against customized Huyou genetic background (C. Chen et al., 2020).

3 | RESULTS

3.1 | Genome assembly and annotation of the 120-year-old "ancestral tree" of Citrus changshan-huyou Y. B. Chang

In this study, a century-old Huyou tree (120-year-old as of 2025) currently growing and protected in Chentang Village, Changshan County, after which Huyou was named by Mr. Yunbing Chang in 1991, was selected for genome sequencing (Figure 1A). This century-old tree was believed by many local Huyou growers and botanists to be the ancestor of most Huyou trees currently cultivated across Zhejiang Province and was thus termed the "ancestral tree." Notably, this "ancestral tree" of Huyou, despite its old age, remains highly vigorous and productive with premium fruits to date (Figure 1A, photographed in 2024).

To determine the chromosome number and ploidy level of Huyou, high-resolution metaphase chromosome preparation assay was performed using root tip cells from tissue-cultured Huyou seedlings. Results showed that the genome of Huyou is composed of 18 chromosomes ($2n = 2x = 18$, Figure 1B), which was confirmed by FISH with 5S rDNA and 18S rDNA repeat sequence probes, and consistent with those reported in other citrus species (X. Wang et al., 2017). *K*-mer frequency distribution analysis ($K = 21$) of Illumina short reads (57.05 Gb, ~189 × coverage) showed a distinct bimodal profile with 3.07% heterozygosity (Figure 1C), compared to 1.5%–2.4% for other hybrid citrus genomes and 0.1%–0.6% for intraspecies accessions (G. A. Wu et al., 2018).

Due to its exceptionally high heterozygosity level, long-read PacBio HiFi sequencing (26.17 Gb, ~86×) and the Hi-C sequencing (31.77 Gb, 105×) were performed to obtain high-

quality chromosome-scale reference assembly for Huyou. The de novo assembly using Hifiasm with HiFi long reads and Hi-C reads as inputs produced two haplotype genomes, HY1 (324 contigs, total size at 326.65 Mb with contig N50 of 21.15 Mb) and HY2 (235 contigs, total size at 322.79 Mb with contig N50 of 18.77 Mb) (Table 1). GC contents for HY1 (56.6%) and HY2 (55.02%) were comparable with each other. The largest contigs for HY1 and HY2 were at 46.45 Mb and 34.26 Mb, respectively. The assembled contigs were anchored into nine pairs of chromosomes with Hi-C scaffolding. Distinct chromosomal groupings on HI-C interaction heatmap were observed in Figure 1D, indicating the high quality of genome anchoring. The chromosomes were further ordered and named using *C. sinensis* (v3.0 from the Citrus Genome Database http://citrus.hzau.edu.cn/download.php) as reference. Particularly, 91.8% and 94.4% of the contigs for HY1 and HY2, respectively, could be assigned to chromosomes. Using the eukaryota_odb10 database as reference, the Benchmarking Universal Single-Copy Orthologs (BUSCO) revealed completeness of 98.1% and 97.7% for HY1 and HY2, respectively. Using HY1 as the reference, genome alignment of HY1 and HY2 revealed overall well-conserved synteny between the two haplotype genomes, albeit some large structural variations were observed mainly in the central region of each chromosome, such as a large insertion in chr4 and large inversions in chr8 and chr9 (Figure 1E). Despite the overall synteny, the mosaic pattern and gaps in the genomic alignment plot indicated significant genetic variations between HY1 and HY2 (Figure 1E), consistent with the high heterozygosity in Huyou observed in this study.

Repetitive sequences were identified in the two haplotype genomes HY1 (181.5 Mb or 55.56%) and HY2 (179.4 Mb or 55.58%) (Table 1). In both genomes, LTR represented the largest TE family, followed by DNA, LINE, and Rolling Circle sequentially (Supporting Information File S0). Despite a relatively smaller genome size and smaller total TE content, HY2 displayed higher number of TEs in all TE classes than that in HY1 (Table 1). The repeat sequences were masked before gene model prediction using both homology search (*C. sinensis*, *C. ichangensis*, *C. medica*) and RNA sequencing data (10 Gb), which led to the identification of 42,091 and 42,050 protein-coding genes for HY1 and HY2, respectively, with comparable mean length for transcripts, CDS, exon, and average exon number per gene (Table 1). HY1 displayed relatively higher number of micro-RNA (161 miRNA) and tRNA (1003) but relatively lower number of small nuclear RNA (1393 snoRNA) than that in HY2 (150 miRNA, 790 tRNA, 1463 snoRNA). However, the number of ribosomal RNA in HY2 (2686 rRNA) was much higher than that in HY1 (1590 rRNA). The distribution of TEs, genes, and non-coding RNAs in HY1 and HY2 is displayed in Figure 2A. Overall, we observed higher gene density in the distal regions of each chromosome compared to the central region, whereas TEs tend to be enriched in the central regions. In summary, we obtained two haplotype-resolved Huyou assemblies with high quality and completeness.

3.2 | Ancestral tracing of Huyou haplotype genomes HY1 and HY2

Due to the high heterozygosity level and significant structural variations between the haplotype-resolved genomes, the ancestral origin of each block (100 kb) of the HY1 and HY2 genomes was determined using a k-mer-based tracing approach. The distribution of differential k-mers of homologous chromosomes indicated that HY1 and HY2 could be clearly separated into 2 distinct progenitors (Figure 2B), except for chr9 in HY1, which displayed a mixture of origin and was clustered with HY2. Using the citrus ancestor genomes mandarin (*C. reticulata*), pummelo (*C. maxima*), and citron (*C. medica*) as references, the genomic origin of HY1 and HY2 was traced and quantified. Results showed that 87.8% of HY1 genome originated from mandarin, followed by 7.3% from pummelo, 0.2% from citron, and 4.7% from an unknown source (could not be explicitly assigned) (Table 2; Supporting Information File S1). In contrast, 85.0% of HY2 genome was derived from pummelo, followed by 2.9% from mandarin, 0.3% from citron, and 11.9% from an unknown source (Table 2; Figure 2A). These results strongly suggested that Huyou may have originated from a hybridization between mandarin (HY1) and pummelo (HY2). At the chromosome level, chr1–chr7 of HY1 were predominantly from mandarin, whereas chr8 (11.9% pummelo) and chr9 (54.1% pummelo) of HY1 contained significant components from pummelo, which may have resulted from chromosome recombination. Mapping the genomic origin to the chromosomes indicated clear evidence of chromosome recombination in HY1 genome, whereby one distal region of chr8 and the central chr9 were replaced by pummelo genome (Figure 2A). For the HY2 genome, all chromosomes were predominantly from pummelo, with minor introgression of mandarin on chr1 (3.3% mandarin), chr3 (7.6%), chr5 (4.8%), chr8 (1.2%), and chr9 (3.6%) (Table 2). In contrast to HY1, which retained clear segmental recombination boundary, the genome introgressions in HY2 were generally mosaic along the chromosomes (Figure 2A), which may provide further clue on the evolution history of Huyou following the hybridization of its pummelo and mandarin parents. Furthermore, the LTR insertion age analysis also indicated distinct profiles for HY1 and HY2 (Figure 2C), providing additional support for two divergent progenitors. Specifically, most of the LRTs in HY1 displayed concentrated age with a single peak of insertion time, whereas the insertion time of LRTs in HY2 was relatively dispersed and contained an additional minor peak at a much older age (Figure 2C).

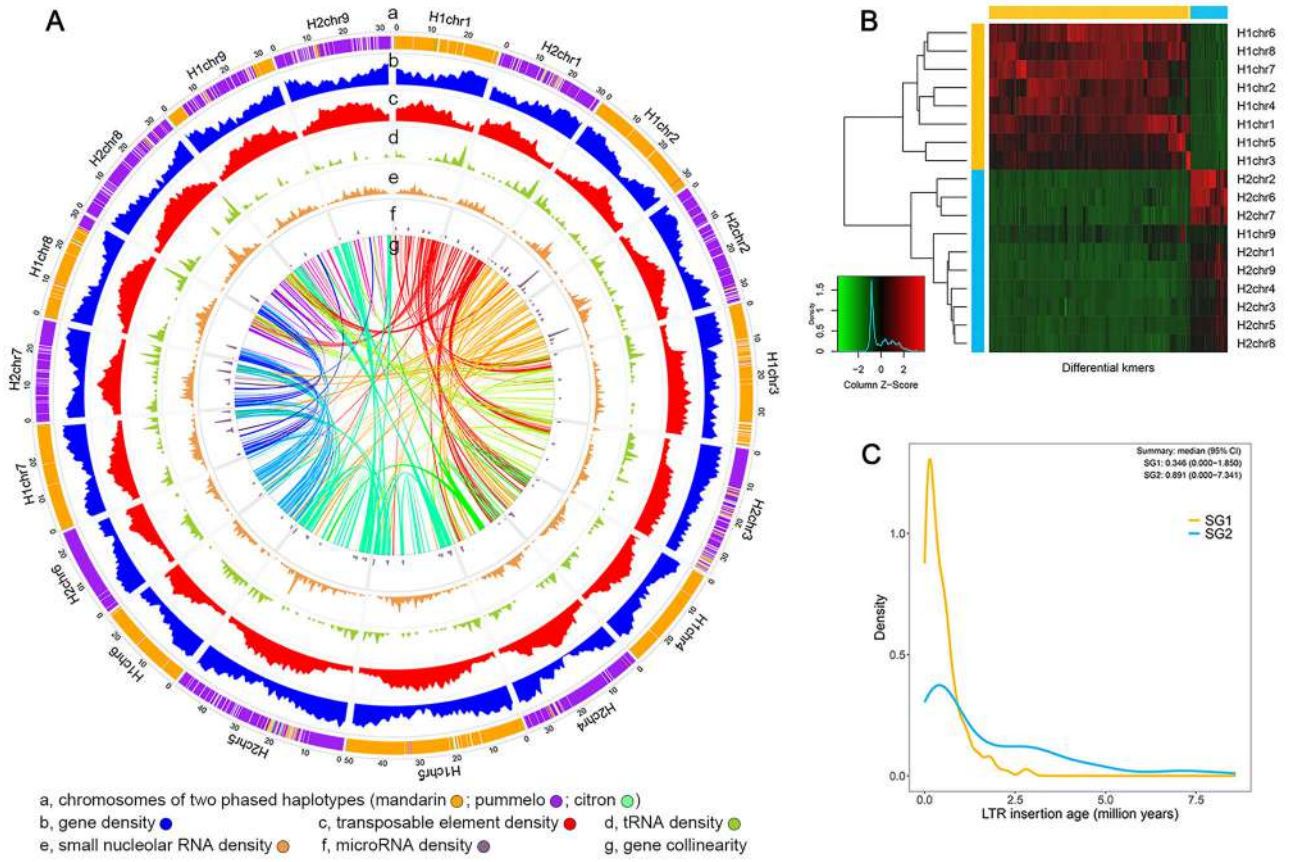


FIGURE 2 Ancestral tracing and comparative genome analysis of Huyou. (A) a, circles from outside to inside represent chromosomes of two phased haplotypes (mandarin, orange; pummelo, purple; citron, green; undetermined, white); b, gene density; c, transposable element density; d, tRNA density; e, small nucleolar RNA density; f, microRNA density; g, the curved lines in the center indicate patterns of gene collinearity. All distributions are drawn in a window size of 100 kb. (B) The histogram of differential k-mers among homoeologous chromosomes of HY1 and HY2. (C) Insertion time of subgenome-specific LTR-RTs (output from SubPhaser program, see method for details).

3.3 | Phylogeny dating and gene family analyses of HY1 and HY2

To deduce the divergence time of HY1 and HY2, single copy genes (2754 in total, Supporting Information File S2) for 13 previously published citrus genomes and six citrus-related genomes, together with HY1 and HY2, were identified and used for phylogeny development. Five calibration points (annotated in tree in Figure 3A: 3.49∼7.39, 3.92∼11.0, 12.8∼13.0, 7.57∼25.92, and 13.5∼30.0 Mya) were applied to estimate the divergence time. As shown in Figure 3A, the six citrus-related species were resolved as basal branches (group F), followed by the early diverging citrus species *P. trifoliata* (group G). The rest of the citrus species could be divided into 5 major groups (A-E highlighted in Figure 3A). Three of those corresponded to the well-recognized citrus ancestors: mandarin (A: *C. reticulata*), pumelo (C: *C. maxima*), and citron (E: *C. medica*). Consistent with the genome tracing results, HY1 and HY2 were grouped with mandarin and pumelo, respectively (Figure 3A). The divergence time of HY1 was estimated at 2.0 Mya, earlier than the split between *C. clementina* and *C. reticulata* but after the divergence of *C. sinensis*, while HY2 was estimated to diverge at 2.18 Mya, before the split between *C. grandis* and *C. maxima* and after the branching of *Citrus hongheensis* (Figure 3A). Based on the unique evolutionary position and ancestral compositions, we suggest that HY1 and HY2 represented 2 novel genomes that have not been reported before. Using the earliest diverging *P. trifoliata* as reference, ks peak fitting showed that HY1, HY2, and other Citrus species displayed a single ks peaks at similar position (Figure 3B), implying no additional large-scale whole genome duplication events. Gene family analyses based on the developed phylogeny revealed more than three-fold expansion (+345) than contraction (-102) in the branch leading to HY1. In contrast, slightly more expansion (+255) than contraction (-162) was observed in the branch leading to HY2 (Figure 3A). The percentage of single copy genes in HY1 and HY2 (6.5%) was relatively lower than that in other citrus species (8.1%–12.1%), probably caused by genome annotation bias (bar plot in Figure 3A). Relatively higher percentage of HY2-specific genes (0.6% or 261) was identified than that in HY1 (0.4% or 158). The numbers of genes

TABLE 2 The ancestral origin of each chromosome of haplotype-resolved Huyou.

			Percentage (%) of sequences from ancestor species				Origin tracing	
	Chromosome	Length (Mb)	Pummelo	Citron	Mandarin	Unknown	Closest	Ancestor
HY1	H1chr1	29.23	0.00	0.03	9.34	0.38	Clementina	Mandarin
	H1chr2	32.43	0.03	0.00	10.65	0.13	Clementina	Mandarin
	H1chr3	39.69	0.10	0.00	11.71	1.43	Clementina	Mandarin
	H1chr4	29.54	0.00	0.01	9.68	0.17	Sinensis	Mandarin
	H1chr5	50.35	0.10	0.03	16.06	0.60	Reticulata	Mandarin
	H1chr6	25.60	0.00	0.00	8.44	0.10	Sinensis	Mandarin
	H1chr7	29.80	0.00	0.00	9.88	0.07	Sinensis	Mandarin
	H1chr8	30.36	1.20	0.03	8.37	0.52	Sinensis	Mandarin
	H1chr9	32.74	5.91	0.03	3.65	1.33	Sinensis	Pummelo
	Total	299.73	7.34	0.15	87.78	4.73	–	–
HY2	H2chr1	30.06	8.25	0.03	0.33	1.25	Maxima	Pummelo
	H2chr2	34.26	9.89	0.00	0.13	1.21	Grandis	Pummelo
	H2chr3	35.93	8.46	0.07	0.90	2.36	Grandis	Pummelo
	H2chr4	35.11	10.63	0.00	0.23	0.66	Grandis	Pummelo
	H2chr5	49.49	13.12	0.13	0.79	2.20	Grandis	Pummelo
	H2chr6	24.99	8.10	0.00	0.00	0.10	Grandis	Pummelo
	H2chr7	28.27	8.49	0.00	0.00	0.79	Grandis	Pummelo
	H2chr8	33.75	9.55	0.03	0.13	1.36	Grandis	Pummelo
	H2chr9	32.95	8.45	0.03	0.39	1.94	Grandis	Pummelo
	Total	304.81	84.95	0.30	2.90	11.86	–	–

Note: Genome tracing was implemented using a k-mer-based approach by mapping sequencing data of three reference ancestral genomes mandarin (*Citrus reticulata*), pummelo (*Citrus maxima*), and citron (*Citrus medica*) to HY1 and HY2 using 100 kb window. See the method section and Supporting Information File S1 for details.

for five important metabolic pathways (map00900: terpenoid biosynthesis, map00906: carotenoid biosynthesis, map00940: phenylpropanoid biosynthesis, map00941: flavonoid biosynthesis, and map04626: plant-pathogen interaction) related to fruit quality were counted (heatmap in Figure 3A; Supporting Information File S3). Overall, we observed a general expansion of gene members in all target pathways for citrus species, compared to the basal citrus-related species and the early diverging species *P. trifoliata*, implying a potential human selection impact. Particularly, we noticed that group A and C, encompassing HY1 and HY2, contained the most expanded gene members compared to other citrus species, except for *C. maxima* Huazhouyou, which instead displayed relatively contracted gene families in these pathways. Further analysis is needed to investigate if this is caused by genome annotation bias or not. Noteworthy, a total of 158 and 261 species-specific genes were identified for HY1 and HY2, respectively (Supporting Information File S4). Enrichment analyses showed that gene ontology (GO) terms related to photosynthesis, signal transduction, and response to stimulus were significantly enriched in the HY1-specific genes (Figure 3C). In contrast, no significantly enriched GO term could be detected in the HY2-specific genes.

3.4 | Synteny and structural variations analysis of HY1 and HY2

The synteny and structural variations of HY1 and HY2 were inferred in reference to their ancestral genomes: mandarin and pummelo, respectively (Figure 4A). Overall, we observed much less large structural variations between mandarin and HY1 compared to that between HY1 and HY2, consistent with our ancestor tracing results that HY1 was derived from mandarin. For example, the large inversions identified between HY1 and HY2 on chromosomes chr6, chr7, and chr8 were not observed between mandarin and HY1, providing direct evidence that HY1 may have originated from a mandarin ancestor. Similar patterns could be found for other chromosomes and other structural variations as well. In contrast to the strong synteny conservation between HY1 and its ancestor genome, mandarin, we observed significant large structural variations between HY2 and the pummelo genome (*C. grandis* L. Osbeck.cv. Wanbaiyou) (Figure 4A). Noteworthy, these large structural variations, such as the inversions on chromosomes chr1, chr5, and chr9, were specific to the pummelo genome used in this study but not present between HY1 and HY2 (Figure 4A), implying that these large structural variations between HY2 and pummelo were mainly due to genome

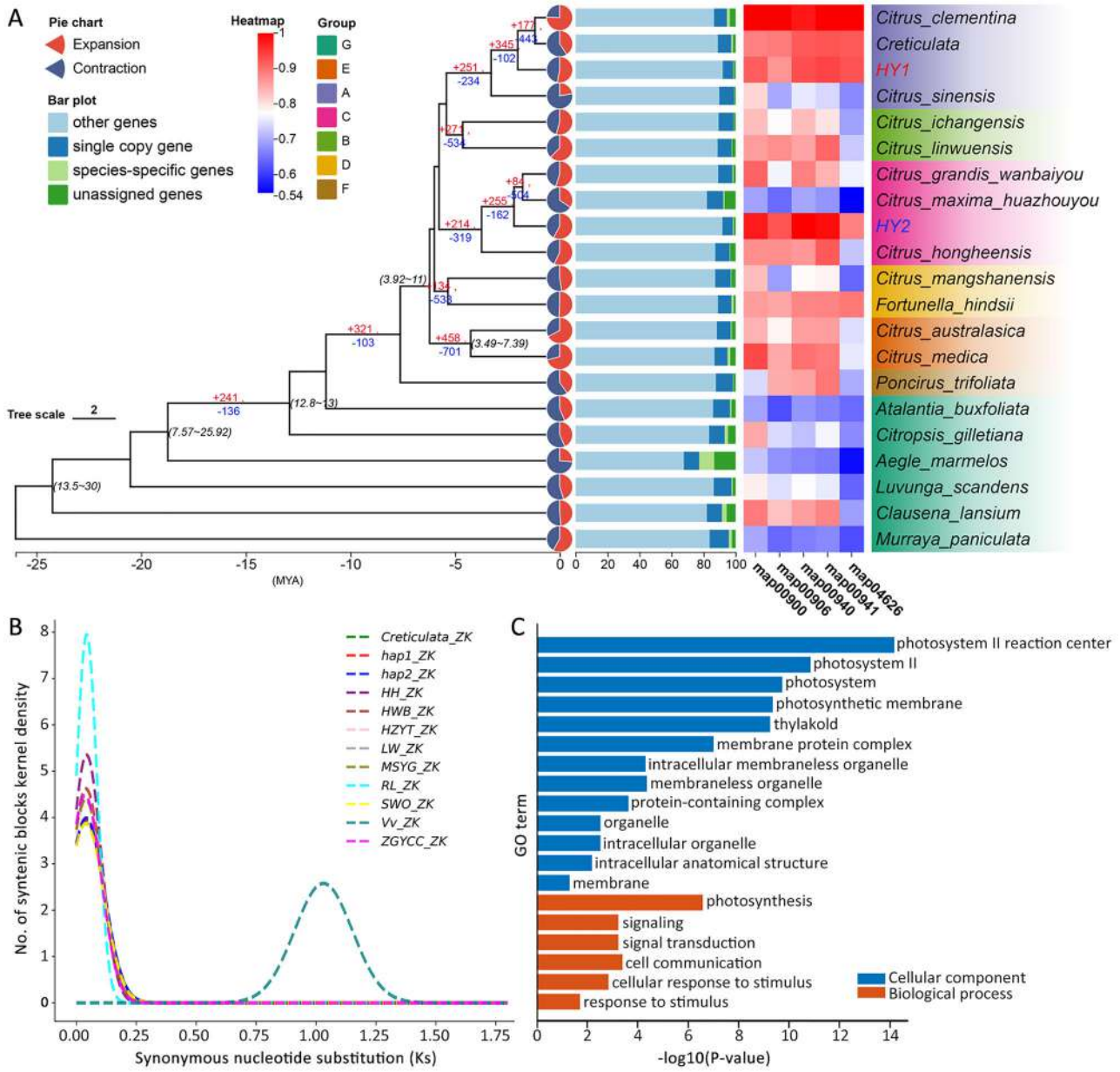


FIGURE 3 Phylogeny dating and gene family evolution analyses. (A) Displays the species phylogeny of HY1 and HY2 in references to other citrus species and citrus-related species. The tree topology was obtained using OrthoFinder, calibrated at five nodes (age annotated in tree) using mcmctree program based on single copy genes. Identified branching groups were highlighted correspondingly. Gene family expansion and contraction at key nodes (numbers) and each taxa (pie charts) were indicated. The percentage of single copy genes, species-specific genes, unassigned genes, and other genes were annotated as bar charts for each taxa. The heatmap represents gene counts for five key Kyoto encyclopedia of genes and genomes (KEGG) pathways: map00900- terpenoid biosynthesis, map00906- carotenoid biosynthesis, map00940- phenylpropanoid biosynthesis, map00941- flavonoid biosynthesis, and map04626- plant-pathogen interaction (Supporting Information File S2, S3). B. Ks distribution of collinear gene pairs of citrus species with the early diverging *P. trifoliata* (ZK) species as reference (HWB *Citrus grandis* wanbaiyou, HZYT *Citrus maxima* huazhouyou, SWO *Citrus sinensis*, MSYG *Citrus mangshanensis*, ZGYCC *Citrus ichangensis*, LW *Citrus linwuensis*, HH *Citrus hongheensis*, RL *Citrus medica*, and Vv *Vitis vinifera*). *Vitis vinifera* was used as an outgroup reference. C. Gene ontology enrichment analysis of HY1-specific genes. GO terms with corrected p-value < 0.05 were included as significant. No significant enrichment was identified for HY2-specific genes (Supporting Information File S4).

rearrangement in *C. grandis* L. Osbeck.cv. Wanbaiyou but not in the pummelo ancestor. Further analyses are needed to investigate if these large genome rearrangements are conserved in other pummelo genomes or not.

The genomic variations between HY1 and HY2 were examined specifically to investigate the allelic diversity in Huyou. We identified 1,996,631 single-nucleotide polymorphisms, 150,487 insertions (1.66 Mb), and 139,519 deletions (1.57

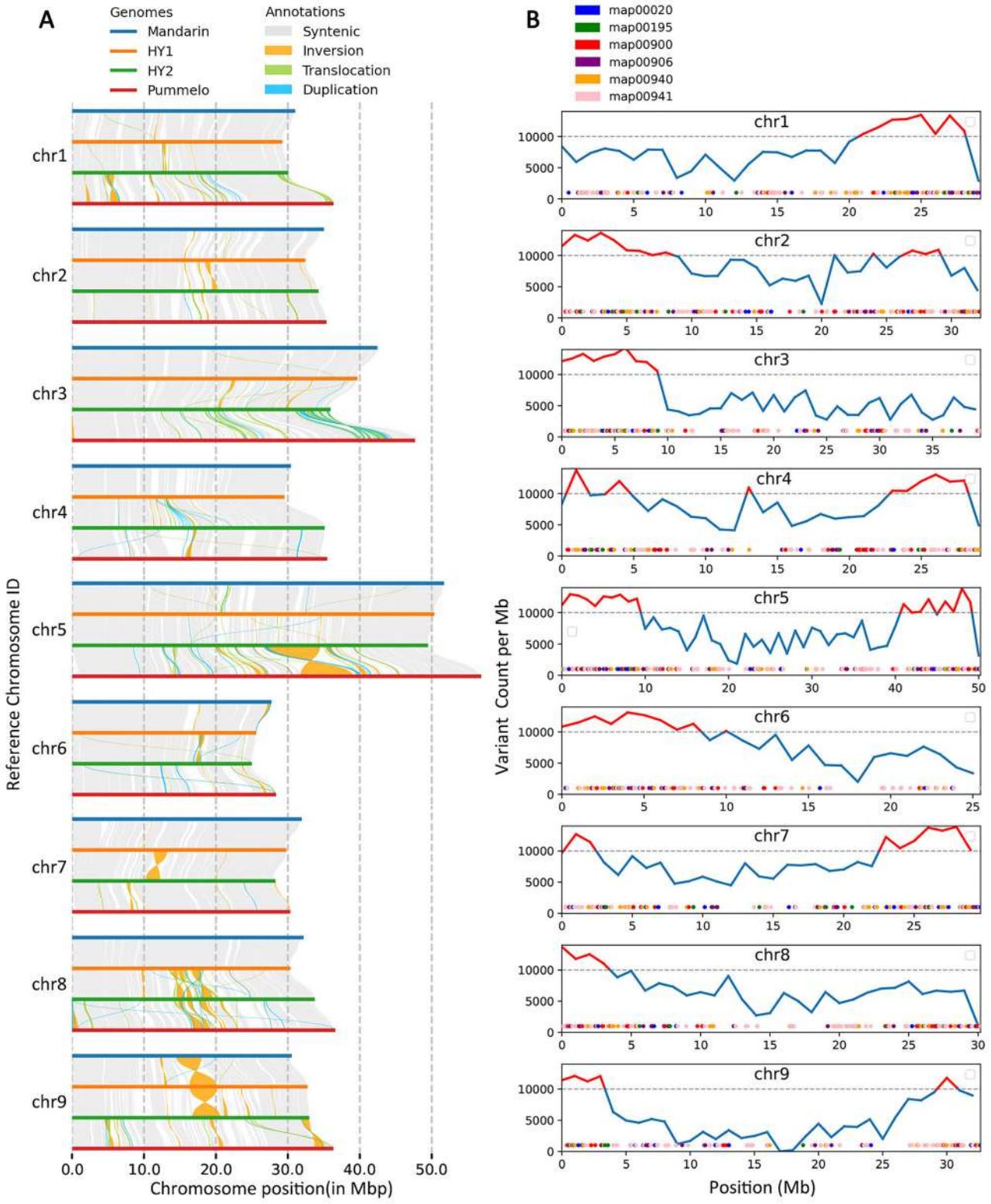


FIGURE 4 Synteny and genetic variations of HY1, HY2, mandarin, and pummelo genomes. (A) Display the synteny and structural variations of HY1, HY2, and their ancestral genomes, mandarin (*Citrus reticulata*) and pummelo (*Citrus grandis*). Pair-wise sequential whole genome alignments in the order of mandarin-HY1-HY2-pummelo were performed. Synteny (gray) and large structural variations (inversions: orange, translocations: green, duplications: blue) were identified using SyRI. (B) Displays the distribution of the allelic genetic variants (single-nucleotide polymorphisms [SNPs] and indels based on SyRI outputs) between HY1 and HY2 and important pathway genes (map00020- citrate cycle, map00195- photosynthesis, map00900- terpenoid biosynthesis, map00906- carotenoid biosynthesis, map00940- phenylpropanoid biosynthesis, map00941- flavonoid biosynthesis).

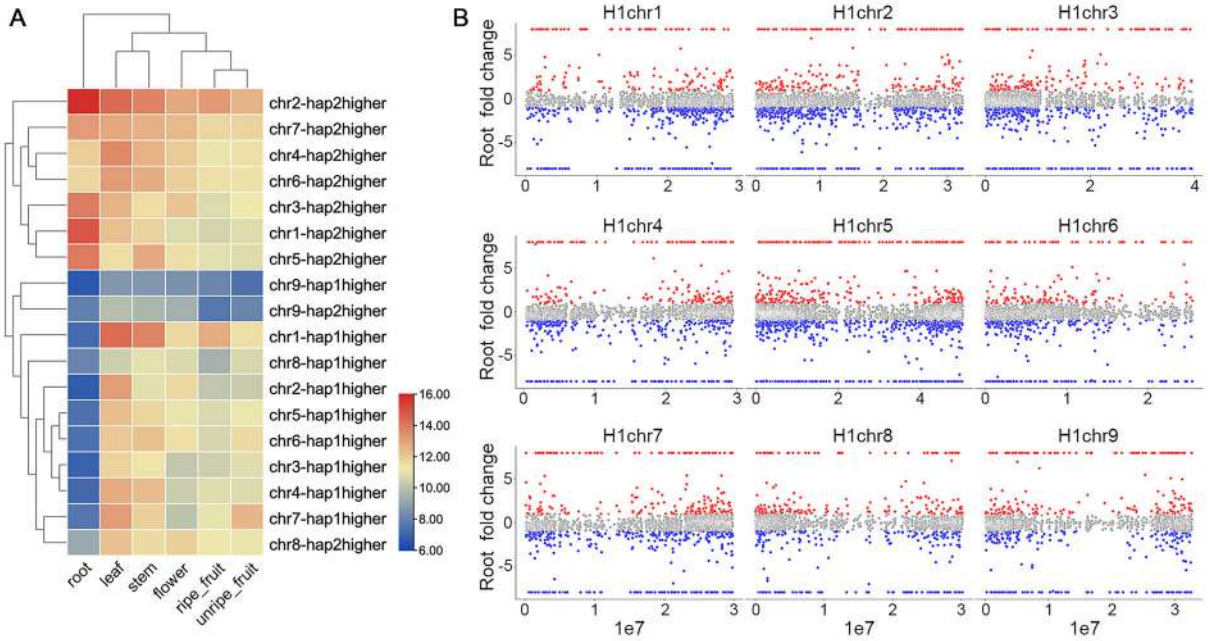
Mb) (Supporting Information File S5). These variations led to the classification of 25,503 highly diverged regions, accounting for ~163.49 Mb, or 50%, of the HY1 genome, consistent with heterozygosity. Despite these divergent regions, majority of the genome (252.08 Mb or 77.2%) was found conserved in syntenic regions, implying a generally conserved synteny between the ancestral genomes for HY1 and HY2 before their hybridization that resulted in Huyou. For structural variations, we identified a total of 73 inversions (21.04 Mb), 592 translocations (6.35 Mb), and 938 duplications (8.14 Mb), which were mainly located in the central region of each chromosome (Supporting Information File S5). Three large inversions between HY1 and HY2 were observed in the middle regions of chr9 and chr8, followed by a few moderate inversions on chr1, chr2, chr3, chr4, and chr6. In addition, a genomic region in the middle of chr5 in HY2 seems to be expanded through significant duplications, which may have contributed to the chromosome extension of chr5 in HY2 compared to chr5 in HY1 (Figure 4A). Lastly, the number of genetic variations between HY1 and HY2 was calculated per 1 Mb window along each chromosome. The obtained variation distribution profile revealed a relatively higher level of genetic diversity in the distal regions of each chromosome compared to the central regions (Figure 4B). Mapping the genes from the six important metabolic pathways (map00020- Citrate cycle, map00195- Photosynthesis, map00900- Terpenoid biosynthesis, map00906- Carotenoid biosynthesis, map00940- Phenylpropanoid biosynthesis, and map00941- Flavonoid biosynthesis) to the genome showed that many of these genes were located within the identified divergent regions, suggesting a high allelic diversity for important fruit quality characters in Huyou.

3.5 | AI of gene expression profiles in different Huyou tissues

As the most heterozygous citrus genome to date, Huyou provides an invaluable opportunity to study genome evolution. To investigate the AI post ancient hybridization, we retrieved public transcriptome data of six different tissues (root, stem, leaf, flower, unripe fruit, and ripe fruit) in Huyou from a previous study (Miao et al., 2024) and compared the expression levels of collinear genes (26,092 in total) between HY1 and HY2. The percentages of genes preferentially expressed in HY1 (HY1higher ($\log_2^{(HY1/HY2)} \geq 1$) and HY2 (HY2higher $\log_2^{(HY1/HY2)} \leq -1$) were calculated across each chromosome and tissue (Supporting Information File S6). Overall, we observed relatively higher numbers of HY2-preferential genes than HY1-preferential genes across most tissues and chromosomes (Figure 5A), implying a clear transcriptional bias for HY2 genes following hybridization. Noteworthy, we found that the HY2 allelic dominance was exceptionally prominent in root tissue (HY2higher at 12.5% versus HY1higher at 7.24%) compared to that in the other tissues (leaf, stem, flower, unripe fruit, and ripe fruit), which only displayed slight HY2 allelic dominance (11.67% vs. 11.10% on average), implying a strong tissue-specific pattern. The exceptional HY2 allelic dominance in root tissue across the nine chromosomes could be seen in Figure 5B, relatively more evident on chromosomes chr1–7 and relatively mild on chr8 and chr9. This might be caused by the large segmental replacements observed on chr8 and chr9. Indeed, at the chromosome level, chr9 displayed the lowest level of total allelic divergence (HY1higher + HY2higher = 16.77%), followed by chr8 (21.29%) and other chromosomes (22.30%–24.27%).

Due to the exceptional HY2 allelic dominance in the root tissue, the collinear genes with transcriptional preference from HY1 (1354 HY1higher genes) and HY2 (2432 HY2higher genes) were used for KEGG pathway enrichment analysis (corrected p-value $\leq$ 1-e5). Interestingly, these two sets of genes shared some commonly enriched pathways that are mainly related to antioxidants biosynthesis, including flavonoid, phenylpropanoid, isoflavonoid, carotenoid, benzoxazinoid, flavone, flavonol, alkaloid, ascorbate, and steroid biosynthesis (Figure 5C). It is well-known that antioxidant biosynthesis plays critical role in plants' adaptation to various biotic and abiotic stresses. The AI of genes within these pathways in the root tissue may have important implication in Huyou's environmental adaptation following its ancient genome hybridization. In addition to these common pathways, HY1 preferential genes were uniquely enriched with amino acid and linoleic acid metabolism and the biosynthesis of cutin, wax, diterpenoid, triterpenoid, sesquiterpenoid, stilbenoid, diarylheptanoid, and gingerol, which are closely linked to environmental adaptation (Figure 5C). In contrast, HY2 preferential genes were uniquely enriched with protein families related to signaling and cellular processes, including transporter genes, sphingolipid metabolism, genes responsible for fructose, mannose, pentose, and glucuronate interconversions, and glycerolphospholipid and peptidoglycan metabolism. In addition, the biosynthesis of anthocyanin, folate, and various secondary metabolites was also enriched in HY2 preferential genes but not in HY1 preferential genes.

In addition to the root tissue, the AI genes from HY1 (2053 HY1higher genes) and HY2 (2084 HY2higher genes) in ripe fruit tissue were also investigated. Overall, we observed strikingly similar enrichment profiles to that in the root tissue (Figure 5D). In detail, like that in the root tissue, the commonly enriched pathways from HY1 and HY2 were also prominently related to antioxidant biosynthesis, such as flavonoid, phenylpropanoid, isoflavonoid, flavone, flavonol, isoquinoline alkaloid, carotenoid, benzoxazinoid, and glucosinolate. Other common pathway terms included lipid, cofactors, and vitamin, amino acid, porphyrin, and



C

Root

Common terms in both Hap1 and Hap2	Hap1 Selected set (1354 total)	Hap1 Background (21789 total)	Hap2 Selected set (2432 total)	Hap2 Background (21789 total)
Flavonoid biosyn.	136	1065	192	1065
Phenylpropanoid biosyn.	159	1376	241	1376
Isoflavonoid biosyn.	129	1038	178	1038
Biosyn. of other secondary metabolites	229	2289	378	2289
Enzymes with EC numbers	437	5177	705	5177
Carotenoid biosyn.	105	817	153	817
Benzoxazinoid biosyn.	97	737	132	737
Metabolism of cofactors and vitamins	197	1986	307	1986
Metabolism of other amino acids	166	1641	271	1641
Lipid metabolism	285	3239	493	3239
Flavone and flavonol biosyn.	83	644	115	644
Isoquinoline alkaloid biosyn.	127	1182	200	1182
Amino acid metabolism	295	3428	504	3428
Tryptophan metabolism	131	1241	202	1241
Ascorbate and aldarate metabolism	85	708	144	708
Glucosinolate biosyn.	73	573	102	573
Carbohydrate metabolism	266	3157	445	3157
Porphyrin and chlorophyll metabolism	83	751	129	751
Steroid biosyn.	81	736	130	736

Hap1 unique terms	Selected set	Background
Cyanoamino acid metabolism	92	792
Cutin, suberine and wax biosyn.	68	549
alpha-Linolenic acid metabolism	81	699
Diterpenoid biosyn.	78	693
Sesquiterpenoid and triterpenoid biosyn.	75	674
Linoleic acid metabolism	52	411
Stilbenoid, diarylheptanoid and gingerol biosyn.	75	680
Environmental information processing	372	4830

Hap2 unique terms		
Transporters	324	2109
Protein families: signaling and cellular processes	1057	8229
Sphingolipid metabolism	90	430
Fructose and mannose metabolism	86	408
Pentose and glucuronate interconversions	112	617
Biosyn. of various secondary metabolites	42	164
Anthocyanin biosyn.	62	290
Glycerophospholipid metabolism	153	935
Folate biosyn.	56	258
Metabolism	1116	9004
Peptidoglycan biosyn. and degradation proteins	17	42

D

Ripe fruit

Common terms in both Hap1 and Hap2	Hap1 Selected set (2053 total)	Hap1 Background (21789 total)	Hap2 Selected set (2084 total)	Hap2 Background (21789 total)
Enzymes with EC numbers	635	5177	611	5177
Lipid metabolism	424	3239	412	3239
Flavonoid biosyn.	173	1065	168	1065
Phenylpropanoid biosyn.	209	1376	211	1376
Isoflavonoid biosyn.	166	1038	162	1038
Biosyn. of other secondary metabolites	310	2289	315	2289
Metabolism of cofactors and vitamins	273	1986	274	1986
Flavone and flavonol biosyn.	110	644	110	644
Amino acid metabolism	420	3428	413	3428
Isoquinoline alkaloid biosyn.	173	1182	170	1182
Metabolism of other amino acids	225	1641	240	1641
Carbohydrate metabolism	383	3157	405	3157
Cyanoamino acid metabolism	122	792	128	792
Carotenoid biosyn.	122	817	143	817
Porphyrin and chlorophyll metabolism	113	751	120	751
Benzoxazinoid biosyn.	111	737	123	737
Glucosinolate biosyn.	91	573	97	573

Hap1 unique terms	Selected set	Background
alpha-Linolenic acid metabolism	123	699
Cutin, suberine and wax biosyn.	100	549
Steroid biosyn.c	113	736
Nitrogen metabolism	82	505
Environmental information processing	543	4830
Stilbenoid, diarylheptanoid and gingerol	103	680
Sesquiterpenoid and triterpenoid biosyn.	102	674

Hap2 unique terms		
Tryptophan metabolism	189	1241
Biosyn. of various secondary metabolites	45	164
Anthocyanin biosyn.	64	290
Ascorbate and aldarate metabolism	119	708
Sphingolipid metabolism	82	430
Pentose and glucuronate interconversions	105	617
Transporters	279	2109
Metabolism of terpenoids and polyketides	249	1912
Folate biosyn.	52	258
Peptidoglycan biosyn. and degradation	16	42

FIGURE 5 Transcriptional allelic imbalance between HY1 and HY2 in different tissues. (A) Clustering of the percentages of HY1-preferential ($\log_2^{(HY1/HY2)} \geq 1$) and HY2-preferential ($\log_2^{(HY1/HY2)} \leq -1$) genes six tissues (root, stem, leaf, flower, unripe fruit, and ripe fruit) in each chromosome (Supporting Information File S7). Only the collinear genes between HY1 and HY2 were counted. The percentages of HY1 and HY2 preferential genes were calculated relative to the total number of gene in each chromosome. (B) Displays the distribution of HY1 and HY2

(Continues)

FIGURE 5 (Continued)
preferential genes along each chromosome for the root transcriptome. (C) Kyoto encyclopedia of genes and genomes (KEGG) enrichment of allelic imbalance collinear genes in root with a corrected p-value $< 1 \times 10^{-5}$ was shown. (D) KEGG enrichment of allelic imbalance collinear genes in ripe fruit with a corrected p-value $< 1 \times 10^{-5}$ was shown.

chlorophyll metabolism, which were also detected in the root tissue (Figure 5D). The unique pathway terms in HY1 (linolenic acid metabolism and the biosynthesis of cutin, suberine, wax, stilbenoid, diarylheptanoid, gingerol, sesquiterpenoid, and triterpenoid) and HY2 (tryptophan, ascorbate, aldarate, sphingolipid, pentose, glucuronate, terpenoids metabolisms, and the biosynthesis of secondary metabolites, anthocyanin, folate, and peptidoglycan) in ripe fruit tissue were also similar to those in the root tissue (Figure 5D).

3.6 | Tissue-specific AI in key pathways

Carotenoid biosynthesis pathway (map00906) plays a critical role in citrus fruit quality, controlling fruit color and flavor. To investigate the genomic characteristics of key Huyou candidate genes in this pathway, we first investigated the AI by mapping the ratio of HY1 and HY2 collinear gene transcription to this pathway. While the major components of the pathway displayed no allelic preference, we observed a clear tissue-specific AI for some key genes responsible for the astaxanthin biosynthesis specifically (Figure 6). For instance, we found one enzyme (EC 1.14.15.24) encoding beta-carotene 3-hydroxylase displayed a consistent HY2 dominance in the root tissue but a HY1 dominance in the stem, flower, unripe fruit, and ripe fruit tissue and no imbalance in leaf. Another enzyme (EC 1.23.5.1) encoding violaxanthin de-epoxidase displayed HY1 dominance in root, leaf, and ripe fruit but no imbalance in the other tissue. Conversely, the enzyme (EC 1.2.3.14) encoding abscisic-aldehyde oxidase display HY2 dominance in stem, leaf, and unripe fruit but not in other tissue. These observations provided direct support of various AIs present in Huyou genome and shed valuable insights into genome evolution following ancient hybridization.

To characterize the complete profile of the carotenoid biosynthesis pathway in Huyou, we further extracted the gene copy number information of key genes (represented by 39 KEGG orthologs) in HY1, HY2, and another eight citrus species and explored their transcriptional profiles in six Huyou tissues. Overall, we observed conserved gene presence and copy number pattern across different citrus species, with cytochrome P450 (CYP701, CYP97C1, CYP707A, CYP97A3, and CYP82G1) as the most expanded protein families within the carotenoid pathway. Copy number variations (CNVs) were observed in all KO orthologs except for two single copy genes K15744 (zeta-carotene isomerase) and K17911 (beta-carotene isomerase) (Figure 7A). Similarly, HY1 and HY2 displayed varied copy numbers in 33 out of the 39 KO orthologs. Noteworthy, Huyou (HY1 or HY2) displayed the highest copy numbers in 21 out of the 39 KO orthologs (Figure 7A). At the gene expression level, we observed varied transcriptional profiles across the six tissues investigated (Figure 7B). Interestingly, we found that over half of the KO orthologous genes in Huyou displayed the highest expression in the flower tissue, followed by the root tissue. Instead, out of the 39 KO orthologs, we found 17 KOs displayed above median transcription in fruit tissues (either unripe or ripe), of which six KOs (FDFT1, LCYB, LCYE, CCS1, AOG, and DWARF27) displayed the highest expression in fruit tissues (highlighted in red rectangles in Figure 7C), which require more attention for their potential critical roles in fruit development, despite that some of these genes may not be expanded in Huyou (Figure 7C). At the pathway map level, we found most of the genes leading to terpenoid biosynthesis (C30, C15, C5, C10, and C20) were expanded compared to other species, while this pattern was not prominent for carotenoid biosynthesis (C40) (Figure 7C). Taken together, CNV and transcriptome analyses revealed unique genomic characteristics for the carotenoid biosynthesis pathway in Huyou, which may have contributed to its distinct fruit flavor against other citrus species.

4 | DISCUSSION

As a unique natural hybrid citrus landrace first discovered in Zhejiang Province in China, Huyou is widely grown in the region and holds important agricultural, economic, and medicinal values. It has many premium fruit quality characteristics, such as golden skin, abundant bioactive components, and extended storage life, which has intrigued researchers and local growers to investigate its genomic origin. Several previous studies have employed traditional molecular marker genotyping and sequencing approach to investigate Huyou's origin but only gained limited information (L. Chen et al., 2002; S. Chen et al., 2006; Y. Li et al., 2019; C. J. Xu et al., 2006). In this study, we leveraged PacBio long-read and Hi-C sequencing technologies and presented two high-quality haplotype-resolved genome assemblies, HY1 and HY2, of a 120-year-old Huyou "ancestral tree." Assisted by the haplotype-resolved assemblies, we performed in-depth ancestor tracing, phylogeny dating, genome structure, and transcriptome analyses, providing a clear view of

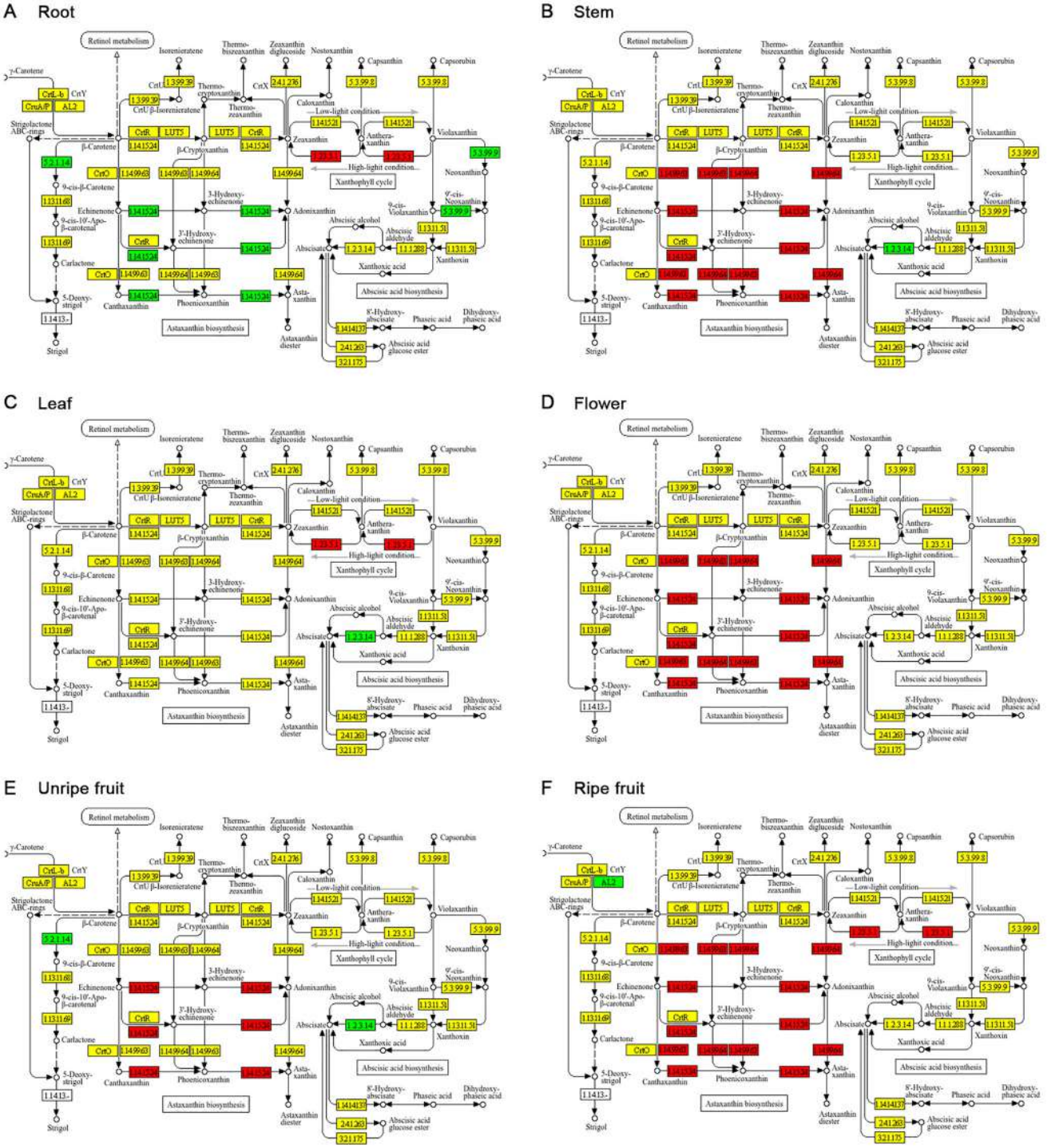


FIGURE 6 Displays the transcriptome allelic imbalance of HY1 and HY2 for map00906 pathway. The ratios of HY1 and HY2 collinear genes ($\log_2^{[HY1/HY2]}$) were mapped to Kyoto encyclopedia of genes and genomes (KEGG) pathway map00906 using TBtools. Transcriptome data in six Huyou tissues (root, stem, leaf, flower, unripe fruit, and ripe fruit) from a previous study was downloaded and analyzed. Yellow, red, and green highlights indicated neutral, HY1 dominant, and HY2 dominant, respectively. Only the Astaxanthin biosynthesis module of map00906 were displayed.

Huyou's genetic makeup. Our results explicitly showed that Huyou originated from the hybridization of mandarin and pummelo, followed by large chromosomal recombination and genetic divergence during evolution. Citrus species are known for their complex hybridization history. Based on whole genome survey of 58 worldwide diverse citrus germplasms, hybrid citrus species derived from two or more ancestral species have been found to have generally high genome heterozygosity at 1.5%–2.4% (G. A. Wu et al., 2018). In this study, we highlighted that Huyou displayed exceptional

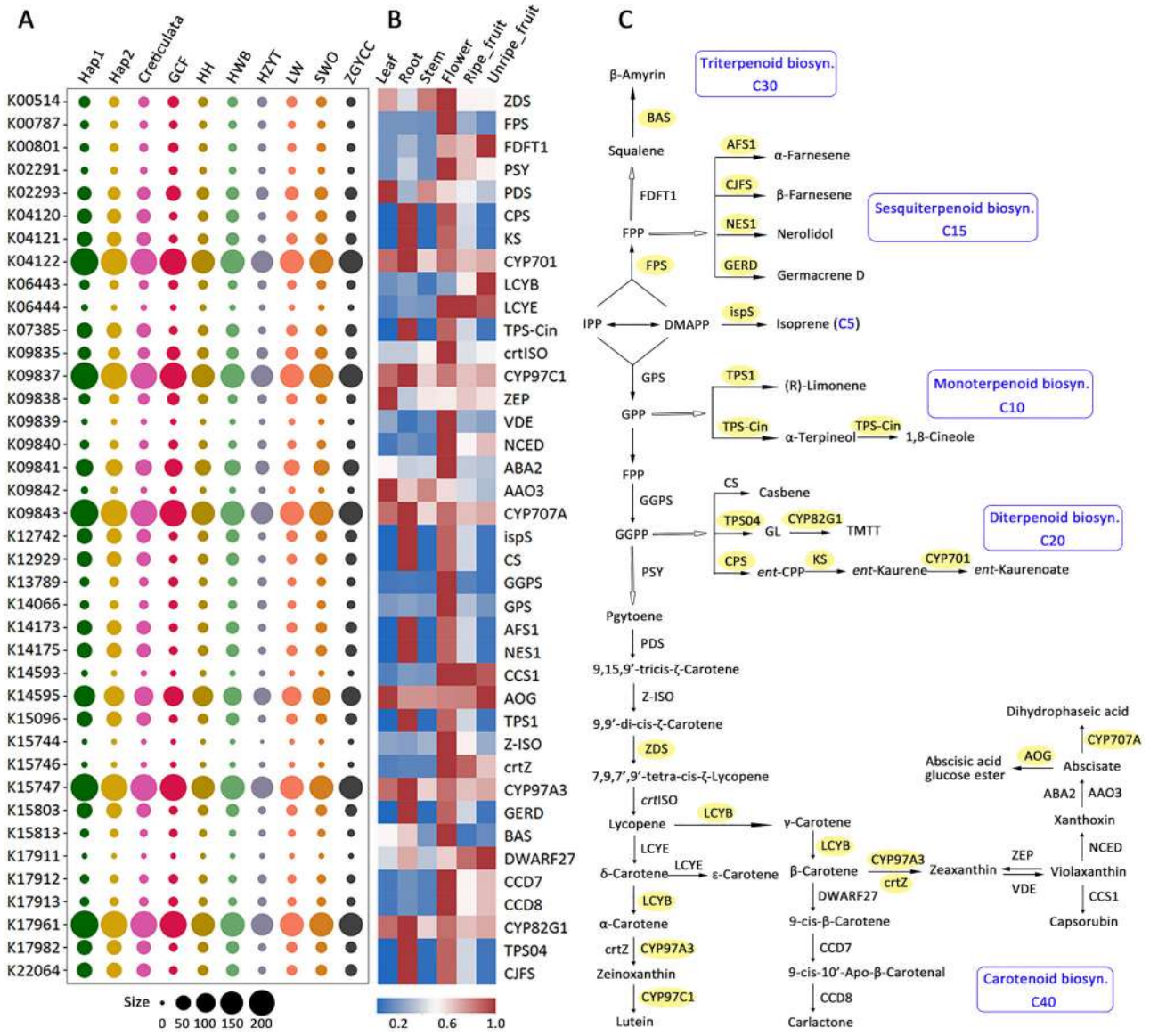


FIGURE 7 The copy number variations (CNVs) and transcriptome profiles for carotenoid pathway in Huyou. A. Displays the CNV in HY1, HY2, and other eight citrus species included in this study. Gene counts were based on results from OrthoFinder analysis in Figure 3. Genes were assigned to KO orthologs based on Kyoto encyclopedia of genes and genomes (KEGG) annotation. B. Heatmap displaying the transcription level of each KO ortholog in 6 Huyou tissues. The transcriptome values for multiple genes were summed and normalized (divided by maximum value) across the six tissues. (C) Display the metabolic map of the carotenoid pathway. Genes whose copy numbers are highest in Huyou (either HY1 or HY2) were highlighted in yellow. Genes with above median expression in fruit tissues were enclosed in red rectangles.

genome heterozygosity of 3.07%, among the highest across known citrus genomes. This will make Huyou an interesting subject for genome evolution studies by broad citrus researchers in the future. The haplotype genomes HY1 and HY2 displayed unique genomic composition, structural variations, and phylogeny position, representing two novel ancestral citrus genomes not reported before. The genetic diversity observed in Huyou genome would significantly contribute to the currently limited citrus genetic pool. Specifically, the genomic origin information of Huyou (ancestor tracing to 100 kb window) gives citrus breeder an unprecedented picture of its genomic component. This will inform and guide the potential breeding design for breeders interested in exploiting the genetic diversity of Huyou. In addition, the genetic variations within HY1 and HY2 and those against their ancestral genomes study will enable genetic mining of novel alleles for genes with known and novel functions in citrus. Potential quantitative trait loci, markers, and traits that are novel to Huyou need to identified to allow breeders to harness the genetic potential of Huyou more effectively, accelerating the development of improved citrus varieties with unique traits.

The explicit resolution of Huyou's genome origin in this study highlights the power of haplotype-resolved genomes in dissecting the complex evolutionary origin of citrus species. In the *Citrus* genus, there are a total of 30 species, most of which have been derived from the hybridization across three ancestral species: mandarin (*C. reticulata*), pummelo (*C. maxima*), and citron (*C. medica*). The ancient hybrids were subjected to further reticulate hybridization during citrus domestication, leading to a complex history of admixture in current citrus cultivars and landraces characterized by high heterozygosity (G. A. Wu et al., 2018). Tracing the evolution and domestication history of citrus species has been extremely challenging in traditional citrus genome studies based on short read sequencing, which often produced a mosaic genome assembly and failed to account for allelic variants, leading to incomplete or inaccurate representations of the genome. By leveraging a k-mer-based tracing approach on the haplotype-resolved genomes, we can resolve and quantify the ancestral components of Huyou to specific genomic regions. In addition, the haplotype genome tracing also allows us to dissect the segmental insertions, deletions, inversions, recombinations, and other structural variations between HY1 and HY2, particularly the recombination events on chr8 and chr9. These unprecedented genomic insights are critical for the accurate interpretation of the genomic and phenotypic features of Huyou. Our study was inspired by a recent study in which the haplotype-resolved lemon genome was traced to be an admixture of three ancestor genomes of citron, mandarin, and pummelo (Bao et al., 2023). Notably, numerous studies highlight the increasing trend of generating chromosome-level and haplotype-resolved genome assemblies across the plant kingdom, such as potato (P. Sun et al., 2022), lychee (Hu et al., 2022), tea (W. Y. Zhang et al., 2021), and strawberry (Jin et al., 2023). The advantages of haplotype-resolved genomes for citrus genomes are being increasingly appreciated in several recent citrus genome research including *C. limon* (Mario et al., 2021), *C. australis* (Nakandala et al., 2023), *C. changshanensis* (Miao et al., 2024), and *Papeda* (F. Wang et al., 2025). However, we noticed that only the collapsed genome sequence was reported and analyzed in several haplotype-resolved genome studies, which, in our opinion, should be avoided, particularly for those organisms with high heterozygosity such as Citrus species. We expect an increasing number of haplotype-resolved citrus genomes to be generated and published, which are positioned to revolutionize citrus genome evolution and domestication research.

The SVs such as CNVs, insertions/deletions, inversions, and translocations detected between HY1 and HY2 may important biological implications. SVs in citrus species are often responsible for phenotypic diversity related to environmental adaptation (stress and disease resistance), speciation, and domestication traits. For example, SVs were found associated with *Citrus apomixis* and domestication (N. Wang et al., 2022). TE insertions significantly affect allelic-specific expression in sweet orange (Scaglione et al., 2025). In addition, allelic-specific SVs in Huanglongbing disease-tolerant and -sensitive Valencia sweet orange mutants (B. Wu et al., 2023). Similarly, the unique SVs detected in Huyou haplotype genomes may be associated with environmental adaptation to local region. They may also have contributed to the special and premium fruit flavor of Huyou, as evidenced by the enriched photosynthesis, signal transduction, and response to stimulus GO terms in HY1-specific genes. In addition, we found most of the gene families within the carotenoid pathway displayed CNVs between HY1 and HY2 and across other citrus species, supporting the prevalence of CNVs in key pathway related to citrus fruit quality. These CNVs, particularly for genes highly expressed in the fruit tissues, may have contributed to the distinct fruit characteristics and may be exploited in future citrus breeding.

AI refers to the unequal expression of alleles from a heterozygous locus, leading to differential gene expression. This is another critical insight gained from haplotype-resolved genome assembly. In this study, we detected abundant levels of genes affected by AI, ranging from 20% to 24% (HY1 and HY2 higher) of the collinear genes. A recent study identified only 296 imprinted genes in the seeds of two pairs of reciprocal citrus crosses (He et al., 2025). Similarly, only limited numbers of AI genes were detected in other crops such as *Arabidopsis*, rapeseed, castor bean, capsella, flax, rice, maize, and sorghum (reviewed by He et al. [2025]). Most of these studies attribute AI to epigenetic causes. Consistent with our observation, 1324 AI genes were detected in Pinus densiflora with a genome heterozygosity of 1.6% (Jang et al., 2024), at a comparable level with Huyou. In addition to epigenetic cause, we reason that AI may also be caused by allelic variations, lower in self-pollinating plants but higher in cross-pollinating species. The number of AI genes may also be directly related to genome heterozygosity, whereby a higher genome heterozygosity of 3.07% in Huyou may have resulted in the exceptionally high level of AI. Based on gene expression analysis of collinear genes between HY1 and HY2, we observed clear evidence of AI (HY2 dominant over HY1) across all six tissues investigated. Another notable finding is that AI in Huyou tends to be imbalanced across tissues; that is, the HY2 allelic dominance was much more prominent in some tissues, such as root, over the others. The tissue- and stage-specific pattern of AI seems to be a common observation (He et al., 2025; Jang et al., 2024). Metabolic pathway enrichment analyses of AI genes in Huyou root tissue revealed a strong overrepresentation of genes related to various antioxidant production, such as flavonoid, phenylpropanoid, isoflavonoid, carotenoid, benzoxazinoid, flavone, flavonol, alkaloid, ascorbate, and steroid biosynthesis. These antioxidants have well-known biological functions in plants' adaptation to various biotic and abiotic

stresses. The AI of genes within these pathways in the root tissue may have important implication in Huyou's environmental adaptation following its ancient genome hybridization. It is interesting to observe that these antioxidant biosynthesis terms were over-represented in both HY1-higher and HY2-higher genes, suggesting an overall evolutionary preference of AI genes related to environmental adaptation. At the plant tissue level, we observed overall conservation of AI functions across tissues such as root and ripe fruit, despite a significantly higher degree of AI in root. This novel insight needs to be verified in future analysis in other species. In plants, especially those with high heterozygosity, AI plays a significant role in genome evolution, influencing traits such as disease resistance, growth patterns, and adaptation to environmental changes (Djari et al., 2024; Jang et al., 2024; P. Li et al., 2023). The impact of AI on plants' genome evolution and environmental adaptation has been well recognized in several recent haplotype-resolved genome studies. For example, a study in *Populus tomentosa* identified heterozygous alleles associated with adaptive traits, highlighting the role of AI in evolutionary processes (P. Li et al., 2023). Similarly, haplotype-resolved assemblies of grapevine identified structural variations between haplotypes associated with allelic expression linked to disease resistance and fruit quality (Djari et al., 2024). Another recent study in *Pinus densiflora* revealed the functional contributions of AI to flowering and abiotic stress-related traits (Jang et al., 2024). In addition to the overlapping AI functions, haplotype-specific AI functions were also observed. The biological implications of our observed AI patterns remain to be characterized, particularly for citrus genome research. As far as the authors are concerned, our study is pioneering to demonstrate the presence of strong AI in citrus hybrid genome. Based on our observation of AI in Huyou, we reason that AI should be prevalent across the various citrus genomes due to the presence of frequent hybridization events. Given the still limited haplotype-resolved genome assemblies for citrus species, the importance of AI in citrus genome evolution and its interplay with phenotypic diversity in citrus deserves much more attention in future research.

AUTHOR CONTRIBUTIONS

Zhanghui Zeng: Conceptualization; data curation; funding acquisition; investigation; project administration; writing—original draft. **Yingjie Luo**: Data curation; formal analysis; investigation; methodology. **Haifei Hu**: Data curation; formal analysis; investigation; methodology. **Lan Lan**: Data curation; formal analysis; investigation. **Baojin Guo**: Data curation; investigation. **Ping Zhou**: Data curation; formal analysis. **Cong Tan**: Data curation; formal analysis; investigation. **Xiaoping Huang**: Data curation; formal analysis; investigation. **Tuo Qi**: Data curation; formal analysis. **Zhehao Chen**: Data curation; formal analysis. **Zhiming Yu**: Data curation; formal analysis. **Lilin Wang**: Data curation; formal analysis. **Taihe Xiang**: Conceptualization; investigation; project administration; supervision. **Chengdao Li**: Conceptualization; project administration; supervision. **Yong Jia**: Conceptualization; data curation; formal analysis; investigation; project administration; supervision; writing—original draft.

ACKNOWLEDGMENTS

We acknowledge the relevant citrus research communication for making the citrus genomic data available to the public, particularly the recently published Huyou transcriptome data (Miao et al., 2024).

Open access publishing facilitated by Murdoch University, as part of the Wiley - Murdoch University agreement via the Council of Australasian University Librarians

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The raw sequencing data have been deposited at NCBI with project ID PRJNA1090545 (release upon acceptance). The genome assemblies and gene annotations are available at https://figshare.com/s/ba2b2081a2512770c31a.

ORCID

Zhanghui Zeng https://orcid.org/0000-0002-4959-6015
Yingjie Luo https://orcid.org/0009-0008-7466-7517
Chengdao Li https://orcid.org/0000-0002-9653-2700
Yong Jia https://orcid.org/0000-0002-0394-3966

REFERENCES

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology*, *23*(1), 258. https://doi.org/10.1186/s13059-022-02823-7

Andrews, S. (2012). *FastQC: A quality control tool for high throughput sequence data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bao, Y., Zeng, Z., Yao, W., Chen, X., Jiang, M., Sehrish, A., Wu, B., Powell, C. A., Chen, B., Xu, J., Zhang, X., & Zhang, M. (2023). A gap-free and haplotype-resolved lemon genome provides insights into flavor synthesis and huanglongbing (HLB) tolerance. *Horticulture Research*, *10*(4), uhad020. https://doi.org/10.1093/hr/uhad020

Baril, T., Galbraith, J., & Hayward, A. (2024). Earl grey: A fully automated user-friendly transposable element annotation and analysis pipeline. *Molecular Biology and Evolution*, *41*(4), msae068. https://doi.org/10.1093/molbev/msae068

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and*

Evolution, *17*(4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Chan, P. P., Lin, B. Y., Mak, A. J., & Lowe, T. M. (2021). tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research*, *49*(16), 9077–9096. https://doi.org/10.1093/nar/gkab688

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, *13*(8), 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Chen, L., Hu, X., & Zhao, S. (2002). Molecular research on Huyou origin. *Acta Horticulturae Sinica*, *29*(3), 276–277. https://doi.org/10.3321/j.issn:0513-353X.2002.03.021

Chen, S., Yang, H., Zheng, Y., Chen, Y., & Qiu, Y. (2006). Preliminary identification of *Citrus changshanhuyou* elite genotypes by molecular markers. *Journal of Molecular Cell Biology*, *39*(6), 502–508. https://www.ncbi.nlm.nih.gov/pubmed/17348202

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170–175. https://doi.org/10.1038/s41592-020-01056-5

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics*, *22*(10), 1269–1271. https://doi.org/10.1093/bioinformatics/btl097

Djari, A., Madignier, G., Di Valentin, O., Gillet, T., Frasse, P., Djouhri, A., Hu, G. J., Julliard, S., Liu, M. C., Zhang, Y., Regad, F., Pirrello, J., Maza, E., & Bouzayen, M. (2024). Haplotype- resolved genome assembly and implementation of VitExpress, an open interactive transcriptomic platform for grapevine. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(23), e2403750121. https://doi.org/10.1073/pnas.2403750121

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems*, *3*(1), 99–101. https://doi.org/10.1016/j.cels.2015.07.012

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Gmitter, F. G., Chen, C., Machado, M. A., de Souza, A. A., Ollitrault, P., Froehlicher, Y., & Shimizu, T. (2012). Citrus genomics. *Tree Genetics & Genomes*, *8*(3), 611–626. https://doi.org/10.1007/s11295-012-0499-2

Goel, M., Sun, H., Jiao, W. B., & Schneeberger, K. (2019). SyRI: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, *20*(1), 277. https://doi.org/10.1186/s13059-019-1911-0

Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, *36*(9), 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025

Guo, J.-j., Gao, Z.-p., Xia, J.-l., Ritenour, M. A., Li, G.-y., & Shan, Y. (2018). Comparative analysis of chemical composition, antimicrobial and antioxidant activity of citrus essential oils from the main cultivated varieties in China. *LWT-Food Science and Technology*, *97*, 825–839. https://doi.org/10.1016/j.lwt.2018.07.060

He, J. J., Hu, G., Shen, M. Y., Fan, Y. J., Shi, X. S., Wu, X. M., Guo, W. W., Xia, Q. M., & Xie, K. D. (2025). Comprehensive analysis of imprinted genes in citrus endosperm and their contributions to seed development. *Plant Journal*, *122*(6), e70290. https://doi.org/10.1111/tpj.70290

Hu, G., Feng, J., Xiang, X. U., Wang, J., Salojärvi, J., Liu, C., Wu, Z., Zhang, J., Liang, X., Jiang, Z., Liu, W., Ou, L., Li, J., Fan, G., Mai, Y., Chen, C., Zhang, X., Zheng, J., Zhang, Y., … Li, J. (2022). Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. *Nature Genetics*, *54*(1), 73–83. https://doi.org/10.1038/s41588-021-00971-3

Huang, Y., He, J., Xu, Y., Zheng, W., Wang, S., Chen, P., Zeng, B., Yang, S., Jiang, X., Liu, Z., Wang, L., Wang, X., Liu, S., Lu, Z., Liu, Z., Yu, H., Yue, J., Gao, J., Zhou, X., … Xu, Q. (2023). Pangenome analysis provides insight into the evolution of the orange subfamily and a key gene for citric acid accumulation in citrus fruits. *Nature Genetics*, *55*, 1964–1975. https://doi.org/10.1038/s41588-023-01516-6

Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., & Bruyere, C., … French-Italian Public Consortium for Grapevine Genome, C. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, *449*(7161), 463–467. https://doi.org/10.1038/nature06148

Jang, M.-J., Cho, H. J., Park, Y.-S., Lee, H.-Y., Bae, E.-K., Jung, S., Jin, H., Woo, J., Park, E., Kim, S.-J., Choi, J.-W., Chae, G. Y., Guk, J.-Y., Kim, D. Y., Kim, S.-H., Kang, M.-J., Lee, H., Cheon, K.-S., Kim, I. S., … Kim, S. (2024). Haplotype-resolved genome assembly and resequencing analysis provide insights into genome evolution and allelic imbalance in. *Nature Genetics*, *56*(11), 2551–2561. https://doi.org/10.1038/s41588-024-01944-y

Jia, K.-H., Wang, Z.-X., Wang, L., Li, G.-Y., Zhang, W., Wang, X.-L., Xu, F.-J., Jiao, S.-Q., Zhou, S.-S., Liu, H., Ma, Y., Bi, G., Zhao, W., El-Kassaby, Y. A., Porth, I., Li, G., Zhang, R.-G., & Mao, J.-F. (2022). SubPhaser: A robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *New Phytologist*, *235*(2), 801–809. https://doi.org/10.1111/nph.18173

Jiang, J., Yan, L., Shi, Z., Wang, L., Shan, L., & Efferth, T. (2019). Hepatoprotective and anti-inflammatory effects of total flavonoids of Qu Zhi Ke (peel of *Citrus changshan-huyou*) on non-alcoholic fatty liver disease in rats via modulation of NF-κB and MAPKs. *Phytomedicine*, *64*, 153082. https://doi.org/10.1016/j.phymed.2019.153082

Jin, X., Du, H. Y., Zhu, C. M., Wan, H., Liu, F., Ruan, J. W., Mower, J. P., & Zhu, A. D. (2023). Haplotype-resolved genomes of wild octoploid progenitors illuminate genomic diversifications from wild relatives to cultivated strawberry. *Nature Plants*, *9*(8), 1252–1266. https://doi.org/10.1038/s41477-023-01473-2

Keilwagen, J., Hartung, F., & Grau, J. (2019). GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Methods in Molecular Biology*, *1962*, 161–177. https://doi.org/10.1007/978-1-4939-9173-0_9

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645. https://doi.org/10.1101/gr.092759.109

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, Article 323. https://doi.org/10.1186/1471-2105-12-323

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, P., Xiao, L., Du, Q., Quan, M., Song, Y., He, Y., Huang, W., Xie, J., Lv, C., Wang, D., Zhou, J., Li, L., Liu, Q., El-Kassaby, Y. A., & Zhang, D. (2023). Genomic insights into selection for heterozygous alleles and woody traits in. *Plant Biotechnology Journal*, *21*(10), 2002–2018. https://doi.org/10.1111/pbi.14108

Li, Y., Zhou, S., Feng, M., Lin, Y., Li, L., Ma, G., & Xiang, T. (2017). Regeneration of plants from ancestry tree of *Citrus changshan-huyou* Y. B. Chang via tissue culture. *Bangladesh Journal of Botany*, *46*(3), 1233–1240.

Li, Y., Zhou, S., Tong, X., Wang, L., & Xiang, T. (2019). Analysis of the origin of *Citrus changshan-huyou* by DNA barcode. *International Journal of Agriculture and Biology*, *22*(5), 973–978. https://doi.org/10.17957/IJAB/15.1157

Liu, H., Wang, X., Liu, S., Huang, Y., Guo, Y.-X., Xie, W.-Z., Liu, H., Tahir Ul Qamar, M., Xu, Q., & Chen, L.-L. (2022). Citrus pangenome to breeding database (CPBD): A comprehensive genome database for citrus breeding. *Molecular Plant*, *15*(10), 1503–1505. https://doi.org/10.1016/j.molp.2022.08.006

Mario, D., Marco, M., Mirko, M., Chiara, C., Michela, T., Deng, Z. N., Alessandro, C., Marco, C., Gaetano, D., Stefano, L., Luca, B., & Aalessandra, G. (2021). The haplotype-resolved reference genome of lemon (*Citrus limon* L. Burm f.). *Tree Genetics & Genomes*, *17*(6), 46. https://doi.org/10.1007/s11295-021-01528-5

Miao, C., Wu, Y., Wang, L., Zhao, S., Grierson, D., Xu, C., Chen, W., & Chen, K. (2024). Haplotype-resolved chromosome-level genome assembly of Huyou (*Citrus changshanensis*). *Scientific Data*, *11*(1), 605. https://doi.org/10.1038/s41597-024-03437-3

Nakandala, U., Masouleh, A. K., Smith, M. W., Furtado, A., Mason, P., Constantin, L., & Henry, R. J. (2023). Haplotype resolved chromosome level genome assembly of *Citrus australis* reveals disease resistance and other citrus specific genes. *Horticulture Research*, *10*(5), uhad058. https://doi.org/10.1093/hr/uhad058

Nawrocki, E. P., Kolbe, D. L., & Eddy, S. R. (2009). Infernal 1.0: Inference of RNA alignments. *Bioinformatics*, *25*(10), 1335–1337. https://doi.org/10.1093/bioinformatics/btp157

Peng, Z., Bredeson, J. V., Wu, G. H. A., Shu, S. Q., Rawat, N., Du, D. L., Parajuli, S., Yu, Q. B., You, Q., Rokhsar, D. S., Gmitter, F. G., & Deng, Z. N. (2020). A chromosome-scale reference genome of trifoliate orange (*Poncirus trifoliata*) provides insights into disease resistance, cold tolerance and genome evolution in. *Plant Journal*, *104*(5), 1215–1232. https://doi.org/10.1111/tpj.14993

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*(1), 1432. https://doi.org/10.1038/s41467-020-14998-3

Scaglione, D., Ciacciulli, A., Gattolin, S., Caruso, M., Marroni, F., Casas, G. L., Jurman, I., Licciardello, G., Catara, A. F., Rossini, L., Licciardello, C., & Morgante, M. (2025). Deep resequencing unveils novel SNPs, InDels, and large structural variants for the clonal fingerprinting of sweet orange [*Citrus sinensis* (L.) Osbeck]. *The Plant Genome*, *18*(1), e20544. https://doi.org/10.1002/tpg2.20544

Sheng, L., Shen, D., Luo, Y., Sun, X., Wang, J., Luo, T., Zeng, Y., Xu, J., Deng, X., & Cheng, Y. (2017). Exogenous gamma-aminobutyric acid treatment affects citrate and amino acid accumulation to improve fruit quality and storage performance of postharvest citrus fruit. *Food Chemistry*, *216*, 138–145. https://doi.org/10.1016/j.foodchem.2016.08.024

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Strijk, J. S., Hinsinger, D. D., Roeder, M. M., Chatrou, L. W., Couvreur, T. L. P., Erkens, R. H. J., Sauquet, H., Pirie, M. D., Thomas, D. C., & Cao, K. (2021). Chromosome-level reference genome of the soursop (*Annona muricata*): A new resource for Magnoliid research and tropical pomology. *Molecular Ecology Resources*, *21*(5), 1608–1619. https://doi.org/10.1111/1755-0998.13353

Sun, H., Jiao, W.-B., Krause, K., Campoy, J. A., Goel, M., Folz-Donahue, K., Kukat, C., Huettel, B., & Schneeberger, K. (2022). Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nature Genetics*, *54*(3), 342–348. https://doi.org/10.1038/s41588-022-01015-0

Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., Xi, Z., Wang, X., & Liu, J. (2022). WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Molecular Plant*, *15*(12), 1841–1851. https://doi.org/10.1016/j.molp.2022.10.018

Suyama, M., Torrents, D., & Bork, P. (2006). PAL_2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*, W609–W612. https://doi.org/10.1093/nar/gkl315

Wang, F., Wang, S., Wu, Y., Jiang, D., Yi, Q., Zhang, M., Yu, H., Yuan, X., Li, M., Li, G., Cheng, Y., Feng, J., Wang, X., Cheng, C., Zhu, S., & Liu, R. (2025). Haplotype-resolved genome of a papeda provides insights into the geographical origin and evolution of *Citrus*. *Journal of Integrative Plant Biology*, *67*(2), 276–293. https://doi.org/10.1111/jipb.13819

Wang, L., He, F., Huang, Y., He, J., Yang, S., Zeng, J., Deng, C., Jiang, X., Fang, Y., Wen, S., Xu, R., Yu, H., Yang, X., Zhong, G., Chen, C., Yan, X., Zhou, C., Zhang, H., Xie, Z., … Xu, Q. (2018). Genome of Wild Mandarin and Domestication History of Mandarin. *Molecular Plant*, *11*(8), 1024–1037. https://doi.org/10.1016/j.molp.2018.06.001

Wang, N., Song, X. T., Ye, J. L., Zhang, S. Q., Cao, Z., Zhu, C. Q., Hu, J. B., Zhou, Y., Huang, Y., Cao, S., Liu, Z. J., Wu, X. M., Chai, L. J., Guo, W. W., Xu, Q., Gaut, B. S., Koltunow, A. M. G., Zhou, Y. F., & Deng, X. X. (2022). Structural variation and parallel evolution of apomixis in citrus during domestication and diversification. *National Science Review*, *9*(10), nwac114. https://doi.org/10.1093/nsr/nwac114

Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., Wu, G., Tian, S., Chen, C., Liu, Y., Yu, H., Yang, X., Lan, H., Wang, N., Wang, L., Xu, J., Jiang, X., Xie, Z., Tan, M., … Xu, Q. (2017). Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature Genetics*, *49*(5), 765–722. https://doi.org/10.1038/ng.3839

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: A toolkit for detection and evolutionary analysis

of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), e49. https://doi.org/10.1093/nar/gkr1293

Wu, B., Yu, Q., Deng, Z., Duan, Y., Luo, F., & Gmitter Jr, F. (2023). A chromosome-level phased genome enabling allele-level studies in sweet orange: A case study on citrus Huanglongbing tolerance. *Horticulture Research*, *10*(1), uhac247. https://doi.org/10.1093/hr/uhac247

Wu, G. A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., Takita, M. A., Labadie, K., Poulain, J., Couloux, A., Jabbari, K., Cattonaro, F., Del Fabbro, C., Pinosio, S., Zuccolo, A., … Rokhsar, D. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology*, *32*(7), 656–662. https://doi.org/10.1038/nbt.2906

Wu, G. A., Terol, J., Ibanez, V., López-García, A., Pérez-Román, E., Borredá, C., Domingo, C., Tadeo, F. R., Carbonell-Caballero, J., Alonso, R., Curk, F., Du, D. L., Ollitrault, P., Roose, M. L., Dopazo, J., Gmitter, F. G., Rokhsar, D. S., & Talon, M. (2018). Genomics of the origin and evolution of *Citrus*. *Nature*, *554*(7692), 311–316. https://doi.org/10.1038/nature25447

Xie, J., Chen, Y., Cai, G., Cai, R., Hu, Z., & Wang, H. (2023). Tree visualization by one table (tvBOT): A web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Research*, *51*(W1), W587–W592. https://doi.org/10.1093/nar/gkad359

Xu, C. J., Bao, L., Zhang, B., Bei, Z. M., Ye, X. Y., Zhang, S. L., & Chen, K. S. (2006). Parentage analysis of huyou (*Citrus changshanensis*) based on internal transcribed spacer sequences. *Plant Breeding*, *125*(5), 519–522. https://doi.org/10.1111/j.1439-0523.2006.01263.x

Xu, Q., Chen, L.-L., Ruan, X., Chen, D., Zhu, A., Chen, C., Bertrand, D., Jiao, W.-B., Hao, B.-H., Lyon, M. P., Chen, J., Gao, S., Xing, F., Lan, H., Chang, J.-W., Ge, X., Lei, Y., Hu, Q., Miao, Y., … Ruan, Y. (2013). The draft genome of sweet orange (*Citrus sinensis*). *Nature Genetics*, *45*(1), 59–66. https://doi.org/10.1038/ng.2472

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, Z. H., & Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, *23*(1), 212–226. https://doi.org/10.1093/molbev/msj024

Yin-bin, C. (1991). A new species of genus citrus from China. *Bulletin of Botanical Research*, *11*(2), 5–7.

Zhang, H., Song, L., Wang, X., Cheng, H., Wang, C., Meyer, C. A., Liu, T., Tang, M., Aluru, S., Yue, F., Liu, X. S., & Li, H. (2021). Fast alignment and preprocessing of chromatin profiles with Chromap. *Nature Communications*, *12*(1), 6566. https://doi.org/10.1038/s41467-021-26865-w

Zhang, H., Wang, F., Zeng, C., Zhu, W., Xu, L., Wang, Y., Zeng, J., Fan, X., Sha, L., Wu, D., Cheng, Y., Zhang, H., Chen, G., Zhou, Y., & Kang, H. (2022). Development and application of specific FISH probes for karyotyping Psathyrostachys huashanica chromosomes. *BMC Genomics*, *23*(1), Article 309. https://doi.org/10.1186/s12864-022-08516-6

Zhang, J., Sun, C., Yan, Y., Chen, Q., Luo, F., Zhu, X., Li, X., & Chen, K. (2012). Purification of naringin and neohesperidin from Huyou (*Citrus changshanensis*) fruit and their effects on glucose consumption in human HepG2 cells. *Food Chemistry*, *135*(3), 1471–1478. https://doi.org/10.1016/j.foodchem.2012.06.004

Zhang, W. Y., Luo, C., Scossa, F., Zhang, Q. H., Usadel, B., Fernie, A. R., Mei, H. W., & Wen, W. W. (2021). A phased genome based on single sperm sequencing reveals crossover pattern and complex relatedness in tea plants. *Plant Journal*, *105*(1), 197–208. https://doi.org/10.1111/tpj.15051

Zheng, W., Zhang, W., Liu, D., Yin, M., Wang, X., Wang, S., Shen, S., Liu, S., Huang, Y., Li, X., Zhao, Q., Yan, L., Xu, Y., Yu, S., Hu, B., Yuan, T., Mei, Z., Guo, L., Luo, J., … Ma, Z. (2023). Evolution-guided multiomics provide insights into the strengthening of bioactive flavone biosynthesis in medicinal pummelo. *Plant Biotechnology Journal*, *21*(8), 1577–1589. https://doi.org/10.1111/pbi.14058

Zhou, C., McCarthy, S. A., & Durbin, R. (2023). YaHS: Yet another Hi-C scaffolding tool. *Bioinformatics*, *39*(1), btac808. https://doi.org/10.1093/bioinformatics/btac808

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zeng, Z., Luo, Y., Hu, H., Lan, L., Guo, B., Zhou, P., Tan, C., Huang, X., Qi, T., Chen, Z., Yu, Z., Wang, L., Xiang, T., Li, C., & Jia, Y. (2026). Highly heterozygous *Citrus changshan-huyou* Y. B. Chang originated from ancient hybridization between mandarin and pummelo and displayed distinct tissue-specific allelic imbalance. *The Plant Genome*, *19*, e70232. https://doi.org/10.1002/tpg2.70232