the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) **117**, 1264–1280 doi: 10.1111/tpj.16543

RESOURCE

Chromosomal-scale genomes of two *Rosa* species provide insights into genome evolution and ascorbate accumulation

Dan Zong[1,†], Huan Liu[2,†], Peihua Gan[1,†], Shaojie Ma[1], Hongping Liang[2], Jinde Yu[1], Peilin Li[1], Tao Jiang[1], Sunil Kumar Sahu[2], Qingqing Yang[1], Deguo Zhang[1], Laigeng Li[3], Xu Qiu[1], Wenwen Shao[2], Jinlong Yang[4], Yonghe Li[5,*], Xuanmin Guang[2,*] and Chengzhong He[1,*]

[1]*Key Laboratory for Forest Genetic and Tree Improvement and Propagation in Universities of Yunnan Province, Southwest Forestry University, Kunming 650224, China,*

[2]*State Key Laboratory of Agricultural Genomics, Key Laboratory of Genomics, Ministry of Agriculture, BGI Research, Shenzhen 518083, China,*

[3]*National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 20032, China,*

[4]*BGI Research, Kunming 650106, China, and*

[5]*Yunnan Agricultural University, Kunming 650201, China*

Received 14 July 2022; revised 7 October 2023; accepted 30 October 2023; published online 15 November 2023.
*For correspondence (e-mail hecz@swfu.edu.cn; guangxuanmin@genomics.cn; swfclyh@126.com).
†These authors contributed equally to this work.

SUMMARY

***Rosa roxburghii* and *Rosa sterilis*, two species belonging to the Rosaceae family, are widespread in the southwest of China. These species have gained recognition for their remarkable abundance of ascorbate in their fresh fruits, making them an ideal vitamin C resource. In this study, we generated two high-quality chromosome-scale genome assemblies for *R. roxburghii* and *R. sterilis*, with genome sizes of 504 and 981.2 Mb, respectively. Notably, we present a haplotype-resolved, chromosome-scale assembly for diploid *R. sterilis*. Our results indicated that *R. sterilis* originated from the hybridization of *R. roxburghii* and *R. longicuspis*. Genome analysis revealed the absence of recent whole-genome duplications in both species and identified a series of duplicated genes that possibly contributing to the accumulation of flavonoids. We identified two genes in the ascorbate synthesis pathway, *GGP* and *GalLDH*, that show signs of positive selection, along with high expression levels of GDP-D-mannose 3′, 5′-epimerase (*GME*) and GDP-L-galactose phosphorylase (*GGP*) during fruit development. Furthermore, through co-expression network analysis, we identified key hub genes (*MYB5* and *bZIP*) that likely regulate genes in the ascorbate synthesis pathway, promoting ascorbate biosynthesis. Additionally, we observed the expansion of terpene synthase genes in these two species and tissue expression patterns, suggesting their involvement in terpenoid biosynthesis. Our research provides valuable insights into genome evolution and the molecular basis of the high concentration of ascorbate in these two *Rosa* species.**

Keywords: *Rosa roxburghii* and *Rosa sterilis*, chromosome-scale genome, haplotype, hybridization, genome evolution, ascorbate biosynthesis.

INTRODUCTION

Rosa roxburghii ($2n = 2x = 14$), commonly known as chestnut rose, is a wild fruit crop belonging to the genus *Rosa* of the Rosaceae family. This deciduous shrub is characterized by its rose-like flowers and spiky fruits and mainly grows at a high latitude of 1000–1600 m in the mountains of Southwest China (Figure 1a) (Chen & Kan, 2018; Xu et al., 2004). With a history of medicinal use spanning centuries, *R. roxburghii* is valued for its exceptionally high levels of flavonoids and polysaccharides in its fruits (Chen & Kan, 2018; Yuan et al., 2020). Flavonoids extracted from *R. roxburghii* are effective antioxidants and exhibit

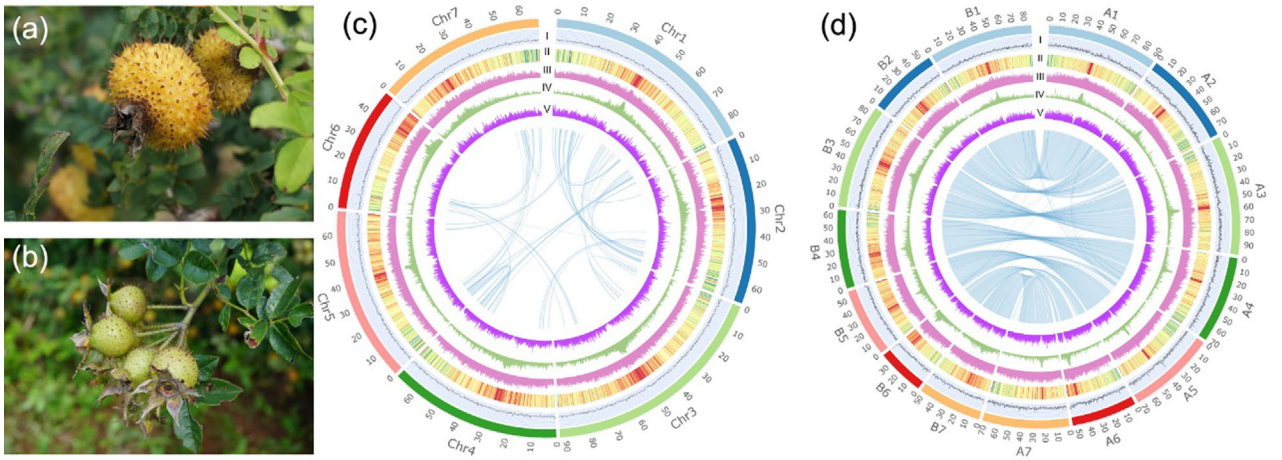


Figure 1. Overview of *Rosa roxburghii* and *R. sterilis* genome assembly.
(a) Image of *R. roxburghii* fruits.
(b) Image of *R. sterilis* fruits.
(c) The genome landscape of *R. roxburghii*.
(d) The genome landscape of *R. sterilis*, showing distributions of (I) GC percentage, (II) protein-coding genes, (III) repeat sequences, (IV) LTR-Gypsy, (V) LTR-Copia. Collinear blocks were shown in the center.

anti-apoptosis properties by reducing reactive oxygen species (ROS) and regulating genes that participate in programmed cell death, thus reducing the side-effects of radiotherapy (Xu et al., 2017). *R. roxburghii's* ripening fruit and leaves contain extremely high levels of triterpenes. The fruit of *R. roxburghii* can be consumed fresh or in various processed forms, including juice, dried products, canned items, yogurt, and even wine (Chen, Xu, et al., 2021).

Rosa sterilis ($2n = 2x = 14$), a new species identified in Guizhou Province in 1985, is believed to be a hybrid species closely related to *R. roxburghii* (Figure 1b) (Liang et al., 1989). *R. sterilis* was named "sterilis" owing to its characteristics of pollen abortion and thus production of seedless fruits. *R. sterilis* exhibits distinct variations from *R. roxburghii* in several characteristics including *R. sterilis* having a smaller fruit size and a relatively lower ascorbic acid (AsA) content. However, despite these differences, the fruits of *R. sterilis* share a comparable taste and are known for their high nutritional content, making them highly desirable in the food industry (Liu et al., 2016).

One of the most remarkable features of both *R. roxburghii* and *R. sterilis* is their exceptional abundance of AsA. *R. roxburghii*, in particular, stands out as the "king of fruits" renowned for its remarkably high ascorbate concentration, reaching up to 2585 mg/100 g fresh weight (FW). This concentration is five times higher than that found in kiwifruit, making it an excellent source for ascorbate production. Due to its impressive ascorbate content, *R. roxburghii* has gained significant recognition as an important commercial fruit plant (Figure 1a) (Li et al., 2022; Liu et al., 2016). Ascorbate, also known as vitamin C, is an essential metabolite in most living organisms. It is the most abundant soluble antioxidant in plants, playing an important role in antioxidant systems by detoxifying the free radicals generated by photosynthesis and UV from sunlight (Fenech et al., 2019; Smirnoff, 2018). It also functions as an enzyme cofactor, modulating a wide range of metabolic pathways, and as a signal molecule in the regulation of senescence and stress tolerance (Arrigoni & De Tullio, 2000; Barth et al., 2004; López-Carbonell et al., 2006; Noctor, 2006; Smirnoff & Wheeler, 2000). Many animals, including humans, lack the ability to synthesize ascorbate and need to obtain it from their diets, especially from plant sources (Smirnoff, 2018), while the concentration of ascorbate varies between species and within different tissues of the same plant. There are four well-defined biosynthesis pathways of AsA in plants: L-galactose, L-gulose, myo-inositol, and D-galacturonate (Fenech et al., 2019). All genes involved in the L-galactose pathway have been characterized and it is the predominant biosynthesis pathway in plants (Wang et al., 2023). Previous research has shown that *GMP*, *GME*, *GGP*, and some transcription factors including AP2, MYB, and HD-zip play key roles in the biosynthesis of AsA in plants (Liu et al., 2022). Further exploration is needed to better understand the genetic factors that account for variations in AsA content among different plant species.

Some ascorbate acid biosynthesis genes, including *GME*, *GGP*, *GPP*, *GDH*, *GLDH*, *GUR*, and *MIOX*, were identified through *de novo* transcriptome assemblies of *R. roxburghii*, and the subsequent quantitative reverse transcriptase-polymerase chain reaction of these genes showed increasing expression during *R. roxburghii* fruit

development, in accordance with its high ascorbate concentration (Yan et al., 2015). Furthermore, a series of genes including *DHAR*, *GME*, *GalLDH*, and *GGP* were experimentally demonstrated to have a positive correlation with the accumulation of AsA in *R. roxburghii* fruit (Huang et al., 2014; Liu et al., 2013; Yan et al., 2023). Although the mechanism of AsA production in *R. roxburghii* has been preliminarily explored, further genome-based research in this area could provide a more comprehensive and meaningful understanding.

Despite the availability of transcriptome data for *R. roxburghii* (Huang et al., 2019), the lack of a reference genome for *R. roxburghii* and *R. sterilis* hampers the characterization of comprehensive AsA biosynthesis and genome evolution. A previous genome survey of *R. roxburghii* revealed a genome size of 480.97 Mb and a transposable element (TE) content of 29.20% (Lu et al., 2016). Genetic research based on molecular markers suggests that *R. sterilis* may have originated from hybridization between *R. roxburghii* and *R. longicuspis* (Deng et al., 2015). Here, we conducted genome sequencing and assembly of both *Rosa* species at the chromosome level. We successfully separated two haplotypes of *R. sterilis* and our analysis offered molecular evidence for the hybrid origin of *R.sterilis*. Using RNA-seq analysis, we observed high expression of ascorbate synthesis genes in fruits of both species. Additionally, we identified several positively selected genes in *R. roxburghii*, which may explain its high ascorbate content. The genome data of these two *Rosa* species could be used to improve fruit quality and accelerate molecular breeding.

RESULTS

Genome assembly and annotation

Using flow cytometry, we estimated the genome size and ploidy level of the two *Rosa* species. Both species were diploid and the genome sizes were ~0.487G and ~0.522G for *R. roxburghii* and *R. sterilis*, respectively (Figure S1). Then, using MGISEQ-500 sequence data, a genomic survey for both species was performed and the estimated genome size for *R. roxburghii* was 497 Mb with a heterozygosity of 0.69% and that for *R. sterilis* was 522.3 Mb with a heterozygosity of 4.78%, which were consistent with the flow cytometry estimates (Table S1; Figure S2). In addition, *K*-mer analysis revealed that both species were diploid (Figure S3). These two genomes were assembled by a combination of three sequencing technologies: single-molecule real-time sequencing (Oxford Nanopore PromethION, Oxford, UK), paired-end sequencing (MGISEQ-500, Shenzhen, China), and Hi-C sequencing (Phase Genomics, Inc., Seattle, WA, USA) (Tables S2 and S3). The initial assembly of *R. roxburghii* had a genome size of 504 Mb and contig N50 of 30.2 Mb, while *R. sterilis* had a genome size of 981.2 Mb and a contig N50 of 27.4 Mb (Table S4).

By using Hi-C data, contigs from *R. roxburghii* were anchored onto seven pseudochromosomes with a scaffold N50 of 68.9 Mb, spanning 501 Mb covering 99.40% of the draft assembly (Figure 1c; Figure S4). Contigs from the *R. sterilis* assembly were anchored onto 14 pseudochromosomes with a scaffold N50 of 72.53 Mb, and a total size of 978.9 Mb, representing 99.77% of the contig genome size (Figure 1d; Table 1; Figure S5; Table S5). To assess the completeness of our genome assemblies, we aligned MGISEQ-500 data to our assemblies and found that 97.47 and 97.49% of the reads mapped to the *R. roxburghii* and *R. sterilis* assemblies, respectively (Table S6). Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis revealed that 2287 of the core eudicot genes (98.3%) were complete in *R. roxburghii*, while 2294 (98.6%) genes were complete in *R. sterilis*, with 1826 of them (78.5%) found in multiple copies (Table S7). Furthermore, we evaluated the quality of the genomes using the LTR assembly index (LAI), and obtained high LAI values of 18.32 and 24.89 for *R. roxburghii* and *R. sterilis*, respectively (Table S8), indicating high base accuracy and desirable continuity in our genome assemblies.

We annotated the repetitive sequences of both species with a combination of homology-based and *de novo* prediction methods. A total of 244.83 Mb (48.58%) and 470.45 Mb (47.95%) repetitive elements were identified, among which 209.45 Mb (41.56%) and 401.26 Mb (40.89%) belonged to long terminal repeat retrotransposons for *R. roxburghii* and *R. sterilis*, respectively (Figures S6 and S7; Table S9). The proportion of LTRs in these two species exceeded that found in other *Rosa* species, such as *R. chinensis* (28.3%), *R. multiflora* (19.3%), and *R. rugosa* (26.7%)

Table 1 Assembly and annotation for *Rosa roxburghii* and *Rosa sterilis* genome

Species	*R. roxburghii*	*R. sterilis*
Assembly		
Assembled genome size (bp)	504 018 164	981 156 797
GC content	38.73%	38.76%
Contig N50	30 235 923	27 374 109
Number of contigs	42	83
Scaffold N50	68 894 232	72 534 269
Complete BUSCOs	98.3%	98.6%
Gene annotation		
Repeat region	50.30%	49.9%
Number of protein-coding genes	40 020	79 545
Average gene length (bp)	2743	2667
Average exon length (bp)	242.37	250.06
Average intron length (bp)	433.32	441.51
Hi-C		
Anchored size	501 415 290	978 932 316
Anchored rate	99.46%	99.77%
Anchored genes	39 803	79 212
Number of chromosomes	7	14

(Chen, Su, et al., 2021; Hibrand Saint-Oyant et al., 2018; Nakamura et al., 2018). Using a combination of *de novo* prediction, homology search, and transcriptome-based methods, we annotated 40 020 and 79 545 protein-coding genes, including 94 and 96.5% of BUSCO genes, in *R. roxburghii* and *R. sterilis*, respectively (Table S10). For *R. roxburghii*, 96.06% of protein-coding genes could be assigned functions in at least one of the six databases (NR, TrEMBL, InterPro, KEGG, SwissProt, and GO), and the rate for *R. sterilis* was 94.64% (Table S11). In addition, we identified 94 microRNAs, 644 tRNAs, 496 small nuclear RNAs, and 703 ribosomal RNAs in the *R. roxburghii* assembly; in *R. sterilis*, there were 176 microRNAs, 1222 tRNAs, 956 small nucleus RNAs, and 1206 ribosomal RNAs (Table S12).

Phasing of *R. sterilis* haplotypes

The assembled genome size of *R. sterilis* was found to be nearly twice as large as the size estimated by genome survey and flow cytometry methods. This unexpected result, along with the high number of duplications observed in the BUSCO evaluation, suggests that the assembly has the potential to resolve haplotypes. In addition, we performed genome synteny analysis among *R. roxburghii*, *R. sterilis*, and *Fragaria vesca*, which belong to the same family of Rosaceae, which revealed high collinearity among these three species with a 1:2:1 syntenic depth and almost complete 1:1 synteny for *R. sterilis*' intragenomic (Figure 2a; Figures S8–S10). These results provide compelling evidence that we successfully assembled the two haplotypes of the *R. sterilis* genome. Given that *R. sterilis* has only seven chromosomes, our assembly of 14 pseudochromosomes represents its two distinct haplotypes. To accomplish this separation, we used SubPahser, a tool specifically designed for haplotype genome separation (Jia et al., 2022) (Figure S11). Based on the specific *K-mer* analysis, we successfully phased the genome into two distinct parts, as supported by both clustering and principal component analysis results (Figure S11a–c). The distribution of subgenome-specific *K-mers* and the phylogeny of LTRs (long terminal repeats) further provided evidence for the presence of two haplotype genomes (Figure S11d–f). Consequently, we assigned the longer and more complete homologous chromosomes to haplotype group A (RsterA), which contained pseudochromosomes A1–7, and the remaining chromosomes were categorized as haplotype group B (RsterB), which consisted of B1–7 (Table S5). The genome size of RsterA was 536.36 Mb with 47.92% of them were repeats and 46 158 genes in it. The haplotype RsterB had a genome length of 442.57 Mb with a higher repeat ratio of 52.07%, and a total of 33 054 genes were predicted. BUSCO analysis showed that 97.4 and 83.8% of core orthologs were identified in RsterA and RsterB, respectively. The LAI value for RsterA was 29.32, which was much larger than that for RsterB (21.22). Finally, we used the haplotype genome RsterA representing *R. sterilis* in the subsequent genomic analysis.

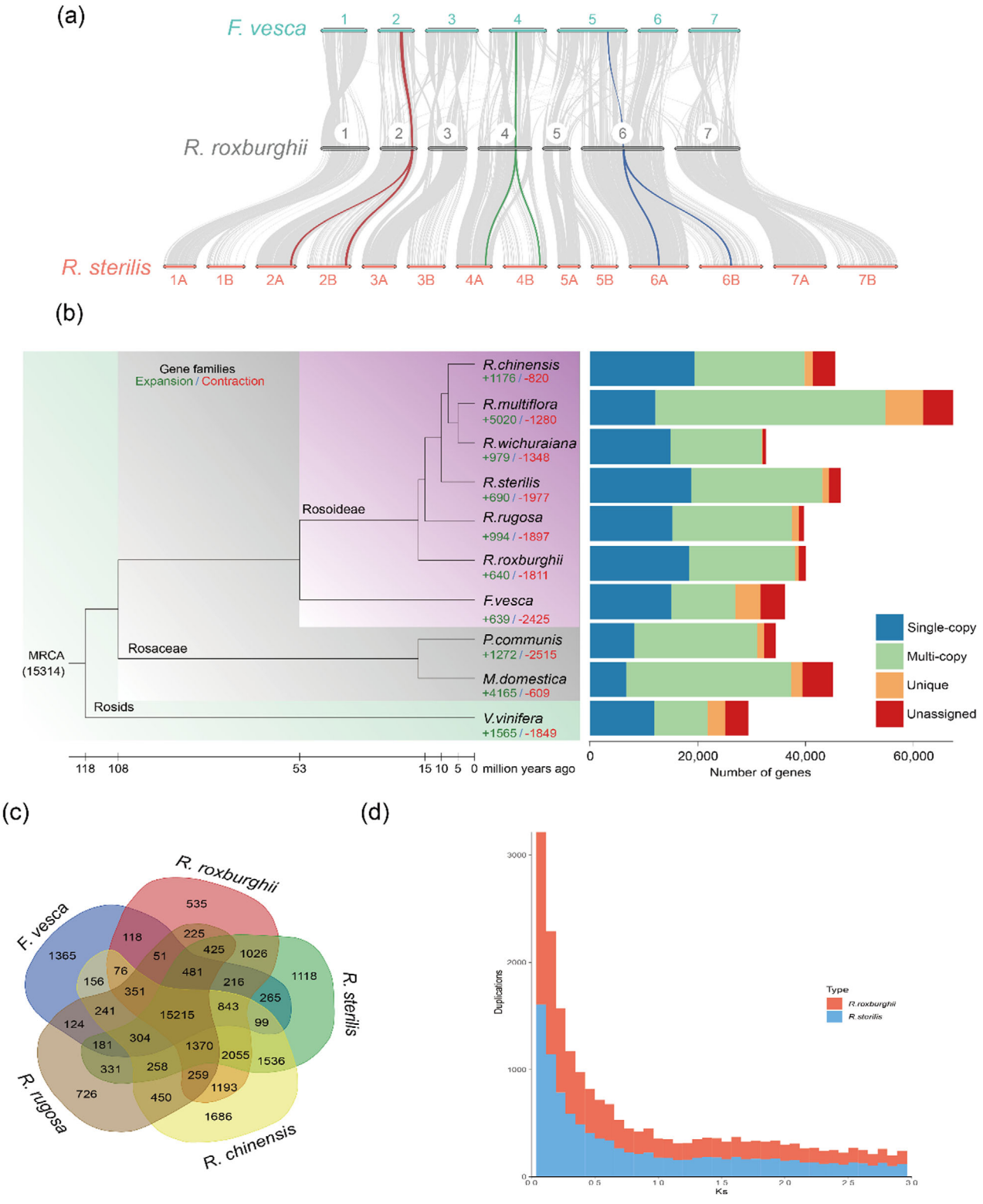
To identify the differences between RsterA and RsterB, we used SyRi to detect the structure variations between these two haplotypes (Goel et al., 2019) (Figure S12). Our analysis revealed a total of 212 687 SVs between these two haplotypes, including 303 inversions, 85 697 duplications, 82 043 inverted duplications, 22 630 translocations, and 22 014 inverted translocations. The median lengths for inversions, duplications, inverted duplications, translocations, and inverted translocations were 1815, 857, 837, 779, and 767 bp, respectively (Table S13). Among all the SVs, the mean length of inversions was found to be the longest, affecting 18 782 822 bp of RsterA and 13 508 094 bp of RsterB. This suggests that inversions play a significant role in the structural variation between the two haplotypes of *R. sterilis*.

Using the RNA data, we conducted an exploration of paralog distribution and expression in the two haplotype genomes of *R. sterilis*. By comparing the expression levels of homologous genes between the two haplotypes, we found no significant differences in the gene expression of paralogs (Figure S13). This suggests that there is no haplotype dominance within the *R. sterilis* genome.

Phylogenetics and genome evolution

To gain further insights into the evolutionary history of both species, we performed gene family clustering using *R. roxburghii*, *R. sterilis*, and eight other representative eudicot species, including seven from Rosaceae (*Rosa multiflora*, *R. chinensis*, *R. rugosa*, *R. wichuraiana*, *Fragaria vesca*, *Malus domestica*, *Pyrus communis*), and the outgroup eudicot species *Vitis vinifera* (Figure 2b). A total of 39 394 orthogroups encompassing 385 590 genes were found between these species. Within these orthogroups, 743 were single-copy orthologs, indicating their conserved nature across the species (Figure 2b; Table S14). Of these, 15 215 gene families were shared among five species (*R. roxburghii*, *R. sterilis*, *F. vesca*, *R. rugosa*, and *R. chinensis*) (Figure 2c). A total of 535 and 1116 families were

Figure 2. Genome evolution and phylogenetic analysis of *Rosa roxburghii* and *R. sterilis*.
(a) Syntenic blocks between genomic regions from *Fragaria vesca*, *R. roxburghii*, and *R. sterilis*.
(b) Phylogenetic tree of 10 eudicots species constructed with maximum likelihood method, based on single-copy ortholog genes. The distribution of genes in each species is shown in the right panel.
(c) Venn diagram of gene families in *R. roxburghii*, *R. sterilis*, and other Rosaceae species.
(d) Synonymous substitution rate (*Ks*) distribution among *R. roxburghii* and *R. sterilis*.



unique to *R. roxburghii* and *R. sterilis*, respectively. To elucidate the evolutionary relationships, a phylogenetic tree was constructed based on single-copy orthologs, confirming that *R. roxburghii* diverged earliest among all *Rosa* species, indicating its divergence from the common ancestor. *R. sterilis* formed a separate clade and exhibited

a sister group relationship with *R. chinensis*, *R. wichuraiana*, and *R. multiflora*. Furthermore, we calculated the divergence times of the species in the phylogenetic tree, and the results showed that *R. roxburghii* and *R. sterilis* diverged approximately 15 million years ago (Mya), and the divergence time between *R. chinensis* and *R. sterilis* was 9.52 Mya.

We proceeded to investigate the expansion and contraction of gene families in both *Rosa* species. Our analysis revealed the expansion of 640 gene families and the contraction of 1811 gene families in *R. roxburghii*. Kyoto Encyclopedia of Genes and Genome (KEGG) pathway enrichment analysis highlighted that these expanded gene families were associated with various metabolic pathways, including pyruvate metabolism, ubiquinone and other terpenoid-quinone biosynthesis, valine, leucine, and isoleucine biosynthesis, and phenylalanine, tyrosine and tryptophan biosynthesis (Table S15). These expanded gene families potentially play a role in the accumulation of ascorbate and flavone compounds in R. roxburghii fruits. In *R. sterilis*, we identified 690 expanded gene families and 1977 contracted gene families. The expanded genes were predominantly involved in metabolic pathways, such as monoterpenoid biosynthesis (Table S16).

To investigate the occurrence of whole-genome duplication (WGD) events in the genomes of *R. roxburghii* and *R. sterilis*, we calculated the synonymous substitutions (*Ks*) of paralogs for *R. roxburghii* and *R. sterilis* (Figure 2d). Analysing the distribution frequencies of *Ks* values, we observed that the two species exhibited a shared peak at approximately 1.5, which represents the ancient WGT events for the eudicots. However, we did not detect any significant recent WGD events during the evolution of *R. roxburghii* and *R. sterilis*, which is consistent with previous reports that no WGD events have occurred in other *Rosa* species (Chen, Su, et al., 2021; Chen, Xu, et al., 2021; Xiang et al., 2017). Furthermore, the observed 1:1:2 syntenic relationship among *F. vesca*, *R. roxburghii*, and *R. sterilis* provided additional evidence that there was no recent WGD in *Rosa* species (Figure 2a).

We carried out ancestral genome reconstruction of *R. roxburghii*, *R. sterilis*, *R. chinensis*, *R. rugosa* and *F. vesca*, based on orthologous and paralogours genes among these genomes. Based on the obtained reconstructed ancestral chromosomes, we deduced the possible evolutionary trajectories for these *Rosa* species. The results showed nonrecent fusion and fission among these Rosacea species (Figure S14).

Molecular evidence for the hybrid origin of *R. sterilis*

As it was previously reported that *R. sterilis* may have a hybrid origin with *R. roxburghii* and *R. longicuspis* as its parents (Deng et al., 2015), we were motivated to explore its origin using whole-genome data. Since a reference genome for *R. longicuspis* was not available, we assembled its RNA-seq data and successfully obtained 29 514 genes. These gene sets from RsterB and *R. longicuspis* were then added to the existing dataset comprising 10 other species for gene family identification. Using the 454 single-copy genes, we constructed a phylogenetic tree of the species, revealing that RsterA and *R. longicuspis* formed a sister group, while RsterB and *R. roxburghii*

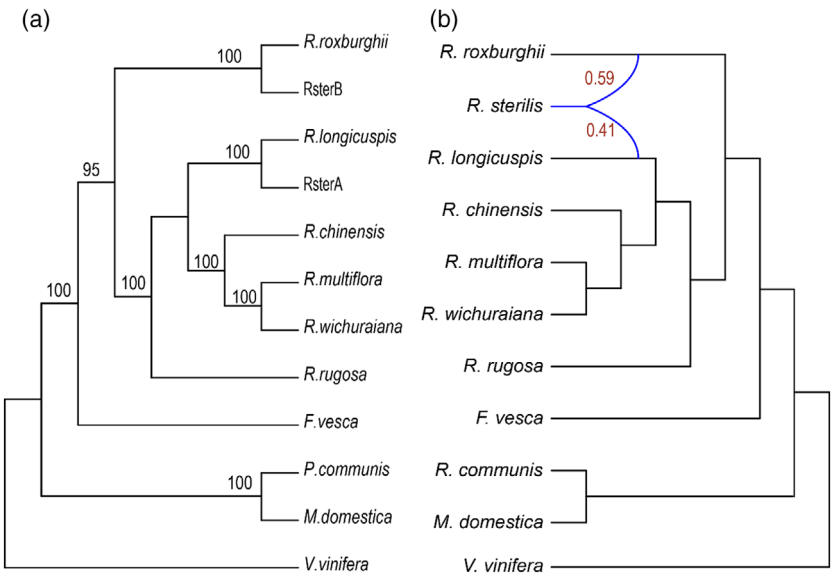


Figure 3. Phylogenomic analysis of *Rosa* species and the Hybridization inference of *R. sterilis*.
(a) Phylogenetic tree of haplotype genomes of *R. sterilis* constructed with maximum likelihood method, based on single-copy ortholog genes.
(b) Phylonetwork inferred by Phylonet. The proportion of the genome introgressed is shown on the hybrid branch.

constituted another smaller clade (Figure 3a). We further used PhyloNet to investigate the hybridization events in our phylogeny (Wen et al., 2018). Three network searches were carried out to identify reticulation events, and the network with only one hybridization event exhibited the highest likelihood value based on three information criteria. Utilizing this best species network, we identified a hybridization event between the clades of *R. longicuspis* and *R. roxburghii*, confirming the hybrid origin of *R. sterilis* (Figure 3b). It was determined that 41% of the *R. longicuspis* genome was transferred to *R. sterilis*, while that ratio for *R. roxburghii* was 59%. Furthermore, we added three more species (*R. persica*, *R. damascena*, and *R. lucieae*) for PhyloNet analysis and successfully detected a hybridization event between the clades of *R. longicuspis* and *R. roxburghii*, confirming the hybrid origin of *R. sterilis* (Figure S15). The analysis of sequence similarity supported these findings as well, with the most pronounced sequence similarity between RsterA and *R. longicuspis*, and RsterB with *R. roxburghii* (Figure S16). In addition, genome synteny analysis by SyRI revealed that 76.46% of the sequences exhibited collinearity between RsterB and *R. roxburghii*, which was three times higher than that observed between RsterA and *R. roxburghii* (23.92%). These findings support the notion that *R. roxburghii* and *R. longicuspis* are the most likely progenitors of *R. sterilis*. Additionally, the peak *Ks* value between *R. roxburghii* and RsterB was 0.006, which was 10 times higher than that between *R. longicuspis* and RsteA (0.0006). Meanwhile, the *Ks* peak value between RsterA and RsterB was 0.04, which is greater than that between RsterA and *R. roxburghii* (0.027) (Figure S17). Using the molecular clock rate calculated for eudicots (Ma & Bennetzen, 2004), which was estimated at 1.3×10^{-8} substitutions per synonymous site per year, we calculated the divergence times between *R. roxburghii* and RsterB, *R. roxburghii* and RsterA, *R. longicuspis* and RsteA, and the two-haplotype genomes of *R. sterilis* to 0.46, 1.04, 0.046, and 3.08 Mya, respectively. These results indicate that the divergence time between *R. roxburghii* and RsteA occurred earlier than that of RsteB, while the divergence time between *R. longicuspis* and RsterA was more recent than that of RsterB.

We employed MCMCtree to estimate the divergence time between *R. roxhurghii* and RsterB. The analysis suggested that the divergence between these two species occurred approximately 0.43–1.14 Mya. This estimated timeframe aligns with the inferences made from the *Ks* (synonymous substitution rate) values, further supporting the occurrence of hybridization events around that period.

Gene duplication contributes to flavonoid compound accumulation

Gene duplication is a common phenomenon in plants and plays a significant role in driving evolutionary innovation and functional diversification (Panchy et al., 2016). We used DupGen_finder (Qiao et al., 2019) and identified a total of 32 287 and 38 083 duplicated genes from *R. roxburghii* and *R. sterilis*, respectively (Table S17). These duplicated genes were classified into five categories based on their duplication mechanisms: tandem duplicates (TDs), WGD duplicates (WGDs), transposed duplicates (TRDs), dispersed duplicates (DSDs) and proximal duplicates (PDs). For *R. roxburghii*, DSD was the most abundant type, followed by PD, TRD, WGD, and TD (Figure 4a). This distribution may reflect the evolutionary history and genomic changes in *R. roxburghii*. Furthermore, KEGG enrichment analysis of these duplicated genes in *R. roxburghii* revealed their involvement in important biological processes, including plant–pathogen interactions and the synthesis of secondary metabolites, such as flavonoid biosynthesis, phenylpropanoid biosynthesis and glycosphingolipid biosynthesis (Figure 4b). Meanwhile, the duplicated genes had different distributions in their categories for *R. sterilis*, compared with *R. roxburghii*. DSD duplicates were the major type of tandemly duplicated genes, followed by dispersed duplicates and transposed duplicates. The two species had similar numbers of TD, TRD, and WGD genes. Notably, the duplicated genes in both species exhibited enrichment in important biological processes, such as plant-pathogen interaction, flavonoid biosynthesis, isoflavonoid biosynthesis, and monoterpenoid biosynthesis (Figure 4b). This suggests that gene duplication events have led to an increased copy number of genes related to secondary metabolites, particularly flavonoids and terpenoids. These findings indicate the potential significance of gene duplication in the evolution and adaptation of *R. sterilis*, particularly in the context of secondary metabolite production and plant defense mechanisms.

Ascorbate accumulation in fruits

Both *R. roxburghii* and *R. sterilis* have extremely high ascorbate concentrations in their fruits, ranging from 1300 to 3500 mg/100 g FW (He et al., 2016; Huang et al., 2014). To explore genes that might contribute to the high accumulation of AsA, we carried out comparative genome analysis and found evidence that two genes (*GGP*, *GalLDH*) in the AsA pathway were positively selected for *R. roxburghii* ($P < 0.05$ by Fisher's exact test). This indicates that these genes have experienced selection pressures specific to *R. roxburghii*. Additionally, we observed a noteworthy expansion of the *GalLDH* gene in *R. roxburghii*, resulting in a total of five gene copies, whereas other Rosaceae species typically possess one or two copies of this gene.

The iodine titration method was used to quantify the ascorbate content in fruits of different developing stages from June to October for both species. The results indicate that the ascorbate concentration in the fruits gradually increases over time, with the most significant increase

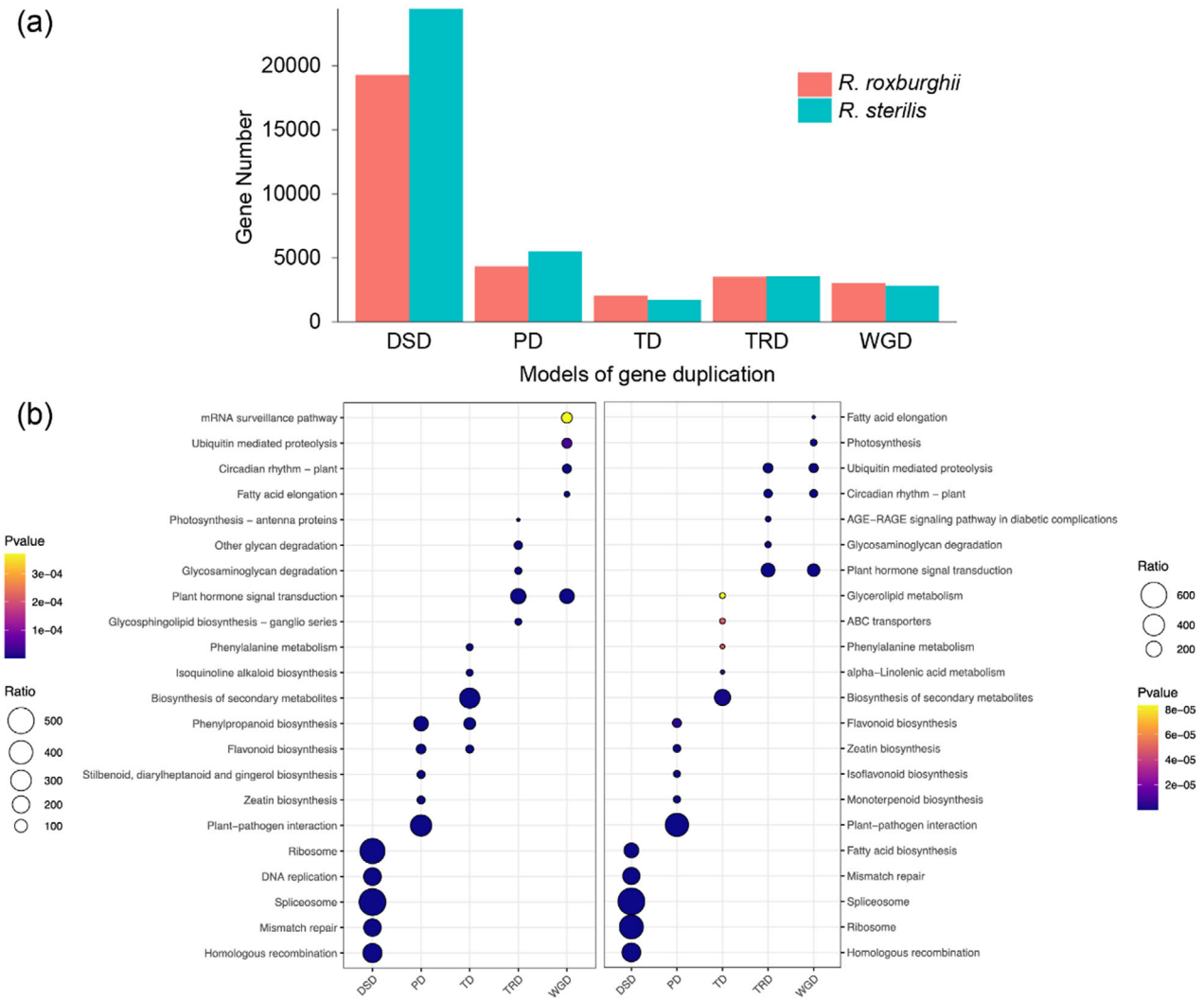


Figure 4. Classification and function of duplicated genes in *Rosa roxburghii* and *R. sterilis*.
(a) Number of duplicated genes types in *R. roxburghii* and *R. sterilis*.
(b) KEGG enrichment of duplicated genes types. Genes from *R. roxburghii* are on the left side, while *R. sterilis* genes are shown on the right side.

occurring from July to August (Figure 5a). Therefore, we performed RNA sequencing of fruits for both species, from June to October, to identify the key genes that participate in ascorbate biosynthesis (Table S18). Intriguingly, a series of genes in ascorbate biosynthesis pathways were highly expressed during the fruit maturation process for both species, including glucose-6-phosphateisomerase (*PGI*), mannose-6-phosphate isomerase (*MPI*), L-galactose dehydrogenase (*GDH*), and D-galacturonate reductase (*GalUR*). Interestingly, two key genes in the L-galactose pathway, GDP-L-galactose phosphorylase (*GGP*) and GDP-D-mannose 3′, 5′-epimerase (*GME*), were found to be highly expressed in *R. roxburghii* fruits, which was positively correlated with the high abscisic acid content in fruits (Figure 5b; Tables S19 and S20). Additionally, other genes in pathways related to ascorbate biosynthesis and metabolism, such as ascorbate oxidase and dehydroascorbate reductase (Broad et al., 2020; Smirnoff & Wheeler, 2000), were found to be actively expressed in both species. Some of these genes exhibited higher expression levels in *R. roxburghii* than in *R. sterilis*, indicating potential differences in ascorbate metabolism between the two species. These findings highlight the importance of core genes in the ascorbate biosynthesis pathway and their dynamic expression patterns during fruit development. The differential expression of these genes may contribute to the higher ascorbate levels observed in *R. roxburghii* fruits than in *R. sterilis* fruits.

To identify the genes that were strongly correlated with AsA accumulation, a weighted gene coexpression network analysis (WGCNA) was carried out. All the filter-expressed genes were assigned to 11 modules and three modules were significantly ($P < 0.05$) relevant to two

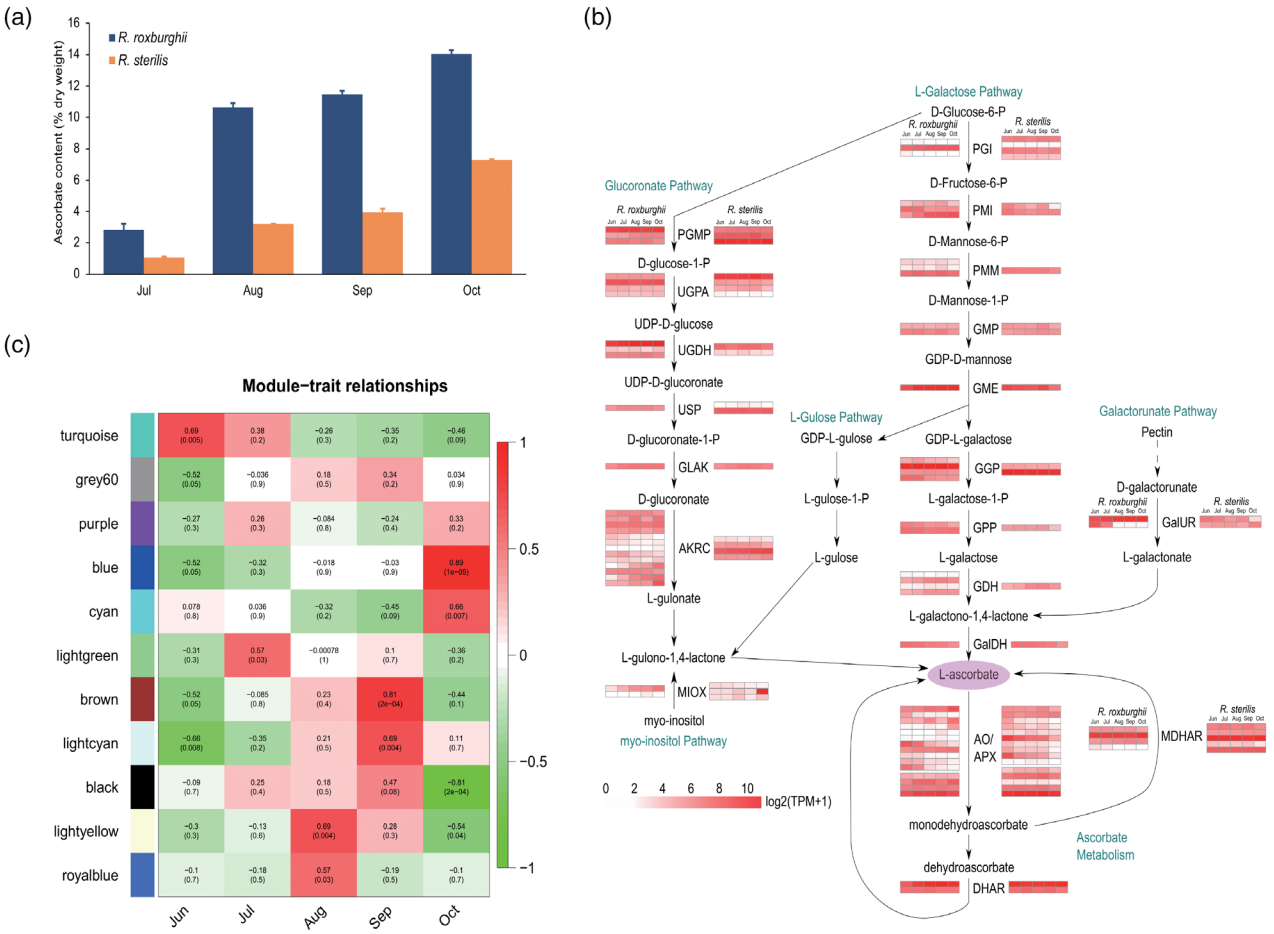


Figure 5. Expression of ascorbate synthesis and metabolism genes in *Rosa roxburghii* and *R. sterilis* fruits.
(a) Ascorbate content in *R. roxburghii* and *R. sterilis* fruits of different development stages, from July to October.
(b) Expression of genes involved in ascorbate biosynthesis and metabolism in *R. roxburghii* (left) and *R. sterilis* (right).
(c) Weighted gene coexpression network analysis of relationships between traits and modules. Different colors represent different modules.

developmental stages of the fruits, which were abundant in AsA accumulation (Figure 5c, light yellow 0.69, light cyan 0.69, brown 0.81). These modules comprised a total of 2002 genes, and KEGG pathway enrichment analysis revealed their involvement in metabolic pathways, biosynthesis of secondary metabolites, carbon metabolism, and oxidative phosphorylation (Table S21). In these three modules, we found that seven genes were specifically related to AsA biosynthesis and regeneration pathways, suggesting their potential importance in AsA accumulation (Table S22). To further investigate the key genes within these modules, we focused on identifying the hub genes. Hub genes are highly connected within a module and are considered crucial in regulating the biological processes associated with that module. Among those hub genes, we identified 24 transcription factors and these transcription factors were highly correlated with the accumulation of AsA ($r > 0.8$). Interestingly, there were two TFs, *MYB5* and bZIP, among these hub genes. The bZIP gene was reported to indirectly regulate *GGP* and promote AsA biosynthesis (Liu et al., 2023). In *Pyrus betulaefolia*, an experiment verified that MYB5 could promote the expression of *DHAR2* to activate AsA biosynthesis (Xing et al., 2019).

As *R. roxburghii* exhibited a higher ascorbate content than *R. sterillis*, we conducted WGCNA for *R. sterillis*. This systematic analysis revealed three distinct modules that exhibited statistical significance ($P < 0.05$) and strong relevance to two distinct developmental stages of the fruit (Figure S18). Remarkably, these modules encompassed a total of 2265 genes, and KEGG enrichment analysis revealed that these genes were related to ribosomes, terpenoid backbone biosynthesis, and ubiquitin-mediated proteolysis (Table S21). Furthermore, we performed an in-depth exploration of the hub genes within these three modules, uncovering those three specific genes – *APX*, *GalDH*, and *UGPA* – played pivotal roles in the AsA pathway. Intriguingly, *APX* gene was also found as the key hub gene in *R. roxburghii*. Among the hub genes, 92 were

identified as TFs with bZIP being one of the prominent TFs shared with *R. roxburghii*. This finding underscores potential regulatory elements contributing to AsA biosynthesis in both species.

Analysis of terpene synthase genes

As previously reported, the triterpenoid content in *R. roxburghii* fruits is relatively high (Li et al., 2022), and we performed a genome-wide analysis of terpene synthase (TPS) genes among different species, including four Rosaceae species (*R. roxburghii*, *R. sterilis*, *R. chinensis*, *F. vesca*), three other monocot and eudicot species (*A. thaliana*, *V. vinifera*, *Oryza sativa*), and three outgroup species (*Picea abies*, *Selaginella moellendorffii*, *Amborella trichopoda*). TPS-a is mainly involved in sesquiterpene synthase (sesTPS), while TPS-b and TPS-g mainly correspond to monoterpene synthase (monoTPS), and TPS-c and TPS-e/f mainly direct diterpene synthase (diTPS) (Shen et al., 2022). Using homologous comparison and phylogenetic research, we identified 47 and 112 TPS genes in the *R. roxburghii* and *R. sterilis* genomes, respectively. A vast majority of them were classified into the TPS-a subfamily. There were 71 TPS-a genes in the *R. sterilis* genome, representing a significant expansion (Figure 6a; Tables S23 and S24), while TPS-d and TPS-h genes were absent in both species and other *Rosa* species. We noticed that TPS-a was the dominant TPS genes among all the monocot and eudicot species used in our research, while no TPS-a genes were found in the gymnosperm species (*P. abies*, *S. moellendorffii*) and early angiosperm species (*A. trochopoda*). TPS genes were distributed in five clades and each subfamily of genes tended to cluster on a branch in the phylogenetic tree (Figure 6b). Analysis of chromosome distribution showed that many TPS genes are located in tandem arrays for both species (Figure S19), as previously reported in other species (Li et al., 2021; Wang, Liang, et al., 2022), suggesting that tandem duplication events likely participate in the evolution of these TPS genes.

We investigated the expression of TPS genes in different organs and developmental stages for both species (Tables S23 and S24). Although TPS-a represents the majority of TPS genes, most of its copies exhibited relatively low expression in roots and flowers (Figure 6c). TPS-g genes were actively expressed in leaves and flowers in both species, and some TPS-b and TPS-e genes exhibited relatively high expression. These genes may contribute to the accumulation of terpene metabolites in both species, especially in their fruits.

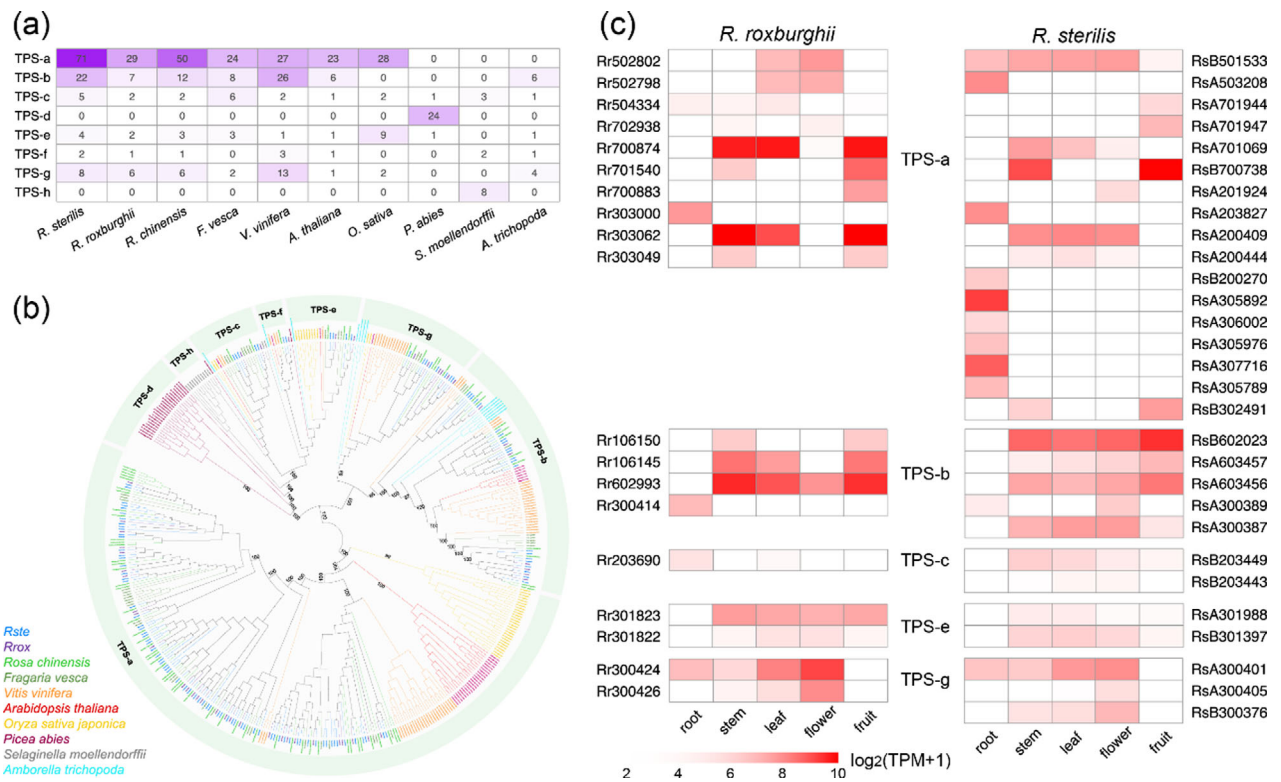


Figure 6. Expression and evolution of TPS genes in *Rosa roxburghii* and *R. sterilis*.
(a) Number of different types of TPS genes among 10 different species.
(b) Phylogenetic tree of TPS genes.
(c) Expression of TPS genes in root, stem, leaf, flower, and fruit of *R. roxburghii* and *R. sterilis*.

DISCUSSION

Rosa roxburghii and *R. sterilis* are renowned for their exceptional levels of ascorbate in fresh fruits, making them highly valued in the food industry. The first high-quality reference genomes of *R. roxburghii* and *R. sterilis* allowed us to gain insights into the evolutionary history of *Rosa* species and unravelled the underlying mechanisms responsible for the remarkable ascorbate accumulation in their fruits. Here, we generated high-quality chromosome-level reference genomes for these two species, with contig N50 values of 30.2 Mb and 27.4 Mb for *R. roxburghii* and *R. sterilis*, respectively. To our knowledge, these assemblies are of relatively high quality and continuity among available plant genome assemblies.

We further successfully separated the highly heterozygous genome into two haplotypes derived from different parents. Interestingly, we observed no significant biased expression or asymmetric gene loss between these two haplotypes of *R. sterilis*. This finding contrasts with what has been observed in certain allopolyploid species, where gene expression and gene loss can exhibit pronounced asymmetry between different subgenomes (Schnable et al., 2011; Shen et al., 2022). Our results suggest a balanced and coordinated gene regulation and retention mechanism within the hybrid genome of *R. sterilis*, contributing to its stability and functionality. This might indicate a rather young hybrid origin for *R. sterilis*. Through ancestral genome reconstruction of related Rosaceae species, we gained insights into the chromosomal structure and variation in *R. roxburghii* and *R. sterilis*. Surprisingly, we found that these two species exhibited a relatively conserved chromosome structure, with minimal rearrangements observed throughout the evolutionary history of the *Rosa* genus. This finding suggests marked stability in the chromosomal architecture of *Rosa* species and implies the conservation of essential genomic regions and functional elements.

Published studies on homoploid hybrid species have provided valuable insights into genetic mixing and variation, adaptive traits, speciation, ecological niches, and reproductive isolation (Chase et al., 2010; Moran et al., 2021; Wang et al., 2021; Wang, Kang, et al., 2022). However, there are still lingering questions regarding the macroevolutionary significance of the hybrid process, particularly concerning its influence on evolutionary trajectories. Additionally, there is a pressing need to investigate how life history and ecological factors either shape or have historically shaped the frequency of hybridization across geological epochs and within various plant phylogenies (Stull et al., 2023). Previous research analysing chloroplast and two nuclear gene fragments indicated that *R. sterilis* originated from the natural hybridization of *R. longicuspis* and *R. roxburghii*. This result might be limited by the use of partial molecular data, which may compromise the accuracy of the findings. To gain a more comprehensive understanding of the origin of *R. sterilis*, we utilized whole-genome data, which provides a more extensive and detailed view of the species' evolutionary history. We constructed a phylogenetic tree based on single-copy genes from all published *Rosa* species, and together with sequence similarity distribution analysis, both results revealed that *R. sterilis* haplotype B is more closely related to *R. roxburghii*, while haplotype A diverged from a common ancestor that is closely related to *R. longicuspis*. The PhyloNet genome analysis results further supported that *R. sterilis* originated as a natural hybrid, with over 50% genetic material inherited from *R. longicuspis* and the rest from *R. roxburghii*. Moreover, based on the *Ks* value and divergence time, we infer that the hybridization events may have occurred approximately 1 Mya. The recently arising hybrid species *R. sterilis*, along with the availability of its genome and the genomes of its parent species, has provided a unique opportunity to investigate *Rosa* species hybridization at the whole-genome level. This comprehensive approach has significantly enhanced our understanding of the *Rosa* genus and has offered valuable resources for further research on hybridization. Future studies on the population and reproductive biology of these three species are expected to provide insights into hybrid fertility, along with a deeper understanding of the genetic and ecological factors that contribute to hybrid speciation.

The fruits of *R. roxburghii* and *R. sterilis* are rich in flavonoid compounds, which act as excellent antioxidants, preventing ROS-mediated DNA damage and cell apoptosis (Xu et al., 2017, 2020; Yuan et al., 2020). Therefore, we investigated the genes that promote the accumulation of flavonoids. We found that many duplicated genes were related to the pathway of flavonoid biosynthesis pathway, and duplicated genes in both the PD and TD categories were enriched in flavonoid biosynthesis in *R. roxburghii* and *R. sterilis*. These genes were greatly expanded in the two species, suggesting that they contribute to advanced flavonoid accumulation in fruits.

Ascorbate is the most abundant water-soluble antioxidant in plants, and functions as an important coenzyme in many enzymatic reactions. In this research, we detected and confirmed that both species have an impressively high concentration of ascorbate in their fruits, and *R. roxburghii* fruits have an even higher amount of ascorbate than those of *R. sterilis*. Previous studies reported that *GGP* is the key modulator of this pathway (Bulley et al., 2009, 2012; Fenech et al., 2019), and *GME* collaborates with *GGP* as an important regulator (Bulley et al., 2009; Tao et al., 2018). RNA-seq results revealed that many genes showed an increasing trend according to ascorbate accumulation in fruits. Moreover, some important genes involved in ascorbate biosynthesis and metabolism, especially *GGP* and

GME, are actively expressed in fruits in different developmental stages for both species. The expression of *GGP* and *GME* in *R. roxburghii* fruits was higher than that in *R. sterilis* fruits and is increased during fruit development, in accordance with the ascorbate concentrations in their fruits, suggesting that the active expression of certain genes in the ascorbate biosynthesis pathway is related to high ascorbate concentration in their fruits. Based on WGCNA, three modules were identified as having a high correlation with AsA accumulation. Some of the transcription factors and AsA synthesis genes belong to the hub genes in these modules, and two positively selected genes also indicated the special AsA for these two *Rosa* species. We observed a slight expansion of the *GalLDH* gene in *R. roxburghii*. This expansion of *GalLDH* has potential functional significance and suggests a unique evolutionary trajectory for this particular gene in *R. roxburghii* compared to other related species. These results should be further verified through experiments.

Terpenoids, a large and diverse family of chemical compounds in plants, play crucial roles in growth, development, and stress responses (Chen et al., 2011; Gershenzon & Dudareva, 2007; Tholl, 2015; Yamaguchi, 2008). In our research, we identified 47 and 112 TPS genes in the *R. roxburghii* and *R. sterilis* genomes, respectively, mainly classified into the TPS-a subfamily, which represents sesquiterpene synthase (Chen et al., 2011). Tandem duplication events appear to have played a significant role in the evolution of TPS genes. Some TPS genes, including those representing kaurene synthase and acyclic terpenes, were highly expressed in fruits, indicating their potential involvement in terpenoid biosynthesis. Further investigation is needed to determine their specific molecular functions.

In summary, our study represents a significant advancement in the understanding of *R. roxburghii* and *R. sterilis*. We successfully generated high-quality chromosome-level genomes for both species, providing valuable resources for future research in the field. Our analysis confirmed the hybrid nature of *R. sterilis* and separated its genome into two distinct haplotypes. Based on the RNA-seq data and genome data, we shed light on its special ascorbate accumulation mechanism and provided insights into genes that participate in flavonoid biosynthesis. These findings contribute to the broader knowledge of *Rosa* species evolution and differentiation, while also providing valuable insights for studying the functional genomics of high ascorbate production in plants.

METHODS

Flow cytometry and polyploidy estimation

Young leaves were placed in 0.8 ml of pre-cooled MG dissociation solution, and the tissue was quickly chopped vertically with a sharp blade. Then, was filtered through a 40 μm pore size filter. The tomato was used as an internal reference, and the suspension sample and reference sample were mixed in an appropriate ratio. BD FACScalibur flow cytometer was used and the stained cell nuclei suspension samples were put on the machine to detect the fluorescence intensity emitted by propidium iodide using 488 nm blue light excitation and 10 000 particles were collected for each assay. The coefficient of variation CV% was controlled within 5%. Plotting and analysis were performed using Modifit 3.0 analysis software. Ploidy was tested using the CyFlow Ploidy Analyzer ploidy analyzer from Partec, and known diploid plants were used as controls.

Genome survey and sequencing

Fresh young leaves from mature plants of *R. roxburghii* and *R. sterilis* were collected from Southwestern Forest University, Yunnan Province, China. Genomic DNA was extracted for MGISEQ and Oxford Nanopore sequencing. For MGISEQ sequencing, 300 bp short-read pair-end libraries were constructed and sequenced on the MGISEQ platform. 122 Gb of clean reads were obtained for *R. roxburghii*, and 132 Gb for *R. sterilis*. Genome size, heterozygosity, and repeat content were calculated with k-mer method using GCE (ftp://ftp.genomics.org.cn/pub/gce) and Genomescope (Vurture et al., 2017). For Nanopore sequencing, genomic DNA was fragmented to ~20 kb to construct long-read libraries as previously published, and the libraries were sequenced on the Oxford Nanopore platform. 61.94 Gb of clean subreads were obtained for *R. roxburghii* with N50 value of 33.05 kb, and for *R. sterilis* we obtained 45.43 Gb clean subreads with a 33.24 kb N50 value.

Genome assembly

We used NextDenovo (https://github.com/nextomics) to assemble the genomes. The draft assemblies were polished by Racon (Vaser et al., 2017) for three times, followed by Pilon (Walker et al., 2014) to correct the potential errors with MGISEQ clean reads for two times. The final *R. roxburghii* genome size was 504.02 Mb, with an contig N50 value of 30.24 Mb. *R. sterilis* assembly has a genome size of 981.16 Mb, with 27.37 Mb length of contig N50 size. To access the quality of the assemblies, we use bwa-mem tool of BWA (https://github.com/lh3/bwa) to map the short pair-end reads to the assemblies, and BUSCO analysis was performed with embryophyte_odb9 database (Manni et al., 2021).

For chromosome-level assembly, we used fresh young leaves of both species to construct Hi-C libraries and sequence them on MGISEQ platform. MboI was used to digest the chromatin, 146.60 Gb and 127.37 Gb clean reads were obtained for *R. roxburghii* and *R. sterilis*, respectively. The reads were anchored to scaffolds with Juicer (Durand et al., 2016) and 3D-DNA pipeline (Dudchenko et al., 2017). The interactions were visualized as heat maps with Juicebox.

Genome annotation

We first used a combination of homology comparison and de novo prediction to create repetitive libraries for both species. LTR Finder (Xu & Wang, 2007), RepeatScout (Price et al., 2005), and RepeatModeler (Flynn et al., 2020) were used to search repetitive sequences in genomes, and we used TRF and RepeatMasker to do homology-based prediction with Repbase TE library. The repetitive sequences were further annotated, classified, and mask the genomes by RepeatMasker. Meanwhile, LTR Retriever, together with LTR Finder, was used to calculate LAI (LTR Assembly Index)

which represents assembly continuity by evaluating the assembly of repeat sequences (Ou et al., 2018).

For the transcriptome-assisted predictions, RNA sequencing was done with different organs for both species. We used Trinity (Grabherr et al., 2011) for transcriptome assembly based on their assembled genomes. Meanwhile, we used Trinity to assemble the RNA-seq of *longicuspis* and predicted its genes by using TransDecoder (https://github.com/TransDecoder/TransDecoder). In homologous predictions, protein sequences from *R. rugosa*, *R. wichuraiana*, *A. thaliana*, *R. chinensis*, and *F. vesca* were used. MAKER pipeline (Campbell et al., 2014) was used for protein-coding genes annotation, and de novo gene models were generated by AUGUSTUS (Stanke, et al., 2008) and SNAP (Korf, 2004). We used BUSCO analysis to check the completeness of the protein-coding genes and further assigned them to BLAST against NR, InterPro, SwisssProt, TrEMBL, and KOG databases. Blast2GO and KofamKOALA (Aramaki et al., 2020) were used to do function and pathway analysis based on GO and KEGG databases, and KEGG annotation results (lipid metabolism, energy metabolism, pigment synthesis, pathogen defending) were used for Circos maps.

Structure variation analysis

A series of tools from MUMmer4 (Marcais et al., 2018) was used to analyse the differences between *R. roxburghii* genome and the two haplotypes of *R. sterilis*. We used nucmer to perform a comparison between syntenic chromosomes, and delta-filter was used to filter the alignment blocks with parameter: –m –i 90 – l 100. Then show-coords programs were used to convert them into tab-delimeted files. Finally, SyRI was employed to identify those SVs including: inversions, duplications, translocations, inverted translocations, and inverted duplications (Goel et al., 2019).

Ancestral genome reconstruction

The orthologous and paralogous genes among *R. roxburghii*, *R. sterilis*, *R. chinensis*, *R. rugosa*, *F. vesca*, and *P. persica* (as outgroup species) were identified with BlastP, based on a cumulative identity percentage of 60% and a cumulative alignment length percentage of 70%, as described in previous research (Murat et al., 2015). Syntenic blocks among species were identified and filtered with MCScanX (Python version) (Tang et al., 2008), with a parameter of minsize = 5 minspan = 20, and filtered blocks were further used for ancestral genome reconstruction via MLGO (Hu et al., 2014). Figures of chromosome structures were carried out with TBtools (Chen et al., 2020).

Comparative genome analysis

Eleven species, including *R. roxburghii* and two haplotypes of *R. sterilis*, were selected for the phylogenetic analysis. Orthologous groups were identified with OrthoFinder2 (Emms & Kelly, 2019). The protein sequences from single-copy genes were aligned with MUSCLE (Edgar, 2004), and the poorly aligned regions were removed by Gblocks (Talavera & Castresana, 2007). We used IQtree2 (Nguyen et al., 2015) to select the best model (JTT+F+R3) and constructed a maximum likelihood phylogenetic tree. The ML tree was further used to calculate divergence time with the PAML MCMCTree program (Yang, 2007) with GTR model. The estimated divergence time of *M. domestica*–*P. communis* (12–20 Mya) and *M. domestica*–*V. vinifera* (110–124 Mya) in TimeTree (Kumar et al., 2017) was used as calibrators. Café (De Bie et al., 2006). was used to identify the gene family expansions and contractions. The phylogeny tree was visualized by iTOL (https://itol.embl.de/).

MCScanX was used for syntenic comparison between chromosomes and different species. We use LASTZ to do all-to-all protein comparison, and the results were used by MCScanX to identify the syntenic blocks. The homologous gene pairs from syntenic blocks were aligned with MUSCLE, and the *Ks* values were calculated by ParaAT (Zhang et al., 2012) and *KaKs*_calculator. Images of the syntenic blocks and other features were produced with Circos (Krzywinski et al., 2009).

Ascorbate quantification

Rosa roxburghii and *R. sterilis* fresh fruits from June, July, August, September, and October of 2020 were oven-dried and ground into powder, then ~0.2 g powder was used for ascorbate quantification using a titration method. Iodine standard solution was used to oxidize ascorbate in the titration test, and 10 g L^{-1} starch solution was used as an indicator. Blank tests were also executed for standardization. At least six fresh fruits from top branches were selected as samples, and three biological repeats were set for each group.

RNA-seq and analysis and genes co-expression network analysis

Root, stem, leaf, flower tissues, and fruits from different development stages, were collected from *R. roxburghii* and *R. sterilis* plants. Three biological repeats were executed for each sample. cDNA was synthesized from the RNA samples extracted from these tissues using QIAGEN® Genomic kits. 200 bp/400 bp libraries were constructed for MGISEQ platform sequencing, and 100 bp pair-end reads were generated. FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) and SOAPnuke (Chen et al., 2018) were used for data filtering, and bowtie2 (Langmead & Salzberg, 2012) to align the reads to the annotated genes. Next, we use RSEM (Li & Dewey, 2011) to calculate the TPM and differential expression of genes in different tissues.

Co-expression networks were carried out using WGCNA (v1.47) package in R (Langfelder & Horvath, 2008). Gene expression were fed into WGCNA to construct co-expression modules with default settings, and eigengenes in these modules were calculated. The co-expression relationship between two genes was measured by a weighted value. Genes that were most correlated (correlation coefficients >0.8) with each module and development (correlation coefficients >0.2) were defined as hub genes.

Gene family analysis

Sequences of TPS genes were accessed from previous research (Zhang et al., 2020), And two specific TPS domains, PF01397 representing the N-terminal and PF03936 representing the C-terminal domain from the Pfam database (http://pfam.xfam.org/), were used to predict the TPS genes with Hidden Markov Model (HMM) search tool (version 3.1) (*E*-value < 1e-5).

Gene predictions were further checked with functional annotation with NR and Swissprot database. MAFFT (Rozewicki et al., 2019) and MUSCLE were used to generate multi-sequence alignments, and maximum likelihood phylogenetic trees were constructed by IQTree2 and MEGA X (Kumar et al., 2018). Phylogenetic trees were visualized with iTOL.

AUTHOR CONTRIBUTIONS

CH, YL, and DZ conceived the project and were responsible for the project initiation. DZ, CH, and LL designed the experiments. PG, PL, QY, DZ, JYang, and XQ contributed

to sample preparation. DZ, XG, HLiang, HLiu, PG, SM, JYu, TJ, and QY performed the experiments and the data analysis. DZ, PG, XG, and CH wrote the original draft. LL, WS, and SKS, revised the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

This work was supported by Yunnan Provincial Expert Workstation (No. 202005AF150020), Yunnan Provincial "Ten-Thousand Program" for Leading Industry Innovation (No. YUWR-CYJS-2020-018), Special Funds for Local Scientific and Technological Development Guided by the Central Government (No. YDZX201953000002845). We thank Dr. Jing Pan for help on genome analysis. We thank the Guangdong Provincial Academician Workstation of BGI Synthetic Genomics, BGI-Shenzhen, Guangdong, China. Finally, we are thankful to the China National GeneBank for producing sequencing data.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The whole-genome sequencing data and the two chromosome-level genome assemblies have been deposited to the CNSA (CNGB Nucleotide Sequence Archive) with accession CNP0004212 (https://db.cngb.org/cnsa/).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Flow cytometry analysis of the genome size and ploidy of the *R. roxburghii* and *R. sterilis*. (a) Histogram of relative fluorescence intensity for *R. roxburghii*, (b) histogram of relative fluorescence intensity for *R. sterilis*, (c) ploidy distribution of *R. sterilis*, (d) ploidy distribution of *R. roxburghii*.

Figure S2. K-mer distribution of the *R. roxburghii* (a) and *R. sterilis* (b) genome. K-mer = 17.

Figure S3. Smudgeplots for *R. roxburghii* (a) and *R. sterilis* (b) genome.

Figure S4. Heatmap for Hi-C assembly of *R. roxburghii* genome.

Figure S5. Heatmap for Hi-C assembly of *R. sterilis* genome.

Figure S6. The distribution of transposable elements in *R. roxburghii* genome, found by de novo methods (a) and homologous comparison (b).

Figure S7. The distribution of transposable elements in *R. sterilis* genome, found by de novo methods (a) and homologous comparison (b).

Figure S8. Homologous dot-plot between *R. roxburghii* and *R. sterilis* genomes.

Figure S9. Homologous dot-plot between *R. roxburghii* and *F. vesca* genomes.

Figure S10. Homologous dot-plot in *R. sterilis* intra genomes.

Figure S11. Haplotype genome phasing and characterization of *R. sterils*. (a) Number of differential 15-mers in the chromosomes. (b) Unsupervised hierarchical clustering. (c) Principal component analysis of 15mers demonstrated that the genome is successfully phased into two haplotype genomes. (d) Chromosomal characteristics. (e) Insertion time of the two haplotype genomes. (f) Phylogenetic tree of 1000 Gypsy LTR-RTs randomly selected from the subgenome-specific LTR-RTs.

Figure S12. Structural variations between RsterA and RsterB.

Figure S13. Characteristics of syntenic genes between two haplotypes of *R. sterilis*. (a) Number of DEGS between RsterA and RsterB in different tissues. (b) Expression of syntenic genes between RsterA and RsterB in different tissues.

Figure S14. Ancestral genome reconstruction of five Rosaceae species.

Figure S15. Phylogenomic analysis of *Rosa* species and the Hybridization inference of *R. sterilis*. (a). Phylogenetic tree of haplotype genomes of *R. sterilis* constructed with maximum likelihood method, based on single-copy ortholog genes. (b) Phylonetwork inferred by Phylonet.

Figure S16. The sequence similarity between the *Rosa* genus species and the two haplotype genesets of R. sterilis. (a) The distribution of sequence similarity between *Rosa* species and RsterA. (b) The distribution of sequence similarity between *Rosa* species and RsterB.

Figure S17. Distribution of synonymous substitution levels (*Ks*) of syntenic orthologous.

Figure S18. WGCNA analysis of relationships between traits and modules. Different colors represent different modules.

Figure S19. Distribution of TPS genes on chromosomes of *R. roxburghii* (a) and *R. sterilis* (b) genome.

Table S1. *K*-mer statistics of *R. roxburghii* and *R. sterilis* genomes by genome survey analysis.

Table S2. Summary of *R. roxburghii* and *R. sterilis* genome sequencing data.

Table S3. Statistics of Oxford Nanopore sequencing data for *R. roxburghii* and *R. sterilis*.

Table S4. The summary of preliminary assmbly of *R. roxburghii* and *R. sterilis* genome.

Table S5. The assembled length and gene number of *R. roxburghii* and *R. sterilis* based on Hi-C assembly.

Table S6. Re-mapping of short-reads data to *R. roxburghii* and *R. sterilis* genome assembly.

Table S7. Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis of *R. roxburghii* and *R. sterilis* haplotype genome.

Table S8. LTR Assembly Index (LAI) value of *R. roxburghii* and *R. sterilis* genome assembly.

Table S9. Statistics of repetitive sequences in *R. roxburghii* and *R. sterilis* and its two haplotypes genome.

Table S10. Statistics of protein-coding gene annotation in *R. roxburghii* and *R. sterilis* haplotype genomes.

Table S11. Statistics for gene function annotation of *R. roxburghii* and *R. sterilis*.

Table S12. Statistics of non-coding RNA sequences in *R. roxburghii* and *R. sterilis* genome.

Table S13. Summary of structural variations between RsterA and RsterB.

Table S14. Statistics of gene family clustering in 10 species with OrthoFinder.

Table S15. KEGG enrichment analysis of the expanded gene families in *R. roxburghii* genome.

Table S16. KEGG enrichment analysis of the expanded gene families in *R. sterilis*.

Table S17. Statistics of different types of duplicated genes.

Table S18. Statistics of RNA-seq data.

Table S19. Expression of ascorbate synthesis and metabolism-related genes in *R. roxburghii* fruits (TPM).

Table S20. Expression of ascorbate synthesis and metabolism-related genes in *R. sterilis* fruits (TPM).

Table S21. KEGG enrichment analysis of genes in three modules for *R. roxburghii* and R. sterilis fruits.

Table S22. Expression of ascorbate synthesis and metabolism-related hub genes in *R. roxburghii* fruits (TPM).

Table S23. Classification and expression of TPS genes in *R. roxburghii* (TPM).

Table S24. Classification and expression of TPS genes in *R. sterilis* (TPM).

REFERENCES

Aramaki, T., **Blanc-Mathieu, R.**, **Endo, H.**, **Ohkubo, K.**, **Kanehisa, M.**, **Goto, S.** ***et al.*** (2020) KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, **36**, 2251–2252.

Arrigoni, O. & **De Tullio, M.C.** (2000) The role of ascorbic acid in cell metabolism: between gene-directed functions and unpredictable chemical reactions. *Journal of Plant Physiology*, **157**, 481–488.

Barth, C., **Moeder, W.**, **Klessig, D.F.** & **Conklin, P.L.** (2004) The timing of senescence and response to pathogens is altered in the ascorbate-deficient Arabidopsis mutant vitamin c-1. *Plant Physiology*, **134**, 1784–1792.

Broad, R.C., **Bonneau, J.P.**, **Hellens, R.P.** & **Johnson, A.A.T.** (2020) Manipulation of ascorbate biosynthetic, recycling, and regulatory pathways for improved abiotic stress tolerance in plants. *International Journal of Molecular Sciences*, **21**, 1790.

Bruna, T., **Hoff, K.J.**, **Lomsadze, A.**, **Stanke, M.** & **Borodovsky, M.** (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, **3**, lqaa108.

Bulley, S., **Wright, M.**, **Rommens, C.**, **Yan, H.**, **Rassam, M.**, **Lin-Wang, K.** ***et al.*** (2012) Enhancing ascorbate in fruits and tubers through over-expression of the L-galactose pathway gene GDP-L-galactose phosphorylase. *Plant Biotechnology Journal*, **10**, 390–397.

Bulley, S.M., **Rassam, M.**, **Hoser, D.**, **Otto, W.**, **Schunemann, N.**, **Wright, M.** ***et al.*** (2009) Gene expression studies in kiwifruit and gene over-expression in Arabidopsis indicates that GDP-L-galactose guanyltransferase is a major control point of vitamin C biosynthesis. *Journal of Experimental Botany*, **60**, 765–778.

Campbell, M.S., **Holt, C.**, **Moore, B.** & **Yandell, M.** (2014) Genome annotation and curation using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, **48**, 4.11.1–4.11-39.

Chase, M.W., **Paun, O.** & **Fay, M.F.** (2010) Hybridization and speciation in angiosperms: a role for pollinator shifts? *BMC Biology*, **8**, 45.

Chen, C., **Chen, H.**, **Zhang, Y.**, **Thomas, H.R.**, **Frank, M.H.**, **He, Y.** ***et al.*** (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, **13**, 1194–1202.

Chen, F., **Su, L.**, **Hu, S.**, **Xue, J.Y.**, **Liu, H.**, **Liu, G.** ***et al.*** (2021) A chromosome-level genome assembly of rugged rose (*Rosa rugosa*) provides insights into its evolution, ecology, and floral characteristics. *Horticulture Research*, **8**, 141.

Chen, F., **Tholl, D.**, **Bohlmann, J.** & **Pichersky, E.** (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal*, **66**, 212–229.

Chen, G. & **Kan, J.** (2018) Ultrasound-assisted extraction, characterization, and antioxidant activity in vitro and in vivo of polysaccharides from chestnut rose (*Rosa roxburghii* Tratt) fruit. *Journal of Food Science and Technology*, **55**, 1083–1092.

Chen, Y., **Chen, Y.**, **Shi, C.**, **Huang, Z.**, **Zhang, Y.**, **Li, S.** ***et al.*** (2018) SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience*, **7**, 1–6.

Chen, Y., **Xu, Y.**, **Lin, X.**, **Yu, Y.**, **Wen, J.** & **Fu, M.** (2021) Comparisons of nutritional and bioactive components of *Rosa roxburghii* fruits of different maturities, altitudes and regions. *Journal of Food Safety and Quality*, **12**, 9556–9564.

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**, 1269–1271.

Deng, H., **Gao, X.**, **Li, X.** & **Zhou, H.** (2015) Molecular evidence for hybridization origin of *Rosa* × *sterilis* (Rosaceae). *Journal of Plant Resources and Environment*, **24**, 10–17.

Dudchenko, O., **Batra, S.S.**, **Omer, A.D.**, **Nyquist, S.K.**, **Hoeger, M.**, **Durand, N.C.** ***et al.*** (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**, 92–95.

Durand, N.C., **Robinson, J.T.**, **Shamim, M.S.**, **Machol, I.**, **Mesirov, J.P.**, **Lander, E.S.** ***et al.*** (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems*, **3**, 99–101.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.

Emms, D.M. & **Kelly, S.** (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**, 238.

Fenech, M., **Amaya, I.**, **Valpuesta, V.** & **Botella, M.A.** (2019) Vitamin C content in fruits: biosynthesis and regulation. *Frontiers in Plant Science*, **9**, 2006.

Flynn, J.M., **Hubley, R.**, **Goubert, C.**, **Rosen, J.**, **Clark, A.G.**, **Feschotte, C.** ***et al.*** (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 9451–9457.

Gershenzon, J. & **Dudareva, N.** (2007) The function of terpene natural products in the natural world. *Nature Chemical Biology*, **3**, 408–414.

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**, 277.

Grabherr, M.G., **Haas, B.J.**, **Yassour, M.**, **Levin, J.Z.**, **Thompson, D.A.**, **Amit, I.** ***et al.*** (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

He, J.-Y., **Zhang, Y.-H.**, **Ma, N.**, **Zhang, X.-L.**, **Liu, M.-H.** & **Fu, W.-M.** (2016) Comparative analysis of multiple ingredients in *Rosa roxburghii* and *R. sterilis* fruits and their antioxidant activities. *Journal of Functional Foods*, **27**, 29–41.

Hibrand Saint-Oyant, L., **Ruttink, T.**, **Hamama, L.**, **Kirov, I.**, **Lakhwani, D.**, **Zhou, N.N.** ***et al.*** (2018) A high-quality genome sequence of *Rosa chinensis* to elucidate ornamental traits. *Nature Plants*, **4**, 473–484.

Hu, F., **Lin, Y.** & **Tang, J.** (2014) MLGO: phylogeny reconstruction and ancestral inference from gene-order data. *BMC Bioinformatics*, **15**, 354.

Huang, M., **Xu, Q.** & **Deng, X.X.** (2014) L-ascorbic acid metabolism during fruit development in an ascorbate-rich fruit crop chestnut rose (*Rosa roxburghii* Tratt). *Journal of Plant Physiology*, **171**, 1205–1216.

Huang, X., **Yan, H.**, **Zhai, L.**, **Yang, Z.** & **Yi, Y.** (2019) Characterization of the *Rosa roxburghii* Tratt transcriptome and analysis of MYB genes. *PLoS One*, **14**, e0203014.

Jia, K.H., **Wang, Z.X.**, **Wang, L.**, **Li, G.Y.**, **Zhang, W.**, **Wang, X.L.** ***et al.*** (2022) SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *The New Phytologist*, **235**, 801–809.

Korf, I. (2004) Gene finding in novel genomes. *BMC Bioinformatics*, **5**, 59.

Krzywinski, M., **Schein, J.**, **Birol, I.**, **Connors, J.**, **Gascoyne, R.**, **Horsman, D.** ***et al.*** (2009) Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.

Kumar, S., **Stecher, G.**, **Li, M.**, **Knyaz, C.** & **Tamura, K.** (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, **35**, 1547–1549.

Kumar, S., **Stecher, G.**, **Suleski, M.** & **Hedges, S.B.** (2017) TimeTree: a resource for timelines, Timetrees, and divergence times. *Molecular Biology and Evolution*, **34**, 1812–1819.

Langfelder, P. & **Horvath, S.** (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Langmead, B. & **Salzberg, S.L.** (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Li, B. & **Dewey, C.N.** (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.

Li, H., **Fang, W.**, **Wang, Z.** & **Chen, Y.** (2022) Physicochemical, biological properties, and flavour profile of *Rosa roxburghii* Tratt, *Pyracantha fortuneana*, and *Rosa laevigata* Michx fruits: a comprehensive review. *Food Chemistry*, **366**, 130509.

Li, J., **Wang, Y.**, **Dong, Y.**, **Zhang, W.**, **Wang, D.**, **Bai, H.** ***et al.*** (2021) The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis. *Horticulture Research*, **8**, 53.

Liang, G.Y., **Alexander, I.G.** & **Peter, G.W.** (1989) Pentacyclic triterpenes from the fruits of Rosa sterilis. *Journal of Natural Products*, **52**, 162–166.
Liu, M.H., **Zhang, Q.**, **Zhang, Y.H.**, **Lu, X.Y.**, **Fu, W.M.** & **He, J.Y.** (2016) Chemical analysis of dietary constituents in *Rosa roxburghii* and *Rosa sterilis* fruits. *Molecules*, **21**, 1204.
Liu, W., **An, H.M.** & **Yang, M.** (2013) Overexpression of *Rosa roxburghii* L-galactono-1, 4-lactone dehydrogenase in tobacco plant enhances ascorbate accumulation and abiotic stress tolerance. *Acta Physiologiae Plantarum*, **35**, 1617–1624.
Liu, X., **Bulley, S.M.**, **Varkonyi-Gasic, E.**, **Zhong, C.** & **Li, D.** (2023) Kiwifruit bZIP transcription factor AcePosF21 elicits ascorbic acid biosynthesis during cold stress. *Plant Physiology*, **192**, 982–999.
Liu, X., **Wu, R.**, **Bulley, S.M.**, **Zhong, C.** & **Li, D.** (2022) Kiwifruit MYBS1-like and GBF3 transcription factors influence L-ascorbic acid biosynthesis by activating transcription of GDP-L-galactose phosphorylase 3. *The New Phytologist*, **234**, 1782–1800.
López-Carbonell, M., **Munné-Bosch, S.** & **Alegre, L.** (2006) The ascorbate-deficient vtc-1 Arabidopsis mutant shows altered ABA accumulation in leaves and chloroplasts. *Journal of Plant Growth Regulation*, **25**, 137–144.
Lu, M., **An, H.** & **Li, L.** (2016) Genome survey sequencing for the characterization of the genetic background of *Rosa roxburghii* Tratt and leaf ascorbate metabolism genes. *PLoS One*, **11**, e0147530.
Ma, J. & **Bennetzen, J.L.** (2004) Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 12404–12410.
Manni, M., **Berkeley, M.R.**, **Seppey, M.**, **Simao, F.A.** & **Zdobnov, E.M.** (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, **38**, 4647–4654.
Marcais, G., **Delcher, A.L.**, **Phillippy, A.M.**, **Coston, R.**, **Salzberg, S.L.** & **Zimin, A.** (2018) MUMmer4: a fast and versatile genome alignment system. *PLoS Computational Biology*, **14**, e1005944.
Moran, B., **Payne, C.**, **Langdon, Q.**, **Powell, D.**, **Brandvain, Y.** & **Schumer, M.** (2021) The genomic consequences of hybridization. *eLife*, **4**, e69016.
Murat, F., **Zhang, R.**, **Guizard, S.**, **Gavranović, H.**, **Flores, R.**, **Steinbach, D.** ***et al.*** (2015) Karyotype and gene order evolution from reconstructed extinct ancestors highlight contrasts in genome plasticity of modern rosid crops. *Genome Biology and Evolution*, **7**, 735–749.
Nakamura, N., **Hirakawa, H.**, **Sato, S.**, **Otagaki, S.**, **Matsumoto, S.**, **Tabata, S.** ***et al.*** (2018) Genome structure of *Rosa multiflora*, a wild ancestor of cultivated roses. *DNA Research*, **25**, 113–121.
Nguyen, L.-T., **Schmidt, H.A.**, **von Haeseler, A.** & **Minh, B.Q.** (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.
Noctor, G. (2006) Metabolic signalling in defence and stress: the central roles of soluble redox couples. *Plant, Cell & Environment*, **29**, 409–425.
Ou, S., **Chen, J.** & **Jiang, N.** (2018) Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Research*, **46**, e126.
Panchy, N., **Lehti-Shiu, M.** & **Shiu, S.H.** (2016) Evolution of gene duplication in plants. *Plant Physiology*, **171**, 2294–2316.
Price, A.L., **Jones, N.C.** & **Pevzner, P.A.** (2005) De novo identification of repeat families in large genomes. *Bioinformatics*, **21**(Suppl 1), i351–i358.
Qiao, X., **Li, Q.**, **Yin, H.**, **Qi, K.**, **Li, L.**, **Wang, R.**, **Zhang, S.** & **Paterson, A.H.** (2019) Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biology*, **20**, 38.
Rozewicki, J., **Li, S.**, **Amada, K.M.**, **Standley, D.M.** & **Katoh, K.** (2019) MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research*, **47**, W5–w10.
Schnable, J.C., **Springer, N.M.** & **Freeling, M.** (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 4069–4074.
Shen, Y., **Li, W.**, **Zeng, Y.**, **Li, Z.**, **Chen, Y.**, **Zhang, J.** ***et al.*** (2022) Chromosome-level and haplotype-resolved genome provides insight into the tetraploid hybrid origin of patchouli. *Nature Communications*, **13**(1), 3511.
Smirnoff, N. (2018) Ascorbic acid metabolism and functions: a comparison of plants and mammals. *Free Radical Biology & Medicine*, **122**, 116–129.
Smirnoff, N. & **Wheeler, G.L.** (2000) Ascorbic acid in plants: biosynthesis and function. *Critical Reviews in Biochemistry and Molecular Biology*, **35**, 291–314.
Stanke, M., **Diekhans, M.**, **Baertsch, R.** & **Haussler, D.** (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, **24**(5), 637–644.
Stull, G.W., **Pham, K.K.**, **Soltis, P.S.** & **Soltis, D.E.** (2023) Deep reticulation: the long legacy of hybridization in vascular plant evolution. *The Plant Journal*, **114**, 743–766.
Talavera, G. & **Castresana, J.** (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, **56**, 564–577.
Tang, H., **Bowers, J.E.**, **Wang, X.**, **Ming, R.**, **Alam, M.** & **Paterson, A.H.** (2008) Synteny and collinearity in plant genomes. *Science*, **320**, 486–488.
Tao, J., **Wu, H.**, **Li, Z.**, **Huang, C.** & **Xu, X.** (2018) Molecular evolution of GDP-D-mannose epimerase (GME), a key gene in plant ascorbic acid biosynthesis. *Frontiers in Plant Science*, **9**, 1293.
Tholl, D. (2015) Biosynthesis and biological functions of terpenoids in plants. *Advances in Biochemical Engineering/Biotechnology*, **148**, 63–106.
Vaser, R., **Sovic, I.**, **Nagarajan, N.** & **Sikic, M.** (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, **27**, 737–746.
Vurture, G.W., **Sedlazeck, F.J.**, **Nattestad, M.**, **Underwood, C.J.**, **Fang, H.**, **Gurtowski, J.** ***et al.*** (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, **33**, 2202–2204.
Walker, B.J., **Abeel, T.**, **Shea, T.**, **Priest, M.**, **Abouelliel, A.**, **Sakthikumar, S.** ***et al.*** (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**, e112963.
Wang, S., **Liang, H.**, **Wang, H.**, **Li, L.**, **Xu, Y.**, **Liu, Y.** ***et al.*** (2022) The chromosome-scale genomes of *Dipterocarpus turbinatus* and *Hopea hainanensis* (Dipterocarpaceae) provide insights into fragrant oleoresin biosynthesis and hardwood formation. *Plant Biotechnology Journal*, **20**, 538–553.
Wang, Y., **Wang, Z.**, **Geng, S.**, **Du, H.**, **Chen, B.**, **Sun, L.** ***et al.*** (2023) Identification of the GDP-L-galactose phosphorylase gene as a candidate for the regulation of ascorbic acid content in fruits of *Capsicum annuum* L. *International Journal of Molecular Sciences*, **24**, 7529.
Wang, Z., **Jiang, Y.**, **Bi, H.**, **Lu, Z.**, **Ma, Y.**, **Yang, X.** ***et al.*** (2021) Hybrid speciation via inheritance of alternate alleles of parental isolating genes. *Molecular Plant*, **14**, 208–222.
Wang, Z., **Kang, M.**, **Li, J.**, **Zhang, Z.**, **Wang, Y.**, **Chen, C.** ***et al.*** (2022) Genomic evidence for homoploid hybrid speciation between ancestors of two different genera. *Nature Communications*, **13**, 1987.
Wen, D., **Yu, Y.**, **Zhu, J.** & **Nakhleh, L.** (2018) Inferring phylogenetic networks using PhyloNet. *Systematic Biology*, **67**, 735–740.
Xiang, Y., **Huang, C.H.**, **Hu, Y.**, **Wen, J.**, **Li, S.**, **Yi, T.** ***et al.*** (2017) Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Molecular Biology and Evolution*, **34**, 262–281.
Xing, C., **Liu, Y.**, **Zhao, L.**, **Zhang, S.** & **Huang, X.** (2019) A novel MYB transcription factor regulates ascorbic acid synthesis and affects cold tolerance. *Plant, Cell & Environment*, **42**, 832–845.
Xu, P., **Liu, X.**, **Xiong, X.**, **Zhang, W.**, **Cai, X.**, **Qiu, P.** ***et al.*** (2017) Flavonoids of *Rosa roxburghii* Tratt exhibit anti-apoptosis properties by regulating PARP-1/AIF. *Journal of Cellular Biochemistry*, **118**, 3943–3952.
Xu, Q., **Wen, X.** & **Deng, X.** (2004) A simple protocol for isolating genomic DNA from chestnut rose (*Rosa roxburghii* tratt) for RFLP and PCR analyses. *Plant Molecular Biology Reporter*, **22**, 301–302.
Xu, S.J., **Wang, X.**, **Wang, T.Y.**, **Lin, Z.Z.**, **Hu, Y.J.**, **Huang, Z.L.** ***et al.*** (2020) Flavonoids from *Rosa roxburghii* Tratt prevent reactive oxygen species-mediated DNA damage in thymus cells both combined with and without PARP-1 expression after exposure to radiation in vivo. *Aging*, **12**, 16368–16389.
Xu, Z. & **Wang, H.** (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, **35**, W265–W268.
Yamaguchi, S. (2008) Gibberellin metabolism and its regulation. *Annual Review of Plant Biology*, **59**, 225–251.
Yan, X., **Zhang, X.**, **Lu, M.**, **He, Y.** & **An, H.** (2015) De novo sequencing analysis of the *Rosa roxburghii* fruit transcriptome reveals putative ascorbate biosynthetic genes and EST-SSR markers. *Gene*, **561**, 54–62.
Yan, Y., **Liu, Y.**, **Lu, M.**, **Lu, C.**, **Ludlow, R.A.**, **Yang, M.** ***et al.*** (2023) Gene expression profiling in *Rosa roxburghii* fruit and overexpressing RrGGP2 in tobacco and tomato indicates the key control point of AsA biosynthesis. *Frontiers in Plant Science*, **13**, 1096493.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yuan, H., **Wang, Y.**, **Chen, H.** & **Cai, X.** (2020) Protective effect of flavonoids from *Rosa roxburghii* Tratt on myocardial cells via autophagy. *3 Biotech*, **10**, 58.

Zhang, L., **Chen, F.**, **Zhang, X.**, **Li, Z.**, **Zhao, Y.**, **Lohaus, R.** ***et al.*** (2020) The water lily genome and the early evolution of flowering plants. *Nature*, **577**, 79–84.

Zhang, Z., **Xiao, J.**, **Wu, J.**, **Zhang, H.**, **Liu, G.**, **Wang, X.** ***et al.*** (2012) ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochemical and Biophysical Research Communications*, **419**, 779–781.