Genomic insights into the evolution and adaptation of the memorial roses (*Rosa lucieae*; Rosaceae) in saline environments

Ji-Hyeon Jeon[1], Younhee Shin[2*] & Seung-Chul Kim[1*]

[1]Department of Biological Sciences, Sungkyunkwan University, Suwon 16419, Republic of Korea;

[2] Research and Development Center, Insilicogen Inc., Yongin 16954, Republic of Korea

Running head: Genomic insights into the evolution of *Rosa lucieae*

*Corresponding authors: Seung-Chul Kim, sonchus96@skku.edu; Younhee Shin, yhshin@insilicogen.com

ABSTRACT

Rosa lucieae (syn. *Rosa wichurana*; Rosaceae), also known as the memorial rose, is one of the major wild progenitors of modern rose cultivars. In breeding roses, *R. lucieae* has contributed not only to white blooms and shiny leaves, but most of all, to succeed in various harsh conditions and high resistance to diseases and pests. The wild *R. lucieae* inhabits littoral sand banks or rocks, where drought stress and high salinity may inhibit plant growth. In this study, we assembled the chromosome-level genome sequence of a wild accession of *R. lucieae* and identified potential genes that may aid in the evolution of salt tolerance. The assembly is 544.7 Mb length and contains 36,876 protein-coding genes. The gene family evolution analysis revealed that rapid evolution of stress tolerant functions in *R. lucieae*, such as Ca^{2+}-mediated signaling, homeostasis, protein repair, or root growth. Moreover, the gene expression analysis found that some genetic functions related to salt tolerant responses were enriched, including MAPK and phytohormone signaling. The first high-quality genome assembly of wild *R. lucieae* not only contributes significantly to our understanding of genome organization and some adaptive mechanisms in *R. lucieae*, but it also may enhance our efforts in rose breeding programs.

INTRODUCTION

The roses are the flowering plants belonging to the genus *Rosa* in the family Rosaceae, encompassing 120–200 wild species and more than 30,000 cultivars (Rehder, 1940; Nybom, 2009). As renowned for their beauty and fragrance, roses have been beloved for millennia along with the history of human civilization, and cultivated for ornamental, aromatizing, therapeutic and culinary purposes in global (Rehder, 1940; Yu and Ku, 1985; Gudin, 2003; Nybom, 2009). Despite importance of roses, rose cultivation has faced various hurdles to overcome, including environmental changes, diseases and pests, hybrid incompatibility, and productivity (Gerard, 1897; Van Fleet, 1908; Gudin, 2003; Nybom, 2009). Breeding hardy wild roses, however, developed many robust hybrid cultivars which can cope with those problems (Van Fleet, 1908; Gudin, 2003). *Rosa lucieae* Franch. & Rochebr. ex Crép., also well known as its synonym, *Rosa wichurana* Crép., or the memorial rose, has been cherished for centuries as one of the major progenitors contributed to developing many modern rose cultivars (Gerard, 1897; Van Fleet, 1908; Yu and Ku, 1985; Crespel and Mouchotte, 2003; Marriott, 2003; Nybom, 2009). *R. lucieae* is a rambler rose vigorously growing even in the bad climate and on bare soil or rocks (Gerard, 1897; Rehder, 1940; Yu and Ku, 1985), highly resistant to the prevalent rose diseases such as black spot (*Diplocarpon rosae*; Noack, 2003; Byrne et al., 2005), and hybridizes easily (Gerard, 1897; Van Fleet, 1908). Such characteristics of *R. lucieae* have led the successful development of thousands of robust *R. lucieae* hybrid cultivars (Gerard, 1897; Van Fleet, 1908; Rehder, 1940; Yu and Ku, 1985; Byrne et al., 2005). Consequently, in the classification scheme of modern roses, the Hybrid Wichurana is now identified as one of the cultivar groups, bred by crossing *R. lucieae* and other roses and characterized by its rambling habit and wide variety of flower shapes and colors (Cairns, 2003).

R. lucieae has been classified as a member of the section *Synstylae* within genus *Rosa*,

where the species are closely related but diversified with unique characteristics to adapt to various ecological habitats (Rehder, 1940; Wissemann, 2003). Unlike most of other *Rosa* section *Synstylae* species, the natural habitats of *R. lucieae* are open and exposed rocky slopes or banks, especially along the sea shores and riverside at low elevation (Rehder, 1940; Ohba, 1993; Ku and Robertson, 2003). In these habitats, both abiotic stresses such as high light, high salinity and heavy metals, and biotic stresses caused by various herbivores and pathogens restrict plant growths and introduction. Nevertheless, *R. lucieae* has adapted to harsh environments of those habitats, evolving to resist and tolerate to various abiotic and biotic stresses. In addition to their ornamental characteristics, stress resistance and resilience in various environmental conditions are pivotal in breeding roses, efficiently increasing productivity and durability (Van Fleet, 1908; Gudin, 2003; Debener and Linde, 2009; Leus et al., 2018). Utilizing wild roses including *R. lucieae* in the rose breeding programs will pave the way for introducing the stress tolerant traits to cultivated roses (Van Fleet, 1908; Noack, 2003; De Cock et al., 2007). Understanding genetic and genomic architectures of roses also promises improvement in breeding roses for the resistance by allowing the applications such as marker-assisted selection and genetic engineering (Debener and Hibrand-Saint Oyant, 2009; Debener and Linde, 2009; Gahlaut et al., 2021). Recently, several transcriptomic approaches illustrated genetic functions and mechanisms of the stress responses in resistant wild roses, including disease resistance (Xiang et al., 2019) and cold tolerance (Zhang et al., 2016; Zhuang et al., 2021). Additionally, the genome of *Rosa rugosa* provided insights into the evolution of stress response in harsh environments (Chen et al., 2021). Nevertheless, the genomic intricacies of roses have impeded the extensive understanding on the genomic bases of key characteristics in the rose cultivation (Gahlaut et al., 2021).

The salt stress is one of major abiotic stresses in plants, which induces osmotic stress, ion toxicity, oxidative stress, and decreasing in photosynthesis (Zhu, 2001; Munns and Tester,

2008). In agriculture, nearly 20% of irrigated lands in global have been salinized, and recurrent irrigation with brackish groundwater or return flow caused from limited sources of fresh water accumulates salts in agricultural lands (Ghassemi et al., 1995; Tilman et al., 2002; Munns, 2005). Meanwhile, the increased level of soil salinity has severely affected the decline in productivity and quality of crops, resulting in consistent and complex economic losses (Munns and Gilliham, 2015). Roses are generally salt sensitive with lower yield, fewer flowers, and poor growths even in relatively low salt concentrations (Cabrera and Perdomo, 2003; Cabrera et al., 2009; Niu and Sun, 2017). Nevertheless, the genetic or genomic studies of roses not only lags behind other rosaceous crops (e.g., apples, peaches, and strawberries; Shulaev et al., 2008; Foucher, 2009; Gahlaut et al., 2021), but the salt tolerance of roses is also poorly understood compared to the aesthetic features (e.g., flower forms, colors, and fragrance), which have been priorly considered in breeding roses (Gudin, 2003; Bao et al., 2021). Given the deteriorating soil salinity which stems from recurrent irrigation, urbanization and climate changes, the salt tolerance become pivotal in the rose cultivation (Cai et al., 2014; Corwin, 2021). Consequently, the advent of high-throughput next-generation sequencing technique have led recent multi-omics investigation into the salt tolerance in various rose cultivars of *Rosa chinensis* (Tian et al., 2018; Bao et al., 2021), *Rosa rugosa* (Chen et al., 2021; Qi et al., 2023), *Rosa* × *damascena* and *Rosa* × *hybrida* (Ren et al., 2024).

In this study, we assembled a chromosome-level genome of a wild accession of *R. lucieae*, inhabiting the saline sea shore, in order to understand the genomic evolution and nature of salt tolerance in *R. lucieae*. The chromosome-level genome of *R. lucieae* and its evolution in salt tolerance with adaptation or acclimation to saline environment, will not only contributes to our understanding of genomic features of wild roses, but it also may enhance our efforts in rose breeding programs.

MATERIALS AND METHODS

Sample Collection and Preparation

To develop a genomic sequence of *R. lucieae*, a wild individual of *R. lucieae* (accession ID: SG001; NCBI BioSample accession: SAMN39614440) was collected from a coastal field in Jeju, Korea, and its voucher specimen was deposited into the Ha Eun Herbarium (SKK) of Sungkyunkwan University. Genomic DNA (gDNA) was isolated from fresh young leaves of the accession using the Exgene Plant SV mini kit (GeneAll Biotechnology Co., Ltd., Seoul, Korea), following the manufacturer's instructions. For the quality control, the isolated gDNA was quantified and assessed using QuantiFluor dsDNA System (Promega Corp., WI, USA) and the 2100 Bioanalyzer System (Agilent Technologies, Inc., CA, USA). Additionally, flower buds (BioSample accession: SAMN40548981) and young leaves (BioSample accession: SAMN40548982) were collected from the same individual to provide evidence for the splice site prediction. The samples were stored in RNAlater solution (Thermo Fisher Scientific Inc., MA, USA) immediately after the collection to stabilize their RNAs. Total RNAs were isolated using Maxwell RSC Plant RNA Kit and Instrument (Promega Corp., WI, USA) by Macrogen, Inc. (Seoul, Korea). For the quality control, the isolated RNAs were quantified and assessed using Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher Scientific Inc., MA, USA) and the 2100 Bioanalyzer System.

Molecular Sequencing and Genome Assembly

The gDNA was prepared into respective libraries for Illumina HiSeq short-read and PacBio SMRT long-read sequencing platforms. The short-read sequencing library was prepared by random fragmentation, followed by 5′- and 3′-adapter ligations, using the TruSeq

Nano DNA kit (Illumina, Inc., CA, USA). The fragmented library was sequenced via Illumina HiSeq platform (Illumina, Inc., CA, USA) by Macrogen, Inc. (Seoul, Korea) after the size selection. The PacBio SMRTbell library was prepared using SMRTbell Express Template Prep Kit (Pacific Biosciences of California, Inc., CA, USA), following the manufactures instructions. The large fragments longer than 20 kb were selected using BluePippin system (Sage Science, Inc., MA, USA), and the large-insert library was sequenced via PacBio Sequel platform (Pacific Biosciences of California, Inc., CA, USA) by DNALink, Inc. (Seoul, Korea). The total RNAs were prepared into sequencing libraries by purifying poly-A containing mRNAs, random fragmentation, cDNA synthesis, and 5′- and 3′-adapter ligations, using the TruSeq Stranded mRNA kit (Illumina, Inc., CA, USA). The fragmented libraries were sequenced via Illumina NovaSeq platform (Illumina, Inc., CA, USA) by Macrogen, Inc. (Seoul, Korea) (Supplementary Table S1).

The genome size, heterozygosity and repeat content of *R. lucieae* was estimated with *k*-mer analysis using Jellyfish v2.1.3 (Marçais and Kingsford, 2011) and GenomeScope v2.0 (Ranallo-Benavidez et al., 2020), after preprocessing short reads by trimming the sequencing adapters, PCR dimers, and low-quality sequences using BBDuk v38.26 (Bushnell, 2014) and removing contaminant reads by mapping reads to NCBI RefSeq complete bacterial, viral and organellar genomes using bowtie2 v2.3.4 (Langmead and Salzberg, 2012). PacBio subreads were error-corrected into consensus reads using SMRT Link v2.3 (Pacific Biosciences of California, Inc., CA, USA), and preprocessed short reads were mapped to the consensus reads to remove contaminant reads using bwa v0.7 (Li, 2013). The preprocessed PacBio consensus reads were *de-novo* assembled into primary contigs using a diploid-aware assembler FALCON, and the primary contigs were phased and updated using a diploid assembler FALCON-Unzip (Chin et al., 2016). The phased primary contigs of the diploid genome assembly were curated to remove haplotigs to complete the draft genome assembly using

Purge haplotigs v1.1.1 (Roach et al., 2018). Error sequences in the draft genome assembly were corrected using Pilon v1.23, after mapping preprocessed short reads to the draft genome assembly using bwa aligner (Li, 2013; Walker et al., 2014). The *de novo* genome assembly was assessed for the completeness based on the universal single-copy orthologs using BUSCO v5.3.0 and eudicots_odb10, embryophyta_odb10, and viridiplantae_odb10 datasets (Manni et al., 2021). The *de novo* genome assembly was scaffolded into chromosomes using RagTag v2.1.0 (Alonge et al., 2022), referenced from the chromosome-level assembly of *R. lucieae* 'Basye's Thornless' (GenBank accession: GCA_024704745.1; Zhong et al., 2021). The scaffolded chromosomes were assessed with the genome collinearity analysis with the reference assembly using MUMmer4 v4.0.0rc1 (Marçais et al., 2018), examining any genomic gaps or putative misassembled contigs.

Genome Annotation

The repeat sequences of genome of *R. lucieae* were *de novo* predicted and developed into the *de novo* repeat library using RepeatModeler v1.0.8 (Smit and Hubley, 2008), which employing complementary repeat annotation methods like RECON v1.08 (Bao & Eddy, 2002), RepeatScout v1.0.5 (Price et al., 2005), and TRF v4.04 (Benson, 1999). The repeat regions of draft genome were identified, classified into subclasses based on the Dfam database v2.0 (Hubley et al., 2016) and RepBase database v23.10 (Bao et al., 2015), and masked using RepeatMasker v4.0.7 (Smit et al., 2013).

Protein-coding genes of *R. lucieae* genome were predicted and annotated integrating *ab initio*, RNA-seq evidenced, and homology-based gene predictions. For *ab initio* gene prediction, the rosaceous RefSeq gene models of *R. chinensis* 'Old Blush' (RefSeq accession: GCF_002994745.2; Raymond et al., 2018), *R. rugosa* (GCF_958449725.1), *Fragaria vesca*

(GCF_000184155.1; Shulaev et al., 2011), *Prunus persica* 'Lovell' (GCF_000346465.2; IPGI, 2013), and *Malus domestica* 'Golden Delicious' (GCF_002114115.1; Daccord et al., 2017) were trained and optimized using AUGUSTUS v3.1.0 (Stanke & Waack, 2003). With meta-parameters of rosaceous reference gene models, the ab initio gene model of *R. lucieae* was predicted within the draft genome using AUGUSTUS v3.1 (Stanke and Waack, 2003). For homology-based prediction, protein-coding genes from reference genomes (Supplementary Table S2) were aligned to the draft genome of *R. lucieae* to identify homologous regions using Exonerate v2.4.0 (Slater & Birney, 2005). For RNA-seq evidenced prediction, each RNA-seq accession was aligned to the draft genome of *R. lucieae* using STAR v2.5.1a (Dobin et al., 2013). Leveraging the RNA-seq alignment, reference gene models of *R. chinensis*, *R. rugosa*, and *F. vesca*, and evidence from ab initio and homology-based gene predictions, an integrated consensus gene model was predicted using GeMoMa v1.9 (Keilwagen et al., 2018) based on conserved coding-sequence and intron positions.

A set of protein-coding genes of *R. lucieae* were annotated in their functions. The gene ontologies (GOs; Ashburner et al., 2000) and definitions were annotated using blastp v2.15.0 (Camacho et al., 2009), Blast2GO v6.0.3 (Conesa et al., 2005), InterProScan v5.52 (Jones et al., 2014), and Trinotate v4.0.2 (Bryant et al., 2017), with databases including NCBI non-redundant protein sequence (nr; accessed on 12 Feb 2024), Gene Ontology Annotation (GOA; v2023.08; Camon et al., 2004), Pfam (v36.0; El-Gebali et al., 2019), KEGG (Kanehisa et al., 2012), Swiss-Prot (v2024_01; Boutet et al., 2007), and EggNOG (v5.0; Huerta-Cepas et al., 2019) databases.

Comparative Genomic and Evolutionary Analyses

To infer the genetic evolution of *R. lucieae*, the orthologous gene groups were

identified using OrthoMCL v2.0.9 (Li et al., 2003), from the gene models of *R. lucieae* and other reference species including *R. chinensis*, *R. rugosa*, *F. vesca*, *P. persica*, *M. domestica*, and *Populus trichocarpa* (RefSeq accession: GCF_000002775.4; Tuskan et al., 2006). The phylogenetic inference and divergence time estimation were analyzed with conserved single-copy orthologs. Divergence times of *R. lucieae* and reference species were estimated by running two Markov chain Monte Carlo chains for 10,000,000 generations using BEAST2 v2.6.3 (Bouckaert et al., 2014) with the partitioned alignment and JTT amino-acid exchange rate matrix from IQ-TREE v2.2.0 (Chernomor et al., 2016) and ModelFinder (Kalyaanamoorthy et al., 2017), discarding the first 10% of trees as post-burn-in. The divergence time estimation was calibrated using the oldest fossil records within Rosaceae (113.0–100.5 Mya; Peppe et al., 2008) and Eurosids (145.0–100.5 Mya; Ward, 1899; Mellon et al., 1963). Based on identified orthologous groups and estimated divergence time, gene family evolution in *R. lucieae* and other rosaceous species were inferred using CAFE v5.1.0 (Mendes et al., 2020). To infer the significantly evolved gene functions, the GO and KEGG orthology (KO) enrichment analyses on expanded or contracted gene families were performed using the R package clusterProfiler v4.4.4 (Wu et al., 2021).

The genomic synteny among the genome of *R. lucieae* and the reference genomes of its closely related species in Rosoideae (i.e., *R. chinensis*, *R. rugosa*, and *F. vesca*) was analyzed to infer the structural genomic variations within their evolution. The pairwise genetic homologies between species were assessed using blastp, and homology-based collinear blocks were identified using MCScanX (Wang et al., 2012). The collinearity blocks between species were visualized using SynVisio (Bandi and Gutwin, 2020). To detect the gene duplications within the genome of *R. lucieae*, the paralogous collinear blocks were also identified using blastp and MCscanX.

Results

Molecular Sequencing, Genome Size Estimation and Assembly

A total of 374.6 million paired-end short-sequence reads with 56.6 Gbp (123.4× coverage depths of the estimated haploid genome size) were generated through Illumina HiSeq platform, and 2.7 million long-sequence reads with 28.7 Gbp (62.6× coverage depths of the estimated genome size) were generated through PacBio Sequel platform (Supplementary Table S1). Each of two transcriptome samples from the young leaves and flower buds was sequenced into 9.0-Gbp data with 89 million paired-end short-sequence reads through Illumina NovaSeq platform (Supplementary Table S1). The *k*-mer analysis using genomic short reads estimated that haploid genome size of *R. lucieae* was 458.3 Mbp, of which repeat content was 59.61% (273.2 Mbp), and the heterozygosity rate was 1.35% (Table 1; Supplementary Figure S1).

The error-corrected PacBio subreads were *de novo* assembled into primary 2,067 contigs with a total of 619.5 Mbp lengths, and this heterozygous primary assembly was curated into the 1,356 contigs with a total of 544.9 Mbp lengths. Finally, error sequences of the curated assembly were corrected with short sequence reads, establishing the 544.5 Mbp-length draft *de novo* genome assembly with 1,356 contigs (Supplementary Table S3). The BUSCO assessment identified more than 96% of BUSCO eudicots-, embryophyte-, and viridiplantae-ortholog datasets within the draft assembly of *R. lucieae* respectively (Table 1). The *de novo* assembly was subsequently scaffolded into seven chromosomal sequences and 187 unplaced contigs, finalizing the 544.7 Mb-length chromosomal genomic sequences of *R. lucieae* (Figure 1; Table 1; Supplementary Table S4). The genome collinearity analysis with the reference chromosomes indicated no genomic gaps, missing data, or possible mis-assembly within the draft genome assembly of *R. lucieae* in this study (Supplementary Figure S2).

Genome Annotation

Within the genome of *R. lucieae*, 49.67% of sequences (270.5 Mbp) were identified as repeat sequences. Among them, most of repeat sequences were interspersed repeats (97.4%), of which a majority (54.1%) were identified as LTR elements, 15.5% were identified as DNA elements, and 24.5% were unclassified (Supplementary Table S5).

A total of 36,876 protein-coding genes were identified and annotated within the genome of *R. lucieae* (Table 1), the number of which is close to the annotations of other *Rosa* genomes (*R. lucieae*: 32,674; *R. chinensis*: 30,924–39,669; *R. rugosa*: 29,146–39,704 protein-coding genes; Hibrand Saint-Oyant et al., 2018; Raymond et al., 2018; Chen et al., 2021; Zang et al., 2021; Zhang et al., 2024). Among the protein-coding genes, all the genes except three (36,873 genes; 99.99%) were functionally annotated from the public databases, including nr (36,870 genes; 99.98%), GO (33,795 genes; 91.64%), KEGG (30,378 genes; 82.38%), and Pfam (33,012 genes; 89.52%) databases (Supplementary Table S6).

Ortholog Group Inference and Genome Synteny

Out of 36,876 protein-coding genes identified within the genome of *R. lucieae*, the ortholog group inference analysis classified 32,666 genes into 19,667 gene family groups, of which 14,436 groups were shared among Rosoideae species (i.e., *R. lucieae*, *R. chinensis*, *R. rugosa* and *F. vesca*), 15,790 groups were shared among the genus *Rosa*, and 586 groups were unique to *R. lucieae* (Figure 2A). Among 32,666 classified genes, 31,264 were inter-specific orthologous and the other 1,402 were species-specific paralogous genes (Figure 2B). Based on the gene annotations of Rosoideae genomes and their homologies, the inter-specific collinear blocks were identified to infer the genome synteny. Within the genus *Rosa*, the

genome synteny was highly conserved among species, especially between *R. chinensis* and *R. rugosa*. The genome synteny was also moderately conserved between the genera *Rosa* and *Fragaria*, even though several structural variations were identified (Figure 3).

Phylogeny and Gene Family Evolution Inferences

The ortholog group inference analysis identified 2,228 single-copy orthologous gene families shared among the genome of *R. lucieae*, five other rosaceous reference genomes, and the outgroup genome of *P. trichocarpa*. The Bayesian phylogenetic inference using the single-copy orthologous gene set corroborated the taxonomic classification of the species, resolving robust monophyly of the family Rosaceae, subfamilies Rosoideae and Amygdaloideae, and genus *Rosa* respectively (Figure 2C). Within the genus *Rosa*, the closer relationship between *R. lucieae* and *R. chinensis* was inferred compared to with *R. rugosa*, which is congruent with the previous phylogenetic inferences of the genus *Rosa* using conventional molecular markers (Wissemann and Ritz, 2005; Fougère-Danezan et al., 2015).

Based on the inferred orthologous gene families and species phylogeny, the gene family evolution was inferred within each phylogenetic lineage. While contractions of gene families were predominantly involved in the evolution of most of other rosaceous lineages, the evolution of *R. lucieae* involved the pronounced expansion of gene families, from the stem lineage of genus *Rosa* (Figure 2C). After the divergence with *R. chinensis*, 1,875 gene families were expanded with 2,579 more genes, whereas 385 gene families were contracted with 609 less genes within the lineage of *R. lucieae*. Among the expanded or contracted gene families involved in the *R. lucieae* lineage, 283 gene families were inferred to be rapidly evolving with the significant family-size changes (p-value < 0.01) in this lineage. For the *R. lucieae* genes of rapidly evolving families, the GOs related to plant signaling, secondary

metabolites, salinity stress, root system, homeostasis, calcium ion, protein repair, or defense mechanisms were significantly enriched (Figure 4). Likewise, based on their KEGG orthology annotations, the KEGG pathways related to the steroid hormones including brassinosteroids (BRs), secondary metabolites, and stress responses were significantly enriched (Figure 5).

DISCUSSION

Structural Genomic Evolution of Rosa lucieae

The genome of *R. lucieae* was assembled into a total of 544.7 Mbp sequence scaffolds, including seven chromosomal sequences. The genome size was almost similar to the genome of *R. chinensis* (515.1–541.3 Mbp; sect. *Chinenses*; Raymond et al., 2018; Hibrand Saint-Oyant et al., 2018; Zhang et al., 2024), and slightly to moderately larger than those of *R. rugosa* (401.2–417.1 Mbp; sect. *Rosa*; GCF_958449725.1; Chen et al., 2021; Zang et al., 2021), *Rosa roxburghii* (504.0–529.9 Mbp; subg. *Platyrhodon*; Yang et al., 2024; Zong et al., 2024), and *Rosa persica* (366.1 Mbp; subg. *Hulthemia*; GCA_025331145.1). The genome size profile of these *Rosa* species was corroborated with the nucleic DNA amount investigation, estimating 2C was 1.13 pg in *R. lucieae*, 1.16 pg in *R. chinensis*, 0.98 pg in *R. rugosa*, 0.95 pg in *R. roxburghii*, and 0.84 pg in *R. persica* (Yokoya et al., 2000). Despite of the increased genome size, the repeat content of genome of *R. lucieae* (49.7%; Table 1; Supplementary Table S5) was nearly similar to *R. rugosa* (50.3–54.1%; Chen et al., 2021; Zang et al., 2021), whereas it is notably increased in the genome of *R. chinensis* (61.8–67.9%; Raymond et al., 2018; Hibrand Saint-Oyant et al., 2018; Zhang et al., 2024). Despite the increased repeat content in the genome of *R. chinensis*, the genome synteny was highly conserved between *R. chinensis* and *R. rugosa* (Figure 3). Moreover, despite the closer

phylogenetic relationship of *R. chinensis* to *R. lucieae* than *R. rugosa*, more inter-chromosomal collinear blocks were identified between *R. chinensis* and *R. lucieae*, which also prevailed in the genomic study of *R. lucieae* 'Basye's thornless' (Zhong et al., 2021), compared to between *R. chinensis* and *R. rugosa* (Figures 2C and 3). Given the intra-genomic collinear blocks and increased gene content within the genome of *R. lucieae* (Figure 1; Table 1), putative gene duplication may have influenced the genomic evolution of *R. lucieae*, which aligned with the pronounced expansion of gene families in *R. lucieae* lineage detected from the gene family evolution analysis (Figure 2C).

Evolution of Salt Tolerance-Related Genes in Rosa lucieae

R. lucieae inhabits open and exposed rocky slopes or banks, especially along the sea shores and riverside at low elevation (Rehder, 1940; Ohba, 1993; Ku and Robertson, 2003), where the salts are constantly accumulated in their soils. In these natural habitats, *R. lucieae* has evolved to adapt to saline environments, resisting to and growing vigorously under salt stress. Although *R. rugosa* is another renowned rose species for salt tolerance, the earlier genomic study could not find any difference in genetic component related to salt tolerance, but only find the duplications of two putative genes (*DRIP2* and *PTR3*) related to water deprivation (Chen et al., 2021).

The high concentration of salts can cause both ionic and osmotic stresses in plants. Not only can it directly affect to the cytological and physiological dysfunction, but also involve other secondary cytotoxic effects caused from generation of reactive oxygen species (ROS), imbalance of other essential ions such as K^+, and malnutrition (Zhu, 2001; Botella et al., 2005). In plants, three major mechanisms of salt tolerance are known: homeostasis, detoxification, and growth control (Zhu, 2001; Munns, 2005; Munns and Tester, 2008).

Homeostasis can be established by compartmentalization of Na^+ and Cl^- into the vacuoles, modulation Na^+ influx and efflux, and accumulating compatible osmolytes, which are regulated by root endoderm and xylem, or stress signaling pathway including Ca^{2+}-mediated cascades (Zhu, 2001; Botella et al., 2005; Munns, 2005; Van Zelm et al., 2020). Ionic and oxidative toxicity can be detoxified or prevented by protein protection and repair, antioxidant metabolites, and mitogen-activated protein kinase (MAPK) cascades-induced detoxifying enzymes (Zhu, 2001; Chinnusamy and Zhu. 2003; Botella et al., 2005; Munns, 2005; Van Zelm et al., 2020). Salt stress also can inhibit the plant growth with stomal closure induced by abscisic acid (ABA), restricting the photosynthesis (Botella et al., 2005; Van Zelm et al., 2020). However, the growth-promoting hormones such as jasmonic acid (JA), BR, and gibberellic acid (GA) are responsible for the growth recovery after salt stress (Van Zelm et al., 2020).

To understand whether the remarkable changes of gene components in the genome evolution of *R. lucieae* are putatively related to its adaptive evolution in salt tolerance, the gene set of families expanded or contracted within the lineage of *R. lucieae* were enriched for their GOs and KEGG pathways. Within the rapidly evolving gene families of *R. lucieae*, the GOs related to responses to salt and osmotic stresses, and related to ionic and osmotic homeostasis were significantly enriched (Figure 4), aligning with the potential of genetic-component changes of *R. lucieae* in adaptation to salt stress. Additionally, the GOs related to ABA signaling, Ca^{2+} signaling and ROS (e.g., hydrogen peroxide (H_2O_2) and nitrogen oxide (NO)) signaling, and the KEGG pathways related to MAPK signaling were also significantly enriched, which are principal signaling pathways under salt stress in plants (Figures 4 and 5; Zhu, 2001; Botella et al., 2005; Van Zelm et al., 2020). The evolution of genetic components related to early signaling under salt stress may influenced the adaptation of *R. lucieae*, possibly enabling a rapid response to salt stress. Because early responses of salt stress

including generation of ABA and ROS can damage and impair plant systems, the downstream responses and recovery is another pivotal mechanism in salt tolerance (Van Zelm et al., 2020). The significant enrichment of metabolite-, homeostasis-, and protein repair-related GOs, and secondary metabolite- and protein processing-related KEGG pathways implied the putative evolution of ion- or ROS-detoxification strategies in *R. lucieae* (Figures 4 and 5). Likewise, pronounced enrichment of growth hormone-related GOs and KEGG pathways including JA, BR, and auxin, implying that vigorous growth of *R. lucieae* in saline environment putatively stems from the enhanced growth recovery from salt stress (Figures 4 and 5; Van Zelm et al., 2020). Consider that several studies proposed the possible roles of these growth hormones in stress tolerance (Robert-Seilaniantz et al., 2010), the evolution of growth hormone-related genes are putatively responsible for the salt tolerant tactics specific to *R. lucieae*. The GOs related to growth and development of root system were predominantly enriched within the rapidly evolving gene families of *R. lucieae* (Figure 4). Coupled with inhibiting plant growths, salt stress suppresses root elongation by reducing auxin level. However, the root development in *R. lucieae* can play an important role in adaptation to salt stress by growing roots away from saline environments (Galvan-Ampudia et al., 2013). Additionally, the development of root endodermis including the Casparian strip in *R. lucieae* can restrict apoplastic and transcellular transport of by accumulating lignin or suberin (Van Zelm et al., 2020). Consequently, adaptation of *R. lucieae* in saline environments may have putatively evolved with rapid expansion or contraction of the gene family with gene functions related to sensing high concentration of salt, salt stress signaling, detoxification, growth recovery, and root development. Further investigations of mechanisms and specific genes regulating salt tolerance in *R. lucieae*, and their comparative genetics with orthologous genes of other *Rosa* species will illustrate the evolution of salt tolerance in *R. lucieae*.

ACKNOWLEDGEMENTS

This study was supported by the National Research Foundation of Korea [grant number NRF-2017H1A2A1045787] to the first author.

AUTHOR CONTRIBUTIONS

J.-H.J, Y.S. and S.-C.K designed and coordinated the study. J.-H.J and S.-C.K collected the materials. J.-H.J and Y.S. analyzed molecular data. J.-H.J and S.-C.K wrote the manuscript. The manuscript was reviewed and accepted by all the authors.

DATA AVAILABILITY

Molecular sequencing data and genome assembly generated in this study were deposited into NCBI under BioProject accessions PRJNA1069305.

CONFLICT OF INTERESTS

The authors declare that there are no conflicts of interest.

SUPPLEMENTARY DATA

Supplementary data is available at online in the Supplemental Material section.

Figure S1. *K*-mer distribution in the short-read sequencing data of *Rosa lucieae*.

Figure S2. Syntenic dotplot of two *Rosa lucieae* assemblies, the *R. lucieae* 'Basye's Thornless' assembly (ASM2470474v1) and the wild *R. lucieae* assembly (Rluc_SG001_v1).

Table S1. Summary of whole-genome sequencing of *Rosa lucieae* and subsequent read

preprocessing.

Table S2. List of the reference genomes and their gene models used for the homology-based gene prediction of *Rosa lucieae*.

Table S3. Statistics of the *de novo* genome assembly of *Rosa lucieae*.

Table S4. Statistics of the genome scaffolding of *Rosa lucieae* assembly.

Table S5. Repeat sequences within the genome assembly of *Rosa lucieae*.

Table S6. Statistics of functional annotation of protein-coding genes from the genome of *Rosa lucieae*.

REFERENCES

Alonge M, Lebeigle L, Kirsche M, et al. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology* 23(1): 258.

Ashburner M, Ball CA, Blake JA, et al. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* 25(1): 25–29.

Bandi V, Gutwin C. 2020. Interactive exploration of genomic conservation. In: *Proceedings of the 46th Graphics Interface Conference on Proceedings of Graphics Interface 2020 (GI'20)*. Waterloo: Canadian Human-Computer Communications Society.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 6: 1–6.

Bao Y, Chen C, Fu L, Chen Y. 2021. Comparative transcriptome analysis of *Rosa chinensis* 'Old Blush' provides insights into the crucial factors and signaling pathways in salt stress response. *Agronomy Journal* 113(4): 3031–3050.

Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research* 12(8): 1269–1276.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27(2): 573–580.

Botella MA, Rosado A, Bressan RA, Hasegawa PM. 2005. Plant adaptive responses to salinity stress. In: Jenks MA, Hasegawa PM (eds.). *Plant Abiotic Stress*. Oxford: Blackwell Publishing, 37–70.

Bouckaert R, Heled J, Kühnert D, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10(4): e1003537.

Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. 2007. UniProtKB/Swiss-Prot: the manually annotated section of the UniProt KnowledgeBase. In: Edwards D (ed.). *Plant Bioinformatics: Methods and Protocols*. Totowa: Humana Press, 89–112.

Bryant DM, Johnson K, DiTommaso T, et al. 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Reports* 18(3): 762–776.

Bushnell B. 2014. *BBMap: a fast, accurate, splice-aware aligner*. Berkeley: Lawrence Berkeley National Lab., United States.

Cabrera RI, Perdomo P. 2003. Reassessing the salinity tolerance of greenhouse roses under soilless production conditions. *HortScience* 38(4): 533–536.

Cabrera RI, Solís-Pérez AR, Sloan JJ. 2009. Greenhouse rose yield and ion accumulation responses to salt stress as modulated by rootstock selection. *HortScience* 44(7): 2000–2008.

Cai X, Niu G, Starman T, Hall C. 2014. Response of six garden roses (*Rosa* × *hybrida* L.) to salt stress. *Scientia Horticulturae* 168: 27–32.

Cairns T. 2003. Classification: Horticultural Classification Schemes. In: Roberts AV, Debener T, Gudin S (eds.). *Encyclopedia of Rose Science*. Oxford: Academic Press, 117–124.

Camacho C, Coulouris G, Avagyan V, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10: 1–9.

Camon E, Magrane M, Barrell D, et al. 2004. The Gene Ontology Annotation (GOA) Database: sharing knowledge in UniProt with Gene Ontology. *Nucleic Acids Research* 32(Suppl_1), D262–D266.

Chen F, Su L, Hu S, 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.

Chernomor O, Von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology* 65(6): 997–1008.

Chin CS, Peluso P, Sedlazeck FJ. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods* 13(12): 1050–1054.

Chinnusamy V, Zhu JK. 2003. Plant salt tolerance. In: Hirt H, Shinozaki K (eds.). *Plant Responses to Abiotic Stress*. Berlin: Springer, 241–270.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18): 3674–3676.

Corwin DL. 2021. Climate change impacts on soil salinity in agricultural areas. *European Journal of Soil Science* 72(2): 842–862.

Crespel L, Mouchotte J. 2003. Breeding: Methods of Cross-Breeding. In: Roberts AV, Debener T, Gudin S (eds.). *Encyclopedia of Rose Science*. Oxford: Academic Press, 30–33.

Daccord N, Celton JM, Linsmith G, et al. 2017. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics* 49(7): 1099–1106.

Debener T, Hibrand-Saint Oyant L. 2009. Genetic engineering and tissue culture of roses. In: Folta KM, Gardiner SE (eds.). *Genetics and Genomics of Rosaceae*. New York: Springer, 393–409.

Debener T, Linde M. 2009. Exploring complex ornamental genomes: the rose as a model plant. *Critical Reviews in Plant Sciences* 28(4): 267–280.

Dobin A, Davis CA, Schlesinger F, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1): 15–21.

Edelman DW. 1975. *The Eocene Germer Basin Flora of South-Central Idaho*. MS Thesis, University of Idaho, USA.

Foucher F. 2009. Functional genomics in rose. In: Folta KM, Gardiner SE (eds). *Genetics and Genomics of Rosaceae*. New York: Springer, 381–392.

Fougère-Danezan M, Joly S., Bruneau A, Gao XF, Zhang LB. 2015. Phylogeny and biogeography of wild roses with specific attention to polyploids. *Annals of Botany* 115(2): 275–291.

Gahlaut V, Kumari P, Jaiswal V, Kumar S. 2021. Genetics, genomics and breeding in *Rosa* species. *The Journal of Horticultural Science and Biotechnology* 96(5): 545–559.

Galvan-Ampudia, CS, Julkowska MM, Darwish E, et al. 2013. Halotropism is a response of plant roots to avoid a saline environment. *Current Biology* 23(20): 2044–2050.

Gerard JN. 1897. Rose Hybrids. In: Sargent CS (ed.). *Garden and Forest: A Journal of Horticulture, Landscape Art and Forestry*, Vol. 10. New York: The Garden and Forest Publishing Co., 147.

Ghassemi F, Jakeman AJ, Nix HA. 1995. *Salinisation of Land and Water Resources: human causes, extent, management and case studies*. Wallingford: CAB International.

Gudin S. 2003. Breeding: Overview. In: Roberts AV, Debener T, Gudin S (eds.). *Encyclopedia of Rose Science*. Oxford: Academic Press, 25–30.

Hibrand Saint-Oyant L, Ruttink T, Hamama L, et al. 2018. A high-quality genome sequence of *Rosa chinensis* to elucidate ornamental traits. *Nature Plants* 4(7): 473–484.

Hubley R, Finn RD, Clements J, et al. 2016. The Dfam database of repetitive DNA families. *Nucleic Acids Research* 44(D1): D81–D89.

Huerta-Cepas J, Szklarczyk D, Heller D, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47(D1): D309–D314.

International Peach Genome Initiative. 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics* 45(5): 487–494.

Jones P, Binns D, Chang HY, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9): 1236–1240.

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14(6): 587–589.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* 40(D1): D109–D114.

Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* 19: 1–12.

Ku TC, Robertson KR. 2003. *Rosa*. In: Wu Z, Raven PH, Hong D (eds.). *Flora of China. Volume 9: Pittosporaceae through Connaraceae*. Beijing: Science Press, 339–381.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4): 357–359.

Leus L, Van Laere K, De Riek J, Van Huylenbroeck J. 2018. Rose. In: Van Huylenbroeck J (ed.). *Ornamental Crops*. Cham: Springer International Publishing, 719–767.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*: 1303.3997. Preprint: not peer reviewed.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* 13(9): 2178–2189.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 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(10): 4647–4654.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* 27(6): 764–770.

Marçais G, Delcher AL, Phillippy AM, et al. 2018. MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology* 14(1): e1005944.

Marriott M. 2003. History of roses in cultivation: Modern (Post–1800). In: Roberts AV, Debener T, Gudin S (eds.). *Encyclopedia of Rose Science*. Oxford: Academic Press, 402–409.

Mellon GB, Wall JH, Stelck CR. 1963. Lower Cretaceous section, Belcourt Ridge, northeastern British Columbia. *Bulletin of Canadian Petroleum Geology* 11(1): 64–72.

Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2020. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36(22-23): 5516–5518.

Munns R. 2005. Genes and salt tolerance: bringing them together. *New Phytologist* 167(3): 645–663.

Munns R, Gilliham M. 2015. Salinity tolerance of crops – what is the cost? *New Phytologist* 208(3): 668–673.

Munns R, Tester M. 2008. Mechanisms of salinity tolerance. *Annual Review of Plant Biology* 59: 651–681.

Niu G, Sun Y. 2019. Salt tolerance in roses: a review. *Acta Horticulturae* 1232: 143–150.

Noack R. 2003. Breeding: Selection Strategies for Disease and Pest Resistance. In: Roberts AV, Debener T, Gudin S (eds.). *Encyclopedia of Rose Science*. Oxford: Academic Press, 49–55.

Nybom H. 2009. Introduction to *Rosa*. In: Folta KM, Gardiner SE (eds.). *Genetics and Genomics of Rosaceae*. New York: Springer, 339–351.

Ohba H. 1993. *Rosa* L. In: Iwatsuki K, Boufford DE, Ohba H (eds.). *Flora of Japan*, Vol. 3. Tokyo: Kodansha, 169–177.

Peppe DJ, Hickey LJ, Miller IM, Green WA. 2008. A morphotype catalogue, floristic analysis and stratigraphic description of the Aspen Shale flora (Cretaceous–Albian) of southwestern Wyoming. *Bulletin of the Peabody Museum of Natural History* 49(2): 181–208.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl_1): i351–i358.

Qi S, Wang X, Wu Q, et al. 2023. Morphological, physiological and transcriptomic analyses reveal potential candidate genes responsible for salt stress in *Rosa rugosa*. *Ornamental Plant Research* 3(1): 21.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* 11(1): 1432.

Raymond O, Gouzy J, Just J, et al. 2018. The *Rosa* genome provides new insights into the domestication of modern roses. *Nature Genetics* 50(6): 772–777.

Rehder A. 1940. *Manual of Cultivated Trees and Shrubs Hardy in North America: exclusive of the subtropical and warmer temperate regions*, 2nd edn. New York: Macmillan. Pp. 426–451.

Ren H, Yang W, Jing W, et al. 2024. Multi-omics analysis reveals key regulatory defense pathways and genes involved in salt tolerance of rose plants. *Horticulture Research* 11(5): uhae068.

Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* 19: 1–10.

Robert-Seilaniantz A, Bari R, Jones JD. 2010. A biotic or abiotic stress? In: Pareek A, Sopory SK, Bohnert HJ, Govindjee (eds.). *Abiotic Stress Adaptation in Plants: physiological, molecular and genomic foundation*. Dordrecht: Springer, 103–122.

Shulaev V, Korban SS, Sosinski B, et al. 2008. Multiple models for Rosaceae genomics. *Plant Physiology* 147(3): 985–1003.

Shulaev V, Sargent DJ, Crowhurst RN, et al. 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* 43(2): 109–116.

Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6(1): 1–11.

Smit AFA, Hubley R. 2008–2015. RepeatModeler Open-1.0. Available at https://www.repeatmasker.org/RepeatModeler/

Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. Available at https://www.repeatmasker.org/RepeatMasker/

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19(Suppl_2): ii215–ii225.

Tian X, Wang Z, Zhang Q, et al. 2018. Genome-wide transcriptome analysis of the salt stress tolerance mechanism in *Rosa chinensis*. *PloS One* 13(7): e0200938.

Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S. 2002. Agricultural sustainability and intensive production practices. *Nature* 418(6898): 671–677.

Tuskan GA, Difazio S, Jansson S, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793): 1596–1604.

Van Fleet W. 1908. Breeding wild roses. Journal of Heredity os-4(1): 240–244.

Van Zelm E, Zhang Y, Testerink C. 2020. Salt tolerance mechanisms of plants. *Annual Review of Plant Biology* 71: 403–433.

Walker BJ, Abeel T, Shea T, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* 9(11): e112963.

Wang Y, Tang H, DeBarry JD, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* 40(7): e49.

Ward LF, Jenney WP. 1899. *The Cretaceous formation of the Black Hills as indicated by the fossil plants*. Washington: US Government Printing Office.

Wissemann V. 2003. Classification: Conventional Taxonomy (Wild Roses). In: Roberts AV, Debener T, Gudin S (eds.). *Encyclopedia of Rose Science*. Oxford: Academic Press, 111–117.

Wissemann V, Ritz CM. 2005. The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS-1 and *atpB–rbcL* intergenic spacer (IGS) versus conventional taxonomy. *Botanical Journal of the Linnean Society* 147(3): 275–290.

Wu T, Hu E, Xu S, et al. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2(3): 100141.

Xiang G, Zhang H, Jian H, et al. 2019. De Novo assembly and characterization of the transcriptome of susceptible and resistant rose species in response to powdery mildew. *Scientia Horticulturae* 257: 108653.

Yang J, Zhang J, Yan H, et al. 2024. The chromosome-level genome and functional database accelerate research about biosynthesis of secondary metabolites in *Rosa roxburghii*. *BMC Plant Biology* 24(1): 410.

Yokoya K, Roberts AV, Mottley J, Lewis R, Brandham PE. 2000. Nuclear DNA amounts in roses. *Annals of Botany* 85(4): 557–561.

Yu TT, Ku TC. 1985. *Rosa* L. In: Yu TT (ed.). *Flora Reipublicae Popularis Sinicae*, Vol. 37. Beijing: Science Press, 360–455.

Zang F, Ma Y, Tu X, et al. 2021. A high-quality chromosome-level genome of wild *Rosa rugosa*. *DNA Research* 28(5): dsab017.

Zhang X, Wu Q, Lan L, et al. 2024. Haplotype-resolved genome assembly of the diploid *Rosa chinensis* provides insight into the mechanisms underlying key ornamental traits. *Molecular Horticulture* 4(1): 14.

Zhang X, Zhang J, Zhang W, Yang T, Xiong Y, Che D. 2016. Transcriptome sequencing and de novo analysis of *Rosa multiflora* under cold stress. *Acta Physiologiae Plantarum* 38: 1–13.

Zhong MC, Jiang XD, Yang GQ, et al. 2021. Rose without prickle: genomic insights linked to moisture adaptation. *National Science Review* 8(12): nwab092.

Zhu JK. 2001. Plant salt tolerance. *Trends in Plant Science* 6(2): 66–71.

Zhuang D, Ma C, Xue L, et al. 2021. Transcriptome and de novo analysis of *Rosa xanthina* f. *spontanea* in response to cold stress. *BMC Plant Biology* 21: 1–11.

Zong D, Liu H, Gan P, et al. 2024. Chromosomal□scale genomes of two *Rosa* species provide insights into genome evolution and ascorbate accumulation. *The Plant Journal* 117(4): 1264–1280.

Table 1. Statistics of the genome assembly and annotation of *Rosa lucieae*.

Feature	Value
Genome assembly	
Estimated haploid size (bp)	458,297,472
Estimated repeat sequences (bp)	273,209,581
Estimated heterozygosity (%)	1.35
Assembled genome size (bp)	544,655,161
Number of chromosomes	7
Total size of chromosomes (bp)	520,416,292
Number of unplaced contigs	187
Total size of unplaced contigs (bp)	24,238,869
Contig N50 (bp)	180,204
GC content (%)	38.71
Gap length (bp)	116,200
Complete eudicots_odb10 BUSCOs (%)	96.17
Complete embryophyte_odb10 BUSCOs (%)	96.78
Complete viridiplantae_odb10 BUSCOs (%)	97.17
Genome annotation	
Repeat sequences (bp)	270,481,675
Number of protein-coding genes	36,876
Average gene length (bp)	3,651.95
Average CDS length (bp)	1,263.19
Average number of exons per gene	5.16
Average exon length (bp)	244.86
Average intron length (bp)	574.38
Gene content (%)	24.12
CDS content (%)	8.55

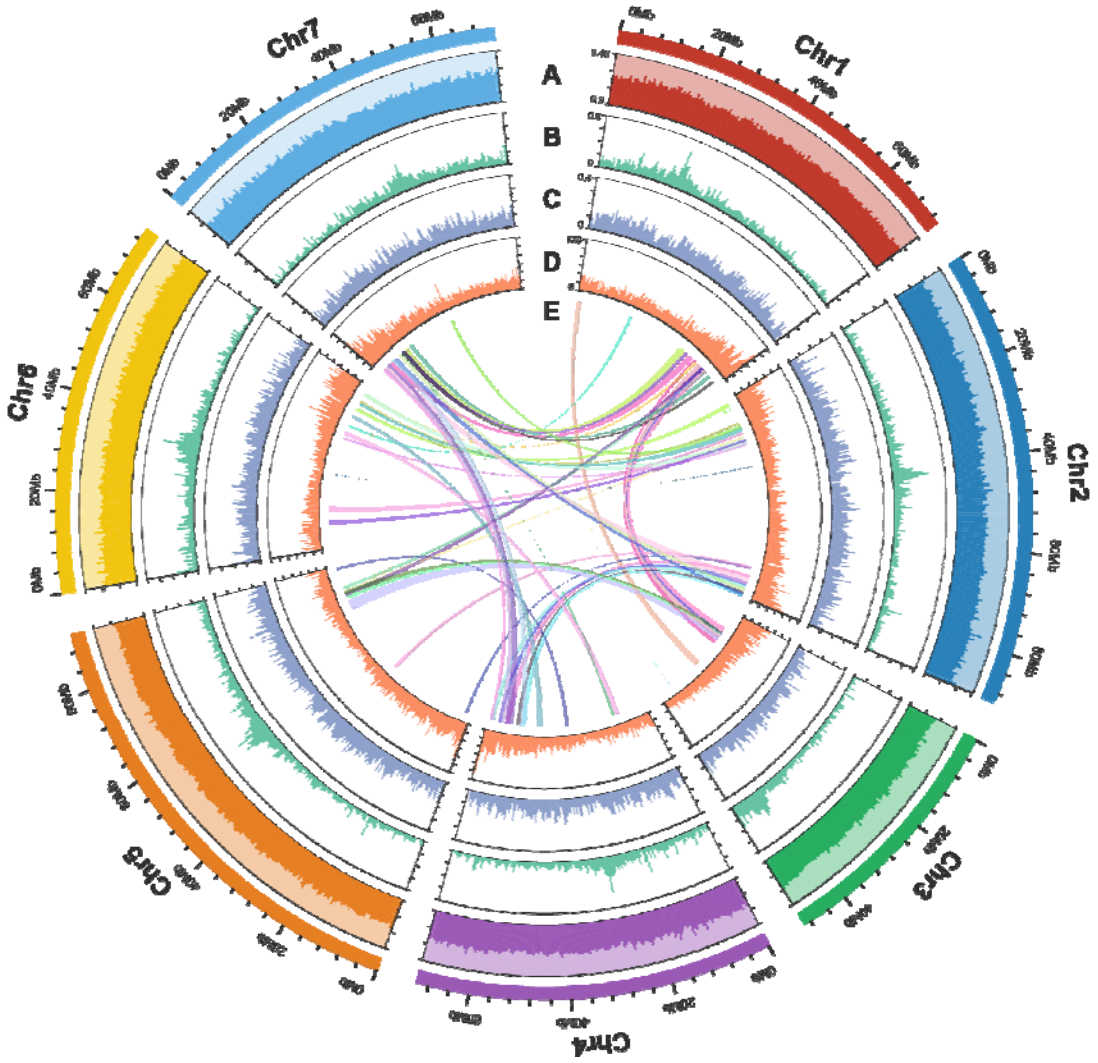


Figure 1. Genome map of *Rosa lucieae*. **A)** GC content, **B)** density of Gypsy, **C)** density of Copia, **D)** density of protein-coding genes, and **E)** collinear relationships within the genome. The GC content and densities of genetic elements were calculated using sliding windows of 500kb without overlaps.

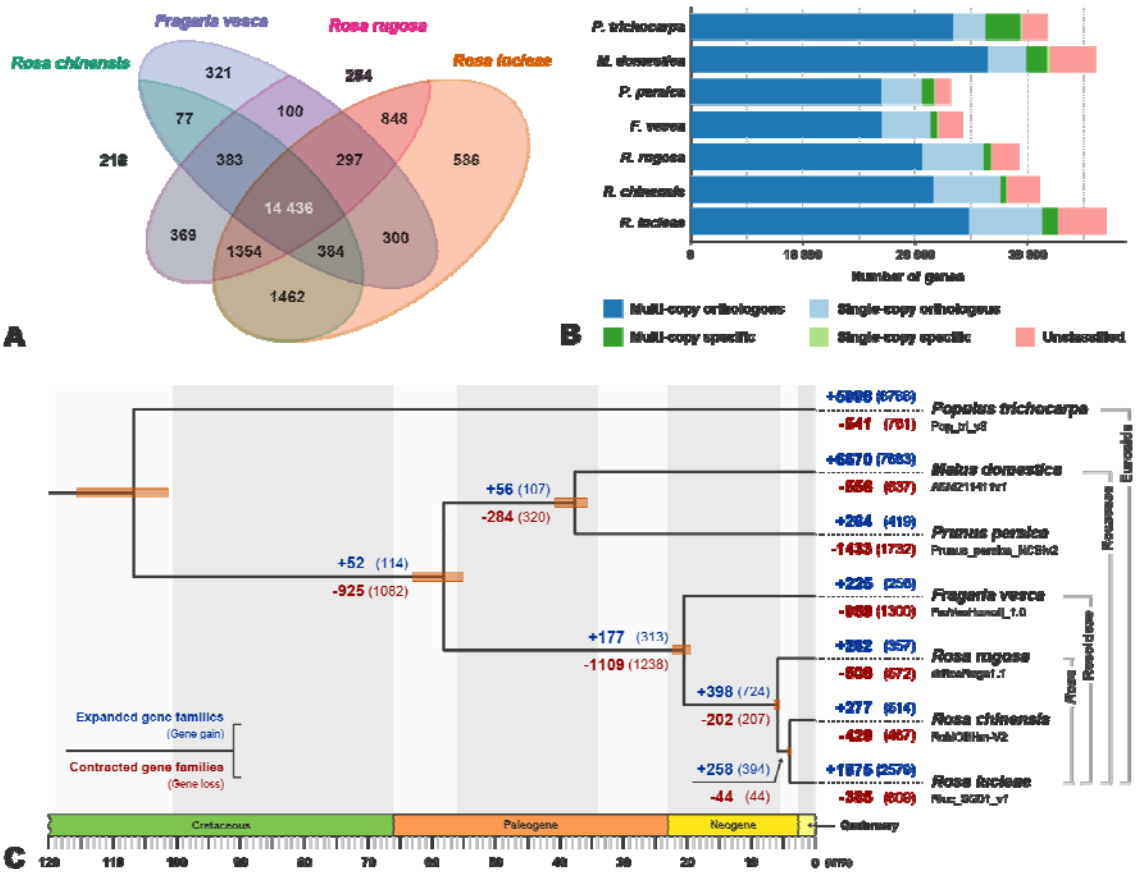


Figure 2. Comparative genomic and evolutionary analyses of *Rosa lucieae*. **A)** Venn diagram of orthologous gene families among *Rosa lucieae*, *Rosa chinensis*, *Rosa rugosa* and *Fragaria vesca*. **B)** Gene family classification profile of protein-coding genes from *R. lucieae* and other close reference species. **C)** Chronogram and gene family evolution of *R. lucieae* and other close reference species. In the Bayesian phylogenetic tree, the posterior probability for every branch was 1.0.

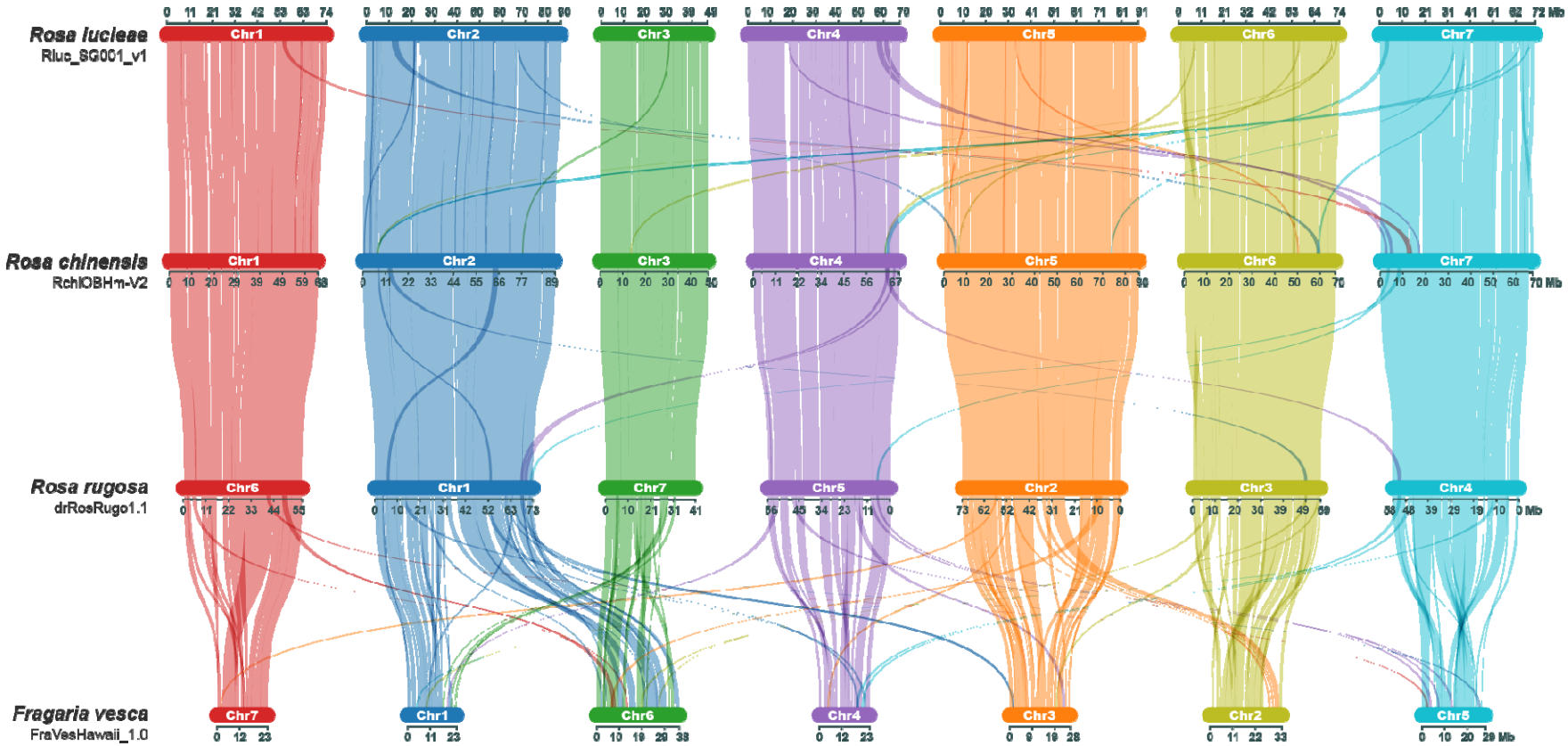


Figure 3. Pairwise genome synteny of *Rosa lucieae*, *Rosa chinensis*, *Rosa rugosa* and *Fragaria vesca*. Each colored link represents the relationships of collinear blocks.

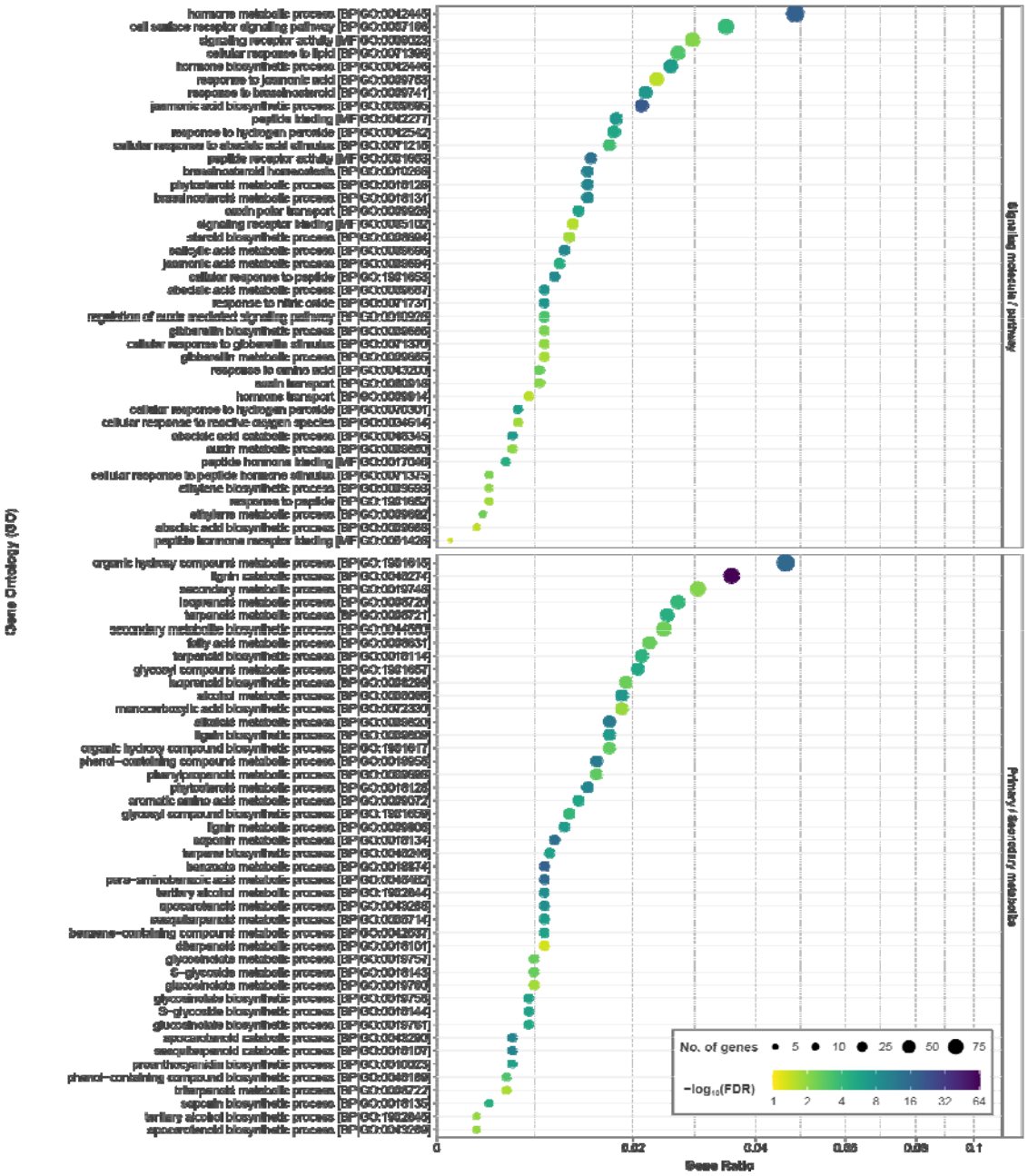
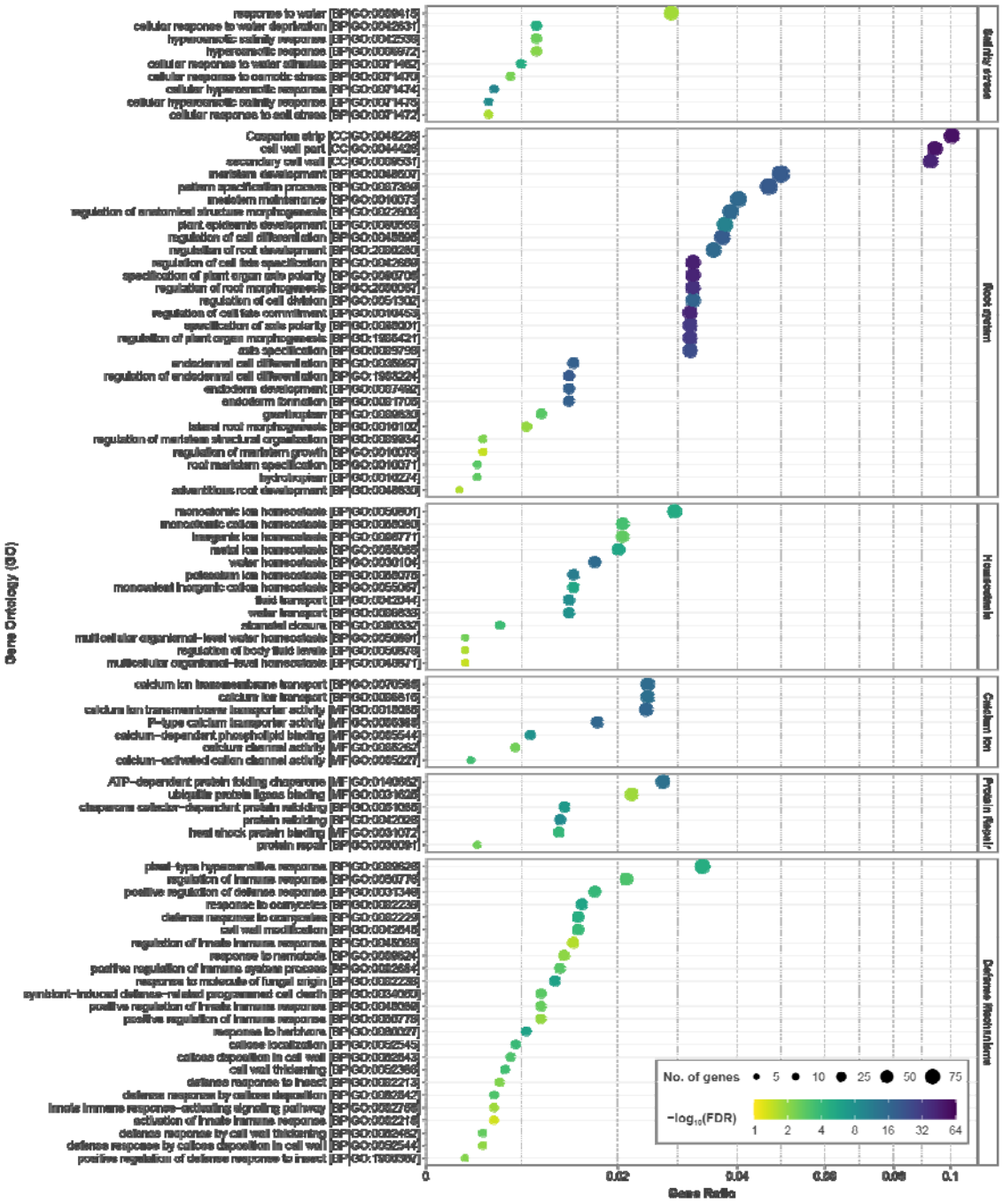


Figure 4. Gene ontology (GO) enrichment analysis on *Rosa lucieae* genes within the rapidly evolving gene families based on gene family evolution analysis. BP: Biological Process; MF: Molecular Function; CC: Cellular Component; FDR: False Discovery Rate.



1

2 **Figure 4.** *Continued*.

3

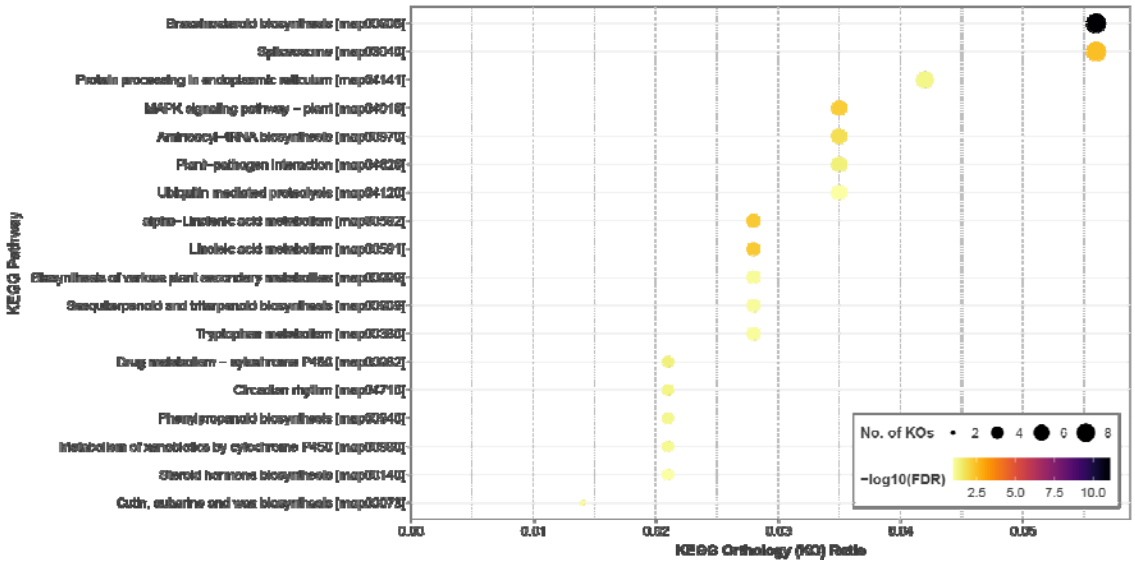


Figure 5. KEGG pathway enrichment analysis on KEGG orthologies of *Rosa lucieae* genes within the rapidly evolving gene families based on gene family evolution analysis. KO: KEGG orthology; FDR: False Discovery Rate.