

Orphan genes are involved in environmental adaptations and flowering process in the rose

Dongna Ma^{1,2#}, Qiansu Ding^{3#}, Zhizhu Zhao⁴, Xiao Han^{1,2} and Jian Mao^{1,2*}

¹ National Engineering Research Center for Cereal Fermentation and Food Biomanufacturing, State Key Laboratory of Food Science and Technology, School of Food Science and Technology, Jiangnan University, Wuxi 214122, China

² Jiangsu Provincial Research Center for Bioactive Product Processing Technology, Jiangnan University, Wuxi 214122, China

³ Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, 55108, USA

⁴ School of Biological Sciences, The University of Hong Kong, Hong Kong, 999077, China

Authors contributed equally: Dongna Ma, Qiansu Ding

* Corresponding author, E-mail: maojian@jiangnan.edu.cn

Abstract

Orphan genes (OGs) are genes with no obvious homology compared to other species and play a key part in new function generation, phenotypic changes, and adaptive evolution. *Rosa chinensis* is an important horticultural variety and the most popular cut flower worldwide, with high economic and ornamental benefits. Herein, 2,586 OGs were identified in the *R. chinensis* genome, accounting for approximately 7.11% of all protein-coding genes. Genetic structure analysis indicated that the OGs had a shorter protein size, fewer exons, lower GC content, and a higher isoelectric point compared to non-orphan genes (NOGs). Transcriptomic analyses revealed that OGs had a stronger tissue-specific expression, with more than 50% specifically expressed in reproductive organs. Weighted gene co-expression network analysis (WGCNA) resulted in 215 OGs distributed in five modules. The co-expression genes of these OGs were engaged in a variety of important biological processes, including photosynthesis, pentose and glucuronate interconversions, linoleic acid metabolism, and phytohormone signaling. It was found that 107 OGs were significantly up- or down-regulated across responses to both abiotic stresses (salt and drought). Fuzzy c-means clustering identified 50 OGs (salt: 22 and drought: 28) had increasing and decreasing expression patterns, suggesting their potential function in interactions with the environmental adaptation. In addition, there were 11 OGs involved in the flowering process of roses. The present study provides the first systematic identification of OGs in *R. chinensis* as well as a comprehensive analysis of the characteristics and potential functions, resulting in valuable clues and new insights into the importance of these new genes.

Citation: Ma D, Ding Q, Zhao Z, Han X, Mao J. 2024. Orphan genes are involved in environmental adaptations and flowering process in the rose. *Tropical Plants* 3: e036 <https://doi.org/10.48130/tp-0024-0036>

Introduction

Genetic diversity due to genetic variation is the cornerstone for ecosystem diversity and species diversity. Genetic variation is a change in the genetic material of an organism that can be passed on to future generations, including chromosomal variation, genetic recombination, genetic mutations, and the emergence of orphan genes (OGs)^[1]. OGs are a group of genes in a taxonomic group for which no obvious homology can be found at the genomic level, also known as 'lineage-specific genes'^[2,3]. The first report of OGs in *Saccharomyces cerevisiae* was released in 1996^[4]. However, in the early stage, researchers knew little about OGs due to the limitation of sequencing ability to obtain the complete genome of many lineages^[5]. Over recent decades, following the speedy advancement of sequencing technologies, complete genomic and transcriptomic sequences of a large number of species have been obtained quickly, accurately, and inexpensively^[6]. Each freshly sequenced genome consists of a portion of OGs, and therefore, OGs in organisms such as the silkworm^[7], sweet orange^[8], and tea plant^[9] have begun to be studied extensively. As research has continued, based on previous work on complete genomes in such species as *Human*, *Drosophila*, and *Arabidopsis thaliana*, researchers have proposed several mechanisms for the origin of OGs

related to gene duplication and subsequent sequence divergence, transposable elements, lateral gene transfer, and *de novo* origination^[10–14].

However, the functions of most OGs are not annotated due to the lack of homologous genes and functional domain information^[15]. Although it is challenging to analyze the biological functions using comparative genomics, some preliminary explorations can be made based on their sequence structural features. OGs have a shorter generation time in comparison to non-orphan genes (NOGs), and therefore, there are some identifiable differences in gene and protein length, exon number, GC content, transcript support, and position preference on chromosomes. In sweet orange, the OGs had a lower average number of exons per gene, both shorter genes and protein length, higher GC content, as well as priority distribution on certain chromosomes, in comparison to NOGs^[8]. Similar results were found in *A. thaliana*^[15], *Drosophila*^[16], *Poaceae*^[17], zebrafish^[18], and tea plant^[9].

A considerable amount of research has now proven that OGs have a crucial impact on growth and development^[19,20]. RNA interference of 200 OGs in *Drosophila melanogaster* induced down-regulation of their expression and was found to result in lethality mainly concentrated in the metaphase^[21]. In plants,

transgenic transfer of the OG (QQS), which is responsible for influencing the process of carbon and nitrogen segregation of proteins and carbohydrates in *A. thaliana*, into soybean was found to affect the protein content of soybean seeds similarly^[22]. It is noteworthy that tissue-specific expression of OGs exists, with a preference in animals for expression in male reproductive tissues such as the testis^[13], exemplified by the species-specific chimeric gene *Sdic1* encoding the sperm-specific kinesin intermediate chain in *D. melanogaster*^[23]. A similar phenomenon is present in plants, where tissue-specific expression of OGs tends to be reflected in the male reproductive system, for instance, in pollen. By way of exemplification, the orphan protein encoded by the *Ms2* gene caused male sterility in wheat, barley, and phragmites^[24]. Besides, OGs have been shown to play an essential part in environmental adaptation. Drought-induced over-expression of the OG (UP12_8740) in cowpea increased its tolerance to osmotic stress and soil drought^[25]. Similarly, preferential expression of OGs under abiotic stress was reported in water flea^[26], *A. thaliana*^[27] and rice^[28].

The Chinese rose (*Rosa chinensis*) is an important horticultural crop and one of the most commonly cultivated ornamentals in the world, with considerable economic value in potted and cut flowers. The bright colors of roses are visually attractive, and therefore, *R. chinensis* is often used as a suitable demonstration plant for the production of new flowering varieties^[29]. *R. chinensis* is also the dominant species used for hybridization and was introduced to Europe in the 18th century as a parent of modern roses^[30]. In the present study, the published *R. chinensis* genome^[31] was used to identify OGs and a comprehensive analysis was performed, including sequence structural features, subcellular localization, gene duplication, and chromosome distribution. In addition, a weighted gene co-expression network analysis (WGCNA) was built to project the functions of these identified OGs. The differences in the expression of OGs at different times under salt and drought stress were also analyzed to explore the adaptation of OGs to the environment. In conclusion, the present results provide valuable clues to reveal the evolution, characterization, and environmental adaptation of OGs in *R. chinensis*.

Materials and methods

Data collection

Genomes and annotation information for *R. chinensis* were obtained from the *Rosa* database (<https://lipm-browsers.toulouse.inra.fr/pub/RchiOBHm-V2/>). Predicted proteins of *Rosa rugosa* were downloaded from the eplant (<http://eplant.njau.edu.cn>). All other *Rosaceae* predicted proteins were downloaded from the NCBI Datasets (www.ncbi.nlm.nih.gov/datasets). Unique transcripts (PUT) assembled from plant mRNA sequences were downloaded from PlantGDB (<https://goblinp.luddy.indiana.edu/prj/ESTCluster>), and 122 plant genomes predicted proteins were extracted from Phytozome (<https://phytozome-next.jgi.doe.gov>), and 77 from eplant (<http://eplant.njau.edu.cn>) were downloaded. UniProtKB was obtained from UniProt (www.uniprot.org/uniprotkb) and NR database were downloaded from NCBI (<ftp://ftp.ncbi.nlm.nih.gov>), individually.

For the prediction of the potential function of OGs, we collected RNA-seq data from publicly available materials to acquire the gene expression levels. Such data consisted of

different tissues or varying stress of *R. chinensis*. The transcriptome data was downloaded from the National Center for Biotechnology Information database (NCBI; www.ncbi.nlm.nih.gov) with BioProject accession number PRJNA546486 (leaf, stem, root, pistil ovary, prickle, and stamen), PRJNA722055 (leaf imposed to drought stress: 0, 30, 60, and 90 d) and PRJNA587482 (root imposed to salt stress: 0, 2, 24, and 48 h).

Identification of OGs

The advancement of comparative genomics has led to a much more improved investigation of the origin and evolution of OGs. A homolog-based search was performed in a pipeline to identify *R. chinensis* within OGs (Fig. 1). Initially, *R. chinensis* protein sequences were scoured against the *R. rugosa* proteome with the BLASTP. Any *R. chinensis* protein sequence with BLASTP hit with an E-value cutoff of 1e-5 was discarded once available. Homology searches were then conducted with genomes of other *Rosaceae* plants, Phytozome, eplant, Plant-PUTs database, Uniprot-KB database, and NR database sequentially with an E-value cutoff of 1e-5. Finally, the genes without homologs in any databases were the OGs^[15,32], whereas all the alternatives with homologous were non-orphan genes (NOGs).

Genic features

To visualize the structural characteristics of the OGs, a genome-wide profile of *R. chinensis* was applied. DAMBE7 software was employed to evaluate the isoelectric points of OGs and NOGs^[33]. Discrepancies among OGs and NOGs, including gene size, length of the protein, size of the exons and introns, number of the exons, and content of GC were computed with the use of in-house Python scripts. The Wilcoxon rank-sum test was then used to identify the significant difference across distinct groups for OGs and NOGs. Information on chromosome localization was retrieved from chromosome sequences and plotted using MapGene2Chrom (http://mg2c.iask.in/mg2c_v2.0). A final BUSCA (Bologna Unified Subcellular Component Annotator) was applied to predict the OGs subcellular localizations^[34].

Gene duplication analysis

Based on previous studies, there are varied models interpreting the origin of OGs^[3,14], of which gene duplication has been considered the dominant mechanism underlying the emergence of OGs^[35]. The present work started with a BLASTP search for homologous genes with an E-value cutoff of 1e-8, followed by the identification of different types of gene duplication using MCscanX which was capable of detecting WGD, tandem duplication, proximal duplication, transposon duplication, and dispersed duplication^[36].

Gene expression analysis

To analyze tissue expression, growth and development, and ability to adapt to the environment of *R. chinensis*, the transcriptome data was downloaded. The raw RNA-seq data was in turn filtered using the Trimmomatic program^[37]. With the aim of identifying differentially expressed genes (DEGs) between treatments, we used the default settings of the abundance_estimates_to_matrix.pl, run_DE_analysis.pl (DESeq2) and analyze_diff_expr.pl modules of the Trinity package. Significant differences in gene expression were ascertained using the $|\log_2FC| \geq 1$ with a false discovery rate (FDR) < 0.05 as thresholds, and RSEM implemented in the Trinity package was employed to calculate FPKM (fragments per kilobase of exon per million fragments mapped)^[38]. A cluster analysis was

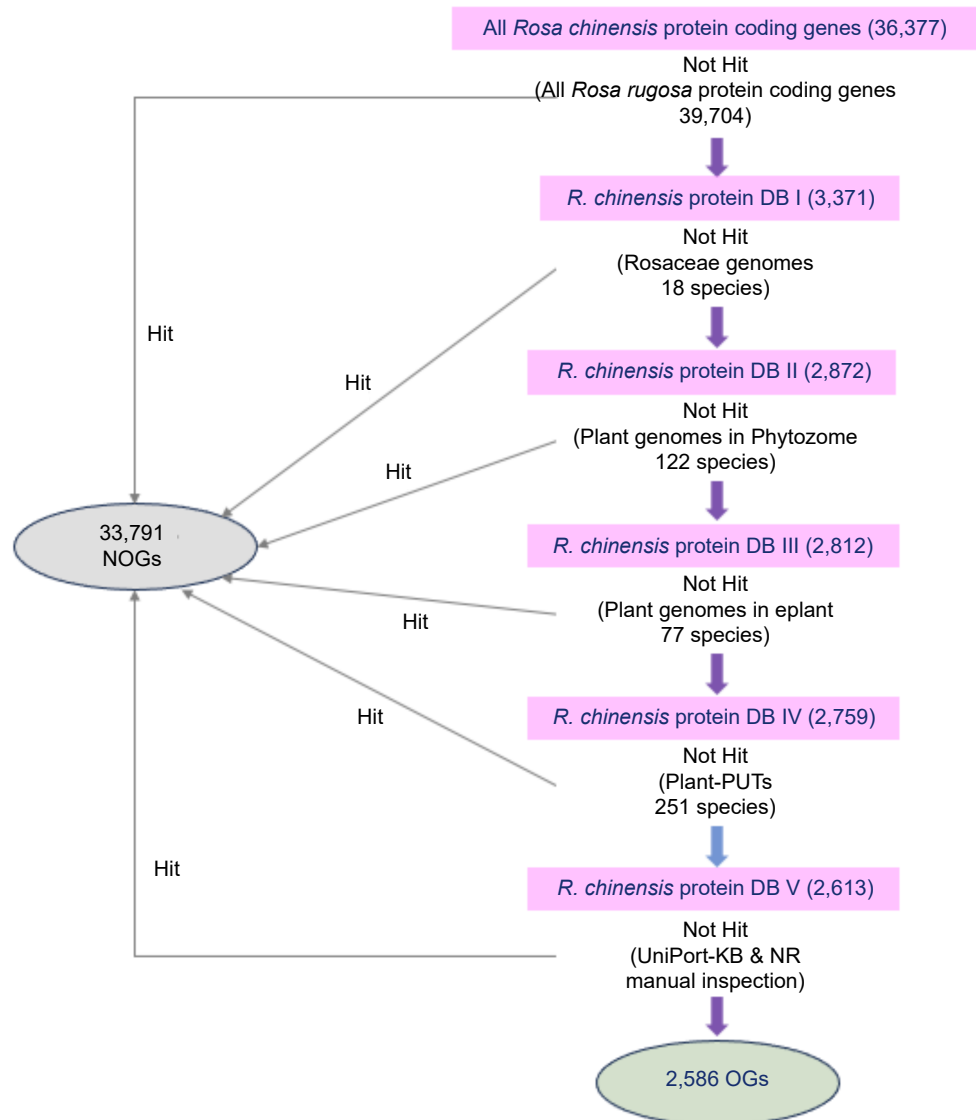


Fig. 1 Procedure for identifying the orphan genes in *R. chinensis* genome. The purple arrows represent a homolog-based search by BLASTP with an E-value cutoff of $1e-5$. The blue arrow represents a homolog-based search by TBLASTN with an E-value cutoff of $1e-5$. Genes without homologs in any databases were identified as OGs (2,586), while genes with homologs were classified as NOGs (33,791).

conducted using R software targeting particular expressed genes based on RNA-seq data and following functional validation was selected. It was assumed that genes with FPKM value > 0.02 were already expressed^[39]. Besides, genes that were exclusively expressed in specific tissue were determined by PaGeFinder software with specificity measure (SPM)^[40], and once the SPM value was ≥ 0.9 , the gene in that tissue was identified specifically.

Weighted gene co-expression network analysis (WGCNA) and function annotation

Following the discard of genes with FPKM < 1 , WGCNA was constructed and the genes were grouped into modules aided by the WGCNA package in R software^[41]. The automatic network builder function block-wise Modules were utilized to construct the network with default parameters. After that, eigengene values were computed for individual modules for every tissue and the module with the highest correlation coefficient, while fulfilling a p -value < 0.05 , was picked to serve as

the tissue-specific module for further analysis. The candidate with the most significant representative gene in each module was assumed to be the module eigengene. Module membership^[10] and gene importance^[32] were calculated for every ME in each tissue-specific module, and once $MM > 0.95$ and $GS > 0.85$, the gene was deemed to be the central gene of the module. KEGG enrichment analysis was carried out on an online platform, OmicShare (www.omicshare.com).

Results

Identification of OGs

The OGs of *R. chinensis* were characterized according to the methodology employed in recent studies (Fig. 1) with the recently published database resources^[15,42,43]. A total of 36,377 annotated protein-coding genes within the *R. chinensis* genome were used for BLASTP with all *R. rugosa* protein-coding genes (39,704) presented in this study. During this procedure,

there were a total of 33,006 genes with significant similarity (E -value $< 1e-5$), and 3,371 genes (DBI) were kept for the follow-up analysis. The NOGs indicating homology were eliminated, and a further scan of the remaining genes were carried out alongside the published genomes of the *Rosaceae* family. Altogether 2872 genes (DBII) were kept for further analysis at this step. The NOGs displaying homology were dropped and the remaining genes were matched with 122 plant genomes in Phytozome for an additional search, yielding 2,812 genes being retained for the following step (DBIII). After comparing these 2,812 genes with the 77 plant genomes in the eplant database, 2,759 genes were not found as homologs (DBIV). Of these 2,759 genes subsequently matched against 251 PlantGDB-assembled Unique Transcripts (PUTs) sequences and no homologs were found for 2,613 genes (DBV). The last step, was to completely erase the impact of false positives from the analysis, leftover genes were examined in the context of UniProt-KB and NR databases, an action that finally left 2,586 genes. These leftover 2,586 genes were labelled as OGs in the *R. chinensis* genome, representing 7.11% of the entire genome of *R. chinensis* (Supplementary Table S1), as opposed to these remaining 33,791 genes whose similarity to the databases was defined as NOGs.

Features of OGs

Aiming at clarifying any significant differences between OGs and NOGs, the analysis highlighted and compared the sequence structural features between 2,586 OGs and 33,791 NOGs found in the present study. The results showed a significantly smaller gene size (Wilcox rank sum test, $p < 2.2e-16$) and protein size (Fig. 2a, Wilcox rank sum test, $p < 2.2e-16$) for OGs than for NOGs (Table 1), where 1,512.34 bp was the OGs gene size and 72.5 amino acids (aa) was the OGs protein size, 2868.43 bp for NOGs gene size and 378.4 aa for NOGs protein size, which suggested that the NOGs protein was 5.22-fold lengthening than OG protein (Table 1). An in-depth analysis on the structural components of the genes revealed that the shorter protein size was predominantly attributed to the lower number of exons (Fig. 2b, Wilcox rank sum test, $p < 2.2e-16$). The exon size (Fig. 2c, Wilcox rank sum test, $p = 7.789e-15$) and intron size (Fig. 2d, Wilcox rank sum test, $p < 2.2e-16$) of OGs, however, were both remarkably bigger than NOGs. The GC content and the isoelectric point of OGs were further comparatively examined with those of NOGs. It was shown that the GC contents of OGs were lower than that of NOGs (Table 1, Fig. 2e, Wilcox rank sum test, $p < 2.2e-16$). In contrast, the isoelectric point of OGs (8.53) was higher than that of NOGs (7.42) (Table 1, Fig. 2f, Wilcox rank sum test, $p < 2.2e-16$). In general, these results suggested that there existed significant differences in genetic features of OGs and NOGs.

For the analysis of OGs' genomic distribution, the OGs were plotted over the chromosomes of *R. chinensis* based on the available information from genome annotation (Supplementary Table S1). A total of 2560 OGs were spread over seven chromosomes. A maximum number of OGs per chromosome was in the order of Chr2 (465), Chr5 (423) and Chr1 (395). The percentages of OGs distributed on the seven chromosomes were 7.97%, 7.01%, 6.84%, 7.37%, 6.91%, 6.41%, and 7.29%, respectively (Fig. 3b), demonstrating that there was no chromosomal preference in the distribution of the OGs in *R. chinensis*. Besides, the distribution density of OGs was higher near the telomeres, and the distribution was relatively balanced on the chromosomes apart from the aggregation phenomenon in some chromosomal regions (Fig. 3c). Generally, the spread of OGs on these seven chromosomes was reasonably uniform.

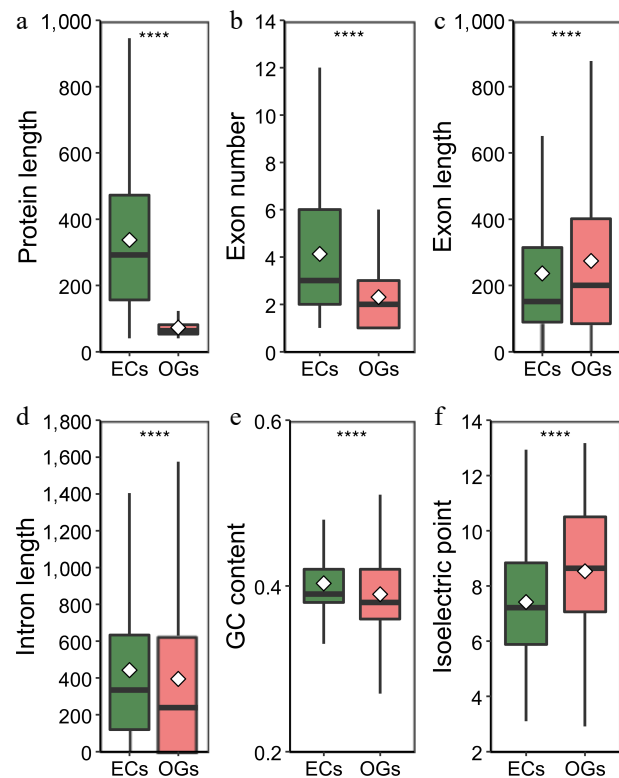


Fig. 2 Analysis and comparison of the structural characteristics of orphan genes (OGs) and non-orphan genes (NOGs). (a) Box-plot comparisons of protein length. (b) Exon number per gene. (c) Exon length. (d) Intron length. (e) GCs content. (f) Isoelectric point. White squares represent the mean value. **** indicate significance levels at $p < 0.0001$.

Table 1. Genic features of orphan genes (OGs) compared with non-orphan genes (NOGs).

Items	OGs		NOGs		Wilcox rank sum test probability
	Mean (SE)	Median	Mean (SE)	Median	
Gene size (bp)	1,512.34 (1,736.26)	880	2,868.43 (2,590.26)	2,239	$< 2.2e-16$
Protein size (aa)	72.5 (32.17)	63	378.4 (306.51)	306	$< 2.2e-16$
Exons per gene	2.34 (1.99)	2	4.82 (4.74)	3	$< 2.2e-16$
Exon size (bp)	347.84 (385.78)	225.5	322.22 (423.62)	164	7.789e-15
Intron size (bp)	571.3 (998)	270	556.31 (731.19)	356	$< 2.2e-16$
Gene GC content (%)	39 (4.85)	38.05	40.31 (4.06)	39.46	$< 2.2e-16$
CDS GC content (%)	41.9 (6.47)	41.46	45.08 (4.54)	44.35	$< 2.2e-16$
Isoelectric point	8.53 (2.34)	8.64	7.42 (1.94)	7.21	$< 2.2e-16$

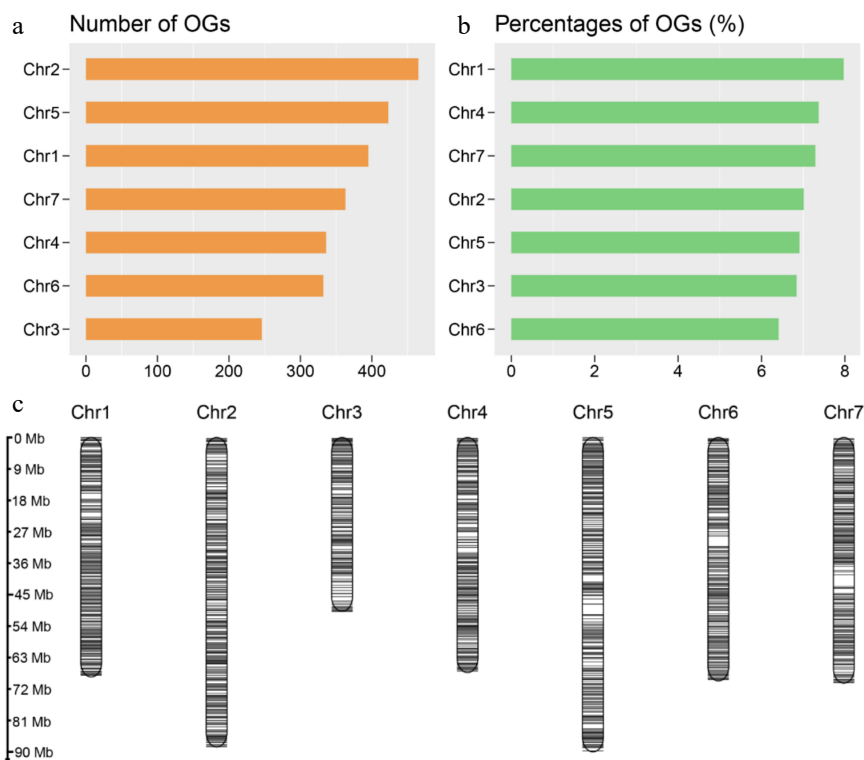


Fig. 3 Orphan genes (OGs) distribution on chromosomes. (a) The numbers of OGs on each chromosome of *R. chinensis*. (b) Percentage of OGs on each chromosome of *R. chinensis*. (c) Chromosomal distribution of the identified OGs. Black horizontal lines represent OGs.

Subcellular localization and gene duplication

Protein functions typically are, to a certain extent, inferred from their subcellular localization. In this study, among the 2,586 OGs that were identified, 788 were positioned in extracellular space, 652 in the chloroplast, 534 in the nucleus, 261 on the organelle membrane, 232 in the endomembrane system, 86 on the plasma membrane, 27 in the mitochondria, and all but eight in the chloroplast thylakoid lumen (Fig. 4).

The origination of OGs is essential in the evolution of the genome. OGs often arise from a combinatorial mix of diverse mechanisms of origin. While gene duplication is believed to be the predominant model for the origin of OGs, the gene duplication theory consists primarily of the generation of the new gene *via* differentiation following duplication. 2,586 OGs derived from the genome of *R. chinensis* were detected in this study, including 274 OGs derived from gene duplication, accounting for approximately 10.6% of the whole OGs ((Supplementary Table S2). There were altogether four OGs originating from whole-genome duplication (WGD). Besides, the number of OGs resulting from tandem duplication, transposed duplication, proximal duplication, and dispersed duplication were 10, 4, 12, and 244 (Supplementary Table S2), respectively.

Expression profiles of OGs

The gene expression pattern for one gene across different tissues must enlighten the corresponding biological function. Preceding transcriptomic data on six tissues subjected to normal growth conditions were reanalyzed. It was found that the transcriptional data contained 2,088 (80.74%) OGs and 29,572 (87.51%) NOGs with FPKM > 0.02. Typically, the expressed amount of OGs is relatively lower than that of NOGs.

Of these, in pistil ovary and stamen, there were the most OGs expressed (Fig. 5a). In addition, 1,139 OGs were identified to be represented across all six tissues (FPKM > 2 in a minimum of one tissue), and 346 OGs were hyper-expressed in all six tissues (FPKM > 2 in all of them). The amount of actively expressed OGs (GeneSpring normalized expression value > 0) followed a generally parallel trend throughout the six tissues and the highest expression level in stamen (Fig. 5b). It was further observed that 214 OGs were specifically expressed in six tissues, out of which 24 were specifically expressed in root, 20 in the stem, 29 in leaf, 21 in prickle, 90 in stamen, and 30 in the pistil ovary (Fig. 5c), such genes may have specific roles within the corresponding tissues. It was obvious that OGs showed more potential expression in stamen (Fig. 5d). A tissue preference for the expression of most OGs was seen according to the expression abundance in each tissue (Fig. 6).

Functional inference of OGs

As the function of OGs cannot be inferred from homologous genes, however, OGs were exclusively expressed in diverse tissues (Fig. 5). The potential functions of OGs were further profiled employing WGCNA, a tool for determining synergistic gene modules. Fifteen modules were defined. Considering different tissues as traits, the modules associated with the optimization of characteristic vector genes and phenotypes were filtered and mapped the heat map of module-trait relationships. Five modules were eventually settled on that had an extremely strong positive correlation to the trait (Fig. 7a). Subsequently, 5,218 hub genes were screened and confirmed in five modules, consisting of 217 OGs. In MEblue (leaf), there were 1,454 hub genes, containing 31 OGs. In the MEbrown model (root), 1,436 hub genes were present, covering 49 OGs.

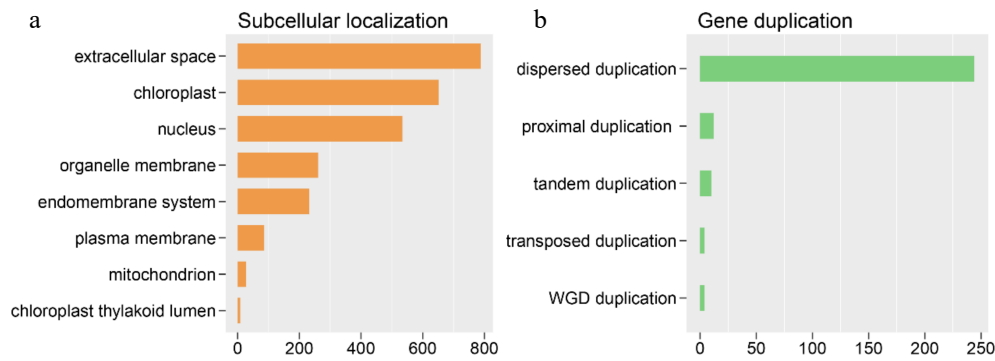


Fig. 4 Orphan gene (OGs) subcellular and gene duplication analysis. (a) OGs assigned to different subcellular locations. (b) The OGs number of different duplication types.

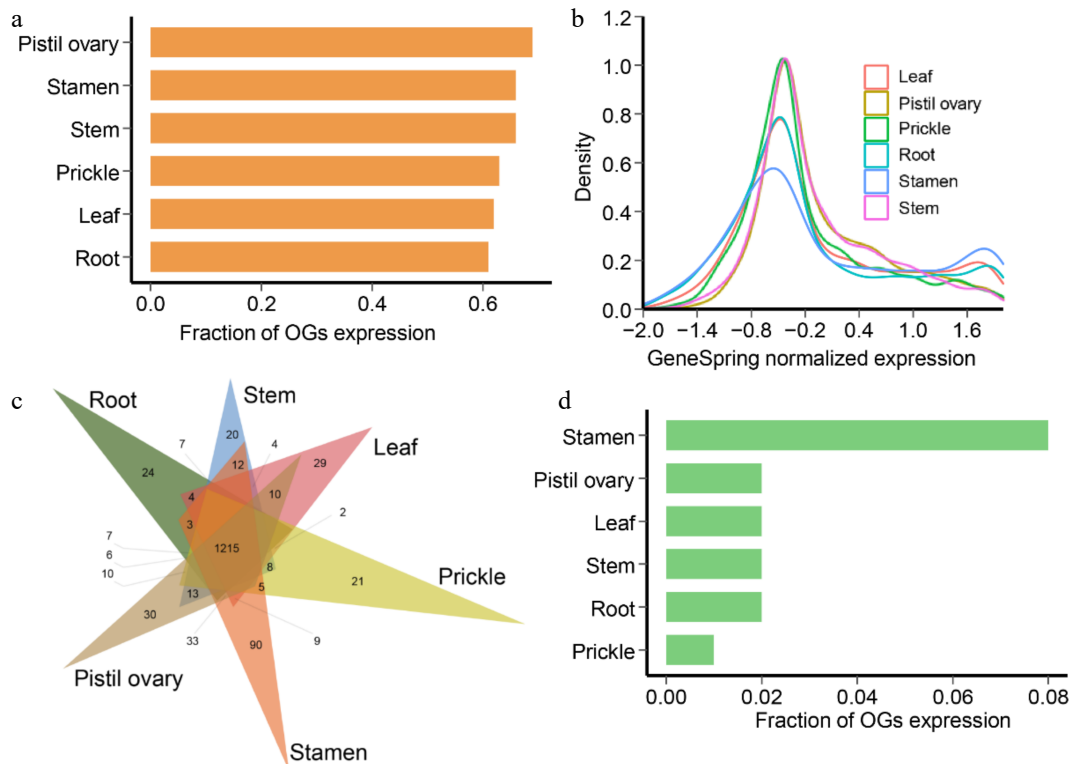


Fig. 5 Gene expression patterns of *R. chinensis* orphan genes (OGs). (a) Fraction of OGs having expression in different tissues. (b) GeneSpring normalized expression levels of OGs in different tissues. (c) Fraction of OGs having tissue-specific expression in different tissues of adult stage. (d) Venn diagram showing the number and relationships of expressed OGs in root, stem, leaf, prickle, stamen, and pistil ovary.

In the MEgreen model (pistil ovary), a total of 250 hub genes existed, including 11 OGs. Among the MEdred model (stem), with 159 hub genes that included one OG. In the METurquoise model (stamen), as many as 1,919 hub genes were found, comprising 125 OGs (Supplementary Table S3). These five modules were followed up with an analysis of KEGG enrichment immediately (p -value < 0.05). In MEblue (leaf), it was predominantly enriched in photosynthesis (ko00195), porphyrin, and chlorophyll metabolism (ko00860), carbon fixation in photosynthetic organisms (ko00710), carotenoid biosynthesis (ko00906), and plant hormone signal transduction (ko04075). In the MEdbrown model (root), it was primarily affluent in glutathione metabolism (ko00480), phenylpropanoid biosynthesis (ko00940), MAPK signaling pathway (ko04016), and plant hormone signal transduction (ko04075). In the MEgreen model

(pistil ovary), linoleic acid metabolism (ko00591), alpha-Linolenic acid metabolism (ko00592), and plant hormone signal transduction (ko04075) were enriched. In the MEdred model (stem), it was mainly enriched SNARE interactions in vesicular transport (ko04130) and isoflavonoid biosynthesis (ko00943). In the METurquoise model (stamen), pentose and glucuronate interconversions (ko00040), phosphatidylinositol signaling system (ko04070), glycerophospholipid metabolism (ko00564), glycolysis/gluconeogenesis (ko00010), galactose metabolism (ko00052) and ether lipid metabolism (ko00565) (Fig. 7b).

In addition, with the aim of probing the potential link between OGs and environmental adaptation, the expression of OGs were reanalyzed in roots under salt stress (0, 2, 24, and 48 h under salt stress) and leaves under drought stress (0, 30, 60, and 90 d under drought stress) using published RNA-seq data. At

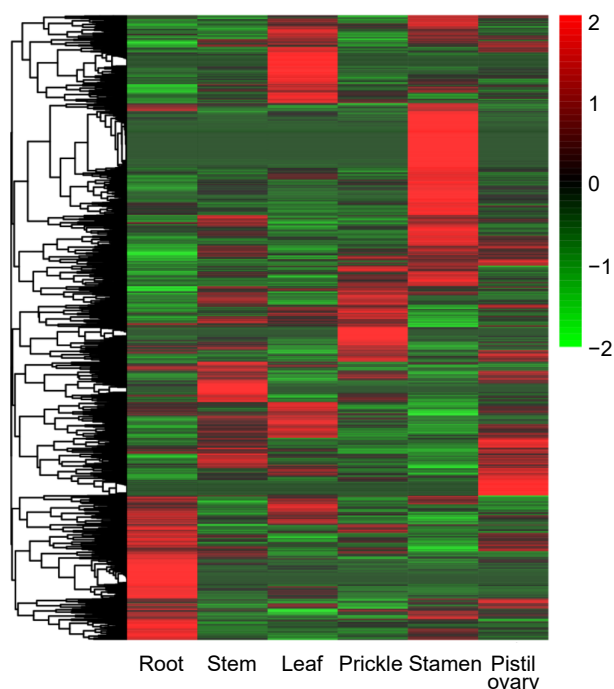


Fig. 6 Expression pattern of orphan genes in different tissues includes root, stem, leaf, prickle, stamen, and pistil ovary of *R. chinensis*.

salt treatment, compared with the control group (0 h under salt stress), 21 (up-regulated: 7, down-regulated: 14), 116 (up-regulated: 56, down-regulated: 60) and 56 (up-regulated: 91, down-regulated: 39) identified OGs in 2 h vs 0 h, 24 h vs 0 h, and 48 h vs 0 h were differentially expressed, correspondingly (Supplementary Table S4). A total of 201 OGs overlapped (201/2586, 7.78%) that were salt responsive in roots, of which 112 OGs (112/201, 55.72%) were up-regulated and 89 OGs (89/201, 44.28%) were down-regulated (Fig. 8a). Fuzzy c-means clustering analysis of all salt-associated DEGs (including OGs and NOGs) was further divided into six Clusters of gene co-expression patterns (Fig. 8b). Two Clusters were focused on with increasing (Cluster 2), and decreasing trends (Cluster 4) of gene expression levels with increasing salt treatment time. Genes with Memberships > 0.7 in Cluster 2 and Cluster 4 were screened for subsequent functional enrichment analysis. There was a total of 440 DEGs in Cluster 2, including 15 OGs, and 336 DEGs in Cluster 4, including 7 OGs (Supplementary Table S5). Nineteen OGs were engaged and enriched in GO terms, consisting 'cellular hyperosmotic salinity response', 'cellular response to hydrogen peroxide', and 'response to salt stress' (p -value < 0.01) (Supplementary Table S6). KEGG enrichment results showed that ABC transporters and brassinosteroid biosynthesis pathway were remarkably enriched (p -value < 0.05) (Table 2).

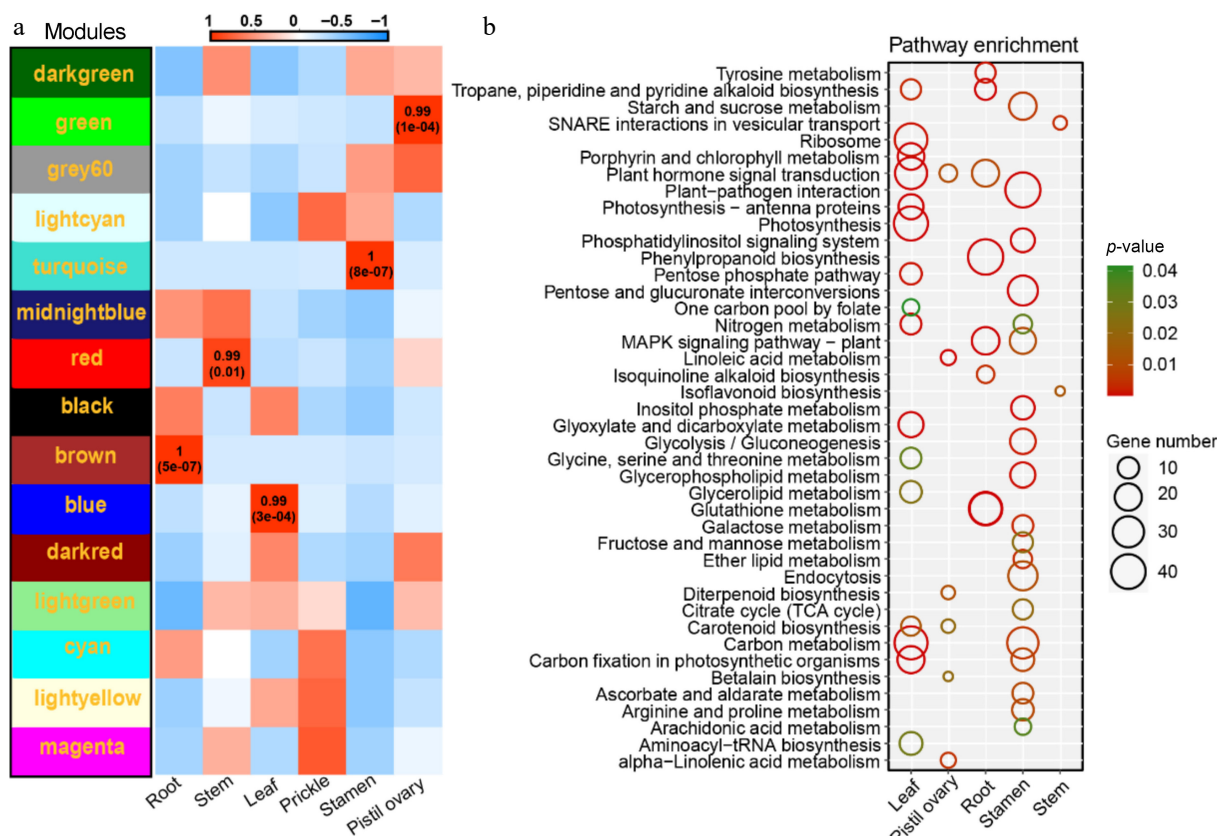


Fig. 7 Co-expression network analyses. (a) Heat map of module-tissue relationship. (b) KEGG enrichment analysis of five tissue-specific modules, include KEGG enrichment analysis result of MEblue module genes (leaf). KEGG enrichment analysis result of MEgreen module genes (pistil ovary). KEGG enrichment analysis result of MEbrown module genes (root). KEGG enrichment analysis result of METurquoise module genes (stamen). KEGG enrichment analysis result of MERed module genes (stem).

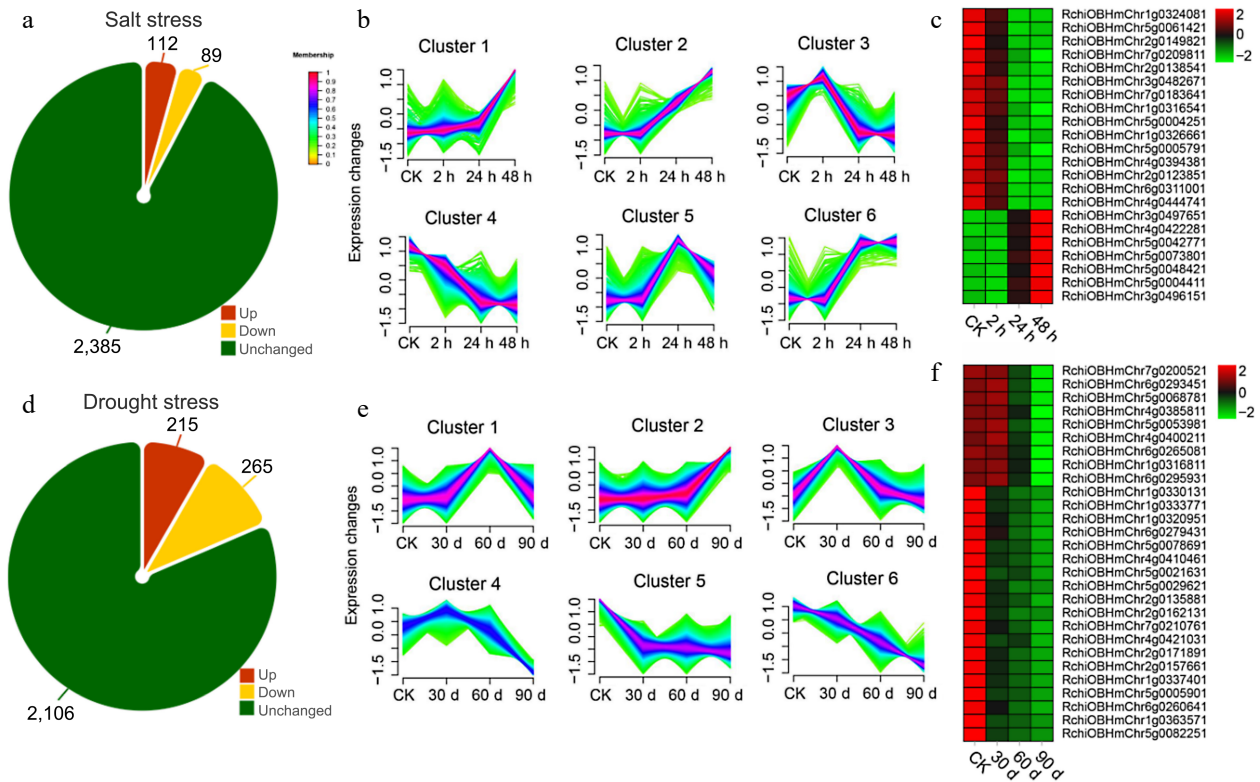


Fig. 8 Transcriptome analysis of orphan genes (OGs) under salt and drought stress. (a) Number of differentially expressed OGs under salt stress in leaves of *R. chinensis*. (b) Trends in the expression of differentially expressed genes at different time points under salt stress. (c) Heat map of the expression of OGs under the trend of pattern Cluster 2 and Cluster 4. (d) Number of differentially expressed OGs under drought stress in roots of *R. chinensis*. (e) Trends in the expression of differentially expressed genes at different time points under drought stress. (f) Heat map of the expression of OGs under the trend of pattern Cluster 1 and Cluster 3.

Table 2. Enriched KEGG pathway for *R. chinensis* orphan genes.

Type	Cluster	KEGG pathway	p-value
Salt stress	Cluster 2	Alanine, aspartate and glutamate metabolism	0.011861
		Base excision repair	0.013442
		Pentose and glucuronate interconversions	0.016805
	Cluster 4	Ubiquinone and other terpenoid-quinone biosynthesis	0.032602
		Glutathione metabolism	0.006557
		ABC transporters	0.007274
		Brassinosteroid biosynthesis	0.014615
		MAPK signaling pathway	0.016903
		Terpenoid backbone biosynthesis	0.023847
		Glycine, serine and threonine metabolism	0.035604
Drought stress	Cluster 1	Carbon metabolism	4.02E-06
		Ubiquinone and other terpenoid-quinone biosynthesis	5.04E-05
		Carbon fixation in photosynthetic organisms	5.22E-05
		Riboflavin metabolism	0.000115
		Pentose phosphate pathway	0.00284
		Glyoxylate and dicarboxylate metabolism	0.008393
		Steroid biosynthesis	0.03522
		DNA replication	0.036059
		Thiamine metabolism	0.043981
		Cluster 3	Pyrimidine metabolism
Aminoacyl-tRNA biosynthesis	0.019032		
Purine metabolism	0.037451		

Upon drought treatment, in contrast to the control group (0 day under drought stress), 101 (up-regulated: 46, down-regulated: 55), 204 (up-regulated: 76, down-regulated: 128) and 352 (up-regulated: 152, down-regulated: 200) OGs in 30 d vs 0 d, 60 d vs 0 d, and 90 d vs 0 d were differentially expressed, respectively (Supplementary Table S7). Altogether, 480 OGs overlapped (480/2586, 18.56%) in response to drought in roots, with 215 OGs (215/480, 44.79%) being up-regulated and 265 OGs (265/480, 55.21%) being down-regulated (Fig. 8d). An analysis of fuzzy c-means clustering centered on two Clusters (Cluster 1: increasing and Cluster 3: decreasing) of gene expression levels as the duration of drought treatment increased. Individual genes with Memberships > 0.7 in Cluster 1 and Cluster 3 were filtered out for follow-up functional enrichment analysis. 344 DEGs, including nine OGs, were enriched in Cluster 1, and 348 DEGs in Cluster 3 including 19 OGs (Supplementary Table S8). A total of 28 OGs were engaged and enriched in GO terms, namely 'photosynthesis', 'reductive pentose-phosphate cycle' and 'regulation of defense response' (p -value < 0.01) (Supplementary Table S9). The results of KEGG enrichment revealed that carbon metabolism and pentose phosphate pathway were markedly enriched (p -value < 0.05) (Table 2). Strikingly, for 107 OGs a response to both categories of stresses was also discovered (Supplementary Table S10), indicating that these genes likely function significantly in stress tolerance.

Petal expansion is the principle procedure by which roses open. Interestingly, it was found that 158 OGs were differentially expressed during petal expansion from the results of Han

Orphan genes in *R. chinensis*

et al. (Supplementary Table S11)^[71]. In the results of Han's study, they used WGCNA to identify two modules MEyellow and MEgreenyellow that may be involved in adaxial–abaxial regulation in rose petals. In the MEyellow model, with 383 hub genes, comprising nine OGs. Within the MEgreenyellow model, a total of 52 hub genes were identified, covering two OGs (Supplementary Table S12). Functional analysis of the hub genes of both modules showed that they mainly include transcription factors such as MYBs and WUSCHEL. They also encode various enzymes such as laccase, cellulose synthase, and trehalose-6-phosphate synthase. All in all, 11 OGs were identified that may participate in adaxial–abaxial regulation and have a significant effect on the flowering open procedure.

Discussion

Identification of OGs in *R. chinensis* genome

The burgeoning field of comparative genomics has accelerated the exploration of the emerging field of OGs, with OGs potentially important for the development, function, and evolution of living organisms being identified in successive species^[7,44–46]. The accurate identification of OGs is an important prerequisite for their functional prediction and analysis. In the present study, 2,586 OGs from the genome of the rose species *R. chinensis* were identified, representing approximately 7.11% of the genome, a proportion consistent with the typical percentage of OGs in organisms^[47]. Similar to the present results, Lin et al.^[15] identified 1,324 OGs in the *A. thaliana* genome covering approximately 4.9% of the entire genome, Zhao and Ma^[9] characterized a total of 1,789 OGs in tea plant accounting for about 3.37% of the genome, and Guo et al.^[28] identified 1,926 OGs in the rice genome representing approximately 4.9% of the whole genome. At present, the authentication of OGs relies primarily on comparing the target genome with published genomes of its homologous species. The more reference genomes are available, the richer the annotation information and the smaller the genome gaps of the target genome. Consequently, the number of reference genomes will affect the number and accuracy of OGs identified, where more reference genomes will likely result in fewer OGs and higher accuracy. However, the limitations of currently available identification tools lead to the possibility that the present study may not reflect the fully authentic OGs in the *R. chinensis* genome. Future improvements in the exclusion of pseudogenes from identification tools and evolutionary analysis of non-conserved genes may further improve the accuracy of identification.

Distinctive gene structure of OGs

The shorter origin of OGs relative to NOGs has led to differences in sequence structural features, including gene size, protein size, number and size of exons and introns, GC content, and isoelectric point. The sequence structures between OGs and NOGs were analyzed and compared to reveal whether these general differences exist in the *R. chinensis* genome. Typically, NOGs have a larger gene size compared with OGs^[17,18,42], and the results were consistent with this (Table 1). The short protein length and few exons of OGs (Fig. 2) were similar to the general characteristics of plant families *Poaceae*^[17], *Brassicaceae*^[27], *Rutaceae*^[8], and *Camellia*^[9]. The decrease in the number of exons in OGs may be an important factor contribut-

ing to the reduction in their average size (Fig. 2), as the average length of exons is somewhat constant^[47]. Consequently, even if the length of exons and introns of OGs was significantly higher than that of NOGs, it had little effect on average size. *De novo* origination might be another reason accounting for this shorter gene size of OGs, due to their short evolutionary time^[48]. In addition, it was speculated that another possible contributor to the difference was the higher proportion of intron-less OGs. Kaessmann suggested that recurrent line-specific expansion may lead to a dramatic enrichment of intron-less genes and the creation of new genes by retrotransposition, a phenomenon that has been demonstrated in zebrafish^[13]. On the other hand, in the *R. chinensis* genome, the GC content of OGs was noticeably less than that of NOGs, in line with *Aegiceria corniculatum*^[49], and the wheat genome^[39]. Notably, the GC content of OGs and NOGs tended to be highly variable across species. In contrast to the present results, the GC content of OGs was markedly higher than that of NOGs in sweet orange^[8], *Bombyx mori*^[7], zebrafish^[18], and tea^[9]. The selection and adaptation of organisms to the external environment and genetic recombination are important drivers of changes in GC content^[50,51]. The isoelectric point is intimately associated with protein function, and its alteration is thought to be a modifying effect on protein function, important for solubility, sub-cellular localization, and protein interactions^[52]. It was revealed that the isoelectric point of OGs was distinctly above that of NOGs, which may also be driven by selection^[53]. For example, in prokaryotes, adaptation to the environment has led to changes in protein isoelectric points^[54].

Expression characteristics of OGs

During evolution, new genetic elements acquired by the genome, such as OGs, are one of the important sources of functional and phenotypic diversity^[55]. The expression patterns of OGs on different tissues for the prediction and understanding of their biological functions are accessible using RNA-seq^[56]. OGs are inconsistently expressed on different tissues, and in general, OGs are highly expressed in the reproductive system of plants and animals^[8,18,57,58]. In the present study, 214 OGs were expressed tissue-specific, of which 90 and 30 OGs were expressed only in reproductive organs, stamen, and pistil ovaries, respectively; in addition, 24, 20, 29, and 21 OGs were expressed in nutritional organs, including roots, stems, leaves, and prickles (Fig. 5). The specific expression of more than 50% of *R. chinensis* OGs on reproductive organs implied their important role in reproductive development, which was largely in line with the expression profile of other plants' OGs^[9,28,49], and more detectable expression was found in stamen (42.06%). Many studies have shown that OGs, or young genes as some researchers call them, were more inclined to be expressed in the male organs. In 2015, Cui refined this doctrine to 'new genes out of the male' and hypothesized that new genes have an important role in reproductive isolation and species differentiation^[1]. The present results also support this hypothesis. In a word, the specific expression of OGs in *R. chinensis* provides important data on their resistance mechanisms, which is resistant to herbivores, pathogens, or mechanical damage, and also prevents water loss, which suggests an important role in the evolution of habitat adaptation in *R. chinensis*.

Functional inference of OGs

It is not feasible to use homology comparisons to infer the possible expression characteristics and functions of OGs, as

they are unique in each species. Under this circumstance, co-expressed gene modules rich in biological information become a dependable vehicle for inferring the biological processes that maybe involved in OGs, as co-expressed genes usually exhibit significant functional similarities^[59]. In the present study, WGCNA was used to identify 217 OGs distributed in five modules (Supplementary Table S3). KEGG analysis revealed that these co-expressed gene modules were engaged in a variety of biologically important processes. These mainly included linoleic acid metabolism, pentose and glucuronate interconversions, photosynthesis, and plant hormone signal transduction. In all, the involvement of OGs in important physiological processes such as growth and development, signal transduction, and metabolism in roses demonstrates that OGs can be functional and are likely to be essential.

OGs involved in pentose and glucuronate interconversions may contribute to male sterility, given the crucial role of anthers in male gametogenesis, while male sterility mutants usually exhibit genetic disruption linked to anther and pollen development^[60,61], which can result from abnormal carbohydrates metabolism^[62,63]. Carbohydrates, such as pentose and glucuronate, play multiple roles in pollen development, serving as a major source of energy for plant metabolism and as important signaling molecules for regulating growth and development^[64,65]. β -glucuronidase, a glycosyl hydrolase, is mainly responsible for the lysosomal degradation of mucopolysaccharides, dermatopietin, and keratin sulfate^[66]. In addition, β -glucuronidase is involved in the metabolism of various endogenous substances, including pentose, glucuronides, porphyrins, starch, and sucrose^[67]. The pentose and glucuronide interconversion pathway, enriched with OGs, is highly correlated with male sterility in studies of cabbage (*Brassica oleracea* L. var. *capitata*)^[68] and cotton (*Gossypium hirsutum* L.)^[69,70] transcriptome analysis, suggesting the involvement of key genes in this pathway.

Rose flower opening is dependent on petal expansion. The adaxial-abaxial regulation of petals led to a heterogeneous distribution of auxin, and the transport and signaling of the phytohormone auxin were involved in the entire development of flowers^[71], and the OGs might be active in determining and maintaining the adaxial-abaxial polarity of petals. On the other hand, floral fragrance is an important feature of ornamental roses, providing sensory pleasure to humans, and monoterpenes are one of the main classes constituting the fragrance of roses^[72]. Usually, monoterpenes are mainly produced in plastids and their substrates are derived from the methylerythritol 4-phosphate pathway^[73]. However, the lack of photosynthesis in rose petals may lead to a reduced flux of substrates produced through the methylerythritol 4-phosphate pathway. Therefore, photosynthesis not only provides roses with energy for growth and development, but also plays a role in the production and dispersal of floral fragrances. Fatty acid derivatives are prominent compounds in the leaves and sepals of roses^[74]. An important class of enzymes involved in the formation of fatty acid-derived volatiles is lipoxygenase (LOX), the enzymatic oxidation of linoleic acid by LOX to produce hexanal. In addition, cleavage of linoleic acid at the 12–13 double bond yields the C12 precursor of jasmonic acid^[75]. It has been shown that increased levels of jasmonic acid methylation led to a delayed degradation of carotenoids and affected the carotenoid content of the yellow rose cultivar *R. hybrida* 'Frisca'^[76].

Rose plants usually grow in subtropical climates where environmental stresses such as drought and salinity are important factors limiting their growth and productivity. It has been shown that OGs were preferentially expressed under abiotic stress, as in yeast^[77] and rice^[26,28]. Using available RNA-seq data, the expression patterns of OGs were screened under salinity and drought stresses, respectively, and observed 201 and 480 OGs stimulated individually, suggesting the possible important role of these stress-responsive OGs in adaptation to extreme environmental stresses (Fig 8). The increased stimulation of OGs under drought is probable to be highly correlated with their native environment, exemplified by the Middle East Taif region, where Damask rose (*Rosa damascena* Mill.) originated^[78]. The roots serve as the primary site of plant perception of soil water deficit, and subsequently, drought response signals from the roots are transmitted to the leaves^[79]. The enrichment of OGs in phytohormone signaling pathways suggested that they may have an essential position in drought sensing and signaling. Surprisingly, a total of 107 OGs responding to both types of stress were found (Supplementary Table S10). *RhEXPA4* is a rose expansion protein gene that regulates leaf growth and bestows drought and salt tolerance in *Arabidopsis*^[80]. This suggests that these 107 OGs may be critical candidates for further studies of environmental adaptations in roses.

In conclusion, even without direct functional evidence, the analysis of co-expressed gene modules implies that OG genes maybe involved in important biological processes such as developmental regulation, signal transduction, metabolism, and stress adaptation in roses.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Mao J; data analysis, draft manuscript preparation: Ma D, Ding Q; critical manuscript revision: Mao J, Ma D, Zhao Z, Han X. All authors read and approved the final manuscript.

Data availability

The data that support the findings of this study are available in the Rosa database repository: <https://lipm-browsers.toulouse.inra.fr/pub/RchiOBHm-V2/>.

Acknowledgments

We appreciate Laboratoire Reproduction et Développement des Plantes for providing their valuable databases to the public. This work was financially supported by Fundamental Research Funds for the Central Universities (JUSRP124005).

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper at (<https://www.maxapress.com/article/doi/10.48130/tp-0024-0036>)

Dates

Received 4 July 2024; Revised 11 August 2024; Accepted 26 August 2024; Published online 23 October 2024

References

- Cui X, Lv Y, Chen M, Nikoloski Z, Twell D, et al. 2015. Young genes out of the male: an insight from evolutionary age analysis of the pollen transcriptome. *Molecular Plant* 8:935–45
- Fischer D, Eisenberg D. 1999. Finding families for genomic ORFans. *Bioinformatics* 15:759–62
- Wissler L, Gadua J, Simola DF, Helmkampf M, Bornberg-Bauer E. 2013. Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome Biology and Evolution* 5:439–55
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, et al. 1996. Life with 6000 genes. *Science* 274:546–67
- Siew N, Fischer D. 2003. Analysis of singleton ORFans in fully sequenced microbial genomes. *Proteins* 53:241–51
- Heather JM, Chain B. 2016. The sequence of sequencers: the history of sequencing DNA. *Genomics* 107:1–8
- Sun W, Zhao XW, Zhang Z. 2015. Identification and evolution of the orphan genes in the domestic silkworm, *Bombyx mori*. *FEBS Letters* 589:2731–38
- Xu Y, Wu G, Hao B, Chen L, Deng X, et al. 2015. Identification, characterization and expression analysis of lineage-specific genes within sweet orange (*Citrus sinensis*). *BMC Genomics* 16:995
- Zhao Z, Ma D. 2021. Genome-wide identification, characterization and function analysis of lineage-specific genes in the tea plant *Camellia sinensis*. *Frontiers in Genetics* 12:770570
- Daubin V, Lerat E, Perrière G. 2003. The source of laterally transferred genes in bacterial genomes. *Genome Biology* 4:R57
- Long M, Betrán E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics* 4:865–75
- Daubin V, Ochman H. 2004. Start-up entities in the origin of new genes. *Current Opinion in Genetics and Development* 14:616–19
- Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Research* 20:1313–26
- Wu DD, Irwin DM, Zhang YP. 2011. *De novo* origin of human protein-coding genes. *PLoS Genetics* 7:e1002379
- Lin H, Moghe G, Ouyang S, Iezzi A, Shiu SH, et al. 2010. Comparative analyses reveal distinct sets of lineage-specific genes within *Arabidopsis thaliana*. *BMC Evolutionary Biology* 10:41
- Domazet-Lošo T, Tautz D. 2003. An evolutionary analysis of orphan genes in *Drosophila*. *Genome Research* 13:2213–19
- Campbell MA, Zhu W, Jiang N, Lin H, Ouyang S, et al. 2007. Identification and characterization of lineage-specific genes within the *Poaceae*. *Plant Physiology* 145:1311–22
- Yang L, Zou M, Fu B, He S. 2013. Genome-wide identification, characterization, and expression analysis of lineage-specific genes within zebrafish. *BMC Genomics* 14:65
- Heinen TJA, Staubach F, Häming D, Tautz D. 2009. Emergence of a new gene from an intergenic region. *Current Biology* 19:1527–31
- Joppich C, Scholz S, Korge G, Schwendemann A. 2009. Umbrea, a chromo shadow domain protein in *Drosophila melanogaster* heterochromatin, interacts with Hip, HP1 and HOAP. *Chromosome Research* 17:19–36
- Chen S, Zhang YE, Long M. 2010. New genes in *Drosophila* quickly become essential. *Science* 330:1682–85
- Li L, Zheng W, Zhu Y, Ye H, Tang B, et al. 2015. QQS orphan gene regulates carbon and nitrogen partitioning across species via NF- κ B interactions. *Proceedings of the National Academy of Sciences of the United States of America* 112:14734–39
- Yeh SD, Do T, Chan C, Cordova A, Carranza F, et al. 2012. Functional evidence that a recently evolved *Drosophila* sperm-specific gene boosts sperm competition. *Proceedings of the National Academy of Sciences of the United States of America* 109:2043–48
- Ni F, Qi J, Hao Q, Lyu B, Luo M, et al. 2017. Wheat *Ms2* encodes for an orphan protein that confers male sterility in grass species. *Nature Communications* 8:15121
- Li G, Wu X, Hu Y, Muñoz-Amatriaín M, Luo J, et al. 2019. Orphan genes are involved in drought adaptations and ecoclimatic-oriented selections in domesticated cowpea. *Journal of Experimental Botany* 70:3101–10
- Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, et al. 2011. The ecoresponsive genome of *Daphnia pulex*. *Science* 331:555–61
- Donoghue MT, Keshavaiah C, Swamidatta SH, Spillane C. 2011. Evolutionary origins of *Brassicaceae* specific genes in *Arabidopsis thaliana*. *BMC Evolutionary Biology* 11:47
- Guo WJ, Li P, Ling J, Ye SP. 2007. Significant comparative characteristics between orphan and nonorphan genes in the rice (*Oryza sativa* L.) genome. *Comparative and Functional Genomics* 2007:21676
- Dong X, Jiang X, Kuang G, Wang Q, Zhong M, et al. 2017. Genetic control of flowering time in woody plants: roses as an emerging model. *Plant Diversity* 39:104–10
- Martin M, Piola F, Chessel D, Jay M, Heizmann P. 2001. The domestication process of the Modern Rose: genetic structure and allelic composition of the rose complex. *Theoretical and Applied Genetics* 102:398–404
- Raymond O, Gouzy J, Just J, Badouin H, Verdenaud M, et al. 2018. The *Rosa* genome provides new insights into the domestication of modern roses. *Nature Genetics* 50:772–77
- Zhang G, Wang H, Shi J, Wang X, Zheng H, et al. 2007. Identification and characterization of insect-specific proteins by genome data analysis. *BMC Genomics* 8:93
- Xia X. 2018. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Molecular Biology and Evolution* 35:1550–52
- Savojardo C, Martelli PL, Fariselli P, Profiti G, Casadio R. 2018. BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic Acids Research* 46:W459–W466
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* 18:292–98
- Wang Y, Tang H, DeBarry JD, Tan X, Li J, et al. 2022. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* e49
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–20
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29:644–52
- Ma S, Yuan Y, Tao Y, Jia H, Ma Z. 2020. Identification, characterization and expression analysis of lineage-specific genes within *Triticeae*. *Genomics* 112:1343–50
- Pan JB, Hu SC, Wang H, Zou Q, Ji ZL. 2012. PaGeFinder: quantitative identification of spatiotemporal pattern genes. *Bioinformatics* 28:1544–45
- Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559
- Toll-Riera M, Bosch N, Bellora N, Castelo R, Armengol L, et al. 2009. Origin of primate orphan genes: a comparative genomics approach. *Molecular Biology and Evolution* 26:603–12
- Yang X, Jawdy S, Tschaplinski TJ, Tuskan GA. 2009. Genome-wide identification of lineage-specific genes in *Arabidopsis*, *Oryza* and *Populus*. *Genomics* 93:473–80
- Kapas S, Clark AJL. 1995. Identification of an orphan receptor gene as a type 1 calcitonin gene-related peptide receptor. *Biochemical and Biophysical Research Communications* 217:832–38
- Prabh N, Rödelsperger C. 2019. *De novo* divergence, and mixed origin contribute to the emergence of orphan genes in *Pristionchus nematodes*. *G3 Genes|Genomes|Genetics* 9:2277–86
- Zhang W, Gao Y, Long M, Shen B. 2019. Origination and evolution of orphan genes and *de novo* genes in the genome of *Caenorhabditis elegans*. *Science China Life Sciences* 62:579–93

47. Arendsee ZW, Li L, Wurtele ES. 2014. Coming of age: orphan genes in plants. *Trends in Plant Science* 19:698–708
48. Neme R, Tautz D. 2013. Phylogenetic patterns of emergence of new genes support a model of frequent *de novo* evolution. *BMC Genomics* 14:117
49. Ma D, Ding Q, Guo Z, Zhao Z, Wei L, et al. 2021. Identification, characterization and expression analysis of lineage-specific genes within mangrove species *Aegiceras corniculatum*. *Molecular Genetics and Genomics* 296:1235–47
50. Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159:907–11
51. Lassalle F, Périán S, Bataillon T, Nesme X, Duret L, et al. 2015. GC-content evolution in bacterial genomes: the biased gene conversion hypothesis expands. *PLoS Genetics* 11:e1004941
52. Kiraga J, Mackiewicz P, Mackiewicz D, Kowalczyk M, Biecek P, et al. 2007. The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. *BMC Genomics* 8:163
53. Alendé N, Nielsen JE, Shields DC, Khaldi N. 2011. Evolution of the isoelectric point of mammalian proteins as a consequence of indels and adaptive evolution. *Proteins* 79:1635–48
54. Nandi S, Mehra N, Lynn AM, Bhattacharya A. 2005. Comparison of theoretical proteomes: Identification of COGs with conserved and variable pI within the multimodal pI distribution. *BMC Genomics* 6:116
55. Chen S, Krinsky BH, Long M. 2013. New genes as drivers of phenotypic evolution. *Nature Reviews Genetics* 14:645–60
56. Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10:57–63
57. Begun DJ, Lindfors HA, Kern AD, Jones CD. 2007. Evidence for *de novo* evolution of testis-expressed genes in the *Drosophila yakuba/Drosophila erecta* clade. *Genetics* 176:1131–37
58. Wu DD, Wang X, Li Y, Zeng L, Irwin DM, et al. 2014. "Out of pollen" hypothesis for origin of new genes in flowering plants: study from *Arabidopsis thaliana*. *Genome Biology and Evolution* 6:2822–29
59. Obayashi T, Kinoshita K. 2009. Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA Research* 16:249–60
60. Jeong HJ, Kang JH, Zhao M, Kwon JK, Choi HS, et al. 2014. Tomato *Male sterile 10³⁵* is essential for pollen development and meiosis in anthers. *Journal of Experimental Botany* 65:6693–709
61. Geng X, Ye J, Yang X, Li S, Zhang L, et al. 2018. Identification of proteins involved in carbohydrate metabolism and energy metabolism pathways and their regulation of cytoplasmic male sterility in wheat. *International Journal of Molecular Sciences* 19:324
62. Yue J, Ren Y, Wu S, Zhang X, Wang H, et al. 2014. Differential proteomic studies of the genic male-sterile line and fertile line anthers of upland cotton (*Gossypium hirsutum* L.). *Genes and Genomics* 36:415–26
63. Liu H, Wang J, Li C, Qiao L, Wang X, et al. 2018. Phenotype characterization and analysis of expression patterns of genes related mainly to carbohydrate metabolism and sporopollenin in male-sterile anthers induced by high temperature in wheat (*Triticum aestivum*). *Crop and Pasture Science* 69:469–78
64. Kunz S, Pesquet E, Kleczkowski LA. 2014. Functional dissection of sugar signals affecting gene expression in *Arabidopsis thaliana*. *PLoS ONE* 9:e100312
65. Hirsche J, García Fernández JM, Stabentheiner E, Großkinsky DK, Roitsch T. 2017. Differential effects of carbohydrates on *Arabidopsis* pollen germination. *Plant and Cell Physiology* 58:691–701
66. Sudan C, Prakash S, Bhomkar P, Jain S, Bhalla-Sarin N. 2006. Ubiquitous presence of beta-glucuronidase (GUS) in plants and its regulation in some model plants. *Planta* 224:853–64
67. Witcher DR, Hood EE, Peterson D, Bailey M, Bond D, et al. 1998. Commercial production of β -glucuronidase (GUS): a model system for the production of proteins in plants. *Molecular Breeding* 4:301–12
68. Tian A, Zhang E, Cui Z. 2021. Full-length transcriptome analysis reveals the differences between floral buds of recessive genic male-sterile line (RMS3185A) and fertile line (RMS3185B) of cabbage. *Planta* 253:21
69. Zhang Y, Chen J, Liu J, Xia M, Wang W, et al. 2015. Transcriptome analysis of early anther development of cotton revealed male sterility genes for major metabolic pathways. *Journal of Plant Growth Regulation* 34:223–32
70. Li Y, Qin T, Wei C, Sun J, Dong T, et al. 2019. Using transcriptome analysis to screen for key genes and pathways related to cytoplasmic male sterility in cotton (*Gossypium hirsutum* L.). *International Journal of Molecular Sciences* 20:5120
71. Han Y, Yong X, Yu J, Cheng T, Wang J, et al. 2019. Identification of candidate adaxial-abaxial-related genes regulating petal expansion during flower opening in *Rosa chinensis* "old blush". *Frontiers in Plant Science* 10:1098
72. Tholl D, Gershenzon J. 2015. The flowering of a new scent pathway in rose. *Science* 349:28–29
73. Hemmerlin A, Harwood JL, Bach TJ. 2012. A *raison d'être* for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Progress in Lipid Research* 51:95–148
74. Caissard JC, Bergougnoux V, Martin M, Mauriat M, Baudino S. 2006. Chemical and histochemical analysis of 'Quatre Saisons Blanc Mousseux', a moss rose of the *Rosa × damascena* group. *Annals of Botany* 97:231–38
75. Dudareva N, Pichersky E, Gershenzon J. 2004. Biochemistry of plant volatiles. *Plant Physiology* 135:1893–902
76. Glick A, Philosoph-Hadas S, Vainstein A, Meir A, Tadmor Y, et al. 2007. Methyl jasmonate enhances color and carotenoid content of yellow-pigmented cut rose flowers. *Acta Horticulturae* 755:243–50
77. Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, et al. 2012. Proto-genes and *de novo* gene birth. *Nature* 487:370–74
78. Al-Yasi H, Attia H, Alamer K, Hassan F, Ali E, et al. 2020. Impact of drought on growth, photosynthesis, osmotic adjustment, and cell wall elasticity in Damask rose. *Plant Physiology and Biochemistry* 150:133–39
79. Schachtman DP, Goodger JQD. 2008. Chemical root to shoot signaling under drought. *Trends in Plant Science* 13:281–87
80. Lü P, Kang M, Jiang X, Dai F, Gao J, et al. 2013. RhEXPA4, a rose expansin gene, modulates leaf growth and confers drought and salt tolerance to *Arabidopsis*. *Planta* 237:1547–59



Copyright: © 2024 by the author(s). Published by Maximum Academic Press on behalf of Hainan University. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.