

Resource Article: Genomes Explored

Chromosome-level genome of a leaf vegetable *Glebionis coronaria* provides insights into the biosynthesis of monoterpenoids contributing to its special aroma

Sen Wang[†], Anqi Wang[†], Hengchao Wang, Fan Jiang, Dong Xu, and Wei Fan^{*†} 

Guangdong Laboratory for Lingnan Modern Agriculture (Shenzhen Branch), Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangdong, China

^{*}Corresponding author: Email: fanwei@caas.cn

[†]These authors contributed equally to this work.

Abstract

Glebionis coronaria is a popular vegetable with special aroma and a medical plant in East Asia and Mediterranean, but its biological studies and breeding have been hindered by the lack of reference genome. Here, we present a chromosome-level reference genome of *G. coronaria*, with assembled genome size of 6.8 Gb, which is the largest among all the published genomes of diploid Asteraceae species. The large genome size of *G. coronaria* is mainly caused by the recent widespread explosions of long-terminal-repeat retrotransposons. Analyses of macro-synteny and synonymous mutation rate distribution indicate that the *G. coronaria* genome experienced a whole-genome triplication at 40–45 million years ago, shared with all Asteraceae species. In subtribe Artemisiinae, *Glebionis* arose before the divergence of *Chrysanthemum* from *Artemisia*, and *Glebionis* species evolved much faster than *Chrysanthemum* and *Artemisia* species. In *G. coronaria*, the synthesis genes of monoterpenoids 8-oxocitronellyl enol and isopipitenone were expanded, and the higher expressions of these expanded genes in leaves and stems may contribute to its special aroma. The *G. coronaria* genomic resources will promote the evolution studies of Asteraceae, the metabolism mechanism studies of bioactive compounds, and the breeding improvement of agronomic traits in *G. coronaria*.

Key words: *Glebionis coronaria*, reference genome, transposable element, vegetable, aroma

1. Introduction

Glebionis coronaria is a popular leaf vegetable and also a medical plant in East Asia and Mediterranean.¹ *Glebionis coronaria* is an annual plant propagated by seeds, and its yellow capitulum flowers are similar to the flowers of some *Chrysanthemum* species. Therefore, many previous studies named it as *Chrysanthemum coronarium*,² but recent molecular phylogeny studies propose to place *G. coronaria* in a separate genus *Glebionis*.³ As *G. coronaria* contain abundant β -carotene, iron, calcium, and other nutrients and have special pleasant aroma,⁴ the shoots are consumed as vegetable in China, Korea, and Japan. Moreover, the essential oils extracted from *G. coronaria* contain many secondary metabolites such as camphor, pinene, and chrysanthenyl, which have antimicrobial, antioxidant, antiviral, and antimycotic activities.⁵ Thus, *G. coronaria* has also been used as a traditional medicine to treat pain, constipation, and cough in ancient times.⁶

The production of secondary metabolites can help plants adapt to changing environments and promote their survival and reproduction. As a large group of plant secondary metabolites, terpenoids are widespread in various plants and participate in photosynthesis (carotene), membrane fluidity (sterol), hormone metabolism (gibberellin, abscisic acid), etc.⁷

The major component of plant terpenoids is species specific, such as taxol (*Taxus chinensis*) and artemisinin (*Artemisia annua*), contributing to their specific biological characteristics.⁸ The major compounds of essential oils in Asteraceae plants are also terpenoids, especially the monoterpenoids (C10) which are synthesized from two isoprene (C5) units.⁹ Monoterpenoids are often volatile and have specific smell, contributing to the flower scent of many ornamental plants,¹⁰ and some monoterpenoids like camphor and linalool are widely used in spice and perfume industry.¹¹ Many monoterpenoids can be detected in the essential oils of *G. coronaria*, but the genes involved in the synthesis of monoterpenoids are not clear.

Glebionis coronaria belongs to the subfamily Asteroideae of the largest family Asteraceae of flower plants. In the past decade, over 20 Asteraceae genomes have been published (Supplementary Fig. S15), including 6 species of subfamily Cichorioideae like *Lactuca sativa*,¹² 4 species of subfamily Carduoideae like *Cynara cardunculus*,¹³ and 18 species of Asteroideae such as *Erigeron canadensis*,¹⁴ *Helianthus annuus*,¹⁵ and *A. annua*.¹⁶ These genomic studies revealed that the ancestor of all Asteraceae species experienced a whole-genome triplication (WGT) at 40–45 million years ago (MYA),^{12,15} the ancestor of Heliantheae alliance experienced

an additional whole-genome duplication (WGD) at ~29 MYA,¹² and the yacon-unique WGD occurred at 5.6–5.8 MYA.¹⁷ The genome size of *G. coronaria* is estimated to be ~7 Gb by *C*-value,¹⁸ which is much larger than other Asteraceae diploid species. However, the reason for large genome of *G. coronaria* is not clear. In this study, we present a chromosome-level reference genome of *G. coronaria*, investigate its phylogeny history and genome evolution, and explore the biosynthesis genes of monoterpenoids contributing to its special aroma.

2. Materials and methods

2.1. Karyotype analysis

A local cultivar named ‘fanji’ of *G. coronaria* widely grown in Hebei, Anhui, Jiangsu, and Guangdong provinces of China was selected for genome sequencing. The seeds of *G. coronaria* were incubated at 25°C in culture dish, and the fresh root tips of seedlings were sampled for karyotype analysis by fluorescence *in situ* hybridization. Chromosome numbers were counted by staining with the fluorescence dye DAPI and hybridization with telomere-specific oligonucleotide probes. Ploidy was determined by hybridization with 18S rDNA and 5S rDNA-specific probes.

2.2. Genome sequencing

Young leaves of a 30-day-old plant of *G. coronaria* grown in greenhouse were sampled for genomic DNA extraction by Tiangen Hi-DNAsecure Plant Kit (DP350) according to the provided protocols. The purity and integrity of extracted DNA were checked by agarose gel electrophoresis and Thermo Fisher Nanodrop 2000 ultraviolet spectrophotometer, and the high-quality DNA sample with main fragment length >23,000 bp and A260/280 1.8–2.0 was used for long-read sequencing library preparation. Then, genomic DNA was sheared into 10,000–15,000 bp fragments by Covaris g-TUBEs according to the provided protocols, and converted to SMRT dumbbell libraries using PacBio SMRTbell Express Template Prep Kit 2.0 according to the provided protocols. The prepared SMRT sequencing libraries were then sequenced on PacBio Sequel II with the Circular Consensus Sequencing (CCS) mode. Due to the large genome size of *G. coronaria*, a total of six SMRT libraries were prepared and sequenced.

Fresh young leaves of the same 30-day-old plant of *G. coronaria* used for genomic sequencing were also used for High throughput *in situ* Chromatin conformation capture (Hi-C) sequencing. First, nuclear DNA of sampled young leaves was cross-linked by soaking in formaldehyde solution for 15 min. Second, the cross-linked genomic DNA was also extracted using Tiangen Hi-DNAsecure Plant Kit (DP350) and *in vitro* digested by the restriction endonuclease MboI at motifs GATC. Third, the digested DNA ends were repaired and marked with biotin, and spatially proximal ends were ligated to form circles. Fourth, the circular DNA fragments were purified and sheared into 350 bp inserts via Covaris S220 Focused Ultrasonicator, and the inserts with biotin mark were pull down by Streptavidin Magnetic Beads. Last, the biotin-marked spatially proximal DNA inserts were converted to short-read sequencing library by Truseq DNA Library Prep Kit, and sequenced on Illumina NovaSeq 6000 sequencer in paired-end 150-bp mode.

2.3. Transcriptome sequencing

Total RNA of root, stem, and leaf tissues of 60-day-old *G. coronaria* plants were extracted using QIAGEN RNeasy Plant Mini Kit according to the provided protocols. The purity and integrity of extracted RNA sample were checked by agarose gel electrophoresis, Thermo Fisher Nanodrop 2000 ultraviolet spectrophotometer, and Agilent 2100 Bioanalyzer. High-quality RNA samples with RIN >8 and clear 18S rDNA and 25S rDNA bands were used for transcriptome sequencing. The mRNAs in total high-quality RNA were reverse transcribed to cDNAs using NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module and PacBio Iso-Seq Express Oligo Kit, according to the provided protocols. Then, the cDNA fragments with lengths of 500–6,000 bp were converted to isoform sequencing (Iso-Seq) libraries by PacBio SMRTbell Express Template Prep Kit 2.0. The Iso-Seq libraries of root, stem, and leaf tissues were barcoded and sequenced on PacBio Sequel II sequencer.

2.4. Genome assembly

Before genome assembly, the genomic HiFi sequencing reads with average quality <99% were discarded, and the sequenced reads derived from a Hi-C library were also quality filtered to remove adapter contamination, low-quality or ambiguous bases. Then, we used GCE version 1.02¹⁹ to estimate the genome size, repeat content, and heterozygosity of *G. coronaria* by *K*-mer analysis (*K* = 17) of genomic HiFi reads. Because the estimated heterozygosity is high (~0.5%, [Supplementary Fig. S2](#)), we used the Hifiasm version v0.16.1 with Hi-C mode²⁰ to obtain haplotype-resolved assemblies of diploid *G. coronaria* [haplotype 1 (hap1) and haplotype 2 (hap2)], through combining HiFi reads with Hi-C data. The completeness of hap1 and hap2 contig assemblies was assessed using BUSCO version 5.2.2²¹ with eudicots_odb10 database. The contig set of hap2 assembly having slightly higher contiguity and BUSCO completeness was selected as the reference genome assembly of *G. coronaria* ([Supplementary Table S3](#)). Then the organelle and microbe-derived fragments were identified and removed to get the nucleus contigs, by searching against the 403,174 prokaryote and 23,229 organelle genomes downloaded from the NCBI database, using Minimap version 2.20²² and the criterion of identity >0.95 and coverage >0.95.

To obtain the nine pseudochromosome scaffolds of *G. coronaria* through proximity ligation data, we mapped the quality-filtered Hi-C sequencing reads to the reference genome to generate the valid Hi-C contact matrixes among contig bins (size 100,000 bp) using HiC-Pro version 3.1.0.²³ Then, the contigs >1,000,000 bp were assembled into chromosome-level scaffolds based on the Hi-C linkage information among contig ends, using EndHiC version 1.0²⁴ in multi-round mode and manual correction of mis-joined scaffolds according to Hi-C heatmap.

2.5. Annotation of repeat elements

We conducted a comprehensive identification of tandem repeats (TRs) and transposon elements (TEs) in *G. coronaria* genome. TRF version 4.07²⁵ was used to identify TRs in *G. coronaria* genome. TEs in *G. coronaria* genome were classified into three groups and identified through three steps: (i) Intact TEs including long-terminal-repeat retrotransposons (LTR-RTs), DNA transposons, and Helitron transposons were predicted according to their structural characteristics

using EDTA version 2.0.0,²⁶ and these intact TEs were clustered as a reference TE library for *G. coronaria*; (ii) Homology TEs in the intact TE-masked genome were identified by sequence similarity to the known TEs in intact TE library, Repbase (plant lineage) database version 26.05, and TE protein database, using RepeatMasker version 4.1.2 (<http://www.repeatmasker.org>); (iii) Denovo TEs were identified in the Intact and Homology TE-masked genome mainly by high copy number, in which a TE library was firstly created using RepeatModeler version 2.0.2²⁷ and classified using TERL,²⁸ and the classified TE library was used to identify Denovo TEs using RepeatMasker. The identified Intact, Homology, and Denovo TEs were combined into a non-redundant TE annotation of *G. coronaria* genome. The statistics of TEs at class, order, superfamily, and family level were done using in-house scripts, and the insertion time of intact LRT-RTs was estimated by the sequence divergence of LTR pairs using LTR_retriever.²⁹ Then, the *G. coronaria* genome sequences were soft masked (uppercase to lowercase) at all TEs with length over 80 bp for gene prediction.

2.6. Annotation of protein-coding genes and non-coding RNA genes

Protein-coding genes in the TE-masked *G. coronaria* genome were predicted using Augustus version 3.4.0,³⁰ which integrated the supporting evidence from mRNA transcripts and homologue proteins. The species-specific gene training parameters used by Augustus were obtained from BUSCO assessment of *G. coronaria* genome assembly completeness. The transcript supporting hints used by Augustus were generated by aligning the PacBio Iso-Seq full-length transcripts of root, stem, and leaf tissues to genome using GMAP version 2020-10-27,³¹ and converting the alignments with identify and coverage >95% to hints file using Augustus script blat2hints.pl. The homologue protein supporting hints used by Augustus were produced by aligning the proteome of *A. annua*, *C. nankingense*, *E. canadensis*, and *H. annuus* (Supplementary Table S7) to *G. coronaria* genome using Exonerate version 2.2.0,³² and converting the best predicted gene structures to hints file using Augustus script exonerate2hints.pl. BUSCO was also used to assess the completeness of *G. coronaria* gene set with eudicots_odb10 database.

Due to the abundant TE content of *G. coronaria* genome, a post-filtering of transposon genes was conducted for the predicted gene set of *G. coronaria*, in which all genes were firstly functionally annotated by searching against the NCBI-NR database using Diamond version 0.8.28,³³ and the genes with terms ‘retrovirus’, ‘transposon’, ‘copia’, ‘gypsy’, ‘transposae’, ‘gag-pol’, ‘integrase’, etc. were identified as potential transposon genes and removed from the gene set. Thereafter, the functional annotation of the filtered protein-coding gene set of *G. coronaria* was done by searching against NCBI-NR and KEGG databases using Diamond, and the protein domain annotation was done using InterProScan version 5.52-86.³⁴ The genes encoding tRNAs and rRNAs were predicted using tRNAscan-SE version 2.0 and RNAmmer version 1.2, respectively.

2.7. Phylogeny reconstruction and divergence time estimation

To reconstruct the phylogenetic history of *G. coronaria* in Asteraceae, we firstly identify the orthogroups (including orthologue and recent paralogue genes) of *G. coronaria*

and six Asteroideae species *C. seticuspe* (Plant GARDEN CsGojo-0_v1), *A. annua* (Global Pharmacopoeia Genome Database Phase0), *E. canadensis* (NCBI C_canadensis_v1), *S. sonchifolius* (NCBI ASM2352597v1), *S. rebaudiana* (FigShare 15169491.v1), *H. annuus* (NCBI HanXRQr2.0-SUNRISE), one Cichorioideae species *L. sativa* (NCBI Lsat_Salinax_v7), and one outgroup species *Coffea canephora* (NCBI AUK_PRJEB4211_v1) (Supplementary Tables S11 and S12) using OrthoFinder version 2.5.2³⁵ with parameters ‘-M msa -A mafft -T fasttree -1 -y’. The rooted species tree was inferred using STAG and STRIDE methods invoked in OrthoFinder, and species divergence time was estimated using the RelTime method in MEGA11³⁶ with one calibration of the divergence of *Coffea* and Asteraceae at 95–106 MYA, which was obtained from TimeTree (www.timetree.org).

2.8. Genome polyploidization analysis

WGD events during *G. coronaria* evolution were determined based on the macro-synteny at chromosome-scale as well as the distribution of synonymous mutation rate (Ks) for syntenic genes within and between species. Firstly, the all-vs-all alignments of the proteome sequences of *C. seticuspe*, *A. annua*, *G. coronaria*, *E. canadensis*, and *H. annuus* were generated using Diamond³³ in the orthogroup finding by OrthoFinder.³⁵ Then, the proteome alignments were used as input for MCScanX³⁷ to identify the inter- and intra-species syntenic genomic blocks. The type of duplicate genes within species was determined using the duplicate_gene_classifier in MCScanX.³⁷ R packages circlize and ggplot2 were used to draw the inter- and intra-species synteny dot plot and circle plot. The Ks values of intra-species paralogue genes located in syntenic blocks with more than five genes, and inter-species reciprocal best orthologue genes located in syntenic blocks with more than five genes, were calculated using KsKs_Calculator³⁸ with the GMYN model. Ks distribution curves were drawn in Microsoft Excel 2016.

2.9. Analysis of genes involved in terpenoid synthesis

The terpenoid synthesis genes in *G. coronaria* genome were identified by homology alignment. Firstly, we downloaded the known genes involved in the synthesis pathway of terpenoid backbones (map00900) and monoterpenoids (map00902) from KEGG database. Then, the protein-coding genes of *G. coronaria* were aligned to the downloaded known terpenoid synthesis genes using Diamond version 0.8.28³³ with parameter settings ‘blastp --more-sensitive --evaluate 0.00001’. The *G. coronaria* genes with best alignment identity >80% and coverage >60% were retained as potential genes involved in terpenoid synthesis. Furthermore, the potential terpenoid synthesis genes were checked for the existence of N-terminal domain pfam01397 and metal-binding domain pfam03936 of terpenoid synthases using HMMER version 3.1b2, and the genes having both domains were identified as final terpenoid synthesis genes in *G. coronaria*.

To compare the copy number of terpenoid synthesis genes among *G. coronaria* and other Asteroideae species, the above method was also used to identify the terpenoid synthesis genes in *A. annua*, *C. seticuspe*, *E. canadensis*, *H. annuus*, and *S. rebaudiana*. For the terpenoid synthesis gene families with more members in *G. coronaria*, we also used Muscle version 3.8.31 to conduct multiple sequence alignment of

the genes from six Asteroideae species, and used Fasttree version 2.1.11 to construct an unrooted tree to investigate the gene phylogenetic history. Besides, the recently expanded terpenoid synthesis genes were also checked for their expression levels in root, stem, and leaf of *G. coronaria*, by counting the mapped full-length transcripts from these tissues to genes.

3. Results

3.1. High-quality reference genome and annotation

Karyotype analysis by fluorescence *in situ* hybridization showed that the sequenced material of *G. coronaria* is a diploid with $2n = 18$ chromosomes (Supplementary Fig. S1), consistent with the previous report.³⁹ The estimated genome size was ~6.9 Gb with heterozygosity rate of ~0.5%, using K-mer analysis of 160 Gb PacBio HiFi reads (~23.5×) by GCE¹⁹ (Supplementary Fig. S2 and Table S1). To resolve the heterozygous problem of *G. coronaria*, the 160 Gb HiFi reads and 100 Gb Hi-C reads (~15.5×) were assembled using Hifiasm²⁰ to generate two haplotype-resolved contig assemblies (hap1 and hap2), and the contig set of hap2 assembly with slightly higher BUSCO completeness and contiguity was chosen as the reference genome (Supplementary Table S3). The reference genome of *G. coronaria* includes 5,135 contigs, with N50 size of 3.8 Mb and total length of 6.8 Gb (Table 1, Supplementary Table S3). The assembly size is similar to the estimated genome size, suggesting high completeness of the genome assembly. Then, 85.2% of these contigs were further anchored into nine pseudochromosomes by EndHiC,²⁴ with scaffold N50 size of 610.7 Mb (Table 1, Supplementary Tables S2 and S3, Supplementary Fig. S3). Considering that the genome-wide Hi-C contact heatmap had no obvious mistake (Fig. 1b), and the BUSCO completeness of eudicots_odb10 was 94.6% and duplicate rate was 6.4% (Table 1, Supplementary Table S3), the reference genome assembly of *G. coronaria* is in high quality.

Table 1. Statistics of genome assembly and annotation of *Glebionia coronaria*

Genomic feature	Value
Estimated genome size (Gb)	6.7
Assembled genome size (bp)	6,799,396,618
Contig N50 size (bp)	3,868,525
Scaffold N50 size (bp)	610,719,037
Anchored to chromosome (%)	85.2%
Telomeres assembled (%)	27.8%
BUSCO completeness of genome	94.6%
Length of TRs (bp)	407,425,333
Percent of TRs (%)	5.9%
Length of TEs (bp)	6,332,052,017
Percent of TEs (%)	93.1%
Number of tRNA genes	8,241
Number of rRNA genes	1,708
Number of protein-coding genes	76,090
Total CDS length (bp)	69,217,854
Percent of total CDS length (%)	1.0%
BUSCO completeness of gene set	94.8%

A comprehensive annotation of repeat elements showed that 5.9% of *G. coronaria* genome were TRs, and 93.1% were TEs (Table 1, Supplementary Table S4). After masking of TEs with length >80 bp, 104,192 gene models were predicted using Augustus³⁰ with the supporting evidence of full-length transcript mappings generated by PacBio Isoseq (Supplementary Table S6) and homology protein alignments of well-annotated Asteroideae species (Supplementary Table S7). Of these gene models, 28,102 transposon genes were identified by searching against NCBI-NR database and removed to generate the final 76,090 protein-coding genes, with average CDS length 910 bp and exon number 4.5 per gene (Table 1, Supplementary Table S8). The BUSCO completeness of the gene set was 94.8% and duplicate rate was 8.0% which are comparable to those of the reference genome (Table 1), indicating that both the reference genome and gene set of *G. coronaria* are in high quality and have no obvious heterozygous fragment contamination. In addition, 86.8% of protein-coding genes were functionally annotated by at least one hit from NCBI-NR, KEGG, InterPro, and GO databases (Supplementary Table S10). Furthermore, 8,241 tRNA genes and 1,708 rRNA genes were identified in the genome (Table 1, Supplementary Table S9).

3.2. Recent explosion of LTR-RTs results in the large genome of *G. coronaria*

The genome size of *G. coronaria* is the largest among the published genomes of Asteraceae diploid species, which can be ascribed to the highest TE content up to 93% of *G. coronaria* genome (Fig. 2a). The main TEs in *G. coronaria* genome are LTR-RTs, accounting for 84.2% of genome, and the main LTR-RTs are *Gypsy* and *Copia* superfamily which occupy 37.8% and 20.2% of genome, respectively (Fig. 2a, Supplementary Table S5). The DNA transposon superfamilies Mutator and MC-EnSpm, and Helitron transposons occupy 1.9%, 1.6%, and 3.1% of *G. coronaria* genome, respectively (Supplementary Table S5). Besides, the distribution of TEs is overwhelming along the whole chromosomes (Fig. 1a). Thus, the widespread insertion of LTR-RTs contributes greatly to the large genome of *G. coronaria*.

Further analyses of the TEs in *G. coronaria* genome showed that up to 23% were structurally intact TEs. Within the intact TEs, 45% were *Gypsy* LTR-RTs, 23% were *Copia* LTR-RTs, and 28% were other LTR-RTs (Fig. 2b). In addition, the insertions of these intact LTR-RTs were widespread across the whole chromosomes (Supplementary Fig. S4). Intact TEs were normally inserted recently and still have high activity. In *G. coronaria* genome, most intact LTR-RTs were generated in the past 1 million years, and especially an obvious explosion of *Gypsy* and *Copia* LTR-RTs occurred at ~0.1 MYA (Fig. 2c). The earth at ~0.1 MYA was entering into the Pleistocene ice age, and the explosion of LTR-RTs in *G. coronaria* could create more genetic diversity and novel genes to promote its adaptability to cold environments. In the *G. coronaria* genome, the 20 largest LTR-RT families collectively account for over 25% of TEs, and some *Cypsy* or *Copia* families occupy even more than 100 Mb genomic regions (Fig. 2d, Supplementary Fig. S5). Therefore, the large genome size of *G. coronaria* is mainly caused by the recent explosion of LTR-RTs, and the *G. coronaria* genome may continue increase due to the activity of abundant intact TEs.

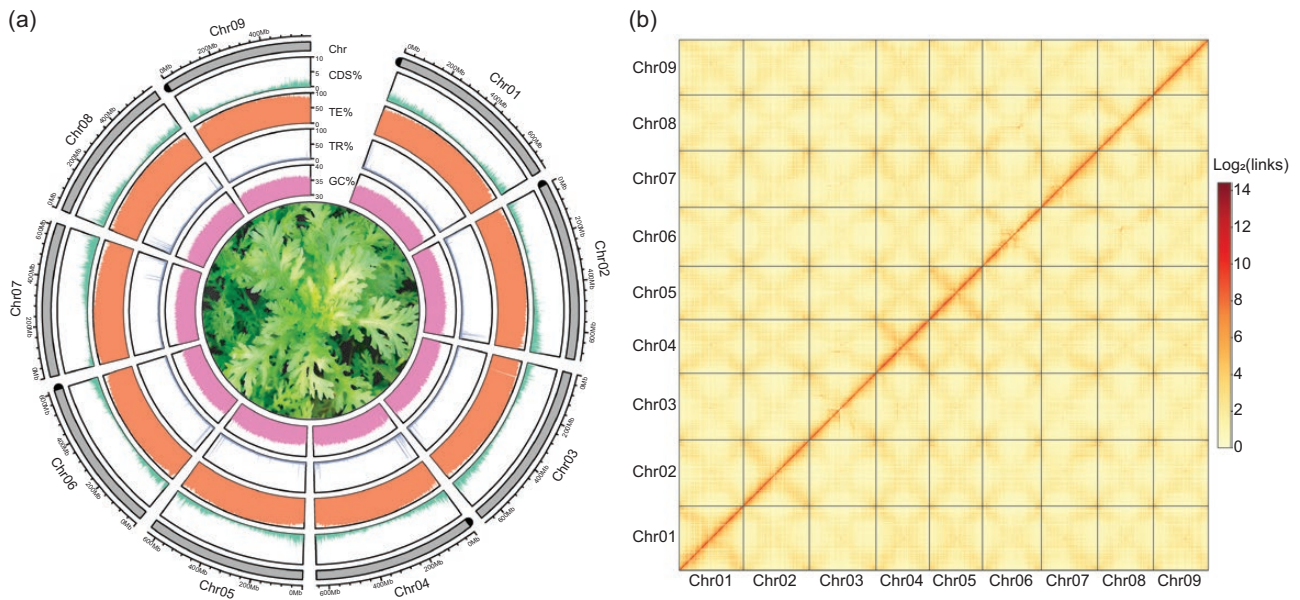


Figure 1. Genome assembly and annotation of *G. coronaria*. (a) Circular view of genomic feature distribution along nine pseudochromosomes. The heights of filled lines in the tracks GC%, TR%, TE%, and CDS% indicate the percent of GC content, accumulated length of TRs, TEs, and gene-coding regions in each 1-Mb sliding window, respectively. In the Chr track, black semicircles at chromosome ends indicate telomeres. (b) Genome-wide Hi-C contact heatmap. Colours are proportional to the Log₂-transformed number of Hi-C links in each 3-Mb bin from the same genomic region or between two 3-Mb bins from two different genomic regions.

3.3. *Glebionis* arose before *Chrysanthemum* and evolved fast in Asteroideae

Previous molecular phylogeny studies have moved *G. coronaria* from the genus *Chrysanthemum* to the genus *Glebionis*, which includes only two species *G. coronaria* and *G. segetum*.⁴⁰ Up to now, the detailed phylogeny history of *G. coronaria* is still not clear. To investigate the evolution history of *G. coronaria* in the subfamily Asteroideae, we selected 6 representative Asteroideae species with published genomes, *C. seticuspe*,⁴¹ *A. annua*,¹⁶ *E. canadensis*,¹⁴ *S. sonchifolius*,¹⁷ *S. rebaudiana*,⁴² *H. annuus*,¹⁵ 1 Cichorioideae species *L. sativa*,¹² and 1 outgroup species *C. canephora*,⁴³ and cluster their genes into 37,696 orthogroups using OrthoFinder2³⁵ (Supplementary Tables S11 and S12). Then, a rooted phylogeny tree was inferred using STAG and STRIDE methods invoked in OrthoFinder2³⁵ based on the gene trees of 2,456 orthogroups with at least 88.9% of species having single-copy genes in any orthogroup (Fig. 3a, Supplementary Fig. S6). Within the subtribe Artemisiinae, *G. coronaria* is a sister to the ancestor of *A. annua* and *C. seticuspe*, suggesting *Glebionis* arose before *Artemisia* and *Chrysanthemum*. In addition, the branch length of *G. coronaria* is much longer than that of *A. annua* and *C. seticuspe* (Fig. 3a), indicating that *G. coronaria* genome has mutated faster and evolved quicker after divergence.

We further estimated the species divergence time in the Asteroideae phylogeny tree (Fig. 3b), using the RelTime method in MEGA11³⁶ with one calibration constraint, 95–106 Ma, between coffee and Asteraceae (obtained from TimeTree database). In the estimated time tree, *G. coronaria* diverged from the ancestor of *A. annua* and *C. seticuspe* at 7.55 MYA, and later *A. annua* diverged from *C. seticuspe* at 4.50 MYA (Fig. 3b). The inferred phylogeny history of *G. coronaria* in Asteroideae supports the placement of *G. coronaria* in the genus *Glebionis*, which arose much earlier than the genus *Chrysanthemum*.

3.4. Whole-genome triplication of Asteraceae ancestor shapes the genome of *G. coronaria*

Genome polyploidization and TE explosion are major driving forces of plant genome evolution, and for the large genome of *G. coronaria* it is necessary to investigate the past genome duplication or triplication events along its evolution history. Because macro-syntenic fragments are important signs of genome duplication, we firstly identified the intra-species syntenic genes in *G. coronaria* and other four Asteroideae species *C. seticuspe*, *A. annua*, *E. canadensis*, and *H. annuus* using MCScanX.³⁷ The intra-species synteny plot of *G. coronaria* shows that some genomic blocks are in triplicate (Fig. 4a, Supplementary Fig. S7), for example the syntenic region among the middle part of Chr02, the middle part of Chr07, and the left part of Chr08, indicating the existence of whole-genome triplication during the genome evolution of *G. coronaria*. Besides, the synonymous mutation rate (*K*_s) distribution of intra-species syntenic genes in *G. coronaria* shows only one peak at ~1.5, similar to other Asteroideae species (Fig. 4c). This *K*_s peak is corresponding to the widely reported whole-genome triplication (WGT1) event occurred at 40–45 MYA in the ancestor of Asteraceae.^{12,15}

In addition to WGT1, the genomes of some *Chrysanthemum* species were reported to undergo recent WGD or segmental duplication (SD) events.^{41,44} For *G. coronaria* and *A. annua*, no sign of recent WGD or SD can be found from intra-species paralogue *K*_s distribution or inter-species macro-syntenic analysis (Fig. 4b and c). Besides, only 7.2% of duplicated genes are predicted to be derived from WGD or SD by MCScanX (Supplementary Table S13), similar to the BUSCO duplicate rate (6.4%) of genome (Supplementary Table S3), both of which indicates no recent WGD event occurred for *G. coronaria*. The nine chromosomes of *G. coronaria* are overall one-to-one to the nine chromosomes of *C. seticuspe* or *A. annua* in the inter-species synteny dot plot (Fig. 4b,

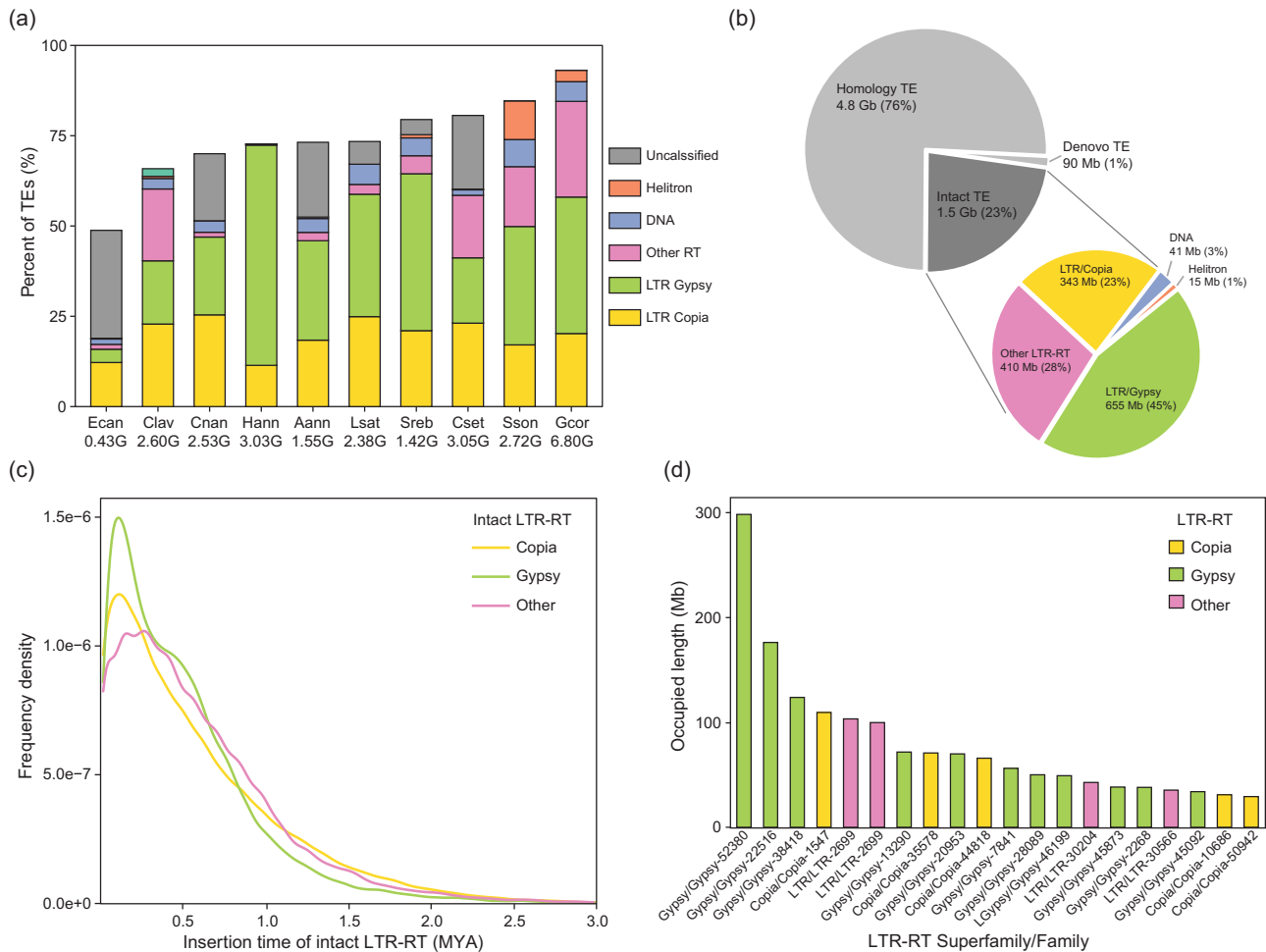


Figure 2. Recent expansion of TEs in *G. coronaria* genome. (a) Accumulated percent of different TE types in the genomes of 10 representative species in subfamily Asteroideae. The TE contents of *E. canadensis* (Ecan), *C. lavandulifolium* (Clav), *C. nankingense* (Cnan), *H. annuus* (Hann), *A. annua* (Aann), *L. sativa* (Lsat), *S. rebaudiana* (Sreb), *C. seticuspe* (Cset), and *S. sonchifolius* (Sson) are obtained from the corresponding genome papers. Numbers below species name indicate genome sizes. (b) Percentages of intact, homology, and Denovo TEs within all TEs in *G. coronaria* genome and the percentages of different TE types within intact TEs. Intact TEs were identified by structural characteristics using EDTA, homology TEs were identified by similarity to known TEs using RepeatMasker, and Denovo TEs were identified by copy number using RepeatModeler. (c) Distribution of the number of intact LTR-RTs along their insertion time in *G. coronaria* genome, estimated by the sequence divergence of LTRs using LTR_retriever. (d) Occupied length of 20 largest LTR-RT families in *G. coronaria* genome.

Supplementary Figs S8 and S9), indicating no recent WGD occurred in these Asteroideae species. The previously reported SD1 in *C. seticuspe* is also observed at the Ks peak ~ 0.2 . The inter-species Ks peak of *A. annua* vs. *C. seticuspe* is on the left of inter-species Ks peaks of *G. coronaria* vs. *A. annua* and *G. coronaria* vs. *C. seticuspe* (Fig. 4c), indicating that *G. coronaria* arose before the divergence of *A. annua* from *G. coronaria*. This is consistent with the above inferred phylogeny history and divergence time of these three species in the subtribe Artemisiinae (Fig. 3). Therefore, *G. coronaria* did not experience recent WGD after the shared WGT1 event in the ancestor of Asteraceae.

3.5. Expansion of 8-oxocitronellyl enol and isopiperitenone synthesis genes contribute to the special aroma of *G. coronaria*

The main aroma compounds in Asteraceae plants are volatile terpenoids, such as monoterpenoids (C10), sesquiterpenoids (C15), and diterpenoids (C20), all of which consist of multiple isoprene (C5) units.⁷ The precursors of terpenoid

synthesis in plants, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), are synthesized through mevalonate (MVA) pathway and methylerythritol phosphate (MEP)/deoxy-xylulose phosphate (DOXP) pathway (Fig. 5a, adapted from KEGG pathway 00900). In the MVA pathway, Acetyl-CoA is converted to IPP through five consecutive reactions in cytosol.⁴⁵ In the MEP/DOXP pathway, glyceraldehyde phosphate and pyruvate are converted to IPP and DMAPP through seven consecutive reactions in chloroplast.⁴⁵ IPP and DMAPP are further synthesized to geranyl diphosphate (GPP), geranyl geranyl diphosphate (GGPP), farnesyl diphosphate (FPP), etc. which can be converted to various monoterpenoids, sesquiterpenoids, and diterpenoids by various terpenoid synthases.⁸

The leaves and stems of *G. coronaria* have special aroma, and the major components of essential oils of *G. coronaria* are monoterpenoids.^{1,4} We identified all the genes involved in the synthesis of monoterpenoids and their precursors IPP and DMAPP, by searching the homologues of the known terpenoid synthesis genes in *G. coronaria* genome. There are totally 65

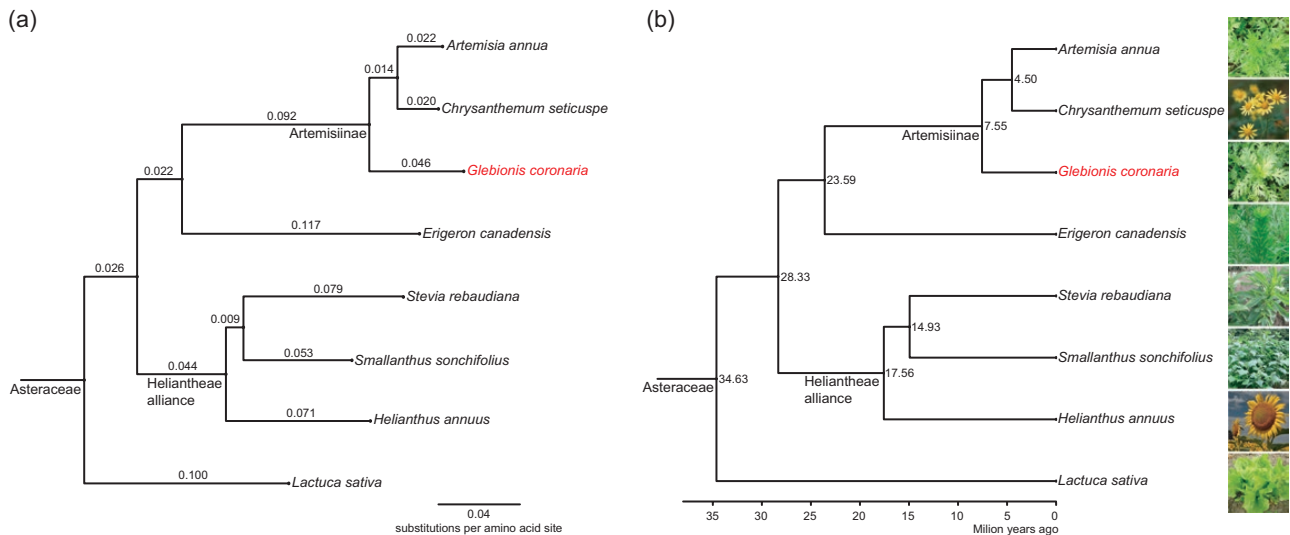


Figure 3. Phylogeny of *G. coronaria* and seven other representative species of Asteraceae. (a) Phylogenetic tree of 8 Asteraceae species built on the gene trees of 2,456 orthogroups with at least 88.9% of species having single-copy genes in any orthogroup, using the STAR and STRIDE methods invoked in OrthoFinder2 with parameters '-M msa -A mafft -T fasttree -1 -y'. Float numbers on branches indicate number of substitutions per amino acid site. (b) Time tree of eight Asteraceae species, estimated using RelTime method of MEGA11 with one calibration of the divergence of Coffea and Asteraceae at 95–106 MYA. Float numbers at internodes refer to the estimated species divergence time.

genes involved in the synthesis of terpenoid precursors (IPP, GPP, FPP, etc.), and 131 genes involved in the synthesis of monoterpenoids (Supplementary Tables S14 and S15) in *G. coronaria*. In particular, the copy numbers of four precursor synthesis genes of MVA pathway (*HMGCS*, *MVK*, *ipk*, and *ispH*) and two monoterpenoid synthesis genes *8-oxocitronellyl enol synthase* and *isopiperitenol dehydrogenase* are much higher in *G. coronaria* than those in the other five Asteroideae species (Fig. 5b, Supplementary Tables S14 and S15). Besides, most copies of these six genes were duplicated after the species divergence of *G. coronaria*, indicating species-specific expansions of these genes (Fig. 5c and d, Supplementary Figs S10–S13). The expansion of these genes may enable *G. coronaria* to produce more terpenoid precursors and monoterpenoids 8-oxocitronellyl enol and isopiperitenone.

Transcriptome sequencing showed that the expression levels of expanded terpenoid synthesis genes in *G. coronaria* were much higher in leaves and stems than those in roots (Supplementary Fig. S14). The major components of the volatile oils extracted from leaves and stems of *G. coronaria* are monoterpenoids geraniol, limonene, and their derivatives.^{5,46} The geraniol derivative 8-oxocitronellyl enol and the limonene derivative isopiperitenone have been reported to have special aroma and scent in *Chrysanthemum* plants.⁴⁷ Therefore, the expansion of synthesis genes of 8-oxocitronellyl enol and isopiperitenone may strengthen the synthesis of 8-oxocitronellyl enol and isopiperitenone and contributes to the special aroma of *G. coronaria*.

4. Discussion

This study generated a high-quality reference genome and annotation for *G. coronaria*, which has the largest genome size (6.8 Gb) among all the published genomes of diploid Asteraceae species. Our analysis showed the large genome size of *G. coronaria* is mainly caused by the recent explosion of LTR-RTs. Phylogenetic analysis of Asteroideae species supports the current taxonomic placement of *G.*

coronaria in *Glebionis* but not *Chrysanthemum*.⁴⁰ In the subtribe Artemisiinae, the genus *Glebionis* arose much earlier (~7.55 MYA) and evolved much faster than the genus *Chrysanthemum* and *Artemisia*. Synteny analysis and Ks distribution indicate that *G. coronaria* genome experienced the WGT1 at 40–45 MYA, shared with all Asteraceae species, and no recent WGD occurred. The *G. coronaria* genomic resources can be used as a model to study the relationship between TE activity and genome evolution, and promote the phylogeny, selection, and evolution studies of Asteraceae.

The special aroma of *G. coronaria* is an important agronomic trait of this vegetable. In the *G. coronaria* genome, we identified a total of 65 genes involved in the synthesis of terpenoid precursors, and 131 genes involved in the synthesis of monoterpenoids. The synthesis genes of monoterpenoids 8-oxocitronellyl enol and isopiperitenone show species-specific expansion in *G. coronaria*. The higher expressions of these expanded genes in leaves and stems of *G. coronaria* may produce more 8-oxocitronellyl enol and isopiperitenone and contribute to the special aroma of *G. coronaria*. Besides, *G. coronaria* also contains abundant unsaturated spiroketal enol compounds, tonghaosu, which has antifeeding activity and protects *G. coronaria* against insects.⁴⁸ Moreover, the essential oils of *G. coronaria* are also used as traditional medicine, due to their antimicrobial, antioxidant, antiviral, and antimycotic activities.^{5,49,50} The generated reference genome here will promote the in-depth studies of the genes involved in the metabolism of these bioactive compounds in *G. coronaria*, the breeding improvement of agronomic traits, and the application of *G. coronaria* in agrochemical and medical industries.

Funding

The present work was supported by the Agricultural Science and Technology Innovation Program, the Elite Young Scientist Program of Chinese Academy of Agricultural Sciences, and the fund of Key Laboratory of Shenzhen (ZDSYS201411181701111640).

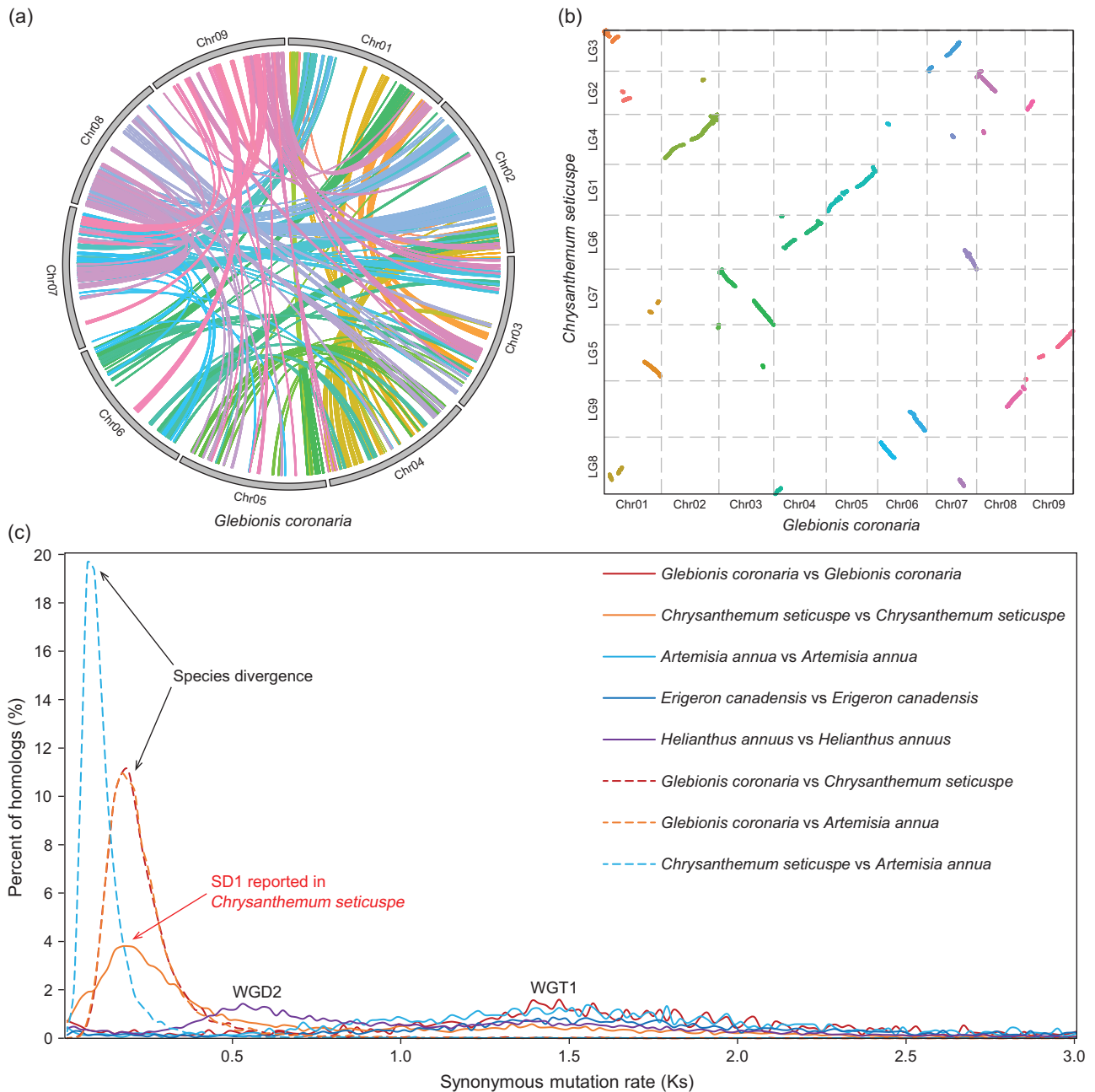


Figure 4. Whole-genome triplication of *G. coronaria*. (a) Circle plot of intra-species synteny blocks in *G. coronaria*, obtained using MCScanX with the all-vs-all alignment of protein-coding genes as input. Each line represents a synteny block with ≥ 5 genes, and all lines are coloured according to the corresponding chromosome pairs. (b) Dot plot of inter-species synteny between *C. seticuspe* and *G. coronaria*. Each dot represents a synteny block with ≥ 15 genes, and all dots are coloured according to the corresponding chromosome pairs. (c) Distribution of synonymous mutation rate (Ks) of homologue gene pairs for five Asteroidae species. Ks of intra-species paralogue or inter-species orthologue gene pairs obtained using MCScanX was calculated using KaKs_Calculator with the GMYN model. The previously reported whole-genome triplication (WGT1) for Asteraceae ancestor, whole-genome duplication (WGD2) for the ancestor of Heliantheae alliance, segmental duplication (SD1) for *C. seticuspe*, and the species divergence events are marked on the corresponding Ks peaks.

Acknowledgements

The authors thank Yazhi Qin for help in planting.

Authors' contributions

Anqi Wang and Sen Wang prepared samples for sequencing. Sen Wang and Anqi Wang assembled genome and performed structural and functional annotation. Sen Wang, Hengchao Wang, and Fan Jiang performed phylogeny and genome evo-

lution analyses. Anqi Wang, Sen Wang, and Dong Xu performed the analyses of terpenoid synthesis genes. Wei Fan supervised the whole research and provided constructive suggestions for data analysis. Sen Wang and Anqi Wang wrote the manuscript draft and all authors revised the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

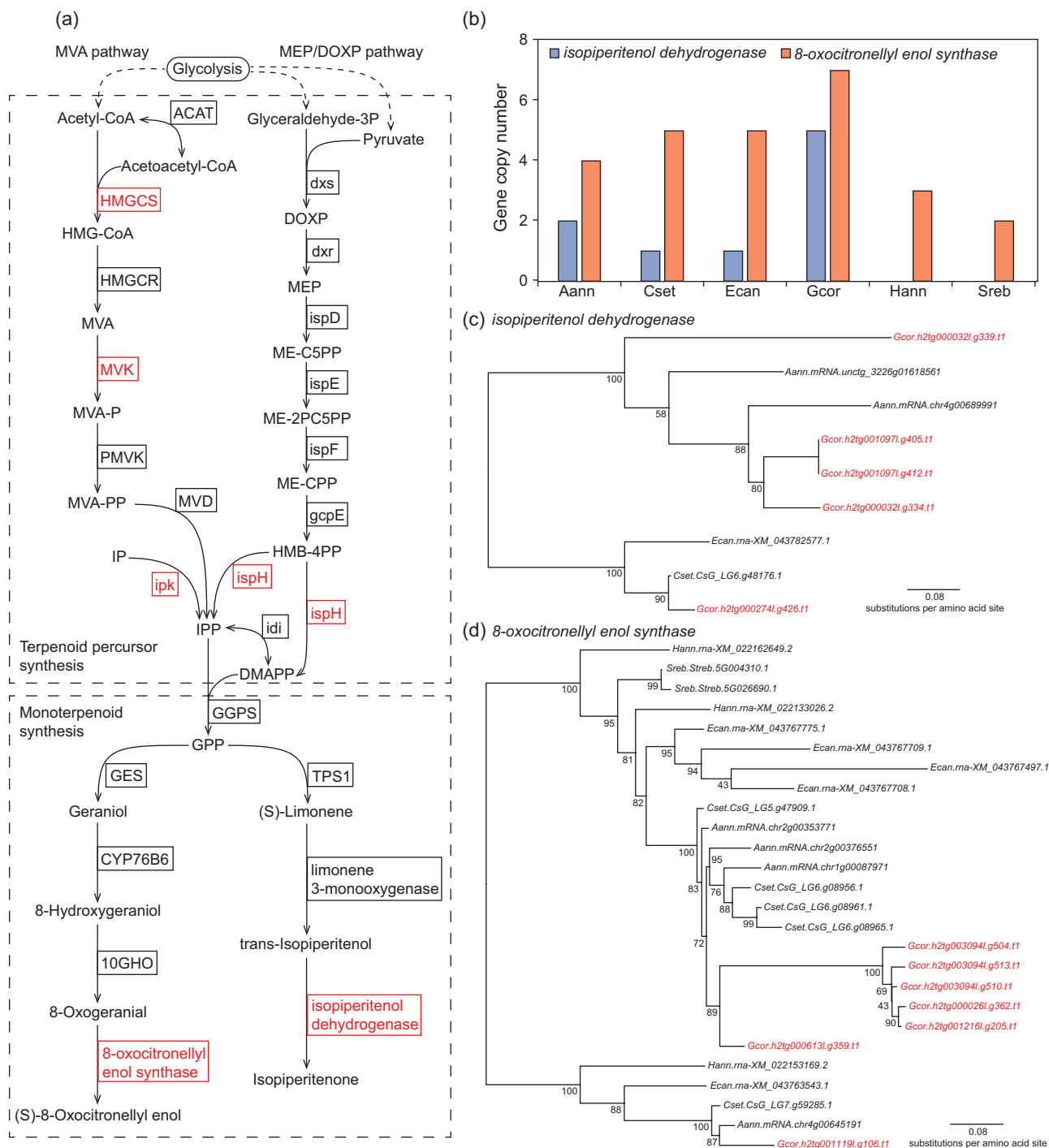


Figure 5. Expansion of terpenoid synthesis genes in *G. coronaria*. (a) Synthesis pathway of terpenoid precursor IPP from MVA and MEP/DOXP, and monoterpene synthesis. The full names of abbreviated enzymes are given in [Supplementary Tables S14 and S15](#). (b) Gene copy number of *isopiperitenol dehydrogenase* and *8-oxocitronellyl enol synthase* in six Asteroidae species *A. annua* (Aann), *C. seticuspe* (Cset), *E. canadensis* (Ecan), *G. coronaria* (Gcor), *H. annuus* (Hann), and *S. rebaudiana* (Sreb). (c) Phylogeny tree of *isopiperitenol dehydrogenase* gene in Asteroidae species Aann, Cset, Ecan, and Gcor (highlighted in red). (d) Phylogeny tree of *8-oxocitronellyl enol synthase* gene in Asteroidae species Aann, Cset, Ecan, Gcor (highlighted in red), Hann, and Sreb. Phylogeny tree was constructed using FastTree with the multiple protein sequence alignment generated by Muscle. Branch lengths indicate the phylogenetic distances (number of substitutions per amino acid site), and integers at internodes refer to the percent of bootstraps supporting the corresponding splits.

Data availability

The genome assembly of *G. coronaria* has been deposited at DDBJ/ENA/GenBank under the accession JANFOE000000000, and the genomic annotation data are

deposited at FigShare (10.6084/m9.figshare.20294385). The HiFi reads, Hi-C reads, and full-length transcripts have been deposited in NCBI SRA under the BioProject accession PRJNA851195.

Supplementary data

Supplementary data are available at *DNARES* online.

References

- Flamini, G., Cioni, P.L. and Morelli, I. 2003, Differences in the fragrances of pollen, leaves, and floral parts of garland (*Chrysanthemum coronarium*) and composition of the essential oils from flowerheads and leaves, *J. Agric. Food Chem.*, **51**, 2267–71.
- Turland, N.J. 2004, Proposal to conserve the name *Chrysanthemum coronarium* (Compositae) with a conserved type, *Taxon*, **53**, 1072–4.
- Li, W., Jiang, S., Wang, J., Yu, Y. and Zhu, Z. 2021, Complete chloroplast genome and phylogenetic analysis of *Glebionis coronaria* (L.) Cass. ex Spach (Asteraceae), *Mitochondrial DNA B Resour.*, **6**, 2693–4.
- Haouas, D., Cioni, P., Flamini, G., Ben Halima, M. and Ben Hamouda, M. 2016, Variation of chemical composition in flowers and leaves essential oils among natural population of Tunisian *Glebionis coronaria* (L.) Tzvelev (Asteraceae), *Chem. Biodivers.*, **13**, 1251–61.
- Servi, H. 2021, Chemical composition and biological activities of essential oils of two new chemotypes of *Glebionis* Cass., *Turk. J. Chem.*, **45**, 1559–66.
- Wijaya, E.J., Nathanael, J., Carolan, O., Adiyanto, S.A., Bun, W.B. and Sahamastuti, A.A.T. 2020, A review of phytochemical properties and therapeutic activities of *Glebionis coronaria*, *Indones. J. Life Sci.*, **2**, 44–55. ISSN: 2656-0682 (online).
- Bohlmann, J. and Keeling, C.I. 2008, Terpenoid biomaterials, *Plant J.*, **54**, 656–69.
- Pichersky, E. and Raguso, R.A. 2018, Why do plants produce so many terpenoid compounds? *New Phytol.*, **220**, 692–702.
- Yaoita, Y., Kikuchi, M. and Machida, K. 2012, Terpenoids and related compounds from plants of the family Compositae (Asteraceae), *Nat. Prod. Commun.*, **7**, 533–8.
- Muhlemann, J.K., Klempien, A. and Dudareva, N. 2014, Floral volatiles: from biosynthesis to function: floral volatiles, *Plant Cell Environ.*, **37**, 1936–49.
- Masyita, A., Mustika Sari, R., Dwi Astuti, A., et al. 2022, Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives, *Food Chem. X*, **13**, 100217.
- Reyes-Chin-Wo, S., Wang, Z., Yang, X., et al. 2017, Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce, *Nat. Commun.*, **8**, 1–11: 14953.
- Acquadro, A., Portis, E., Valentino, D., Barchi, L. and Lanteri, S. 2020, “Mind the gap”: Hi-C technology boosts contiguity of the globe artichoke genome in low-recombination regions, *G3*, **10**, 3557–64.
- Laforest, M., Martin, S.L., Bisailon, K., et al. 2020, A chromosome-scale draft sequence of the Canada fleabane genome, *Pest Manag. Sci.*, **76**, 2158–69.
- Badouin, H., Gouzy, J., Grassa, C.J., et al. 2017, The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution, *Nature*, **546**, 148–52.
- Liao, B., Shen, X., Xiang, L., et al. 2022, Allele-aware chromosome-level genome assembly of *Artemisia annua* reveals the correlation between ADS expansion and artemisinin yield, *Mol. Plant.*, **15**, 1310–28.
- Fan, W., Wang, S., Wang, H., et al. 2022, The genomes of chicory, endive, great burdock and yacon provide insights into Asteraceae palaeo-polyploidization history and plant inulin production. *Mol. Ecol. Resour.*, **22**, 3124–40.
- Garcia, S., Hidalgo, O., Jakovljević, I., et al. 2013, New data on genome size in 128 Asteraceae species and subspecies, with first assessments for 40 genera, 3 tribes and 2 subfamilies, *Plant Biosyst.*, **147**, 1219–27.
- Binghang, L., Shi, Y., Yuan, J., et al. 2013, Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv*, 1308.2012v2.
- Cheng, H., Jarvis, E.D., Fedrigo, O., et al. 2022, Haplotype-resolved assembly of diploid genomes without parental data, *Nat. Biotechnol.*, **40**, 1332–35.
- Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A. and Zdobnov, E.M. 2021, BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes, *Mol. Biol. Evol.*, **38**, 4647–54.
- Li, H. 2018, Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics*, **34**, 3094–100.
- Servant, N., Varoquaux, N., Lajoie, B.R., et al. 2015, HiC-Pro: an optimized and flexible pipeline for Hi-C data processing, *Genome Biol.*, **16**, 1–11: 259.
- Wang, S., Wang, H., Jiang, F., et al. 2021, EndHiC: assemble large contigs into chromosomal-level scaffolds using the Hi-C links from contig ends. *arXiv*, 2111.15411v1.
- Benson, G. 1999, Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res.*, **27**, 573–80.
- Ou, S., Su, W., Liao, Y., et al. 2019, Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline, *Genome Biol.*, **20**, 1–18: 275.
- Flynn, J.M., Hubley, R., Goubert, C., et al. 2020, RepeatModeler2 for automated genomic discovery of transposable element families, *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 9451–7.
- da Cruz, M.H.P., Domingues, D.S., Saito, P.T.M., Paschoal, A.R. and Bugatti, P.H. 2021, TERL: classification of transposable elements by convolutional neural networks, *Brief. Bioinform.*, **22**, 1–19.
- Ou, S. and Jiang, N. 2018, LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons, *Plant Physiol.*, **176**, 1410–22.
- Stanke, M., Schöffmann, O., Morgenstern, B. and Waack, S. 2006, Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources, *BMC Bioinf.*, **7**, 1–11: 62.
- Wu, T.D., Reeder, J., Lawrence, M., Becker, G. and Brauer, M.J. 2016, GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. In: Mathé, E. and Davis, S., eds. *Statistical genomics: methods and protocols*, pp. 283–334. Springer New York: New York, NY.
- Slater, G.S. and Birney, E. 2005, Automated generation of heuristics for biological sequence comparison, *BMC Bioinf.*, **6**, 1–11: 31.
- Buchfink, B., Xie, C. and Huson, D.H. 2015, Fast and sensitive protein alignment using DIAMOND, *Nat. Methods*, **12**, 59–63.
- Jones, P., Binns, D., Chang, H.-Y., et al. 2014, InterProScan 5: genome-scale protein function classification, *Bioinformatics*, **30**, 1236–40.
- Emms, D.M. and Kelly, S. 2019, OrthoFinder: phylogenetic orthology inference for comparative genomics, *Genome Biol.*, **20**, 1–14: 238.
- Tamura, K., Stecher, G. and Kumar, S. 2021, MEGA11: Molecular Evolutionary Genetics Analysis Version 11, *Mol. Biol. Evol.*, **38**, 3022–7.
- Wang, Y., Tang, H., Debarry, J.D., et al. 2012, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, *Nucleic Acids Res.*, **40**, e49.
- Wang, D., Zhang, Y., Zhang, Z., Zhu, J. and Yu, J. 2010, KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies, *Genomics Proteomics Bioinformatics*, **8**, 77–80.
- Paciolla, C., D’Emerico, S., Tommasi, F. and Scrugli, A. 2010, Karyomorphological and biochemical studies in *Glebionis coronaria* (L.) Spach and *Glebionis segetum* (L.) Fourreau from Italy, *Plant Biosyst.*, **144**, 563–7.
- Cano, E., Musarella, C.M., Cano-Ortiz, A., et al. 2017, Morphometric analysis and bioclimatic distribution of *Glebionis coronaria* s.l. (Asteraceae) in the Mediterranean area, *Phytokeys*, **81**, 103–26.
- Nakano, M., Hirakawa, H., Fukai, E., et al. 2021, A chromosome-level genome sequence of *Chrysanthemum seticuspe*, a model species for hexaploid cultivated chrysanthemum, *Commun. Biol.*, **4**, 1–11: 1167.

42. Xu, X., Yuan, H., Yu, X., et al. 2021, The chromosome-level *Stevia* genome provides insights into steviol glycoside biosynthesis, *Hortic. Res.*, **8**, 1–10: 129.
43. Denoeud, F., Carretero-Paulet, L., Dereeper, A., et al. 2014, The coffee genome provides insight into the convergent evolution of caffeine biosynthesis, *Science*, **345**, 1181–4.
44. Song, C., Liu, Y., Song, A., et al. 2018, The *Chrysanthemum nankingense* genome provides insights into the evolution and diversification of *Chrysanthemum* flowers and medicinal traits, *Mol. Plant*, **11**, 1482–91.
45. Lange, B.M., Rujan, T., Martin, W. and Croteau, R. 2000, Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes, *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 13172–7.
46. Ivashchenko, I.V. 2017, Chemical composition of essential oil and antimicrobial properties of *Chrysanthemum coronarium* (Asteraceae), *Biosyst. Divers.*, **25**, 119–23.
47. Xiao, Z., Fan, B., Niu, Y., et al. 2016, Characterization of odor-active compounds of various *Chrysanthemum* essential oils by gas chromatography–olfactometry, gas chromatography–mass spectrometry and their correlation with sensory attributes, *J. Chromatogr. B*, **100**, 152–62.
48. Yin, B.L., Fan, J.F., Gao, Y. and Wu, Y.L. 2003, Progress in molecular diversity of tonghaosu and its analogs, *Arkivoc*, **2**, 70–83.
49. Marongiu, B., Piras, A., Porcedda, S., et al. 2009, Chemical and biological comparisons on supercritical extracts of *Tanacetum cinerariifolium* (Trevir) Sch Bip. with three related species of chrysanthemums of Sardinia (Italy), *Nat. Prod. Res.*, **23**, 190–9.
50. Abd-Alla, H.I., Albalawy, M.A., Aly, H.F., Shalaby, N.M.M. and Shaker, K.H. 2014, Flavone composition and antihypercholesterolemic and antihyperglycemic activities of *Chrysanthemum coronarium* L., *Z. Naturforsch. C*, **69**, 199–208.