Molecular Plant
Research Article

CellPress
PARTNER JOURNAL

The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits

Chi Song[1,9], Yifei Liu[2,9], Aiping Song[3,9], Gangqiang Dong[4,9], Hongbo Zhao[5,9], Wei Sun[1], Shyam Ramakrishnan[6], Ying Wang[7], Shuaibin Wang[8], Tingzhao Li[4], Yan Niu[7], Jiafu Jiang[3], Bin Dong[5], Ye Xia[7], Sumei Chen[3], Zhigang Hu[2], Fadi Chen[3,*] and Shilin Chen[1,*]

[1]Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

[2]College of Pharmacy, Hubei University of Chinese Medicine, Wuhan 430065, China

[3]College of Horticulture, Nanjing Agricultural University, Key Laboratory of Landscape Agriculture, Ministry of Agriculture, Nanjing 210095, China

[4]Amway (China) Botanical R&D Center, Wuxi 214115, China

[5]Department of Ornamental Horticulture, School of Landscape Architecture, Zhejiang Agriculture and Forestry University, Hangzhou 311300, China

[6]Amway Corporation, Buena Park, CA 90622-5940, USA

[7]Wuhan Benagen Tech Solutions Company Limited, Wuhan 430070, China

[8]Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, The Chinese Academy of Sciences, Guangzhou 510650, China

[9]These authors contributed equally to this article.

***Correspondence: Fadi Chen (chenfd@njau.edu.cn), Shilin Chen (slchen@icmm.ac.cn)**

https://doi.org/10.1016/j.molp.2018.10.003

ABSTRACT

The Asteraceae (Compositae), a large plant family of approximately 24 000–35 000 species, accounts for ~10% of all angiosperm species and contributes a lot to plant diversity. The most representative members of the Asteraceae are the economically important chrysanthemums (*Chrysanthemum* L.) that diversified through reticulate evolution. Biodiversity is typically created by multiple evolutionary mechanisms such as whole-genome duplication (WGD) or polyploidization and locally repetitive genome expansion. However, the lack of genomic data from chrysanthemum species has prevented an in-depth analysis of the evolutionary mechanisms involved in their diversification. Here, we used Oxford Nanopore long-read technology to sequence the diploid *Chrysanthemum nankingense* genome, which represents one of the progenitor genomes of domesticated chrysanthemums. Our analysis revealed that the evolution of the *C. nankingense* genome was driven by bursts of repetitive element expansion and WGD events including a recent WGD that distinguishes chrysanthemum from sunflower, which diverged from chrysanthemum approximately 38.8 million years ago. Variations of ornamental and medicinal traits in chrysanthemums are linked to the expansion of candidate gene families by duplication events including paralogous gene duplication. Collectively, our study of the assembled reference genome offers new knowledge and resources to dissect the history and pattern of evolution and diversification of chrysanthemum plants, and also to accelerate their breeding and improvement.

Key words: *Chrysanthemum*, genome evolution, whole-genome duplication, nanopore sequencing, flower evolution, medicinal plant

Song C., Liu Y., Song A., Dong G., Zhao H., Sun W., Ramakrishnan S., Wang Y., Wang S., Li T., Niu Y., Jiang J., Dong B., Xia Y., Chen S., Hu Z., Chen F., and Chen S. (2018). The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits. Mol. Plant. **11**, 1–10.

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

Please cite this article in press as: Song et al., The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits, Molecular Plant (2018), https://doi.org/10.1016/j.molp.2018.10.003

INTRODUCTION

Chrysanthemums are among the most famous traditional Chinese flowers and are popular around the world as cut flowers and pot plants. Their commercial value is largely attributed to the diversity of ornamental traits created by targeted breeding, which has been primarily focused on petal colors and the shapes of ray florets (Shinoyama et al., 2012). Chrysanthemums are also important as health foods and as anti-inflammatory herbs in traditional Chinese medicine. Recent pharmacological studies have focused on their various biological effects including antioxidative, cardiovascular protective, vasorelaxant, and anti-human immune deficiency virus activities (Kim and Lee, 2005; Woo et al., 2008; Wang et al., 2014; Sun et al., 2015). However, despite their economic importance and beneficial properties, chrysanthemum evolution and the underlying molecular basis of the diversity of chrysanthemums are poorly understood. Multiple lines of evidence indicate that both polyploidy and hybridization are prevalent in the genus *Chrysanthemum*, suggesting that recurrent hybridizations between several wild progenitor species contributed to the evolutionary novelty of the present hybrid chrysanthemum cultivar *Chrysanthemum* × *morifolium* (Yang et al., 2006; Huang et al., 2016; Ma et al., 2016). Interestingly, the diploid genome of *Chrysanthemum nankingense* (Nakai) Tzvel (Chinese name *Juhuanao*, $2n = 2x = 18$) may represent one of the key progenitor genomes (Yang et al., 2006; Ren et al., 2014).

Despite tremendous improvements in sequencing technologies, the analysis of entire plant genomes remains comparatively difficult due to their large size and a high repeat content (Jiao and Schneeberger, 2017). However, updated single-molecule sequencers with exceptional long-read capabilities have ushered in a new era of genome sequencing whereby the read lengths exceed those of most genomic repeats. The mini-sequencer MinION from Oxford Nanopore Technologies (Oxford, UK) is notable for its low-cost sequencing and innovative nanopore sequencing technology (de Lannoy et al., 2017), and provides a powerful platform for analyzing and assembling plant and human genomes that is comparable with other single-molecule real-time sequencing platforms, such as Pacific Biosciences (Menlo Park, CA, USA) sequencers (Schmidt et al., 2017; Jain et al., 2018). Hence, nanopore devices have emerged as a suitable option for large sequencing projects that require long-read sequencers (Leggett and Clark, 2017).

Here, we deployed nanopore sequencing technology to perform *de novo* sequencing and assembly of the genome of *C. nankingense*, an out-crossing perennial plant with a highly complex genome of approximately 3.07 Gb that features high heterozygosity and long and highly similar repeats (Supplemental Figures 1–3 and Supplemental Note 3). We resolved most of the *C. nankingense* genome, including both its genes and repetitive elements, demonstrating the efficiency of this approach in assembling a complex genome. We performed phylogenetic and comparative genomic analyses to investigate whole-genome duplication (WGD) events and expansion/contraction of gene families. The evolutionary patterns observed at both the whole-genome and the gene family levels provide unprecedented insights into the diversification of chrysanthemum genomes in the context of both ornamental and medicinal traits. Thus, the *C. nankingense* reference genome will be a valuable resource for genetic studies and improvement of chrysanthemums, including genome-assisted breeding of novel cultivars with desired traits and exceptional ecological fitness.

RESULTS

Genome Sequencing, Assembly, and Annotation

We performed genomic sequencing with an Oxford Nanopore device using a total of 39 flow cells, yielding approximately 5.7 million single-molecule nanopore long reads (average read length, 17.7 kb) with a total data volume of 105.2 Gb (Supplemental Table 1), of which 99.5 Gb were retained for subsequent analysis after data filtering using the Oxford Nanopore Metrichor base caller. A hybrid assembly strategy was applied in which the nanopore long reads were corrected and assembled prior to using 362.3 Gb of Illumina short reads for further polishing (Supplemental Table 2). The assembly yielded 24 051 sequence contigs (N50 = 130.7 kb) with a total size of 2.53 Gb, representing ~82% of the estimated genome size. The completeness of the genome assembly was further evaluated and verified (Supplemental Tables 3 and 4; Supplemental Note 4).

An integrated strategy was applied to predict the protein-coding gene content of the *C. nankingense* genome using transcripts generated from various tissues for validating and constructing gene models (Supplemental Table 5). After removing nonfunctional annotations, a final set of 56 870 protein-coding genes was retained, with an average coding-sequence length of 1.0 kb and an average of 4.6 exons per gene (Supplemental Table 6). Among the genomes in the Asteraceae family, the number of genes in the *C. nankingense* genome was similar to the numbers estimated for the sunflower (52 232 genes) and horseweed (44 592 genes) genomes, but higher than the number of genes in the lettuce (38 919 genes) and artichoke (28 310 genes) genomes. Among the annotated *C. nankingense* genes, 50 549 (~88.9%) were functionally classified (Supplemental Table 7) and 48 027 (~84.4%) were assigned to transcripts. We further annotated noncoding RNA (ncRNA) genes (Supplemental Table 8), yielding 2076 transfer RNA (tRNA) genes, 55 ribosomal RNA (rRNA) genes, 1504 small nuclear RNA genes, and 579 microRNA genes.

Repetitive Content and Recent Bursts of LTR Retrotransposons

Through a combination of approaches, we annotated 69.6% of the assembly as repetitive elements (Supplemental Table 9). The long terminal repeat retrotransposons (LTRs) were the most abundant, and most LTRs were LTR/*Copia* elements, which occupied 25.4% of the genome, followed by the LTR/*Gypsy* repeats (21.5%, Supplemental Table 9). Besides the main groups of LTR elements, 3.2% of the genome was annotated as DNA elements and 1.3% as long interspersed nuclear elements, whereas the remainder was either assigned to other repeat families or could not be assigned (Supplemental Table 9). Among the several sequenced Asteraceae species, the repetitive elements identified in the *C. nankingense* genome were comparable with those found in sunflower (Badouin et al., 2017) and lettuce (Reyes-Chin-Wo et al., 2017), which also have genomes with a high content of repetitive DNA.

Please cite this article in press as: Song et al., The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits, Molecular Plant (2018), https://doi.org/10.1016/j.molp.2018.10.003

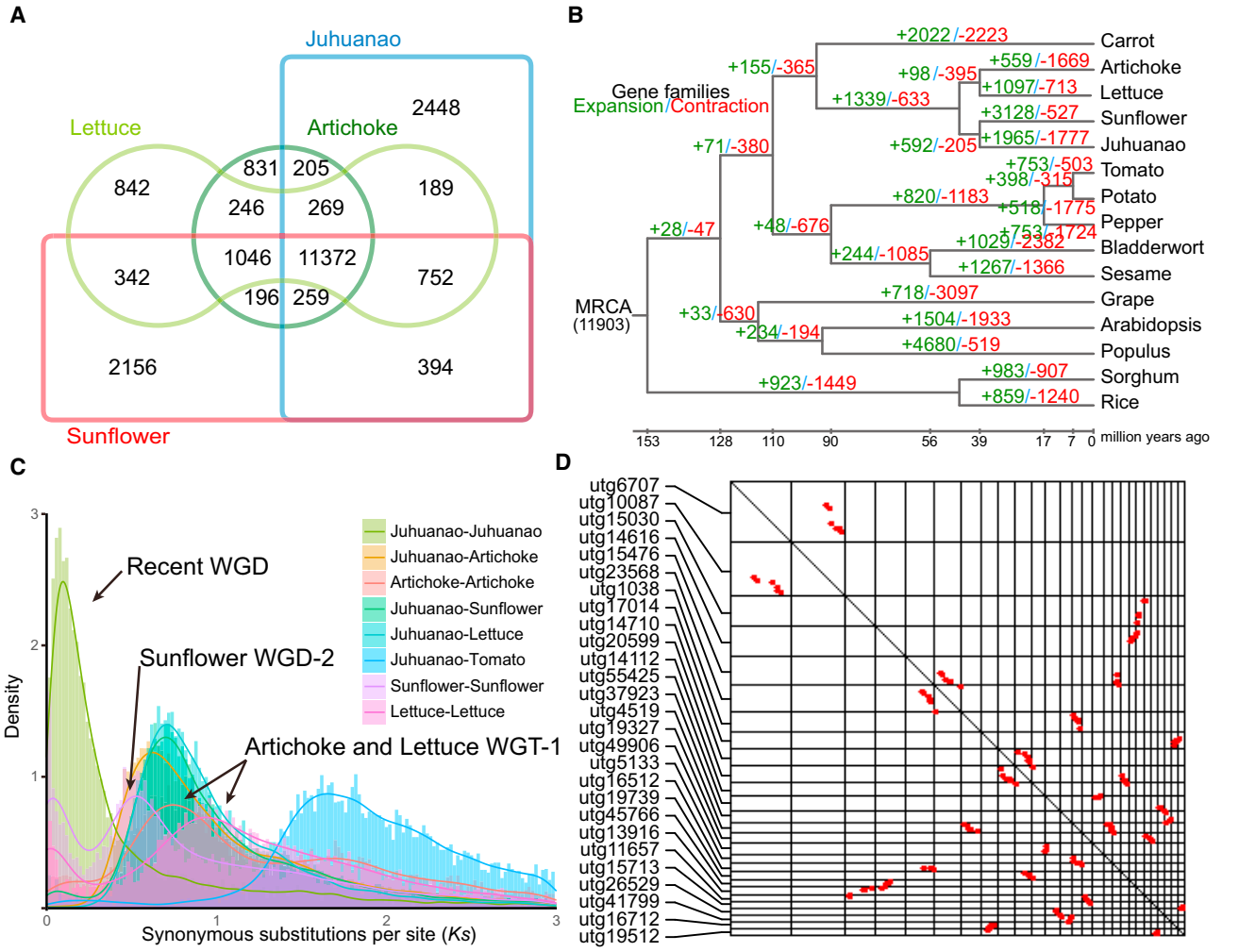

Figure 1. Comparative Genomic Analysis of *Chrysanthemum* and Other Plant Species.

(A) Number of gene families shared between Juhuanao (*C. nankingense*) and three other species in the Asteraceae family.

(B) Expansion and contraction of gene families among 15 plant species. The number at the root (11 903) denotes the total number of gene families predicted in the most recent common ancestor (MRCA). The divergence time was estimated for each node (Mya).

(C) Distribution of synonymous substitution rates (*Ks*) for pairs of syntenic paralogs in Juhuanao and three other plants.

(D) Dot plots of paralogs identified across scaffolds in the *C. nankingense* genome.

We further identified 30 639 and 24 276 intact *Gypsy* and *Copia* retrotransposons, respectively. The estimated burst of *Copia* retrotransposition occurred earlier than the burst of *Gypsy* retrotransposition, although the amplification of both peaked around 1 million years ago (Mya) (Supplemental Figure 4 and Supplemental Note 5) after the divergence of chrysanthemum from sunflower (see next section). The predicted time of the LTR retrotransposon burst in *C. nankingense* was comparable with that recently reported in *Triticum urartu* (~1 Mya), which carries a large wheat progenitor genome (~5 Gb) (Ling et al., 2018). Thus, these data suggest that recent insertions of LTR elements may have contributed to the increase in chrysanthemum genome size.

Genome Evolution and Expansion of Gene Families

To investigate the relationship between gene families and distinct traits of chrysanthemum, we compared the *C. nankingense* genome with 14 other plant genomes (Supplemental Table 10). Using the predicted proteomes of these plants, a total of 39 414 orthologous gene families consisting of 418 703 genes were identified, including a core set of 161 163 genes belonging to 5278 gene families shared among the 15 plant species and 11 372 gene families shared among four Asteraceae species (Figure 1A; Supplemental Figure 5; Supplemental Table 11). We found a total of 1939 gene families containing 8009 genes unique to *C. nankingense* (Supplemental Table 12). In addition, gene family evolution analysis revealed that 1965 gene families in *C. nankingense* underwent expansion, whereas 1777 gene families underwent contraction (Figure 1B and Supplemental Table 13). Among the *C. nankingense* expanded gene families, functional annotation demonstrated that they were mainly enriched in gene ontology (GO) functional categories such as transferase activity (GO: 0016758) and terpene synthase activity (GO: 0010333), suggesting possible roles related to the production of secondary metabolites. We constructed a phylogenomic tree and estimated the divergence times of these plants using genes derived from

Please cite this article in press as: Song et al., The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits, Molecular Plant (2018), https://doi.org/10.1016/j.molp.2018.10.003

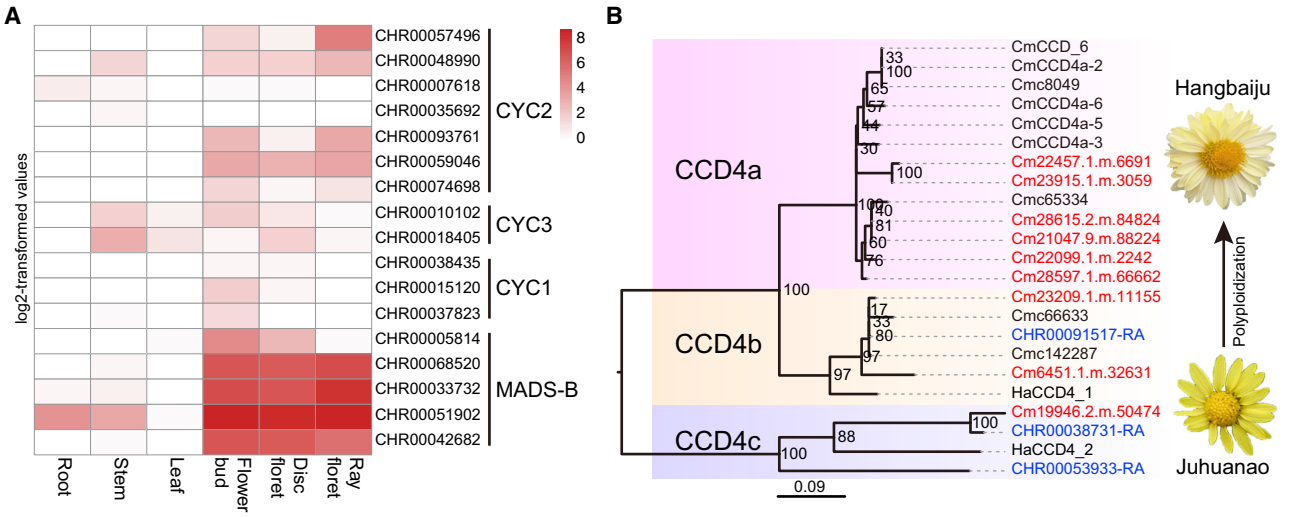

Figure 2. Genes Involved in Chrysanthemum Flower Evolution.

(A) Flower-specific expression of *CYC* and MADS-box B class genes.

(B) Evolution of *CCD4a* genes in relation to recent hybridization and polyploidization of domesticated chrysanthemums. Gene copies derived from the *C. nankingense* genome and the *C.* × *morifolium* "Hangbaiju" genome are marked in blue and red, respectively. See Supplemental Figure 10 for a full version of this tree with additional homologous genes from other representative plant species.

500 single-copy families. The estimated divergence time of chrysanthemum (Juhuanao) and sunflower was ~38.8 Mya (Figure 1B and Supplemental Figure 6). Their most recent common ancestor (MRCA) diverged from the MRCA of artichoke and lettuce ~45.7 Mya (Supplemental Figure 6).

We further investigated WGD events during chrysanthemum evolution. We identified 8176 paralogous gene pairs that covered 21.4% of the *C. nankingense* genome assembly. We used the *Ks* values of the duplicated genes to calculate the age distribution of the duplication events and found a peak at ~0.1 (Figure 1C), suggesting that the most recent WGD event occurred ~5.8 Mya. We performed the same analysis with the genome data of sunflower, artichoke, lettuce, and tomato to determine whether the recent WGD event in *C. nankingense* was species specific or shared between different plant species. This analysis showed that the recent WGD event in *C. nankingense* was distinct from the WGD-2 event (~31.8 Mya) identified in sunflower (Figure 1C), although both chrysanthemum and sunflower share a whole-genome triplication (WGT-1, ~57 Mya), an event that is also shared by lettuce and artichoke (Badouin et al., 2017) and that is superimposed on the ancestral palaeohexaploidy event that occurred in all eudicots (WGT-γ) (Salse, 2016). The inferred WGD event in the *C. nankingense* genome was further supported by dot-plot analysis of representative scaffolds in which numerous paralogs derived from this event were identified (Figure 1D). Using transcriptome data for the polyploid chrysanthemum cultivar *C.* × *morifolium* "Hangbaiju" (Chinese name *Hangbaiju*) (Supplemental Table 14), we identified additional genomic duplications associated with domesticated chrysanthemums (Supplemental Figure 7).

Expansion of Genes Involved in Chrysanthemum Flower Evolution

The genome assembly of *C. nankingense* allows us to identify the complete catalog of genes involved in pathways underlying important biological features. Asteraceae species have typical composite flowers consisting of two types of ray florets and disc florets. The *TCP* transcription factor clade *CYCLOIDEA2* (*CYC2*) appears to be crucial for the regulation of ray floret development in Asteraceae (Garcês et al., 2016), whereas the MADS-box B class genes *DEFICIENS* (*DEF*)/*APETALA3* (*AP3*) and *GLOBOSA* (*GLO*)/*PISTILLATA* (*PI*) participate in controlling petal and stamen development. We found seven *CYC2*-like genes in *C. nankingense*, which is comparable with the eight copies identified in sunflower, but more than the three and four copies identified in lettuce and artichoke, respectively, suggesting that *CYC2*-like genes play crucial roles in composite flower development in the two closely related taxa chrysanthemum and sunflower (Supplemental Figure 8). Among the MADS-box B class genes, the *GLO*/*PI* clade appeared to expand in chrysanthemum (Supplemental Figure 9). We further examined the transcript levels of these genes in different chrysanthemum tissues. We observed significant expression of both *CYC* and MADS-box B class genes in flowers, but *CYC2* genes were the only *CYC* genes expressed in ray florets (Figure 2A).

In chrysanthemums, the carotenoid cleavage dioxygenase-related gene *CCD4a* is specifically expressed in ray petals. Its translation product converts carotenoids into colorless compounds, resulting in a white petal color (Ohmiya et al., 2006). Interestingly, when comparing *C. nankingense* and *C.* × *morifolium* cultivars, we found no *CCD4a* gene copies in the *C. nankingense* genome but multiple copies of *CCD4a* in the *C.* × *morifolium* transcriptomes (Figure 2B and Supplemental Figure 10) including in *C.* × *morifolium* "Hangbaiju", one of the cultivars investigated in this study (with white ray florets, Supplemental Figure 1), suggesting that the evolution of the *CCD4a* clade was probably associated with the recent hybridization and polyploidization event in cultivated

Please cite this article in press as: Song et al., The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits, Molecular Plant (2018), https://doi.org/10.1016/j.molp.2018.10.003

chrysanthemums. Analysis of transcriptomic data for both leaf and flower tissues of *C.* × *morifolium* "Hangbaiju" further revealed the specific expression of *CCD4a* genes in flowers (Supplemental Figure 11).

Genes Involved in Flavonoid Biosynthesis

A significant number of flavonoids identified in diverse chrysanthemum species and cultivars (Supplemental Figure 12 and Supplemental Table 15) are the active constituents in medicinal and food preparations of chrysanthemum materials (Lin and Harnly, 2010). We annotated genes involved in flavonoid biosynthesis in the *C. nankingense* genome and performed a comparative analysis with other annotated plant genomes, including those of sunflower, lettuce, and artichoke, which belong to the Asteraceae family. Comparison of the gene distribution patterns revealed the enrichment of genes within flavonoid biosynthesis-associated pathways in the *C. nankingense* genome (Supplemental Figure 13 and Supplemental Table 16). For example, two key genes in the flavonoid biosynthesis pathway (ko00941), chalcone synthase (*CHS*) and chalcone isomerase (*CHI*), are significantly expanded in the *C. nankingense* genome (17 *CHS* genes and eight *CHI* genes, Supplemental Figure 13). Further analysis revealed the differential expression of these genes among different tissues but significantly high expression levels in flowers (Supplemental Figure 14), suggesting the spatial and temporal regulation of flavonoid synthesis.

The expression of flavonoid biosynthetic genes is regulated by dimers of MYB and bHLH, which act as transcriptional activators or repressors (Dubos et al., 2010; Goossens et al., 2017). In the *C. nankingense* genome, we identified 187 MYB and 203 bHLH genes, which was a little less than the 213 MYBs and 223 bHLHs annotated in the sunflower genome, but significantly more than the number of copies found in both the lettuce and artichoke genomes (Supplemental Table 17). Members of the R2R3 MYB subgroups 5, 6, and 7 were identified as activators of proanthocyanidin, anthocyanidin, and flavonol pathways in *Arabidopsis thaliana* and other plants (Huang et al., 2013; Wang et al., 2017), whereas the subgroup 4 members act as repressors of phenylpropanoid pathways (Dubos et al., 2010). Based on the phylogenetic reconstruction and subgroup delimitation of *C. nankingense* MYB genes and previously characterized genes from *Arabidopsis*, we identified two members each of subgroups 6, 7, and 4, in the *C. nankingense* genome, five of which were expressed in different tissues, suggesting the potentially diverse roles of MYBs in the regulation of flavonoid biosynthesis (Supplemental Figure 15).

We also investigated genes encoding UDP-glucuronosyl and -glucosyltransferase (UGT) because they are significantly expanded (PF00201 and GO: 0016758, $P < 0.001$, Supplemental Table 13) and catalyze a decoration step—flavonoid glycosylation—and thus are potentially involved in the diversification of flavonoids in chrysanthemum (Supplemental Figure 12). A phylogenetic analysis identified numerous flavonoid UGT gene copies in the C. *nankingense* genome, which was consistent with those found in the phytochemical data (Supplemental Figure 16).

Diversification of Terpenoid Biosynthesis-Associated Genes

Volatiles found in the aerial parts of chrysanthemum represent another type of valued constituent used in medicinal preparations and food. Terpenes are the predominant class of volatiles (Wang et al., 2008). The primary drivers of terpene diversification are the terpenoid synthases (TSs), which include *trans*-isoprenyl diphosphate synthases and squalene synthases (SQSs), terpene synthases (TPSs), and triterpene cyclases (TCCs), which generate scaffold diversity, and cytochrome P450-dependent monooxygenases (CYPs), which modify and further diversify the compound scaffolds (Boutanaev et al., 2015). Among the chrysanthemum expanded gene families, the genes involved in the biosynthesis of sesquiterpenoids, triterpenoids (ko00909, $P < 0.001$), and monoterpenoids (ko00902, $P < 0.001$) are highly enriched. The TS and CYP families are among the most highly enriched functional categories (Supplemental Table 13). In total, we identified 219 TS genes (7 SQSs, 158 TPSs, and 54 TCCs) and 708 CYP genes in the *C. nankingense* genome, which indicates that there was a significant expansion of TPS genes, specifically in the TPS-a and TPS-b subfamilies (Figure 3A). Compared with chrysanthemum, significantly smaller numbers of TS and CYP genes were identified in the other three Asteraceae genomes, sunflower, lettuce, and artichoke (Supplemental Table 18).

Terpenoid biosynthesis genes in plants are generally tandemly organized in metabolic gene clusters (Osbourn, 2010; Nützmann and Osbourn, 2014). We examined TS and CYP gene clusters in the *C. nankingense* genome, and identified 186 gene clusters, including seven containing TS/CYP gene pairs representing five different gene combinations (Figure 3B and Supplemental Table 19). Surprisingly, in addition to the TS/CYP combinations already identified in other sequenced eudicots (Boutanaev et al., 2015), we found two new combinations, TPS-a/CYP99 and TPS-g/CYP79/CYP76, in chrysanthemum (Figure 3B). We further investigated the expression of genes in these TS/CYP pairs (Figure 3C). Interestingly, several TPS and CYP genes were co-expressed in different chrysanthemum tissues. The *CHR00048430* and *CHR00048432* genes in the TPS-a/CYP99 cluster were mainly co-expressed in the roots, whereas the two genes *CHR00011805* and *CHR00011811* in the TPS-g/CYP79/CYP76 cluster were co-expressed in flowers (specifically in the disc floret) (Figure 3C).

DISCUSSION

The *de novo* assembly of large genomes with a high degree of heterozygosity and repeat content remains a challenge (Claros et al., 2012; Kajitani et al., 2014). However, sequencing platforms such as the PacBio RS II, Sequel, and Oxford Nanopore systems are designed to generate long reads, which greatly facilitates the sequence assembly process and allows high-quality assemblies to be generated. Using nanopore long reads, we assembled ~82% of the estimated 3.07 Gb genome of *C. nankingense*, (Supplemental Tables 3 and 4; Supplemental Note 4). The completeness and quality of our genome assembly is comparable to that of the sunflower genome (80% of the estimated genome size) assembled using long reads generated with the PacBio RS II platform (Badouin

Please cite this article in press as: Song et al., The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits, Molecular Plant (2018), https://doi.org/10.1016/j.molp.2018.10.003

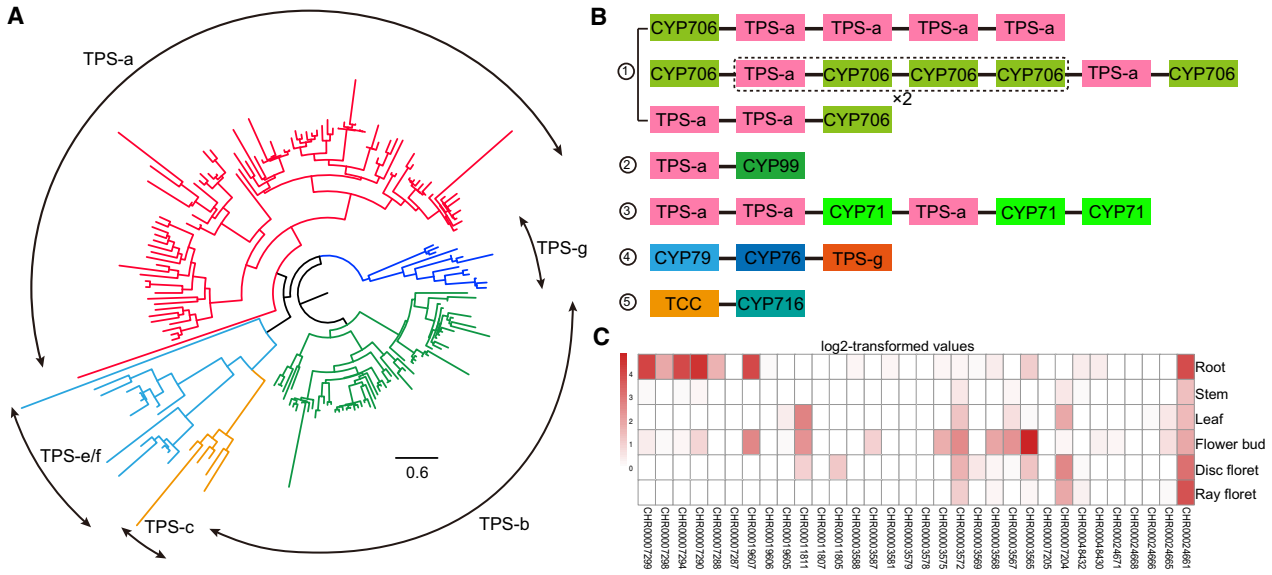

Figure 3. The TPS Genes and TS/CYP Gene Clusters Identified in the *C. nankingense* Genome.
(A) The phylogenetic tree shows a total of 158 TPSs. The TPS subfamilies are labeled.
(B) Seven clusters of TS/CYP genes representing five different gene combinations were identified. These include the eudicot-type TPS-a/CYP706 and two new combinations, TPS-a/CYP99 and TPS-g/CYP79/CYP76. The tandem unit TPS-a/CYP706/CYP706/CYP706 is repeated twice.
(C) Tissue-specific expression of genes that belong to the seven TS/CYP gene clusters.

et al., 2017). In theory using a hybrid strategy to assemble a genome with relatively high heterozygosity will generate a consensus between different haplotypes. In practice, we expect this to be not present in the final assembly of chrysanthemum, as the variation resulting from heterozygosity is much lower than the nanopore error rate. That is, Pilon will treat SNPs and small indels as sequencing errors to be corrected if they are not present in the Illumina data. Currently, the application of the Oxford Nanopore sequencing platform to plants remains in its infancy (Schmidt et al., 2017; Michael et al., 2018). Further improvements of the nanopore chemistry (e.g., the recently developed R9.4 nanopores) and the sequencing accuracy prior to polishing will potentially expand the application of this platform in plants, and allow the avoidance altogether of a hybrid assembly incorporating short reads (Schmidt et al., 2017).

We identified an abundance of repetitive elements in the *C. nankingense* genome (69.6% of the assembly), among which the LTR retrotransposons (*Copia* and *Gypsy*) were the most abundant (Supplemental Table 9). A high content of repetitive elements is a shared feature among several known large genomes (~3 Gb) of the Asteraceae family. In sunflower, more than 75% of the genome consists of LTRs with a relatively high proportion (59%) of *Gypsy* repeats (Badouin et al., 2017). In lettuce, 74.2% of the genome is occupied by repetitive regions, of which the LTR/*Gypsy* subfamilies (33.9%) are the most abundant (Reyes-Chin-Wo et al., 2017). We found recent bursts of both *Copia* and *Gypsy* retrotransposons in the *C. nankingense* genome (Supplemental Figure 4). In the tea (Xia et al., 2017) and *Picea abies* (Nystedt et al., 2013) genomes, long-lasting and uninterrupted LTR retrotransposon bursts may have led to extreme increases in genome size due to the lack of efficient DNA removal mechanisms. However, the recent LTR retrotransposon bursts in the *C. nankingense* genome might have also significantly contributed to genome size. These observations indicate the critical role of transposable elements in the evolution of large plant genomes.

Rounds of WGD or polyploidization have played an important role in plant genome evolution (Panchy et al., 2016). In our study, the most recent WGD event in the *C. nankingense* genome turned out to be distinct from the WGD-2 event reported for the sunflower genome (Figure 1C), suggesting that a lineage-specific WGD occurred after the chrysanthemum-sunflower divergence from the MRCA. Moreover, chrysanthemum and other Asteraceae plants including sunflower, lettuce, and artichoke shared the WGT-1 (Badouin et al., 2017) and the ancestral palaeohexaploidy WGT-γ event (Salse, 2016). Therefore, the *C. nankingense* genome underwent at least three rounds of WGDs, resulting in an expansion of genomic size and content. Notably, both natural and cultivated chrysanthemums are mixed populations of diploid and polyploid individuals, illustrating the prevalence of polyploidy in this large evolutionary lineage. We identified an additional genomic duplication event in *C.* × *morifolium* based on analysis of transcriptomic data (Supplemental Figure 7), which is possibly indicative of a hexaploidization event in *C.* × *morifolium*. However, it is unclear whether this event is actually involved in the autopolyploidy or allopolyploidy, because the domestication of different *C.* × *morifolium* cultivars was generally independent due to variable breeding and hybridizing processes (Yang et al., 2006).

C. nankingense genomic data and the corresponding transcriptomic data provide important insights into the ancient WGD events and the recent tandem duplication and recombination events that potentially contributed to trait diversification among chrysanthemums. The duplicated *CYC2*

Please cite this article in press as: Song et al., The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits, Molecular Plant (2018), https://doi.org/10.1016/j.molp.2018.10.003

clade genes are known to be involved in Asteraceae inflorescence development. These genes are associated with the complex capitulum structure and the regulation of flower differentiation including ray floret development (Tähtiharju et al., 2012; Juntheikki-Palovaara et al., 2014; Garcês et al., 2016). Our finding that there are seven *CYC2*-like genes in *C. nankingense* further supports this observation (Supplemental Figure 8). Moreover, multiple *CCD4a* gene copies are found in cultivated chrysanthemum genomes but not in the *C. nankingense* progenitor genome, which indicates that the *CCD4a* genes contribute to the white color formation in chrysanthemum petals (Ohmiya et al., 2006; Sasaki et al., 2017) and suggests that the evolution of this gene clade is linked to a recent polyploidy event in cultivated chrysanthemums.

Furthermore, we elucidated the expansion of gene families in relation to the diversification of secondary metabolites, including those involved in both flavonoid and terpenoid biosynthesis (Figure 3 and Supplemental Figure 13). Specifically, numerous tandem duplications of genes giving rise to the widespread presence of both TS and CYP gene clusters in the *C. nankingense* genome greatly contributed to the terpenoid diversity in chrysanthemum (Figure 3 and Supplemental Table 17). Among the known plant genomes, the *C. nankingense* genome appears to carry the highest number of genes encoding TPS (Figure 3A) (e.g., 113 TPSs in *Eucalyptus grandis*) (Myburg et al., 2014). To assess the potential benefit of TPS and CYP gene expansion and identify additional recombination events, we characterized seven different clusters of TS/CYP gene pairs (Figure 3B). Notably, the TPS-a/CYP99 combination was originally identified in monocots (Boutanaev et al., 2015). Our analysis revealed that both genes in this pair are co-expressed in the roots of *C. nankingense* (Figure 3C), suggesting that this gene pair also has a functional role in eudicots.

In summary, our study provides critical information that adds to the limited genomic resources of chrysanthemums, an important evolutionary lineage with remarkable ornamental and medicinal applications worldwide. Given the frequent hybridization and polyploidization of chrysanthemums, our analysis of a potential progenitor genome will further contribute to the understanding of the complex reticulate history of chrysanthemum evolution, including the unresolved origins and domestication of numerous cultivated chrysanthemums.

METHODS

Plant Materials, Library Construction, and Sequencing

The wild *C. nankingense* plant used for genomic sequencing was collected from Nanjing, Jiangsu Province, China, the main geographical area of this species. The *C.* × *morifolium* "Hangbaiju" cultivar was sampled from Tongxiang, Zhejiang Province, China, the main area where this cultivar is planted. Total genomic DNA was isolated from fresh leaves of *C. nankingense* using the conventional cetyltriethylammonium bromide method (Doyle and Doyle, 1987). Total RNA was extracted from roots, stems, leaves, and flowers from both samples using the HiPure Plant RNA Kit according to the manufacturer's instructions (Magen, Guangzhou, China).

A total of 10 μg of high molecular weight DNA was used for Oxford Nanopore library preparation (Supplemental Note 2). The purified library was loaded onto primed R9.4 Spot-On Flow cells (Oxford Nanopore Technologies) for sequencing on a GridION X five sequencer with a running time of 48 h. Base calling analysis of unprocessed data was performed using the Oxford Nanopore Albacore software (v2.1.3). For genome polishing during assembly, we also generated ~100× Illumina short reads on the HiSeq2000 platform (Illumina, San Diego, CA, USA) with insert sizes of 180 bp and 300 bp (Supplemental Note 2).

We selected the PacBio Sequel platform (Pacific Biosciences, Menlo Park, CA, USA) for full-length RNA sequencing of *C.* × *morifolium* "Hangbaiju" (Supplemental Note 2). A total of 3 μg of RNA per sample was used for library preparation with insert sizes of 0.5–4 kb and >4 kb. RNA-seq analysis was conducted using the Sequel platform according to standard protocols (Supplemental Note 2). Further Illumina RNA-seq reads for mixed tissues of *C. nankingense* were obtained for gene prediction analysis.

Genome Assembly and Quality Assessment

Before assembly, we estimated the genome size of *C. nankingense* by performing *k*-mer (k = 19) analysis, resulting in an estimated size of ~3.07 Gb. Another estimate was obtained using flow cytometry, which indicated a haplotype genome size of ~3.24 Gb (Supplemental Note 3).

Nanopore long reads flagged as "passing" were corrected using Canu (v1.6) (Koren et al., 2017) and then used as input for SMARTdenovo (https://github.com/ruanjue/smartdenovo) assembly. The parameters for assembly were: -p jvh -k 17 -J 500 -t 32 -c 1. After finishing the initial assembly, iterative polishing was conducted using Pilon (v1.22) (Walker et al., 2014) in which adapter-trimmed paired-end Illumina reads were aligned with the raw assembly. The Pilon program was run with default parameters to fix bases, fill gaps, and correct local misassemblies.

We aligned the Illumina short reads to the assembled genome using BWA, resulting in a mapping rate of 97.3%. Transcripts assembled from Illumina RNA-seq reads were then included in the alignment; 23 736 of 26 422 long transcripts (>1000 bp) were mapped with >50% sequence coverage. We finally performed both CEGMA and BUSCO (v2.0) assessments on the assembly (Supplemental Note 4).

Structural and Functional Annotation of Genes

Putative protein-coding genes in the *C. nankingense* genome were predicted using the Maker package (v2.31.8) with protein references from published plant genomes and the transcriptome data generated in this study. We also included *de novo* predictions of gene structures obtained using Augustus software (v3.0.3) (Stanke et al., 2006). The rRNAs were predicted using RNAmmer (v1.2) (Lagesen et al., 2007), tRNAs were predicted using tRNAscan-SE (v1.23) (Lowe and Eddy, 1997), and other ncRNA sequences were identified using the Perl program Rfam_scan.pl (v1.0.4) by inner calling using Infernal (v1.1.1) (Nawrocki and Eddy, 2013).

Functional annotation of the protein-coding genes was carried out by performing BlastP (e-value cut-off 1e−05) searches against entries in both the NCBI nr and SwissProt databases. Searches for gene motifs and domains were performed using InterProScan (v5.28) (Jones et al., 2014). The GO terms for genes were obtained from the corresponding InterPro or Pfam entry. Pathway reconstruction was performed using KOBAS (v2.0) (Xie et al., 2011) and the KEGG database.

Annotation of Repetitive Sequences

We identified *de novo* repetitive sequences in the *C. nankingense* genome using RepeatModeler (v1.0.4) (https://github.com/rmhubley/RepeatModeler) based on a self-blast search. We further used RepeatMasker (v4.0.5) (http://www.repeatmasker.org/) to search for known repetitive sequences using a cross-match program with a Repbase-derived RepeatMasker library and the *de novo* repetitive sequences constructed by RepeatModeler. The integration times (t) of intact LTRs were estimated

using the equation $t = K/2r$, where K is the number of nucleotide substitutions per site between each LTR pair and r is the nucleotide substitution rate, which was set to 1×10^{-8} substitutions per site per year (Strasburg and Rieseberg, 2008).

Gene Family and Phylogenomic Analysis

Orthologous gene clusters in the genomes of *C. nankingense* and 14 other representative plants (Supplemental Table 10) were identified using the OrthoMCL program (Li et al., 2003). We determined gene family expansion or contraction using CAFÉ (v3.0) (De Bie et al., 2006). The gene families with >200 members per species were selected based on this analysis.

Alignments from MUSCLE were converted to coding sequences, and RAxML (v8.2.10) (Stamatakis, 2014) was used to construct the phylogenetic trees. The Bayesian Relaxed Molecular Clock method was used to estimate species divergence times using the program MCMCTREE (v4.0) within the PAML package (v4.8) (Yang, 2007). Published times for tomato–potato (<7.4 Mya, >7.2 Mya), sorghum–rice (<55 Mya, >35 Mya), and monocot–dicot (<240 Mya, >130 Mya) divergence were used to calibrate the divergence time. PAML was utilized to calculate the value under evolutionary pressure based on 500 single-copy gene families.

Investigation of Whole-Genome Duplication

We detected and compared WGD events in both the *C. nankingense* genome and three other plant genomes (sunflower, lettuce, and tomato). We identified paralogous gene pairs using Blast-based methods and determined syntenic paralogs using MCScanX (http://chibba.pgml.uga.edu/mcscan2/MCScanX.zip). We also identified orthologs between the four species. We calculated the number of synonymous substitutions per synonymous site (*Ks*) for gene pairs based on the NG method of Yang implemented in the PAML program (v4.8). The synonymous substitution rate of 8.25×10^{-9} mutations per site per year for asterids was applied to calculate the ages of the WGDs (Badouin et al., 2017). Dot-plot analysis of syntenic blocks with at least five gene pairs ($Ks < 0.5$) was conducted using the dot plotter program within the MCScanX package. To investigate the potential WGD of *C.* × *morifolium* "Hangbaiju" in comparison with *C. nankingense* using transcriptome data, we aligned the proteins of both chrysanthemum species and constructed gene families using OrthoMCL. We further filtered families with gene number >10 and used an average linkage hierarchical clustering approach (Badouin et al., 2017) to correct for the redundancy of *Ks* values.

Analysis of Flower Trait-Related and Secondary Metabolite-Associated Genes

We investigated the evolution of genes related to flower trait variation in the *C. nankingense* and *C.* × *morifolium* cultivars. Homologous genes (*CYC*, MADS-box B class genes *DEF*/*AP3* and *GLO*/*PI*, and *CCD4*) in these species were extracted using BlastP (e-value cut-off 1e−50). Redundant sequences and sequences without conserved motifs were removed. MAFFT (v7.397) (Katoh et al., 2002) was used for multiple sequence alignments (–maxiterate 1000 –localpair), and RAxML (v8.2.10) was used for tree building with bootstrapping set to 500.

We identified genes potentially involved in the biosynthesis of flavonoids (ko00940, ko00941, ko00942, ko00944) in chrysanthemums and other representative plants using the KAAS web service in KEGG (http://www.genome.jp/tools/kaas/) (Supplemental Note 6 and Supplemental Table 10).

We identified both TS and CYP genes in the *C. nankingense* genome as previously described (Boutanaev et al., 2015). In brief, to identify TS genes we retrieved and aligned protein sequences of related accessions using BlastP (e-value cut-off 1e−05). The aligned sequences were searched for conserved domains (TPS: PF01397 and PF03936; TCC: PF13243 and PF13249; SQS: PF00494) using HMMER (v3.1b2) (http://hmmer.org/). The predicted TS candidates were further grouped into subfamilies (Chen et al., 2011). To identify CYP genes, we used the CYPs from rice and *Arabidopsis* (http://drnelson.uthsc.edu/P450seqs.dbs.html) as queries to search for homologs and conserved domains (PF00067). We grouped CYPs into clans and families as previously described (Nelson and Werck-Reichhart, 2011). Due to the large size of the *C. nankingense* genome, we defined groups of TS/CYP genes as a cluster if fewer than three other genes were inserted between two tandem TS/CYP genes based on their genomic coordinates.

ACCESSION NUMBERS

We have developed a *Chrysanthemum* genome database at http://www.amwayabrc.com and all of the data generated in this project, including those related to genome assembly, gene prediction, gene functional annotations, and transcriptome sequencing, can be downloaded from our database by visiting http://www.amwayabrc.com.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This study was supported by the National Key Research and Development Program of China (2016YFD0801102) and the National Natural Science Foundation of China (31870198 and 31400278).

AUTHOR CONTRIBUTIONS

S.C., F.C., G.D., T.L., and C.S. conceived the study. G.D., H.Z., A.S., Y.X., S.R., J.J., B.D., S.C., and Z.H. collected samples and performed experiments. Y.W. and C.S. assembled the genome. Genome annotation and evolutionary analyses were completed by Y.W., Y.N., S.W., W.S., and Y.L. Y.L. and W.S. wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

No conflict of interest declared.

Received: July 25, 2018
Revised: September 25, 2018
Accepted: October 10, 2018
Published: October 17, 2018

REFERENCES

Badouin, H., Gouzy, J., Grassa, C.J., Murat, F., Staton, S.E., Cottret, L., Lelandais-Brière, C., Owens, G.L., Carrère, S., Mayjonade, B., et al. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature **546**:148–152.

Boutanaev, A.M., Moses, T., Zi, J., Nelson, D.R., Mugford, S.T., Peters, R.J., and Osbourn, A. (2015). Investigation of terpene diversification across multiple sequenced plant genomes. Proc. Natl. Acad. Sci. U S A **112**:E81–E88.

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

Claros, M.G., Bautista, R., Guerrero-Fernández, D., Benzerki, H., Seoane, P., and Fernández-Pozo, N. (2012). Why assembling plant genome sequences is so challenging. Biology **1**:439–459.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFÉ: a computational tool for the study of gene family evolution. Bioinformatics **22**:1269–1271.

de Lannoy, C., de Ridder, D., and Risse, J. (2017). The long reads ahead:de novogenome assembly using the MinION. F1000Res **6**:1083.

Please cite this article in press as: Song et al., The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits, Molecular Plant (2018), https://doi.org/10.1016/j.molp.2018.10.003

Doyle, J.J., and Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. **19**:11–15.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. Trends Plant Sci. **15**:573–581.

Garcês, H.M.P., Spencer, V.M.R., and Kim, M. (2016). Control of floret symmetry by *RAY3*, *SvDIV1B*, and *SvRAD* in the capitulum of *Senecio vulgaris*. Plant Physiol. **171**:2055–2068.

Goossens, J., Mertens, J., and Goossens, A. (2017). Role and functioning of bHLH transcription factors in jasmonate signalling. J. Exp. Bot. **68**:1333–1347.

Huang, C.-H., Zhang, C., Liu, M., Hu, Y., Gao, T., Qi, J., and Ma, H. (2016). Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. Mol. Biol. Evol. **33**:2820–2835.

Huang, W., Sun, W., Lv, H., Luo, M., Zeng, S., Pattanaik, S., Yuan, L., and Wang, Y. (2013). A R2R3-MYB transcription factor from *Epimedium sagittatum* regulates the flavonoid biosynthetic pathway. PLoS One **8**:e70778.

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. **36**:338–345.

Jiao, W.-B., and Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. Curr. Opin. Plant Biol. **36**:64–70.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics **30**:1236–1240.

Juntheikki-Palovaara, I., Tähtiharju, S., Lan, T., Broholm, S.K., Rijpkema, A.S., Ruonala, R., Kale, L., Albert, V.A., Teeri, T.H., and Elomaa, P. (2014). Functional diversification of duplicated CYC2 clade genes in regulation of inflorescence development in *Gerbera hybrida* (Asteraceae). Plant J. **79**:783–796.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. **24**:1384–1395.

Katoh, K., Misawa, K., Kuma, K.-I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30**:3059–3066.

Kim, H.J., and Lee, Y.S. (2005). Identification of new dicaffeoylquinic acids from *Chrysanthemum morifolium* and their antioxidant activities. Planta Med. **71**:871–876.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. **27**:722–736.

Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.

Leggett, R.M., and Clark, M.D. (2017). A world of opportunities with nanopore sequencing. J. Exp. Bot. **68**:5419–5429.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**:2178–2189.

Lin, L.-Z., and Harnly, J.M. (2010). Identification of the phenolic components of chrysanthemum flower (*Chrysanthemum morifolium* Ramat). Food Chem. **120**:319–326.

Ling, H.-Q., Ma, B., Shi, X., Liu, H., Dong, L., Sun, H., Cao, Y., Gao, Q., Zheng, S., Li, Y., et al. (2018). Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. Nature **557**:424–428.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.

Ma, Y.-P., Chen, M.-M., Wei, J.-X., Zhao, L., Liu, P.-L., Dai, S.-L., and Wen, J. (2016). Origin of *Chrysanthemum* cultivars—evidence from nuclear low-copy LFY gene sequences. Biochem. Syst. Ecol. **65**:129–136.

Michael, T.P., Jupe, F., Bemm, F., Motley, S.T., Sandoval, J.P., Lanz, C., Loudet, O., Weigel, D., and Ecker, J.R. (2018). High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. Nat. Commun. **9**:541.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., et al. (2014). The genome of *Eucalyptus grandis*. Nature **510**:356–362.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics **29**:2933–2935.

Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. Plant J. **66**:194–211.

Nützmann, H.-W., and Osbourn, A. (2014). Gene clustering in plant specialized metabolism. Curr. Opin. Biotechnol. **26**:91–99.

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. Nature **497**:579–584.

Ohmiya, A., Kishimoto, S., Aida, R., Yoshioka, S., and Sumitomo, K. (2006). Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. Plant Physiol. **142**:1193–1201.

Osbourn, A. (2010). Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. Trends Genet. **26**:449–457.

Panchy, N., Lehti-Shiu, M.D., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. Plant Physiol. **171**:2294–2316.

Ren, L., Sun, J., Chen, S., Gao, J., Dong, B., Liu, Y., Xia, X., Wang, Y., Liao, Y., Teng, N., et al. (2014). A transcriptomic analysis of *Chrysanthemum nankingense* provides insights into the basis of low temperature tolerance. BMC Genomics **15**:844.

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., Xia, L., Froenicke, L., Lavelle, D.O., Truco, M.J., et al. (2017). Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. Nat. Commun. **8**:14953.

Salse, J. (2016). Ancestors of modern plant crops. Curr. Opin. Plant Biol. **30**:134–142.

Sasaki, K., Mitsuda, N., Nashima, K., Kishimoto, K., Katayose, Y., Kanamori, H., and Ohmiya, A. (2017). Generation of expressed sequence tags for discovery of genes responsible for floral traits of *Chrysanthemum morifolium* by next-generation sequencing technology. BMC Genomics **18**:683.

Schmidt, M.H., Vogel, A., Denton, A.K., Istace, B., Wormit, A., van de Geest, H., Bolger, M.E., Alseekh, S., Maß, J., Pfaff, C., et al. (2017). De Novo assembly of a new *Solanum pennellii* accession using Nanopore sequencing. Plant Cell **29**:2336–2348.

Shinoyama, H., Aida, R., Ichikawa, H., Nomura, Y., and Mochizuki, A. (2012). Genetic engineering of chrysanthemum (*Chrysanthemum morifolium*): current progress and perspectives. Plant Biotechnol. **29**:323–337.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312–1313.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. **34**:W435–W439.

Strasburg, J.L., and Rieseberg, L.H. (2008). Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris*—large effective population sizes and rates of long-term gene flow. Evolution **62**:1936–1950.

Sun, H., Zhang, T., Fan, Q., Qi, X., Zhang, F., Fang, W., Jiang, J., Chen, F., and Chen, S. (2015). Identification of floral scent in chrysanthemum cultivars and wild relatives by gas chromatography-mass spectrometry. Molecules **20**:5346–5359.

Tähtiharju, S., Rijpkema, A.S., Vetterli, A., Albert, V.A., Teeri, T.H., and Elomaa, P. (2012). Evolution and diversification of the CYC/TB1 gene family in Asteraceae—a comparative study in Gerbera (Mutisieae) and sunflower (Heliantheae). Mol. Biol. Evol. **29**:1155–1166.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One **9**:e112963.

Wang, G., Dong, C., Sun, Y.-A., Xie, K., and Zheng, H. (2008). Characterization of volatile components in dry chrysanthemum flowers using headspace-liquid-phase microextraction-gas chromatography. J. Chromatogr. Sci. **46**:127–132.

Wang, N., Xu, H., Jiang, S., Zhang, Z., Lu, N., Qiu, H., Qu, C., Wang, Y., Wu, S., and Chen, X. (2017). MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*). Plant J. **90**:276–292.

Wang, S., Hao, L.-J., Zhu, J.-J., Zhang, Q.-W., Wang, Z.-M., Zhang, X., and Song, X.-M. (2014). Study on the effects of sulfur fumigation on chemical constituents and antioxidant activity of *Chrysanthemum morifolium* cv. Hang-ju. Phytomedicine **21**:773–779.

Woo, K.-S., Yu, J.-S., Hwang, I.-G., Lee, Y.-R., Lee, C.-H., Yoon, H.-S., Lee, J.-S., and Jeong, H.-S. (2008). Antioxidative activity of volatile compounds in flower of *Chrysanthemum indicum*, *C. morifolium*, and *C. zawadskii*. J. Korean Soc. Food Sci. Nutr. **37**:805–809.

Xia, E.-H., Zhang, H.B., Sheng, J., Li, K., Zhang, Q.J., Kim, C., Zhang, Y., Liu, Y., Zhu, T., Li, W., et al. (2017). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. Mol. Plant **10**:866–877.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C.-Y., and Wei, L. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. **39**:W316–W322.

Yang, W., Glover, B.J., Rao, G.-Y., and Yang, J. (2006). Molecular evidence for multiple polyploidization and lineage recombination in the *Chrysanthemum indicum* polyploid complex (Asteraceae). New Phytol. **171**:875–886.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586–1591.