

# Multiple Polyploidization Events across Asteraceae with Two Nested Events in the Early History Revealed by Nuclear Phylogenomics

Chien-Hsun Huang,<sup>1</sup> Caifei Zhang,<sup>1</sup> Mian Liu,<sup>1</sup> Yi Hu,<sup>2</sup> Tiangang Gao,<sup>3</sup> Ji Qi,<sup>\*,1</sup> and Hong Ma<sup>\*,1</sup>

<sup>1</sup>State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development, Ministry of Education Key Laboratory of Biodiversity Sciences and Ecological Engineering, Institute of Plant Biology, Institute of Biodiversity Sciences, Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai, China

<sup>2</sup>Department of Biology, Huck Institutes of the Life Sciences, Pennsylvania State University, State College, PA

<sup>3</sup>State Key Laboratory of Evolutionary and Systematic Botany, Institute of Botany, the Chinese Academy of Sciences, Beijing, China

\*Corresponding authors: E-mails: hongma@fudan.edu.cn; qij@fudan.edu.cn.

Associate editor: Hideki Innan

## Abstract

Biodiversity results from multiple evolutionary mechanisms, including genetic variation and natural selection. Whole-genome duplications (WGDs), or polyploidizations, provide opportunities for large-scale genetic modifications. Many evolutionarily successful lineages, including angiosperms and vertebrates, are ancient polyploids, suggesting that WGDs are a driving force in evolution. However, this hypothesis is challenged by the observed lower speciation and higher extinction rates of recently formed polyploids than diploids. Asteraceae includes about 10% of angiosperm species, is thus undoubtedly one of the most successful lineages and paleopolyploidization was suggested early in this family using a small number of datasets. Here, we used genes from 64 new transcriptome datasets and others to reconstruct a robust Asteraceae phylogeny, covering 73 species from 18 tribes in six subfamilies. We estimated their divergence times and further identified multiple potential ancient WGDs within several tribes and shared by the Heliantheae alliance, core Asteraceae (Asteroideae–Mutisioideae), and also with the sister family Calyceraceae. For two of the WGD events, there were subsequent great increases in biodiversity; the older one preceded the divergence of at least 10 subfamilies within 10 My, with great variation in morphology and physiology, whereas the other was followed by extremely high species richness in the Heliantheae alliance clade. Our results provide different evidence for several WGDs in Asteraceae and reveal distinct association among WGD events, dramatic changes in environment and species radiations, providing a possible scenario for polyploids to overcome the disadvantages of WGDs and to evolve into lineages with high biodiversity.

**Key words:** Asteraceae, divergence time estimation, orthologous nuclear gene, phylogeny, transcriptome, whole-genome duplication.

## Introduction

The evolution of new life forms and the increase in biodiversity require initial genetic variation within populations and subsequent natural selection (Stebbins 1999) or fixation (Kimura 1983; Hughes 2007). Genetic variations are generated by mutation and recombination; if the variations are beneficial, then they are more likely to be passed on to new generations. Systematic comparisons of genes, proteins, and their interactions between different organisms are beginning to yield insights into the molecular basis of diversity (Sharma et al. 2011; Zhang et al. 2013). The rapid advances of genomics also greatly facilitate the understanding about the relatedness and differences between organisms at a genome-wide level (Paterson et al. 2010).

Whole-genome duplications (WGDs), or polyploidizations, duplicate all genes simultaneously and provide abundant genetic materials for evolutionary processes such as

chromosomal rearrangement (Pontes et al. 2004; Madlung et al. 2005), neofunctionalization (the acquisition of a new function in a duplicate copy) (Blanc and Wolfe 2004a), subfunctionalization (the division of the original function in the two duplicates) (Cusack and Wolfe 2007), and gene conservation owing to dosage effects (the increased production of a beneficial gene product) (Freeling 2009; Bekaert et al. 2011; Hudson et al. 2011). These processes contribute to genomic novelty, speciation, adaptive radiation, and organismal complexity (Stebbins 1940; Levin 1983; Soltis et al. 2003; Rieseberg and Willis 2007; Maere and Van de Peer 2010; Mayrose et al. 2011; Arrigo and Barker 2012). Such changes likely allow organisms to consequently take advantage of new ecological opportunities or to cope with new environmental challenges (Ohno 1970; Hahn 2009; Maere and Van de Peer 2010; Schranz et al. 2012; Fawcett et al. 2013). For example, the angiosperm family Brassicaceae

© The Author 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

experienced rapid and nested radiations of five major lineages based on a recently resolved phylogeny (Huang et al. 2015) after WGD at the crown node of this family (Henry et al. 2006; Schranz and Mitchell-Olds 2006; Edger et al. 2015).

However, recently formed polyploids appear to be an evolutionary dead end. Observations supporting this idea include increased abnormalities through improper meiotic chromosome pairing between two subgenomes of a recent polyploid, resulting in genomic instability that detrimentally affects fertility and fitness (Madlung et al. 2005). Even when the polyploids are stable, suitable mating partners are often lacking, leading to the failure in establishing a viable population for propagation (Levin 1975; Vanneste et al. 2014). The genomic instability and the minority cytotype may be responsible for the lower speciation rates and higher extinction rates of polyploids compared with diploids (Mayrose et al. 2011).

Nevertheless, ancient WGDs have been detected in many evolutionary lineages associated with major species radiations, including vertebrates, fish, and flowering plants (e.g., Levin 1983; McLysaght et al. 2002; Soltis et al. 2003; Dehal and Boore 2005; Rieseberg and Willis 2007; Mayrose et al. 2011). In particular, Jiao et al. (2011) found that the ancestor of all extant angiosperms likely experienced an ancient WGD event, as did the ancestor of the extant seed plants. WGD events have also been found in many specific plant lineages by analyses of sequenced genomes of *Arabidopsis* (*Arabidopsis thaliana*), soybean (*Glycine max*), poplar (*Populus trichocarpa*), and several grasses including rice (*Oryza sativa*), sorghum (*Sorghum bicolor*), *Brachypodium distachyon*, and maize (*Zea mays*) (Blanc et al. 2000; Paterson et al. 2000; Vision et al. 2000; Simillion et al. 2002; Bowers et al. 2003; Paterson et al. 2004; Maere et al. 2005; Tuskan et al. 2006; Schmutz et al. 2010; Tang et al. 2010). The fact that two of the most successful groups of life forms, angiosperms, and vertebrates, are paleopolyploids (Putnam et al. 2008; Jiao et al. 2011) argues strongly for the long-term positive effects of polyploidizations. Considering the aforementioned contrasting fates of polyploids and the resulting controversies (Soltis and Burleigh 2009; Soltis et al. 2009; Abbasi 2010; Van de Peer et al. 2010), the effect of WGD and the underlying mechanisms in evolution and diversification remains an area of vigorous research.

The sunflower family (Asteraceae or Compositae) is one of the largest angiosperm families, with 24,000–35,000 species, accounting for ~10% of all angiosperm species, with diverse habitats, floral structures, and biochemistry (Funk et al. 2009b). Species of Asteraceae range from tiny annual herbs to large trees and have predominated for millions of years in numerous biomes worldwide, primarily in open habitats with seasonal climates such as the Mediterranean region, prairies, steppes, mountains, and deserts (Funk et al. 2009b; Mandel et al. 2015). Asteraceae species are generally known for their attractive inflorescences, the capitula or flower heads, which promote outcrossing, pollination specialization and fitness and are thought to contribute to the evolutionary success and rapid tribal radiation of this large family (Marshall and Abbott 1984; Stuessy et al. 1986; Sun and Ganders 1990;

Endress 1999; Cubas 2004; Sargent 2004; Andersson 2008). Several other characters were also proposed to possibly promote the adaptive radiation of Asteraceae, such as the pappus that assists seed dispersal and secondary metabolites providing defense against herbivores (Stuessy and Garver 1996; Jeffrey 2007), but there is little evidence for a link between these innovative traits and species radiation.

A paleopolyploidization event early in Asteraceae history was proposed to have affected the genes involved in the innovation of the capitulum (Chapman et al. 2008; Chapman et al. 2012). Barker et al. (2008) analyzed distributions of Ks values for paralogs of 18 species in four tribes, one in each of four subfamilies (Asteroideae, Cichorioideae, Carduoideae, and Mutisioideae) and found that Ks peaks at ~0.75 in these species after correction for rate heterogeneity, suggesting that a WGD event might be shared by these four subfamilies. Another putative event was reported for the clade encompassing Heliantheae s. l. and Eupatorieae by comparisons of phylogeny with chromosome numbers (Smith 1975; Robinson 1981; Yahara et al. 1989; Berry et al. 1995; Gentzittel et al. 1995; Baldwin et al. 2002). Barker et al. (2008) also analyzed Ks distributions using EST datasets of several *Helianthus* species and detected evidence for a WGD event with corrected Ks of ~0.37 shared by members of this genus. A putative event also observed in the Ks distribution of *Gerbera hybrida* with Ks of ~0.56 (Barker et al. 2008). Thus, ancient duplication events have been reported for Asteraceae, but the relatively small number of taxon samples did not allow precise positioning of those events. In addition, it is not clear whether these events are related to the evolutionary success of Asteraceae in species diversity, habitats, and morphology.

Morphological and molecular analyses both indicate Asteraceae to be monophyletic (Panero and Funk 2008; Funk et al. 2009b; Panero et al. 2014). To date, the most comprehensive phylogeny is a meta-tree (Funk and Specht 2007; Funk et al. 2009a) constructed using a base tree of 10 chloroplast loci (Panero and Funk 2008; Baldwin 2009; Funk and Chan 2009; Pelsner and Watson 2009), including ~900 of the 1700 genera in the family. However, genetic information provided by phylogenetic markers used in these previous studies were insufficient to resolve relationships of several areas. Comparing with chloroplast sequences, nuclear genes provide genetic information from both parents and are known to be highly effective in resolving phylogeny (Zhang et al. 2012; Wickett et al. 2014; Zeng et al. 2014; Huang et al. 2015; Yang et al. 2015); however, large datasets for Asteraceae are quite limited. Advances in next-generation sequencing technology has provided a means to produce large datasets, facilitating the phylogenetic reconstruction with abundant nuclear markers and the investigation of possible polyploidy events using large scale information from transcriptomic datasets.

Here, we generated transcriptomes for 62 Asteraceae species and two close relatives as outgroups (Calyceraceae *Nastanthus ventosus* and Menyanthaceae *Menyanthes trifoliata*, members of Asterales), and retrieved public datasets of 11 additional Asteraceae members, resulting in a total of 75

species for the investigation of the phylogeny and polyploidizations of Asteraceae. By carefully screening to avoid possible hidden-paralogs, which are single-copy genes resulting from loss of distinct paralogs in different taxa after duplication, 175 nuclear genes and a subset of 71 genes were selected to reconstruct a family-wide Asteraceae phylogeny. A highly supported phylogeny including 18 tribes of six subfamilies is produced from both concatenation and coalescence methods. We estimated divergence times with extensive outgroups and 12 fossil constraints, and found this family to originate during late Cretaceous with a rapid radiation of at least 10 core-subfamilies within 10 My. We also present evidence for multiple ancient WGD events, including nested duplications in the early history of this family, and several independent events in different lineages. The resulting history of Asteraceae reveals two rounds of WGDs that each coincide with an extinction event in the Earth history and followed by two kinds of biodiversity burst in the family. Altogether, our results reveal repeated WGDs during the evolution of Asteraceae, and also support the hypothesis that the success of this proliferative family might have resulted from the combined effect of WGD and rapidly changing environments as indicated by the mass extinction events. The results here on the history of Asteraceae may shed new light on the evolutionary success of polyploids and mechanisms associated with species diversification.

## Results

### A Highly Supported Family-Wide Asteraceae Phylogeny Using Nuclear Orthologs

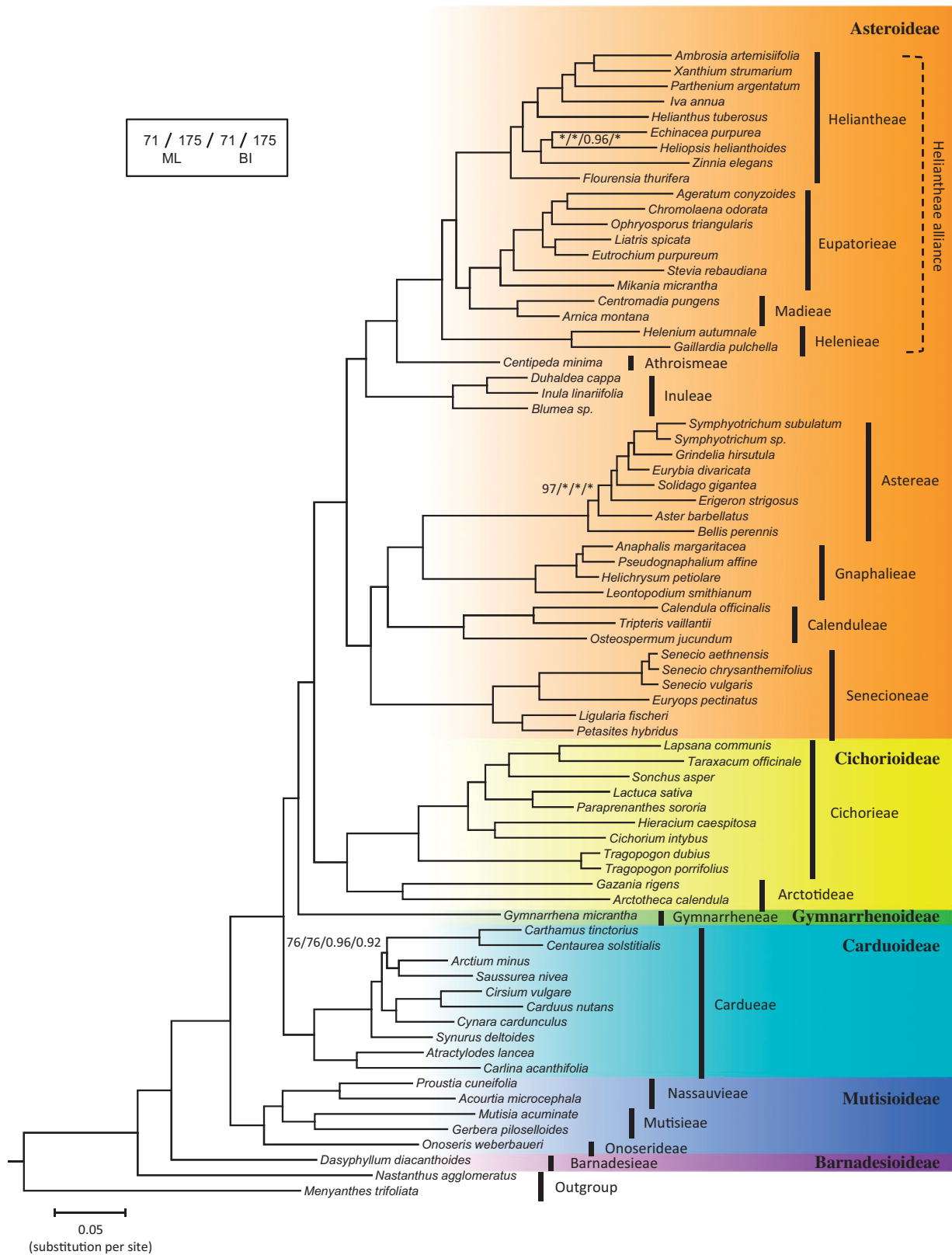
Low-copy nuclear genes have been successfully used to reconstruct phylogenetic relationship of land plants (Wickett et al. 2014), angiosperms (Zhang et al. 2012; Zeng et al. 2014), Brassicaceae (Huang et al. 2015), Caryophyllales (Yang et al. 2015), and many others. Here, we sequenced transcriptomes for 62 Asteraceae species (supplementary table S1, Supplementary Material online) in three largest subfamilies (Asteroideae, Cichorioideae, and Carduoideae), a small subfamily (Gymnarrhenoideae), and two early divergent subfamilies (Mutisioideae and Barnadesioideae), representing 98% of the species richness of the family (Funk et al. 2009b). Also two species *Nastanthus ventosus* (Calyceraceae) and *Menyanthes trifoliata* (Menyanthaceae) were included as closely related outgroups within Asterales. The illumina HiSeq2000 sequencing yielded 13,245,776 to 62,351,587 reads, and the subsequent *de novo* assembly of these datasets resulted in 25,492 to 119,998 contigs with average lengths of 546–1064 nucleotides. Datasets for 11 other Asteraceae species were retrieved from public resources (supplementary table S2, Supplementary Material online). Among the 75 organisms included in this study (supplementary table S2, Supplementary Material online), 18 species representing distinct tribes and an outgroup were adopted for selection of single-copy putative orthologs by using OrthoMCL (Li et al. 2003) and subsequent filtering procedures (see Materials and methods section). A total of 175 orthologous groups (OGs) were identified (supplementary table S3, Supplementary

Material online) for phylogenetic tree reconstructions; among them, 71 OGs having gene length longer than 900 amino acids on an average were also used in separate analyses.

Both maximum-likelihood (ML) and Bayesian inference (BI) analyses using concatenated supermatrices of the two sets of OGs lead to trees with the same topology for all 75 species (fig. 1). Furthermore, coalescence analyses with gene trees from both sets of OGs (supplementary fig. S1, Supplementary Material online) resulted in phylogenetic relationships completely identical to the supermatrix trees. Overall, the topological hypothesis here presents monophyly of each subfamily and tribe and supports their relative positions that were previously defined using morphological characters and/or chloroplast sequences (Panero and Funk 2008; Funk et al. 2009b; Panero et al. 2014). The selected nuclear markers seemed also effective in resolving relationships within tribes; our results largely agree with relationships obtained using chloroplast sequences (Funk et al. 2009a) and also provide resolution for some genera within tribes Astereae, Eupatorieae, and Heliantheae of Asteroideae. Detailed descriptions about the phylogeny and comparison with previous reported topologies are presented in supplementary material, Supplementary Material online. It is noteworthy that our 71 OGs contain sufficient taxonomic signals to infer identical and robust Asteraceae phylogeny and could be considered as phylogenetic marker genes for further studies. Future investigation using nuclear markers with more taxa in various subgroups will aid in reconstructing a more complete phylogeny.

### Estimation of Divergence Time Suggests the Asteraceae Origin in Late Cretaceous

Molecular clock analysis was performed to estimate divergence times of Asteraceae by penalized-likelihood (PL) method implemented in r8s (Sanderson 2003) using concatenation of 71 nuclear markers (fig. 1). We used a total of 12 fossil calibrations in this analysis and included 27 more outgroup species in addition to the two species used in phylogeny reconstruction for implementation of fossils of other plant lineages (supplementary fig. S2 and table S2, Supplementary Material online). To strive for a relatively accurate estimation of divergence times for this large and diverse family, four Asteraceae fossils were used here (supplementary fig. S2, Supplementary Material online). Assignments of fossils (supplementary table S4, Supplementary Material online) were in accordance with Smith et al. (2010) and Magallón et al. (2015) except those of Asteraceae described in supplementary material, Supplementary Material online. A pollen fossil of *Tubulifloridites lilliei* type A (76 – 66 Mya) was recently found in deposits from the late Cretaceous of Antarctica (Barreda et al. 2015, 2016; Panero 2016; Panero and Crozier 2016) (also see supplementary material, Supplementary Material online), but its assignment to specific lineage was controversial. To test effect of different possible assignments, we performed estimations with three different positions of calibration for *T. lilliei* type A (supplementary fig. S3, Supplementary Material online): (1) crown Asteraceae, which is the nearest position applicable in the context of our samplings regarding the original



**Fig. 1.** Asteraceae phylogeny using 175 or 71 low-copy orthologous nuclear genes inferred by ML and BI methods. Values indicate the BS support and posterior probabilities from ML and BI analyses, respectively. Branches without values are with maximum support from all analyses, and asterisks indicate maximum support from the corresponding analysis.

proposed assignment to crown *Dasyphyllum* based on the single-most parsimonious tree (Barreda et al. 2015), (2) at the most recent common ancestor (MRCA) of Asteraceae + Calyceraceae, and (3) at the MRCA of Asteraceae + Calyceraceae + Menyanthaceae. The estimated mean and 95% highest posterior distribution (HPD) ages of several nodes are listed in [supplementary table S5, Supplementary Material online](#).

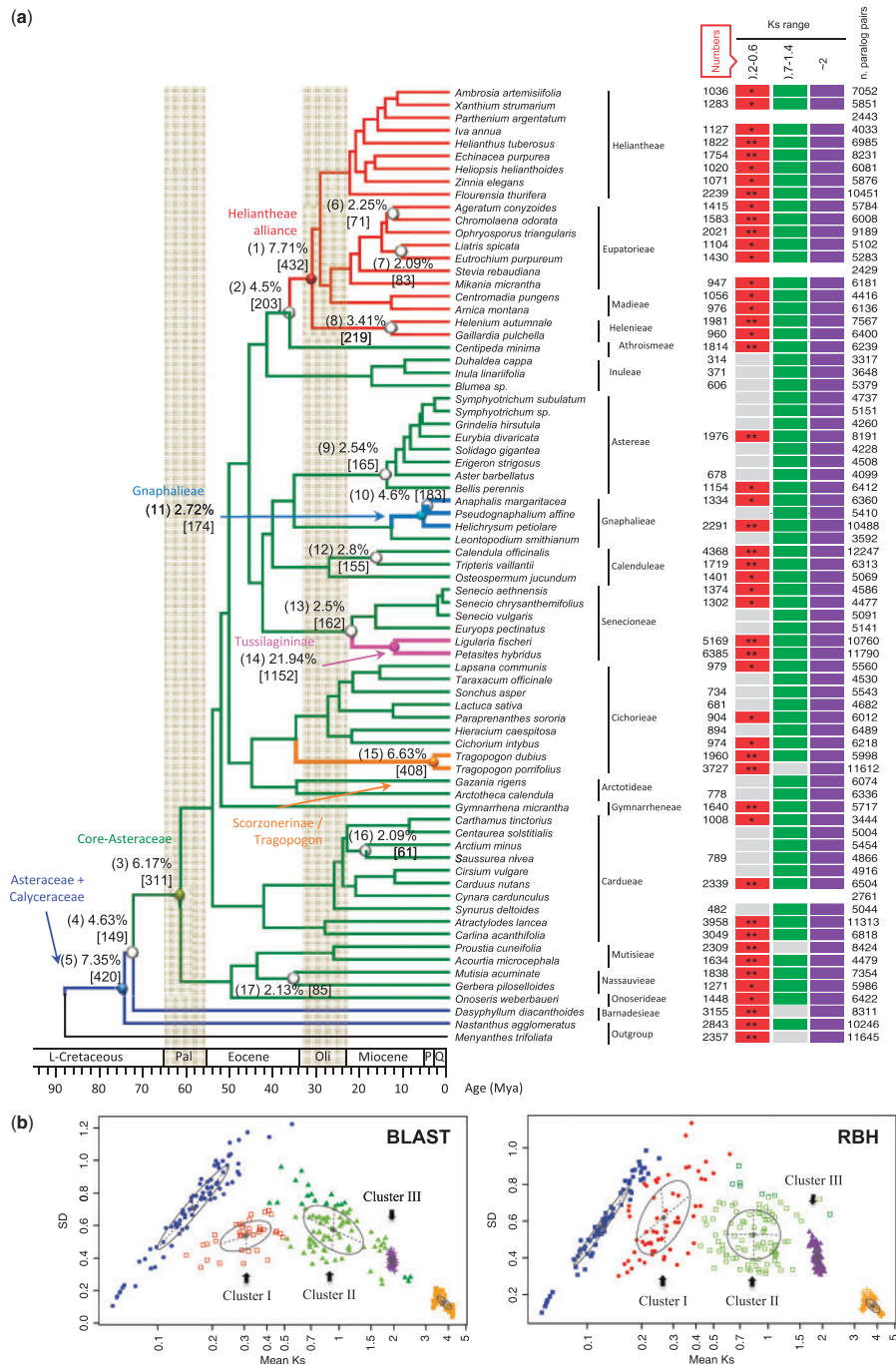
Using the 71 nuclear gene dataset with the above calibrations, our estimated ages showed ~20 Mya for crown Asteraceae using all three calibration sets ([supplementary fig. S3, Supplementary Material online](#)); in particular, the age of ancestor of Asteraceae + Calyceraceae + Menyanthaceae using calibration 3 was much older than the age constraint of *T. lilliei* type A at the same node, thus *T. lilliei* type A calibration at this position had no practical influence in the estimation. Results from calibrations 1 and 2 were very similar to each other with only 2–6 My differences. Our ages of crown Asteraceae and the MRCA of Asteraceae + Calyceraceae using *T. lilliei* type A with calibration 1 were 10–14 My younger than scenario 1 of Barreda et al. (2015) and this is possibly due to the different position for the *T. lilliei* type A fossil in the context of our incomplete sampling, whereas the difference in the age of core Asteraceae (from Mutisioideae to Asteroideae, but not Barnadesioideae) is smaller (~6 Mya). Therefore, our current results should be viewed as tentative and the ages of Asteraceae deep nodes could be older if more *Dasyphyllum* or Barnadesioideae species are sampled. Our ages using *T. lilliei* type A with calibration 2 are very close to scenario 2 of Barreda et al. (2015) when the position of *T. lilliei* type A is the same in the two analyses. Interestingly, the ages reported by a recently published paper (Panero and Crozier 2016) using chloroplast sequences without implementing *T. lilliei* type A are very close to our results using this fossil, showing only 7 My (calibration 1) and 1 My (calibration 2) differences in the age of crown Asteraceae. These age estimates all support that Asteraceae very likely originated during late Cretaceous, older than earlier estimates (Bremer et al. 2004; Kim et al. 2005; Barres et al. 2013; Beaulieu et al. 2013; Magallón et al. 2015; Tank et al. 2015).

Our results showed that branches leading to crown nodes of Asteraceae, subfamilies, even most of the tribes included here were nested during Paleocene to Eocene. It was the hottest and most humid period in Cenozoic Era including Early Eocene Climate Optimum (EECO) and Paleocene–Eocene Thermal Maximum (PETM) (Zachos et al. 2008). Especially, most subfamilies of core Asteraceae (i.e., 10 out of 13 subfamilies, except Barbadesioideae and Famatinanthoideae) diverged within ~10 My (61–50 and 57–47 Mya for calibrations 1 and 2, respectively; [supplementary table S5, Supplementary Material online](#)). Most tribes included here diverged during Eocene, including Heliantheae alliance separated from their sister clade in very late Eocene, followed by subsequent divergences of these tribes during Oligocene.

### Ks Distributions of Paralogs Suggest Multiple Polyploidizations in Asteraceae History

The large number of simultaneously generated duplicates (paralogs) from WGDs tends to exhibit a cluster of Ks values, providing a means to detect potential WGD (Blanc and Wolfe 2004b; Kagale et al. 2014; Vanneste et al. 2014; Cannon et al. 2015). It was previously proposed by Barker et al. (2008) that Asteraceae experienced ancient WGDs from the Ks analysis of paralogs from 18 species representing four tribes, one of each of four subfamilies (the three largest plus Mutisioideae). To detect further evidence for ancient polyploidy in Asteraceae, we identified paralog pairs by using BLAST to identify the reciprocal best hits (RBH) for each species (see Materials and methods section) and monitored their Ks distributions ([supplementary fig. S4, Supplementary Material online](#)). We followed Kagale et al. (2014) to use log-transformed Ks values in the analyses for better resolution on the distribution of low Ks values. Furthermore, we performed mixture model analyses to identify the potential WGD events by assuming that each Ks peak represents a large-scale duplication event following a Gaussian distribution (Blanc and Wolfe 2004b; Schlueter et al. 2004). Datasets with the number of total RBH gene pairs fewer than 3000 were not included in analyses below ([fig. 2a](#), white columns) to avoid possible biases due to the large differences between the number of paralog pairs in these species and those of others. Overall, the results provide evidence for WGDs in the history of most of the species examined here ([fig. 2](#) and [supplementary fig. S4, Supplementary Material online](#)).

We also employed a recently reported clustering analysis using statistical modeling with means and standard deviations of Gaussian components from multiple species (Kagale et al. 2014) as an exploratory tool for potential ancient events and their Ks ranges. Peaks of Ks > 3 might have resulted from WGDs well before the diversification of Asteraceae with possible saturation effect and are not further discussed in this study. Our result included three major clusters ([fig. 2b](#)): Cluster I (Ks 0.2–0.6), Cluster II (Ks 0.7–1.4), and Cluster III (Ks ~2). We mapped peaks of each species onto our phylogeny to ascertain whether members of the same clade have Ks peaks within the same range of values ([fig. 2a](#)). For recent duplications with Ks of 0.2–0.6 (red blocks), we denoted with asterisks when the numbers of paralogs were > 900 (one asterisk) or 1500 (two asterisks). Specifically, all species of Eupatorieae, Madieae, Heliantheae, and Helenieae showed peaks in this Ks range, suggesting the possibility of a shared WGD in their common ancestor. We also observed obvious presence of peaks for many of the members of subfamily Mutisioideae, tribe Calenduleae, and a subtribe of Senecioneae (i.e., Tussilaginae, represented here by *Ligularia fischeri* and *Petasites hybridus*) ([fig. 2a](#)); these peaks were all with large numbers of paralogs, providing evidence for large-scale parallel/independent duplications in these lineages. Notably, Barker et al. (2008) found a substantial amount of rate heterogeneity across the Asteraceae with a nearly 30% difference in background molecular evolutionary rate between the fastest (Heliantheae) and slowest



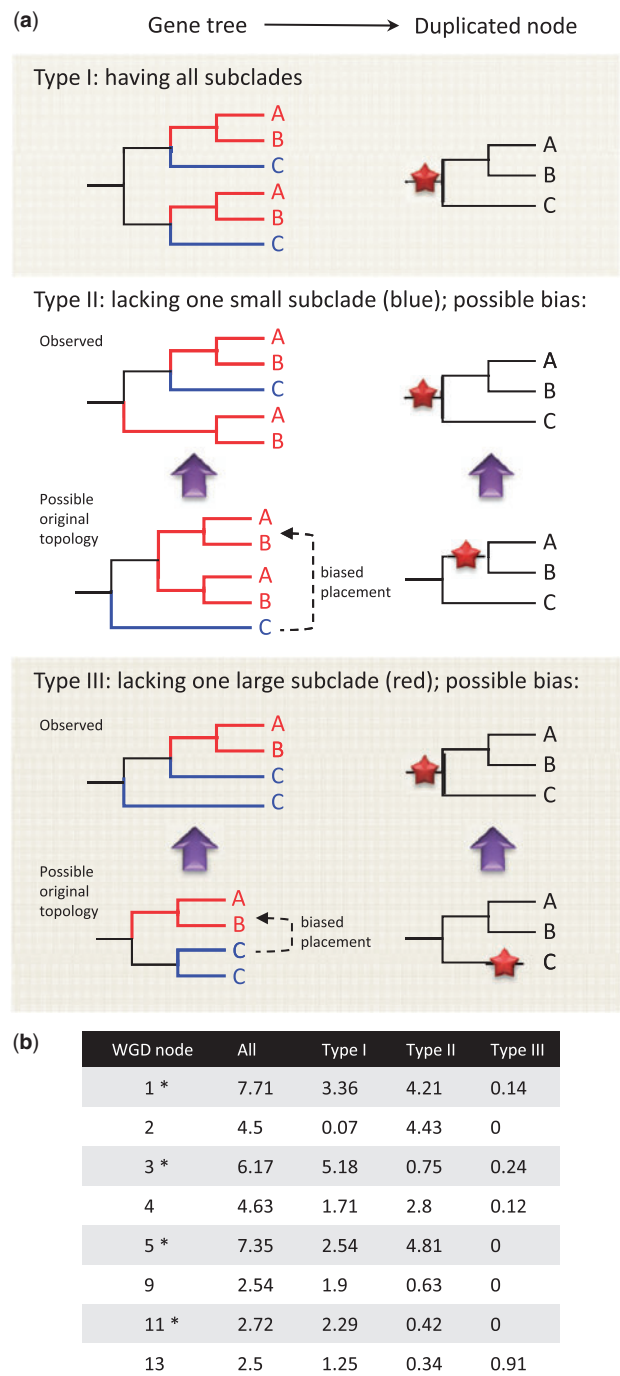
**Fig. 2.** Identifying and mapping gene duplications by Ks distribution and phylogenomic analyses. (a) *Right:* the presence of peaks was determined according to Ks distributions from RBH gene pairs. Blocks colored with red, green or purple indicate the presence of peak in the corresponding Ks range. For those with Ks of 0.2 – 0.6, the peaks mark with one or two asterisks are considered as representative or large duplications when the number of paralogs is higher than 900 or 1500, respectively. For those with more than one component in this Ks range, we take the counts of the ones most prominent or corresponding to a peak in the contour. Gray blocks denote the absence of a recognized peak within the range. Numbers of total paralog pairs are provided for each species (the rightmost column); datasets with numbers fewer than 3000 were not used here. *Left:* WGDs inferred by phylogenomic analysis. Values are percentage of total duplicated gene families for the adjacent nodes; only those > 2% are shown with the actual numbers of duplicated gene trees presented in parenthesis. Gene duplications were recorded only when the two sub-branches shared two or more taxa with BP values > 50 for the node. Results with BP > 70 share the same conclusion (supplementary fig. S6, Supplementary Material online). WGD events with stronger evidence are highlighted with colored circles at the node followed by colored branches, and the ambiguous ones are depicted with open circles. Phylogeny here is according to our chronogram in supplementary fig. S2 Supplementary Material online. (b) Scatterplot of parameters from the fitted Gaussian mixture model of Ks distributions of BLAST (left) or RBH (right) gene pairs. For each Gaussian component of the fitted model, the mean of  $\log_2(Ks)$  value was plotted against standard deviation (SD), whereas X-axis labels here are Ks value instead of  $\log_2(Ks)$ . Classification using Mclust package resulted in five clusters in both datasets, while Clusters I, II, and III are in the range of Ks between 0.2 – 2.

(Carduoideae) lineages among their samplings, providing an explanation for the wide-range in mean  $K_s$  of clusters and peaks of species that are suggested to share common ancient duplication event.

All Asteraceae species showed  $K_s$  peaks between  $\sim 0.7 - 1.4$ , except *Tragopogon porrifolius*, *Tragopogon dubius*, *Proustia cuneifolia*, and *Dasyphyllum diacanthoides*, all of which have in their  $K_s$  distributions a very strong peak near by that could possibly mask an adjacent peak. These common peaks supported a potential paleopolyploidization across the family or at least core Asteraceae (Asteroideae–Mutisioideae), consistent with a previous study (Barker et al. 2008), which observed  $K_s$  value of 0.74 lying within ranges of our Cluster II. The third group of  $K_s$  peaks at  $\sim 2$  in Cluster III is shared not only by members of Asteraceae but also by the two outgroup species (fig. 2a and supplementary fig. S4, Supplementary Material online), suggesting an ancient duplication before the split of Asteraceae and these Asterales lineages. In addition to these three clusters, there is another group of  $K_s$  peaks (fig. 2b, blue dots) between  $\sim 0.1$  and 0.3; although some of these might not represent true WGD, others might be evidence for relatively recent lineage-specific WGDs.

### Placement of Polyploidizations on the Asteraceae Phylogeny Using Gene Family Trees

Results from  $K_s$  distributions and clustering analysis suggest that the ancestors of Asteraceae and some major branches might have experienced one or more ancient WGDs. To test this hypothesis and to investigate the timing of such possible WGDs, we performed a phylogenomic analysis as done in several recent studies (Jiao et al. 2011; Jiao et al. 2012; Cannon et al. 2015; Li et al. 2015; Yang et al. 2015) and recognized as a reliable method for investigating the presence and positions of WGDs (Kellogg 2016). We used all the datasets in determining the phylogeny with six additional outgroup species (supplementary table S2, Supplementary Material online). Homologous gene groups (families) were identified using BLASTp and OrthoMCL among these 81 species, and gene families with species coverage less than 85% were removed. The remaining 6,700 gene families were subjected to ML tree reconstructions followed by comparison with the organismal tree (fig. 1) to detect gene duplications with timings relative to lineage divergence (for detail see Materials and methods section). The result is presented as the percentage of duplicated gene families in fig. 2a adjacent to nodes. Nodes potentially having WGD events with  $> 2\%$  duplicated gene families were listed; among them, there were five showing prominent proportions ( $> 5\%$ ) of gene duplications (nodes 1, 3, 5, 14, and 15; see below), which also correspond well to the observations of  $K_s$  peaks shared by clades. We further investigated each of the duplicated gene trees with three possible topologies regarding the presence or absence of the two subclades following a node (fig. 3a); among them, Type I topology provides stronger support than Types II and III for a gene duplication at the node of interest (see supplementary material, Supplementary Material online for detailed descriptions). Note that only clades with more than three species were included in this follow-up investigation.



**Fig. 3.** Distribution of three types of topologies of duplicated homologous trees. (a) Three possible topologies of duplicated homologous trees. For the two sub-clades derived from one node, red clade represents the large sub-clade (containing A and B), whereas blue clade indicates the small sub-clade (C). Resulting duplication nodes based on Types II and III have the possibilities to be biased in misplacement of the member of small sub-clade due to long-branch attraction or others. Detailed descriptions are in supplementary material, Supplementary Material online. (b) Percentages of the three types of homologous trees supporting duplications for the corresponding WGD nodes. Nodes with  $> 2\%$  Type I trees are denoted with asterisks. Note that only clades with more than three species are available for this investigation.

Among the 6.17% and 7.35% trees supporting duplications at nodes 3 and 5, respectively, we found 5.18% and 2.54% trees have the highly supportive Type I topology (fig. 3b), while the 4.63% supporting trees for node 4 only contains 1.71% of Type I. Barker et al. (2008) observed an ancient WGD before the divergence of core Asteraceae. Here, we found signals for ancient WGD events for three adjacent nodes at the base of the family (nodes 3–5), while those of the core-Asteraceae and Asteraceae–Calyceraceae (which is not shared by Goodeniaceae; unpublished data) were relatively strong and constitute “nested duplications”. The signal for crown Asteraceae remains ambiguous and awaits further investigations with more samplings. The detection of a possible additional WGD shared by Asteraceae and its sister family Calyceraceae is intriguing. Truco et al. (2013) found *Lactuca sativa* to be an ancient hexaploid based on the high-density linkage map of its genome as compared with grape, while hexaploids are probably from two successive genome duplications rather than a single triplication event (Harlan and deWet 1975; Ramsey and Schemske 1998). The nested duplications of core-Asteraceae and Asteraceae–Calyceraceae provide another case of successive events.

In addition, WGD at nodes 1, 14, and 15 were supported by relatively high percentage of corresponding trees. There were 7.71% genes trees showing duplications at MRCA of the Heliantheae alliance (node 1) with 3.36% having Type I topology. Previously a WGD was proposed to be shared by members of Heliantheae alliance as supported by cytology, phylogeny and Ks distributions (Smith 1975; Robinson 1981; Yahara et al. 1989; Berry et al. 1995; Gentzmittel et al. 1995; Baldwin et al. 2002; Barker et al. 2008); our results further restricted its timing to that after the divergence of Athroismeae and the Heliantheae alliance. In our result, there were also 4.5% homologs duplicated at the node containing *Centipeda* (node 2), but most of them belonged to Type II (fig. 3b), which had the possibility to be inferred wrongly from the derived node due to possible biased placement of *Centipeda* in a group with one of the duplicated clades. One possible event was at the MRCA of *Ligularia*/*Petasites* (21.94%, node 14), representing a duplication event for ancestor of subtribe Tussilaginatae. It was suggested that the base chromosome number for this subtribe was  $x = 30$  (Robinson et al. 1997), supporting a possible WGD early in its evolution. The other potential event was shared by the two *Tragopogon* species (6.63%, node 15), intimating a possible WGD at the level ranging from subtribe Scorzonerinae to this genus in the context of our samplings, whereas allopolyploidizations have been observed in members of *Tragopogon* (Soltis et al. 2012, 2016).

We also detected WGD signal for a Gnaphalieae subclade except *Leontopodium* (node 11), having 2.72% supporting trees with 2.29% belonging to Type I. Interestingly, the three genera are relatively large in this tribe. However, obvious peaks can be observed on Ks distribution of *Anaphalis margaritacea* and *Helichrysum petiolare*, but not *Pseudognaphalium affine*. For nodes 9 and 13, only low percentage (< 2%) of gene duplications were inferred by Type I gene trees, and 6 and 2 of their members, respectively, do not have peaks at the Ks

range of Cluster I. Other nodes (6, 7, 8, 10, 12, 16, and 17) showing 2–4.6% of gene trees showing duplication had only two terminals in our samplings and did not permit further investigations for types of topology; future investigations with more taxa are beneficial for a robust conclusion of WGDs for these clades. Nevertheless, it is interesting that all species of Calenduleae and Mutisioideae examined here showed strong peaks within Ks ranges of Cluster I, but only fewer than 20 homologs duplicated at ancestors of these crown groups; therefore, WGDs might have occurred for part of these clades (e.g., node 12 for *Calendula* and *Tripteris* and node 17 for *Gerbera* and *Mutisia*), although the evidence is not strong. Nevertheless, these results intimate multiple possible independent WGDs for members of these clades.

### Detection of Two Major Net Diversification Rate Accelerations during Asteraceae History

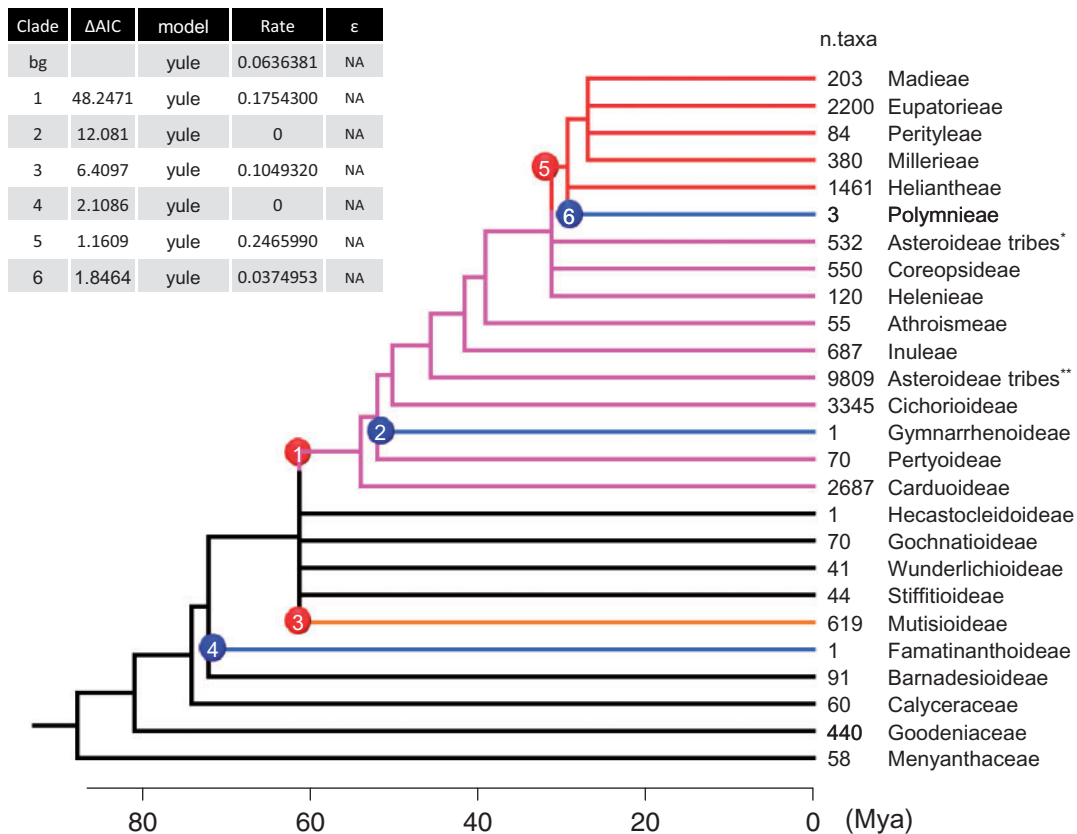
Comparing with the numerous species in the Asteraceae family, our transcriptome data lack sufficient taxon samplings to profile species richness for various subfamilies or tribes. Thus, we took the summarized meta-tree of Funk et al. (2009b) based on several previous studies (Panero and Funk 2008; Baldwin 2009; Funk and Chan 2009; Pelsner and Watson 2009) for a hypothesis on phylogeny with comprehensive inclusion of major lineages and their species richness (supplementary fig. S5, Supplementary Material online), which was then subjected to diversification rate shift analysis using Modeling Evolutionary Diversification Using Stepwise Akaike Information Criterion (AIC) (MEDUSA; Alfaro et al. 2009). Our result showed three positions of acceleration (red circles) and three occasions of slowdown (blue circles) in net diversification rate (fig. 4). The slowdowns happened at branches leading to specific subfamilies and tribes with fewer than three species. There was moderate acceleration in the Mutisioideae lineage. One of the major acceleration events was located at the node leading to Asteroideae–Carduoideae clade with almost 3-fold of the background rate (0.175 vs. 0.064 species/My), and the other was at the clade of tribes Madieae, Eupatorieae, Perityleae, Millerieae, and Heliantheae (i.e., the Heliantheae alliance clade except the basal tribe Helenieae) with 4-fold of the background rate (0.247 species/My). In our analysis, we did not include some of the basal subfamilies and thus the results in rate shift are tentative, but we could identify a potential relationship between WGDs and net diversification rate shifts in accordance with WGD analyses in the context of our samplings, demonstrating two accelerations each after two of the duplication events reported in this study (fig. 5).

### Discussion

The previously proposed Asteraceae phylogeny was based on analyses using chloroplast sequences (Funk et al. 2009a, 2009b). Our phylogenetic analyses using both concatenation and coalescence methods with two sets of orthologous nuclear genes from 64 newly generated transcriptomes and additional public datasets yielded a hypothesis with strong supports for monophyly of all subfamilies and tribes included



The appropriate AICC-threshold for a tree of 26 tips is 0.8278956



**Fig. 4.** Detecting diversification rate shifts by MEDUSA. Number of taxa (n.taxa) and phylogeny of lineages not sampled here are according to Funk et al. (2009b) and is also illustrated as a hypothesized phylogeny in supplementary fig. S5, Supplementary Material online. Partitions with higher diversification rate are colored with orange, red or pink, and those with lower rate are with blue. Positions of rate shifts are depicted with red (acceleration) and blue (slowdown) circles numbered by the order of rate shifts added by the stepwise AIC procedure. Terminals representing the combination of multiple tribes are indicated with \* (Neurolaeneae, Tageteae, Chaenactideae, and Bahieae) and \*\* (Corymbieae, Doronicum, Abrotanella, Senecioneae, Calenduleae, Gnaphalieae, Astereae, and Anthemideae). The improved AIC scores ( $\Delta$ AIC), fitted models, estimated net diversification rates (Rate), and relative extinction rates ( $\epsilon$ ) are shown on the table in the upper left corner. NA, relative extinction rate cannot be estimated because the shifts fitted to the pure-birth (yule) model; bg, background.

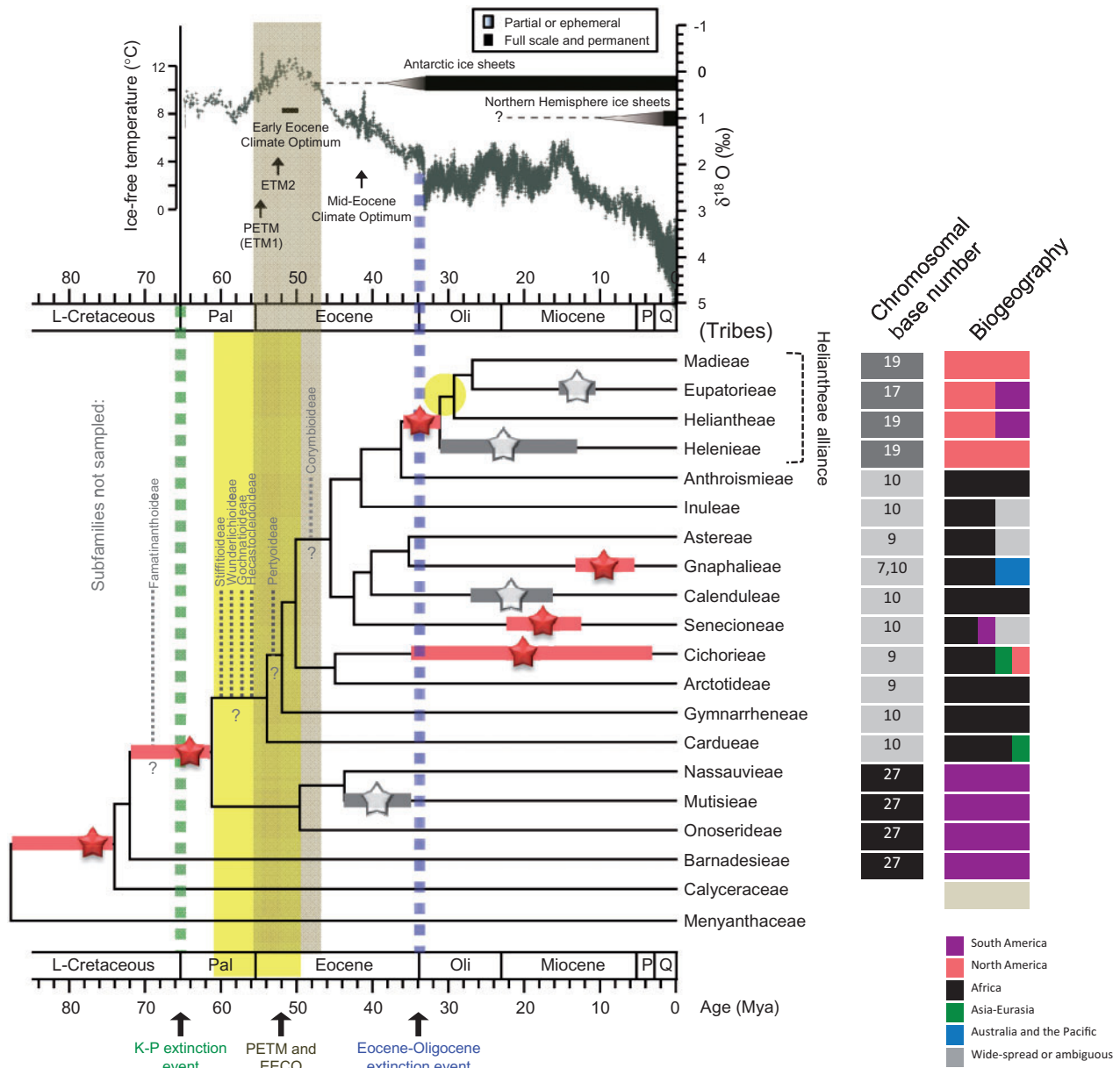
here. The nuclear gene-based phylogenetic relationships among tribes and subfamilies are in complete agreement with the current conclusions according to chloroplast sequences (Panero and Funk 2008; Funk et al. 2009a; Panero et al. 2014), showing an example for the consistency between results using the nuclear and chloroplast genomes. We have also resolved relationships among genera of Asteroideae, including some that were previously ambiguous from analyses using chloroplast sequences, demonstrating the effectiveness of these informative nuclear markers.

In addition to a robust phylogeny, our results on divergence times of Asteraceae at a family-wide scale (supplementary figs. S2 and S3, Supplementary Material online) and the detection of strong evidence for multiple polyploidizations across the family (figs. 2 and 3) allow us to reconstruct the evolutionary history of Asteraceae (see fig. 5) combining the observations on species diversifications (fig. 4; Panero and Crozier 2016) (see below).

Using both Ks distribution and gene tree analyses, we found evidence for multiple WGD events across the family. Our findings support two nested duplications: one at core

Asteraceae and the other before the split of Asteraceae from Calyceraceae. We also observed several additional events: at the crown node of the Heliantheae alliance, at clades Tussilaginae and Tragopogon-Scorzonerinae, and within Gnaphalieae. A subclade within Mutisioideae or Calenduleae might each have an ancient WGD, but other possibilities also existed with members of these clades.

In combination with the chronograms here, two WGDs occurred along the backbone around the time of global mass extinction events, shortly before species radiations with two kinds of evidence (a radiation of 10 subfamilies and high species richness; see below for more details). The earlier duplication event near the basal node of Asteraceae is close to the Cretaceous–Paleogene (KT) mass extinction event (fig. 5 and supplementary fig. S2, Supplementary Material online). Net diversification rate acceleration was detected by analysis using the MEDUSA program subsequent to the mass extinction and WGD events (figs. 4 and 5). This differs from the estimated core shifts in net diversification rate in the Phytomelanin Fruit (PF) and Vernonioid clades by Panero and Crozier (2016), but not near the base of the family, likely



**Fig. 5.** A proposed evolutionary history of Asteraceae. The top part of this figure shows a stacked deep-sea benthic foraminiferal oxygen-isotope curve adapted from fig. 2 in Zachos et al. (2008) corresponding to the evolution of global climate over the last 65 My. Red stars indicate the phylogenetic positions of major putative paleopolyploidizations presented here, pink strips specify the estimated window in which duplications occurred (according to the ages in supplementary fig. S2, Supplementary Material online). Open stars and gray strips are WGD events showing consistency between results from Ks distributions and phylogenomic analyses but with more ambiguous supports from the percentage of duplicated homolog trees. Cases having inconsistent results for the two analyses are not shown in this figure. The positions of species radiations are highlighted as referred to a core-shift in net diversification rate (yellow circle) (fig. 4; Panero and Crozier 2016) and rapid branching of at least 10 subfamilies of core-Asteraceae (yellow strip). Chromosomal base numbers and biogeography distributions (or optimization) are according to Semple and Watanabe (2009) and Funk et al. (2009b). Green and blue dotted lines indicate KT extinction event and Eocene-Oligocene extinction event, respectively. The warm period of PETM and EECO is highlighted with a brown strip. Subfamilies not sampled here are integrated with gray dashed lines according to the hypothesized position in the meta-tree by Funk et al. (2009b). Pal, Paleocene; Oli, Oligocene; P, Pliocene; Q, quaternary; ETM, Eocene thermal maximum.

due to their samplings of more basal lineages. However, dramatic loss of ancient taxa can bias the result when determining diversification dynamics using only extant taxa (Quental and Marshall 2010; Morlon 2014). This is particularly relevant because the early diverging Asteraceae lineages had originated during the period of global catastrophe and evolved under the subsequent stressful environments, which resulted the

extinction of ~67% of the species (Jablonski and Chaloner 1994). On the contrary, at least 10 out of 13 subfamilies diverged very rapidly within 10 My (fig. 5, yellow strip); a similar time frame was also reported in studies using chloroplast sequences (within 13–10 Mya) (Barreda et al. 2015; Panero and Crozier 2016). This radiation of subfamilies is indicative of a burst in biodiversity.

The descendant clades after WGD and KT extinction events can be further divided into two groups showing distinct patterns of evolution. According to the observations of extant species, members of the more proliferative clades (Asteroideae, Cichoroideae, and Carduoideae) have a relatively low chromosomal base numbers of 9–10 in contrast to 27 of the basal clades (e.g., Mutisioideae, and other basal subfamilies not sampled here); furthermore, most of these large clades are distributed in Africa and Eurasia, whereas the members of the basal clades are found in South America, the proposed origin of Asteraceae (Funk et al. 2009a). It is possible that after polyploidization the common ancestor of these large clades might have experienced a large-scale chromosome rearrangement (Inoue et al. 2015), resulting in fewer chromosomes and other changes that facilitated adaptation to new habitats. On the contrary, Mutisioideae retained a high chromosomal base number (27) and remained in South America (Funk et al. 2009b); nevertheless, polyploidization might still have played a role in enhancing the ability of Mutisioideae species to adapt to environmental changes in their original habitat. In comparison, the relatively long branch of Barnadesioideae for the split of its genera after divergence from other subfamilies (Panero and Crozier 2016; our unpublished data) might be the result of extinction of ancient relatives under stressful environment.

Similarly, the WGD event at the base of the Heliantheae alliance was also close to the Eocene–Oligocene boundary (fig. 5), when an extinction event named *Grande Coupure* occurred in Europe (Hooker et al. 2004) under a short period of abrupt cooling icehouse climate (Zachos et al. 2001). The extant species of this WGD clade have largely distributed to North America (in contrast to Africa for the basal sister tribes), and their chromosomal base numbers are about twice of those of their sister clade (17–19 comparing with 7–10; fig. 5), providing additional support for paleopolyploidization at the MRCA of the Heliantheae alliance (Smith 1975; Robinson 1981; Yahara et al. 1989; Berry et al. 1995; Gentzmittel et al. 1995; Baldwin et al. 2002). Following this node is the diversification acceleration contributed by the great species richness (fig. 4), which is also supported by the core shift in net diversification rate of the PF clade by Panero and Crozier (2016). In addition, there is a period of delay between this WGD and the subsequent acceleration in diversification rate, consistent with the idea that significant lag-times often exist between WGDs and species radiations (Schrantz et al. 2012). The extraordinary species richness of this lineage again demonstrates the possible role of the combined effect of WGD (and its subsequent modifications) and dramatic environmental changes on evolution. Notably, the subsequent global grassland expansion along with increasing aridity and seasonal change during Oligocene likely created various open niches for multiple lineages and facilitated diversifications after these events.

A possible correlation between polyploidy and KT extinction event has been proposed previously (Fawcett et al. 2009; Vanneste et al. 2014). The authors analyzed 38 plant genomes and three transcriptomes, with evidence for 20 independent WGDs across 19 observed plant families, and revealed a wave of successful genome duplications that were significantly and

non-randomly clustered between 55 and 75 Mya (Vanneste et al. 2014), showing a strong indication of an association with the KT boundary (66 Mya). In the two analyses, they dated the ages of WGD paralogs with correction for substitution rate heterogeneity among different lineages, using *Lactuca sativa* as the only representative of Asterales; therefore, they could not locate the positions of WGD along the Asteraceae phylogenetic branches. The older ages of Asteraceae from this study as well as Panero and Crozier (2016) suggest that the WGD identified in Fawcett et al. (2009) and Vanneste et al. (2014) very possibly occurred after the divergence of Asteraceae (i.e., within the family), as is well supported by our result (61.3 Mya for node 3, fig. 2a). There was also a broader clustering centered at ~22.9 Mya (Vanneste et al. 2014), also consistent with our results of a WGD at the clade of Heliantheae alliance. Overall, our findings in Asteraceae revealed that at least two independent scenarios occurred within a family (Asteraceae) with very high biodiversity for nearly ten percent of angiosperms, providing support for the hypothesis of major geological events and polyploidy. Vanneste et al. (2014) suggested that environmental stresses around the time of polyploidizations might have promoted the establishment of polyploids and the realization of their evolutionary potential, thus alleviating the rigid contrast in the proposed evolutionary fates of polyploids. Notably, several other WGDs supported by analyses here are not related to a mass extinction event and no obvious species radiations could be detected. Thus, establishment of polyploids are not restricted to the time of stressful environments, but stressful environments could allow polyploids manifest evolutionary potential, perhaps competing better with diploid counterparts and becoming the origins of species radiation.

In summary, the ability of Asteraceae to become one of the most successful plant families might be facilitated by multiple WGD events, particularly those that occurred near geological times of dramatic changes in the Earth environments. The WGD events likely allowed the evolution of a large number of new gene functions; at the same time, the stressful environments after mass extinctions might be strongly selective helping in establishing sustainable populations successfully (Voss et al. 2012; Chao et al. 2013) while also generating many open niches. Following WGD, some descendants of the recently formed polyploids were provided with new niches with different environmental conditions and diversified into very large subfamilies or tribes, whereas other descendants remained in relatively stable habitats.

## Materials and Methods

### Taxon Sampling and Transcriptomics

Information for the taxa included is listed in [supplementary table S2, Supplementary Material](#) online. Total RNA extraction with ZR plant RNA MiniPrep kit and subsequent sequencing by illumina HiSeq2000 and assembling using Trinity (Grabherr et al. 2011) and TGICLv2.1 (Pertea et al. 2003) were performed as described in Huang et al. (2015). Redundant contigs in each transcriptome were removed using CD-HIT 4.6 (Fu et al. 2012) with the threshold of 0.98.

Sequence Read Archive (SRA) data were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov/sra/>; last accessed August 5, 2016).

### Selection of Putative Orthologs as Phylogenetic Markers

We used OrthoMCL v1.4 (Li et al. 2003) to identify the single-copy orthologous sets (SCOS) from 18 Asteraceae species of different tribes of Asteraceae (*Arctotheca calendula*, *Bellis perennis*, *Carduus nutans*, *Centipeda minima*, *Centromadia pungens*, *Dasyphyllum diacanthoides*, *Eutrochium purpureum*, *Gymnarrhena micrantha*, *Helenium autumnale*, *Inula linariifolia*, *Leontopodium smithianum*, *Mutisia acuminata*, *Onoseris weberbaueri*, *Osteospermum jucundum*, *Proustia cuneifolia*, *Senecio vulgaris*, *Taraxacum officinale*, and *Xanthium strumarium*) and one outgroup species *Acicarpha spathulata*. The resulting 271 SCOS among these 19 species were used as seeds in HaMStR (Ebersberger et al. 2009) to find orthologs in other species with  $E$  values less than  $e^{-20}$ . Nucleotide sequences of the 271 OGs were aligned and trimmed using MUSCLE v3.8.31 (Edgar 2004) and trimAl v1.2 (Capella-Gutiérrez et al. 2009) with the default settings, respectively.

Possible sequence biases of long-branch attraction and saturation of partitions were detected by TreSpEx (Struck 2014), in which the long-branch attraction was determined by standard deviation and the average of upper quartile of long-branch scores, and the saturation degree was determined by slopes and  $R^2$  values of linear regressions of patristic distances against uncorrected distances  $p$ , with alignments of the 271 gene sequences. Genes showing high degrees of misleading signals as indicated by the four parameters were pruned, resulting in 175 OGs (supplementary table S3, Supplementary Material online). To avoid the possible effect of hidden-paralogs, persisting single-copy genes resulted from the loss of distinct paralogs in different species, a filtering process was performed by using the ML tree of 175-gene concatenation as a basic topology to investigate each of the 175 single-gene trees and remove the gene of particular species that shows significant differences in groupings within major clades. The filtered set of 175 OGs, and the 71 OGs with alignment length > 900 nucleotides, were both used in subsequent phylogeny reconstructions. Nucleotide sequences used for phylogenetic analyses in this study are available in the TreeBASE website (<http://treebase.org/treebase-web/home.html>) under the accession code S19664.

### Phylogeny Reconstruction

Alignments of the nucleotide sequences of the 175 and 71 OGs were concatenated using SeaView (Gouy et al. 2010). Subsequent phylogenetic analyses were performed with GTRGAMMAI model as suggested by Modeltest (Posada and Crandall 1998) for both 175- and 71-gene sets. Phylogenetic reconstructions using concatenations of 175 and 71 genes by BI and ML methods and coalescence analysis were carried out using MrBayes 3.2.1 (Ronquist and Huelsenbeck 2003), RAxML 7.0.4 (Stamatakis 2006), and

Astral 4.4.4 (Mirarab et al. 2014), respectively, as described before (Huang et al. 2015).

### Fossil Calibrations and Divergence Time Estimation

We used 12 fossil constraints in our divergence time estimation (supplementary table S4, Supplementary Material online) according to the assignments by Smith et al. (2010) and Magallón et al. (2015), except those of the four Asteraceae fossils as described in supplementary material, Supplementary Material online. All the fossil calibrations were implemented as minimum constraint, whereas the tricolpate pollen grains (about 125 Mya; Hickey and Doyle 1977) were used to fix the age of crown group eudicots (Forest 2009; Magallón and Castillo 2009). Divergence times were estimated using r8s v1.7.1 (Sanderson 2002) as previously described (Huang et al. 2015), with the ML tree with branch lengths generated by RAxML using 71 OGs as the input tree. *Oryza sativa* was the most distant outgroup species in our analysis and was pruned as required by r8s. Mean and 95% HPD ages of nodes of interest were estimated and summarized across the 100 BS trees (supplementary table S5, Supplementary Material online).

### Synonymous Substitution Rate of Paralogs and Mixture Model Analysis

Paralogs were determined from gene pairs identified by using BLASTP with an  $E$ -value cutoff of  $e^{-5}$  followed by filtering with > 70% match in alignment length and > 50% sequence identity, resulting in “BLAST” gene pairs; among them, the RBHs within each species were used as “RBH” pairs. We estimated  $K_s$  values between pairs of paralogs according to Blanc and Wolfe (2004b) by calculations based on codon alignments using the ML method implemented in codeml of the PAML package (Yang 2007) under the F3x4 model (Goldman and Yang 1994). To identify significant peaks, Gaussian mixture models were fitted to the distributions of  $\log_2$  transformed  $K_s$  values using the R package Mclust (Fraley and Raftery 2002, 2003), which determined the best-fitted model according to Bayesian Information Criterion (BIC). Histograms of  $\log_2(K_s)$  distribution and the estimated components of the fitted model of each species are provided in supplementary fig. S4, Supplementary Material online. Observations on  $K_s$  distributions found overfitting or skewed peaks owing to the multiple overlapping Gaussian components in specific cases; such complications were resolved as follows to combine the overlapping components as Gaussian mixtures. For a Gaussian mixture distribution with mixing proportions  $\pi_i$  that sum to 1, means  $\mu_i$ , and standard deviations  $\sigma_i$ , the combined mean were determined to be as  $\sum_i \pi_i \mu_i$  and the combined SD as  $\sqrt{\sum_i \pi_i (\sigma_i^2 + \mu_i^2) - (\sum_i \pi_i \mu_i)^2}$  (Kagale et al. 2014). The resulting mean and SD were then plotted as single points in fig. 2b and single component in supplementary fig. S4, Supplementary Material online. A scatterplot was generated with the mean value of the  $\log_2(K_s)$  and SD of each Gaussian component, and their classifications were determined using Mclust with the fitted model selected via BIC.

## Locating Gene Duplications by Comparing Gene Family Trees with Species Tree

Homologs of 81 species (supplementary table S2, Supplementary Material online) were identified by performing an all-against-all BLASTp comparison among the protein sequences of them with *e*-value cutoff of  $10^{-5}$ . In Paranoïd 4.1 (O'Brien et al. 2005) was then applied on each BLAST match to calculate a global protein identity to filter out those with poor similarities (<50%) or gene coverage (<50%). Homologous groups were predicted using OrthoMCL v1.4 (Li et al. 2003) with the inflation value of 2.0. For each group composing of four or more genes, protein sequences were aligned using MUSCLE v3.8.31 (Edgar 2004) with default parameters and were further trimmed by using trimAl v1.4 (Capella-Gutiérrez et al. 2009) with the options “-gt 0.1 -resoverlap 0.75 -seqoverlap 80”. The alignment of corresponding nucleotide sequences was built guided by the trimmed protein alignment. Then, phylogenies for homolog groups with species coverage higher than 85% were reconstructed by RAxML with GTRCAT model for 100 BS replicates.

The resulting gene trees were reconciled with the species trees to calculate the numbers of gene duplication at each node, a method commonly used in many recent studies in detecting the presence and positions of large duplication events (Jiao et al. 2012; Cannon et al. 2015; Li et al. 2015; Yang et al. 2015) and is recognized as one of the more reliable method (Kellogg 2016). In detail, in each gene family tree, a node was recognized as valid for further analyses when the BS support was higher than 50 (fig. 2) or 70 (supplementary fig. S6, Supplementary Material online). Then, the last common ancestor (LCA) was determined for each of the two sub-branches. A gene duplication was then counted when meeting two requirements: (1) the LCAs of the two sub-branches had the same or close depth (the difference is smaller than or equals to one), where the depth of node was defined by the number of steps traveling from the node to the root on the species tree; (2) the two sub-branches shared two or more taxa. Numbers of gene duplications were summarized on the species tree by iterating all single gene family trees.

## Diversification Rate Shift

Diversification rates were investigated by MEDUSA (Alfaro et al. 2009) using ultrametric tree data based on our date phylogeny with the missing taxa incorporated based on Funk et al. (2009b); details in the implemented phylogeny and species richness are in fig. 4 and supplementary fig. S5, Supplementary Material online. For the branches not sampled in this study (whose ages are unknown), because they could diverge from any points along the corresponding main branches, we set these clades to be branched out from their corresponding stem nodes, that is to set their ages the same as their basal sister clades, in order to eliminate their effect on the species richness of the sampled and dated clades here. The age of Goodeniaceae was set to the midpoint of Calyceraceae and Menyanthaceae close to those obtained by Panero and Crozier (2016) and our unpublished data. Corrected AIC criterion and mixed model of birth-death or Yule were implemented in the calculation.

## Supplementary Materials

Supplementary figures S1–S6 and tables S1–S5 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We thank Drs Holly Forbes, Xuejun Ge, Jinshuang Ma, Michael Monosov, Weimin Ni, Jun Wen, Ji Yang, Liping Zeng, Ning Zhang, and Shu Zhang for help with sampling of plant materials. We also thank the two anonymous reviewers for providing valuable comments. This work was supported by grants from the National Natural Science Foundation of China (Grant no. 91131007 to H.M. and 31371330 and 31570224 to J.Q.); the China Postdoctoral Science Foundation (Grant no. 2014M551316); and funds from the State Key Laboratory of Genetic Engineering at Fudan University.

## References

- Abbasi AA. 2010. Piecemeal or big bangs: correlating the vertebrate evolution with proposed models of gene expansion events. *Nat Rev Genet.* 11:166.
- Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc Natl Acad Sci USA.* 106:13410–13414.
- Andersson S. 2008. Pollinator and nonpollinator selection on ray morphology in *Leucanthemum vulgare* (oxeye daisy, Asteraceae). *Am J Bot.* 95:1072–1078.
- Arrigo N, Barker MS. 2012. Rarely successful polyploids and their legacy in plant genomes. *Curr Opin Plant Biol.* 15:140–146.
- Baldwin BG. 2009. Heliantheae alliance. In: Funk VA, Susanna A, Stuessy T, Bayer RJ, editors. Systematics, evolution, and biogeography of Compositae. Vienna: IAPT. p. 689–730.
- Baldwin BG, Wessa BL, Panero JL. 2002. Nuclear rDNA evidence for major lineages of Helioid Heliantheae (Compositae). *Syst Bot.* 27:161–198.
- Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol Biol Evol.* 25:2445–2455.
- Barreda VD, Palazzesi L, Tellería MC, Olivero EB, Raine JL, Forest F. 2015. Early evolution of the angiosperm clade Asteraceae in the Cretaceous of Antarctica. *Proc Natl Acad Sci USA.* 112:10989–10994.
- Barreda VD, Palazzesi L, Tellería MC, Olivero EB, Raine JL, Forest F. 2016. Reply to Panero: robust phylogenetic placement of fossil pollen grains: the case of Asteraceae. *Proc Natl Acad Sci USA.* 113:E412.
- Barres L, Sanmartín I, Anderson CL, Susanna A, Buerki S, Galbany-Casals M, Vilatersana R. 2013. Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). *Am J Bot.* 100:867–882.
- Beaulieu JM, Tank DC, Donoghue MJ. 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. *BMC Evol Biol.* 13:80.
- Bekaert M, Edger PP, Pires JC, Conant GC. 2011. Two-phase resolution of polyploidy in the *Arabidopsis* metabolic network gives rise to relative and absolute dosage constraints. *Plant Cell* 23:1719–1728.
- Berry ST, Leon AJ, Hanfrey CC, Challis P, Burkholz A, Barnes SR, Rufener GK, Lee M, Caligari PD. 1995. Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. *Theor Appl Genet.* 91:195–199.
- Blanc G, Barakat A, Guyot R, Cooke R, Delseny M. 2000. Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* 12:1093–1101.

- Blanc G, Wolfe KH. 2004a. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16:1679–1691.
- Blanc G, Wolfe KH. 2004b. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667–1678.
- Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438.
- Bremer K, Friis EM, Bremer B. 2004. Molecular phylogenetic dating of Asterid flowering plants shows early cretaceous diversification. *Syst Biol*. 53:496–505.
- Cannon SB, McKain MR, Harkness A, Nelson MN, Dash S, Deyholos MK, Peng Y, Joyce B, Stewart CN, Rolf M, et al. 2015. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Mol Biol Evol*. 32:193–210.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Chao DY, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE. 2013. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* 341:658–659.
- Chapman MA, Leebens-Mack JH, Burke JM. 2008. Positive selection and expression divergence following gene duplication in the sunflower CYCLOIDEA gene family. *Mol Biol Evol*. 25:1260–1273.
- Chapman MA, Tang S, Draeger D, Nambesani S, Shaffer H, Barb JG, Knapp SJ, Burke JM. 2012. Genetic analysis of floral symmetry in Van Gogh's sunflowers reveals independent recruitment of CYCLOIDEA genes in the Asteraceae. *PLoS Genet*. 8:e1002628.
- Cubas P. 2004. Floral zygomorphy, the recurring evolution of a successful trait. *Bioessays* 26:1175–1184.
- Cusack BP, Wolfe KH. 2007. When gene marriages don't work out: divorce by subfunctionalization. *Trends Genet*. 23:270–272.
- Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol*. 3:e314.
- Ebersberger I, Strauss S, von Haeseler A. 2009. HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evol Biol*. 9:157.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.
- Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci USA*. 112:8362–8366.
- Endress PK. 1999. Symmetry in flowers: diversity and evolution. *Int J Plant Sci*. 160:S3–S23.
- Fawcett JA, Maere S, Van de Peer Y. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous – tertiary extinction event. *Proc Natl Acad Sci USA*. 106:5737–5742.
- Fawcett JA, Van de Peer Y, Maere S. 2013. Significance and biological consequences of polyploidization in land plant evolution. In: Leitch IJ, editor. *Plant genome diversity*. Vienna: Springer. p. 277–294.
- Forest F. 2009. Calibrating the tree of life: fossils, molecules and evolutionary timescales. *Ann Bot*. 104:789–794.
- Fraley C, Raftery AE. 2002. Model-based clustering, discriminant analysis and density estimation. *J Am Stat Assoc*. 97:611–631.
- Fraley C, Raftery AE. 2003. Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *J Classif*. 20:263–286.
- Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol*. 60:433–453.
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152.
- Funk VA, Anderberg AA, Baldwin BG, Bayer RJ, Bonifacino JM, Breitwieser I, Brouillet L, Carbajal R, Chan R, Coutinho AXP, et al. 2009a. Compositae metatrees: the next generation. In: Funk VA, Susanna A, Stuessy TF, Bayer RJ, editors. *Systematics, evolution, and biogeography of Compositae*. Vienna: IAPT. p. 747–777.
- Funk VA, Chan R. 2009. Introduction to the Cichoroideae. In: Funk VA, Susanna A, Stuessy TF, Bayer RJ, editors. *Systematics, evolution, and biogeography of Compositae*. Vienna: IAPT. p. 335–342.
- Funk VA, Specht C. 2007. Meta-trees: grafting for a global perspective. *Proc Biol Soc Wash*. 120:232–240.
- Funk VA, Susanna A, Stuessy TF, Robinson H. 2009b. Classification of Compositae. In: Funk VA, Susanna A, Stuessy TF, Bayer RJ, editors. *Systematics, evolution, and biogeography of Compositae*. Vienna: IAPT. p. 171–192.
- Gentzmittel L, Vear F, Zhang YX, Berville A, Nicolas P. 1995. Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). *Theor Appl Genet*. 90:1079–1086.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 11:725–736.
- Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 27:221–224.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 29:644–652.
- Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered*. 100:605–617.
- Harlan JR, deWet JM. (ref1 co-authors). 1975. On Ö. Winge and a Prayer: the origins of polyploidy. *Bot Rev*. 41:361–390.
- Henry Y, Bedhomme M, Blanc G. 2006. History, protohistory and prehistory of the *Arabidopsis thaliana* chromosome complement. *Trends Plant Sci*. 11:267–273.
- Hickey LJ, Doyle JA. 1977. Early Cretaceous fossil evidence for angiosperm evolution. *Bot Rev*. 43:3–104.
- Hooker JJ, Collinson ME, Sille NP. 2004. Eocene–Oligocene mammalian faunal turnover in the Hampshire Basin, UK: calibration to the global time scale and the major cooling event. *J Geol Soc Lond*. 161:161–172.
- Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP, et al. 2015. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol Biol Evol*. 33:394–412.
- Hudson CM, Puckett EE, Bekaert M, Pires JC, Conant GC. 2011. Selection for higher gene copy number after different types of plant gene duplications. *Genome Biol Evol*. 3:1369–1380.
- Hughes AL. 2007. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* 99:364–373.
- Inoue J, Sato Y, Sinclair R, Tsukamoto K, Nishida M. 2015. Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. *Proc Natl Acad Sci USA*. 112:14918–14923.
- Jablonski D, Chaloner WG. 1994. Extinctions in the fossil record [and discussion]. *Philos Trans R Soc B: Biol Sci*. 344:11–17.
- Jeffrey C. 2007. Compositae: introduction with key to tribes. In: Kadereit JW, Jeffrey C, editors. *Families and genera of vascular plants, Flowering plants, Eudicots, Asterales*, vol. VIII. Berlin: Springer-Verlag. p. 61–87.
- Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, et al. 2012. A genome triplication associated with early diversification of the core eudicots. *Genome Biol*. 13:R3–R3.
- Jiao Y, Wickett NJ, Ayyampalayam S, Chandrabali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97–100.
- Kagale S, Robinson SJ, Nixon J, Xiao R, Huebert T, Condie J, Kessler D, Clarke WE, Edger PP, Links MG, et al. 2014. Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell* 26:2777–2791.
- Kellogg EA. 2016. Has the connection between polyploidy and diversification actually been tested? *Curr Opin Plant Biol*. 30:25–32.

- Kim K-J, Choi K-S, Jansen RK. 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Mol Biol Evol.* 22:1783–1792.
- Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.
- Levin DA. 1975. Minority cytotype exclusion in local plant populations. *Taxon* 24:35–43.
- Levin DA. 1983. Polyploidy and novelty in flowering plants. *Am Nat.* 122:1–25.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS. 2015. Early genome duplications in conifers and other seed plants. *Sci Adv.* 1:e1501084.
- Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L. 2005. Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J.* 41:221–230.
- Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA.* 102:5454–5459.
- Maere S, Van de Peer Y. 2010. Duplicate retention after small- and large-scale duplications. In: Dittmar K, Liberles D, editors. Evolution after gene duplication. Hoboken, NJ: Wiley. p. 31–56.
- Magallón S, Castillo A. 2009. Angiosperm diversification through time. *Am J Bot.* 96:349–365.
- Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207:437–453.
- Mandel JR, Dikow RB, Funk VA. 2015. Using phylogenomics to resolve mega-families: an example from Compositae. *J Syst Evol.* 53:391–402.
- Marshall D, Abbott R. 1984. Polymorphism for outcrossing frequency at the ray floret locus in *Senecio vulgaris* L. II: confirmation. *Heredity* 52:331–336.
- Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333:1257.
- McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early chordate evolution. *Nat Genet.* 31:200–204.
- Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30:i541–i548.
- Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecol Lett.* 17:508–525.
- O'Brien KP, Remm M, Sonnhammer ELL. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* 33:D476–D480.
- Ohno S. 1970. Evolution by gene duplication. Berlin: Springer-Verlag.
- Panero JL. 2016. Phylogenetic uncertainty and fossil calibration of Asteraceae chronograms. *Proc Natl Acad Sci USA.* 113:E411.
- Panero JL, Crozier BS. 2016. Macroevolutionary dynamics in the early diversification of Asteraceae. *Mol Phylogen Evol.* 99:116–132.
- Panero JL, Freire SE, Ariza Espinar L, Crozier BS, Barboza GE, Cantero JJ. 2014. Resolution of deep nodes yields an improved backbone phylogeny and a new basal lineage to study early evolution of Asteraceae. *Mol Phylogen Evol.* 80:43–53.
- Panero JL, Funk VA. 2008. The value of sampling anomalous taxa in phylogenetic studies: major clades of the Asteraceae revealed. *Mol Phylogen Evol.* 47:757–782.
- Paterson AH, Bowers JE, Burow MD, Draye X, Elsiek CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming R, et al. 2000. Comparative genomics of plant chromosomes. *Plant Cell* 12:1523–1540.
- Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA.* 101:9903–9908.
- Paterson AH, Freeling M, Tang H, Wang X. 2010. Insights from the comparison of plant genome sequences. *Annu Rev Plant Biol.* 61:349–372.
- Pelser PB, Watson LE. 2009. Introduction to Asteroideae. In: Funk VA, Susanna A, Stuessy TF, Bayer RJ, editors. Systematics, evolution, and biogeography of Compositae. Vienna: IAPT. p. 495–502.
- Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, et al. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19:651–652.
- Pontes O, Neves N, Silva M, Lewis MS, Madlung A, Comai L, Viegas W, Pikaard CS. 2004. Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. *Proc Natl Acad Sci USA.* 101:18240–18245.
- Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.
- Quental TB, Marshall CR. 2010. Diversity dynamics: molecular phylogenies need the fossil record. *Trends Ecol Evol.* 25:434–441.
- Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu Rev Ecol Syst.* 29:467–501.
- Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* 317:910–914.
- Robinson H. 1981. A revision of the tribal and subtribal limits of the Heliantheae (Asteraceae). *Smithson Contrib Bot.* 51:1–102.
- Robinson H, Carr GD, King RM, Powell AM. 1997. Chromosome numbers in Compositae, XVII: Senecioneae III. *Ann Mo Bot Gard.* 84:893–906.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol.* 19:101–109.
- Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.
- Sargent RD. 2004. Floral symmetry affects speciation rates in angiosperms. *Proc Biol Sci.* 271:603–608.
- Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC. 2004. Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47:868–876.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183.
- Schranz EM, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr Opin Plant Biol.* 15:147–153.
- Schranz ME, Mitchell-Olds T. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 18:1152–1165.
- Semple JC, Watanabe K. 2009. A review of chromosome numbers in Asteraceae with hypotheses on chromosomal base number evolution. In: Funk VA, Susanna A, Stuessy TF, Bayer RJ, editors. Systematics, evolution, and biogeography of Compositae. Vienna: IAPT. p. 61–72.
- Sharma B, Guo C, Kong H, Kramer EM. 2011. Petal-specific subfunctionalization of an APETALA3 paralog in the Ranunculales and its implications for petal evolution. *New Phytol.* 191:870–883.
- Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA.* 99:13627–13632.
- Smith E. 1975. The chromosome numbers of North American *Coreopsis* with phyletic interpretations. *Bot Gaz.* 136:78–86.
- Smith SA, Beaulieu JM, Donoghue MJ. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc Natl Acad Sci USA.* 107:5897–5902.
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *Am J Bot.* 96:336–348.

- Soltis DE, Buggs R, Barbazuk B, Chamala S, Chester M, Gallagher JP, Schnable PS, Soltis PS. 2012. The early stages of polyploidy: rapid and repeated evolution in *Tragopogon*. In: Soltis PS, Soltis DE, editors. Polyploidy and genome evolution. Berlin (Germany): Springer. p. 271–292.
- Soltis DE, Burleigh JG. 2009. Surviving the K–T mass extinction: new perspectives of polyploidization in angiosperms. *Proc Natl Acad Sci USA*. 106:5455–5456.
- Soltis DE, Soltis PS, Tate JA. 2003. Advances in the study of polyploidy since *Plant speciation*. *New Phytol*. 161:173–191.
- Soltis DE, Visger CJ, Marchant DB, Soltis PS. 2016. Polyploidy: pitfalls and paths to a paradigm. *Am J Bot*. DOI: 10.3732/ajb.1500501.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stebbins GL. 1940. The significance of polyploidy in plant evolution. *Am Nat*. 74:54–66.
- Stebbins GL. 1999. A brief summary of my ideas on evolution. *Am J Bot*. 86:1207–1208.
- Struck TH. 2014. TreSpEx-detection of misleading signal in phylogenetic reconstructions based on tree information. *Evol Bioinform*. 10:51–67.
- Stuessy TF, Garver D. 1996. The defensive role of pappus in heads of Compositae. In: Caligari PDS, Hind DJN, editors. Compositae: Biology & Utilization. Royal Botanic Gardens, Kew: Proceedings of the International Compositae Conference.
- Stuessy TF, Spooner DM, Evans KA. 1986. Adaptive significance of ray corollas in *Helianthus grosseserratus* (Compositae). *Am Midl Nat*. 115:191–197.
- Sun M, Ganders FR. 1990. Outcrossing rates and allozyme variation in rayed and rayless morphs of *Bidens pilosa*. *Heredity* 64:139–143.
- Tang H, Bowers JE, Wang X, Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc Natl Acad Sci USA*. 107:472–477.
- Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchliff CE, Brown JW, Sessa EB, Harmon LJ. 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol*. 207:454–467.
- Truco MJ, Ashrafi H, Kozik A, van Leeuwen H, Bowers J, Wo SRC, Stoffel K, Xu H, Hill T, Van Deynze A, et al. 2013. An ultra-high-density, transcript-based, genetic map of Lettuce. *G3 Genes Genom Genet*. 3:617–631.
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604.
- Van de Peer Y, Maere S, Meyer A. 2010. 2R or not 2R is not the question anymore. *Nat Rev Genet*. 11:166.
- Vanneste K, Baele G, Maere S, Van de Peer Y. 2014. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Res*. 24:1334–1347.
- Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290:2114–2117.
- Voss N, Eckstein RL, Durka W. 2012. Range expansion of a selfing polyploid plant despite widespread genetic uniformity. *Ann Bot*. 110:585–593.
- Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci USA*. 111: E4859–E4868.
- Yahara T, Kawahara T, Crawford DJ, Ito M, Watanabe K. 1989. Extensive gene duplications in diploid Eupatorium (Asteraceae). *Am J Bot*. 76:1247–1253.
- Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GK-S, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y, et al. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol*. 32:2001–2014.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
- Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292:686–693.
- Zachos JC, Dickens GR, Zeebe RE. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451:279–283.
- Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat Commun*. 5:4956.
- Zhang N, Zeng L, Shan H, Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol*. 195:923–937.
- Zhang R, Guo C, Zhang W, Wang P, Li L, Duan X, Du Q, Zhao L, Shan H, Hodges SA, et al. 2013. Disruption of the petal identity gene *APETALA3-3* is highly correlated with loss of petals within the buttercup family (Ranunculaceae). *Proc Natl Acad Sci USA*. 110:5074–5079.