

# Resolution of Brassicaceae Phylogeny Using Nuclear Genes Uncovers Nested Radiations and Supports Convergent Morphological Evolution

Chien-Hsun Huang,<sup>1</sup> Renran Sun,<sup>1</sup> Yi Hu,<sup>2</sup> Liping Zeng,<sup>1</sup> Ning Zhang,<sup>3</sup> Liming Cai,<sup>1</sup> Qiang Zhang,<sup>4</sup> Marcus A. Koch,<sup>5</sup> Ihsan Al-Shehbaz,<sup>6</sup> Patrick P. Edger,<sup>7</sup> J. Chris Pires,<sup>8</sup> Dun-Yan Tan,<sup>9</sup> Yang Zhong,<sup>1</sup> and Hong Ma<sup>\*1</sup>

<sup>1</sup>State Key Laboratory of Genetic Engineering and Collaborative Innovation Center of 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, The Huck Institute of the Life Sciences, Pennsylvania State University

<sup>3</sup>Department of Botany, National Museum of Natural History, MRC 166, Smithsonian Institution, Washington, DC

<sup>4</sup>Guangxi Key Laboratory of Plant Conservation and Restoration Ecology in Karst Terrain, Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and Chinese Academy of Sciences, Guilin, China

<sup>5</sup>Biodiversity and Plant Systematics, Centre for Organismal Studies Heidelberg, Heidelberg University, Heidelberg, Germany

<sup>6</sup>Missouri Botanical Garden, St. Louis

<sup>7</sup>Department of Horticulture, Michigan State University, East Lansing

<sup>8</sup>Division of Biological Sciences, University of Missouri, Columbia

<sup>9</sup>Xinjiang Key Laboratory of Grassland Resources and Ecology, College of Grassland and Environment Sciences, Xinjiang Agricultural University, Ürümqi, China

\*Corresponding author: E-mail: hongma@fudan.edu.cn.

Associate editor: Hongzhi Kong

## Abstract

Brassicaceae is one of the most diverse and economically valuable angiosperm families with widely cultivated vegetable crops and scientifically important model plants, such as *Arabidopsis thaliana*. The evolutionary history, ecological, morphological, and genetic diversity, and abundant resources and knowledge of Brassicaceae make it an excellent model family for evolutionary studies. Recent phylogenetic analyses of the family revealed three major lineages (I, II, and III), but relationships among and within these lineages remain largely unclear. Here, we present a highly supported phylogeny with six major clades using nuclear markers from newly sequenced transcriptomes of 32 Brassicaceae species and large data sets from additional taxa for a total of 55 species spanning 29 out of 51 tribes. Clade A consisting of Lineage I and *Macropodium nivale* is sister to combined Clade B (with Lineage II and others) and a new Clade C. The ABC clade is sister to Clade D with species previously weakly associated with Lineage II and Clade E (Lineage III) is sister to the ABCD clade. Clade F (the tribe Aethionemeae) is sister to the remainder of the entire family. Molecular clock estimation reveals an early radiation of major clades near or shortly after the Eocene–Oligocene boundary and subsequent nested divergences of several tribes of the previously polytomous Expanded Lineage II. Reconstruction of ancestral morphological states during the Brassicaceae evolution indicates prevalent parallel (convergent) evolution of several traits over deep times across the entire family. These results form a foundation for future evolutionary analyses of structures and functions across Brassicaceae.

**Key words:** ancestral character reconstruction, Brassicaceae, divergence time estimation, orthologous nuclear gene, phylogeny, transcriptome.

## Introduction

Although understanding of gene functions has relied heavily on studies of model organisms, such as the model plant *Arabidopsis thaliana* (L.) Heynh since 1940s (Meyerowitz and Somerville 1994; Meinke et al. 1998; Somerville et al. 2002; Flavell 2009), increasing efforts have been made on the comparison of different species (Gale and Devos 1998; Mitchell-Olds and Clausen 2002; Paterson et al. 2010). In particular, it is important to investigate whether gene functions

uncovered in a model organism are conserved among a wider group of species (Nardmann and Werr 2007; Paterson et al. 2010; Movahedi et al. 2011; Shrestha et al. 2014). Comparative studies can provide insights into the conservation and divergence of morphological and physiological characteristics (Sharma et al. 2011; Zhang et al. 2013). Therefore, a group of plants with a wide range of evolutionary, ecological, morphological, and physiological properties would be an excellent system for such comparative studies. In addition, members of

© The Author 2015. 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

the group should include model system(s) to provide a foundation for understanding gene functions, and inclusion of crop plants would facilitate the application of the generated new insights in agriculture.

Brassicaceae is an attractive candidate as an evolutionary model family in many ways. First, Brassicaceae is one of the most well-known and diverse plant families, with ~3,700 species and many important agricultural and horticultural members. Recently, Brassicaceae has been delineated into 51 tribes and 321 genera (Al-Shehbaz 2012; Al-Shehbaz et al. 2014; German and Friesen 2014). The family is distributed worldwide, with numerous members being relatively easy to access in temperate regions of the Northern Hemisphere (Al-Shehbaz et al. 2006). Second, much of the current knowledge of gene functions in developmental and physiological processes of flowering plants comes from studies using the model plant *A. thaliana* (Meinke et al. 1998). Additional functional studies have been carried out with related crops such as *Brassica* species (Mathur et al. 2013; Kim et al. 2014; Liu et al. 2014).

Besides *Arabidopsis* and *Brassica* species, other Brassicaceae taxa have also been investigated for various aspects in plant biology, such as biotic and abiotic stress tolerance, genome evolution, self-incompatibility, morphological studies, and so on. To date, this family contains abundant genetic and genomic resources including 19 publicly available genome sequences (Gong et al. 2005; Amasino 2009; Mummenhoff et al. 2009; Dassanayake et al. 2011; Hu et al. 2011; Rushworth et al. 2011; Claire-Lise and Nathalie 2012; Wu et al. 2012; Haudry et al. 2013; Slotte et al. 2013; Verbruggen et al. 2013; Anderson et al. 2014; Halimaa et al. 2014; Vekemans et al. 2014). In addition, Brassicaceae experienced whole-genome duplications and organismal radiation in its early evolutionary history (Vision et al. 2000; Simillion et al. 2002; Henry et al. 2006; Schranz and Mitchell-Olds 2006; Couvreur et al. 2010; Edger et al. 2015), providing great opportunities to investigate these important evolutionary processes. In short, Brassicaceae has many advantages to be a model family for comparative and evolutionary studies (Franzke et al. 2011).

To facilitate the use of Brassicaceae as an evolutionary model family, a phylogeny with well-supported relationships is crucial. Numerous previous phylogenetic studies (Al-Shehbaz et al. 2006; Bailey et al. 2006; Beilstein et al. 2006, 2008; Koch et al. 2007; Franzke et al. 2009, 2011; German et al. 2009; Khosravi et al. 2009; Couvreur et al. 2010; Warwick et al. 2010; Hohmann et al. 2015) provide the current understanding of Brassicaceae phylogeny (see the *BrassiBase* website <http://brassibase.cos.uni-heidelberg.de/>; Koch et al. 2012; Kiefer et al. 2014; supplementary fig. S1a, Supplementary Material online). Notably, Beilstein et al. (2006) analyzed the chloroplast *ndhF* gene from 113 species of 101 genera and identified 3 major lineages (i.e., Lineages I–III and LI–LII) with low to moderate supports. Group of additional genera associated with LII with low support was referred to as Expanded Lineage II (EII) by Franzke et al. (2011).

However, the deep relationships of Brassicaceae still lack strong support, possibly due to the rapid radiation event(s) in the early evolutionary history (Al-Shehbaz et al. 2006; Bailey et al. 2006; Koch et al. 2007; Franzke et al. 2009; Couvreur et al. 2010) associated with polyploidization events (Lysak et al. 2005; Mandáková and Lysak 2008; Mandáková, Joly, et al. 2010) and origin of novel traits (Edger et al. 2015). Even at the tribal level multigene approaches to resolve a reliable and highly confident phylogenetic backbone are rare (such as Karl and Koch 2013). Therefore, numerous uncertainties remain among tribes and genera of Brassicaceae, particularly along the backbone.

Low or single-copy nuclear genes among many species have been tested recently as effective phylogenetic markers (Zhang et al. 2012; Zimmer and Wen 2013; Zeng et al. 2014; Yang et al. 2015). Although nuclear genes have more complex structure and are harder to amplify compared with chloroplast genes, rapid advances of next-generation sequencing technology have made large data sets accessible, allowing high-throughput selection of nuclear genes as phylogenetic markers. Recent studies used transcriptome data sets to successfully reconstruct phylogenies of various scales from genus to angiosperm-wide, even land plants (Wen et al. 2013; Kagale et al. 2014; Wickett et al. 2014; Zeng et al. 2014; Yang et al. 2015).

To provide nuclear gene-sequence resources for evolutionary studies of Brassicaceae, here we generated 33 transcriptome data sets, including 32 Brassicaceae species and 1 outgroup species *Cleome serrulata* from Cleomaceae, the sister family of Brassicaceae. Combined with 10 public genomic data sets and 13 additional collaborative transcriptome data sets, a total of 55 large data sets of Brassicaceae species were used to select a set of marker genes for phylogenetic reconstructions. A total of 113 orthologous nuclear markers were obtained from 3 groups of orthologs identified using different initial gene sets. Genes with putative misleading signal were pruned after analyses for long-branch attraction and saturation of partitions. Both concatenation and coalescence analyses were performed with consistent results. Finally, a strongly supported phylogeny was obtained and provides for the first time a highly resolved backbone with six major clades named Clade A–F, further expanding the LI and LII as Clade A and B, respectively, and defining two new clades (Clades C and D) with several species of the previous polytomous EII.

Using the new phylogenetic hypothesis, divergence times were estimated with multiple fossil constraints, thereby strongly supporting the view that nested radiations contributed substantially to the Brassicaceae diversity. We further used the new phylogenetic hypothesis as an evolutionary framework for reconstructing the ancestral states of 19 morphological characters, including fruit lengths and flattening, cauline leaf, and trichome presence and shape, revealing multiple parallel/convergent evolutionary events. Our transcriptomic sequence resources and the robust phylogeny support Brassicaceae as a model family for further evolutionary and other comparative studies.

## Results

### Selection of 113 Low-Copy Orthologous Nuclear Genes from 55 Brassicaceae Data sets

In this study, we obtained transcriptome data sets of 32 Brassicaceae species and *Cl. serrulata* as an outgroup (supplementary table S1, Supplementary Material online). Augmented with 10 whole-genome sequences obtained from public databases, and 13 transcriptome data sets generated in the laboratory of one of the authors (J.C.P), our phylogenetic analyses (below) included a total of 55 Brassicaceae species, covering 29 of the 51 tribes (supplementary table S2, Supplementary Material online).

Three groups of nuclear markers were selected separately from the orthologs shared by different sets of species to avoid method-dependent biases. First, we use a set of previously described low-copy nuclear genes (Zhang et al. 2012) shared by nine angiosperms (*Zea mays*, *Sorghum bicolor*, *Oryza sativa*, *Glycine max*, *Medicago truncatula*, *Populus trichocarpa*, *Vitis vinifera*, *A. thaliana*, and *Solanum lycopersicum*) shown to be effective in resolving the deep angiosperm phylogeny (Zeng et al. 2014). Then, two additional groups of low-copy nuclear genes were identified separately; one was a group of genes shared by the above-mentioned nine and seven angiosperm genomes (*A. thaliana*, *Po. trichocarpa*, *Prunus persica*, *V. vinifera*, *Mimulus guttatus*, *O. sativa*, and *Sor. bicolor*), and the other group consists of single-copy genes overlapping between the above nine angiosperms and five Brassicaceae species (*Arabidopsis lyrata*, *A. thaliana*, *Brassica rapa*, *Capsella rubella*, and *Eutrema salsugineum*). For additional information on the identification of orthologs, please see Materials and Methods. All three sets of putative orthologs were further analyzed as described below.

A major difficulty in reconstructing Brassicaceae phylogeny is that polyploidization events or gene/genome duplications (and subsequent gene losses) affect many species across Brassicaceae (Lysak et al. 2005; Mandáková and Lysak 2008; Mandáková, Joly et al. 2010; Kagale et al. 2014; Hohmann et al. 2015), resulting in hidden paralogs (remaining single copy genes after the loss of distinct paralogs in different taxa), which should be avoided in phylogenetic analysis. Thus, we performed phylogenetic analysis of individual genes to identify such hidden paralogs, using a well-supported relationship of 13 Brassicaceae species in terms of the three major lineages (i.e., LI to LIII) (supplementary fig. S1, Supplementary Material online), a consensus from all previous studies (see Franzke et al. 2011 and references therein). Notably, we focused on the most basic topology of the 13 species regardless of the relationships among and within different lineages to minimize possible bias of this selection. After examining each of the single-gene trees, 28, 45, and 40 genes were selected from the three above-mentioned groups of putative orthologs, respectively. A total of 113 low-copy orthologous nuclear genes had 177,171 nucleotides and 59,057 amino acids (supplementary table S3, Supplementary Material online).

To further test the possible effect of sequence bias, TreSpEx (Struck 2014) was used to analyze long-branch (LB) attraction (determined by heterogeneity or longest branches) and saturation (determined by slope or  $R^2$  of linear regression) of partitions, resulting in 10 and 13 genes with heterogeneity or longest branches (supplementary fig. S2a and b, Supplementary Material online), and 9 and 2 genes to be saturated inferred by slope or  $R^2$  of linear regression, respectively (supplementary fig. S2c and d, Supplementary Material online). For the identity of these genes, see supplementary table S4, Supplementary Material online. Pruning such genes with putative misleading signals from each analysis resulted in four data sets. In addition, an analysis of LB attraction generated a heat map for taxon-specific, LB attraction scores of each gene (supplementary fig. S3, Supplementary Material online). Here, relatively few genes were found with possible misleading signals, suggesting that most selected genes are reliable phylogenetic markers.

### A Highly Supported Family-Wide Phylogeny of Brassicaceae

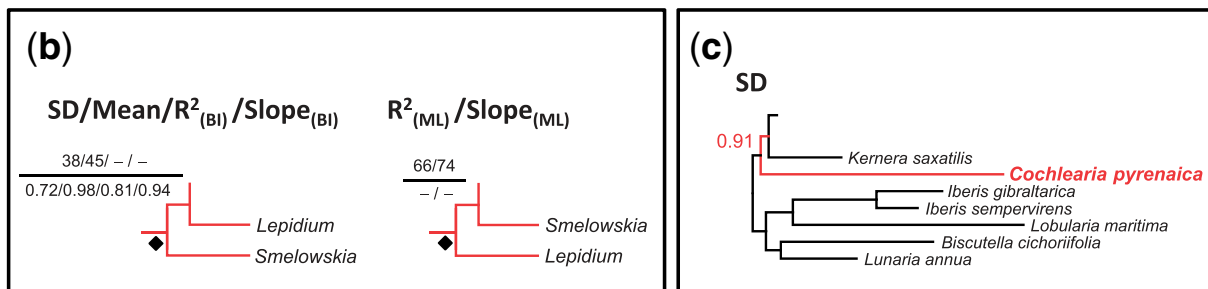
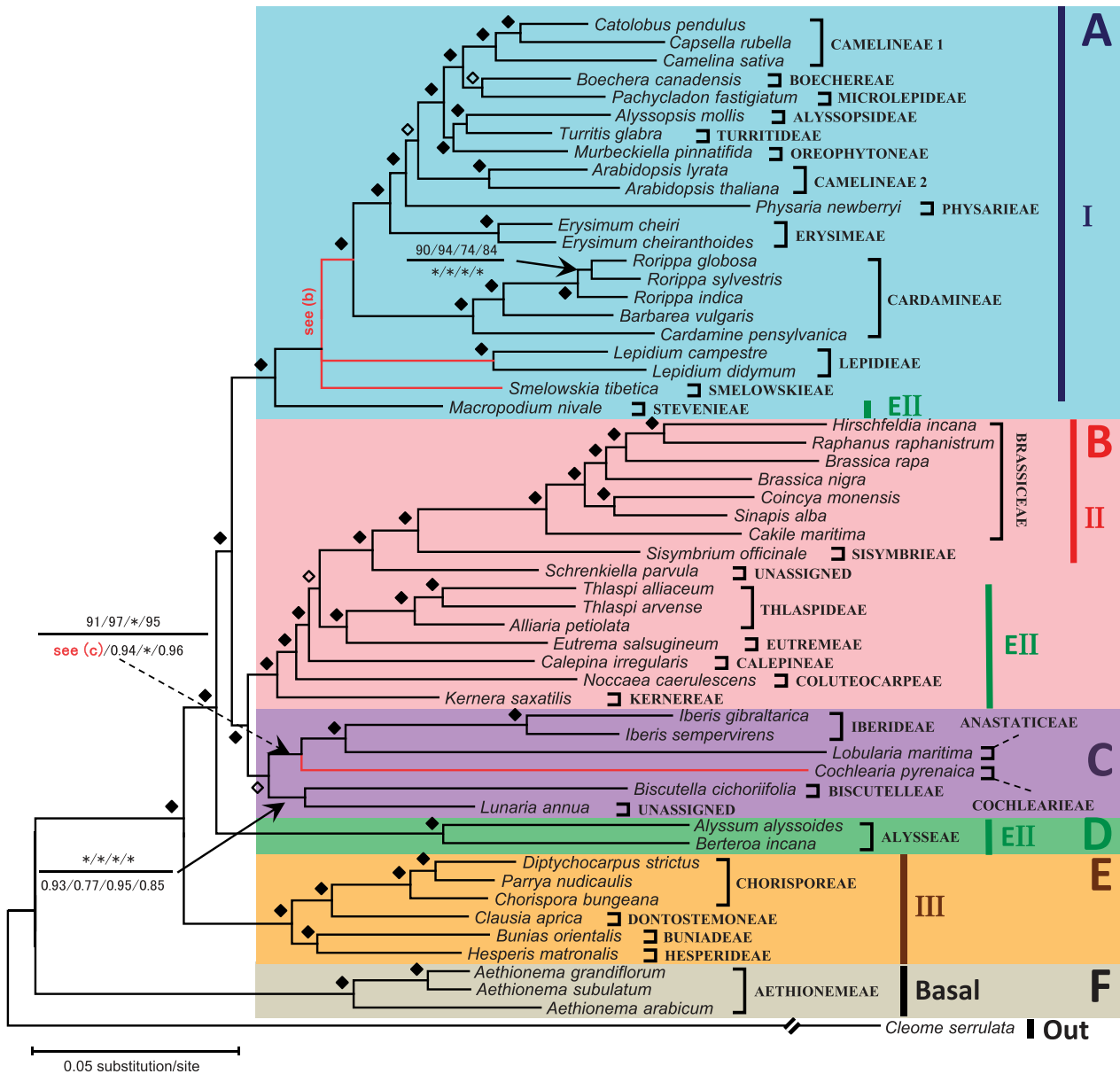
Maximum likelihood (ML) analysis and Bayesian inference (BI) were applied for phylogenetic reconstruction using sequence concatenations of each of the three gene groups from different selections (supplementary figs. S4–S6, Supplementary Material online), the combined 113 gene supermatrix (supplementary fig. S7, Supplementary Material online), and each of the four data sets without those genes with significant sequence bias (i.e., LB and saturation) from separate tests (fig. 1). ML analyses under different models were also performed using 113 gene concatenation (supplementary fig. S8, Supplementary Material online). Furthermore, considering the concerns that the concatenation method might result in biased phylogeny with strong support (Seo 2008), we also constructed a Brassicaceae phylogeny using the multispecies coalescent analysis (Mirarab et al. 2014) with the 113 markers (supplementary fig. S9, Supplementary Material online), in which the distance matrices for multilocus sequence data were estimated separately and these matrices were subsequently combined to reconstruct the phylogeny.

The Brassicaceae phylogeny is summarized in figure 2 according to the trees reconstructed by concatenation and coalescent analyses with 113 genes (supplementary figs. S7 and S9, Supplementary Material online), and trees using markers after bias tests (fig. 1). Generally, these results agree with each other for the vast majority of positions with high supports and the relevant relationships are shown as solid lines. However, positions of four tribes exhibiting conflicts here (fig. 3) require further investigations and are presented as dashed lines in figure 2. Detailed description for each clade is provided in the supplemental text, Supplementary Material online.

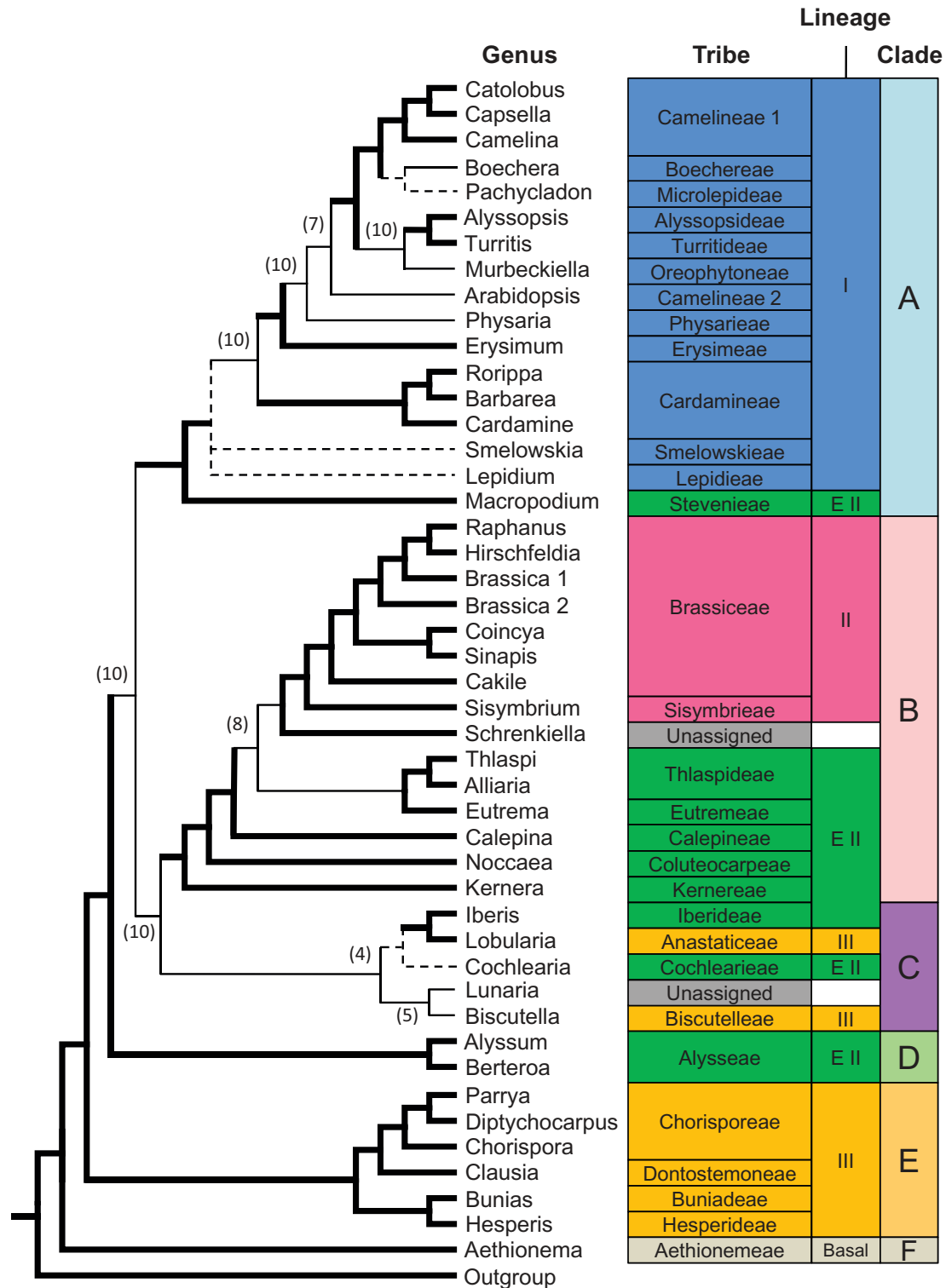
At the family level, there are six well-supported major clades (A–F) in all nine topologies. The relationships among the six major clades are consistent among the trees,

(a)

Clades:



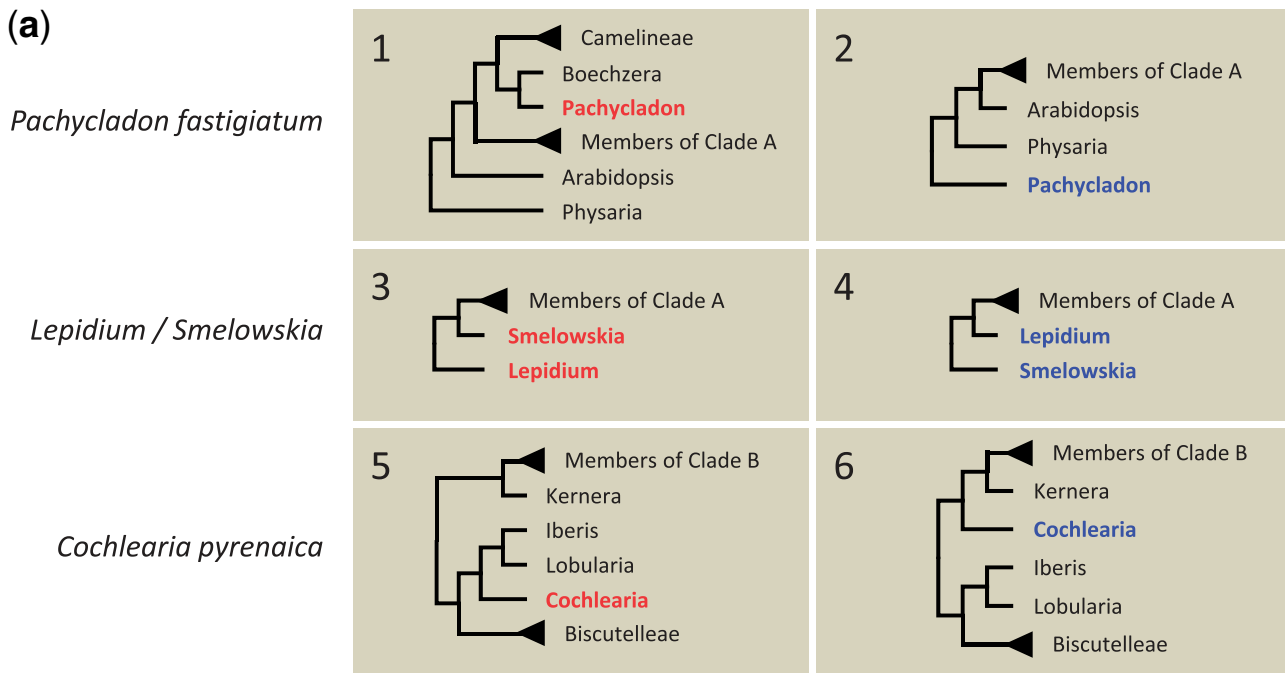
**Fig. 1.** Phylogenetic tree from ML and BI analyses of concatenations of gene sets after sequence bias tests. Solid and open diamonds indicate maximum (BS = 100 or PP = 1) and supports of  $\geq 90$  (BS) or 0.9 (PP) in all tests, respectively. Values correspond to the statistical support of bootstrap value of ML analysis (BS; above) and Bayesian posterior probabilities (PP; below) in the sequence of the tests of [supplementary figure S2a-d](#), [Supplementary Material](#) online, respectively. Stars (\*) indicate maximum support in at least one of the analyses. A complete phylogeny consistent in all tests is in (a). Relationships between *Smelowskia* and *Lepidium* and their supports in each test are shown in (b). Position of *Cochlearia* and its support in BI analysis of data set screened by standard deviation of LB score (SD) are shown in (c). I, II, III, and EII indicate Lineage I, II, III, and Expanded Lineage II, respectively.



**Fig. 2.** A summarized Brassicaceae phylogeny. This model is proposed according to our results from concatenation (using 113 genes as markers, supplementary fig. S7, Supplementary Material online, and trees after bias tests, fig. 1) and coalescent methods (supplementary fig. S9, Supplementary Material online). Structures consistent in all reference trees are drawn in thick (all showing maximum support) or thin solid lines (with values showing the number of results with maximum support). Dashed lines indicate uncertainties, and the details are presented in figure 3.

notably with maximum support in the trees using concatenation matrices (fig. 1 and supplementary fig. S7, Supplementary Material online). Clade A is sister to a combined clade of Clade B and Clade C. Then, the combined ABC clade is sister to Clade D, with Clade E being the sister of the

ABCD clade. *Aethionema* is the basal-most group as Clade F, consistent with previous conclusions (Al-Shehbaz et al. 2006). We then performed the approximately unbiased (AU) test to confirm the pattern of the five major clades (A–E) of the core Brassicaceae. Among the 48 alternative topologies analyzed,



**(b)**

Figure		S7		1								S9
Analysis		113 genes		Bias/SD		Bias/Mean		Bias/R <sup>2</sup>		Bias/Slope		Coalescence
Method		ML	BI	ML	BI	ML	BI	ML	BI	ML	BI	ML
<i>Pachycladon fastigiatum</i>	Topology	1	1	1	1	1	1	1	1	1	1	2
	Support	100	1	100	1	100	1	100	0.97	100	1	93
<i>Lepidium / Smelowskia</i>	Topology	3	3	4	4	4	4	3	4	3	4	3
	Support	76	1	38	0.72	45	0.98	66	0.81	74	0.94	74
<i>Cochlearia pyrenaica</i>	Topology	5	6	6	6	5	5	5	5	5	5	6
	Support	89	1	91	0.91	97	0.94	100	0.98	95	0.96	82

Topologies		Support
1,3,5	≥ 80 or 0.8	
	< 80 or 0.8	
2,4,6	≥ 80 or 0.8	
	< 80 or 0.8	

**FIG. 3.** Summary of conflicting results in this study. There are four tribes that encountered conflicting results in our analyses, and each of the topologies is illustrated in (a). Topologies and supports for these tribes from each reconstruction are summarized in (b). Number of each topology in (a) corresponds to those in (b). Most results for *Pachycladon* and *Cochlearia* are topologies 1 and 5, respectively, with high supports and are presented in figure 2 (dashed lines).

only 2 topologies cannot be rejected ( $P > 0.05$ ; supplementary table S5, Supplementary Material online). The topology in our hypothesis has a relatively high  $P$  value of 0.961 as compared with that of the other one ( $P = 0.06$ ), in which Clade C is sister to Clade A instead of Clade B.

The three previously proposed lineages (Bailey et al. 2006; Beilstein et al. 2006, 2008; Koch et al. 2007; Franzke et al. 2009) are placed in different clades here. Specifically, all species of LI and LII analyzed here are in Clades A and B, respectively, whereas those of LIII are mostly in Clade E, with a few in

the newly defined Clade C. The EII was defined by Franzke et al. (2011) as a group that included LII and a number of paraphyletic taxa near LII. Relationships of the latter species were difficult to resolve and they were often presented with a bush-like structure (Franzke et al. 2011 and references therein). The EII species sampled here are separately placed in Clades A, B, C, and D. Generally, our phylogeny is in good agreement with the highly supported aspects of topologies using chloroplast genes (Beilstein et al. 2006, 2008; Hohmann et al. 2015) and contains additional well-supported relationships, forming a highly resolved backbone in particular.

To add 22 more species lacking a transcriptome data set to the phylogeny, we choose three genes from our 113 markers, which lack introns and thus are relatively easy to amplify from genomic DNAs, with sufficient lengths (~1,500 bp) (supplementary table S3, Supplementary Material online). Five recently published genomes of Brassicaceae (Haudry et al. 2013; Phytozome v10.1) are also included using 113 markers. The resulting 81 species spans 32 Brassicaceae tribes (including 3 more tribes; supplementary table S2, Supplementary Material online) representing the vast majority of Brassicaceae species (2,662 of the 3,700 species). As shown in supplementary figure S10, Supplementary Material online, the deep relationships are the same as those from the analyses using only genes from large data sets (fig. 2) with high supports (bootstrap values > 87). Phylogenies of taxa with new genome data sets are all maximally supported, whereas the nodal supports of the species with only polymerase chain reaction (PCR)-amplified sequences varied (bootstrap values 18 to 100). Placements of all additional species are as expected. Notably, species of the LII tribe Isatideae group well with other LII tribes (Sisymbrieae and Brassiceae). Species of the LIII tribes Anthonieae and Euclidieae are all placed within Clade E, in a subclade containing Buniadeae and Hesperideae. However, their phylogenetic positions should be considered as tentative and needs further validation.

### Age Estimation Reveals an Early Radiation Leading to the Divergences of Major Clades Near the Eocene–Oligocene Boundary and the Subsequent Radiations

Molecular clock analysis was performed to estimate divergence time using concatenation of 113 nuclear markers (supplementary fig. S7, Supplementary Material online) by penalized likelihood (PL) implemented in r8s (Sanderson 2003). In addition to the data set with only *Cl. serrulata* as the outgroup species in phylogeny reconstruction (fig. 1), we also included orthologous genes from 54 more outgroup taxa (supplementary tables S2, Supplementary Material online) to allow the use of 18 non-Brassicaceae calibrations (supplementary table S6, Supplementary Material online), including one Brassicales fossil *Dressiantha bicarpelata* (Gandolfo et al. 1998). Assignments and ages of these fossils are according to Smith et al. (2010) and Magallón et al. (2015) (supplementary table S6, Supplementary Material online). An Oligocene fossil *Thlaspi primaevum* (Becker 1961; Wing 1987; Manchester and O'Leary 2010) was assigned to Brassicaceae and used as a minimum constraint of the divergence of

*Thlaspi* and *Alliaria* in Brassicaceae (Beilstein et al. 2010). However, assignment of this fossil based on fruit character is uncertain due to the prevalence of morphological homoplasy across the family, especially for characters associate with fruit; such difficulties have led to wrong taxonomic assignments including that in *Thlaspi* (Mummenhoff et al. 1997; Franzke et al. 2010). Thus, separate analyses were performed without and with the fossil *T. primaevum* (supplementary fig. S11a and b, Supplementary Material online). The estimated mean and 95% highest posterior density (HPD) ages of several nodes of interest are presented in table 1. In general, analyses with or without the fossil *T. primaevum* produced similar results with differences of about 5 My in the divergence times of this family and core Brassicaceae (supplementary fig. S11, Supplementary Material online, and table 1). Figure 4 illustrates the correspondence of our estimates to palaeoclimate proposed by Zachos et al. (2008).

Our results suggest that the mean age of Aethionemeae is 37 or 42 My from analysis without or with *T. primaevum* (table 1), and then the five major clades, accounting for the majority of the family (i.e., core Brassicaceae), diverged during a very short period during Oligocene (figure 4). Interestingly, rates of the subsequent tribal divergences are distinct in different clades. Several tribes of Clade B and those of Clade C continued to diverge rapidly during Oligocene as nested radiations. Clade E separated from others early in Brassicaceae history, but its tribes diverged during middle to late Miocene, significantly later than those in Clades B and C, showing a lag time between the radiation of major clades and divergences of tribes within the clade. In Clade A, divergence of *Macropodium* was early in middle to late Oligocene, while the other tribes separated subsequently during Miocene. It is noteworthy that divergences of almost all the EII tribes sampled here occurred soon after the early radiation of Brassicaceae (figure 4). For example, in Clade B, EII tribes diverged during middle to late Oligocene whereas those of LII diverged later during middle Miocene. The deep divergences might explain the difficulties in placing these groups previously using limited markers (Franzke et al. 2011 and references therein).

### Ancestral Character Reconstruction Reveals Numerous Convergent Evolution Events

Relying on the high resolution of deep nodes with branch lengths of the phylogenetic hypothesis here, ancestral states of Brassicaceae were reconstructed for 19 morphological characters, such as fruit length–width ratio, fruit flattening, and dehiscence. Others included presence and shape of petals, base of cauline (stem) leaves, shape of trichomes (hair cells), and the presence of multicellular glands. Coding of these characters and the states for each genus is presented in supplementary tables S7 and S8, Supplementary Material online, and the reconstructed results are shown in figure 5 and supplementary figures S12–S22, Supplementary Material online. It should be noted that the reconstructed characters are tentative because of the limited sampling here.

**Table 1.** Mean and 95% HPD Age Estimates of Nodes of Interest across 100 Bootstrap Replicates.

Clade	<i>Thlaspi primaevum</i> Fossil	
	Not Used	Used
MRCA Cleomaceae/Brassicaceae	54.3-55.5-56.8	58.4-59.5-60.7
MRCA Clades ABCDEF (CG Brassicaceae)	36.3-37.1-37.8	41.4-42.0-42.7
MRCA Clades ABCDE (CG core-Brassicaceae)	29.1-29.7-30.3	34.6-35.1-35.6
MRCA clades ABCD	27.9-28.5-29.0	33.5-34.0-34.4
MRCA Clades ABC (also as MRCA <i>Arabidopsis</i> /Brassica)	26.9-27.4-28.0	32.6-33.1-33.5
MRCA Clade A	23.1-23.6-24.0	27.7-28.1-28.6
MRCA <i>Lepidium</i> / <i>Arabidopsis</i>	20.5-20.9-21.3	24.5-24.8-25.2
MRCA <i>Camelina</i> / <i>Arabidopsis</i>	14.3-14.6-14.9	17.0-17.2-17.5
MRCA Clades BC	26.1-26.6-27.1	31.9-32.3-32.8
MRCA Clade B	23.3-23.7-24.2	30.1-30.5-30.8
MRCA <i>Noccaea</i> /Brassica	22.1-22.5-22.9	29.2-29.5-29.8
MRCA <i>Calepina</i> /Brassica	21.3-21.7-22.1	28.6-28.9-29.2
MRCA <i>Thlaspi</i> /Brassica	20.8-21.2-21.6	28.1-28.4-28.7
MRCA <i>Schrenkiella</i> /Brassica	18.6-19.0-19.4	25.0-25.3-25.6
MRCA <i>Sisymbrium</i> /Brassica	16.9-17.2-17.5	22.5-22.8-23.1
MRCA Brassiceae	12.6-12.8-13.1	16.5-16.7-16.9
MRCA Clade C	25.1-25.7-26.2	30.7-31.1-31.6
MRCA <i>Cochlearia</i> /Iberis	23.8-24.3-24.8	29.0-29.4-29.9
MRCA <i>Lobularia</i> /Iberis	21.7-22.2-22.7	26.4-26.9-27.4
MRCA Biscutelleae	20.9-21.8-22.6	25.4-26.4-27.4
MRCA Clade E	16.8-17.2-17.6	19.3-19.7-20.1

NOTE.—Values in bold are mean ages (Mya) and are bracketed by lower and upper bounds of 95% HPD interval.

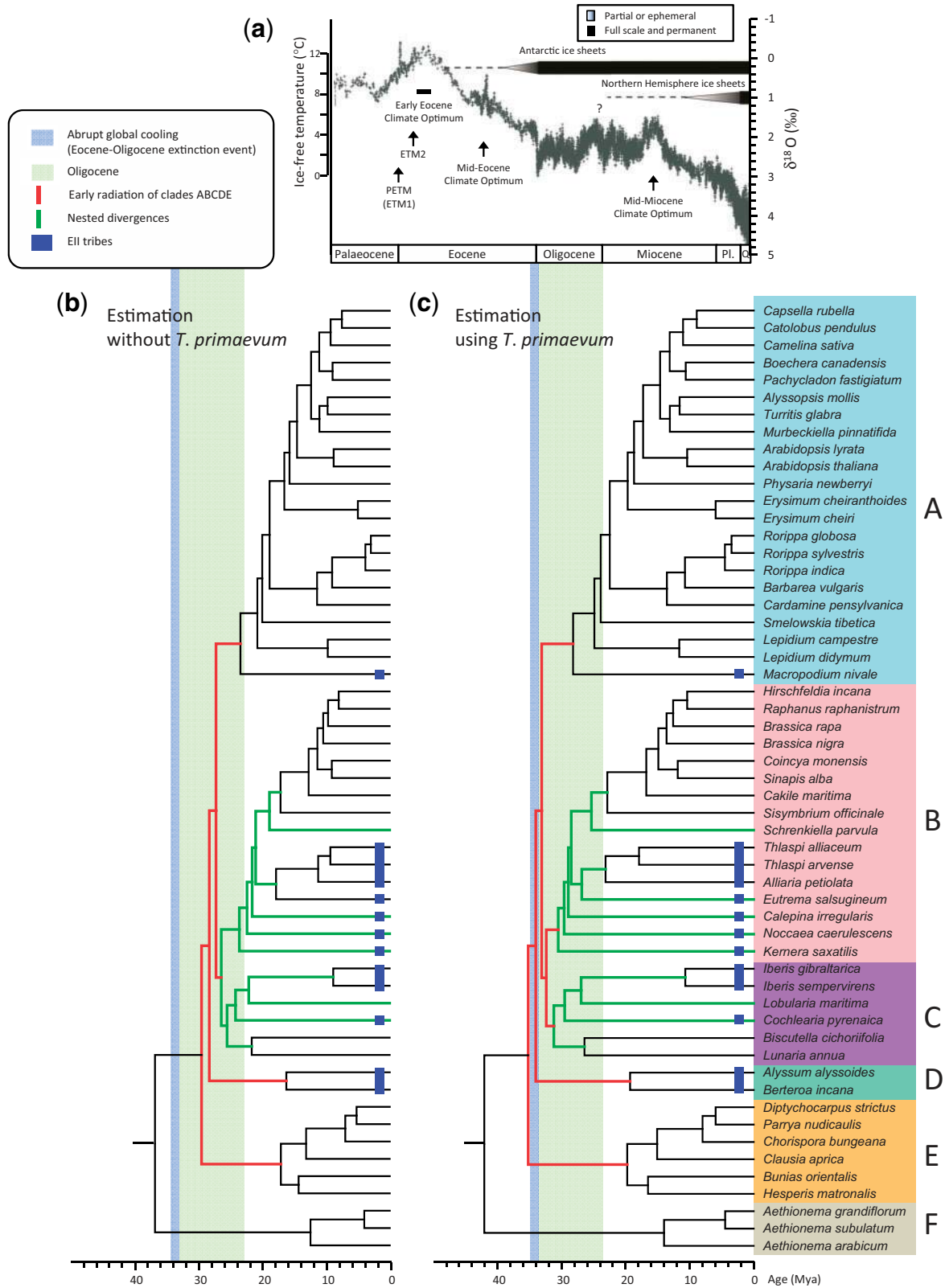
Most characters examined here show generally parallel evolution consistent with previous observations (Al-Shehbaz et al. 2006; Franzke et al. 2011 and references therein), but with strong support from the phylogeny. For fruit length, silicle (short) is ancestral for the Brassicaceae crown group (CG) (fig. 5a) with 0.65 proportional likelihoods. Silique (long) evolved early in Clades A and E and within Clade B, with an obvious transition during late Oligocene to early Miocene. Proportional likelihoods for the most recent common ancestors (MRCA) of Clade A, MRCA of LII tribes and *Schrenkiella* (together constituting the derived tribes of Clade B), and MRCA of Clade E with siliques are 0.64, 0.77 and 0.79, respectively. On the contrary, fruits of Clade C and most of the early divergent tribes (previously designated EII) of Clade B have retained the silicle state. Also, some lineages in Clades A and E show evidence of reversion from silique to silicle. Ancestors of CG Brassicaceae possibly had latiseptate fruits (flattened parallel to septum; 0.54 proportional likelihoods) (fig. 5b) retaining along the branches to tribes of Clade A, although most of the extant species of this clade are not latiseptate, indicating a high degree of plasticity for this trait which might result from frequent parallel evolution. A trend of terete (not flattened, cylindrical) evolved along the branches to the derived tribes of Clade B, and the MRCA of LII (0.56 proportional likelihoods) possibly had terete fruits during early Miocene.

The ancestral state for the base of cauline leaves of this family was likely petiolate with 0.67 proportional likelihoods (fig. 5c). Auriculate (ear-shaped), amplexicaul (clasping the stem), or sagittate (arrowhead-shaped) cauline leaves evolved

along the branches to late divergent tribes of Clade A and were most likely the character state of MRCA Camelineae/Physarieae or even MRCA Camelineae/Erysimeae (0.6 and 0.52 proportional likelihoods, respectively) in middle Miocene. In addition to Clade A, sessile cauline leaves are present in many extant taxa of the basal tribes of Clade B, whereas petiolate cauline leaves seemed to be the ancestral character of this clade (0.78 proportional likelihoods). A more complete sampling is needed to determine how this highly plastic trait evolved. Simple trichomes were associated with the ancestors of CG Brassicaceae (0.72 proportional likelihoods) and maintained in all species of Clade B and most species of Clade C (fig. 5d) except for *Lobularia*, which has evolved branch trichomes. Branch trichomes also evolved in Clade A independently after the divergence of some early-separating tribes, showing a character conversion during early Miocene (0.54 proportional likelihoods for MRCA Camelineae/Erysimeae). Some extant species of Clades D and E also show branch trichomes, indicating convergent evolution in multiple lineages.

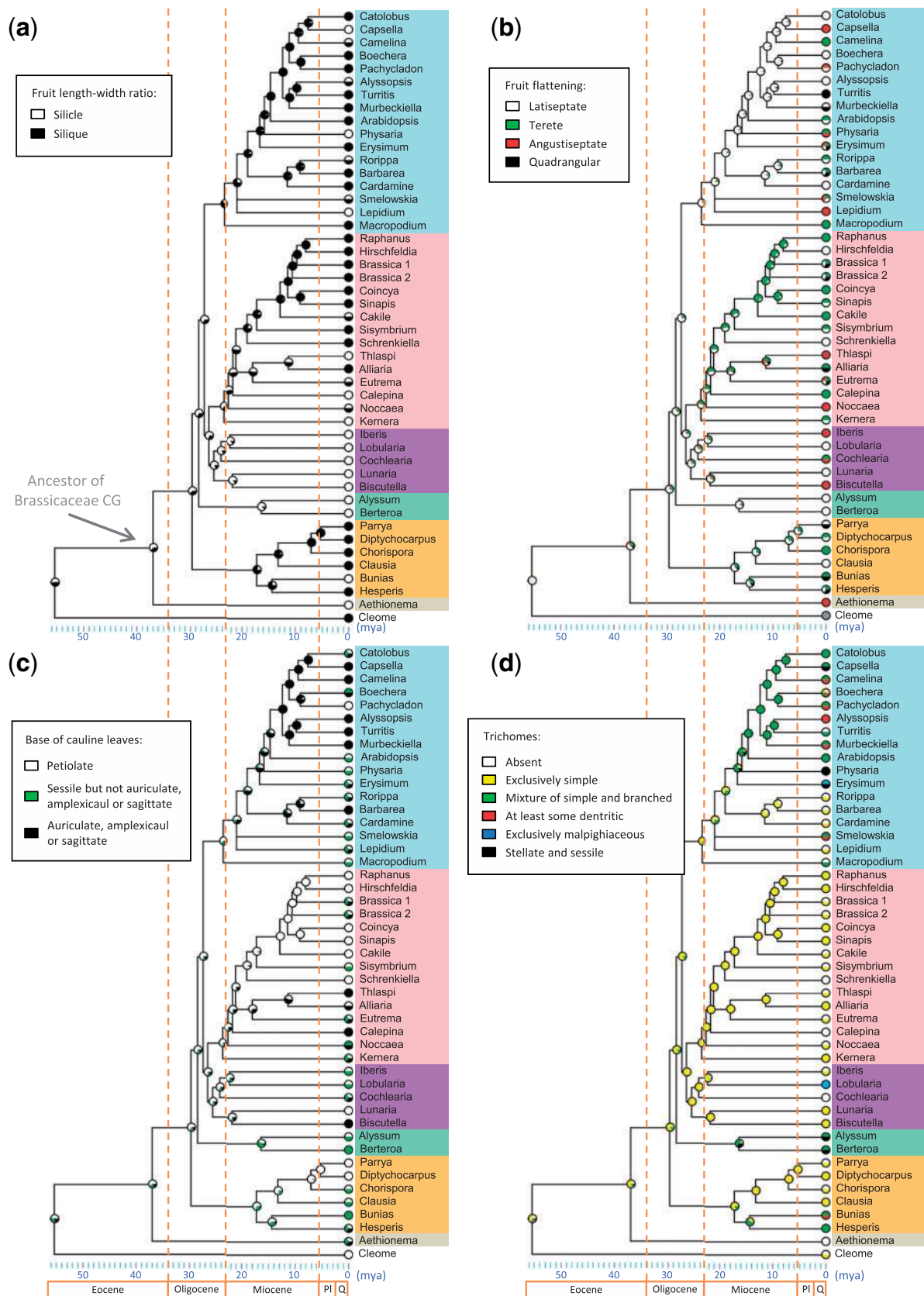
Multicellular glands evolved along the branches to Clade E only and have been maintained in most of the genera of this clade (fig. 5e). Fruit segmentation is only observed in the genera of tribe Brassiceae and genus *Chorispora* (fig. 5f). The phylogenetic distance between the two tribes strongly supports convergence. Ancestral seeds of Brassicaceae likely did not have wings (0.91 proportional likelihoods; fig. 5g), while winged seed was traced to the ancestor of Chorisporaeae (0.63 proportional likelihoods) in late Miocene. Extant species of Clade A show both winged and





**FIG. 4.** Chronogram of Brassicaceae inferred using r8s. We use the phylogeny resulting from 113 gene concatenation in our age estimation. (a) A stacked deep-sea benthic foraminiferal oxygen-isotope curve shows the evolution of global climate over the last 65 My, as modified from figure 2 in Zachos et al. (2008). Chronogram of the subclade regarding Brassicaceae and *Cl. serrulata* was extracted from supplementary figure S11, Supplementary Material online, in which the Brassicaceae fossil *T. primaevum* was not implemented (b) or was implemented (c) in the estimation. Blue strips highlight tribes of EII. Branches in red indicate the radiation leading to five major clades of core Brassicaceae, and those in green denote the nested divergences subsequently (no longer than 3 My for each of the indicated branches). PETM: Palaeocene–Eocene thermal maximum; ETM: Eocene thermal maximum; Pl: Pliocene. Reprinted by permission from Macmillan Publishers Ltd: Nature (451, 279–283), copyright (2008).

Downloaded from https://academic.oup.com/mbe/article/39/2/394/2579467 by guest on 16 August 2022



**FIG. 5.** Evolutionary patterns of selected morphological characters. (a) Fruit length–width ratio, (b) fruit flattening (c) base of cauline leaves, (d) trichome shape, (e) multicellular glands, (f) fruit segmentation, (g) seed wing and (h) stigma lobing in the context of our dated topology. The character reconstructions based on dated topology resulted from figure 4b and we used genera as terminals instead of species. Area of different colors in the nodes represents the proportional likelihoods for each character state to be ancestral. Results regarding other 11 characters are presented in supplementary figure S12–S22, Supplementary Material online. Pl: Pliocene; Q: Quaternary. Information for the state of fruit flattening of *Cleome* is lacking.

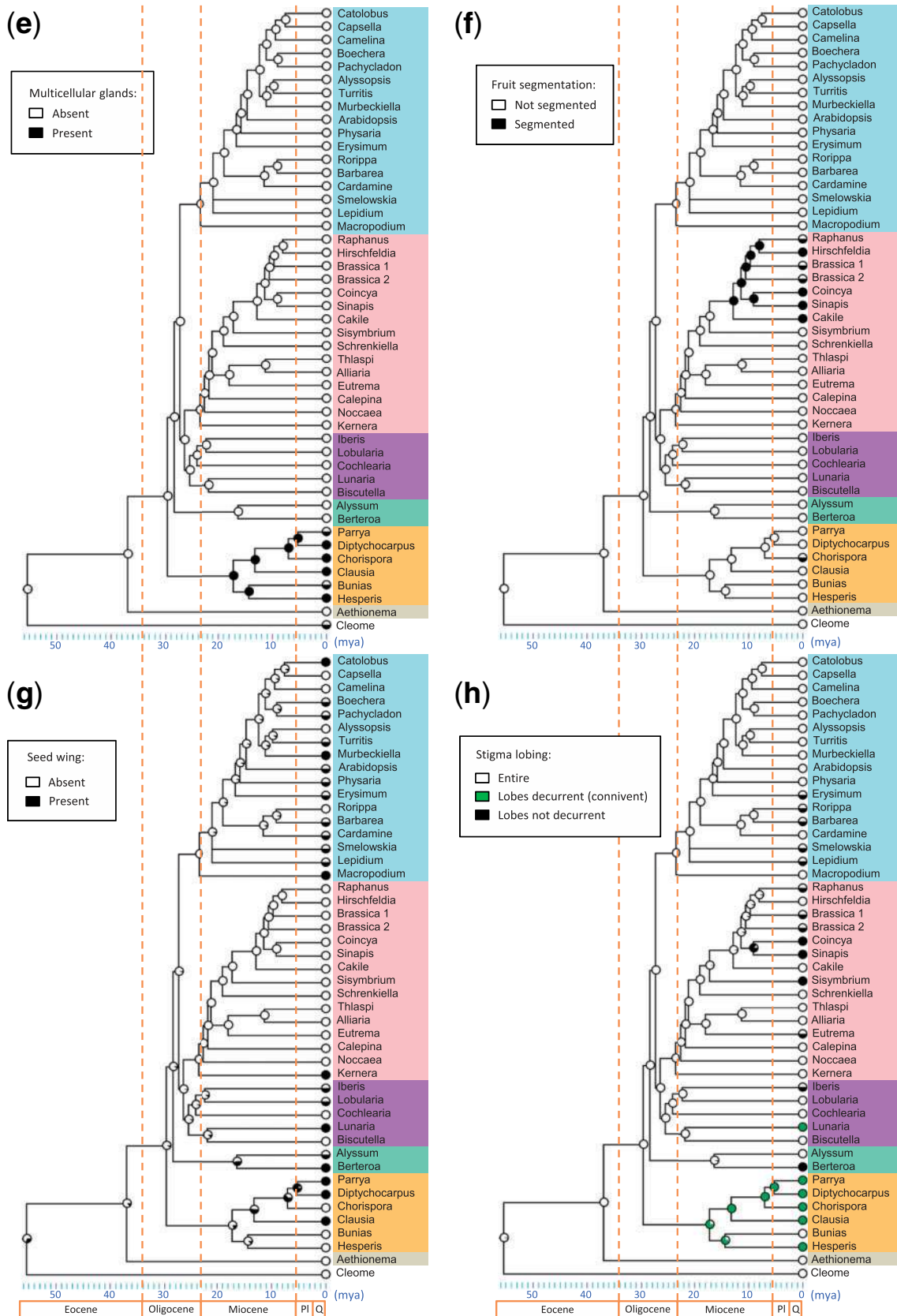


FIG. 5. Continued.

Downloaded from https://academic.oup.com/mbe/article/33/2/394/2579467 by guest on 16 August 2022

**Table 2.** Comparison of Age Estimates with Previous Results.

	This study	Hohmann et al. (2015)	Edger et al. (2015)	Beilstein et al. (2010)	Couvreur et al. (2010)	Franzke et al. (2009)
<b>Markers</b>						
	113 low-copy nuclear orthologs	Chloroplast genome	1,155 single-copy nuclear genes	<i>ndhF</i> + <i>PHYA</i>	Eight genes from nuclear, cp, and mtDNA	<i>nad4</i>
<b>Calibrations</b>						
No. of fossil calibration	18	19	4	2	4	1
<i>Dressiantha bicarpelata</i> (Brassicales)	Yes	Yes	No	No	Yes	Yes
<i>Thlaspi primaevum</i> (Brassicaceae)	No	Yes	No	No	Yes	No
<b>Node age</b>						
MRCA Cleomaceae/ Brassicaceae	54.3-55.5-56.8	58.4-59.5-60.7	ND	23.3-42.8-62.1	54.4-64.5-76.5	ND
CG Brassicaceae	36.3-37.1-37.8	41.4-42.0-42.7	32.4	16.8-31.8-45.9	45.2-54.3-64.2	24.2-37.6-49.4
CG core Brassicaceae	29.1-29.7-30.3	34.6-35.1-35.6	ND	ND	39.4-46.9-54.3	20.9-32.3-42.8

NOTE.—Values in bold are mean ages (Mya) and are bracketed by lower and upper bounds of 95% HPD interval. cp: chloroplast; mtDNA: mitochondrial DNA; ND: not determined.

wingless seeds; however, their common ancestor was presumably wingless (>0.9 proportional likelihoods for all nodes in this clade), suggesting one or more convergence events possibly during late Miocene and afterwards.

In the ancestral ovaries, stigma was likely entire, and this has been retained through much of the subsequent diversification (fig. 5h) with some changes. Particularly, ancestors of Clade E possibly had a transformation into a lobed (and decurrent) stigma during early to middle Miocene (0.76 proportional likelihoods), but afterwards *Bunias* reverted back to the ancestral entire state. Multiple convergence events also occurred in other clades. In addition, the ancestral fruits of Brassicaceae were likely dehiscent and had walls that were neither woody/corky nor thin papery (supplementary figs. S12 and S13). Such fruit characters did not change during the divergence of deep branches, but later parallel evolution likely occurred in several lineages as supported by the morphologies of extant taxa.

Generally, the majority of characters examined here are highly plastic with multiple changes and sometimes reversions across the tribes of Brassicaceae. On the contrary, fruit segmentation and multicellular glands are phylogenetically restricted to particular clades and maintained by most of their descendants. Fruit lengths, the base of cauline leaves, trichome shape, and stigma lobing exhibit a combination of convergent evolution and maintenance by vertical inheritance.

## Discussion

The phylogeny presented here is robust for most of the relationships, having withstood multiple tests and the use of a coalescent method. The results include well-resolved relationships not only among the major clades (A–F in fig. 2), but also for 29 tribes, illustrating the effectiveness of using

conserved nuclear genes for phylogenetic reconstruction at the level of major subdivisions of a relatively large family. Our results not only are in good agreement with relatively well-supported previous relationships using chloroplast genes (Beilstein et al. 2006, 2008), but also support major clades that not only expand the LI (Clade A) and LII (Clade B) groups but also define new relationships (Clades C and D). Our topology of major clades is mostly consistent to a recently reported phylogeny using transcriptome data sets (Kagale et al. 2014), which contained 23 Brassicaceae species from 19 genera in 13 tribes, with the only difference being the position of Lepidieae. The phylogenetic position of Lepidieae is also uncertain with conflicts between some of our analyses. Nevertheless, our analyses provide strong support for the separation of Lepidieae before Cardamineae from most other tribes, unlike that in the previous study, with Lepidieae diverging after Cardamineae. In the previous study, Kagale et al. (2014) used *A. thaliana* as a reference genome to identify putative orthologs in each of the other species, and concatenated the resulting 213 genes to reconstruct Brassicaceae phylogeny using ML analysis. It is possible that the difference between results of the two studies is due to the differences in methods for identifying marker genes and in taxon selection.

The positions of the four tribes remain uncertain in the context of our samplings (fig. 3). Interestingly, a complex evolution history was reported for genera of these tribes. Both *Lepidium* and *Cochlearia* have been implicated with a reticulate history (Koch et al. 1998, 1999; Lee et al. 2002; Koch 2012), and allopolyploidization events have been found for both *Lepidium* and *Pachycladon* (Lee et al. 2002; Joly et al. 2009; Mandáková, Heenan et al. 2010). For *Pachycladon*, previous studies with five single-copy nuclear genes or using *CHS* sequence revealed an allopolyploid origin (Joly et al. 2009;

Mandáková, Heenan et al. 2010). Further analyses using *CHS* and *PHYA* from more taxa provided a good resolution and showed that the two genomes of *Pachycladon* are closely related to *Crucihimalaya* and *Smelowskia*, respectively (Beilstein et al. 2010; Zhao et al. 2010). A similar conclusion can be reached when we examined the 113 single-gene trees here, which often indicated *Pachycladon* to be sister to either *Boechera* or *Smelowskia*. It is known that *Boechera* was phylogenetically near *Crucihimalaya* (Beilstein et al. 2010; Zhao et al. 2010); although *Crucihimalaya* was not included in our study, the close relationship between *Pachycladon* and *Boechera* or *Smelowskia* in our 113-single gene trees is still consistent with the previous conclusions.

We also inspected the topologies of *Cochlearia* in 113 single-gene trees, and found that a relatively large number of topologies place *Cochlearia* near members of Clade C (28 out of 35 trees with support > 70, and 40 out of all 113 trees), especially *Iberis* and/or *Lobularia*, consistent to our concatenation results (figs. 1a and 2, and supplementary fig. S7, Supplementary Material online). In addition, there are 11, 8, and 8 single-gene trees, respectively, with sisterhood of *Cochlearia* either to *Kerneria*, *Noccaea*, or *Alyssum*. These are consistent with the coalescence results of *Cochlearia* being sister to members of Clades B (supplementary fig. S9, Supplementary Material online, and few concatenation results such as fig. 1c). One possible explanation is an ancient allopolyploidization event near the origin of those early-diverging EII tribes. Resolving the phylogeny of species with a complex evolutionary history such as *Pachycladon* and *Cochlearia* requires more sequences and/or additional taxa sampling.

Besides the four tribes with conflicting results, there are 19 tribes not sampled in this study including four LI tribes, one LII tribes, and 12 EII tribes. Consistent with previous studies (Franzke et al. 2011 and references therein; Kagale et al. 2014; Hohmann et al. 2015), the bootstrap supports for monophyly of LI and LII are much stronger than those of LIII and EII. In our phylogeny, Clades A and B containing LI and LII, respectively, received maximum support, consistent with previous results (figs. 1 and 2, and supplementary figs. S4, S9, and S10, Supplementary Material online; Franzke et al. 2011 and references therein), implying that remaining tribes of these two lineages might also belong to their corresponding clades. The positions of EII tribes were poorly supported in previous studies. Our phylogeny provides increased support for several EII tribes, with placements distributed across Clades A, B, C, and D. Therefore, it is not possible to speculate on relationships of the remaining EII tribes without further investigations. Sampling of LIII tribes is complete with the taxa lacking transcriptome data sets (supplementary fig. S10, Supplementary Material online). Four of the six LIII tribes form a monophyletic Clade E, whereas Biscutelleae and Anastaticae group in Clade C with some of the EII species (fig. 2). Species of the two other LIII tribes were included using three genes by PCR amplification and group well in Clade E with most of those of the same lineage (supplementary fig. S10, Supplementary Material online). Although tribes of LIII were previously referred to as monophyletic (Franzke et al.

2009; Warwick et al. 2010), there were also results showing that Biscutelleae and Anastaticae separated from other LIII tribes (Couvreur et al. 2010; Warwick et al. 2010). However, the current sampling for genera and tribes in Clade C is limited. Thus, future studies are needed to determine the relationships of additional tribes not included here.

Previous studies about the age of Brassicaceae remain controversial possibly due to the lack of informative markers recovering a resolved phylogeny, and regarding fossils used to calibrate the ages of various nodes in the phylogeny (Koch et al. 2000, 2001; Franzke et al. 2009; Beilstein et al. 2010; Couvreur et al. 2010). The published ages of CG Brassicaceae varied widely, from 15 to 60 My; similarly, those of CG core Brassicaceae ranged from 11 to 47 My (table 2; Koch et al. 2000). Our estimation takes the advantages of a resolved phylogeny based on abundant informative nuclear sequences with extra tests to detect possible misleading signals (e.g., hidden paralogs, LB, and saturation of partitions), and that of using a total of 19 fossil constraints, including one Brassicales fossil (Gandolfo et al. 1998) and one putative Brassicaceae fossil (Becker 1961; Beilstein et al. 2010; Manchester and O'Leary 2010; Magallón et al. 2015). The effect of the latter fossil under debate was also investigated by performing estimations with or without this fossil; surprisingly the variations between the two results are limited. In contrast, nodal ages here are younger than the findings of Beilstein et al. (2010), which also used *T. primaevum* as one of the constraints for their estimation, with differences up to 17 My (without *T. primaevum* calibration) and 12 My (with *T. primaevum*) (table 2). The differences between the two studies might be due to the phylogenetic markers used; the previous study used sequences of one chloroplast gene *ndhF* and one nuclear gene *PHYA* for their estimations, whereas ours uses the supermatrix of 113 low-copy nuclear sequences.

Our result on the age of CG Brassicaceae lies within the range estimated via synonymous substitution rate in previous studies (60–30 My; Koch et al. 2000, 2001). Our estimates are even closer to those using supermatrix of multiple genes from different genomes (Couvreur et al. 2010; Edger et al. 2015; Hohmann et al. 2015). Specifically, these studies and ours found CG Brassicaceae and the tribe Aethionemeae originated around middle to late Eocene, when the Earth is warm and moist and tropical forests spread worldwide (Zachos et al. 2001), just like the environment of its sister families Cleomaceae and Capparaceae primarily distributed in tropical regions (Pennington et al. 2009; Angiosperm Phylogeny Website <http://www.mobot.org/MOBOT/research/APweb/>). Our results show an early radiation of major clades near or shortly after the Eocene–Oligocene boundary, followed by nested divergences of several tribes during a short period in Oligocene. These also corresponded to a higher diversification rate during 32 to 22 Mya as indicated by lineage-through-time plot presented in Couvreur et al. (2010). Interestingly, there was an extinction event during Eocene–Oligocene boundary called *Grande Coupure* in Europe (Hooker et al. 2004), a short period of abrupt cooling icehouse climate (Zachos et al. 2001). This might explain the long branch leading to the core Brassicaceae after

divergence of Aethionemeae; this event could cause the extinction of many early species contemporary to the ancestor of core Brassicaceae. During Oligocene, there was a global expansion of grassland, with an increase in the deciduous and drought-adaptive flora in Europe (Jacobs et al. 1999; Morley 2000). Also, the African tectonic plate continued to push north into Eurasia during the Oligocene, resulting in the formation of mountains in Europe (Pfiffner 1992; Schmid et al. 1996). North America, East and South Asia also experienced the formation of new mountains (Stanley 2004). The combined effect of paleoclimate (increasing aridity and seasonal change) and tectonic changes might have resulted in increasingly diversified habitats that promoted (or maintained) the nested radiations during Oligocene.

A resolved and dated phylogeny also allowed us to reconstruct ancestral morphological characters of Brassicaceae here, with most characters showing no obvious clade-specific patterns and providing clear evidence of multiple convergence events across the family. Although our samplings are incomplete at the levels of tribes and genera, some interesting phenomena could nevertheless be found according to our current topology. We observe that tribes (specifically the genera sampled here) in Clades B and E could each be divided into two groups according to the occurrences of distinct characters presented on their extant species. In Clade B, one group included the tribes of LII and *Schrenkiella*, and the other with the basal EII tribes. The second group contains more taxa with silicles, sessile cauline leaves, and entire stigma, which are also traced to be ancestral with 0.77, 0.94, and 0.99 proportional likelihoods, respectively (fig. 5a, c, and h), whereas the taxa in the first group tend to have siliques, petiolate cauline leaves, and lobed stigma. Additionally, an angustiseptate or indehiscent fruit is present only in the second group (fig. 5b and supplementary fig. S12, Supplementary Material online).

There are two monophyletic groups in Clade E, the first contains Chorisporeae (including *Chorispora*, *Diptychocarpus*, *Parrya*) and Dontostemoneae (only *Clausia* is sampled here), and the second includes Buniadeae (*Bunias*) and Hesperideae (*Hesperis*). Most species of the first group have simple trichomes and winged seeds, whereas taxa in the second clade possess branch trichomes and generally wingless seeds (fig. 5d and g). Characters of these two subclades are also estimated as ancestral for the corresponding nodes with proportional likelihoods ranging from 0.98 to 0.52. It is possible that dramatic changes in the character states in Clades B and E might have occurred during late Oligocene to early Miocene or later.

Trichome morphology has been studied extensively as a model for control of plant cell shape (Hülkamp 2000, 2004), with insights into possible regulators for this character. A series of *A. thaliana* genes encode transcriptional regulators, such as *ZWICHEL* (*ZWI*), *STICHEL* (*STI*), and *ANGUSTIFOLIA* (*AN*). Also, *GLABRA3* (*GL3*) and *ENHANCER OF GL3* (*EGL3*) code for basic helix-loop-helix proteins, whereas *TRIPTYCHON* (*TRY*), *CAPRICE* (*CPC*), and *CAPRICE TRIPTYCHON 1* (*ETC1*) act as negative regulators (Hülkamp 2000, 2004; Schwab et al. 2000). Some of these

genes are involved in parallel, partially redundant pathways. In *A. thaliana*, double mutants with any two of the *zwi*, *sti*, or *an* mutations produce exclusively simple trichomes (Hülkamp 2000, 2004). Similarly, the *gl3 egl3* mutant develops glabrous leaves, but single mutant still has trichomes. It is worth noting that trichomes of the outgroup species *Cleome* are exclusively simple, whereas species of the basalmost *Aethionema* are glabrous, suggesting that evolution of branch trichomes occurred after the split of *Aethionema* from the ancestor of core Brassicaceae (Henry et al. 2006; Schranz and Mitchell-Olds 2006).

Trichome morphology has also been shown to correlate with phylogeny and holds promise in the delimitation of monophyletic groups (Al-Shehbaz et al. 1999; Beilstein et al. 2006, 2008). Our investigations revealed that dendritic, malpighiaceae, stellate, and sessile trichomes likely have evolved independently several times. Beilstein et al. (2006) suggested that stellate trichomes might have arisen from a single evolutionary origin based on the observation on *Physaria* and *Alyssum* using a phylogeny without a basal resolution. Our phylogeny indicates that the two genera belong to distant clades (Clades A and D; fig. 2), and thus stellate trichomes probably likely evolved at least twice. Interestingly, species of *Alyssum* are found in Mediterranean, and those of *Physaria* are found in the southwestern United States adjacent to Mexico and southern South America (Rollins and Banerjee 1975, 1976, 1979), both distributed in arid habitats. Therefore, morphologically complex stellate trichomes might be adaptive for arid habitats. In addition, it should be noted that both *Schrenkiella parvula* and *E. salsugineum*, which is sister to Thlaspidaceae, are adapted to high-salt habitats and were previously thought to be members of the same genus, *Thellungiella*. The positions of these two species in two separate clades suggest that their adaptation to growth in salty soils is likely independent.

The well-supported topology here provides an important foundation for future studies, including investigations on speciation–extinction rates, biogeography, genome duplication, phenotypic and gene evolution, and many more. Here we have shown that phylogenetically informative orthologous nuclear genes can be identified from large data sets and used to resolve the deep relationships of Brassicaceae. Further investigation with many more taxa in this important family is now both feasible and worthwhile.

## Materials and Methods

### Taxon Sampling and Transcriptomics

Information for the taxa included in this study is listed in supplementary table S2, Supplementary Material online, according to the classification proposed by Al-Shehbaz (2012). Total RNA was extracted from flower buds or leaves using the ZR Plant RNA MiniPrep kit (Zymo Research, Orange, CA). Paired-end reads of 2 × 100 were generated by using Illumina HiSeq2000. Short reads were assembled into longer contigs de novo using Trinity (Grabherr et al. 2011) with default parameters. TGICLv2.1 (Pertea et al. 2003) was used with the parameter being -p 0.98, -l 40, -v to obtain longer

cDNA sequence. Whole-genome data sets and the Sequence Read Archive data of Brassicaceae species were retrieved from Phytozome (<http://phytozome.jgi.doe.gov/pz/portal.html>), UCSC Genome Browser (<http://mustang.biol.mcgill.ca:8885>), and GenBank (<http://www.ncbi.nlm.nih.gov/sra/>). Transcriptome data sets of 13 other species were generated by the research group of one of the authors (Pires JC, unpublished data; [supplementary table S2, Supplementary Material online](#)).

For amplification of sequences from additional species, total genomic DNA was isolated from ~100 mg dried leaf tissue in a CTAB extraction method (Doyle JJ and Doyle JL 1987). Then, 1–3 gene sequences of the additional 22 species without a transcriptome data or whole-genome sequences were obtained by PCR amplification. Primers of *At3g54630* and *At2g47990* were according to Ding et al. (2012), and those of *At5g20600* were designed according to the regions (>20 bp) highly conserved in *A. thaliana*, *A. lyrata*, *B. rapa*, *Ca. rubella*, *E. salsugineum*, and *Sc. parvula* (Sun 2014). Sequences of primers are listed in [supplementary table S9, Supplementary Material online](#).

### Selection of Putative Orthologous Genes

Phylogenetic markers used in this study were obtained using three different approaches. The first group of genes followed our previous studies (Zhang et al. 2012; Zeng et al. 2014; unpublished data) that identified, respectively, 59 and 47 low-copy orthologous nuclear genes conserved across angiosperms, using HaMStR (Ebersberger et al. 2009) and OrthoMCL v1.4 (Li et al. 2003) from the sequenced genomes of a total of 9 angiosperms (*Z. mays*, *Sor. bicolor*, *O. sativa*, *G. max*, *Me. truncatula*, *Po. trichocarpa*, *V. vinifera*, *A. thaliana*, and *Sol. lycopersicum*). These were shown to be effective in resolving deep relationships among major angiosperm lineages. The homologs of these 106 genes were then retrieved from 13 representative Brassicaceae species ([supplementary fig. S1b, Supplementary Material online](#)) and used for single-gene phylogenies reconstruction using RAxML (Stamatakis 2006). The rationale for this preliminary selection is due to the recurrent polyploidization events or gene/genome duplications in Brassicaceae as described in the Results section. In a single-gene tree, when species belonging to the same lineage are grouped together, but not requiring specific relationships among and within lineages, this gene was retained as a putative ortholog. This resulted in 28 genes as the first gene group for further phylogenetic analyses.

The second group of genes was selected from two databases. One contains 4,180 orthologous nuclear gene groups (OGs) shared by nine angiosperm species with sequenced genomes (*Z. mays*, *Sor. bicolor*, *O. sativa*, *G. max*, *Me. truncatula*, *Po. trichocarpa*, *V. vinifera*, *A. thaliana*, and *Sol. lycopersicum*) identified by HaMStR (Ebersberger et al. 2009; Deep Metazoan Phylogeny, <http://www.deep-phylogeny.org/hamster/>). The other data set is 1,989 low-copy OGs found using 7 angiosperm species with sequenced genome (*A. thaliana*, *Po. trichocarpa*, *Pr. persica*, *V. vinifera*, *Mi. guttatus*, *O. sativa*, and *Sor. bicolor*) identified by OrthoMCL v1.4 (Li et al.

2003) with default parameters (Zhang et al. 2012); in that study clusters with 7–9 genes and at least one in each of the 7 angiosperms were identified as low copy. We selected the OGs shared by these two databases as the second group of markers, representing conserved low-copy OGs across angiosperms. To avoid repeated selection of the same genes, those in the first group were excluded. Among the resulting 830 OGs, only 289 OGs were retained for further analyses based on the coverage of more than 80% of the 13 representative species ([supplementary fig. S1b and table S2, Supplementary Material online](#)) and encoding proteins longer than 300 amino acids. Single-gene trees of the 289 OGs from the 13 Brassicaceae species were then examined for groupings of taxa to one of the three lineages as described above, and 45 genes remained as the second group of markers.

Following a similar procedure used for the second group, the third group was filtered from those overlapping between 4,180 OGs identified by HaMStR as described above, and 6,552 OGs found as single-copy orthologous genes (i.e., one gene in each species) in 5 Brassicaceae species with sequenced genomes (*A. lyrata*, *A. thaliana*, *B. rapa*, *Ca. rubella*, and *E. salsugineum*) identified using OrthoMCL v1.4 (Li et al. 2003) with default parameters. OGs from this method are conserved across angiosperm yet especially maintained as single copy in Brassicaceae. Then, 316 genes that were found in more than 80% of the same 13 species as above and coding for proteins with more than 500 amino acids were chosen for further analyses. Single-gene trees of the 316 OGs were then investigated for the grouping of the 13 species according to the three lineages. Finally, 40 genes remained, forming the third group of markers. It is noted that *Biscutella cichoriifolia* contains only seven genes of this group and was therefore excluded in this preliminary analysis. The three groups together contain 113 genes; for their characteristics see [supplementary table S3, Supplementary Material online](#).

### Phylogenetic Analyses

OGs of 113 genes from 56 species (55 Brassicaceae species and 1 outgroup species) were obtained by HaMStR (Ebersberger et al. 2009) with *E* values of less than  $e^{-20}$ . Nucleotide sequences were aligned using MUSCLE v3.8.31 (Edgar 2004) with default parameters, and the sequences of low quality were removed. The poorly aligned regions were further trimmed by using the trimAl v1.2 software (Capella-Gutiérrez et al. 2009) with default settings. Nucleotide sequences were concatenated by SeaView (Gouy et al. 2010). For genes with more than one copy from a single species, the one with the shortest branch was used for further analyses.

The best-fit nucleotide substitution model for our 113 gene sequences was GTRGAMMA according to the results of Modeltest (Posada and Crandall 1998). BI trees were produced by MrBayes 3.2.1 (Ronquist and Huelsenbeck 2003) with the setting of 5,000,000 generations and stopval = 0.01, under GTRGAMMA model with one cold and three incrementally heated Markov Chain Monte Carlo (MCMC) run simultaneously (Ronquist and Huelsenbeck 2003). The MCMC convergence in BI was monitored by AWTY

(Nylander et al. 2008; <http://ceb.csit.fsu.edu/awty>). Trees were sampled every 100 generations. The first 25% of the trees were discarded as burn-in. The remaining trees were used for generating the consensus tree. ML trees were generated using RAxML 7.0.4 (Stamatakis 2006). For ML analyses, different general time reversible models were performed with 113 gene concatenation, and resulting topologies and bootstrap scores are found to be equivocal (supplementary fig. S8, Supplementary Material online). So, for ML analyses with other sets of genes, we used GTRCAT instead of GTRGAMMA for greater computational efficiency. The nucleotide sequence alignments of each of the 113 orthologs of 56 species have been deposited in the TreeBASE under the accession code 18202 (<http://treebase.org/treebase-web/home.html>).

### Assessing the Confidence of Tree Selection

To test for statistic support of the alternative relationships among the 5 major clades of core Brassicaceae (A–E) reported in this study, 48 topologies were tested against our best ML tree (supplementary table S5, Supplementary Material online). For each topology, per site log-likelihoods were computed in RAxML with the GTRGAMMA model, because the GTRCAT model cannot be used to calculate per site log-likelihoods and GTRGAMMA is the best-fit model for our 113 gene sequences as mentioned above. AU test, approximate Bayesian posterior probability test, bootstrap probability test, Kishino–Hasegawa test, weighted Kishino–Hasegawa test, Shimodaira–Hasegawa test, and weighted Shimodaira–Hasegawa test were then performed using CONSEL v0.1j (Shimodaira and Hasegawa 2001).

### Detection of Sequence Biases

TreSpEx, recently developed for detection of misleading signals, was used to examine possible LB attraction and saturation of partitions, as detailed in Struck (2014). In brief, alignments of the 113 gene sequences were analyzed by TreSpEx. The standard deviation and the average of upper quartile of LB scores were calculated for each of the 113 genes, and then density plots were generated using the R package. Genes with high values (distributed in the right shoulders of the density plots) in either of the two parameters were considered as having LBs, and then ML and BI reconstructions were performed excluding either group of the genes. Also, a heat map of taxon-specific LB scores for each of the 113 genes was generated with hierarchical clustering using R.

TreSpEx also detects the degree of saturation of each gene as determined by the slopes or  $R^2$  values of linear regressions of patristic distances against uncorrected distances  $p$ . Density plots of the slope or  $R^2$  values were generated with the use of R. These two indices are negatively correlated with the degree of saturation, so genes with either a low slope or  $R^2$  value, which indicates a high saturation degree, were removed before subsequent ML and BI analyses.

### Coalescent Estimates of Phylogenetic Tree

Astral.4.4.4 (Mirarab et al. 2014) was used to perform multilocus bootstrapping. It identifies the tree that maximizes the number of induced quartet trees in the set of gene trees that are shared by the species tree. The 113 single-gene ML trees were reconstructed by RAxML 7.0.4 under GTRCAT model as mentioned above. Then, 100 bootstrap replicates of each of the 113 genes were subjected to the coalescent analysis.

### Analysis of Morphological Characters

Characters analyzed in this study include the following: Bbase of cauline leaves, trichome shape, multicellular glands, inflorescences flower arrangement, petal presence, petal shape, stamens position, stigma lobing, fruit length–width ratio, fruit flattening, fruit segmentation, fruit dehiscence, fruit wall, gynophore in fruit, septum in mature fruit, seed wing, seed mucilage, habit, and duration. Information of the generic morphology of the above 19 characters was obtained from the Brassicaceae of the World website (<http://flora.huh.harvard.edu/Brassicaceae/>). Matrix of morphological characters for ancestral character reconstruction using genera as terminals is presented in supplementary table S8, Supplementary Material online; for coding see supplementary table S7, Supplementary Material online. The ancestral character was analyzed using Mesquite (Maddison WP and Maddison DR 2004; <http://mesquiteproject.org>) with the likelihood model in the context of our dated topology without fossil *T. primaevum* (fig. 4b) to detect potential synapomorphies.

### Fossil Calibrations and Divergence Time Estimation

Our estimation used a total of 19 fossil calibrations; assignments and ages of the fossils are according to Smith et al. (2010) and Magallón et al. (2015) (supplementary table S6, Supplementary Material online). There are two calibrations implemented here at the branches leading up to and within Brassicaceae respectively. One is fossil *D. bicarpelata* (89.3 Mya; Gandolfo et al. 1998), which could be a stem representative or crown member of Brassicales (Magallón et al. 2015), and is used to calibrate the stem node Brassicales here (with Malvaceae species as the sister lineage here). Age of the putative Brassicaceae fossil *T. primaevum* (Becker 1961; Manchester and O’Leary 2010) was implemented here conservatively, corresponding to the upper boundary of Oligocene (23.03 Mya) to constrain the minimum age of the divergence of *Thlaspi* and *Alliaria* (Beilstein et al. 2010; Magallón et al. 2015). Considering the uncertainty on the assignment of *T. primaevum*, age estimations were performed using two sets of calibrations with the only difference being whether this fossil was used or not. To allow the application of the 18 fossil calibrations outside the Brassicaceae CG, a total of 55 outgroup species (supplementary table S2, Supplementary Material online) were included by using their homologs of the marker genes. All the fossil calibrations were implemented as minimum constraint in our PL analysis, except that the tricolpate pollen grains (about 125 Mya; Hickey and Doyle 1977) were used to fix the age of CG



eudicots, which is widely assumed to be unlikely older than the earliest tricolpate pollen fossil, a synapomorphy of eudicots and preserved in most of its early diverging lineages and never found in the sediments lower than the boundary of Barremian-Aptian (Forest 2009; Magallón and Castillo 2009).

We used PL method implemented in r8s v1.7.1 (Sanderson 2002) for the estimates, as constant substitution rate across the phylogenetic tree was rejected ( $P < 0.01$ ) for all cases by likely ratio test conducted in PAUP 4.0 beta10 (Swofford 2002). The ML tree reconstructed using 113 genes with branch length generated by RAxML was used as the input tree. When presenting the results, the most distant outgroup species *O. sativa* was pruned as required by r8s. Cross-validation was tested to determine the best smoothing value for our data set. After testing a range of smoothing parameters from 0.01 to 100,000 (cvstart = -2; cvinc = 0.5; cvnum = 15), the smoothing parameter was determined as 0.1. The low smoothing value also indicates a large deviation from the strict molecular clock hypothesis. One hundred BS replicates with branch lengths were also generated using RAxML for calculating the confidence age intervals. Mean and 95% HPD ages of nodes of interest were estimated and summarized across the 100 BS trees (table 1). The algorithm of TN was used and all other parameters were set as default in all above PL analyses.

## Supplementary Material

Supplementary figures S1 to S22, tables S1–S9, and text are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We thank Dr Wenju Zhang (Fudan University) for the assistance in material collection, Dr Fan Lu (Fudan University) and Dr Guiling Zhou (Xinjiang Agricultural University) for specimen identification. We also thank Dr Ji Qi (Fudan University) for help in transcriptome sequence assembly. This work was supported by grants from the National Natural Science Foundation of China (grant number 91131007); the China Postdoctoral Science Foundation (grant number 2014M551316); funds from the State Key Laboratory of Genetic Engineering at Fudan University; USA National Science Foundation grant (DEB 1146603 to J.C.P.); NSF Plant Genome postdoctoral grant (NSF-NPGI 1202793 to P.P.E.); and German Research Foundation (DFG 2302-13/1 to M.A.K.).

## References

- Al-Shehbaz IA. 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon* 61:931–954.
- Al-Shehbaz IA, Beilstein MA, Kellogg EA. 2006. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Syst Evol* 259:89–120.
- Al-Shehbaz IA, German DA, Mummenhoff K, Moazzeni H. 2014. Systematics, tribal placements, and synopses of the *Malcolmia* S.L. segregates (Brassicaceae). *Harv Pap Bot* 19:53–71.
- Al-Shehbaz IA, O’Kane SL, Price RA. 1999. Generic placement of species excluded from *Arabidopsis* (Brassicaceae). *Novon* 9:296–307.
- Amasino R. 2009. Floral induction and monocarpic versus polycarpic life histories. *Genome Biol* 10:228–228.
- Anderson JT, Wagner MR, Rushworth CA, Prasad KVSK, Mitchell-Olds T. 2014. The evolution of quantitative traits in complex environments. *Heredity* 112:4–12.
- Bailey CD, Koch MA, Mayer M, Mummenhoff K, O’Kane SL, Warwick SI, Windham MD, Al-Shehbaz IA. 2006. Toward a global phylogeny of the Brassicaceae. *Mol Biol Evol* 23:2142–2160.
- Becker HF. 1961. Oligocene plants from the upper Ruby River Basin, southwestern Montana. *Geol Soc Am Mem* 82:1–127.
- Beilstein MA, Al-Shehbaz IA, Kellogg EA. 2006. Brassicaceae phylogeny and trichome evolution. *Am J Bot* 93:607–619.
- Beilstein MA, Al-Shehbaz IA, Mathews S, Kellogg EA. 2008. Brassicaceae phylogeny inferred from phytochrome A and *ndhF* sequence data: tribes and trichomes revisited. *Am J Bot* 95:1307–1327.
- Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S. 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 107:18724–18728.
- 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.
- Claire-Lise M, Nathalie V. 2012. The use of the model species *Arabidopsis halleri* towards phytoextraction of cadmium polluted soils. *N Biotechnol* 30:9–14.
- Couvreur TLP, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA, Mummenhoff K. 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol Biol Evol* 27:55–71.
- Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Yun DJ, Bressan RA, Zhu JK, Bohnert HJ, et al. 2011. The genome of the extremophile crucifer *Thellungiella parvula*. *Nat Genet* 43:913–918.
- Ding M, Zeng L, Ma H, Zhang N. 2012. The use of low-copy nuclear genes for reconstructing the phylogeny of low-level taxonomic hierarchies: evidence from Brassicaceae. *Plant Diversity Resour* 34:211–221.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15.
- Ebersberger I, Strauss S, von Haeseler A. 2009. HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evol Biol* 9:157–157.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
- Edger PP, Heide-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 U S A* 112:8362–8366.
- Flavell R. 2009. Role of model plant species. In: Gustafson JP, Langridge P, Somers DJ, editors. *Plant genomics*. New York: Humana Press. p. 1–18.
- Forest F. 2009. Calibrating the tree of life: fossils, molecules and evolutionary timescales. *Ann Bot* 104:789–794.
- Franzke A, German D, Al-Shehbaz IA, Mummenhoff K. 2009. *Arabidopsis* family ties: molecular phylogeny and age estimates in Brassicaceae. *Taxon* 58:425–437.
- Franzke A, Koch MA, Couvreur TLP, Lysak MA, Mummenhoff K. 2010. On the age of the mustard family (Brassicaceae). *Nature* 467:755.
- Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K. 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci* 16:108–116.
- Gale MD, Devos KM. 1998. Plant comparative genetics after 10 years. *Science* 282:656–659.
- Gandolfo M, Nixon K, Crepet W. 1998. A new fossil flower from the Turonian of New Jersey: *Dressiantha bicarpellata* gen. et sp. nov. (Capparales). *Am J Bot* 85:964.
- German DA, Friesen N, Neuffer B, Al-Shehbaz IA, Hurka H. 2009. Contribution to ITS phylogeny of the Brassicaceae, with special reference to some Asian taxa. *Plant Syst Evol* 283:33–56.
- German DA, Friesen NW. 2014. *Shehbazia* (Shehbazieae, Cruciferae), a new monotypic genus and tribe of hybrid origin from Tibet. *Turczaninowia* 17:17–23.

- Gong Q, Li P, Ma S, Indu Rupassara S, Bohnert HJ. 2005. Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. *Plant J*. 44:826–839.
- 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.
- Halimaa P, Blande D, Aarts MGM, Tuomainen M, Tervahauta A, Kärenlampi S. 2014. Comparative transcriptome analysis of the metal hyperaccumulator *Noccaea caerulescens*. *Front Plant Sci*. 5:213.
- Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM, et al. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat Genet*. 45:891–898.
- 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.
- Hohmann N, Wolf EM, Lysak MA, Koch MA. 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell*. doi:10.1105/tpc.15.00482.
- 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.
- Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet*. 43:476–481.
- Hülkamp M. 2000. Cell morphogenesis: how plants split hairs. *Curr Biol*. 10:R308–R310.
- Hülkamp M. 2004. Plant trichomes: a model for cell differentiation. *Nat Rev Mol Cell Biol*. 5:471–480.
- Jacobs BF, Kingston JD, Jacobs LL. 1999. The origin of grass-dominated ecosystems. *Ann Mo Bot Gard*. 86:590–643.
- Joly S, Heenan PB, Lockhart PJ. 2009. A Pleistocene inter-tribal allopolyploidization event precedes the species radiation of *Pachycladon* (Brassicaceae) in New Zealand. *Mol Phylogenet Evol*. 51:365–372.
- 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.
- Karl R, Koch MA. 2013. A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. *Ann Bot*. 112:983–1001.
- Khosravi AR, Mohsenzadeh S, Mummenhoff K. 2009. Phylogenetic relationships of Old World Brassicaceae from Iran based on nuclear ribosomal DNA sequences. *Biochem Syst Ecol*. 37:106–115.
- Kiefer M, Schmickl R, German DA, Mandáková T, Lysak MA, Al-Shehbaz IA, Franzke A, Mummenhoff K, Stamatakis A, Koch MA. 2014. BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. *Plant Cell Physiol*. 55:e3.
- Kim J, Lee J, Choi JP, Park I, Yang K, Kim MK, Lee YH, Nou IS, Kim DS, Min SR, et al. 2014. Functional innovations of three chronological mesohexaploid *Brassica rapa* genomes. *BMC Genomics* 15:606.
- Koch M, Haubold B, Mitchell-Olds T. 2001. Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. *Am J Bot*. 88:534–544.
- Koch M, Huthmann M, Hurka H. 1998. Isozymes, speciation and evolution in the polyploid complex *Cochlearia* L. (Brassicaceae). *Bot Acta*. 111:411–425.
- Koch M, Mummenhoff K, Hurka H. 1999. Molecular phylogenetics of *Cochlearia* L. (Brassicaceae) and allied genera based on nuclear ribosomal ITS DNA sequence analysis contradict traditional concepts of their evolutionary relationship. *Plant Syst Evol*. 216:207–230.
- Koch MA. 2012. Mid-Miocene divergence of *Lonopsisidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). *Taxon* 61:76–92.
- Koch MA, Dobeš C, Kiefer C, Schmickl R, Klimeš L, Lysak MA. 2007. Supernetwork identifies multiple events of plastid *trnF*(GAA) pseudogene evolution in the Brassicaceae. *Mol Biol Evol*. 24:63–73.
- Koch MA, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol*. 17:1483–1498.
- Koch MA, Kiefer M, German DA, Al-Shehbaz IA, Franzke A, Mummenhoff K, Schmickl R. 2012. BrassiBase: tools and biological resources to study characters and traits in the Brassicaceae-version 1.1. *Taxon* 61:1001–1009.
- Lee JY, Mummenhoff K, Bowman JL. 2002. Allopolyploidization and evolution of species with reduced floral structures in *Lepidium* L. (Brassicaceae). *Proc Natl Acad Sci U S A*. 99:16835–16840.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 13:2178–2189.
- Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, et al. 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun*. 5:3930.
- Lysak MA, Koch MA, Pecinka A, Schubert I. 2005. Chromosome triplication found across the tribe Brassicaceae. *Genome Res*. 15:516–525.
- Maddison WP, Maddison DR. 2004. Mesquite: a modular system for evolutionary analysis. Available from: <http://mesquiteproject.org>.
- 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.
- Manchester SR, O’Leary EL. 2010. Phylogenetic distribution and identification of fin-winged fruits. *Bot Rev*. 76:1–82.
- Mandáková T, Heenan PB, Lysak MA. 2010. Island species radiation and karyotypic stasis in *Pachycladon* allopolyploids. *BMC Evol Biol*. 10:367.
- Mandáková T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA. 2010. Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell*. 22:2277–2290.
- Mandáková T, Lysak MA. 2008. Chromosomal phylogeny and karyotype evolution in  $x=7$  crucifer species (Brassicaceae). *Plant Cell*. 20:2559–2570.
- Mathur V, Tytgat TOG, Hordijk CA, Harhangi HR, Jansen JJ, Reddy AS, Harvey JA, Vet LEM, Van Dam NM. 2013. An ecogenomic analysis of herbivore-induced plant volatiles in *Brassica juncea*. *Mol Ecol*. 22:6179–6196.
- Meinert DW, Cherry JM, Dean C, Rounsley SD, Koornneef M. 1998. *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 282:662–682.
- Meyerowitz EM, Somerville CR. 1994. *Arabidopsis*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.
- 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.
- Mitchell-Olds T, Claus MJ. 2002. Plant evolutionary genomics. *Curr Opin Plant Biol*. 5:74–79.
- Morley RJ. 2000. Origin and evolution of tropical rain forests. New York: John Wiley & Sons.
- Movahedi S, Van de Peer Y, Vandepoele K. 2011. Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in *Arabidopsis* and rice. *Plant Physiol*. 156:1316–1330.
- Mummenhoff K, Franzke A, Koch M. 1997. Molecular data reveal convergence in fruit characters used in the classification of *Thlaspi* s. l. (Brassicaceae). *Bot J Linn Soc*. 125:183–199.

- Mummenhoff K, Polster A, Mühlhausen A, Theißen G. 2009. *Lepidium* as a model system for studying the evolution of fruit development in Brassicaceae. *J Exp Bot.* 60:1503–1513.
- Nardmann J, Werr W. 2007. The evolution of plant regulatory networks: what *Arabidopsis* cannot say for itself. *Curr Opin Plant Biol.* 10:653–659.
- Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* 24:581–583.
- Paterson AH, Freeling M, Tang H, Wang X. 2010. Insights from the comparison of plant genome sequences. *Annu Rev Plant Biol.* 61:349–372.
- Pennington RT, Lavin M, Oliveira-Filho A. 2009. Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. *Annu Rev Ecol Evol Syst.* 40:437–457.
- Perlea 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.
- Pfiffner OA. 1992. Alpine orogeny. In: Blundell D, Freeman R, Mueller S, editors. *A continent revealed: the European Geotraverse*. Cambridge: European Science Foundation. p. 180–190.
- Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Rollins RC, Banerjee UC. 1975. Atlas of the trichomes of *Lesquerella* (Cruciferae). *Publ Bussey Inst Harvard Univ.* 1975:3–48.
- Rollins RC, Banerjee UC. 1979. Trichome patterns in *Physaria*. *Publ Bussey Inst Harvard Univ.* 1979:65–77.
- Rollins RC, Banerjee UC. 1976. Trichomes in studies of the Cruciferae. In: Vaughn JG, McLeod AJ, Jones BMG, editors. *The biology and chemistry of the Cruciferae*. New York: Academic Press. p. 145–166.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Rushworth CA, Song BH, Lee CR, Mitchell-Olds T. 2011. *Boechera*, a model system for ecological genomics. *Mol Ecol.* 20:4843–4857.
- 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.
- Schmid SM, Pfiffner OA, Froitzheim N, Schönborn G, Kissling E. 1996. Geophysical-geological transect and tectonic evolution of the Swiss-Italian Alps. *Tectonics* 15:1036–1064.
- Schranz ME, Mitchell-Olds T. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell.* 18:1152–1165.
- Schwab B, Folkers U, Ilgenfritz H, Hülskamp M. 2000. Trichome morphogenesis in *Arabidopsis*. *Philos Trans R Soc B Biol Sci.* 355:879–883.
- Seo TK. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol Biol Evol.* 25:960–971.
- 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.
- Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
- Shrestha R, Gómez-Ariza J, Brambilla V, Fornara F. 2014. Molecular control of seasonal flowering in rice, *Arabidopsis* and temperate cereals. *Ann Bot.* 114:1445–1458.
- Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 99:13627–13632.
- Slotte T, Hazzouri KM, Agren JA, Koenig D, Maumus F, Guo YL, Steige K, Platts AE, Escobar JS, Newman LK, et al. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet.* 45:831–835.
- Smith SA, Beaulieu JM, Donoghue MJ. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc Natl Acad Sci U S A.* 107:5897–5902.
- Somerville CR, Meyerowitz EM, Dangi J, Stitt M. 2002. *The Arabidopsis book*. Rockville (MD): American Society of Plant Biologists.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stanley SM. 2004. *Earth system history*. New York: W. H. Freeman and Company.
- Struck TH. 2014. TreSpEx—detection of misleading signal in phylogenetic reconstructions based on tree information. *Evol Bioinform Online.* 10:51–67.
- Sun R. 2014. *Resolving deep relationships of Brassicaceae using phylogenomics and low-copy nuclear genes*. Shanghai (China): Fudan University.
- Swofford DL. 2002. PAUP\*. Phylogenetic analysis using parsimony (\*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
- Vekemans X, Poux C, Goubet PM, Castric V. 2014. The evolution of selfing from outcrossing ancestors in Brassicaceae: what have we learned from variation at the *S*-locus? *J Evol Biol.* 27:1372–1385.
- Verbruggen N, Juraniec M, Bialiardini C, Meyer CL. 2013. Tolerance to cadmium in plants: the special case of hyperaccumulators. *BioMetals* 26:633–638.
- Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290:2114–2117.
- Warwick SI, Mummenhoff K, Sauder CA, Koch MA, Al-Shehbaz IA. 2010. Closing the gaps: phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. *Plant Syst Evol.* 285:209–232.
- Wen J, Xiong Z, Nie ZL, Mao L, Zhu Y, Kan XZ, Ickert-Bond SM, Gerrath J, Zimmer EA, Fang XD. 2013. Transcriptome sequences resolve deep relationships of the grape family. *PLoS One* 8:e74394.
- 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 U S A.* 111:E4859–E4868.
- Wing SL. 1987. Eocene and Oligocene floras and vegetation of the Rocky Mountains. *Ann Mo Bot Gard.* 74:748–784.
- Wu HJ, Zhang Z, Wang JY, Oh DH, Dassanayake M, Liu B, Huang Q, Sun HX, Xia R, Wu Y, et al. 2012. Insights into salt tolerance from the genome of *Thellungiella salsuginea*. *Proc Natl Acad Sci U S A.* 109:12219–12224.
- Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GKS, 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.
- 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 U S A.* 110:5074–5079.
- Zhao B, Liu L, Tan D, Wang J. 2010. Analysis of phylogenetic relationships of Brassicaceae species based on *Chs* sequences. *Biochem Syst Ecol.* 38:731–739.
- Zimmer EA, Wen J. 2013. Reprint of: Using nuclear gene data for plant phylogenetics: progress and prospects. *Mol Phylogeny Evol.* 66:539–550.