Molecular Phylogenetics, Temporal Diversification, and Principles of Evolution in the Mustard Family (Brassicaceae)

Thomas L. P. Couvreur,†,[1,2,3] Andreas Franzke,†,[4] Ihsan A. Al-Shehbaz,[5] Freek T. Bakker,[2] Marcus A. Koch,*,[4,6] and Klaus Mummenhoff[1]

[1]FB Biologie/Chemie, Botanik, Universität Osnabrück, Osnabrück, Germany

[2]Nationaal Herbarium Nederland–Wageningen Branch, Biosystematics Group, Wageningen University, Wageningen, The Netherlands

[3]The New York Botanical Garden, Bronx, NY

[4]Botanical Garden and Herbarium (HEID), Heidelberg Institute of Plant Sciences, Heidelberg, Germany

[5]Heidelberg Institute of Plant Sciences, Biodiversity and Plant Systematics, University of Heidelberg, Heidelberg, Germany

[6]Missouri Botanical Garden, St Louis, MO

***Corresponding author:** E-mail: mkoch@hip.uni-hd.de.

†T.L.P.C. and A. F. contributed equally to the work.

Associate editor: Peter Lockhart

Research article

Abstract

Brassicaceae is an important family at both the agronomic and scientific level. The family not only inlcudes several model species, but it is also becoming an evolutionary model at the family level. However, resolving the phylogenetic relationships within the family has been problematic, and a large-scale molecular phylogeny in terms of generic sampling and number of genes is still lacking. In particular, the deeper relationships within the family, for example between the three major recognized lineages, prove particularly hard to resolve. Using a slow-evolving mitochondrial marker (*nad4* intron 1), we reconstructed a comprehensive phylogeny in generic representation for the family. In addition, and because resolution was very low in previous single marker phylogenies, we adopted a supermatrix approach by concatenating all checked and reliable sequences available on GenBank as well as new sequences for a total 207 currently recognized genera and eight molecular markers representing a comprehensive coverage of all three genomes. The supermatrix was dated under an uncorrelated relaxed molecular clock using a direct fossil calibration approach. Finally, a lineage-through-time-plot and rates of diversification for the family were generated. The resulting tree, the largest in number of genera and markers sampled to date and covering the whole family in a representative way, provides important insights into the evolution of the family on a broad scale. The backbone of the tree remained largely unresolved and is interpreted as the consequence of early rapid radiation within the family. The age of the family was inferred to be 37.6 (24.2–49.4) Ma, which largely agrees with previous studies. The ages of all major lineages and tribes are also reported. Analysis of diversification suggests that Brassicaceae underwent a rapid period of diversification, after the split with the early diverging tribe Aethionemeae. Given the dates found here, the family appears to have originated under a warm and humid climate approximately 37 Ma. We suggest that the rapid radiation detected was caused by a global cooling during the Oligocene coupled with a genome duplication event. This duplication could have allowed the family to rapidly adapt to the changing climate.

Key words: Brassicaceae, supermatrix, molecular dating, diversification, whole genome duplication, phylogeny.

Introduction

The Brassicaceae is an important family that includes several species of crops (e.g., *Brassica* spp.), weeds (e.g., *Capsella, Lepidium, Sisymbrium, and Thlaspi*), ornamentals (e.g., *Hesperis, Lobularia, and Matthiola*), and the model organism for flowering plants *Arabidopsis thaliana* (L.) Heynh. In recent years, many molecular phylogenetic analyses at the tribal (Bailey et al. 2002; Warwick and Sauder 2005; Warwick et al. 2007, 2008; Zunk et al. 1999) and family levels (Koch et al. 2001, 2007; Bailey et al. 2006; Beilstein et al. 2006, 2008; Koch and Mummenhoff 2006; Franzke et al. 2009) have been published (for a review see Al-Shehbaz et al. 2006; Koch and Al-Shehbaz 2009). These studies have contributed significantly to a better understanding of the systematics and evolution of the family as well as the delimitation and relationships of the tribes. In all family-wide analyses, *Aethionema* (including *Moriera*) was sister to the rest of the family that formed three weakly defined major lineages each consisting of several tribes (Al-Shehbaz et al. 2006; Bailey et al. 2006; Beilstein et al. 2006, 2008; Franzke et al. 2009). However, there is a lack of resolution in the backbone of the Brassicaceae tree (e.g., relationship between the major lineages as well as relationships of other groups to these major lineages), regardless of the molecular marker(s) used. In an attempt to resolve that backbone, Franzke et al. (2009) sequenced a slow-evolving mitochondrial (mt) maker, *nad4* intron 1, which was shown to be 23 times slower than the rDNA ITS sequences in the family (Yang, Lai, Tai, Li 1999). They used a small sample of 49 genera representing all tribes recognized by Al-Shehbaz

Table 1. Different Age Estimates for the Crown Node of the Brassicaceae Under Different Methods.

Reference	Crown Age of Brassicaceae (node at Split Between *Aethionema* and the Rest of Brassicaceae)	Method
Koch et al. (2000, 2001)	30–60 Ma	Synonymous mutation rates
Ermolaeva et al. (2003)	24–40 Ma	*Arabidopsis* genome duplication
Schranz and Mitchell-Olds (2006)	34 Ma	*Arabidopsis* genome duplication
Henry et al. (2006)	24–40 Ma	*Arabidopsis* genome duplication
Fawcett et al. (2009)	ca. 40 Ma	*Arabidopsis* genome duplication
Franzke et al. (2009)	15 Ma	Secondary calibration/relaxed molecular clock

et al. (2006), but the *nad4* intron 1 resolution among tribes remained low. However, their findings showed a strong phylogenetic congruence with those based on nuclear rDNA ITS (Bailey et al. 2006), chloroplast *ndhF* (Beilstein et al. 2006), and nuclear *phy*A (Beilstein et al. 2008) sequence data.

The lack of basal resolution in the Brassicaceae perhaps resulted from the radiation events during the early stages of its evolution (Bailey et al. 2006; Koch et al. 2007; Franzke et al. 2009). Under such a scenario, one would ideally expect to obtain short branches subtended by longer ones independent of the DNA marker(s) used, which suggests that diversification and evolutionary rates in the family were not constant during its evolution. Indeed, one would expect to infer higher diversification rates at the early stages of the phylogenetic tree for the family, followed by a decrease in rates, and eventually subsequent additional radiation events. Alternatively, the observed lack of basal resolution in the Brassicaceae phylogenetic trees could be the product of "among-tree conflict," possibly caused by processes such as recombination or ancient hybridization.

Molecular dating and age estimates of the family have been controversial (table 1). Using estimates of synonymous mutation rates of the nuclear marker chalcone synthase (*Chs*) with alcohol dehydrogenase (*Adh*) as well as the chloroplast marker maturase (*matK*) with *Chs*, Koch et al. (2000, 2001) dated the crown age of the Brassicaceae to be between 30 and approximately 60 Ma. Estimations of the origin of Brassicaceae based on timing of the genome duplication in *Arabidopsis* yielded ages situated around 34 Ma (Schranz and Mitchell-Olds 2006), 38–30 Ma (Ermolaeva et al. 2003), or 40–24 Ma (Henry et al. 2006). A significantly younger age of 15 Ma (1–35) was estimated by Franzke et al. (2009) using a secondary calibration approach under a relaxed uncorrelated molecular clock.

Concepts on evolutionary processes at the family level are developing quickly in the Brassicaceae. There is not only a model of the ancestral karyotype (Lysak et al. 2006, 2007) but also a model to reconstruct genome structure evolution in a modularized way (Schranz and Mitchell-Olds 2006; Schranz et al. 2007). Both concepts are closely related to and influenced by ideas on genome size evolution (Lysak et al. 2009). However, any of these studies strongly depend on reliable evolutionary hypothesis in terms of phylogeny but also modes of evolution. What are the most important milestone accomplishments in crucifer phylogenetics during the past two decades? In principle, and aside from the wealth of knowledge on the model species *Arabidopsis* and *Brassica*, they are at least four: 1) achieving a new infrafamiliar classification based on phylogenetically circumscribed genera and tribes and genera, 2) phylogenetic circumscription of the order Brassicales and the determination of Cleomaceae as the closest and sister family to Brassicaceae, 3) unravelling general principles of crucifer evolution by exploring species- or genus-specific evolutionary histories, and 4) detailed information on karyotype and genome evolution across the family including genetic and cytogenetic maps as well as whole-genome DNA sequence data sets.

This study has three major objectives that open up new perspectives into the evolutionary history of the Brassicaceae. First, we evaluated the potential of the mt marker *nad4* intron1 to resolve Brassicaceae relationships under an almost complete generic sampling of the family. Second, by means of a supermatrix approach, we generated a comprehensive evolutionary framework for the whole family, both in number of characters and generic representation to test recent tribal delimitation. Third and most importantly, we tested the hypothesis of an early radiation in the history of the family by dating the resulting supermatrix tree using a primary calibration point and undertaking analyses of diversification within the family.

Material and Methods

Molecular Markers and Taxon Sampling

Nad4 *Intron 1 Mitochondrial Marker Data Set*. The first intron of the mt gene for NADH subunit 4 (*nad4* intron 1) was selected for this analysis because it is considered to be slow evolving (Yang, Lai, Tai, Ma, et al. 1999). Sampling for this data set was built on Franzke et al. (2009). A total of 287 taxa were sequenced for *nad4* intron 1. Of these, 45 taxa had identical *nad4* intron 1 sequences (supplementary table S1, Supplementary Material online). We thus used just one taxon per identical species, reducing the number of taxa to 257 species representing 235 currently recognized genera. This represents the largest sampling to date of Brassicaceae genera for a single phylogenetic marker (supplementary table S2, Supplementary Material online). Experimental procedures to generate *nad4* intron 1 sequences followed Franzke et al. (2009), and in summary we added 235 new sequences here.

The complete list of taxa sampled as well as their GenBank numbers is given in supplementary table S2

(Supplementary Material online). The two outgroups were *Moringa oleifera* (Moringaceae) and *Cleome viscosa* (Cleomaceae).

Supermatrix Data Set. We adopted a supermatrix approach following Bailey et al. (2006). This matrix was constructed by concatenating unpublished and available GenBank sequences for seven additional commonly used genes in Brassicaceae phylogenetics (supplementary table S3, Supplementary Material online). These include ITS, *chs*, *adh* of the nuclear genome (nDNA), and *matK*, *trnL−F*, *ndhF*, and *rbcL* of the plastid genome (cpDNA). With the addition herein of *nad4* intron 1 (mtDNA), the resulting sample represents a comprehensive coverage of all three genomes. Incongruence between gene trees has been shown not to be a major issue within Brassicaceae phylogenetics with most markers returning similar relationships on the broad family scale (Beilstein et al. 2006; Franzke et al. 2009; Koch and Al-Shehbaz 2009). Indeed, concatenating sequence data in Brassicaceae, via supermatrix or supernetwork approaches, has been done repeatedly in the past though at smaller taxon sampling scales (Bailey et al. 2006; Koch et al. 2007; Lysak et al. 2009). Moreover, because of the size of the data set and the large proportion of unequal taxon sampling between the different markers used, topology-based incongruence tests (e.g., Kishino and Hasegawa test; Kishino and Hasegawa 1989) are either impossible to apply or prohibitively computer intensive (e.g., incongruence length difference test; Farris et al. 1994). Here we choose to generate separate maximum likelihood (ML)-based gene trees (see below) and scanned them for well-supported incongruencies.

For the taxon sampling, we included as many genera as possible. We also included old names of genera that have been recently reduced into synonymy (supplementary table S4, Supplementary Material online). These older names are retained because they are still used in all floras, monographs, and herbaria. However, the numerous nomenclatural adjustments (supplementary table S4, Supplementary Material online), which were published in scattered journals within the past 3 years, are compiled herein for the first time to update the reader and serve as a guide for future studies. The complete list of taxa used for the supermtrix analysis as well as GenBank numbers for all sequences is given in supplementary table S4 (Supplementary Material online).

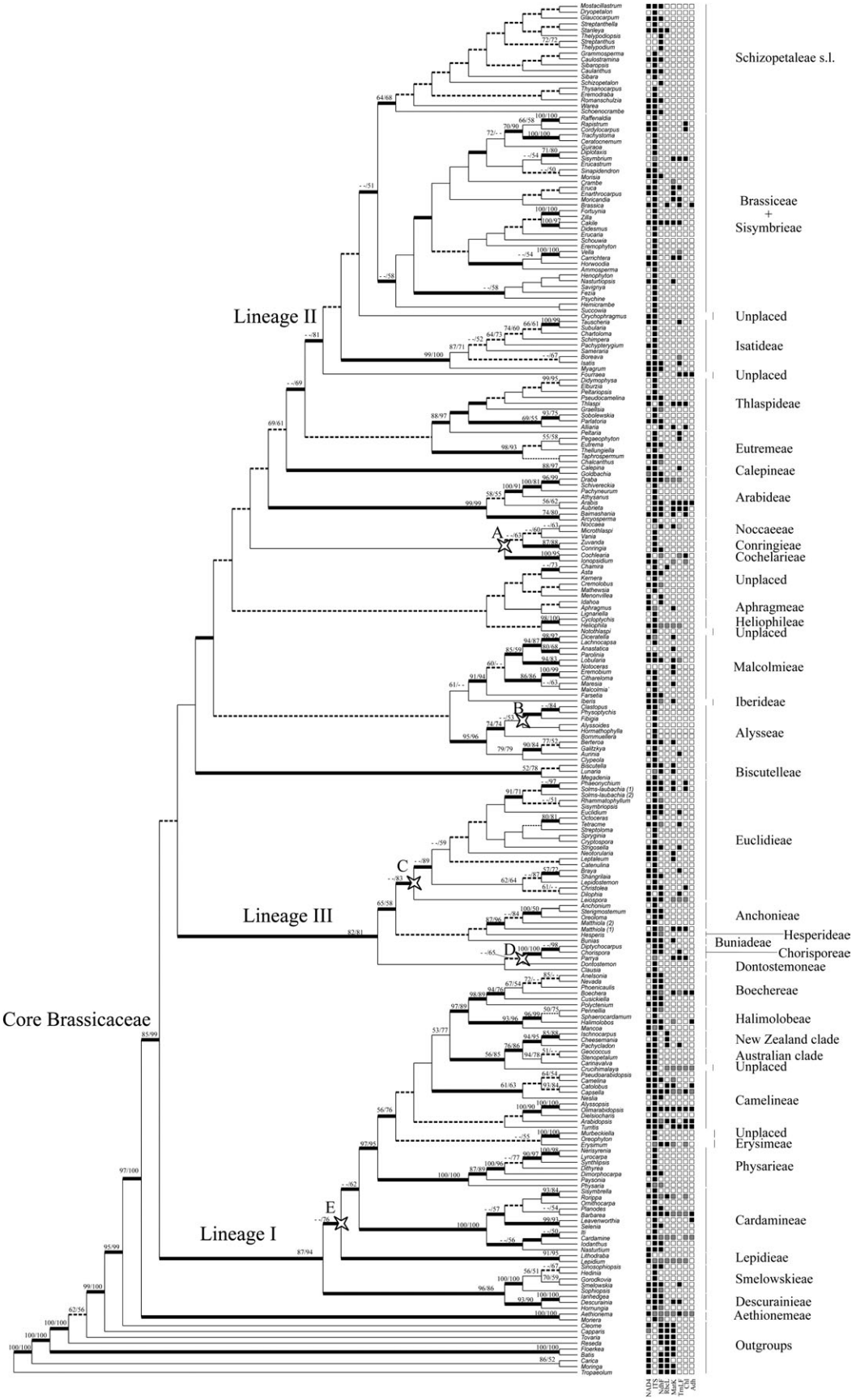
For the sampling of markers per taxon, we generally adopted the "bottom up-top down" approach (Wiens et al. 2005; Pirie et al. 2008) that combines the conservative and slow-evolving markers for a small number of taxa used as placeholders with the more variable and faster evolving markers for the entire study group. ITS sequence data has been used extensively in Brassicaceae systematics mainly for resolving relationship within tribes (Koch et al. 1999, 2001, 2003; Mummenhoff et al. 2001; Bailey et al. 2002; Warwick et al. 2002; Warwick, Al-Shehbaz, Sauder, Harris, et al. 2004; Warwick, Al-Shehbaz, Sauder, Murray, et al. 2004; Warwick and Sauder 2005; Al-Shehbaz et al. 2006; Koch and Mummenhoff 2006; Warwick, Al-Shehbaz, et al. 2006; Warwick et al. 2007, 2008; Nagpal et al. 2008) and was therefore selected here. A total of 58 generic representatives were newly sequenced for ITS, whereas 154 additional sequences were retrieved from GenBank (supplementary table S4, Supplementary Material online). The remaining genes included in this study, including *nad4* intron 1, are generally considered as slow-evolving relative to the ITS and hence deemed more useful in resolving "deeper" phylogenetic relationships.

As *nad4* intron 1 appeared to be very slow evolving within Brassicaceae (see Results), a selection of taxa for this gene was made. Per well-supported tribe identified based on ITS sequences, one to four *nad4* intron 1 sequences were selected (supplementary table S4, Supplementary Material online). In particular, we focused our *nad4* intron 1 subsampling on taxa that were also represented by several other genes, especially *ndhF* (fig. 1). For the *matK* marker, a total of 33 sequences (for 33 genera) were generated according to Abdel Khalik (2002) and those for 25 other genera were obtained from GenBank. All sequences were aligned using MUSCLE on the online server of the European Bioinformatics Institute at http://www.ebi.ac.uk/Tools/muscle/index.html. Aligned sequences were then manually checked and adjusted using PAUP* (version 4.10b; Swofford 2002).

To minimize missing data, but producing composite taxa in the same time, we concatenated sequences from different species of the same genus (supplementary table S3, Supplementary Material online) following Springer et al. (2004). This was only done if evidence for monophyly of the genus or group existed in literature (supplementary table S3, Supplementary Material online). We feel it is inevitable to follow this approach when extracting information from GenBank because the relevant data were not often obtained by targeted sequencing efforts. Thus, sequences for a given gene generated from identical specimen vouchers or species are rare in GenBank. The alternative approach would be to break down the composite taxa and treat them as separate terminals (Malia et al. 2003). One major drawback of this method is that, as there will be no sequences in common between the decomposed taxa, discovery of synapomorphies for that taxon is hampered (Springer et al. 2004). This in turn leads to highly unresolved trees and failure to reconstruct previously well-supported clades (Malia et al. 2003). The use of composite taxa has been successfully undertaken in other large-scale phylogenetic studies (Simpson et al. 2002; McMahon and Sanderson 2006; Burki et al. 2007; Baker et al. 2009).

Parsimony Analysis

Following Pirie et al. (2008), a two-stage heuristic search strategy was used for the *nad4* intron 1 data set to find the set of most parsimonious trees. First, 100 parsimony ratchet searches (Nixon 1999) of 100 generations each were performed, using PAUPrat (Sikes and Lewis 2001) to find islands of shortest trees. For the supermatrix data set, the number of parsimony ratchet searches was reduced


Schizopetaleae s.l.
Brassiceae + Sisymbrieae
Unplaced
Isatideae
Unplaced
Thlaspideae
Eutremeae
Calepineae
Arabideae
Noccaeeae
Conringieae
Cochlarieae
Unplaced
Aphragmeae
Heliophileae
Unplaced
Malcolmieae
Iberideae
Alysseae
Biscutelleae
Euclidieae
Anchonieae
Hesperideae
Buniadeae
Chorisporeae
Dontostemoneae
Boechereae
Halimolobeae
New Zealand clade
Australian clade
Unplaced
Camelineae
Unplaced
Erysimeae
Physarieae
Cardamineae
Lepidieae
Smelowskieae
Descurainieae
Aethionemeae
Outgroups
Lineage II
Lineage III
Lineage I
Core Brassicaceae
A
B
C
D
E

to 50 with 100 generations each. The resulting shortest trees were then used as starting trees for a second round of heuristic searching by applying tree bisection reconnection (TBR) branch-swapping algorithm and limiting the number of saved trees to 10,000. For both data sets, branch support was estimated using jackknife analyses of 100 replicates with "full" heuristic searches of 10 random addition sequences, TBR, saving 10 trees each time. For every marker and every data set, only parsimony-informative (PI) indels were coded following the simple coding model (Simmons and Ochoterena 2000).

ML Analyses

In the past years, ML methods have undergone a huge boost in efficiency of tree-searching algorithms (Guindon and Gascuel 2003; Stamatakis 2006; Zwickl 2006; Morrison 2007). These algorithms are very time efficient allowing for fast and accurate estimation of ML trees, and even bootstrapping, on large data sets. For this study, we used the RAxML web-server program available at the CIPRES portal in San Diego (http://8ball.sdsc.edu:8889/cipres-web/Home.do). This online version implements a very efficient and rapid bootstrap heuristic in RAxML (Stamatakis et al. 2008). For each analysis, the "Maximum likelihood search" and "Estimate proportion of invariable sites" boxes were selected, with a total of 100 bootstrap replicates performed. In the case of the supermatrix, 300 bootstrap replicates were performed.

Bayesian Estimation of Phylogeny

Both Bayesian reconstruction of phylogenetic relationships and molecular dating were determined from the supermatrix data set by applying BEAST v1.4.8 (Drummond and Rambaut 2007). We prefer BEAST to other Bayesian inference programs (e.g., MrBayes; Ronquist and Huelsenbeck 2003) because first it implements a faster likelihood search method and performs well on large analyses (Shapiro et al. 2004; Drummond and Rambaut 2007). Secondly, it allows incorporation of the assumption of a relaxed molecular clock, as time-dependency of the evolutionary process is thought to be of importance in Bayesian estimation of phylogenies, based on DNA sequence data (Drummond et al. 2006). Finally, BEAST allows the incorporation of a starting tree, the use of which allows a quicker converge of the Bayesian Markov chain Monte Carlo (MCMC) run and, therefore, significantly decreases the time needed for analysis of large data sets (Pirie et al. 2008). The ML tree, found with RAxML and rendered ultrametric using the program r8s (Sanderson 2003), was used as a starting tree for independent BEAST runs. As the supermatrix deviated from a strict molecular clock model and rates between adjacent branches were uncorrelated (see Results), a lognormal uncorrelated relaxed clock model was specified in BEAST. We included two speciation process models: birth–death (Gernhard 2008) and pure birth (Yule 1924) as they were shown to yield different age estimates (Gernhard 2008). The Bayes factor as implemented in Tracer 1.4 (Rambaut and Drummond 2003) was used to select the best-fitting model under the smoothed marginal likelihood estimate and with 100 bootstrap replicates (Suchard et al. 2001). The supermatrix data set was not partitioned into individual genes, and a general time reversible model without gamma rate distribution was applied. We are aware of the benefits of data partitioning when using Bayesian analyses (Nylander et al. 2004; Brandley et al. 2005); however, reducing the complexity of such analyses increases the number of tree topology changes suggested during the MCMC, thus leading to better exploration of tree topology space. This approach, in combination with a starting topology, allowed independent runs to reach stationarity and to converge to similar parameter likelihoods under the imposed time limit (see below).

A total of 30 independent runs of 2 million generations each were undertaken on the online cluster of the Computational Biology Service Unit from Cornell University (http://cbsuapps.tc.cornell.edu/beast.aspx). This cluster imposes a time limit of 3 days (72 h) per analysis but allows several runs of the same analysis simultaneously. Analyses were undertaken by sampling every 1000th generation. Tracer 1.4 was used to check for convergence of the model likelihood and parameters between each run until reaching stationarity. The resulting log and tree files were then combined using LogCombiner. Results were considered reliable once the effective sampling size (ESSs) of all parameters were above 100 (see results for the total number of generations). Finally, branches with posterior probabilities (PP) below 0.8 were considered as weak, between 0.8 and 0.95 as moderate, and above 0.95 as strong.

Fossil Calibration and Molecular Dating

Brassicaceae has a fairly poor record of macrofossils (Schulz 1936; Appel and Al-Shehbaz 2003). For this study, we relied on the Turonian taxon fossil *Dressiantha* (ca. 85 Ma; Gandolfo et al. 1998). The set of characters found in *Dressiantha*, such as the presence of a gynophore, unequal petal size, or a bicarpellate gynoecium, suggest an affinity with the order Brassicales (sensu APGII 2003). This fossil has been used in the past for calibration of the Brassicales (Magallon et al. 1999; Magallon and Sanderson 2001). To incorporate this fossil as a primary calibration point, we selected a certain number of outgroup taxa leading up to the crown group of the Brassicales (table 1).

←

Fig. 1. ML cladogram of the 226 ingroup Brassicaceae taxa and nine outgroup taxa with support values indicated. Thick branches: PP $>$ 0.95; normal lines 0.8$<$PP$<$0.95; dash lines PP$<$0.8. Values above branches are bootstrap values of the ML analysis (right) and maximum parsimony (left). Stars indicate nodes of supported conflict between the JK and the ML/Bayes tree topologies. Letters A–D are reported in supplementary fig. S1, Supplementary Material online. The boxes next to names represent sampling of markers per taxon. Black boxes: identical taxon names; grey: taxa with different species names than black ones. Empty boxes: no marker sampled.

Dressiantha was placed at the crown node of Brassicales (split between Tropaeolaceae/Akaniaceae and the rest of the Brassicales families). Moreover, to accommodate for calibration uncertainty, we applied a normal distribution as prior to the calibration node within the BEAST analysis with a mean of 89.5 Ma and standard deviation of 1 (effectively enclosing dates from 86−91 Ma). Although normal prior distributions are used when dating trees under an indirect approach (Ho 2007), we prefer this type of distribution because it does not place a strict minimum age on the calibration. Indeed, the actual dating of the fossil is also subject to uncertainty and by allowing the ages to vary around the mean of the distribution appears as a more realistic choice.

Rates of Diversification

To provide an indication of diversification rates in the family, we generated a semilogarithm lineage-through-time (LTT) plot using the LASER package version 2.2 (Rabosky 2006). The LTT plot was generated on the last 1,000 trees sampled from the posterior to account for phylogenetic uncertainty. For each posterior BEAST tree, the nine outgroups were pruned and the root node of the Brassicaceae was set to 37.6 Ma (the average crown age recovered from the BEAST analysis; table 1) using TreeEdit v1.0a10 (Rambaut 2002a). The mean LTT curve was computed as well as the 95% confidence intervals (CI) and referred to as the "mean Bayes LTT plot." This "plot" was compared with null models of constant rate diversification under two extreme relative extinction rates (speciation λ; extinction μ; extinction rate a $= \mu/\lambda = 0$ and 0.9, with $\lambda = 0.2$). To account for incomplete taxon sampling (Pybus and Harvey 2000), these null distributions were generated by Markov-chain tree simulations using Phyl-O-Gen 1.1 (Rambaut 2002b). Per extinction rate, 1,000 phylogenies were generated to a standing diversity of 3,709 terminals (total number of species recognized in Brassicaceae; Warwick, Francis, et al. 2006) and randomly pruned to 226, reflecting our 226 sampled taxa (see Results). The resulting 226 taxon trees were used to compute mean and 95% CI LTT curves after rescaling the root node to 37.6 Ma. In addition, the mean Bayes LLT plot was compared with the null models (0 and 0.9) by Kolmogorov–Smirnov goodness-of-fit tests. Even though we have significant amount of missing taxa, we have sampled a large number of genera as well as all tribes within the family. Thus, sampling of lineages is complete in the older parts of the phylogeny becoming progressively more incomplete toward the present (Ricklefs et al. 2007). Accordingly, we restrict our interpretations to the older proportions of the LTT plot.

We tested the null hypothesis of no-rate change versus a variable-rate change in diversification, using the ML approach of Rabosky implemented in the LASER package version 2.2 (Rabosky 2006). The test statistic for diversification rate-constancy is calculated as: $\Delta AIC_{RC} = \Delta AIC_{RC} - \Delta AIC_{RV}$, where AIC_{RC} is the Akaike information criterion score for the best-fitting rate-constant diversification model, and AIC_{Rv} is the AIC for the best fitting variable-rate diversification model. Thus, a positive value for ΔAIC_{Rv} indicates that the data are best approximated by a rate-variable model. We tested four different models, of which two are rate-constant and two are rate-variables. 1) The constant-rate birth model (the Yule process; Yule 1924) with one parameter λ and μ set to zero; 2) the constant-rate birth–death model with two parameters λ and μ; 3) a pure birth rate-variable model where the speciation rate λ_1 shifts to rate λ_2 at time t_s, with three parameters (λ_1, λ_2, and t_s); 4) same as 3) but with two shifting points and three different speciation rates (five parameters). The LTT plot derived from the maximum clade credibility (MCC) tree was used for this part.

In addition, we calculated the gamma (γ) statistic (Pybus and Harvey 2000) to test if diversification has decelerated through time. When γ is negative, the internal nodes are closer to the root than expected under a pure-birth model ($\lambda = 0$), thus indicating a decrease in diversification through time. The posterior distribution of the γ-statistic was computed from 1,000 posterior trees from the BEAST analysis. It has been shown, however, that incomplete taxon sampling can lead to a type I error (incorrect rejection of the constant rate null hypothesis). We thus applied the Monte Carlo constant rate test of Pybus and Harvey (2000) as implemented in the LASER package version 2.2 (Rabosky 2006). We compared the empirical γ-statistic distribution computed from the 1,000 posterior trees (from the BEAST analysis above) with the distribution of a γ-statistic of 1,000 simulated phylogenies under a pure birth model.

Finally, we calculated absolute net diversification rates (speciation minus extinction) for the family (3,709 species; Warwick, Francis, et al. 2006) and for the core Brassicaceae (3,652 species; Warwick, Francis, et al. 2006) by using the crown age under a high level of extinction ($\lambda = 0.9$) and no extinction ($\lambda = 0$) following Magallon and Sanderson (2001). These rates were computed using the LASER package version 2.2 (Rabosky 2006).

Results

Nad4 Intron 1 Analyses

The total length, after alignment of the 259 sampled taxa (including the two outgroups), was 1,495 bp with an extra 168 PI indels coded. This led to a total of 389 (23.8%) PI characters. Of 100 ratchet searches, 24 gave a shortest tree of 947 steps (CI = 0.49; retention index = 0.85). The other ratchet searches found trees of 948 to 953 steps long. The resulting jackknife consensus tree was unresolved with more than one-third of sampled taxa (94 of 257) collapsing in a polytomy. The ML bootstrap analysis in RAxML also resulted in a weakly support tree (supplementary fig. S1, Supplementary Material online). Nevertheless, several tribes received high support (e.g., Cardamineae). The sister relationship of *Aethionema* with the rest of the Brassicaceae was strongly supported by bootstrap values while the deeper relationships within the core Brassicaceae remained unresolved.

Supermatrix Analyses

The 226 ingroup taxa sampled belong to 207 genera and 33 presently recognized tribes (Al-Shehbaz et al. 2006;

Table 2. Character Information for the Supermatrix Data Set.

Charset		No. taxa	PI indels	No. charc.	Variable charc.	% variable	PI charc.	% PI	Total PI charc.
adh		15	80	1,986	857	43	460	23	540
chl		26	0	999	400	40	278	28	278
ITS		215	104	642	482	75	334	52	438
	Nuclear		184	3,627	1,739	48	1,072	30	1,256
nad4 intron 1	*mtDNA*	90	70	1,491	363	24.3	100	7	170
ndhF		84	0	2,070	831	40	459	22	459
matK		58	0	1,041	442	42	234	22	234
rbcL		17	0	1,155	92	8	32	3	32
trnL-F		52	21	636	344	54	149	23	170
	cpDNA		21	4,902	1,709	34.9	874	18	895
Total			275	10,020	3,811	38.0	2,046	20	2,321
		Taxa	Characters						
Grand total		226	10,295						

Al-Shehbaz and Warwick 2007; German and Al-Shehbaz 2008). The aligned supermatrix comprising the seven gene sequences (table 2) contained 10,020 characters (analyzed under ML and BEAST), or 10,295 when the 275 PI indels were added (maximum parsimony [MP] analysis). For the ITS partition, several regions were excluded from the analysis because of alignment ambiguities (positions 133–167; 502–529; 598–608; 634–652). In the resulting supermatrix, 75% of data were missing, where 86 taxa had more than 90% of data missing, whereas 34 taxa had 50% or less data missing (supplementary table S3, Supplementary Material online). In total, 2,019 (20%) characters were PI (table 2). The ITS data set provided the most number of PI characters for a single marker (52%; 19% in total, including indels). However, *rbc*L and *nad4* intron 1 had the lowest number of PI characters (3% and 7%, respectively).

For the parsimony analysis, a shortest tree of 12,889 steps long (CI = 0.35 RI = 0.55) was found twice out of the 100 ratchet searches conducted. The length of the other trees found varied from 12,890 to 12,913 steps long. For the RAxML analysis, the final ML optimization likelihood was of −90134.5.

Because of the time limit imposed by the server on the Bayesian phylogeny estimations, no runs reached the 2 million generations specified, and runs were stopped between 1.8 and 1.9 million generations. However, all 30 independent runs reached stationarity within the first 10,000 generations (because the starting tree was already near optimum), and all parameter estimates were consistent between runs. Runs were thus combined, after removing a burn-in of 20,000 generations each, into a single run of 55 million generations. All parameters, including age estimates, reached acceptable ESS values and were thus deemed reliable (ESS > 100). The Bayes factor strongly favored the Yule model over the birth–death model (logarithm of Bayes factor in favor of Yule model = 12 ± 0.46) and, therefore, results under the Yule model are presented here. The tree files generated under the Yule speciation model were combined after a burn-in of 200 trees each. As the resulting combined tree file was too large to analyze in TreeAnnotator (1.5 > Giga bytes), combining of the independent runs was redone. This time it was resampled at a lower frequency of every 5000th tree resulting in a file containing 11,500 trees sampled from the posterior and used to generate the MCC tree in TreeAnnotator.

Phylogeny

After a visual check, no well-supported conflicts (i.e., those receiving bootstrap (BS) > 90%) were found between individual genes trees. Figure 1 represents a cladogram version of the ML tree from the supermatrix analysis. When the PI indels were excluded from the analyses, the same relationships were observed, but some clades received lower support. In general, the jackknife (JK) majority rule consensus, the ML and MCC trees agreed, and well-supported clades were identical in all cases. However, five moderately to strongly supported incongruencies were found between the JK and the ML/Bayes tree topologies, indicated with stars in supplementary figure S2 (Supplementary Material online). Some of them are only moderately supported in all analyses (e.g., conflicts A and B), and some are strongly supported and invert a few relationships at the generic level (e.g., C, D, and F). The most serious conflict identified pertains to the position of the Lepidieae tribe (E). In the JK analysis, the Lepidieae are recovered as sister to the whole of lineage I, whereas in the ML/Bayes analyses it appears as sister to the rest of lineage I excluding Descurainieae and Smelowskieae (supplementary fig. S2A, Supplementary Material online). Finally, certain nodes were highly supported under one optimality criterion but not under the others. For example, the node sustaining the tribe Brassiceae received a JK and ML BS values <51%, whereas the PP was 0.95. Lineage II, grouping the tribes Schizopetaleae, Sisymbrieae, Brassiceae and Isatideae (Koch and Al-Shehbaz 2009) received strong support under ML and Bayesian inference and no support with MP (fig. 1).

The majority of multigeneric tribes (e.g., Arabideae and Cardamineae) received moderate to high support in all three analyses, but a few were poly/paraphyletic with low support. The supermatrix approach presented here did not resolve the backbone of the family based on our data. Resolution with lineages I and III were higher in the Bayesian analysis (62% and 38% of nodes resolved), whereas the overall support within lineage II was half as

Table 3. Mean and 95% of the highest posterior distributions (HPD) crown age estimates of the different multigeneric tribes with their corresponding support values under the 33% Jackknife analysis (MP), maximum likelihood bootstrap analysis (ML), and the BEAST analysis (PP).

TMRCA for Multigeneric Tribes	Mean	95% HPD lower	95% HPD upper	Support (MP/ML/PP)
Alysseae	19.3	12.3	26.7	95/96/1.00
Anchonieae	17.1	10.4	24.7	87/96/0.99
Arabideae	16.8	10.0	23.9	99/99/1.00
AUS/NZ clade	11.4	7.2	15.8	76/86/0.98
Brassiceae/Sisymbrieae	17.3	11.7	23.1	–/55/0.98
Boechereae	8.5	5.2	12.3	98/89/1.00
Brassicaceae	37.6	24.2	49.4	97/100/1.00
Calepineae	16.4	6.6	27.3	88/97/1.00
Camelineae1	19.7	13.0	26.4	–/–/–
Camelineae2	16.1	10.4	22.0	61/63/0.99
Cardamineae	17.7	11.6	24.2	100/100/1.00
Chorisporeae	16.6	9.9	23.6	100/100/1.00
Conringieae	10.9	3.9	18.4	87/88/0.99
Core Brassicaceae	32.3	20.9	42.8	85/99/1.00
Descurainieae	19.2	11.2	28.0	93/90/0.99
Euclidieae	17.4	11.3	24.3	–/83/0.99
Halimolobeae	8.6	5.0	12.5	93/961.00
Heliophileae	11.76	5.53	18.86	98/100/1.00
Isatideae	11.2	6.5	16.7	99/100/1.00
Lepidieae	13.2	4.3	23.0	91/95/1.00
Lineage I	27.3	18.2	36.1	87/94/1.00
Lineage II	28.2	18.1	37.2	82/81/0.99
Lineage III	21.4	14.8	29.3	–/81/0.99
Malcolmieae	19.8	12.8	26.9	91/94/1.00
Moringaceae/Brassicaceae split	72	47.9	90.5	100/100/1.00
Noccaeeae	14.6	7.4	22.6	70/60/0.82
Physarieae	17.1	10.5	23.8	100/100/1.00
Rorippa/*Sisymbrella* split	4.5	0.9	9.2	93/84/1.00
Smelowskieae	9.9	4.2	15.7	100/100/1.00
Schizopetaleae	10.8	6.0	16.5	64/68/0.99
Thlaspideae	9.8	4.2	15.4	88/97/1.00

low (15% of nodes resolved). Certain genera did not cluster with strong support within any clade and are referred to here as "floating". Such genera include *Iberis*, *Notothlaspi*, *Fourraea*, *Orychophragmus*, *Bunias*, and *Clausia*.

A comparison of the results presented here (fig. 1) with previously published, family-wide studies (Warwick and Sauder 2005; Al-Shehbaz et al. 2006; Bailey et al. 2006; Beilstein et al. 2006, 2008; Warwick et al. 2007, 2008; Franzke et al. 2009) reveal that several tribal adjustments are needed and are addressed at the end of the discussion section below.

Molecular Dating

The supermatrix data set deviated from the strict molecular-clock model as indicated by the rate–variation coefficient (value not abutting against zero, mean = 0.621; 95% CI 0.54–0.68). As shown by the rate of covariance, which was centered on zero (mean = 0.067; 95% CI −0.02 to 0.169), the rates between adjacent branches were uncorrelated, and values indicated that a lognormal non-correlated relaxed-clock method best fits the data (Peng et al. 2006; Drummond and Rambaut 2007). Age estimates did not differ under the Yule or birth–death speciation models, and neither model better fitted the data under the Bayes factor (log BF was always <1 for all nodes analyzed). The estimated dates for multigeneric clades presented here are those inferred under the Yule speciation model (table 3).

Rates of Diversification

Goodness-of-fit tests indicated that the mean Bayes LTT plot significantly deviated from expectations under constant diversification models with extinction rates of 0 and 0.9 ($P = 0.0103$; $P < 0.001$, respectively). The LTT plots indicate that the family Brassicaceae did not follow a constant model of diversification during its evolutionary history. When compared with the null models, diversification rates appeared higher in the early stages of the family's evolution than toward the end. Two independent events of diversification change can be identified. One acceleration occurred at around 32 Ma till approximately 22 Ma (grey patch in fig. 2*A* and *B*), after which diversification rates decreased. We fitted several models of diversification to the empirical observations. The diversification rate-constancy statistic ΔAIC_{Rv} was found to be 65.36, indicating that the data better fitted a variable rate than a constant rate model of diversification. The five-parameter model, with two shifts in diversification, was identified as having the lowest AIC value among the other models tested, and therefore is selected as the best-fitted model to our data.

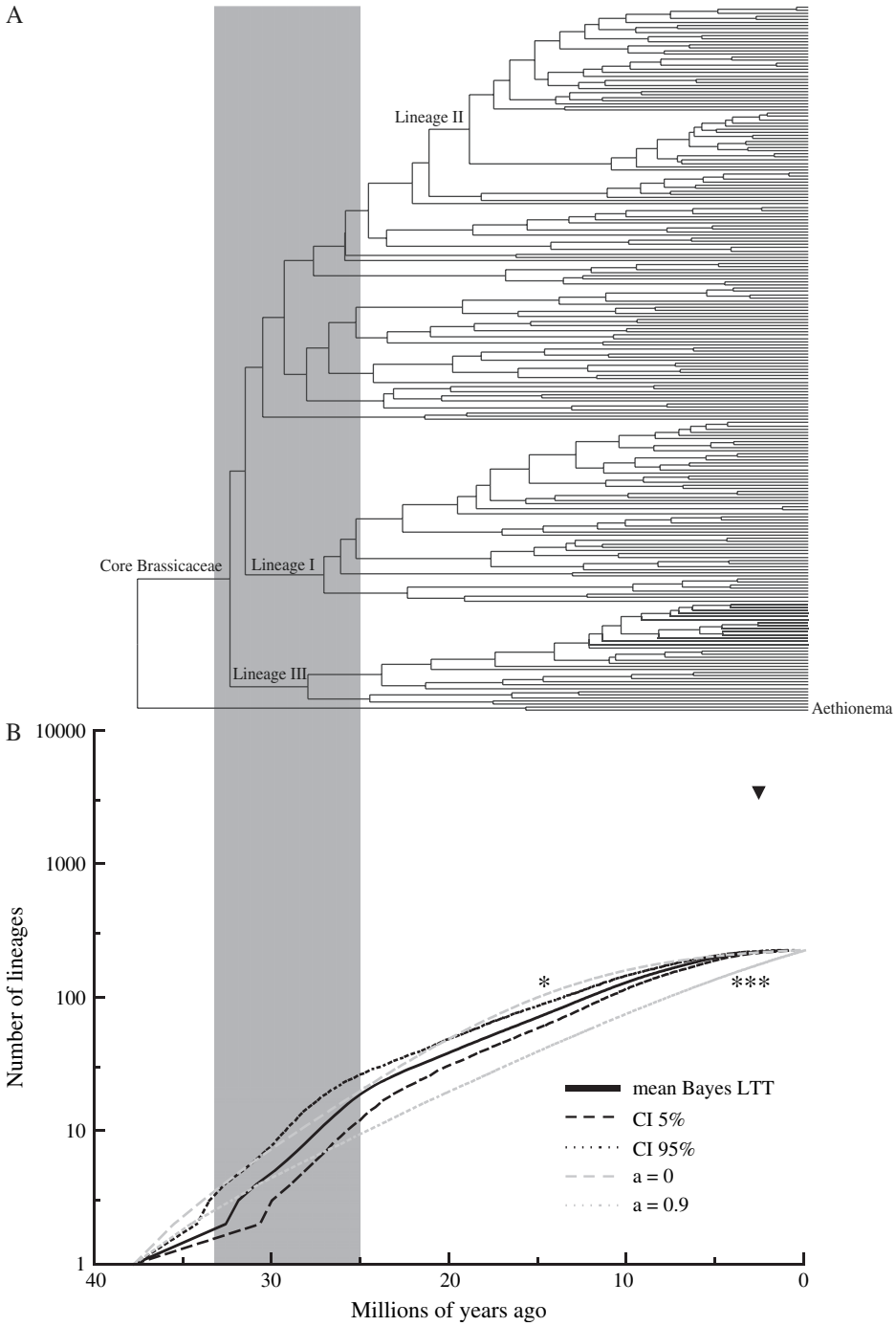


Fig. 2. Diversification of Brassicaceae. (*A*) Maximum clade credibility tree of the Brassicaceae family. (*B*) Semilogarithmic lineage through time plots of the empirical analysis (and 95% CI) and the constant rate diversification simulation analyses. Stars indicate a significant (one star) or highly significant (three stars) deviation from the constant diversification rate under the goodness-of-fit tests. Grey patch indicates period of increased diversification. Inverted triangle: Extant number of Brassicaceae species.

The mean γ-statistic as calculated from the posterior distribution of trees (mean −8.1; 95% CI = 9.8–6.9) was negative indicating a deceleration of diversification through time. However, this value was not significantly higher (supplementary fig. S3, Supplementary Material online) than the critical value found under the simulated constant rate model trees with incomplete taxon sampling (mean −10.76). This indicates that although we have a negative γ-statistic, it is not significantly negative when compared with the null model.

Absolute rates of diversification within Brassicaceae varied from 0.156 ($\lambda = 0.9$) to 0.2 ($\lambda = 0$), whereas the rates in the core Brassicaceae ranged from 0.181 ($\lambda = 0.9$) to 0.223 ($\lambda = 0$).

Discussion

Nad4 Intron 1 Marker and the Supermatrix Approach

Despite the economic and scientific importance of Brassicaceae, a large-scale well-resolved phylogeny of the family

is still lacking. Most studies focused on sampling one or few markers for approximately 40–100 genera (Bailey et al. 2006; Beilstein et al. 2006, 2008; Franzke et al. 2009). We generated a supermatrix following Bailey et al. (2006) using all available data over eight genes and including 226 taxa representing 207 genera. Recently, numerous studies have used GenBank data to infer angiosperm familywide phylogenetic trees (Sanderson and Driskell 2003; McMahon and Sanderson 2006; Higdon et al. 2007). By definition, a supermatrix includes substantial missing data (>60%) considered to pose serious problems for phylogenetic reconstruction and possibly leading to false relationships or severe lack of resolution. However, simulation studies have demonstrated that taxa with high levels of missing data can be accurately placed under most methods of phylogenetic analyses (Wiens 2003, 2006; Wiens and Moen 2008) and therefore should not be systematically excluded from the analysis. The supermatrix approach has been shown to perform well when reconstructing phylogenies of speciose clades (Wiens et al. 2005; Pirie et al. 2008), and it allowed us to generate the largest multigenic phylogenetic tree for the Brassicaceae to date (207 of 338 genera or 61% of total). The resulting tree provides important insights into the systematics of the family on a broad scale and is discussed later. The level of resolution is relatively high, with the large majority of multigeneric tribes receiving high support values (fig. 1 and table 3). This unique large-scale phylogenetic framework of the family allows the testing and/or confirmation of the placement of several genera to tribes (see below). In general, the phylogeny agrees well with previous studies. Unfortunately, the supermatrix approach failed to recover any resolution at the deeper nodes, but this might be expected with early and rapid radiation of the family (see section "Biogeography, Evolution and Diversification in Brassicaceae"). Our analysis strongly supports the sister placement of Aethionemeae tribe to the rest of the Brassicaceae confirming previous molecular studies (Bailey et al. 2006; Beilstein et al. 2006, 2008; Franzke et al. 2009). The relationships between major tribal groupings (i.e., lineages I, II, and III) remain uncertain.

Although the support values returned by the three different optimality criteria used agreed well, a few strongly supported incongruences were found between the jackknife and the ML/Bayes analyses (supplementary fig. S2, Supplementary Material online). Discrepancies between model based and parsimony methods have been shown before (Alfaro et al. 2003). In our case, these differences could be related to the large amount of missing data found within our data set. Indeed, based on simulation analyses, maximum parsimony was unable to accurately place taxa with a large number of missing data (more than 75%) when compared with model-based methods (Wiens 2006). The main reason is probably that ML and Bayesian inference methods are more robust to long-branch attraction (Alfaro et al. 2003; Wiens 2006). For example, the sister relationship found between *Bunias* and *Leiospora* in the MP analysis (conflict C; fig. 1 and supplementary fig. S2, Supplementary Material online) could well be the result of long-branch attraction.

Divergence Dates Estimates

Molecular dating within the Brassicaceae family have been limited and controversial (Koch and Al-Shehbaz 2009). Techniques based on gene duplication, secondary fossil calibrations, or synonymous mutation rates were used to date the origin of the family (table 1). However, no study has applied the more widely used approach of a direct or primary fossil calibration as appropriate fossils in the Brassicaceae or its related families are scarce. Here we present the first familywide divergence-dates estimates based on the fossil *Dressiantha* (Gandolfo et al. 1998), which, to be incorporated, necessitated a sampling of families "leading up to" the Brassicaceae. The crown node of the family was estimated to be 37.6 (24.2–49.4) Ma. This estimate largely agrees with previous dates obtained from gene duplication events or synonymous substitution rates (table 1). Moreover, the split between *Rorippa* and *Sisymbrella* was dated to around 4.5 Ma (table 3), which agrees with the oldest known Brassicaceae macrofossil that can be attributed to a distinct taxon (Mai 1995).

However, the 37.6 Ma date is significantly older than the 15 (1—35) Ma estimated by Franzke et al. (2009) under a relaxed molecular clock using the *nad4* intron1 sequence data on a smaller sample of 45 Brassicaceae genera. In their study, the age estimation of the split between Moringaceae and Brassicaceae was used as a secondary calibration point following Wikström et al. (2001). Interestingly, our results gave almost the same age estimate for that same split of 72 (47.9—90.5) Ma (tables 1 and 3). Thus, the observed discrepancies are unlikely due to different prior calibration points but rather the results of secondary calibrations not being equivalent to primary direct fossil observations. Indeed, Shaul and Graur (2002) showed that secondary calibration could lead to unreliable dates unless a normal prior distribution on the calibration (uncertainty accounted for) is applied, instead of a point calibration (uncertainty unaccounted for). Secondly, our taxon sampling is substantially higher (226 vs. 52 taxa) than that of Franzke et al. (2009). Several studies have shown that taxon sampling plays an important role in estimating divergence dates and that smaller sampling can lead to erroneous age estimates (Linder et al. 2005; Pirie et al. 2005). Thirdly, Franzke et al. (2009) included just one intron (*nad4* intron1) in their analysis, and discrepancies of dates based on single versus multiple markers do occur (Magallon and Sanderson 2005). In dating the origin of angiosperms, Magallon and Sanderson (2005) found that the youngest age of 139.84 Ma was provided by the *psb*A marker and the oldest estimate of 317.65 Ma by the *rbc*L gene. When all markers were combined, the origin was equivalent to the mean of individual markers (Magallon and Sanderson 2005). Accuracy of molecular

dating will depend on how well branch lengths are estimated, and it is generally accepted that single markers can be misleading, whereas combining multiple genes and increasing the number of sequence data will result in better age estimations (Drummond et al. 2006). Finally, little research has been done on the influence of missing data in the supermatrix on age estimations. However, it would appear that such data matrices do not influence the results under a relaxed molecular clock (Douzery et al. 2004). The addition herein of several markers might very well be responsible for the observed discrepancies over the age of the Brassicaceae.

Biogeography, Evolution, and Diversification in Brassicaceae

Several authors suggested that the lack of resolution at the deeper nodes in the Brassicaceae probably resulted from a rapid radiation of lineages during the early history of the family (Al-Shehbaz et al. 2006; Bailey et al. 2006; Franzke et al. 2009). Such cases have been referred to as the "bushes in the Tree of Life", which are created by the rapid splitting of lineages leading to short branches too difficult to resolve even in the presence of significant data (Rokas and Carroll 2006).

Previous phylogenetic studies in the Brassicaceae, based either on one gene with moderate sampling or on several genes with poor taxon sampling, did not resolve the deeper level relationships within the family. In the former case, the lack of resolution could have resulted from fewer data (soft polytomy), and the addition of data could resolve deeper relationships. In the latter case, the addition of extra markers could help resolve these polytomies by breaking up long branches. The concatenation of several markers for a large sampling of Brassicaceae genera, as presented here, did not provide significant extra resolution at the deeper nodes of the tree (fig. 1). Nevertheless, we analyzed the data to test the hypothesis of an early rapid radiation. The LTT plot, presented herein for the first time in the family, indicates that diversification was not constant over time (fig. 2*B*). This is also shown by the rate variable model that was selected as the best-fitting diversification model ($\Delta AIC_{Rv} = 65.36$). Finally, the mean γ-statistic was negative (mean -8.1; 95% CI $= 9.8–6.9$), indicating that rates of diversification decelerated through time (Pybus and Harvey 2000). It would appear as if diversification increased considerably after approximately 32 Ma. This corresponds to the origin of the core Brassicaceae and for the next 10 Ma during which most major lineages evolved (Figs. 2*A* and *B*, grey patch). Indeed, all three lineages within the family (I, II, and II) originated at around the same time (table 3). Finally, at approximately 22 Ma, diversification decelerated. The Brassicaceae is thought to have originated in the Irano-Turanian region, where the highest species diversity is found (Hedge 1976) and where *Aethionema* (ca. 50 spp.), the genus sister to the rest of the family is also most diversified, especially in Turkey (Al-Shehbaz et al. 2006). Moreover, the ancestral area of the Cleomaceae, the sister family of Brassicaceae (Hall, Sytsma, et al. 2002), is also believed to have originated around the Mediterranean or Africa (Inda et al. 2008). At the time of the origin of Brassicaceae (ca. 37 Ma) in the Eocene, a warm and humid climate predominated worldwide (Zachos et al. 2001), including in Turkey (Akgün et al. 2002), with tropical rain forests extending well into Europe (Morley 2003). This would suggest that the Brassicaceae originated as a tropical/subtropical family as its sister families Cleomaceae and Capparaceae that are still largely tropical. A drastic global cooling episode took place between the Late Eocene and Early Oligocene (ca. 33 Ma; the "big chill" or terminal Eocene cooling event), involving the development of permanent continental ice sheets in Antarctica (Coetzee 1993; Zachos et al. 2001). This cooling induced an increase in the deciduous/dry-adapted flora in Europe (Jacobs et al. 1999; Morley 2000) that led to numerous extinctions in 'moister' clades (Morley 2000, 2003). Interestingly, this date also corresponds to the origin of the core Brassicaceae (ca. 32 Ma) and thus with the perceived radiation event in the family (fig. 2*A* and *B* and table 3). Such an event could have first led to the extinction of numerous ancestral tropics-adapted Brassicaceae. This would explain why the core Brassicaceae clade is subtended by a relatively long branch (fig. 2*A*). However, at around 32 Ma, the core Brassicaceae lineage radiated, which apparently did not take place in the Aethionemeae clade. Indeed, diversification rates within the core Brassicaceae range from 0.181 (high level of extinction; $\lambda = 0.9$) to 0.223 (no extinction; $\lambda = 0$). Such levels are comparable with the highest rates found in clades such as Lamiales (0.212, $\lambda = 0$) and Asterales (0.33, $\lambda = 0$) (Magallon and Sanderson 2001). Thus, global cooling, which lasted well into the Oligocene (Jacobs et al. 1999; Morley 2000; Zachos et al. 2001), appears to correlate well with the evolution of some key characters that enabled radiation of the Brassicaceae from approximately 32 Ma onward (fig. 2*B*). Pinpointing such characters will be a challenge in such a morphologically diverse group (Al-Shehbaz et al. 2006); however, as most of this diversity is actually in the fruit morphology, it is not unlikely that for instance the indehiscent mode of fruit opening, is acting only when conditions are favorable (Mühlhausen et al. forthcoming). Adaptation of Brassicaceae to the more arid climates contrasts with its sister families Cleomaceae and Capparaceae that are distributed primarily in the tropical regions with few representatives in the temperate zones. Because of the mid-Tertiary cooling event, the warm and humid environment in which the boreotropical species thrived was replaced by colder and more arid ones, forcing taxa either to migrate to lower latitudes or to face extinction. However, southern migration from Europe to Africa was not possible because of several geographical barriers, such as the Mediterranean Sea and the Alps (Morley 2003). For comparison, taxa that dispersed during the Tertiary via the North Atlantic Land Bridge into North America, before the cooling event, took refuge in the southern part of that continent (Tiffney and Manchester 2001; Morley 2003). Recent evidence would suggest that this scenario is applicable also to the Cleomaceae (Inda et al. 2008). However,

although the Brassicaceae are easily dispersed, it is not possible to determine whether they were present in North America before the cooling event. Donoghue (2008) indicated that it might be easier for plant taxa to migrate than to evolve in situ new gene functions able to cope with the changing environment. In that respect, the Brassicaceae might represent a notable example. One interesting event that took place within the Brassicaceae evolution are whole genome duplications (WGD; Maere et al. 2005) or polyploidizations (Vision et al. 2000; Simillion et al. 2002; Ermolaeva et al. 2003; Henry et al. 2006). Indeed, the *Arabidopsis* genome appears to have undergone at least three distinct WGD, the last one, termed α, occurred between 40 and 30 Ma (Vision et al. 2000; Simillion et al. 2002; Ermolaeva et al. 2003; Henry et al. 2006; Fawcett et al. 2009). It is well established that WGD has played an important role in the evolution and diversification of eukaryotes such as yeasts (Kellis et al. 2004), vertebrates (Ohno 1970; Wang and Gu 2000; Dehal and Boore 2005), and angiosperms (Adams and Wendel 2005; De Bodt et al. 2005; Soltis 2005; Soltis et al. 2009). WGD can provide significant novel sources of genetic material on which mutation, drift, and selection can act, perhaps rendering new evolutionary opportunities relatively fast (De Bodt et al. 2005; Crow and Wagner 2006; Ha et al. 2009; Soltis et al. 2009) and could be associated to rapid diversification in Brassicaceae (Soltis et al. 2009). Indeed, WGD has been shown in recent polyploid Brassicaceae taxa to increase expression diversity of regulatory networks of genes, thus putatively enhancing morphological and adaptive evolution (Ha et al. 2009). Several studies that dated the origin of the last genome duplication in Brassicaceae agree well with the dates presented here (37−32 Ma; tables 1 and 3). The occurrence of genome duplication likely provided a fast way to adapt to climate changes. This duplication, combined with the increase of new available ecological niches, could have favored the rapid adaptive radiation. One prediction would be that this genome duplication likely took place in the core Brassicaceae that radiated but not in the Aethionemeae, a tribe presently included only approximately 50 species. Evidence for such a hypothesis can be drawn from Galloway et al. (1998), who found that the arginine decarboxylase (*adc*) gene family was duplicated in ten core Brassicaceae genera but not in *Aethionema*. However, more extensive studies are needed to confirm this view (Schranz and Mitchell-Olds 2006).

Our data suggest that the ancestor of Brassicaceae originated in a tropical environment and then radiated due to the onset of aridification and global cooling. Such a scenario contrasts with that of Franzke et al. (2009) in which the Brassicaceae originally arose in open and dry areas from tropical humid-adapted Capparaceae/Cleomaceae ancestors. Given the close relationship of Brassicaceae to two primarily tropical plant families, and based on the new age and diversification dates presented here, this later scenario might not be as robust as previously thought. In any case, more familywide data on genome evolution and duplication are needed to adequately test either scenario.

Although our data apparently support an early radiation, it was not robust to the type I error (accepting a hypothesis when it is false) as provided by the γ-statistic. Under a rapid early radiation, we would expect nodes to be more localized at the base of the phylogeny leading to a negative γ-statistic (Pybus and Harvey 2000). In our case, it is not more than that expected under the null model of constant diversification rate (supplementary fig. S2, Supplementary Material online), and using a model designed to take into account the incomplete taxon sampling (Pybus and Harvey 2000). Therefore, our results should to be taken as tentative. To achieve a more robust conclusion and more confident phylogenetic hypothesis, more genera should be included.

Systematics of Brassicaceae

In general, most of the recently recognized tribes (see Al-Shehbazet al. 2006; Al-Shehbaz and Warwick 2007; German and Al-Shehbaz, 2008) are confirmed as monophyletic under extensive generic sampling presented herein. These are Aethionemeae, Schizopetaleae s.l., Eutremeae, Calepineae, Arabideae, Conringieae, Cochlearieae, Heliophileae, Iberideae, Alysseae, Biscutelleae, Euclidieae, Anchonieae, Hesperideae, Chorisporeae, Boechereae, Halimolobeae, Erysimeae, Physarieae, Cardamineae, Lepidieae, Smelowskieae, and Descurainieae.

Many genera, not previously sampled, are assigned to tribes, whereas a few others show new but controversial tribal assignments. Furthermore, many other genera stand alone in separate, well-resolved lineages and, undoubtedly, these ought to be assigned to new tribes. A brief discussion on all these matters is discussed below.

1. *Tribe Malcolmieae:* The present data are in complete agreement with the recent delimitation of the tribe (Al-Shehbaz and Warwick 2007). The genera *Anastatica, Parolinia,* and *Lachnoloma* are assigned herein to this tribe, and they share with the other nine genera characters such as accumbent cotyledons, stellate or sessile trichomes, divided locules, base chromosome number of $x = 11$, or horned fruits.

2. *Tribe Isatideae:* The present tribal delimitation agrees with that of Al-Shehbaz et al. (2006) except for the genera *Glastaria* and *Spirorrhynchus*, neither of which was included in our study. The former genus has not yet been studied, and the latter was assigned by German and Al-Shehbaz (2008) to the tribe Calepineae. The placement herein of the aquatic *Subularia* in the Isatideae is not supported by morphology and needs further study.

3. *Tribe Thlaspideae:* Three genera (*Peltariopsis, Elburzia,* and *Didymophysa*) are added herein to the seven genera previously assigned to this tribe (Al-Shehbaz et al. 2006). The last genus was excluded from the Alysseae (Warwick et al. 2008) but was not reassigned to any other tribe.

4. *Tribe Camelineae*: Our findings show that the tribe is weakly supported and paraphyletic because both tribes Boechereae and Halimolobeae are nested within. These results disagree with the *ndhF* phylogeny of Beilstein et al. (2008) but are in full agreement with Bailey et al. (2006)

and the *phyA* phylogeny of Beilstein et al. (2008), where the Camelineae were not supported as monophyletic. These disagreements highlight the need for further studies, and it is likely that the tribe needs to be divided into a few smaller ones. An alternative solution would be to expand tribe Camelineae.

5. *Tribe Dontostemoneae:* As delimited by Al-Shehbaz and Warwick (2007), the tribe consists of *Clausia* and *Dontostemon*. Our ML and Bayes analyses show that it is paraphyletic, with *Dontostemon* sister to the tribe Chorisporeae. This was however only weakly supported in all analyses.

6. *Tribe Buniadeae:* This monogeneric tribe of two species strongly clusters within lineage III, but its exact placement is not resolved. In the JK analysis, *Bunias* is recovered as sister to *Leiospora* of the tribe Euclidieae.

7. *Tribe Aphragmeae:* This tribe is only moderately recovered under the Bayes analysis (PP = 0.9), and the ML and MP analyses did not provide any support for it. The placement of *Idahoa* herein is new but weakly supported. However, in previous analyses, the position of *Idahoa* was also ambiguous and inconclusive when using the cpDNA *ndhF* and nDNA *phyA* markers (Beilstein et al. 2006, 2008). Obviously, the tribal placement of *Idahoa* requires further studies.

8. *Tribe Noccaeeae:* This tribe received moderate support under both MP and ML analyses but was not recovered under the Bayes one. The close relationship between the Noccaeeae, Cochlearieae, and Conringieae is implied under the Bayes analyses (PP = 0.92) but not under the MP or ML ones.

9. *Tribe Brassiceae:* All previous studies (see Warwick and Sauder, 2005; and references therein) demonstrated that the tribe is monophyletic. The placement herein of *Horwoodia*, *Nasturtiopsis*, and *Sisymbrium* in this tribe is not supported by morphological data, especially the two principal characters on which the tribe is delimited (conduplicate cotyledons and segmented fruits).

10. *Unplaced genera:* Several genera (e.g., *Asta*, *Chamira*, *Cremolobus*, *Kernera*, *Mathewsia*, *Menonvillea*, *Murbeckiella*, *Notothlaspi*, and *Oreophyton*) have not yet been assigned to tribes. Because most of these genera fall in well-resolved lineages distinct from the recognized 33 tribes, new monogeneric or oligogeneric tribes ought to be established. These will be addressed in a separate publication.

Conclusion

A fully resolved phylogenetic framework for the Brassicaceae family will undoubtedly provide a major advancement for the systematics of this taxonomically difficult group (Al-Shehbaz et al. 2006) and will open a new era of evolutionary research on a higher order level. The wealth of genetic and genomic information for the model organism *Arabidopsis thaliana* provide an important comparative framework for studying the molecular and genomic evolution in plants (Hall, Fiebig, et al. 2002; Schranz et al. 2007). Several evolutionary studies in Brassicaceae relied on molecular phylogenies to, for instance, identify suitable model systems (Mitchell-Olds 2001) or test evolutionary hypotheses (chromosome and karyotype evolution, Mandakova and Lysak 2008; evolution of genome size, Lysak et al. 2009).

The supermatrix approach followed here allowed the synthesis of available data as well as generation of the largest and most comprehensive multigene phylogeny of Brassicaceae to date. Such an approach is being widely used for other large plant families (e.g., Arecaceae, Baker et al. 2009, Annonaceae; Chatrou L, personal communication).

Based on 61% generic-level coverage of the family, we found phylogenetic relationships still far from being fully resolved, and in particular with regard to the deeper nodes. We argue that the lack of resolution could be related to a rapid radiation triggered by a combination of genome duplication and climate change. In that case, no data set would be large enough to resolve relationships, although entire genome sequences for a larger number of genera (Eisen and Fraser 2003; Philippe et al. 2005) was recently found to be successful (Moore et al. 2007; Dunn et al. 2008; Hackett et al. 2008). Alternatively, SINEs proved most successful in reconstructing relationships within baleen whale radiations (Nikaido et al. 2006) and thus such homoplasy-free markers could prove useful in Brassicaceae systematics (Tatout et al. 1999; Deragon and Zhang 2005). In parallel, cytogenetic approaches are promising and revealed general patterns of genome evolution (Schranz et al. 2007). A set of carefully selected whole genome sequences over the entire family, combined with robust comprehensive phylogenetic hypothesis and a deep understanding on genome and chromosome evolution, will result in an important in-depth understanding of the evolution of an entire large plant family.

Supplementary material

Figures S1, S2, and S3 and tables S1, S2, S3, and S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank U. Coja, C. Gieshoidt and R. Grupe for technical assistance, Dan Rabosky for his assistance with the program LASER, M. Beilstein for providing DNA samples, and all collectors, herbaria and institutions providing plant material. Marc Sosef is also thanked for logistics. This work was funded by the German Research Foundation (DFG) Mu 1137/7-1, Mu 1137/7-2, and to a minor extent by Ko 2302/11-1.

References

Abdel Khalik KNE. 2002. Biosystematic studies on Brassicaceae (Cruciferae) in Egypt [PhD thesis]. Wageningen University and Research Centre, Biosystematics Group, [Wageningen, The Netherlands].

Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Curr Opin Plant Biol.* 8:135–141.
Akgün F, Akay E, Erdoúan B. 2002. Tertiary terrestrial to shallow marine deposition in central Anatolia: a palynological approach. *Turk J Earth Sci.* 11:127–160.
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, Warwick SI. 2007. Two new tribes (Dontostemoneae and Malcolmieae) in the Brassicaceae (Cruciferae). *Harv Pap Bot.* 12:429–433.
Alfaro ME, Zoller S, Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol Biol Evol.* 20:255–266.
APGII. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot J Linn Soc.* 141:399–436.
Appel O, Al-Shehbaz IA. 2003. Cruciferae. In: Kubitzki K, editor. Families and genera of vascular plants. Springer-Verlag, Berlin (Germany). p. 75–174.
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.
Bailey CD, Price RA, Doyle JJ. 2002. Systematics of the halimolobine Brassicaceae: Evidence from three loci and morphology. *Syst Bot.* 27:318–332.
Baker WJ, Savolainen V, Asmussen-Lange CB, Chase MW, Dransfield J, Forest F, Harley MM, Uhl NW, Wilkinson M. 2009. Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Syst Biol.* 58:240–256.
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.
Brandley MC, Schmitz A, Reeder TW. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Syst Biol.* 54:373–390.
Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland Å, Nikolaev SI, Jakobsen KS, Pawlowski J. 2007. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE.* 2:e790.
Coetzee JA. 1993. African flora since the terminal Jurassic. In: Goldblatt P, editor. Biological relationships between Africa and South America. Yale University Press, New Haven (CT). p. 37–61.
Crespo MB, Lledó MD, Fay MF, Chase MW. 2000. Subtribe Vellinae (Brassiceae, Brassicaceae): a combined analysis of ITS nrDNA sequences and morphological data. *Ann Bot.* 86:53–62.
Crow KD, Wagner GP. 2006. What is the role of genome duplication in the evolution of complexity and diversity? *Mol Biol Evol.* 23: 887–892.
De Bodt S, Maere S, Van de Peer Y. 2005. Genome duplication and the origin of angiosperms. *Trends Ecol Evol.* 20:591–597.
Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3:314.
Deragon JM, Zhang XY. 2005. Short interspersed elements (SINEs) in plants: Origin, classification, and use as phylogenetic markers. Symposium on genome analysis and the molecular systemstics of retroelements. Taylor & Francis Inc, Fairbanks (AK). p. 949–956.
Donoghue MJ. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proc Natl Acad Sci USA.* 105:11549–11555.
Douzery EJP, Snell EA, Bapteste E, Delsuc Fdr, Philippe H. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci USA.* 101: 15386–15391.
Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PloS Biol.* 4:699–710.
Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.
Dunn CW, Hejnol A, Matus DQ, et al. (18 co-authors). 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature.* 452:745–U745.
Eisen JA, Fraser CM. 2003. Phylogenomics: intersection of evolution and genomics. *Science.* 300:1706–1707.
Ermolaeva MD, Wu M, Eisen JA, Salzberg SL. 2003. The age of the Arabidopsis thaliana genome duplication. *Plant Mol Biol.* 51: 859–866.
Farris JS, Källersjö M, Kluge AG, Bult C. 1994. Testing significance of incongruence. *Cladistics.* 10:315–319.
Fawcett JA, Maere S, de Peera YV. 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.
Franzke A, German D, Al-Shehbaz IA, Mummenhoff K. 2009. *Arabidopsis* family ties: molecular phylogeny and age estimates in the Brassicaceae. *Taxon.* 58:425–437.
Galloway G, Malmberg R, Price R. 1998. Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. *Mol Biol Evol.* 15:1312–1320.
Gandolfo MA, Nixon KC, Crepet WL. 1998. A new fossil flower from the Turonian of New Jersey: *Dressiantha bicarpellata* gen. et sp. nov. (Capparales). *Am J Bot.* 85:964–974.
German DA, Al-Shehbaz IA. 2008. Five additional tribes (Aphragmeae, Biscutelleae, Calepineae, Conringieae, and Erysimeae) in the Brassicaceae (Cruciferae). *Harv Pap Bot.* 13:165–170.
Gernhard T. 2008. The conditioned reconstructed process. *J Theor Biol.* 253:769–778.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
Ha M, Kim ED, Chen ZJ. 2009. Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proc Natl Acad Sci USA.* 106:2295–2300.
Hackett SJ, Kimball RT, Reddy S, et al. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science.* 320: 1763–1768.
Hall AE, Fiebig A, Preuss D. 2002. Beyond the Arabidopsis genome: opportunities for comparative genomics. *Plant Physiol.* 129: 1439–1447.
Hall JC, Sytsma KJ, Iltis HH. 2002. Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. *Am J Bot.* 89:1826–1842.
Hedge IC. 1976. A systematic and geographical survey of the Old World Cruciferae. In: Vaughan JG, Macleod AJ, Jones BMG, editors. The biology and chemistry of the Cruciferae. Academic Press, London. p. 1–45.
Heenan PB, Mitchell AD, Koch M. 2002. Molecular systematics of the New Zealand *Pachycladon*(Brassicaceae) complex: generic circumscription and relationship to *Arabidopsis* sens. lat. and *Arabis* sens. lat. *NZ J Bot.* 40:543–562.
Henry Y, Bedhomme M, Blanc G. 2006. History, protohistory and prehistory of the Arabidopsis thaliana chromosome complement. *Trends Plant Sci.* 11:267–273.
Higdon JW, Bininda-Emonds ORP, Beck RMD, Ferguson SH. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evol Biol.* 7.
Ho SYW. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *J Avian Biol.* 38:409–414.

Inda L, Torrecilla P, Catalán P, Ruiz-Zapata T. 2008. Phylogeny of Cleome L. and its close relatives Podandrogyne Ducke and Polanisia Raf. (Cleomoideae, Cleomaceae) based on analysis of nuclear ITS sequences and morphology. *Plant Syst Evol.* 274:111–126.

Jacobs BF, Kingston JD, Jacobs LL. 1999. The origin of grass-dominated ecosystems. *Ann Mo Bot Gard.* 86:590–643.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature.* 428:617–624.

Khosravi AR, Mohsenzadeh S, Mummenhoff K. 2007. Phylogenetic position of *Brossardia papyracea* (Brassicaceae) based on sequences of nuclear ribosomal DNA. *Feddes Repert.* 119:13–23.

Khosravi AR, Mohsenzadeh S, Mummenhoff K. 2008. Analysis of the phylogenetic position of *Acanthocardamum erinaceum* (Brassicaceae) based on ITS-sequences shows that it should be transferred to *Aethionema* as *A. erinaceum* (Boiss.) Khosravi & Mummenhoff. *Nord J Bot.* 26:25–30.

Kishino H, Hasegawa M. 1989. Kishino-Hasegawa (KH) Test: Nonparametric statistical test for comparing two topologies. *J Mol Evol.* 29:170–179.

Koch M, Al-Shehbaz IA. 2002. Molecular data indicate complex intra- and intercontinental differentiation of American Draba (Brassicaceae). *Ann Mo Bot Gard.* 89:88–109.

Koch M, Al-Shehbaz IA, Mummenhoff K. 2003. Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). *Ann Mo Bot Gard.* 90:151–171.

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, Mummenhoff K, Hurka H. 1999. Molecular phylogenetics of Cochlearia (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, 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, Mummenhoff K. 2006. Evolution and phylogeny of the Brassicaceae. *Plant Syst Evol.* 259:81–83.

Linder HP, Hardy CR, Rutschmann F. 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Mol Phylogenet Evol.* 35:569–582.

Lysak MA, Berr A, Pecinka A, Schmidt R, KM, Schubert I. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc Natl Acad Sci USA.* 103:5224–5229.

Lysak MA, Cheung K, Kitschke M, Bureš P. 2007. Ancestral chromosomal blocks are triplicated in Brassicaceae species with varying chromosome number and genome size. *Plant Physiol.* 145:402–410.

Lysak MA, Koch MA, Beaulieu JM, Meister A, Leitch IJ. 2009. The dynamic ups and downs of genome size evolution in Brassicaceae. *Mol Biol Evol.* 26:85–98.

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.

Magallon S, Crane PR, Herendeen PS. 1999. Phylogenetic pattern, diversity, and diversification of eudicots. *Ann Mo Bot Gard.* 86: 297–372.

Magallon S, Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution.* 55:1762–1780.

Magallon SA, Sanderson MJ. 2005. Angiosperm divergence times: The effect of genes, codon positions, and time constraints. *Evolution.* 59:1653–1670.

Mai DH. 1995. Tertiäre Vegetationsgeschichte Europas. Fischer, Jena (Germany).

Malia MJ, Lipscomb DL, Allard MW. 2003. The misleading effects of composite taxa in supermatrices. *Mol Phylogenet Evol.* 27:522–527.

Mandakova T, Lysak MA. 2008. Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). *Plant Cell.* 20:2559–2570.

McMahon MM, Sanderson MJ. 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Syst Biol.* 55:818–836.

Mitchell-Olds T. 2001. *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol Evol.* 16: 693–700.

Moore MJ, Bell CD, Soltis PS, Soltis DE. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci USA.* 104:19363–19368.

Morley RJ. 2000. Origin and evolution of tropical rain forests. John Wiley & Sons, New York.

Morley RJ. 2003. Interplate dispersal paths for megathermal angiosperms. *Perspect Plant Ecol Evol Syst.* 6:5–20.

Morrison DA. 2007. Increasing the efficiency of searches for the maximum likelihood tree in a phylogenetic analysis of up to 150 nucleotide sequences. *Syst Biol.* 56:988–1010.

Mühlhausen A, Polster A, Theißen G, Mummenhoff K. Forthcoming. Evolution of fruit dehiscence in Brassicaceae—examples from *Aethionema* and *Lepidium*. Proceedings of the 5th International Symposium of Brassicas and the 16th Crucifer Genetics workshop. *Acta Horticultura supplement.*

Mummenhoff K, Brüggemann H, Bowman JL. 2001. Chloroplast DNA phylogeny and biogeography of *Lepidium* (Brassicaceae). *Am J Bot.* 88:2051–2063.

Nagpal R, Dar TH, Raina SN. 2008. Molecular systematics of *Brassica* and allied genera in subtribes Brassicinae, Raphaninae, Moricandiinae, and Cakilinae (Brassicaceae, tribe Brassiceae); the organization and evolution of ribosomal gene families. *Bot J Linn Soc.* 157:545–557.

Nikaido M, Hamilton H, Makino H, Sasaki T, Takahashi K, Goto M, Kanda N, Pastene LA, Okada N. 2006. Baleen whale phylogeny and a past extensive radiation event revealed by SINE insertion analysis. *Mol Biol Evol.* 23:866–873.

Nixon KC. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics.* 15:407–414.

Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL. 2004. Bayesian phylogenetic analysis of combined data. *Syst Biol.* 53: 47–67.

Ohno S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin (Germany).

Oldekamp H. 2000. Biosystematische Untersuchungen zur Sippendifferenzierung in mitteleuropäischen *Barbarea*-Arten (Brassicaceae) [Diploma thesis]. [Osnabrück, Germany]: University of Osnabrück.

Peng ZG, Ho SYW, Zhang YG, He SP. 2006. Uplift of the Tibetan plateau: Evidence from divergence times of glyptosternoid catfishes. *Mol Phylogenet Evol.* 39:568–572.

Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005. Phylogenomics. *Ann Rev Ecol Evol Sys.* 36:541–562.

Pirie MD, Chatrou LW, Erkens RHJ, Maas JW, Van der Niet T, Mols J, Richardson JE. 2005. Phylogeny reconstruction and molecular dating in four Neotropical genera of Annonaceae: the effect of taxon sampling in age estimation. In: Bakker FT, Chatrou LW, Gravendeel B, Pelser PB, editors. Plant species-level systematics: new perspectives on pattern and process. Gantner Verlag, Ruggell (Liechtenstein). p. 149–174.

Pirie MD, Humphreys AM, Galley C, et al. (14 co-authors). 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. *Mol Phylogenet Evol.* 48:1106–1119.
Pybus OG, Harvey PH. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc R Soc B—Biol Sci.* 267:2267–2272.
Rabosky DL. 2006. LASER: a maximum likelihood toolkit for detecting temporal shits in diversification rates from molecular phylogenies. *Evol Bioinform Online.* 2:257–260.
Rambaut A. 2002a. TreeEdit version v1.0a10. Available from: http://evolve.zoo.ox.ac.uk/software.html?id=treeedit.
Rambaut A. 2002b. Phy-O-Gen version 1.1. Available from: http://evolve.zoo.ox.ac.uk/software.html?id=phylogen.
Rambaut A, Drummond AJ. 2003. Tracer. Version 1.4. Available from: http://evolve.zoo.ox.ac.uk/.
Ricklefs RE, Losos JB, Townsend TM. 2007. Evolutionary diversification of clades of squamate reptiles. *J Evol Biol.* 20:1751–1762.
Rokas A, Carroll SB. 2006. Bushes in the tree of life. *PloS Biol.* 4: 1899–1904.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572–1574.
Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*. 19:301–302.
Sanderson MJ, Driskell AC. 2003. The challenge of constructing large phylogenetic trees. *Trends Plant Sci.* 8:374–379.
Schranz ME, Mitchell-Olds T. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell*. 18:1152–1165.
Schranz ME, Song B-H, Windsor AJ, Mitchell-Olds T. 2007. Comparative genomics in the Brassicaceae: a family-wide perspective. *Curr Opin Plant Biol.* 10:168–175.
Schulz OE. 1936. Cruciferae. In: Engler A, Harms H, editors. Die natürlichen Pflanzenfamilien. Verlag von Wilhelm Engelmann, Leipzig (Germany). p. 227–658.
Shapiro B, Drummond AJ, Rambaut A, et al. (27 co-authors). 2004. Rise and Fall of the Beringian Steppe Bison. *Science*. 306: 1561–1565.
Shaul S, Graur D. 2002. Playing chicken (*Gallus gallus*): methodological inconsistencies of molecular divergence date estimates due to secondary calibration points. *Gene*. 300:59–61.
Sikes DS, Lewis PO. 2001. PAUPRat: PAUP* implementation of the parsimony ratchet version 1, beta. Distributed by the authors. Department of Ecology and Evolutionary Biology. University of Connecticut, Storrs (CT).
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 USA*. 99:13627–13632.
Simmons MP, Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst Biol.* 49:369–381.
Simpson AGB, Roger AJ, Silberman JD, Leipe DD, Edgcomb VP, Jermiin LS, Patterson DJ, Sogin ML. 2002. Evolutionary history of "Early-Diverging" Eukaryotes: the excavate taxon carpediemonas is a close relative of *Giardia*. *Mol Biol Evol.* 19:1782–1791.
Soltis DE. 2005. Ancient and recent polyploidy in angiosperms. *New Phytol.* 166:5–8.
Soltis DE, Albert VA, Leebens-Mack J, et al. (11 co-authors). 2009. Polyploidy and angiosperm diversification. *Am. J Bot.* 96: 336–348.
Springer MS, Scally M, Madsen O, de Jong WW, Douady CJ, Stanhope MJ. 2004. The use of composite taxa in supermatrices. *Mol Phylogenet Evol.* 30:883–884.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22:2688–2690.
Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol.* 57:758–771.
Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol.* 18:1001–1013.
Swofford DL. 2002. PAUP* Phylogenetic Analysis Using Parsimony (* and other methods), v. 4.0 beta 10. Sinauer Associates, Sunderland (MA).
Tatout C, Warwick S, Lenoir A, Deragon J-M. 1999. SINE insertions as clade markers for wild crucifer species. *Mol Biol Evol.* 16:1614–1621.
Tiffney BH, Manchester SR. 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the northern hemisphere tertiary. *Int J Plant Sci.* 162:S3–S17.
Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in *Arabidopsis*. *Science*. 290:2114–2117.
Wang Y, Gu X. 2000. Evolutionary patterns of gene families generated in the early stage of vertebrates. *J Mol Evol.* 51:88–96.
Warwick SI, Al-Shehbaz IA, Price RA, Sauder C. 2002. Phylogeny of *Sisymbrium* (Brassicaceae) based on ITS sequences of nuclear ribosomal DNA. *Can J Bot.* 80:1002–1017.
Warwick SI, Al-Shehbaz IA, Sauder C, Harris JG, Koch M. 2004. Phylogeny of *Braya* and *Neotorularia* (Brassicaceae) based on nuclear ribosomal internal transcribed spacer and chloroplast trnL intron sequences. *Can J Bot.* 82:376–392.
Warwick SI, Al-Shehbaz IA, Sauder CA. 2006. Phylogenetic position of *Arabis arenicola* and generic limits of *Aphragmus* and *Eutrema* (Brassicaceae) based on sequences of nuclear ribosomal DNA. *Can J Bot.* 84:269–281.
Warwick SI, Al-Shehbaz IA, Sauder CA, Murray DF, Mummenhoff K. 2004. Phylogeny of *Smelowskia* and related genera (Brassicaceae) based on nuclear ITS DNA and chloroplast trnL intron DNA sequences. *Ann Mo Bot Gard.* 91:99–123.
Warwick SI, Francis A, Al-Shehbaz IA. 2006. Brassicaceae: Species checklist and database on CD-Rom. *Plant Syst Evol.* 259:249–258.
Warwick SI, Sauder CA. 2005. Phylogeny of tribe Brassiceae (Brassicaceae) based on chloroplast restriction site polymorphisms and nuclear ribosomal internal transcribed spacer and chloroplast trnL intron sequences. *Can J Bot.* 83:467–483.
Warwick SI, Sauder CA, Al-Shehbaz IA. 2008. Phylogenetic relationships in the tribe Alysseae (Brassicaceae) based on nuclear ribosomal ITS DNA sequences. *Botany-Botanique.* 86:315–336.
Warwick SI, Sauder CA, Al-Shehbaz IA, Jacquemoud F. 2007. Phylogenetic relationships in the tribes Anchonieae, Chorisporeae, Euclidieae, and Hesperideae (Brassicaceae) based on nuclear ribosomal its DNA sequences. *Ann Mo Bot Gard.* 94:56–78.
Wiens JJ. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Syst Biol.* 52:528–538.
Wiens JJ. 2006. Missing data and the design of phylogenetic analyses. *J Biomed Inform.* 39:34–42.
Wiens JJ, Fetzner JW, Parkinson CL, Reeder TW. 2005. Hylid frog phylogeny and sampling strategies for speciose clades. *Syst Biol.* 54:719–748.
Wiens JJ, Moen DS. 2008. Missing data and the accuracy of Bayesian phylogenetics. *J Syst Evol.* 46:307–314.
Wikström N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: calibrating the family tree. *Proc Natl Acad Sci India Sect B (Biol. Sci.).* 268:2211–2220.
Yang YW, Lai KN, Tai PY, Li WH. 1999. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. *J Mol Evol.* 48:597–604.
Yang YW, Lai KN, Tai PY, Ma DP, Li WH. 1999. Molecular phylogenetic studies of *Brassica, Rorippa, Arabidopsis* and allied genera based on the internal transcribed spacer region of 18S-25S rDNA. *Mol Phylogenet Evol.* 13:455–462.

Yule GU. 1924. A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis. *Philos Trans R Soc Lond B*. 213: 21–87.

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.

Zunk K, Mummenhoff K, Hurka H. 1999. Phylogenetic relationships in tribe Lepidieae (Brassicaceae) based on chloroplast DNA restriction site variation. *Can J Bot.* 77:1504–1512.

Zwickl DJ. 2006. Genetic algorithm application for the phylogenetic analyses of large biological sequence datasets under maximum likelihood [PhD thesis]. University of Texas, [Austin (TX)].