A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae

Robert Karl and Marcus A. Koch*

Department of Plant Systematics and Biodiversity, Center for Organismal Studies (COS Heidelberg, Heidelberg University, D-69120 Heidelberg, Germany

* For correspondence. E-mail marcus.koch@cos.uni-heidelberg.de

INTRODUCTION

Rapid radiations with increased rates of diversification, due to increased speciation rates and/or decreased rates of extinction, are generally considered to be influenced by intrinsic traits (so-called key innovations) or extrinsic events (so-called key opportunities) (Hodges, 1997; Moore and Donoghue, 2009; Vamosi and Vamosi, 2011). The emerence of morphological, physiological or behavioural novelties has been linked particularly to subsequent diversification (e.g. floral nectar spurs (Hodges, 1997); flower asymmetry (Sargent, 2004; Kay and Sargent, 2009); dispersal ability and feeding generalization (Phillimore et al., 2006)). In most cases the link between increased diversification and the emergence of a specific trait or event is based on correlations and plausible explanation chains, which is an intuitive and appealing approach (Ree, 2005; Vamosi and Vamosi, 2011), especially when the foundations are laid in a well-resolved phylogenetic framework (Drummond et al., 2012).

Brassicaceae are an angiosperm family in which morphological characters are highly homoplastic and of limited diagnostic value at the level of tribes and sometimes even at the level of genera (Koch and Al-Shehbaz, 2009; Al-Shehbaz, 2012). Consequently, there are few studies attempting to link diversification and speciation in this family with trait or character evolution. Recently, there has been progress in the establishment of a fully resolved phylogenetic tree for Brassicaceae (e.g. Koch et al., 2003, 2007; Al-Shehbaz et al., 2006; Bailey et al., 2006; Beilstein et al., 2008, 2010; German et al., 2009; Couvreur et al., 2010; Warwick et al., 2010). In its current delimitation the Brassicaceae (320 genera with 3660 species) are organized in 49 monophyletic tribes and <3 % of taxa (34 genera with 90 species) remain unassigned to one of these tribes (Al-Shehbaz, 2012) because voucher material either is not suitable for molecular analysis or is unavailable. The systematic backbone of the family is represented by three major evolutionary lineages (Beilstein et al., 2006) with tribe Aethionemeae sister to the rest of the family (Franzke et al., 2011), but many relationships between tribes and core lineages remain uncertain (Warwick et al., 2010; Al-Shehbaz, 2012). Tribe Arabideae belong to an expanded lineage II (Franzke et al., 2009) with various genera being well known, for example, as ornamental plants and often cultivated (e.g. Arabis, Aubrieta or Draba). In this expanded lineage there are another 18 tribes but relationships among these tribes remain uncertain, and only tribe Stevieae has been shown to be sister to Arabideae (German et al., 2009).

There is increasing evidence that the early diversification events in the core family are characterized by rapid radiations (Franzke et al., 2009; Couvreur et al., 2010), and polyploid evolution involving whole genome duplications (WGDs) might have triggered the radiation of many lineages (Lysak et al., 2005, 2009; Koch, 2012; Schranz et al., 2012).

The rapid radiation of Brassicaceae involved colonization and adaptation into newly available ecological niches. This
evolutionary history explains the novel morphological and ecological traits associated with shifts in mating system (Hurka et al., 2005; Ansell et al., 2008; Paetsch et al., 2010; Koelle and Mauricio, 2010), flowering time, flower symmetry and pollination biology (Busch and Zachgo, 2007; Alonso-Blanco et al., 2009), diaspore dispersal and germination (Alonso-Blanco et al., 2009; Linkies et al., 2010), and shifts between annual and perennial growth (Wang et al., 2009). However, all these studies remain speculative linking traits and characters with radiation of the respective lineage, either because of the lack of fully resolved phylogenetic trees or due to speculative uncertainties. Moreover, these studies mainly focused on a single species or genus (for details refer to Franzke et al., 2011). At higher taxonomic levels, the first correlations between trait-evolution and molecular phylogeny were accomplished on trichome branching patterns in the Brassicaceae (Beilstein et al., 2006, 2008) and flower characters have also been analysed in a phylogenetic perspective in tribes Isatideae (Moazzeni et al., 2007) and Brassicaceae (Hall et al., 2011). However, in all of these examples the trait could not be linked to any adaptive evolutionary processes. The lack of adequate experimental and observational studies means that our knowledge of the adaptive significance of these traits is tentative at best.

Reticulate evolution and polyploidization resulting in WGDs (e.g. Lysak et al., 2005, 2009; Soltes et al., 2009; Schranz et al., 2012) are considered as major factors driving diversification. Genomes in Brassicaceae seem highly conserved in structure but not in genome size. The genomes are arranged in defined genomic blocks (Schranz et al., 2006) and polyploidization occurs frequently, increasing block number and total genome size (e.g. Jordon-Thaden and Koch, 2008), but the genomes are stabilized rapidly and are efficiently downsized in most lineages (Lysak et al., 2009). The idea that WGDs initiate speciation by reciprocal gene loss or subfunctionalization (van de Peer et al., 2009) is plausible and it has been demonstrated that WGDs could be associated with the formation of novel, evolutionary important traits (van de Peer et al., 2009; Hoffmann et al., 2012; Schranz et al., 2012 and references therein; Fawcett et al., 2013).

Tribe Arabideae is the largest tribe in Brassicaceae (Al-Shehbaz, 2012), and the clades differ widely in species number and evolutionary putatively important traits such as life-cycle history, geographical and altitudinal distribution, or percentage of polyploid species. Changes in life-cycle strategy might be a particularly important factor in diversification of Arabideae (Jordon-Thaden et al., 2010). The sister relationship between species-poor annual and species-rich perennial clades has been highlighted as an interesting pattern, as it is potentially widespread in Brassicaceae (see Karl et al., 2012) and other plant families (e.g. Ehrhartia, Paeaceae (Verboom et al., 2003); American Lupinus, Fabaceae (Drummond, 2008; Drummond et al., 2012); subtribe Castillejinae, Orobanchaceae (Tank and Olmstead, 2008); and tribe Delphinieae, Ranunculaceae (Jabbour and Renner, 2012)]. Systematics of Arabideae has been challenging as a result of extreme convergence in almost all conceivable morphological characters (Koch et al., 2012a); and most of these characters appear in various combinations in different tribes and evolutionary lineages, making the genus Arabis in its original and past circumscription the most critical taxon of Brassicaceae (Koch et al., 1999, 2000, 2001). The totally revised tribe, currently with 17 genera, can be arranged in several phylogenetically well-defined clades (Karl et al., 2012; Koch et al., 2012a). However, these studies were missing well-resolved relationships between clades, as sequence data [rDNA internal transcribed spacer (ITS) and plastid DNA trnL-trnF] did not resolve the backbone structure of the phylogenetic tree and therefore greatly limited any pilot study in Brassicaceae on diversification in a large monophyletic group.

Here, we aim to present a comprehensive and reliable phylogenetic framework for the tribe, which is achieved by reconstructing a backbone phylogeny focusing on a selected set of nuclear genes. The backbone is used for a tribal phylogenetic reconstruction, including all evolutionary groups and the majority of taxa described so far. The phylogenetic hypothesis is then used to correlate evolutionary processes in space and time. In particular we will highlight the questions of how an increase in speciation rates is linked to life history, and how these rates relate to different elevations in mountainous and alpine habitats, and shifts in ploidy. We present the first complete biogeographical analysis of an entire tribe of Brassicaceae and try to provide insights about its origin, which is crucial for an understanding of the family.

MATERIAL AND METHODS

Taxon sampling and phylogenetic data set composition

To resolve the backbone structure of the phylogenetics of tribe Arabideae we generated a data set of three nuclear loci from 16 representatives of the various clades and groups in the tribe: the ITS regions 1 and 2 of nuclear encoded ribosomal RNA (hereafter ITS region), the chalcone synthase (Chs) gene consisting of two exons and one intron, and the alcohol dehydrogenase (Adh) genome consisting of seven exons and six introns. For this data set Pseudoturriritis turrita from the sister tribe Stevenieae was chosen as an outgroup taxon (e.g., Koch et al., 2012a). Hereafter, this nuclear data set is referred to as the ‘three-marker data set’.

All further analyses were performed with the ITS data set containing a total of 312 accessions. Detailed information about the various accessions is summarized in Supplementary Data Table S1. Additionally, for the same 312 accessions a data set was generated based on the plastid trnL intron-trnL-trnF intergenic spacer region (hereafter named trnL-F). These two data sets include members of all 17 Arabideae genera and 305 different species. This makes these data sets the most comprehensive ones for the whole tribe. Itemized, the data sets comprise 71 species (74 accessions) of Arabis, 13 species (14 accessions) of Aubrieta, 197 species and accessions of Draba and 24 species from the remaining and smaller 14 Arabideae genera (Abdra, Arcyosperma, Athysanus, Baimashania, Borodiniosis, Botschantzevia, Dendoarabid, Drabella, Pachyneurum, Parryodes, Pseododraba, Scapiarabis, Sinoarabis and Tomostima). In addition, both data sets include three taxa from the sister tribe Stevenieae (Macropodium pterospermum, Pseudoturriritis turrita, and Stevenia canescens) as outgroups (Koch et al., 2012a). The annotated alignments for the three data sets are provided in Supplementary Data Figs S1–3 (three-marker data set: S1; ITS region: S2; and trnL-F: S3). For chs and adh the alignments have been trimmed for coding regions only to perform subsequent phylogenetic analysis.
Distribution ranges, life cycles and elevations/habitats of the respective species were carefully compiled from various floras by the authors (Supplementary Data Table S2). In case of the Draba species we employed a data compilation presented earlier (Jordon-Thaden, 2010).

DNA extraction, PCR amplification and sequencing

DNA extractions were carried out according to a slightly modified protocol (Karl et al., 2012) of the CTAB method (Doyle and Doyle, 1987). PCR reactions to amplify the selected regions were performed in a final volume of 25 \( \mu \)L, using 10 \( \mu \)M of each primer, a total of 2.0 mM MgCl\(_2\) and 0.5 U of MangoTaq polymerase (Bioline, Luckenwalde, Germany). The following primers were used: ITS-18F (5′-GGAAGGAGAAGTCGTAAACAAGG-3′), as modified by Mummenhoff et al. (1997); ITS-25R (5′-TCCTCGGTTATTTAGATGTC-3′), designed by White et al. (1990); trnL-for (5′-CGAAAATCGTGAAGCTACCG-3′), published as primer ‘c’ in Taberlet et al. (1991); trnL-rev (5′-GGGATAGGGAGACTTGAAC-3′), published as primer ‘d’ in Taberlet et al. (1991); trnL-Fior (5′-GGTTCATAAGTTCTCATCACC-3′), published as primer ‘e’ in Taberlet et al. (1991); and trnL-F-rev (5′-GATTATGACTGCTTCTGCTCA-3′), designed by Dobeš et al. (2004). Chalcone synthase sequences were first amplified with primers CHS-PRO1-fw (5′-CATCTGGCCCTCATTACAAACCCTACC-3′) and CHS-EX2-TERM-rev (5′-TTAGAAGGAGGCTCGTCAAGACGCT-3′) (Koch et al., 2000). Following our initial sequencing results, we subsequently designed a new and sequence-optimized primer pair ARA-N-CHS-1-for (5′-GGCA CARAGAGCTGATGGA-3′) and ARA-N-CHS-5-rev (5′-AGAGAAGTGAAGACRACWCG-3′), which is more specific for the orthologue copy specific to tribe Arabideae. Alcohol dehydrogenase sequences were amplified with the primer pair ADH-FOR-1 (5′-ACTACTGGTCAGATTATTGCAGTC-3′) and ADH-REV-7 (5′-AGACACCTGGTATGATGC-3′), as designed by Koch et al. (2000). For some accessions it was not possible to amplify the whole fragment at one stretch. In these cases the fragment was amplified in two parts and additionally the primers ADH-REV-4 (5′-CTAACCCAGTAGATAAACCACAAC-3′) and ADH-FOR-4 (5′-GTAGTGTTGTGTTGTTGTTGTAATTACACTGA-3′) (Koch et al., 2000) were utilized. The respective \( \text{cht} \) and \( \text{adj} \) fragments were purified and then cloned into the chemically competent Escherichia coli strain JM110 using the pGEM-T vector system (Promega, Madison, WI, USA). Five positively tested clones (PCR test) of each accession were sequenced using the universal T7 (5′-TATACGACTCACTATAGGG-3′), and SP6 (5′-ATTTAGTGACACTATAGAA-3′) primers (GATC, Konstanz, Germany; or MWG Eurofins, Ebersberg, Germany).

Further sequences for ITS and \( \text{trnL-F} \) were added from our past studies (Jordon-Thaden et al., 2010; Koch et al., 2010, 2012a; Karl et al., 2012).

Phylogenetic reconstruction of the Arabideae backbone structure

A Bayesian Markov chain Monte Carlo (MCMC) analysis (Yang and Rannala, 1997) was performed for the ‘three-marker data set’ using MrBayes v. 3.1.2 (Ronquist and Huelsenbeck, 2003). The data set was divided into three unlinked partitions (\( \text{adj}, \text{cht} \) and ITS) to apply the best fitting nucleotide substitution model to each locus, which were chosen using Modeltest v. 3.7 (Posada and Crandall, 1998). In the following analysis four simultaneous runs were performed with four chains each for 1 million generations, and in each run 1001 trees were saved. The first 25 % of these trees were discarded as burn-in. The temperature of the heated chain was set to 0.01, as this facilitated the most efficient chain-swapping. Finally, posterior probabilities of all splits between the respective runs were compared and cumulative split frequency plots were analysed using the web-based program AWTY (Nylander et al., 2007) to confirm accomplished convergence between the individual Bayesian MCMC runs.

Constraint nodes and ITS and \( \text{trnL-F} \) phylogenetic analyses

In order to assign the well supported backbone structure of Arabideae to the single marker analyses, we applied a constrained topology to the phylogenetic analyses of the ITS data set based on the nuclear encoded three-marker set. We did not apply the constrained topology to the analyses of the \( \text{trnL-F} \) region, because a different genome is represented and because nuclear and plastidic genomes have proved to show different modes of molecular evolution in Arabideae (see Karl et al., 2012; Koch et al., 2012a).

To set up the constrained topology four nodes from the backbone of Arabideae were constrained, according to the phylogenetic reconstruction of the three-marker data set (see Fig. 1): (1) Constrained node 1 (CN1) constrains the monophyly of \( \text{A. alpina} \) and \( \text{A. nordmanniana} \); (2) CN2 constrains CN1 and the \( \text{A. auriculata} \) clade as a monophyletic group; (3) CN3 constrains CN2, \( \text{Aubrieta} \) (including \( \text{Arabis verna} \)), \( \text{Draba} \) (and its segregates) and the \( \text{A. aucteri} \) clade as a monophyletic group; and (4) CN4 constrains CN3 and the main Trabes clade as a monophyletic group.

These constraints also correspond to the finding of our previous studies in the Arabideae (Karl et al., 2012; Koch et al., 2012a), making these constraints a highly reliable phylogenetic framework. However, further nodes, which connect two or more clades, remained unconstrained. We did not constrain the \( \text{Aubrieta} \) clade as sister to \( \text{Draba} \) and the \( \text{A. aucteri} \) clade because of the low posterior probability value (prr < 0.90). Moreover, the position of the \( \text{A. aucteri} \) clade as sister of \( \text{Draba} \) also remained unconstrained because of the incongruency to the plastid DNA tree (Supplementary Data Fig. S4). This pattern is indicative of reticulate evolution and has been demonstrated to have happened in at least two clades of Arabideae (Jordon-Thaden et al., 2010; Koch et al., 2012).

Dated molecular phylogeny

Time-measured phylogenetic analyses of the ITS and \( \text{trnL-F} \) data sets were performed using the software package BEAST v. 1.6.2, (Drummond and Rambaut, 2007), which uses Bayesian MCMC to reconstruct phylogenetic trees. We performed the divergence time estimate calculations with a secondary calibration point for the tribe Arabideae introduced by Couvreur et al. (2010), which sets the age of the Arabideae between 10.0 and 23.9 Mya (mean 16.8 Mya) and fits well with our last estimates (Koch et al., 2012a). In the case of Arabideae, note that applying different divergence time estimates for the family (e.g. Beilstein et al., 2010) has
only very limited effect on estimates for the Arabideae (Koch et al., 2012a). For a more detailed discussion on this topic the interested reader might refer to Franzke et al. (2011). The calculations for the ITS data set used the above-mentioned constraints, whereas the calculations for the trnL-F data set were performed without any constraints. Taxon subsets were specified for Arabideae (all taxa excluding Macropodium pterospermum, Pseudoturritis turrita and Stevenia canescens from tribe Stevenieae) and all clades of the tribe. A lognormal uncorrelated relaxed clock method was applied and a birth–death speciation model (Gernhard, 2008) was selected as tree prior. Four independent runs with 50 million generations were run for each marker on the freely available Bioportal server (www.bioportal.uio.no). In each run 10 001 generations were sampled.

The individual runs were checked in Tracer v.1.5 (Rambaut and Drummond, 2007) as having reached a stationary phase and a sufficient effective sample size (ESS > 200). All runs (log- and trees-files, respectively) were then combined with LogCombiner v.1.6.2. Finally, a single maximum clade credibility (MCC) tree with median branch lengths was produced for each marker, after burning in the first 10 000 trees using TreeAnnotator v.1.6.2.
Lineages through time plots

Lineages through time (LTT) plots were constructed with the above-mentioned single MCC trees, using the ape package (Paradis et al., 2004) of the R software environment (Ihaka and Gentleman, 1996). For each marker (ITS and trnL-F) three LTT plots were constructed. The first included the whole data set with all accessions. For the second LTT plot nearly all species of Draba and its annual segregates (Tomostima clade) were omitted, because this lineage represents >70 % of all species in tribe Arabideae and this high number might bias conclusions at the tribal level. Therefore, one single accession (D. aizoides) was included to represent the stem of Draba and the Tomostima clade. In a third LTT plot, and to differentiate between annuals and perennials, only the annual species from the total data set were considered.

Calculation of diversification rates

Clade-specific net diversification rates were estimated with BayesRate v. 1.4 (Silvestro et al., 2011). The program takes into account uncertainties of divergence times by estimating speciation and extinction rates on a distribution of trees. Furthermore, BayesRate can account for incomplete taxon sampling when computing the clade-specific diversification rates. This overcomes the problem that only half of the Draba species were included, but all other Arabideae genera were nearly complete in the LTT analyses.

Prior to the analyses, 1000 random trees were sampled from the 30 004 post burn-in trees from BEAST with the sub-sample trees feature implemented in BayesRate. For 12 specified clades and groups (see Table 2) speciation rates (S), extinction rates (E), net diversification rates (S – E) and the extinction fraction (a = E/S) were estimated under a birth–death model. Among the specified groups are two groups, which represent the annual and perennial members of Arabideae, respectively. For each marker (ITS and trnL-F) 100 million MCMC iterations (1000 runs with 100 000 iterations each) were run with a burn-in of 10 %. Results were checked and visualized in Tracer v.1.5.

Ancestral area reconstructions (AARs)

Distinct geographical regions were carefully specified and thereby roughly orientated on terrestrial ecozones and their actual outline and boundaries followed largely the major habitat types and ecoregions of the ‘WWF ecoregions 200’ map (Olson et al., 2001). However, the maximum number of discrete geographical units was restricted to 15 by the limitation of the RASP software (Yu et al., 2011), and in some cases adjustments had to be made to accommodate the actual distribution ranges of species and their allocation to the individual geographical regions. Finally, 14 geographical regions were defined (see Fig. 2 and Supplementary Data Table S3): North and Central America (A), South America (B), Iberian Peninsula & North Western Africa (C), Central Mediterranean & Southern Balkan (D), Europe (E), Western and Central Anatolia & Levantine coast (F), Caucasus, Eastern Anatolia & Iranian mountain ranges (G), high mountains of the Arabian peninsula and Eastern Africa (H), Central Asian mountain ranges (I), Eastern Himalaya and Tibet-Chinese mountains (J), Eastern Asia (K) and Siberia & Russian Far East (L). A more detailed characterization of these regions is given in Supplementary Data Table S3.

Reconstructions of ancestral areas, i.e. the distribution area of the ancestor of a monophyletic group (Bremer, 1992), were performed using the program RASP 2.0b, which has both the Statistical DIVA (S-DIVA, Yu et al., 2010) and a Bayesian binary MCMC (BBM) approach implemented. For both approaches 10 000 trees were randomly sampled from the 30 004 post burn-in tree topologies of the BEAST analyses and frequencies of ancestral ranges were calculated and averaged over the sampled trees. The possible ancestral ranges are shown for every node present in the respective MCC trees. In the S-DIVA analysis the maximum number of regions per node was set to 3. This resulted in 300 possible area states, of which 49 area states were included; the remaining 251 area states, which contained non-adjacent regions, were excluded from the S-DIVA analyses.

In the BBM analyses four chains were run for 30 million generations. Every 30 000 generations ancestral area states were sampled. The maximum number of regions was set to 5 without any exclusion of area states, allowing more widespread and non-adjacent area states. The distribution of the virtual outgroup was set to ‘outgroup’. Stevenia canescens and Macropodium pterospermum represent their respective genera in respect of distribution and therefore are stated with a distribution range which exceeds the distribution range of the single species. Additionally, both analyses were performed for a dataset where Arabis caerulea was pruned, to test for the impact of its unresolved position at the base of the Arabideae phylogenetic tree.

Ancestral character state reconstructions

Ancestral character state analyses were run for life cycle strategy (annual versus perennial) and elevation of habitat, using the BBM approach of the program RASP 2.0b. Settings were made analogous to those of the AAR, but the maximum number of states was set to 2.

For the reconstructions of life-cycle strategy annual species were labelled with ‘A’, and perennial ones with ‘B’. Species described as either annuals or perennials were labelled with ‘AB’, indicating both life cycle strategies were possible. Some species, which are described to have an annual life cycle, can grow as short-lived perennials under favourable conditions (Arabis borealis, A. plantsilquai subsp. nemorensis, Draba albertina, D. bifurcata, D. brachystylis, D. corrugata, D. mogollonica and D. stenoloba). Nevertheless, they are stated as annuals. Stevenia canescens represents the genus Stevenia, in which both life cycles are present, and it is therefore labelled with the equivocal character ‘AB’.

For the reconstructions of habitat elevations taxa were labelled as lowland species (A) if they occur generally in lowland habitats and/or ascend to low montane elevations at most. Accordingly, taxa which occur generally in high montane/alpine habitats were labelled as alpine species (B). Occasional occurrences in higher or lower elevations were not taken into account. In contrast, taxa which generally occur both in lowland and in montane/alpine habitats were labelled with the equivocal character state ‘AB’ (e.g. arctic-alpine Arabis alpina, Draba aurea or D. fladnizensis). Stevenia canescens is labelled with the combined character ‘AB’ as it represents the whole genus, which includes several lowland species and one alpine species in Tibet.
Fig. 2. Maximum clade credibility (MCC) tree from the divergence time estimation in BEAST based on ITS sequence information containing 309 taxa of tribe Arabideae. The respective time scale is given below the tree. Vertical bars on the right show the affiliation to different clades of Arabideae. Unassigned taxa are not covered by these bars. The four constrained nodes (nodes 3, 8, 9 and 10, see Fig. 1) are indicated in orange font. For 20 nodes (nodes 1–20) the results of the ancestral area reconstructions [S-DIVA (left) and BBM (right, black-rimmed)] are displayed as pie charts, representing the possible ancestral ranges and their probabilities. Colour codes of specified geographical regions are: A: North & Central America, B: South America, C: Iberian peninsula & NW Africa, D: Central Mediterranean & Southern Balkans, E: Europe, F: Western and Central Anatolia & Eastern Mediterranean, G: Caucasus, Eastern Anatolia & Iranian mountain ranges, H: high mountains of the Arabian peninsula and Eastern Africa, I: Central Asian mountain ranges, J: Eastern Himalaya and Tibet-Chinese mountains, K: Eastern Asia, L: Siberia & Russian Far East. For the two most frequent multi-region areas (FG and CDEF) additional colour codes were specified. Black slices combine all regions with a probability <5 %. All other slices, which do not refer to one of the above mentioned areas, are alternately indicated in white or light grey. Their respective distribution area is given in some cases, when the area is the most probable area for a node or has a probability of 25 % or more.
A comprehensive list presented as Supplementary Data Table S4 registers the applied distribution ranges and character states for life cycle and habitat.

**RESULTS**

**Dataset information and phylogenetic trees**

All new DNA sequences have been submitted to NCBI and accession codes are provided with Supplementary Data Table S1. The alignment of the ‘three-marker data set’ consisted of 2880 characters, of which 2224 were constant, 245 were autapomorphic and 411 were potentially parsimony-informative (Supplementary Data Fig. S1). As most appropriate evolutionary models, the TrN + I + Γ model was applied to the adh partition, the TrNef + I + Γ model to the chs partition and the GTR + I + Γ model to the ITS partition.

The resulting tree is presented in Fig. 1, and had a harmonic mean for the ln likelihood of $-10\, 482.59$ and an average standard deviation of split frequencies of 0.005. Contrary to previous phylogenetic studies in tribe Arabideae, the tree and especially its backbone is fully resolved and all nodes are strongly supported by posterior probability values ($ppr > 0.9$). The Arabis alpina clade (A. montbretiana, A. purpurea and two accessions of A. alpina) and A. nordmanniana are sistered by the A. auriculata clade (A. auriculata and A. nova), Draba and A. aucheri are sister to the Aubrieta clade (including Arabis verna). However, this sister group relationship has a lower support from posterior probability values ($ppr = 0.77$). The ‘three-marker data set’ also revealed the position of the main Arabis clade (A. collina and A. hirsuta) as sister to both groups mentioned above (A. alpina/A. nordmanniana/A. auriculata clades and Draba/A. aucheri/Aubrieta clades), and the early branching position of Sinoarabis setosifolia in Arabideae was confirmed.

The ITS alignment consisted of 656 characters, of which 318 were constant, 99 were autapomorphic and 239 were potentially parsimony-informative (Supplementary Data Fig. S2). As most appropriate evolutionary model, the GTR + I + Γ model was applied to the analyses.

The resulting tree is presented in Fig. 2 and its general outline and the respective positions of the clades are similar to the ‘three-marker tree’, mostly due to the four constrained backbone nodes. The sole exception is the position of the A. aucheri clade, which is sister to the A. alpina/A. nordmanniana/A. auriculata clades instead of being sister to Draba. Whereas the ‘three-marker tree’ is informative in resolving the Arabideae backbone structure, the ITS phylogenetic tree, based on a comprehensive data set, is highly indicative for the composition and internal structure of the individual clades. Nevertheless, support for all above-mentioned clades remained maximal in the comprehensive phylogenetic trees (Supplementary Data Figs S4 and S5). The results are congruent with previous studies in Arabideae (Draba – Jordon-Thaden et al., 2010; main Arabis – Koch et al., 2010; A. alpina clade – Karl et al., 2012; Scapiarabis clade – Koch et al., 2012a) and could affiliate numerous taxa to their respective clades. Moreover, the presented phylogenetic tree could verify several patterns which had been previously suggested: in particular, the inclusion of the monotypic genera Dendroarabis and Pachyneurum in the main Arabis clade (as mentioned in German et al., 2009) and the monophyly of Aubrieta as sister to Arabis verna (as indicated in Karl et al., 2012). Furthermore, it could be shown that Draba and its annual segregates (i.e. Abдра, Athysanus and Tomostina) form two, distinct and well-separated clades in the tribe (see Jordon-Thaden et al., 2010). Posterior probability (ppr) values for the individual nodes are given with Supplementary Data Fig. S5.

The combined trnL–F alignment (Supplementary Data Fig. S3) originally consisted of 1234 characters, but positions 1–5, 533–581 and 1209–1234 were excluded from the analyses due to the number of missing data. Furthermore, positions 219–407, a large indel in the trnL intron, which is only present in the two outgroup taxa Macropodium and Stevenia, were also excluded from the analyses. The final trnL–F alignment consisted of 965 characters, of which 638 were constant, 113 were autapomorphic and 214 were potentially parsimony-informative. As the most appropriate evolutionary model the TVM + I + Γ model was applied to the analysis.

The phylogenetic backbone of Arabideae remains unresolved in the trnL–F tree (presented in Supplementary Data Fig. S4), as most of the nodes only have negligible support from posterior probability values ($ppr < 0.75$). There are some further incongruencies between the plastid DNA (trnL–F) tree and the nuclear ‘three-marker tree’, mainly the structure of the Scapiarabis clade (see Koch et al., 2012a) and the position of the A. aucheri clade. However, both the sister-group relationship of the A. nordmanniana clade to the A. alpina clade and the sister-group relationship between Aubrieta and Draba are also displayed in the trnL–F tree.

**Divergence time estimates and LTT plots**

The evolution of the Arabideae tribe is placed in the Neocene, with its most probable origin in the mid Miocene ($\approx 15$ Mya). Divergence time estimates of the individual clades indicate that they originated in the late stages of the Miocene (6–8 Mya – Draba, main Arabis clade) or later in the Pleistocene or Pliocene. Hence, diversification in the different Arabideae clades took place in the last 5 Myr. The results of the divergence time calculations are listed in Table 1 and the phylogenetic tree is presented in Fig. 2. Mean values and confidence intervals for all major clades are congruent with previous studies of the tribe (Koch et al., 2010, 2012a; Karl et al., 2012).

Divergence times of most clades (e.g. main Arabis, Aubrieta, Arabis alpina or Draba) tend to be somewhat earlier in the analysis running the plastid trnL-F marker, but only two clades (Arabis auriculata clade, Scapiarabis clade) show considerable higher ages analysing trnL–F, which corresponds with their different position in the trnL–F tree compared with the ITS tree.

The LTT plots (Fig. 3) for the ITS marker indicate two periods of increased lineage accumulation during the evolution of Arabideae. Numbers of new lineages increased rapidly in the early stages of the evolution of Arabideae (Mid Miocene, $15–10$ Mya) and more explicitly starting from the Early Pliocene (approx. 5 Mya). Between both periods the increase of new lineages is delayed. The dashed line indicates the number of lineages without the Draba radiation (Fig. 3). It is apparent that the second increase of lineage accumulation starts earlier (approx. 7–8 Mya), if Draba and its annual segregates are fully included. The red line indicates the lineage accumulation of the annual species exclusively. The increase of new annual lineages is
**Table 1. Parameters and results of the divergence time estimate calculations using secondary calibration of node age of tribe Arabideae (Couvreur et al., 2010)**

<table>
<thead>
<tr>
<th>Parameters and results of the divergence time estimate calculations using secondary calibration of node age of tribe Arabideae (Couvreur et al., 2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Based on nDNA ITS</strong></td>
</tr>
<tr>
<td>Runs/generations</td>
</tr>
<tr>
<td>log likelihood</td>
</tr>
<tr>
<td>t_{mrca} of the tribe Arabideae (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of constrained node CN1 (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of constrained node CN2 (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of constrained node CN3 (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of constrained node CN4 (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of Draba and the Tomostima clade (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of the core Draba (w/o Draba hederifolia and Draba verna) (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of Arabis (including Arabis verna) (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of the Arabis alpina clade (including its annual sister species) (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of the Arabis norbmnnana clade (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of the Arabis auriculata clade (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of the Arabis aucheri clade (Ma)</td>
</tr>
<tr>
<td>t_{mrca} of the main Arabis clade (Ma)</td>
</tr>
<tr>
<td>ucd.mean</td>
</tr>
<tr>
<td>meanRate</td>
</tr>
<tr>
<td>N_diversification rates</td>
</tr>
</tbody>
</table>

Mean values are indicated in bold; the lower and upper bounds of the 95% highest posterior density (HPD) interval are given before and after that value, respectively. Enforced bounds are indicated in italics.
incorporated in Fig. 2. The original result files from the ancestral area reconstruction analyses are given with Supplementary Data Tables S5a and b.

Incongruencies between both analyses are present for node 1 (origin of Arabideae) and node 15 (Draba and its annual segregates). For all other nodes the results of both analyses are either congruent (e.g. nodes 5, 11, 18, 19 or 20) or inclusive (e.g. nodes 3, 12 or 16).

The S-DIVA analysis suggests that Arabideae (node 1) originated in Europe, whereas the BBM analysis favours a Central Asian origin (region I). Besides, further Asian distributions (regions K, L and IL) and a European distribution (region E) are mentioned with probabilities >5%. Ancestral area states for node 1 are shifted towards Europe in the BBM analysis, when a different setting for the virtual outgroup (null or wide) is chosen. However, the effect of different virtual outgroup settings is negligible for all other nodes. Moreover, the vague origin of the tribe might be affected by the early branching position of the European A. caerulea, as its exclusion from the analyses increases the probabilities of an Asian origin (regions I (49.88%), IK (13-45%) and IL (13-27%)) in the BBM analysis. For the S-DIVA analysis an ancestral area in the eastern Irano-Turanian region is then indicated [regions GI (51-35%), FGI (20-67%) and EGI (19-82%)]. Generally, the exclusion of A. caerulea does not affect the remaining nodes, but in the S-DIVA analysis the most probable ancestral areas for some early diverging nodes (nodes 2, 3 and 12) are shifted slightly towards the Eastern Mediterranean (regions D or DF, respectively). However, these differences do not change the common picture of early distribution of Arabideae.

All nodes that diverged during the Mid- and Late Miocene (nodes 2, 3, 8, 9, 10 and 12) have their most likely ancestral areas is Anatolia and/or the (Eastern) Mediterranean. In some cases the distribution range is spreading into the adjacent Iranian/Caucasus region (nodes 2, 10 and 12) or northwards into the European region (node 2). In the BBM analysis for several nodes (nodes 3, 8, 9 and 12) the Western Mediterranean (region C) is included into the most likely ancestral areas, a region which is not mentioned for these nodes in the S-DIVA analysis.

For node 15 (Draba and its annual segregates) the S-DIVA analysis indicates a North American origin (region A, 48%), but four

### Table 2. Net diversification rates (speciation rate S – extinction rate E) and the extinction fraction (a = E/S) for 12 clades and groups of Arabideae calculated with BayesRate v. 1.4

<table>
<thead>
<tr>
<th>Clade/Groups</th>
<th>Net diversification rates</th>
<th>Extinction fraction (%)</th>
<th>Net diversification rates</th>
<th>Extinction fraction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ITS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Arabideae</td>
<td>0.283 (0.262)</td>
<td>0.981 (0.890)</td>
<td>0.363 (0.339)</td>
<td>0.806 (0.727)</td>
</tr>
<tr>
<td>Arabideae</td>
<td>0.171 (0.132)</td>
<td>0.542 (0.453)</td>
<td>0.208 (0.182)</td>
<td>0.550 (0.458)</td>
</tr>
<tr>
<td>Draba</td>
<td>0.079 (0.058)</td>
<td>0.280 (0.212)</td>
<td>0.158 (0.134)</td>
<td>0.471 (0.385)</td>
</tr>
<tr>
<td><em>Scapiarabis</em> clade</td>
<td>0.011 (0.008)</td>
<td>0.036 (0.028)</td>
<td>0.014 (0.011)</td>
<td>0.041 (0.032)</td>
</tr>
<tr>
<td>Perennials of <em>A. alpina</em> clade</td>
<td>0.023 (0.018)</td>
<td>0.075 (0.060)</td>
<td>0.024 (0.019)</td>
<td>0.076 (0.061)</td>
</tr>
<tr>
<td>A. <em>nordmanniana</em> clade</td>
<td>0.007 (0.006)</td>
<td>0.023 (0.019)</td>
<td>0.008 (0.007)</td>
<td>0.024 (0.019)</td>
</tr>
<tr>
<td>Annual clades</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Scapiarabis</em> clade</td>
<td>0.010 (0.007)</td>
<td>0.030 (0.024)</td>
<td>0.013 (0.011)</td>
<td>0.033 (0.027)</td>
</tr>
<tr>
<td>Tomostima clade</td>
<td>0.009 (0.006)</td>
<td>0.027 (0.022)</td>
<td>0.010 (0.008)</td>
<td>0.028 (0.023)</td>
</tr>
<tr>
<td>A. <em>auriculata</em> clade</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>A. <em>aucheri</em> clade</td>
<td>0.023 (0.018)</td>
<td>0.073 (0.060)</td>
<td>0.025 (0.020)</td>
<td>0.076 (0.061)</td>
</tr>
<tr>
<td><em>Arabideae</em></td>
<td>0.177 (0.137)</td>
<td>0.563 (0.453)</td>
<td>0.200 (0.168)</td>
<td>0.586 (0.471)</td>
</tr>
<tr>
<td><em>Arabis</em></td>
<td>0.136 (0.106)</td>
<td>0.454 (0.365)</td>
<td>0.154 (0.124)</td>
<td>0.473 (0.383)</td>
</tr>
<tr>
<td>Annuals</td>
<td>0.039 (0.034)</td>
<td>0.155 (0.124)</td>
<td>0.040 (0.035)</td>
<td>0.156 (0.125)</td>
</tr>
<tr>
<td>perennial</td>
<td>0.209 (0.193)</td>
<td>0.760 (0.682)</td>
<td>0.210 (0.196)</td>
<td>0.770 (0.688)</td>
</tr>
<tr>
<td><strong>trnL-F</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annuals</td>
<td>0.062 (0.056)</td>
<td>0.217 (0.180)</td>
<td>0.066 (0.059)</td>
<td>0.221 (0.183)</td>
</tr>
<tr>
<td>Tomostima clade</td>
<td>0.009 (0.007)</td>
<td>0.027 (0.021)</td>
<td>0.010 (0.008)</td>
<td>0.028 (0.023)</td>
</tr>
<tr>
<td>A. <em>auriculata</em> clade</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>A. <em>aucheri</em> clade</td>
<td>0.023 (0.018)</td>
<td>0.074 (0.060)</td>
<td>0.025 (0.020)</td>
<td>0.076 (0.061)</td>
</tr>
<tr>
<td><em>Arabideae</em></td>
<td>0.177 (0.137)</td>
<td>0.563 (0.453)</td>
<td>0.200 (0.168)</td>
<td>0.586 (0.471)</td>
</tr>
<tr>
<td><em>Arabis</em></td>
<td>0.136 (0.106)</td>
<td>0.454 (0.365)</td>
<td>0.154 (0.124)</td>
<td>0.473 (0.383)</td>
</tr>
<tr>
<td>Annuals</td>
<td>0.039 (0.034)</td>
<td>0.155 (0.124)</td>
<td>0.040 (0.035)</td>
<td>0.156 (0.125)</td>
</tr>
<tr>
<td>perennial</td>
<td>0.209 (0.193)</td>
<td>0.760 (0.682)</td>
<td>0.210 (0.196)</td>
<td>0.770 (0.688)</td>
</tr>
</tbody>
</table>

Mean values are indicated in bold, median values in parentheses; the lower and upper bounds of the 95% highest posterior density (HPD) interval are given before and after these values, respectively.

* Large 95% HPD intervals make these estimates less reliable.
areas with a distribution from the eastern Mediterranean to Iran have a combined probability of approx. 30%. The BBM analysis suggests an ancestral area around the Mediterranean (region CDEF, CDF or CEF) for this node. For core Draba (node 16, excluding D. verna and D. hederifolia) both analyses give similar results, indicating an ancestral distribution in the Irano-Turanian floristic region (regions GI, FGI, FG or I). However, in the BBM analysis the most likely ancestral area has a probability of only 13%. Ancestral areas for the three different groups in core Draba (see Jordon-Thaden et al., 2010) are relatively clear and seem unaffected by the reticulate evolution of several species in the genus (Jordon-Thaden et al., 2010). For core Draba group I (node 17), region G (Caucasus/Iran) is indicated as the most likely ancestral distribution in both analyses. The BBM suggests that this distribution might have been extended eastwards into Central Asia (region I). As most probable ancestral area, North America (region A) is shown for core Draba group II (node 19) and eastern Asia (regions J or IJ) for core Draba group III (node 18).

The S-DIVA analysis indicates a broad Euro-Iranian distribution (region EG) as ancestral area for the main Arabis clade (node 4), but the BBM analysis specifies it more to south-eastern Europe (regions D, DE or E). Node 6 (Arabis hirsuta agg.) has its most probable distribution in Europe (region E), whereas for node 7 (North American and Asian Arabis species) a Central Asian distribution is suggested (region I) in both analyses. Analogous to Draba, reconstruction of ancestral areas of these two major groups in the clade was not affected by species with a hybrid history between these groups (see Koch et al., 2010). In summary, the S-DIVA analysis suggests 338 dispersal, 107 vicariance and 19 extinction events, whereas the BBM analysis suggests a slightly higher number of dispersal events (368 dispersal, 69 vicariance and six extinction events).

Reconstruction of changes in life cycle and habitat preferences

Perennialism is indicated as the ancestral life-cycle strategy for Arabideae (Fig. 4, node 1). The results of the ancestral life-cycle reconstructions are graphically presented for 20 nodes of particular interest in the left part of Fig. 4. In addition, the probabilities of the respective nodes are presented in Supplementary Data Table S6. This strategy of perennialism is maintained for node 2, but in all following early-diverging nodes (nodes 3, 8, 9, 10, 12 and 15) annualism is indicated as the most probable ancestral life-cycle strategy, whereas the probability for node 10 (includes the perennial A. nordmanniana clade and the A. alpina clade) is ~50%. However, the probabilities for the remaining five nodes are high, varying between 79.8% (node 3) and 97.0% (node 8). With the exception of nodes 11 (A. alpina clade) and 13 (annual A. verna and Aubrieta), all major nodes that emerged in the Late Miocene or later have high probabilities for perennialism as the ancestral life-cycle strategy. Most prominently, these are the main Arabis clade (nodes 4–7) and core Draba (nodes 16–19).

Fig. 4. Graphical presentation of the results of the ancestral state reconstructions concerning life cycle (left part) and habitat elevation (right part). The results are superimposed on the BEAST MCC trees of the ITS analysis. Probability values are presented as pie charts for 20 main nodes (nodes 1–20). Uncoloured pies refer to the equivocal character states (AB), whereas single character states (A or B) are indicated in their respective colours (annual life cycle = blue, perennial life cycle = red; lowland habitats = green, montane/alpine habitats = brown). Branches are indicated in the above-mentioned colours, if the subsequent node had a probability of >50% for the respective character state. Analogous nodes are indicated in black, if none of the single character states had a probability >50%.
Therefore, the annual life cycle of few species in these clades (e.g. *A. planisiliqua* subsp. *nemorensis*, *D. brachystylis* or *D. nuda*) can be considered as a derived character.

The results of the ancestral reconstructions for habitat elevation are graphically presented for 20 nodes of particular interest in the right part of Fig. 4. In addition, the probabilities of the respective nodes are shown in Supplementary Data Table S6. Ancestral species of Arabideae (nodes 1 and 2) occurred either in high montane/alpine elevations or were distributed in both lowland and alpine habitats, as the character states B and AB are indicated with similar high probabilities. Comparable to the life-cycle strategy reconstructions, for nearly all following early-diverging nodes (nodes 3, 8, 9, 12 and 15) lowlands are indicated as most probable ancestral habitats. With the exception of nodes 13 and 14 (*Aubrieta*clade), for all other nodes montane/alpine elevations are indicated as ancestral habitats. Although there is quite a number of species in the main *Arabis* clade and *Draba*, which occur in lowland or slightly higher elevations, this also applies for these two clades (nodes 4–7 and 16–19).

**DISCUSSION**

*Introductionary comments on the phylogeny and systematics of the Arabideae*

The presented study includes 309 taxa of the Arabideae and presents the most comprehensive phylogenetic analysis of the tribe so far. The included taxa make up >60 % of the currently recognized species of Arabideae. Most of the excluded taxa belong to the species-rich genus *Draba* (about 390 species in total; Al-Shehbaz, 2012), which leaves the remaining 16 genera of Arabideae with much higher taxon coverage (80–100 %). The excluded *Draba* species are predominantly perennials occurring in montane to subnival elevations in North and South America, Central Asia, Europe or the Caucasus, and there is no further indication that they should fall outside their generic boundaries.

The previously proposed (e.g. Koch et al., 2012a) existence of eight evolutionary separate clades (*Draba* and its annual segregates, *Aubrieta*clade, *Arabis alpina* clade, *Arabis norbmanniana* clade, *Arabis auriculata* clade, *Arabis aucheri* clade, main *Arabis* clade and *Scapiarabis* clade) in Arabideae is significantly confirmed herein. However, to accommodate the most recent taxonomic adjustments (Al-Shehbaz, 2012) *Draba* and its annual segregates (the *Tomostima* clade) are regarded as distinct clades. This is concordant with our results and previous studies on *Draba* (Jordon-Thaden et al., 2010). There are only two species of Arabideae which remain unaffiliated to one of the clades (*Arabis caerulea* and *Drabelea muralis*) and we have no indication from morphology and/or molecular data that any of the missing taxa should form another relevant clade in the tribe. Thus, this study provides a robust phylogenetic framework of Arabideae, applicable to both detailed taxonomic work and subsequent evolutionary and comparative studies.

The aim to restore monophyly for all genera has been achieved for 16 out of the 17 genera with the latest taxonomic adjustments from 2012 (Al-Shehbaz, 2012; Koch et al., 2012a). The sole exception is the genus *Arabis*, the paraphyletic circumscription of which is demonstrated. The majority of *Arabis* species are grouped into the ‘main *Arabis*’ clade (55 species analysed, 65 species to be expected to belong to this group altogether).

However, another 20 species (among them *Arabis alpina*, the type species of the genus) are grouped into four other clades, which are closely related to other genera such as *Aubrieta* and *Draba*.

As a consequence, a monophyletic circumscription of *Arabis* could be retained only when the species in the same clade as the generic type, *A. alpina*, are included in a newly circumscribed *Arabis*. However, this would leave the species of the main *Arabis* clade (with well-known species such as *A. blepharophylla*, *A. hirsuta* and *A. procurrens*) in need of a new generic designation. Nonetheless, *A. verna* and *A. caerulea* are in need of new taxonomic placements with the first species best integrated into *Aubrieta*. Finally, the monotypic genera *Dendroarabis* and *Pachyneurum* have been circumscribed and maintained as distinct genera with distinguishing morphological characters (German and Al-Shehbaz, 2008; German et al., 2012), although they cluster clearly in the main *Arabis* clade and their taxonomic inclusion into this group is worthy of consideration.

**Biogeography and origin of the Arabideae clades**

Reconstructions of past geographical patterns revealed that all early diversification events in Arabideae during the Mid and Late Miocene took place in the (western) Irano-Turanian floristic region and/or the (eastern) Mediterranean. Region F (Anatolia and the Levantine coast), in particular, is included as the most likely ancestral area for almost all nodes, whereas the Western Mediterranean, the Caucasus and the Iranian mountains are indicated as potential centres of origin with lower frequencies (Figs 2 and 5). These early diversification events gave rise to all of the presently existing clades and their respective stems. These results are congruent with past hypotheses placing not only the origin of Arabideae into the Irano-Turanian region (Hedge, 1976), but also the origin of the entire family. Species diversity in Brassicaceae is not distributed equally (Koch and Kiefer, 2006); the most important diversification centres are found in the Irano-Turanian region (approx. 150 genera and approx. 900 species with 530 endemics) and the Mediterranean region (ca. 113 genera and approx. 630 species with 290 endemics). Adjacent regions in North America (ca. 99 genera and 778 species with 600 endemics) and in the Sahara-Sindian region (ca. 65 genera and 180 species with 62 endemics) show a significant reduction in species diversity (Hedge, 1976; Al-Shehbaz, 1984; Appel and Al-Shehbaz, 2003). This reduction of species diversity is continued in southern regions (South America with 40 genera and 340 species; Southern Africa with 15 genera and at least 100 species; Australia and New Zealand with 19 genera and 114 species) (Allan, 1961; Marais, 1970; Hewson, 1982; Al-Shehbaz, 1984; Appel and Al-Shehbaz, 2003). This overall distribution pattern might also provide some evidence for the origin of the family in the Irano-Turanian region and follows the basic ideas of Hedge (1976) who claimed that the origin of Brassicaceae occurred in a region encompassing the Mediterranean to the Irano-Turanian territory. In summary, tribe Arabideae could well serve as a representative monophyletic group mirroring crucifer diversification and evolution.

The geomorphological outline of Eurasia during the late Miocene is comparable to today, as major barriers such as the Turgay strait or a continuous Tethys ocean no longer existed (Tiffney and Manchester, 2001; Meulenkamp and Sissingh,
Furthermore, the alpine orogeny was already in progress and the resulting mountain ranges formed a continuous belt from the Balkans to Central Asia (Steininger and Rögl, 1984; Meulenkamp and Sissingh, 2003), which was only temporarily disconnected by narrow gateways (Jolivet et al., 2006) or shallow basins during phases of marine transgressions (Piller...
congruent, especially for all nodes that diverged during the early phases of the evolution of Arabideae, which was accessible also for non-alpine species. Furthermore, it has served as a geographical cradle for the early evolution of Arabideae and the clades within the tribe, which have all evolved in this region or more confined parts of it. In the case of Draba (node 16) the suggested origin of the genus in region G (Iran/Caucus) is congruent with its previously proposed origin in the ‘Turkey—Caucasus—Iran region’ based on trnl-F haplotype statistics (Jordon-Thaden, 2010). Similarly, for the Arabis alpina clade (node 11) the suggested origin in region F (Western and Central Anatolia & Levantine coast) is consistent with previous studies, which determined the origin of Arabis alpina in Central Anatolia based on the distribution of the ancestral trnl-F haplotypes (Koch et al., 2006; Ansell et al., 2011; Karl et al., 2012).

Major migratory events into regions outside the ancestral distribution range of Arabideae did not start before the Pliocene. However, such migratory events occurred in two clades only (Draba, main Arabis clade), if we exclude the recent expansion of Arabis alpina with its migration to Eastern Africa approximately 0.5 Mya (Koch et al., 2006) and its postglacial extension into arctic regions (Koch et al., 2006; Ehrich et al., 2007). Draba, especially core groups II and III, has colonized vast areas in Asia, the Americas and Arctic regions. Jordon-Thaden (2010) proposed migration routes from its Turkey—Caucus—Iran origin to the Mediterranean (group I) and Central Asia (ancestor of groups II and III). North and eventually South America were then colonized from Asia via Beringia (group III) (Koch and Al-Shehbaz, 2002). However, several more recent migratory events within Draba contributed to increased hybridization and the reticulate ties between species from different core groups (Koch and Al-Shehbaz, 2002).

Early range expansions of the main Arabis clade involved migration to Central Europe, whereas the subsequent colonization of East Asia and North America did not occur before the onset of Pleistocene cycles of glaciation and deglaciation (Koch et al., 2010). Furthermore, species of the main Arabis clade did not colonize high mountain areas and alpine habitats in the American Cordillera or the Himalayas as many Draba species did.

Driving factors of evolutionary diversification

As outlined above, distribution and colonization patterns indicate tremendous past success of various clades in Arabideae. Consequently, the question arises if there are traits and characters obviously correlated with colonization and diversification.

It is intriguing that, when superimposed on the phylogenetic trees, the patterns of life cycle and habitat elevation are almost congruent, especially for all nodes that diverged during the early phases of the evolution of Arabideae in the Mid Miocene (Fig. 4). This close correlation between life-cycle strategy and preferred habitat elevation expands the previously raised idea of the pattern of sister relationships between species-poor, annual clades and a larger groups of predominantly perennial species. In addition, the perennial species are prevalent in high montane to alpine elevations, whereas the annual species mainly occur in lowland habitats.

In Arabideae this pattern occurred at least four times: (1) the relationship between Draba (approx. 390 species, perennial dominated) and its annual segregates (Tomostima clade) with about ten species (split 8–11 Mya in the Mediterranean or the Irano–Turansien region); (2) the annual (three species) and perennial (about seven species) members of the Arabis alpina group (split 2.7–3.3 Mya in Anatolia; see Karl et al., 2012); (3) the annual species A. verna as sister to the perennial genus Aubrieta with approx. 15 species (split 2.7–5.2 Mya in the Eastern Mediterranean); and (4) the annual A. auriculata and A. aucheri clades (together five species) as sister groups to the A. nordmanniana and A. alpina clades (together approx. 15 species), which are both dominated by perennials (split 10–13 Mya in Anatolia or the Mediterranean).

Despite the generally species-richer perennial clades, the first, but more moderate phase of increase of lineage accumulation in Arabideae in the Mid Miocene was caused mostly by the formation of annual, lowland lineages. In contrast, the second, more pronounced phase with increase of new lineages, which started in the Pliocene, was almost exclusively caused by the formation of numerous perennial, montane/alpine lineages. During the first phase in the Mid Miocene climatic conditions in the Mediterranean region and neighbouring regions in the East changed from subtropical to more temperate temperature regimes (Flower and Kennett, 1994; Böhme, 2003). Moreover, East Antarctic ice sheet growth and polar cooling also had large effects on global carbon cycling and on the terrestrial biosphere, including aridification of mid-latitude continental regions (Flower and Kennett, 1994; van Dam, 2006). In concordance with life-cycle theory (Cole, 1954) annuals generally seem better adapted to arid conditions with erratic water supply (Schaffer and Gadgil, 1975) in the lowlands. However, these annual lineages remained relatively species-poor compared with the greater diversification in the perennial lineages. The perennials colonized montane and alpine habitats, which in general provide moderate and cooler climatic conditions and a more stable water supply from rainfall and condensation due to orographic effects (Roe, 2005) and are therefore considered better suited for the needs of perennials (Cole, 1954; Schaffer and Gadgil, 1975; Evans et al., 2009). Moreover, montane habitats generally provide a greater abiotic diversity, due to topographic complexity and climatic and ecological gradients (Brown, 2001; Badgley, 2010). As a result, the expansion into such habitats has provided multiple opportunities for diversification, due to the creation of a variety of new ecological niches and the existence of effective barriers to gene flow (Rahbek and Graves, 2001; Hughes and Eastwood, 2006).

In this respect, the switch between life cycle (annual to perennial or vice versa) must be considered as an important evolutionary key innovation. Key innovations are seen as adaptations that enabled the respective taxa to overcome ecological or migratory range restriction (Simpson, 1953). The change in life-cycle strategy made it possible for the perennial lineages to colonize
### TABLE 3. Summary of Arabideae and allocation of evolutionary important features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Core Draba clade</th>
<th>Main Arabis clade</th>
<th>Aubrieta clade</th>
<th>Scapiarabis clade</th>
<th>A. alpina clade</th>
<th>A. nordmanniana clade</th>
<th>Tomostima clade</th>
<th>A. auriculata clade</th>
<th>A. aucheri clade</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of species</td>
<td>≈390</td>
<td>≈70</td>
<td>≈15</td>
<td>12</td>
<td>≈10</td>
<td>≈8</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Perennial life cycle</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ancestors present in high montane/alpine environments</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Major geographical range expansion (i.e. migration from centre of origin)</td>
<td>X</td>
<td>X</td>
<td>–</td>
<td>–</td>
<td>(X)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hybridization</td>
<td>X</td>
<td>X</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>High polyploidization (&gt;4×)</td>
<td>X</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Asexual reproductive modes observed</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Putative geographical origin (crown group)</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Putative geographical origin (stem group)</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Putative time of origin (crown group)</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Putative time of origin (stem group)</td>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

If the feature occurs regularly in a clade it is marked with ‘x’, whereas if it occurs less frequently it is marked with ‘(x)’. If the feature does not occur or only rarely occurs in a clade it is marked with ‘–’.

Sources for the collected information are the data compilations of the authors (see Supplementary Data Table S2) and Jordon-Thaden (2010), as well as the knowledge and resource database for Brassicaceae BrassiBase (Koch et al., 2012b). Information about putative areas and periods of origin for the clades is derived from Fig. 2, Table 1 and Supplementary Data Fig. S4.
higher elevations, and to diversify in these new habitats. The close correlation between both features has also been pointed out for other plant taxa, for example for American Lupinus species, where the switch from an annual to perennial life cycle led to a range expansion into montane/alpine habitats (Drummond et al., 2012), or the genus Anthemis, where the shift from wetter mountain areas to drier coastal plains was accompanied also by a switch from a perennial to annual life cycle (Lo Presti and Oberprieler, 2009).

Usually it is predicted that a key innovation has subsequently caused increased net diversification after the overcoming of previous restrictions (Simpson, 1953; Maynard Smith and Szathmary, 1995; see also Hunter, 1998). These key innovations could, for example, affect mechanisms linked to speciation, e.g. reproductive isolation (Kay et al., 2006) or ecological opportunity (Yoder et al., 2010). Several studies have linked this emerging ecological opportunity to subsequent macroevolutionary diversification and adaptive radiation (e.g. Losos, 2009; Marazzi and Sanderson, 2010; see also Fawcett et al., 2013). Likewise, in the case of American Lupinus the colonization of alpine habitats was followed by increased diversification rates (Drummond et al., 2012), highlighting the shift in life-cycle strategy as an adaptive key innovation. Similar results were also obtained for the perennials in Draba (Jordon-Thaden and Koch, 2008).

By contrast, the actual impact of a single trait is often hard to correlate (Kay et al., 2006), as these traits have underlying correlations with other traits (Vamosi and Vamosi, 2011). Consequently, several studies have indicated that the emergence of a novel adaptive trait (even if it is considered as a key innovation) is not sufficient to directly spur diversification (Levinton, 1988; Davies et al., 2004). Although focusing on major genome duplication events (e.g. as a result of hybridization, reticulation and polyploidization) rather than key innovations, Santini et al. (2009) and Schranz et al. (2012) showed the possibility that a great proportion of diversification is not directly connected to genome duplication and its newly established adaptive key traits. Instead extrinsic events such as episodes of biogeographical or climatic change, formation of new habitats, or biotic interactions have contributed to subsequent diversifications in these groups.

Apart from the emergence of key innovations, dispersal into new environments can also provide ecological opportunity (Yoder et al., 2010), making dispersal events and the extent of geographical distribution ranges important factors in respect of diversification. Arabidaceae represents an excellent case in this regard. The tribe contains several well-resolved sister-group relationships of annual/lowland and perennial/alpine clades and these clades differ remarkably in species number (approx. 10 to 390 species). Species numbers are especially divergent in the perennial/alpine clades, which should all hold potential for subsequent diversification, if we consider the switch to a perennial life cycle and the ability to colonize montane habitats as key factors for diversification success. Within Arabidaceae all perennial/alpine clades originated in the Irano-Turanian floristic region or the adjacent parts of the Mediterranean. However, the clades which stayed mostly in their original distribution areas (Arabis alpina clade – Eastern Mediterranean, Arabis nordmanniana clade – Caucasus, Aubrieta – Eastern Mediterranean, Scapiarabias clade – Central Asian and Chinese mountain ranges) remained species-poor, whereas the species-richer clades (Draba, main Arabis clade) have undergone multiple migratory events (Jordon-Thaden, 2010; Koch et al., 2010). Analogous, the WGD radiation lag-time hypothesis of Schranz et al. (2012) involves one or more dispersal events following the evolution of key novel traits, which first instigates the diversification of the respective group, whereas its sister group remains relatively species-poor in the original distribution area. This pattern indicates the importance of dispersal and colonization success (i.e. geographical range expansions) as a vital factor to maintain high rates of diversification (see Fritz et al., 2011). As an example, for the Scapiarabias clade, which has a comparable age to Draba and the main Arabis clade, but explicitly lowered net diversification rates, it has been discussed that its inability to colonize the high alpine zones (>4000 m a.s.l.) blocked effective dispersal across the mountain chains of the QTP. Migratory routes were then possible only along mountain chains and finally restricted their habitats to a ring around the QTP between the xeric lowlands and the high alpine elevations (Koch et al., 2012a). The causal chain with perennialism as as first key innovation followed by altitudinal range expansion and secondary diversification after geographical range expansion is suggested by the annual Tomostima clade, which has colonized vast lowland areas in America (outside the geographical cradle of Arabidaceae), but remained relatively species-poor.

As a result it can be concluded that in most cases increased diversification cannot be attributed to a single event, e.g. the emergence of a novel adaptive trait. Instead such a key innovation must encounter the adequate environmental context, so that the respective taxonomic group will diversify (see De Queiroz, 2002). In this respect it has been demonstrated that it might be necessary to invade new habitats to play out fully the diversification potential of the acquired key innovation (see also von Hagen and Kadereit, 2003; Drummond et al., 2012).

The highly fragmented habitats of the American Cordillera and the Himalayas, which are home to numerous endemic Draba species, might therefore have contributed to vast diversification in this genus. The diversification in Draba was highly affected by hybridization, polyploidization and alternative reproductive modes (Jordon-Thaden and Koch, 2008), but the genus might have also experienced hybridization between evolutionary lineages from different continents in its early evolutionary history (Koch and Al-Shelbaz, 2002). Hybridization and polyploidization are among the most important factors in speciation (Solís et al., 2009; Solís and Solís, 2009), and in particular for Draba high levels of polyploidy (>4 x) are prevalent, especially for arctic or alpine species (Jordon-Thaden and Koch, 2008). Note here that arctic species of Arabidaceae belong almost exclusively to Draba, and here the percentage of perennial polyploids is highest (Jordon-Thaden and Koch, 2008; Jordon-Thaden et al., 2009). Among the few exceptions is Arabis alpina, which has a wide distribution including arctic regions and is a diploid perennial (Karl et al., 2012). Generally, polyploidy is considered beneficial to maintain taxa in extreme habitats, such as unstable environments (Fawcett et al., 2009), or alpine altitudes and arctic latitudes (Brochmann et al., 2004). In comparison, the main Arabis clade has a considerable number of tetraploids, but higher polyploids are the rare exception (octoploid A. eschscholziana). In all other clades of Arabidaceae (independent of being annual or perennial dominated) diploid species prevail with some rare tetraploids (e.g. Arabis parvula or Aubrieta scardica). Similarly,
hybridization events in the main Arabis clade are set in more recent time frames (see Koch et al., 2010). For the other clades of Arabideae there are no signs that hybridization had played a significant role in promoting species diversity. Moreover, asexual modes of reproduction such as apomixis are another way to impede admixture of derived taxa with their recent ancestors. Within Arabideae apomixis is only known from Draba; members of the other clades are either facultative selfers or outcrossers. In summary, hybridization and polyploidization have played an important role in the evolutionary history and colonization of arctic-alpine habitats in one particular genus, namely Draba, which is the most species-rich clade. For all other clades this is less obvious, and additional data might help to find further correlations.

CONCLUDING REMARKS

Utilizing a well-resolved and comprehensive phylogenetic tree for Arabideae and analyses performed subsequently, we could (1) corroborate a more clear-cut phenomenon for Brassicaceae in the relationship between annual- and perennial-dominated clades (Karl et al., 2012), and (2) reconstruct a biogeographical and evolutionary scenario for Arabideae, which is summarized diagrammatically in Fig. 5.

Early diversification in the tribe was mostly due to the emergence of new annual lineages, which were prevalent in lowland habitats of the Mediterranean and the Irano-Turanian region. The Irano-Turanian region has been suggested as the cradle of Brassicaceae (Hedge, 1976; Franzke et al., 2011) and this study has revealed further evidence to support the evolutionary importance of this region. However, annual lineages remained relatively species-poor during the following epochs and only the switch towards a perennial life-cycle strategy and the ability to colonize high montane/alpine habitats increased diversification in these respective lineages. Table 3 summarizes the nine clades of Arabideae with their presence/absence of features, which are considered as important factors of diversification. In particular, the switch to a perennial life-cycle strategy must be considered a key innovation, as it is the prerequisite for the altitudinal shift and subsequent diversification success. This type of key innovation has been documented for various angiosperm groups (e.g. Apiaceae, Sun et al., 2004; Orobanchaceae, Tank and Olmstead, 2008; Pontederiaceae, Barret and Graham, 1997). Therefore, this transition could be considered an important evolutionary phenomenon more widely than in Brassicaceae alone, and there is a widely held opinion among plant evolutionary biologists that annuals are derived from perennial ancestors and that the shift between these two strategies is unidirectional (e.g. Stebbins, 1957).

However, it was shown that this intrinsic factor did not directly spur the considerably accelerated diversification during the Pliocene and Pleistocene, which was mainly achieved by the emergence of new perennial lineages. Instead, the ability of lineages to colonize new environments, occupy opening niches and adapt to new conditions during the constant change of climatic conditions is emphasized to have driven most of this ‘secondary’ and more recent diversification. In this respect range expansion via migration but also hybridization and polyploidization are again indicated as major factors for plant diversification.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Figures S1–S3: annotated alignments for the three-marker data set (S1), ITS data set (S2) and trnL-F data set (S3). Figure S4: MCC tree from the divergence time estimation in BEAST based on the trnL-F data set with posterior probability values for all nodes with more than 5% support. Figure S5: MCC tree from the divergence time estimation in BEAST based on the ITS data set with posterior probability values for all nodes with more than 5% support. Table S1: internal code numbers, herbarium sources and NCBI GenBank numbers of all accessions used in the study. Table S2: floras, publications and online databases accessed for the Arabideae compilation of distribution, character states and ploidy levels. Table S3: detailed characterization of the geographical regions for the AAR analyses. Table S4: detailed information about distribution, character states and ploidy levels of the 312 accessions used in the ITS and trnL-F datasets. Table S5a: original result file from RASP of the AAR analysis using BBM. Table S5b: original result file from RASP of the AAR analysis using S-DIVA. Table S6: probability values of the ancestral character state reconstructions for the 20 main nodes presented in Fig. 2.

ACKNOWLEDGEMENTS

We thank the gardener team of the Botanical Garden Heidelberg and in particular Torsten Jakob for continuous support with plant cultivation. Thanks go also to Susanne Ball for her substantial help in the laboratory and handling all the material and germplasm collections and to Peter Sack, who organized and helped with the voucher exchanges. We appreciate also a critical discussion and many comments on the final draft version provided by Graham Muir and various re-determinations done by Dmitry German. We are very grateful to Ilhsan Al-Shehbaz, who continuously provided comments for many of the various clades and groups of taxa dealt with. We also thank the curators of the herbaria of the Arnold Arboretum of Harvard University (A), the Botanical Garden and Botanical Museum Berlin (B), the Natural History Museum London (BM), University of Copenhagen (C), Agriculture Canada, Ottawa (DAO), Royal Botanical Garden Edinburgh (E), Gray Herbarium of Harvard University (GH), Botanical Garden and Herbarium Heidelberg (HEID), University of California, Berkeley (JEPS & UC), Royal Botanical Gardens Kew (K), Botanical Museum Lund (LD), V. L. Komarov Botanical Institute, St Petersburg (LE), Botanische Staatsammlung Munich (M), Real Jardin Botanico Madrid (MA), the Missouri Botanical Garden (MO), Botanical Museum Oslo (O), University Osnabrück (OSBU), Uppsala University (UPS), Smithsonian Institution, Washington, DC (US), University of Utah, Salt Lake City (UT), Natural History Museum Vienna (W), University of Natural Resources and Life Sciences, Vienna (WHB), and the Herbarium of the University of Washington, Seattle (WU), for providing voucher material. Moreover, plant or seed material was received from various fellow scientists: Dmitry German (COS, University Heidelberg), Christiane Kiefer (MPIPZ Cologne), Birol Mutlu (Inönü University Malatya) and Ji-Pei Yue (Kunming Institute of Botany). This work was supported by DFG research grants (KO2302 11/1, 12/1 and 13/1) to M.A.K.


