Mutational dynamics and phylogenetic utility of plastid introns and spacers in early branching eudicots

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt

der Mathematisch-Naturwissenschaftlichen Fakultät
der Technischen Universität Dresden

von

Dipl.-Biol. Anna-Magdalena Barniske

geboren am 19. März 1976 in Halle/Saale

Eingereicht am 26.10.2009
Verteidigt am 16.12.2009

Die Dissertation wurde in der Zeit von 02/2005 bis
10/2009 im Institut für Botanik angefertigt

1. Gutachter: Prof. Dr. Christoph Neinhuis, Dresden
2. Gutachter: Prof. Dr. Dietmar Quandt, Bonn

to my grandmother

Table of contents

Acknowledgements

This thesis wouldn't have been possible without the support and interaction of several persons, first of all Prof. Dr. Dietmar Quandt and Prof. Dr. Christoph Neinhuis. Prof. Quandt gave me the opportunity to participate in his project on the evolution of early-diverging eudicots. He supported me in taking delight in a subject new to me. Prof. Neinhuis has offered me the opportunity to be a part of his research group and to conduct the work not only in a modern well-equipped laboratory but also in a warm atmosphere. Both encouraged, supported and helped me especially in difficult situations.

Furthermore I like to thank Prof. Dr. Thomas Borsch, Prof. Dr. Kai Müller, Andreas Worberg, in particular Karsten Salomo and everyone from the eudicots-evolutionary-research group for all kind of help including support in the laboratory or with calculations when needed and proof reading as well as many helpful discussions.

I whish to express my gratitude to the Botanical Garden Dresden in particular the custodian Dr. Barbara Ditsch, for providing fresh plant material in an incomplex way. Further supply of plant material from various colleagues as well as the Botanical Gardens Bonn, Ghent and Talca is gratefully acknowledged.
Additionally I want to warmly thank the whole plant phylogenetics working group who made my work much more pleasant. I enjoyed being a member of this group and benefited a lot from working with the colleagues. My special thanks to Ursula Arndt and Sylvi Malcher for taking care of the official tasks, Volker Buchbender, Dr. Sanna Olsson and Lars Symmank for many fruitful and motivating discussions as well as Dr. Stefan Wanke for his support.

Above all I thank my family and friends for supporting and encouraging me over the years.

This thesis was embedded in the project „Mutational dynamics of non-coding genomic regions and their potential for reconstructing evolutionary relationships in eudicots" (grants BO1815/2 and QU153/2) supported by the Deutsche Forschungsgemeinschaft. The funding of the project is kindly acknowledged.

Introduction

Introduction – The early-diverging eudicots

During the last twenty years major progress has been made towards a better understanding of phylogenetic relationships among angiosperms. An early broad-scale molecular-phylogenetic analysis on the basis of *rbcL* sequence data (Chase & al., 2003; compare Figure 1) clearly revealed three major groups, with eudicots as well as monocots being monophyletic, arisen from a paraphyletic group of "basal" dicotyledonous angiosperms. A number of molecular investigations have consistently recovered the eudicotyledonous clade and increased confidence in its existence (e.g. Savolainen & al., 2000a; Qui & al., 2000; Soltis & al., 2000; Hilu & al., 2003; Kim & al., 2004).With about 200,000 species the eudicot clade contains the vast majority of angiosperm species diversity (Drinnan & al., 1994). As they are characterised by the possession of tricolpate and tricolpate-derived pollen the eudicots have also been called the tricolpate clade (Donoghue & Doyle, 1989). Based on the use of sequence data several lineages, such as Ranunculales, Proteales, Sabiaceae, Buxaceae plus Didymelaceae, and Trochodendraceae plus Tetracentraceae were identified as belonging to the early-diverging eudicots (= "basal eudicots"), while larger groups like asterids, Caryophyllales, rosids, Santalales, and Saxifragales were revealed as being members of a highly supported core clade, the so called "core eudicots" (Chase & al., 1993; Savolainen & al., 2000b; Soltis & al., 2000; 2003; Hilu & al., 2003; Worberg & al., 2007). Furthermore Gunnerales were shown to be the first-branching lineage within core eudicots, having a sistergroup relationship with the remainder of the clade (e.g. Soltis & al., 2003; Worberg & al., 2007).

However, the exact branching order among the several lineages of the eudicots remained difficult to resolve. This thesis is to a great extent concentrated on resolving relationships among the different clades of the early-branching eudicots as well as on clarifying phylogenetic conditions inside distinct lineages, based on phylogenetic reconstructions using sequence data of fast-evolving and non-coding molecular regions.

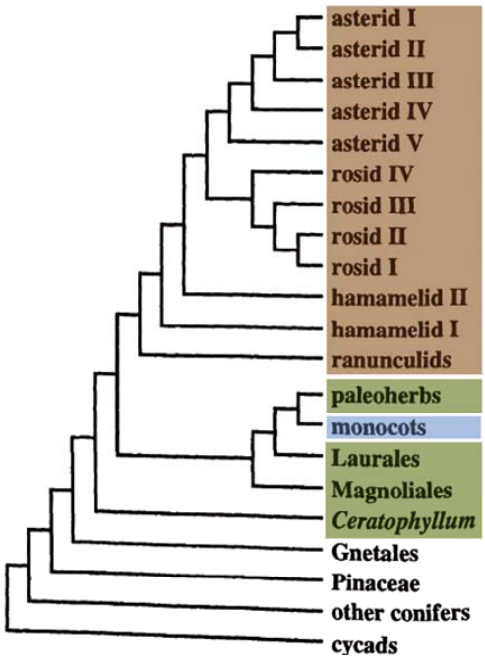


Figure 1: Phylogeny of seed plants based on *rbcL* sequence data taken from Chase & al. (1993). The three major groups of angiosperms are shaded in colour: "basal" dicotyledonous angiosperms (green), monocots (blue), eudicots (brown).

Chapter 1 deals with the placement of Sabiales and Proteales within the "basal" eudicot grade by analyzing a set of nine regions including spacers, group I and group II introns plus the coding *matK* from the large single copy region of the chloroplast genome. Up to now, five different coding regions have been used for reconstructing relationships within the early-diverging eudicots. Analysis of the plastid *rbcL* and *atpB* alone and in combination resulted in the recognition of all lineages (e.g. Chase & al., 1993; Savolainen & al., 2000a), albeit statistical support for their respective placements was not evident. However, close relations of the herbaceous Nelumbonaceae and the woody Platanaceae and Proteaceae emerged. The addition of the nuclear 18S (Hoot & al., 1999; Soltis & al., 2000) and the 26S, completing a four-gene analyses by Kim et al. (2004), resulted in improved support for most terminal clades, recovering the first-branching position of Ranunculales, while the respective placements of clades still needed to be verified. A similar hypothesis was yielded through the application of the rapidly evolving plastid *matK* gene (Hilu & al., 2003), additionally hinting on a sistergroup relationship of

Buxaceae and core eudicots. Worberg & al. (2007) combined the complete *matK* with four non-coding markers from the plastome in their analyses and were thus able to present a highly supported grade of Ranunculales, Sabiales (=Sabiaceae), Proteales (consisting of Nelumbonaceae, Platanaceae and Proteaceae), Trochodendrales (including Trochodendraceae and Tetracentraceae) and Buxales (Buxaceae plus Didymelaceae). As the only exception the position of Sabiales was only moderately supported or differed in model-based approaches, respectively. Thus the placement of Sabiales still remained to be cleared up with confidence. This difficult to resolve relationships inside the early-diverging eudicots were furthermore considered to be well adapted for testing and comparing the utility and performance of different non-coding and fast-evolving genomic partitions like spacers and introns in deep-level reconstructions.

The aim of chapter 2 was to present a thorough reconstruction of phylogenetic relationships within the first-branching clade of the eudicots with an emphasis on the evolution of growth forms inside the group. Currently, the Ranunculales consist of seven families (Ranunculaceae, Berberidaceae, Menispermaceae, Lardizabalaceae, Circaeasteraceae, Eupteleaceae, and Papaveraceae; according to APG II, 2003) comprising predominantly herbaceous groups as well as woody lineages developing trees and lianescent or shrubby forms. A surprising result that emerged due to the increased use of molecular data for systematics is the inclusion of the woody Eupteleaceae, a monogeneric family that was previously placed next to Cercidiphyllaceae (Cronquist, 1981; 1988) or Hamamelididae (Takhtajan, 1997). Although phylogenetic hypotheses agreed in the exclusion of Eupteleaceae and the predominantly herbaceous Papaveraceae s.l. from a core clade, topologies differed in postulating Eupteleaceae being the first-branching lineage (Hilu & al., 2003; Kim & al., 2004; Worberg & al., 2007), assuming a sistergroup relationship between Papaveraceae and the remainder of Ranunculales (Hoot & al., 1999; Soltis & al., 2000) or showing both families as being sister to the core clade (Qiu & al., 2005). Besides the placement of Eupteleaceae, the respective positions of Lardizabalaceae and Menispermaceae as well as of several controversial taxa such as *Glaucidium* and *Hydrastis* were under study.

Finally chapter 3 gives an overview of the phylogenetic conditions within the ranunculaceous tribe Anemoneae. Based on nuclear as well as plastid sequence data the classification system of Tamura (1995), describing the subtribes Anemoninea (including *Anemone*, *Hepatica*, *Pulsatilla* and *Knowltonia*) and Clematidinae (consisting of *Archiclematis*, *Clematis* and *Naravelia*) is tested. Furthermore the placement and

taxonomic rank of distinct lineages within the subtribe Anemoninae were examined. Several phylogenetic investigations (e.g. Hoot, 1995b) discovered two distinct clades within the subtribe, one consisting of the majority of the *Anemone*-species, *Pulsatilla* and *Knowltonia* and another, including various groups of *Anemone* and *Hepatica*. By comparing molecular rates of the distinct lineages taxonomic conclusions were drawn in the present investigation.

Material, methods & related discussion

Molecular markers

Commonly, fast-evolving and non-coding regions were used to infer relationships among species and genera, as practised in chapter 3 by using the nuclear ribosomal ITS1 & 2 and the plastid *atpB-rbcL* spacer-region for reconstructing phylogenetic relationships within a tribe of the eudicot family of Ranunculaceae. This was caused by the assumption of rapidly evolving DNA being inapplicable due to suspected high levels of homoplasy through multiple substitutions and frequent microstructural changes resulting in non-alignability. However, Borsch & al. (2003) were able to present an alignment of the plastid *trnT-F* region (including the *trnT-L* spacer, the *trnL* group I intron and the *trnL-F* spacer) for a broad-scale taxon-sampling comprising basal angiosperms as well as gymnospermous taxa. Resulting phylogentic trees were highly resolved and agreed with multi-gene and three-genome analyses by Qui & al. (1999; 2000) in topology and statistical support. Furthermore the *petD* region (*petB-D* spacer plus *petD* group II intron) was applied to phylogenetic reconstructions and its effectiveness in testing on alternative hypothesis on the "basal" nodes of the angiosperm tree was proven (Löhne & Borsch, 2005). Mutational dynamics in these spacers and introns was shown to follow complex patterns clearly related to structural constraints, such as the introns secondary structure (Quandt & al., 2004; Löhne & Borsch, 2005; Worberg & al., 2007- compare Fig.2). Thus extreme variability was always clearly confirmed to mutational hotspots (H), which could be easily excluded from analyses.

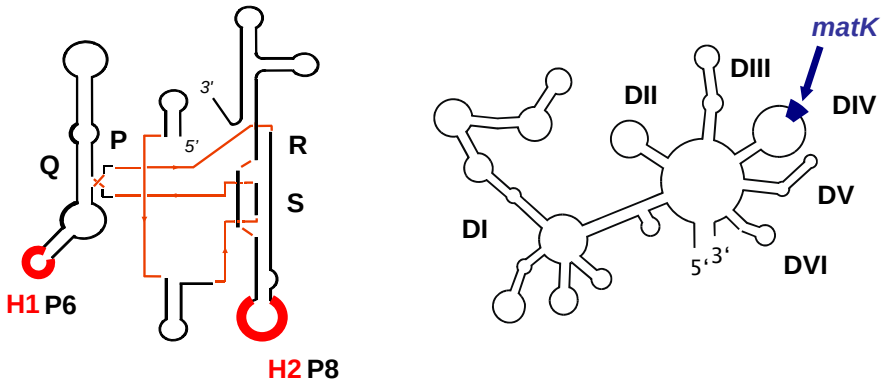


Figure 2: Schematic illustration of group I (left) and group II (right) introns secondary structure based on Cech & al. (1994) and Michel & al. (1989). P, Q, R and S represent highly conserved sequence elements of the group I intron, P6 and P8 (H1 and H2) indicate highly variable elements. DI-DVI denote the six domains of the group II intron. The position of the *matk* gene within domain IV of the *trnK* intron is indicated.

It became clear that combining these non-coding regions from the large single copy region of the chloroplast genome, *trnT-F* or *trnL-F*, respectively, and *petD*, with the fast-evolving and well performing plastid *matK* gene (e.g. Hilu & al., 2003) can lead to further resolved and statistical supported trees inside basal angiosperms as well as within early-diverging eudicots (Borsch & al., 2005; Worberg & al., 2007). Therefore this basic combination of molecular markers was chosen in chapter 2 to infer relationships on the ordinal level. Due to the amplification strategy used the whole *trnK*(*matK*)*-psbA* region, consisting of the *trnK* group II intron inclosing the *matK* open reading frame plus the *psbA* spacer (Figure 3), was included in phylogenetic analyses. Calculations resulted in a well resolved and highly supported phylogeny of Ranunculales. To further improve resolution and support of the branching order inside the early-diverging eudicots as well as to comprehensively investigate phylogenetic utility/structure and pattern of molecular evolution of rapidly evolving and non-coding genomic partitions such as spacers, group I and group II introns, the set of molecular markers used by Worberg & al. (2007) was extended by the addition of the entire *trnK* intron, the *atpB-rbcL* spacer and the *rpl16* region (consisting of the *rps3-rpl16* spacer and the *rpl16* group II intron). All three

genomic regions have already been proven to be reliable molecular markers in reconstructing phylogenetic relationships among angiosperms (e.g. *trnK/matK*: Löhne & al., 2007; Wanke & al., 2007 – *atpB-rbcL*: Renner, 1999; Schütze & al., 2003 – *rpl16*: Kelchner & Clarke, 1997; Downie & al., 2000; Löhne & al., 2007). To achieve comparability the taxon sampling of the study presented in chapter 1 is in conformity with the study of Worberg & al. (2007). Sequence statistics including length, number and quality of characters as well as coded indels was calculated and discussed for overall molecular data under study.

Figure 3: The plastid *trnK*(*matK)-psbA* region in angiosperms.

Phylogenetic reconstructions

Phylogenetic reconstructions were carried out using the programs MrBayes v3.1 (Ronquist & Huelsenbeck, 2003) and PAUP* v.4.0b10/PRAP (Müller, 2004). PRAP allows the use of the parsimony ratchet method (Nixon, 1999) by generating the respective command files which can be implemented with PAUP* v.4.0b10 (Swofford, 2002) in a quick and efficient way. Topologies gained trough Maximum parsimony mostly differed in some crucial points from trees obtained via Bayesian Inference. Therefore trees were shown separately. Since microstructural changes in fast-evolving DNA provide additional information and their utility within phylogenetic reconstructions has been proven in a number of studies (e.g. Löhne & Borsch, 2005; Löhne & al., 2007; Worberg & al., 2007), indels were coded applying the simple indel coding approach published by Simmons & Ochoterena (2000) in all analyses. The percentage of parsimony informative indel characters per data partition analyzed was calculated and discussed intensively (chapter 1 & 2). Addition of indel information rarely resulted in significant topological differences and generally increased statistical support, with the exception of a

few difficult to resolve nodes such as the first-branching position of Eupteleaceae within early-diverging eudicots (chapter 2).

Phylogenetic structure

For measuring phylogenetic structure (chapter 1) of the different partitions the method published by Müller & al. (2006) was applied, which is based on a resampling of an equivalent amount of parsimony-informative characters from the data matrices under study and using a statistic measurement for phylogenetic structure on the basis of mean support across nodes. To represent all partitions of fast-evolving DNA included into calculations four data sets were compiled (*matK*, all spacers, the sole group I intron as well as all group II introns under survey). The analyses were run using the original Perl scripts under Linux and MacOSX.

Molecular rates

Molecular dating was performed on the basis of fossils and geological data (chapter 3) taken from literature. BEAST v1.4 as published by Drummond & Rambaut (2007), applying relaxed molecular clocks within Bayesian MCMC analyses was used.

Results & discussion

Molecular evolution and phylogenetic structure of rapidly-evolving and non-coding DNA

In accordance with the results of several phylogenetic studies using rapidly-evolving and non-coding DNA (e.g. Quandt & al., 2004; Löhne & Borsch, 2005; Worberg & al., 2007) mutational dynamics within the genomic regions used is shown to follow complex patterns closely related to structural constraints. Extreme length variability in spacers and introns is clearly confined to mutational hotspots that can be linked with structural conditions. By comparing three different group II introns within eudicots (chapter 1) it became clear that these hotspots, in large parts consisting of length-variable poly-A/T stretches, are corresponding to the less constrained stem-loop elements and bulges of the introns secondary structure. Furthermore coded length mutations were mostly identified

as being simple sequence repeats, mainly ranging from five to six nucleotides in length. Strikingly, the proportion of coded indels as well as the relative amount of parsimony informative indel characters per region stayed at almost one level for all genomic regions under study within chapter 1 or chapter 2, respectively.

The comparison of the phylogenetic structure being inherent in the different non-coding regions used within phylogenetic reconstructions (spacers, the sole group I intron from the chloroplast genome and the group II introns), the fast-evolving *matK* gene and two slowly-evolving plastid genes (*atpB* and *rbcL*) resulted in the recognition of a higher average phylogenetic signal per informative site inside the non-coding data matrices than in the coding *rbcL* (see chapter 1). This finding corresponds to the results of Müller & al. (2006) who analyzed sequence data of early-diverging angiosperms. The rapidly-evolving *matK* gene was shown to be ranking among the non-coding partitions in this respect. Interestingly, spacers displayed considerable less phylogenetic structure than both, the group I intron as well as the group II introns. This contradicts the assumption of proportion and quality of phylogenetic structure being highest in spacers due to being structurally less constrained than introns.

Phylogeny of early-diverging eudicots

This thesis demonstrates the opportunities and coincidentally the limits of applying rapidly-evolving and non-coding DNA to phylogenetic reconstructions. A prominent example is the exact placement of the Sabiales within the early-diverging eudicot grade. The topology presented in chapter 1 is in mainly congruence with the hypothesis on phylogenetic relationships among early-branching eudicots published by Worberg & al. (2007), showing a grade of Ranunculales, Sabiales, Proteales, Trochodendrales and Buxales. It differs in showing Sabiales as sister to Proteales in all approaches, in contrast to a second-branching position inside early-diverging eudicots and a Bayesian tree displaying *Sabia* and *Meliosma* branching after Proteales. All three hypotheses were tested concerning their likelihood and none of them was shown as being significantly declinable. Albeit the number of characters and parsimony informative sites was doubled in comparison to the analyses carried out by Worberg & al. (2007), the exact position of the Sabiales continues to be an unanswered question.

Phylogenetic relationships and evolution of growth forms inside Ranunculales

A central goal of the studies presented in chapter 1 and 2 was to reliable resolve the branching-order within the early-diverging Ranunculales in order to gain insights into the ancestral conditions of growth forms at the base of the grade. However, relationships among the early-diverging members of the order could not be clarified with confidence. Both approaches emerged on different phylogenetic hypothesis, with the woody *Euptelea* appearing as first-branching lineage (chapter 1 – parsimony analyses, chapter 2) or showing a sistergroup relationship between the predominantly herbaceous Papaveraceae and the remaining members of the order (chapter 1 – Bayesian Inference). Furthermore statistical support as well as topology tests stayed without significance in any case. Within the core clade the branching order was resolved as Lardizabalaceae being sister to the remainder of Ranunculales, followed by Menispermaceae, Berberidaceae and Ranunculaceae, the latter sharing a sistergroup relationship. Due to maximum statistic values this phylogenetic hypothesis seems to be ensured. *Glaucidium* and *Hydrastis* are shown to be early-diverging members of the Ranunculaceae. The family was thus divided into five subfamilies complementing the classification of Ro & al. (1997) by the monogeneric subfamily Glaucidioideae.

Phylogeny and systematics of the Anemoneae

The investigation presented in chapter 3 was carried out to test phylogenetic relationships inside the ranunculaceous tribe Anemoneae. Phylogenetic analyses clearly corroborated the division of the tribe into two subtribes, Anemoninea (including *Anemone*, *Hepatica*, *Pulsatilla* and *Knowltonia*) and Clematidinae (consisting of *Archiclematis*, *Clematis* and *Naravelia*), as presented by Tamura (1995). Inside Anemoninae (= *Anemone* s.l. sensu Hoot & al., 1994) the traditional genera *Knowltonia* and *Pulsatilla* are shown to be deeply nested within the subgenus *Anemone*. In contrast *Hepatica* was revealed as being a very distinct linage within the preliminary subgenus *Anemonidium*, due to significantly differing molecular rates. Therefore the informal classification of Hoot & al. (1994) was complemented by lifting the section *Hepatica* to the subgenus level.

Conclusions

Within this thesis the high utility of fast-evolving and non-coding genomic regions for inferring relationships among early-diverging eudicots has been proven at both, deep phylogenetic levels as well as at the genus or species level, respectively. Combining the non-coding *trnL-F* and the *petD* region with the well-performing *matK* gene is again shown to result in highly resolved and supported topologies inside angiosperms (chapter 2). Microstructural changes, common to rapidly-evolving and non-coding DNA, provide useful additional information within phylogenetics. However, several difficult-to-resolve positions like the exact branching-order inside the early-diverging Ranunculales (chapter 1 & 2) or the respective positions of Sabiales and Proteales (chapter 1) could not be clarified with confidence, not even through redoubling the amount of parsimony informative sites within the comprehensive analyses carried out on early-diverging eudicots (chapter 2) in comparison to the five region investigation by Worberg & al (2007). Therefore it seems that molecular markers should not just be continuously added to analyses but could be selected carefully due to their phylogenetic structure and performance at a certain taxonomic level. Beyond, there should be a balance between high performance and an increased laboratory effort.

It is shown that molecular evolution within non-coding DNA such as spacers and introns follows certain patterns in angiosperms, as indicated by the connection of mutational hotspots to structural and functional constraints. Nevertheless, continuing work should be concentrated on further improving the understanding of mechanisms underlying molecular evolution of genomic regions, this being essential for fully utilizing the information content of non-coding DNA.

Chapter 1

Corroborating the branching order among eudicots: testing for phylogenetic signal among chloroplast introns and spacers

1.1 Abstract

The generally accepted eudicots, comprising about 75% of angiosperm species diversity, were shown to be divided into early-diverging lineages (="basal" eudicots) and a highly supported core clade, the so called core eudicots. Recent phylogenetic studies on early-diverging eudicots based on rapidly-evolving and non-coding plastid regions revealed a highly supported grade of Ranunculales, Sabiales, Proteales, Trochodendrales and Buxales. As the only exception the exact position of the Sabiales remained to be cleared up with confidence. Here we present a phylogenetic analysis based on an extended set of non-coding regions from the chloroplast genome's large single copy region, including one group I intron (*trnL*), three group II introns (*trnK* including *matK*, *petD*, *rpl16*) and four spacers (*trnL-F*, *petB-petD*, *atpB-rbcL*, *rps3-rpl16*). It was carried out to test hypothesis on phylogenetic structure among fast-evolving and non-coding regions and in comparison to coding genes as well as to further corroborate the relationships inside the early-diverging eudicots.

The combined data matrix comprised 14140 aligned sequence and additional 2955 indel characters. Mutational hotspots were shown to correspond to loops and bulges within the secondary structure of introns. Within the first-branching Ranunculales Maximum Parsimony and Bayesian Inference differed in revealing a sistergroup relationship between *Euptelea* and the remaining taxa of the order or Papaveraceae as being the first-branching lineage, respectively. Sabiales and Proteales are found to share a sistergroup relationship in all approaches with moderate to significant statistical support. However, topology tests revealed the hypothesis not being more likely than the alternatives. Analyses of the phylogenetic structure revealed a higher mean phylogenetic signal per informative site within the non-coding partitions than in the slowly-evolving coding *rbcL*, while the well performing *matK* gene is nested within the non-coding partitions. Spacers are shown to display considerably less phylogenetic structure than introns.

Non-coding and fast-evolving regions are shown to be of high utility within deep-level phylogenetics. Furthermore it is proven again that microstructural changes, frequently occurring in less constrained introns and spacers, provide useful information. However, molecular markers should be selected due to their performance. Additionally laboratory effort should be taken into consideration. Further work is needed to improve understanding of mechanisms of molecular evolution.

1.2 Introduction

Considerable progress has been made in recent years towards resolving the phylogenetic relationships among angiosperms. There is a general agreement upon the existence of a eudicot clade (e.g. Donoghue & Doyle, 1989; Chase & al., 1993; Savolainen & al., 2000a; Hilu & al., 2003; Kim & al., 2004; Worberg & al., 2007) that contains about 75% of angiosperm species diversity. Since eudicots share the appearance of tricolpate and tricolpate-dirived pollen (Donoghue & Doyle, 1989; Nandi & al., 1998; Hoot & al., 1999) they have also been called the tricolpate clade (Donoghue & Doyle, 1989). Moreover, molecular studies converged on the sister-group relationship of the Gunnerales to a clade including the remainder of Saxifragales, Vitales, rosids, Berberidopsidales, Santalales, Caryophyllales and asterids, the so called "core eudicots" (e.g. Soltis & al., 2003; Worberg & al., 2007).

The branching order among the early diverging lineages of eudicots, however, remained difficult to resolve. Early analyses of sequence data from the plastid *rbcL* and *atpB* genes (e.g. Chase & al., 1993; Savolainen & al., 2000a) alone and in combination had resulted in the recognition of lineages such as Ranunculales, Proteaceae, Sabiaceae, Trochodendraceae or a Buxaceae-Didymelaceae-clade. Close affinities of the herbaceous large flowered Nelumbonaceae to the woody Platanaceae and Proteaceae were one of the greatest surprises plant molecular phylogenetics. Nevertheless, significant statistical support for many of the respective nodes was not evident. Adding sequence data of nr18S by Hoot & al. (1999) and Soltis & al. (2000) resulted in the first-branching position of Ranunculales within eudicots. Ranunculales as well as Proteales were recovered with high or weak to moderate statistical support, respectively. Additionally, both analyses showed a clade including Buxaceae-Didymelaceae, Trochodendraceae and core eudicots which gained 87 – 88% JK support. The three groups gained maximum support and appeared either as successive sisters or in a tritomy, while the placement of Sabiales and Proteales still remained to be cleared up with confidence. Kim & al. (2004) further added nr26S data but still without much improved trees in their four-gene analysis. Application of the rapidly evolving plastid *matK* gene by Hilu & al (2003) yielded similar hypotheses than the multi-gene analyses, and 91% JK support for a sistergroup relationship of Buxaceae and the core eudicots.

By combining a data set of complete *matK* sequences with non-coding markers (the *trnL* group I intron, the *petD* group II intron, the *trnL-F* and *petB-D* spacers), and adding an

indel matrix, Worberg & al. (2007) were able to raise statistical support for a grade of Ranunculales, Sabiales, Proteales (including Nelumbonaceae and Proteaceae), Trochodendrales and Buxales. Sabiales as an exception gained only moderate Jackknife support (JK 83) as the second branch within the grade in parsimony analyses, while model based approaches depicted inconsistent positions. Bayesian inference gave a switched branching order of Sabiales and Proteales (but no support; 0.52 PP) and Maximum Likelihood lacked resolution for their respective positions.

A number of studies has shown in recent years that rapidly evolving DNA from introns and spacers of the chloroplast genome's large and small single copy region contains high levely of phylogenetic structure to resolve deep nodes in flowering plants (Borsch & al., 2003; 2005; Löhne & Borsch, 2005; Müller & al., 2006). By presenting an alignment of the *trnT-F* region (including the *trnT-L* spacer, the *trnL* group I intron and the *trnL-F* spacer) for a 42 taxon-dataset of "basal angiosperms", Borsch & al. (2003) were able to show that extreme variability is confined to certain mutational hotspots. Phylogenetic trees were well resolved and agreed with multi-gene and three-genome analyses (Qiu & al., 1999; 2000) in terms of topology and statistical support, while in comparison the amount of nucleotides utilized was less than one third. Similar observations were made by Löhne & Borsch (2005) for the group II intron in *petD*.

The detailed comparison of the rapidly evolving *trnT-F* as well as the *matK* region and the more slowly evolving *rbcL* concerning their phylogenetic structure clearly revealed higher amounts of parsimony informative characters per nucleotide sequenced in the data matrices from fast-evolving genomic regions. A resampling of identical numbers of parsimony-informative characters from the three different data partitions and evaluating different statistics of overall tree robustness and phylogenetic signal via a number of significance tests revealed a significantly higher average phylogenetic signal per informative character in the fast evolving DNA (Müller & al., 2006). Phylogenetic structure was highest in informative sites sampled from *trnT-F,* followed by *matK*, and *rbcL*. The conserved *rbcL* was distinctly less useful as a phylogenetic marker. In contrast to the less constrained *matK* gene and *trnT-F* region, displaying a wider spectrum of site rates, the *rbcL* gene showed a few highly homoplastic and rapidly evolving positions and at the same time many very conserved sites. At that time, group II introns were not yet included in comparison. However, with a mosaic like structure (Kelchner, 2002) high levels of phylogenetic structure were expected.

Based on the observations of Müller & al. (2006) made on basal angiosperms, Worberg & al. (2007) carried out an analyses on a taxon-sampling comprising members of all families of the early diverging eudicots, representatives of 19 orders of the core eudicots and members of the basal angiosperms, serving as outgroup taxa. Five fast-evolving markers from the large single-copy region of the chloroplast genome were examined: two transcribed spacers (*petB-D*; *trnL-F*), one group I intron (*trnL*), one group II intron (*petD*) and the coding *matK*. All partitions under survey provided congruent signal for hypothesis on relationships among basal eudicots. It was shown that *trnL-F* as well as *petD*, being rather small genomic regions displaying average sequence length (excluding hotspots) of 755 or 840 nucleotides, respectively, were able to resolve the majority of the eudicot topology, which compares to the markedly longer *rbcL* gene. These findings confirmed experiences made with "basal" angiosperms by Borsch & al. (2003) and Löhne & Borsch (2005).

Beyond Worberg & al. (2007) clearly corroborated the convenience of microstructural changes in fast evolving DNA providing additional information within phylogenetic reconstructions. Their utility concerning deep-level reconstructions within angiosperms has been proven in a number of studies (Löhne & Borsch, 2005; Müller & Borsch, 2005a; Löhne & al., 2007; Worberg & al., 2007). Analyzing the utility of indel information in the *matK* gene and the *trnK* group II intron, lower levels of homoplasy than in substitutions were implied (Müller & Borsch, 2005a; b). Altogether, several empirical considerations on the inclusion of coded indel characters into phylogenetic analyses clearly supported their use.

To get further insights into mutational dynamics and phylogenetic utility of chloroplast introns and spacers, additional partitions were added to the 56 angiosperm taxon set of Worberg & al. (2007). We selected the *atpB-rbcL* intergenic pacer (IGS) as well as the *rpl16* region consisting of the *rps3-rpl16* spacer and the group II intron in *rpl16*. We also sequenced the group II intron in *trnK* in addition to the CDS of the *matK* gene. These genomic regions have already been revealed to be reliable phylogenetic markers (e.g. *atpB-rbcL*: Hoot & Douglas, 1998; Renner, 1999; Schütze & al., 2003 - *rpl16*: Kelchner & Clarke, 1997; Downie & al., 2000; Zhang, 2000; Löhne & al., 2007 – *trnK*/*matK*: Müller & Borsch, 2005a; Löhne & al., 2007; Wanke & al., 2007). This collection of molecular markers was added since these regions could be expected to be well performing because on their structure and molecular evolution. This resulted in a combined data set of nine non-coding and fast-evolving plastid markers in the presented study. The

composition of 3 group II introns, the single group I intron and three spacers resulted in one of the largest data sets of non-coding and fast-evolving regions ever generated.

Hypotheses to be tested using statistical measures: (1) chloroplast introns and spacers have more phylogenetic signal than coding genes, (2) *matK* has an amount and quality of phylogenetic structure comparable to non-coding regions, and considerably more than coding regions such as *rbcL* and *atpB*; (3) amount and quality of phylogenetic structure among non-coding regions is as: IGS > group I intron > group II intron.

This study furthermore aimed at corroborating the relationships among the early-diverging eudicots by extending the set of non-coding and fast-evolving genomic regions. An important question was whether it would be possible to infer the exact placement of the Sabiales.

1.3 Material and methods

Taxon sampling, plant material and molecular markers

The taxon sampling is in conformity with the study of Worberg & al. (2007), which comprised 56 angiosperm species, representing 47 families from 31 orders that were recognized by APG II (2003). In total, 14 outgroup taxa were included into analyses which represent the first branching angiosperms, the magnoliids, Chloranthaceae, *Ceratophyllum*, and *Acorus* (monocots). Among basal eudicots 23 species belonging to 14 families were examined. Within Sabiales a second species of *Sabia* (*Sabia swinhoei*) was additionally chosen to complement *Sabia japonica*. The core eudicots are represented by several families from seven orders: Gunnerales (3 species), Saxifragales (2), Vitales (2), Caryophyllales (2), Dilleniales (1), Santalales (1), and Berberidopsidales (1). Six species belonging to the rosids and five members of the asterids were included.

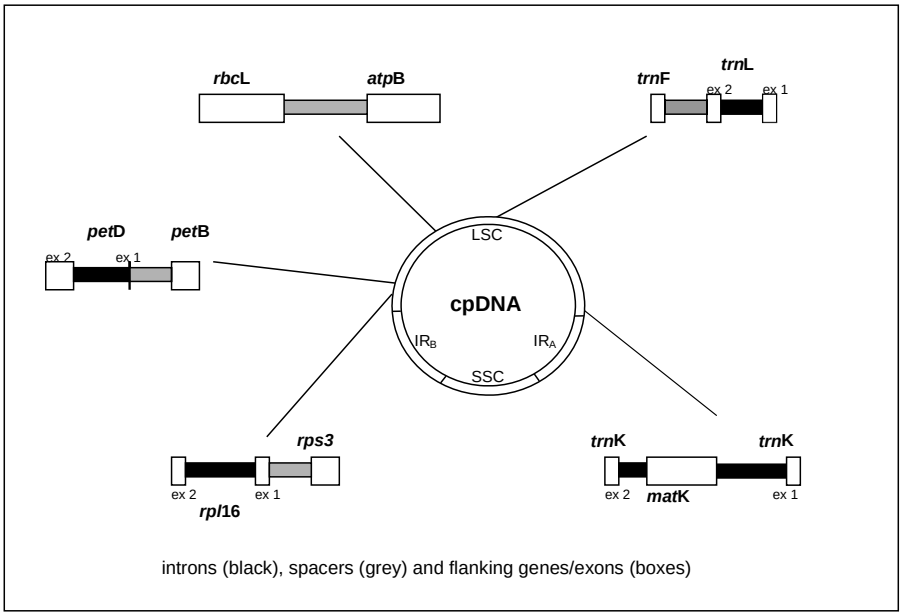


Figure 1: Overview of the cpDNA indicating the positions and organization of the regions studied. Introns (black), spacers (grey) and flanking genes or exons (boxes), respectively.

Molecular data for nine plastid regions was analysed (see Figure 1): the *trnK* group II intron (including *matK*), the *trnL* group I intron, the *trnL-F* IGS, the *petB-petD* IGS, the *petD* group II intron, the *atpB-rbcL* IGS, the *rps3-rpl16* IGS and the *rpl16* group II intron. For amplification and sequencing they were treated as five partitions ("*trnK/matK*" = *trnK* group II intron, including the *matK* gene; "*trnL-F*" = *trnL* group I intron plus the *trnL-F* IGS; "*petD*" = *petB-petD* IGS and the *petD* group II intron; "*atpB-rbcL*" = the *atpB-rbcL* IGS; "*rpl16*" = containing the *rps3-rpl16* IGS and the *rpl16* group II intron). All molecular data for *Sabia swinhoei* was newly generated. For *trnK/matK*, *trnL-F* and *petD* most sequences were taken from Worberg & al. (2007). The major part (45 sequences) of the *trnK/matK* data was updated in this study by completing the upstream and downstream halves of the *trnK* intron. Therefore already existing PCR products were sequenced with additional primers or the missing parts of the *trnK* intron were amplified from the same DNA stock. Molecular data for the *trnK/matK* region of *Aristolochia pistolochia* was obtained from the study by Wanke & al. (2007), while two sequences of basal angiosperms were provided by Löhne & al. (2007). For *trnL-F*, 34 sequences were published by Worberg & al. (2007), whereas 15 were originally generated for the study of Borsch & al. (2003). Sequence data on *petD* was obtained from Worberg & al. (2007, 35 sequences) and Löhne & Borsch (2005; 15 sequences). Most of the molecular data for

atpB-rbcL and *rpl16* was produced in this study. For *atpB-rbcL* 48 sequences were newly generated, while 47 completely new sequences were produced for the *rpl16* partition. Two partial sequences were taken from Löhne & al. (2007) and completed by adding missing parts of the *rps3-rpl16* spacer through amplification and sequencing of the same DNA stock with additional primers. Altogether two single sequences of *trnK/matK*, one of *trnL-F* and one for *atpB-rbcL* as well as complete plastome sequences for *Acorus calamus*, *Amborella trichopoda, Arabidopsis thaliana*, *Atropa belladonna*, *Ceratophyllum demersum*, *Nicotiana tabacum*, *Oenothera elata*, *Panax ginseng*, and *Spinacia oleracea* were obtained from GenBank. The *trnL-F* sequence of *Arabidopsis thaliana* was replaced by *Brassica nigra* since the whole-genome sequence contained obvious sequencing errors. Detailed information on all taxa included in this survey, the respective vouchers and GenBank accession numbers are given in Table 1.

Table 1: Taxa analysed, voucher datails, GenBank accession numbers and references; family assignment according to APG II (2003).

Taxon	Family	Voucher / Herbarium	Garden / Field origin	GenBank Accession Numbers *trnK/matK*	*trnL-F*	*petD*	*atpB-rbcL*	*rpl16*
OUTGROUP								
Chimonanthus praecox (L.) Link	Calycanthaceae	T. Borsch 3396 (BONN)	BG Bonn	(AF542569) This study update	AM397150 Worberg & al. (2007)	AM396524 Worberg & al. (2007)	This study	This study
Hedycarya arborea Forst.	Monimiaceae	A. Worberg 014 (BONN)	BG Bonn	(AM396509) This study update	AM397149 Worberg & al. (2007)	AM396523 Worberg & al. (2007)	This study	This study
Umbellularia californica (Hooker & Arn.)Nutt.	Lauraceae	T. Borsch 3471 (BONN)	BG Bonn	(AF543752) This study update	AY145350 Borsch & al. (2003)	AY590850 Löhne & Borsch (2005)	This study	This study
Magnolia virginiana L.	Magnoliaceae	T. Borsch & C. Neinhuis 3280 (VPI, FR)	USA, Maryland	AB020988 Azuma & al. (1999)	AY145354 Borsch & al. (2003)	-	-	-
Magnolia officinalis Rehder & Wilson	Magnoliaceae	C. Löhne 53 (BONN)	BG Bonn	This study	-	AY590846 Löhne & Borsch (2005)	-	This study
Magnolia officinalis Rehder & Wilson	Magnoliaceae	GenBank	-	-	-	-	AY008970 Kim & al (2000)	-
Chloranthus brachystachys Blume	Chloranthaceae	T. Borsch 3467 (BONN)	BG Bonn	(AF543733) This study update	AY145334 Borsch & al. (2003)	AY590864 Löhne & Borsch (2005)	This study	This study
Acorus gramineus L.	Acoraceae	T. Borsch 3458 (BONN)	BG Bonn	-	AY145336 Borsch & al. (2003)	-	-	-
Acorus calamus L.	Acoraceae	C. Löhne 51 (BONN)	BG Bonn	This study	-	AY590840 Löhne & Borsch (2005)	-	-
Acorus calamus L.	Acoraceae	GenBank	-	-	-	-	NC_007407 Goremykin & al. (2005)	NC_007407 Goremykin & al. (2005)
Ceratophyllum demersum L.	Ceratophyllaceae	T. Wieboldt 16073 (VPI)	USA, Virginia	-	AY145335 Borsch & al.	AY590841 Löhne &	This study	This study

					(2003)	Borsch (2005)		
Ceratophyllum demersum L.	Ceratophyllaceae	GenBank	-	EF614270 Moore & al. (2007)	-	-	-	-
Aristolochia pistolochia L.	Aristolochiaceae	T. Borsch 3257 (FR)	France, Herault	-	AY145341 Borsch & al. (2003)	AY590862 Löhne & Borsch (2005)	-	-
Aristolochia pistolochia L.	Aristolochiaceae	-	-	DQ532062 Wanke & al. (2007)	-	-	This study	This study
Austrobaileya scandens C. White	Austrobaileyaceae	T. Borsch 3464 (BONN)	BG Bonn	DQ185523 Löhne & al. (2007)	AY145326 Borsch & al. (2003)	AY590867 Löhne & Borsch (2005)	This study	AM421606 Löhne & al. (2007)
Nymphaea odorata Aiton ssp. *odorata*	Nymphaeaceae	T. Borsch & V. Wilde 3132 (VPI, BONN)	USA, Georgia	-	AY145333 Borsch & al. (2003)	-	-	-
Nymphaea odorata Aiton ssp. *tuberosa* (Paine)Wiersema & Hellq.	Nymphaeaceae	T. Borsch, B. Hellquist, J. Wiersema 3389 (BONN)	Canada, Manitoba	DQ185549 Löhne & al. (2007)	-	AY590873 Löhne & Borsch (2005)	This study	AM421605 Löhne & al. (2007) – this study update
Amborella trichopoda Baill.	Amborellaceae	T. Borsch 3480 (VPI)	UCLA, Sta. Catarina BG	-	AY145324 Borsch & al. (2003)	AY590876 Löhne & Borsch (2005)	-	-
Amborella trichopoda Baill.	Amborellaceae	GenBank	-	NC_005086 Goremykin & al. (2003)	-	-	N_C005086 Goremykin & al. (2003)	NC_005086 Goremykin & al. (2003)
BASAL EUDICOTS								
Euptelea pleiosperma Siebold & Zucc.	Eupteleaceae	A. Worberg 003 (BONN)	BG Bonn	(AM396510) This study update	AM397151 Worberg & al. (2007)	AM396525 Worberg & sl. (2007)	This study	This study
Akebia quinata Decne.	Lardizabalaceae	T. Borsch 3412 (BONN)	BG Bonn	(AF542587) This study update	AM397152 Worberg & al. (2007)	AM396526 Worberg & al. (2007)	This study	This study
Dicentra eximia (Ker Gawl.)Torr.	Papaveraceae	T. Borsch 3468 (BONN)	BG Bonn	(DQ182345) This study update	AY145361 Borsch & al. (2003)	AY590835 Löhne & Borsch (2005)	This study	This study
Papaver triniaefolium Boiss.	Papaveraceae	A. Worberg 018	BG Bonn	(AM396511)	AM397153	AM396527	This study	This study

		(BONN)		This study update	Worberg & al. (2007)	Worberg & al. (2007)		
				(AF542588)	AM397159	AM396528	This study	This study
Cocculus laurifolius DC.	Menispermaceae	T. Borsch 3406 (BONN)	BG Bonn	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
				(AF542589)	AM397154	AM396529	This study	This study
Stephania delavayi Diels.	Menispermaceae	T. Borsch 3550 (BONN)	BG Bonn	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
				-	AM397155	AM396530	This study	This study
Xanthorhiza simplicissima Woodhouse	Ranunculaceae	T. Borsch 3394 (BONN)	BG Bonn		Worberg & al. (2007)	Worberg & al. (2007)		
Xanthorhiza simplicissima Woodhouse	Ranunculaceae	A-M. Barniske 061 (DR)	BG Dresden	This study	-	-	-	-
				(AF542585)	AM397156	AM396531	This study	This study
Mahonia japonica DC.	Berberidaceae	T. Borsch 3405 (BONN)	BG Bonn	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
				(AF542586)	AM397157	AM396532	This study	This study
Podophyllum peltatum L.	Berberidaceae	T. Borsch 3393 (BONN)	BG Bonn	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
				(AM396512)	AM397158	AM396533	This study	This study
Sabia japonica Maxim.	Sabiaceae	Y-L. Qiu 91025 NCU	NCU	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
Sabia swinhoei Hemsl.	Sabiaceae	Y-L. Qiu 99003 NCU	NCU	This study	This study	This study	This study	This study
				(AM396513)	AM397160	AM396534	This study	This study
Meliosma cuneifolia Franch.	Sabiaceae	A. Worberg 001 (BONN)	BG Bochum	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
				(AM396514)	AM397161	AM396535	This study	This study
Nelumbo nucifera Gaertn. ssp *nucifera var. alba* (Willd.) Borsch & Barthlott	Nelumbonaceae	A.Worberg s.n. (BONN)	BG Bonn	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
Nelumbo nucifera Gaertn. ssp *lutea* (Willd.) Borsch & Barthlott	Nelumbonaceae	T. Borsch & Summers 3220 (FR)	USA, Missouri	(AF543740) This study update	AY145359 Borsch & al. (2003)	AY590836 Löhne & Borsch (2005)	This study	This study
				(AM396515)	AM397162	AM396536	This study	This study
Embothrium coccineum Forst.	Proteaceae	A. Worberg 004 (BONN)	BG Bonn	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
Grevillea banksii R.Br.	Proteaceae	T. Borsch 3413	BG Bonn	(AF542583)	AM397163	AM396537	This study	This study

		(BONN)		This study update (AM396503)	Worberg & al. (2007) AM397164	Worberg & al. (2007) AM396538	This study	This study
Platanus orientalis L.	Platanaceae	A. Worberg 005 (BONN)	BG Bonn	This study update (AF543747)	Worberg & al. (2007) AY145358	Worberg & al. (2007) AY590834	This study	This study
Platanus occidentalis L.	Platanaceae	Slotta s.n. (VPI)	USA, Virginia	This study update (AM396504)	Borsch & al. (2003) AM397165	Löhne & Borsch (2005) AM396539	This study	This study
Tetracentron sinense Oliver	Trochodendraceae	T. Borsch 3494 (BONN)	BG Freiburg	This study update (AF543751)	Worberg & al. (2007) AY145360	Worberg & al. (2007) AY590833	This study	This study
Trochodendron aralioides Siebold & Zucc.	Trochodendraceae	T. Borsch 3478 (BONN)	BG Bonn	This study update (AM396505)	Borsch & al. (2003) AM397166	Löhne & Borsch (2005) AM396540	This study	This study
Didymeles integrifolia J.St.-Hil.	Didymelaceae	J. Rabenantoandro et al. 916 (MO)	Madagascar	This study update (AF543728)	Worberg & al. (2007) AY145357	Worberg & al. (2007) AY590832	This study	This study
Buxus sempervirens L.	Buxaceae	T. Borsch 3465 (BONN)	BG Bonn	This study update (AF542581)	Borsch & al. (2003) AM397167	Löhne & Borsch (2005) AM396541	This study	This study
Pachysandra terminalis Siebold & Zucc.	Buxaceae	T. Borsch 3407 (BONN)	BG Bonn	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
CORE EUDICOTS								
				(AM396506)	AM397168	AM396542	This study	This study
Gunnera tinctoria (Molina) Mirb.	Gunneraceae	N. Korotkov 50 (BONN)	BG Bonn	This study update (AM396507)	Worberg & al. (2007) AM397169	Worberg & al. (2007) AM396543	This study	This study
Myrothamnus flabellifolia Welw.	Myrothamnaceae	A. Worberg 011 (BONN)	BG Bonn	This study update (AF542591)	Worberg & al. (2007) AM397170	Worberg & al. (2007) AM396544	This study	This study
Myrothamnus moschata Baill.	Myrothamnaceae	E. Fischer & W. Höller (BONN)	BG Bonn	This study update (AM396508)	Worberg & al. (2007) AM397171	Worberg & al. (2007) AM396545	This study	This study
Cercidiphyllum japonicum Siebold & Zucc.	Cercidiphyllaceae	T. Borsch s.n. (BONN)	BG Bonn	This study update	Worberg & al. (2007)	Worberg & al. (2007)		
Chrysosplenium alternifolium	Saxifragaceae	T. Borsch s.n.	Germany	(AM396496)	AM397172	AM396546	This study	This study

L.		(BONN)		This study update (AF542593)	Worberg & al. (2007) AM397173	Worberg & al. (2007) AM396547	This study	This study
Vitis riparia A. Gray	Vitaceae	T. Borsch 3458 (BONN)	BG Bonn	This study update (AM396497)	Worberg & al. (2007) AM397174	Worberg & al. (2007) AM396548	This study	This study
Leea coccinea Planch.	Leeaceae	T. Borsch 3418 (BONN)	BG Bonn	This study update (AM396498)	Worberg & al. (2007) AM397175	Worberg & al. (2007) AM396549	This study	This study
Dillenia philippinensis Rolfe	Dilleniaceae	A. Worberg 010 (BONN)	BG Bonn	This study update (DQ182342)	Worberg & al. (2007) AY145362	Worberg & al. (2007) AY590831	This study	This study
Aextoxicon punctatum Ruiz & Pav.	Aextoxicaceae	T. Borsch 3459 (BONN)	BG Bonn	This study update (AM396499)	Borsch & al. (2003) AM397176	Löhne & Borsch (2005) AM396550	This study	This study
Osyris alba L.	Santalaceae	A. Worberg 015 (BONN)	BG Bonn	This study update	Worgerg & al. (2007)	Worberg & al. (2007)		
CARYOPHYLLIDS								
Rhipsalis paradoxa Salm-Dyck.	Cactaceae	A. Worberg s.n. (BONN)	BG Bonn	-	AM397177 Worberg & al. (2007)	AM396551 Worberg & al. (2007)	This study	This study
Rhipsalis floccosa Salm-Dyck.	Cactaceae	GenBank	-	AY015342 Nyffeler (2002)	-	-	-	-
Spinacia oleracea L.	Chenopodiaceae	Genbank	-	NC_002202 Schmitz-Linneweber & al. (2001)	NC_002202 Schmitz-Linneweber & al. (2001)	NC_002202 Schmitz-Linneweber & al. (2001)	NC_002202 Schmitz-Linneweber & al. (2001)	NC_002202 Schmitz-Linneweber & al. (2001)
ROSIDS								
Erodium cicutarium (L.) L'Hér	Geraniaceae	T. Borsch 3483 (BONN)	Germany, Eifel	(AM396500) This study update	AM397178 Worberg & al. (2007)	AM396552 Worberg & al. (2007)	This study	This study
Brassica nigra (L.) W.D.J.Koch	Brassicaceae	Genbank	-	-	AF451579 Yang & al (2002)	-	-	-
Arabidopsis thaliana (L.) Heynh.	Brassicaceae	Genbank	-	NC_000932 Sato & al.	-	NC_000932 Sato & al.	NC_000932 Sato & al.	NC_000932 Sato & al.

				(1999)		(1999)	(1999)	(1999)
				(AM396501)	-	AM396555	This study	This study
Stachyurus chinensis Franch.	Stachyuraceae	A. Worberg s.n. (BONN)	BG Bonn	This study update		Worberg & al. (2007)		
				-	AB066335 Ohi & al. (2003)	-	-	-
Stachyurus chinensis Franch.	Stachyuraceae	GenBank						
				(AF542600)	AM397179	AM396553	This study	This study
		T. Borsch 3415		This study	Worberg &	Worberg & al.		
Coriaria myrtifolia L.	Coriariaceae	(BONN)	BG Bonn	update	al. (2007)	(2007)		
				(AM396502)	AM397180	AM396554	This study	This study
		A. Worberg 012		This study	Worberg &	Worberg & al.		
Larrea tridentata Coult.	Zygophyllaceae	(BONN)	BG Bonn	update	al. (2007)	(2007)		
				NC_002693	NC_002693	NC_002693	NC_002693	NC_002693
Oenothera elata Kunth	Onagraceae	Genbank	-	Hupfer & al. (2000)	Hupfer & al. (2000)	Hupfer & al. (2000)	Hupfer & al. (2000)	Hupfer & al. (2000)
ASTERIDS								
				(AF542608)	AM397181	AM396556	This study	This study
		T. Borsch 3485		This study	Worberg &	Worberg & al.		
Impatiens noli-tangere L.	Balsaminaceae	(BONN)	BG Bonn	update	al. (2007)	(2007)		
				(AF542607)	AM397182	AM396557	This study	This study
		T. Borsch 3419		This study	Worberg &	Worberg & al.		
Ilex aquifolium L.	Aquifoliaceae	(BONN)	BG Bonn	update	al. (2007)	(2007)		
				NC_006290	NC_006290	NC_006290	NC_006290	NC_006290
Panax ginseng C.A. Mey.	Araliaceae	Genbank	-	Kim & Lee (2004)	Kim & Lee (2004)	Kim & Lee (2004)	Kim & Lee (2004)	Kim & Lee (2004)
				NC_004561	NC_004561	NC_004561	NC_004561	NC_004561
				Schmitz-	Schmitz-	Schmitz-	Schmitz-	Schmitz-
Atropa belladonna L.	Solanaceae	Genbank	-	Linneweber & al. (2002)	Linneweber & al. (2002)	Linneweber & al. (2002)	Linneweber & al. (2002)	Linneweber & al. (2002)
				NC_001879	NC_001879	NC_001879	NC_001879	NC_001879
				Shinozaki &	Shinozaki &	Shinozaki &	Shinozaki &	Shinozaki &
Nicotiana tabacum L.	Solanaceae	Genbank	-	al. (1986)	al. (1986)	al. (1986)	al. (1986	al. (1986)

DNA isolation, amplification, and sequencing

DNA was isolated from fresh or silica gel-dried plant material by using the CTAB-method described in Doyle & Doyle (1990). Three extractions were carried out to yield high amounts of genomic DNA (compare Borsch & al., 2003). In cases of suboptimal DNA quality extractions were cleaned using commercially available spin columns (Macherey-Nagel; Düren, Germany). To gain complete sequences of spacers and introns that are necessary to analyze molecular evolution, amplification was carried out using primers that were located sufficiently far away from the region under study. Sequencing was performed using the universal PCR primers and specially designed internal primers in cases of long amplicons or problematic reads due to microsatellite areas. Amplification of *trnK/matK* was done with trnK-Fbryo (forward, Wicke & Quandt, in press) and psbA-R (reverse, Steele & Vilgalys, 1994). Thus it was possible to obtain sequence data of the entire *trnK* as well as of the adjacent *psbA* spacer. The latter was not alignable across angiosperms and will therefore be considered elsewhere. For sequencing the whole fragment several additional primers were designed using SeqState v1.2 (Müller, 2005; see Appendix A). The *trnL-F* partition was amplified and sequenced by using primers trnL-C and trnL-F (Taberlet & al., 1991). For *petD* the existing set of universal primers from Löhne & Borsch (2005) was used. Two universal primers were newly designed for amplifying and sequencing the *atpB-rbcL* region, based on the completely sequenced chloroplast genomes available at GenBank of *Arabidopsis thaliana* (NC_000932), *Nicotiana tabacum* (NC_001879), and *Zea mays* (NC_001666). The forward primer (atpB-rbcLF1) is located about 1240 bp downstream in the *atpB* gene, whereas the reverse primer (atpB-rbcLR) was placed 28 bp downstream the *rbcL* gene. Due to deviating sequences and/or microsatellites several lineage-specific internal primers were designed, such as atpB-rbcL379F (Austrobaileyaceae), GREVatpB-rbcL1700F (Proteaceae) and CA05ar1696F (Cactaceae). For the amplification of the *rpl16* region the newly designed primer rps3Fa (forward) as well as the L16exon2 (reverse) published by Downie & al. (2000) was used. The universal forward primer was designed using complete plastome sequences from GenBank (*Arabidopsis thaliana*, NC_000932; *Nicotiana tabacum*, NC_001879; *Spinacea oleracea*, NC_002202; *Zea mays*, NC_001666). Since the 5'exon of the *rpl16* gene comprises only nine nucleotides it was placed about 487 bp downstream the *rps3* gene to produce complete sequences of the *rpl16* group II intron. As a result the additional inclusion of the *rps3-rpl16* spacer into phylogenetic analyses was possible. Resulting from extensive poly A/T stretches several internal sequencing primers had to be

designed. Two of them, rpl16_690F and rpl16_1900R, are universal for all angiosperm lineages represented in this study. In addition a primer partly annealing to the *rpl16*-5'exon was developed (rpl16_510F, forward), halfway spanning a poly A/T stretch at the beginning of the *rpl16* intron. All primers used in this survey are listed in Appendix A.

Amplification and sequencing reactions were performed in a T3 Thermocycler or Gradient Thermocycler (Biometra; Göttingen, Germany). PCR protocols and reaction conditions followed Löhne & Borsch (2005) for *petD*, Borsch & al. (2003) for *trnL-F*, Wicke & Quandt (in press) for *trnK/matK*. For *atpB-rbcL* as well as for *rpl16* PCR amplifications were performed in 50µl-reactions containing 1U Taq DNA polymerase (SAWADY-Taq-DNA-Polymerase, Peqlab; Erlangen, Germany), 1mM dNTP mix of each 0.25 mM, 1x taq buffer (Peqlab), 1.25-2.5 mM MgCL2 (Peqlab) and 20 pmol of each amplification primer. The following thermal cycling program was used for *atpB-rbcL*: 2 min 94°C, 10 cycles (1 min 94°C, 1 min 55°C dT= -0.50 °C, 3 min 68°C), 20 cycles (1 min 94°C, 1 min 50 °C, 3 min 68°C), 10 min 68°C. The *rpl16* region was amplified applying the PCR protocol outlined in Simões & al. (2004). Amplicons were purified using the NucleoSpin Extract II kit for cleanup of gel extraction (Macherey-Nagel; Düren, Germany) after running them out on a 1.2% agarose gel for 2.5 h at 80 V. Direct sequencing was performed using the DTCS QuickStart Reaction Kit by BeckmannCoulter. Extension products were either run on a BeckmannCoulter CEQ 8000 sequencer, or sequenced by Macrogen Inc., South Korea (www.macrogen.com). Sequences were edited manually with PhyDE v0.995 (Müller & al., 2005).

Alignment, indel coding, and phylogenetic analysis

In addition to substitution events, noncoding regions are characterized by the presence of small structural changes, such as deletions, single sequence repeats, other insertions, and inversions. Recent studies have reconstructed the history of microstructural changes within rapidly evolving spacers and introns located in the single-copy regions of the chloroplast genome (Löhne & Borsch, 2005; Stech & Quandt, 2006), pointing to mutational patterns common to the plastome. Nevertheless, currently available alignment programs (e.g., CLUSTAL X [Thomson & al., 1997], POY [Wheeler & al., 1996-2003], Dialign2 [Morgenstern, 1999]) still fail to recognize these patterns and align sequences comprising insertions and inversions correctly. Instead of using these alignment algorithms and software applications in this study, nucleotide sequences were aligned "by

eye" using PhyDE v0.995. Alignment was carried out by means of the rules pointed out in Kelchner (2000), Borsch & al. (2003) and Olsson & al. (2009). The applied alignment rules are based on motif recognition, taking known mechanisms leading to microstructural changes as well as other similarity-based criteria for homology assessment into consideration (Golenberg & al., 1993; Kelchner & Clark, 1997; Hoot & Douglas, 1998; Graham & al., 2000; Kelchner, 2000; Quandt & al., 2003). Sequence stretches with unclear primary homology were marked as "hotspots" referring to the rules outlined in Olsson & al. (2009.) and excluded from the phylogenetic analyses. For incorporating indel characters into analyses, the simple-indel coding method by Simmons & Ochoterena (2000) was applied via SeqState v1.2. Afterwards the resulting indel matrix was combined with the nucleotide-sequence matrix and used for parsimony analyses and Bayesian Inference (BI).

Molecular data of the five regions was analyzed in different combinations for phylogenetic reconstruction. The following data partitions were surveyed: group I intron sequences, group II intron sequences, group I and II intron sequences, spacer sequences. The dataset analyzed by Worberg & al. (2007) was combined with the *trnK*, *atpB-rbcL* and/or *rpl16* partition. All five regions were studied alone as well.

Calculation of most parsimonious trees (MPTs) was done by using the parsimony ratchet (Nixon, 1999) as implemented in PRAP (Müller, 2004). Ratchet settings were 20 random-addition cycles of 200 ratchet replicates, and upweighting 25% of the characters. In cases with multiple MPTs a strict consensus trees was calculated. Nodes were evaluated by bootstrapping in PAUP* version 4.0b10 for Windows (Swofford, 2002) using 1000 replicates.

BI was done using MrBayes v3.1 published by Ronquist & Huelsenbeck (2003). The GTR + Γ + I model was applied for sequence data, and the restriction site model ("F81") for the indel matrix. Four runs (1,000,000 generations each) with 4 chains each were run simultaneously. Chains were sampled every 10th generation. The consensus tree and the posterior probability (PP) of clades were calculated based upon the trees sampled after the burn-in set at 250,000 generations. TreeGraph (Müller & Müller, 2004) was used for tree drawing. Sequence statistics were calculated using SeqState v1.2 (Müller, 2005). Datasets are deposited on the appended CD.

Evaluation of alternative topologies

For testing the likelihood of the topologies inferred from the dataset used in this study in relation to alternative phylogenetic hypothesis the approximately unbiased test (AU test) as described by Shimodaira (2002) was performed. Log-likelihoods for the trees under survey were calculated using PAUP* version 4.0b10 for Windows (Swofford, 2002), while p-values were generated in CONSEL (Shimodaira & Hasegawa, 2001) using the multi-scale bootstrap technique.

The evaluated alternative hypotheses are illustrated in Figure 3. They refer to the position of *Euptelea* as either being sister to the remaining Ranunculales (4, MP analysis, this study) or second-branching after Papaveraceae (5, BI, this study) as well as to the placement of Sabiales. In the latter case three different scenarios were compared concerning their likelihood: a Sabiales/Proteales sistergroup relationship (1 this study), Sabiales branching-off after Ranunculales (2, MP analysis, Worberg & al., 2007) and Proteales branching-off after Ranunculales (3, BI, Worberg & al., 2007).

Phylogenetic structure

The phylogenetic structure of the different genomic regions used in tree reconstruction was evaluated applying the method developed by Müller & al. (2006), using resampling of an equivalent amount of parsimony-informative characters from four different data matrices (*matK*, all spacers, the sole group I intron as well as all group II introns under survey). For comparing phylogenetic structure of rapidly-evolving DNA and slowly-evolving protein-coding genes additionally molecular data of two plastid genes (*atpB* and *rbcL*) was included into the analyses. Sequences were downloaded from GenBank for an adequate taxon-sampling. Due to missing data several taxa were compensated at the genus-level. Hence, *Umbellularia californica* was replaced by *Laurus nobilis*, and *Rhipsalis paradoxa*/*floccosa* was exchanged by *Pereskia aculeata*. *Myrothamnus moschata* was excluded from analyses and the sequences *Sabia swinhoei* and *Platanus occidentalis* were doubled due to missing data. Taxa enclosed into analyses, family assignment as well as GenBank accession numbers are listed in Table 2. One statistic measurement was used (called *R* in the following) for phylogenetic structure on the basis of mean support across nodes. *R* equals 1 in the case that all branches of a phylogenetic tree received maximal statistical support, whereas it is reaching 0 in a completely unresolved 50%-majority-rule consensus tree. For testing for differences in phylogenetic

structure between the various data partitions a simple significance test was used. In the result a confidence interval was constructed on the basis of the standard error SE. The analyses were run using the original Perl scripts (compare Müller & al., 2006) under Linux and MacOSX.

Table 2: Additional data of two coding plastid regions (*atpB*, *rbcL*) used within measures of phylogenetic structure. Taxa analysed and GenBank accession numbers; family assignment according to APG II (2003). Taxa are listed in alphabetical order.

Taxon	Family	GenBank Accession Numbers		Taxon	Family	GenBank Accession Numbers	
		atpB	*rbcL*			*atpB*	*rbcL*
OUTGROUP				**BASAL EUDICOTS**			
Chimonanthus praecox (L.) Link	Calycanthaceae	AF197605	L12639	*Tetracentron sinense* Oliver	Trochodendraceae	AF093422	L12668
Hedycarya arborea Forst.	Monimiaceae	AJ235490	L12648	*Trochodendron aralioides* Siebold & Zucc.	Trochodendraceae	EU002169	L01958
Laurus nobilis L.	Lauraceae	AJ235518	-	*Didymeles perrieri* Leandri	Didymelaceae	AF092119	AF061994
Umbellularia californica (Hooker & Arn.)Nutt.	Lauraceae	-	DQ182335	*Buxus sempervirens* L.	Buxaceae	AF092110	DQ182333
Magnolia officinalis Rehder & Wilson	Magnoliaceae	-	AY008933	*Pachysandra procumbens* Michx.	Buxaceae	-	AF061993
Magnolia tripetala (L.) L.	Magnoliaceae	AJ235526	-	*Pachysandra terminalis* Siebold & Zucc.	Buxaceae	AF528854	-
Chloranthus japonicus Siebold	Chloranthaceae	AJ235431	L12640	**CORE EUDICOTS**			
Acorus calamus L.	Acoraceae	NC007407	NC007407	*Gunnera manicata* Linden ex Delchev.	Gunneraceae	EU002162	EU002279
Ceratophyllum demersum L.	Ceratophyllaceae	AJ235430	M77030	*Myrothamnus flabellifolia* Welw.	Myrothamnaceae	AF093386	AF060707
Aristolochia macrophylla Lam.	Aristolochiaceae	AJ235399	-	*Cercidiphyllum japonicum* Siebold & Zucc.	Cercidiphyllaceae	AF092112	L11673
Aristolochia pistolochia L.	Aristolochiaceae	-	AF543711	*Chrysosplenium iowense* Rydb.	Saxifragaceae	AJ235432	-
Austrobaileya scandens C. White	Austrobaileyaceae	AJ235403	L12632	*Chrysosplenium japonicum* Siebold & Zucc.	Saxifragaceae	-	AB003269
Nymphaea odorata Aiton	Nymphaeaceae	AJ235544	M77034	*Vitis vinifera* L.	Vitaceae	AM083947	NC007957
Amborella trichopoda Baill.	Amborellaceae	AJ235389	L12628	*Leea guineensis* G. Don	Leeaceae	AJ235520	AJ235783
INGROUP				*Dillenia indica* L.	Dilleniaceae	-	L01903
BASAL EUDICOTS				*Dillenia philippinensis* Rolfe	Dilleniaceae	AY788268	-
Euptelea pleiosperma Siebold & Zucc.	Eupteleaceae	-	AY048174	*Aextoxicon punctatum* Ruiz & Pav.	Aextoxicaceae	AJ235384	X83986
Euptelea polyandra Siebold & Zucc.	Eupteleaceae	AF528850	-	*Osyris lanceolata* Hochst. & Steud.	Santalaceae	AF209641	EF464525
Akebia quinata Decne.	Lardizabalaceae	L37924	L12627	**CARYOPHYLLIDS**			
Dicentra eximia (Ker Gawl.)Torr.	Papaveraceae	L37927	L37917	*Pereskia aculeata* Mill.	Cactaceae	AF209648	AF206805
Papaver orientale L.	Papaveraceae	U86394	L08764	*Spinacia oleracea* L.	Chenopodiaceae	AF528861	NC002202
Cocculus pendulus (J.B. Först. & G. Forst.) Diels	Menispermaceae	FJ026418	FJ026478	**ROSIDS**			
Stephania rotunda Lour.	Menispermaceae	FJ026449	FJ026509	*Erodium chrysanthum* L'Hér. ex DC.	Geraniaceae	EU922030	-
Xanthorhiza simplicissima Woodhouse	Ranunculaceae	AF093394	L12669	*Erodium cicutarium* (L.) L'Hér	Geraniaceae	-	DQ452882
Mahonia aquifolium (Pursh) Nutt.	Berberidaceae	AF528846	-	*Arabidopsis thaliana* (L.) Heynh.	Brassicaceae	NC000932	NC000932
Mahonia bealei (Fortune) Carrière	Berberidaceae	-	L75871	*Stachyurus praecox* Sieb. & Zucc.	Stachyuraceae	AJ235609	DQ307101
Podophyllum peltatum L.	Berberidaceae	AF092109	AF197591	*Coriaria myrtifolia* L.	Coriariaceae	AJ235443	L01897
Sabia campanulata Wall.	Sabiaceae	-	AM183414	*Larrea tridentata* Coult.	Zygophyllaceae	AY935860	Y15022
Sabia swinhoei Hemsl.	Sabiaceae	AF093395	FJ626616	*Oenothera elata* Kunth	Onagraceae	NC002693	NC002693
Meliosma veitchiorum Hemsl.	Sabiaceae	AF209626	AF206793	**ASTERIDS**			
Nelumbo lutea Willd.	Nelumbonaceae	EU642740	DQ182337	*Impatiens noli-tangere* L.	Balsaminaceae	-	AB043516
Nelumbo nucifera Gaertn.	Nelumbonaceae	D89550	FJ626615	*Impatiens repens* Moon	Balsaminaceae	AJ235503	-
Embothrium coccineum Forst.	Proteaceae	AF060429	DQ875857	*Ilex aquifolium* L.	Aquifoliaceae	-	FJ395601
Grevillea baileyana McGill.	Proteaceae	AF060434	-	*Ilex crenata* Thunb.	Aquifoliaceae	AJ235502	-
Grevillea robusta A. Cunn. ex R. Br.	Proteaceae	-	AF197589	*Panax ginseng* C.A. Mey.	Araliaceae	AY582139	AY582139
Platanus orientalis L.	Platanaceae	-	AY858644	*Atropa belladonna* L.	Solanaceae	NC004561	NC004561
Platanus occidentalis L.	Platanaceae	EU642741	L01943	*Nicotiana tabacum* L.	Solanaceae	NC001879	NC001879

1.4 Results

Sequence variability

The 5 partitions as well as the individual introns and spacers studied here differ considerably in their sequence length (see Table 3). By displaying a mean sequence length of 157 nt ranging from 131 to 193 nt the *rps3-rpl16* spacer is one of the shortest regions studied. In contrast the coding *matK* extents from 1495 to 1548 nt, thus being the longest region. Irregularly occurring nucleotide counts deviating from the triplet code within *matK* are most likely an artefact due to insufficient sequence editing appearing in sequence data downloaded from GenBank. The *trnK* group II intron (excluding *matK*) is showing a length variation of 235 nucleotides. It ranges from 867 to 1102 nt, a length extent similar to that of the *rpl16* group II intron (801 – 1122 nt). The latter is missing in *Erodium*, a condition known from two genera of the Geraniaceae as well as several representatives of Goodeniaceae and Plumbaginaceae (Campagna & Downie, 1998). However both regions differ considerably in length variation, proven by their deviating coefficent of variability (*Cv*), which stayed rather low for the *trnK* intron (5.5%), whereas it is raised to 14.8% for the intron within the *rpl16* gene. Comparing the *Cv*-values of all genomic regions surveyed it is conspicuous that the *rpl16* group II intron and the *rps3-rpl16* spacer are showing the highest amount of sequence variation (14.8% and 11.5%, respectively), followed by the *trnL-F* spacer (9.9%) and the *trnL* group I intron (9.3%). Both partitions were recognized as transcription units by Kanno & Hirai (1993). The residual partitions, *trnK/matK*, *petD* and *atpB-rbcL* are characterized by considerably lower *Cv* values that range from 3.5% to 5.6%. In relation to its mean sequence length the *rps3-rpl16* spacer provided the highest number of aligned sequence characters (139 nt and 751 characters, respectively), followed by two more spacers (*trnL-F* spacer: 297 nt and 1185 characters; *atpB-rbcL* spacer: 691 nt and 2387 characters), the *rpl16* intron, exposing 840 nucleotides and 2690 aligned positions, and the *trnK* intron plus the *petB-petD* spacer (860 nt, 2286 characters and 190 nt, 503 characters, respectively). With an amount of 50.9% the coding *matK* provided the highest number of variable characters per aligned position. In contrast it supplied only 3.3% (98 indels) of the overall coded indels, whereas 19.1% (565 indels) were encoded in the non-coding part of the *trnK/matK* partition.

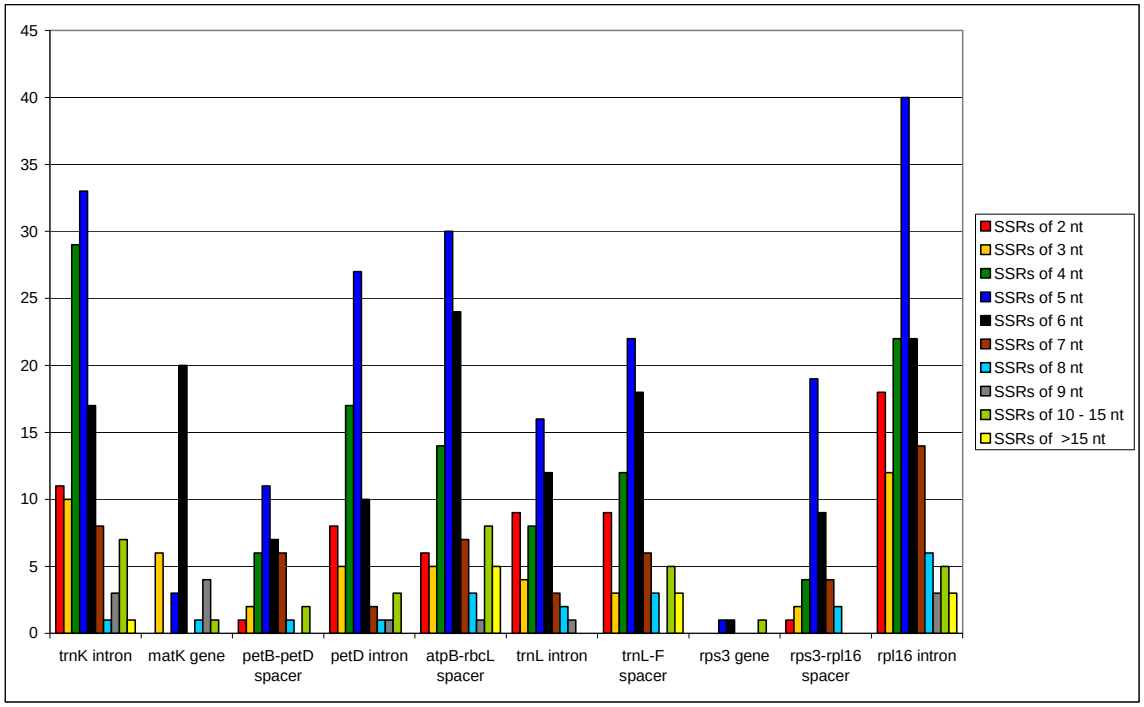


Figure 2: Number and distribution of length of simple sequence repeats within the genomic regions under study. Simple sequence repeats of 2 nucleotides and more were included; repeats of more than 15 nucleotides were summarized. SSRs=simple sequence repeats; nt=nucleotides.

The combined indel matrix provided a set of 2955 characters. Several lineages are characterized by the possession of certain synapomorphic indels, such as Sabiales. All members under survey share indel number 1523 within the *rpl16* intron (alignment position 7157 – 7205) as well as a short deletion (5 nt) localized in the 5'end of the *trnk* intron (indel number 1914, alignment position 8804 – 8808). Other indels are autapomorphic. A prominent example from the *atpB-rbcL* spacer is indel number 192, a deletion unique to *Rhipsalis* (alignment position 1063 – 2567). It is remarkable that by displaying a varying number of characters per region the proportion of coded indels is at least around 24%. A similar pattern is recognized with respect to the relative amount of the parsimony informative indel characters, which is about 25% to 33% for the individual genomic regions. Length mutations were in a large part identified as simple sequence repeats (SSR), mostly comprising 4 to 6 nucleotides and ranging in amount between 36.3 % in the *atpB-rbcL* spacer and 58.3 % in the *petB-petD* spacer (compare Table 3). The non-coding regions mainly include SSRs with a sequence length of 5 nt while *matK* is largely characterized by length mutations of 6 nucleotides, maintaining the open reading frame (see Figure 2). Nevertheless, large insertions are found in several taxa and genomic

regions studied, such as in the *atpB-rbcL* spacer of *Larrea* (82 nt; alignment position 966 – 1047) and *Mahonia* (111 nt, alignment position 2028 – 2138). The latter one is identified as being a tandem repeat of two repetitions.

A number of mutational hotspots (H) were excluded from the analyses due to length-variable poly A/T stretches (microsatellites) or difficulties in motif recognition caused by frequent and overlapping microstructural changes comprising several nucleotides. They were recognized in all partitions surveyed. Detailed information on extension and absolute position (referring to nucleotide positions in the absolute lengths starting at the 5' end of the respective genomic region) of each hotspot are given in Appendix B. Sequence stretches within mutational hotspots are generally ranging from 5 to 30 nt in length. Several very variable, unalignable sequence sections were identified within the 3' part of the *trnk* intron (H9; H10) as well as in the *rpl16* intron (H7; H12; H13), displaying sequence stretches up to 100 nucleotides and more in some taxa. In addition a number of very long autapomorphic insertions were excluded from analyses. A striking example is a length mutation of 391 nucleotides within the *atpB-rbcL* region of *Rhipsalis*, which comprises an inverted copy of 238 nt from the neighbouring *rbcL* gene.

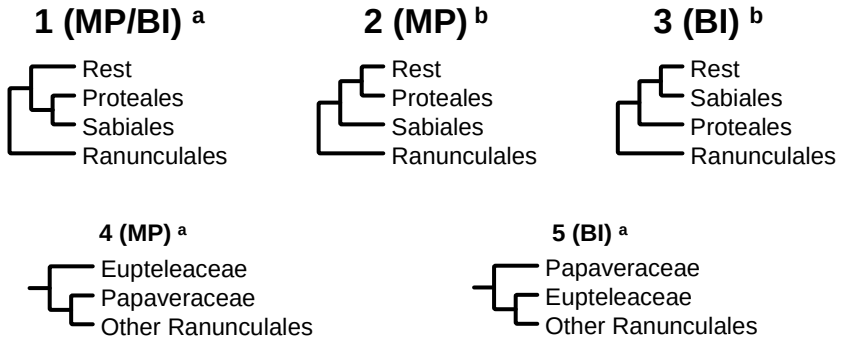
Table 3: Variation and relative contribution of the genomic regions studied. Number and quality of characters, *Cv*, indels coded, number of SSRs, and parsimony informative indels, as well as GC content are calculated with mutational hotspots excluded. SD=Standard deviation, No.-char.=Number of characters, var.-char.=variable characters, inf.-char.=informative characters, *Cv*= coefficent of variability; SSRs=simple sequence repeats, PI=parsimony informative, Ti/Tv ratio=transition/transversion ratio.

Region	**mean sequence length [bp]**	**SD**	***Cv* [%]**	**mean sequence length excl. hotspots [bp]**	**SD**	**No. char.**	**var. char. [%]**	**inf. char. [%]**	**No. of indels coded**	***Cv* [%]**	**No. SSRs [%]**	**PI indels [%]**	**GC-content [%]**
trnK intron	974	54	5.5	860	35	2286	30.8	22.3	565	24.7	42.9	27.3	36.3
matK gene	1525	10	0.7	1524	10	1856	64.7	50.8	98	5.3	41.9	23.5	34.1
petB-petD spacer	198	11	5.6	190	11	503	29.6	18.5	122	24.3	58.3	32.8	29.3
petD intron	722	25	3.5	657	19	1162	43.9	30.3	257	22.1	39.8	27.6	39.1
atpB-rbcL spacer	766	31	4.0	691	58	2387	24.5	17.1	493	20.7	36.3	29.6	31.7
trnL intron	495	46	9.3	450	33	915	37.4	26.6	238	26	40.9	27.3	36.7
trnL-3'exon	48	9	18.8	48	9	50	20	8	0	0	0	0	45.3
trnL-F spacer	364	36	9.9	297	33	1185	27.2	20.5	356	30	45.6	25.8	34.7
rps3-rpl16 spacer	157	18	11.5	139	16	751	16.5	12.3	187	24.9	37.0	31.6	29.9
rpl16 intron	989	146	14.8	840	120	2690	25.9	18.0	634	23.6	46.3	25.4	35.9

Phylogeny of early-divergimg eudicots

The combined data matrix of the genomic regions analysed (*trnK/matK*, *trnL-F*, *petD*, *atpB-rbcL* and *rpl16*) comprised 14140 characters in total (excluding hotspots). Altogether 4833 characters were variable and 3505 parsimony informative. The simple indel coding approach applied on the data matrix supplied 2955 binary indel characters that were added to the dataset. Relative contributions of the five individual partitions are shown in Table 4. The parsimony ratchet analysis resulted in one most parsimonious tree of 24381 steps (CI = 0.476, RI = 0.473) which is shown in Figure 4. Ranunculales are clearly identified as first branching lineage within the eudicot-clade by Maximum Parsimony (MP) (BS 99/95, as in the following the first value refers to statistical support obtained with the binary indel matrix included into analyses) as well as by Bayesian Inference (BI) (PP 1.0/1.0). The topology gained through BI (Figure 5) differs in the placement of Eupteleaceae and Papaveraceae inside the order of Ranunculales. According to MP *Euptelea* is resolved as representing the first branching lineage with weak bootstrap support (BS 67/60), whereas BI shows a sistergroup relationship between Papaveraceae and the remaining taxa of the order. However, statistical support for this hypothesis stayed moderate (PP 0.81/0.91). Ranunculales are followed by a clade comprising Sabiales and Proteales including Nelumbonaceae (BS 100/100, PP 1.0/1.0). This clade, exposing a sistergroup relationship between the two groups, gained moderate to high statistical support in MP and BI, respectively (BS 89/80, PP 0.96/0.92). Sabiales as well as Proteales are clearly identified as being monophyletic, receiving maximal statistical support for the respective nodes.

The family of Nelumbonaceae is resolved as being sister to a Proteaceae plus Platanaceae clade (BS 100/100, PP 1.0/1.0). Branching next are Trochodendrales (BS 70/70, PP 0.87/0.92), followed by Buxales (BS 100/100, PP 1.0/1.0), the latter being sister to the core eudicots. Both orders are shown to be monophyletic with maximum support. Buxales include Buxaceae as well as Didymelaceae.



topology tests

	AU	SH
1	0.903	0.929
2	0.133	0.266
3	0.087	0.201
4	0.257	0.228
5	embedded in 1-3	

Figure 3: Five alternative tree topologies used to perform the approximately unbiased test for the placement of Sabiales (1, 2, 3) and Eupteleaceae (4, 5). Three different topologies concerning the placement of Sabiales as inferred through 1) MP and BI, this study ([a]), 2) MP analyses by Worberg & al. (2007) ([b]), 3) BI by Worberg & al. (2007). 4) Simplified topology of the MP tree (this study), or 5) the BI tree, showing two different positions of Eupteleaceae. p-values are given in the table. AU=approximately unbiased test, SH=Shimodaira-Hasegawa test.

Inside the strongly supported core-clade, Gunnerales are depicted as first lineage, a scenario which received high statistical support in all approaches (BS 99/99, PP 1.0/1.0). The backbone of core eudicots is resolved in MP analyses but bootstrap support is lacking for various nodes, while several major clades were recognized with moderate to high confidence, such as Saxifragales (BS 99/96), Vitales (BS 100/100), rosids (BS 82/79), Caryophyllales (BS 100/100) and asterids (BS 98/96). BI resulted in a topology that mainly differs in the placement of Dilleniales, Santalales and Berberidopsidales. Statistical values of the respective nodes of the backbone that received no bootstrap support in MP were raised to a significant level while the five major lineages gained maximum statistical support.

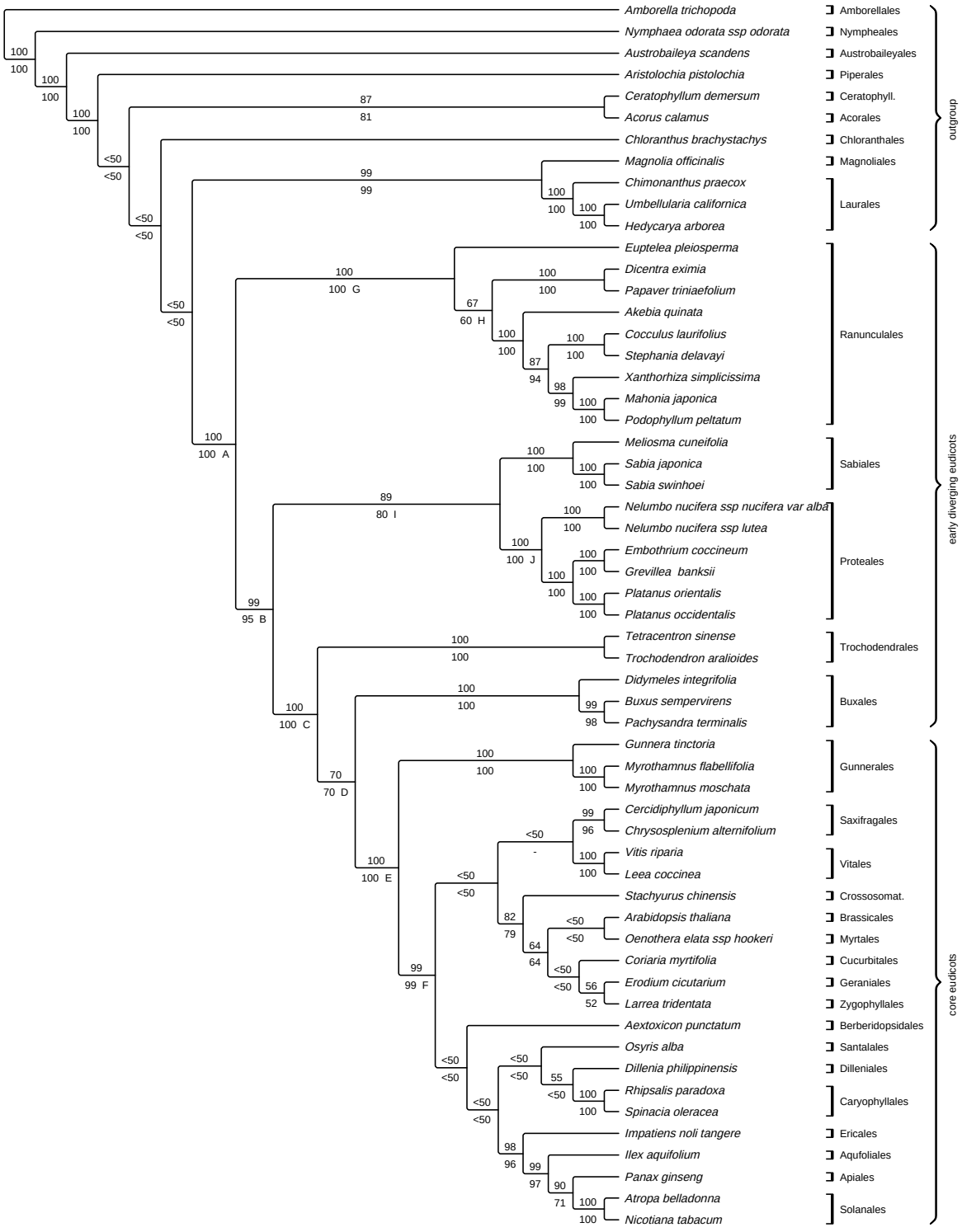


Figure 4: Strict consensus tree based on substitutions and indels of all 5 regions, inferred with MP. Values above and below branches are Bootstrap percentages, referring to substitutions plus indels or to substitutions only, respectively. Letters below branches indicate single evaluated nodes (compare Table 4).

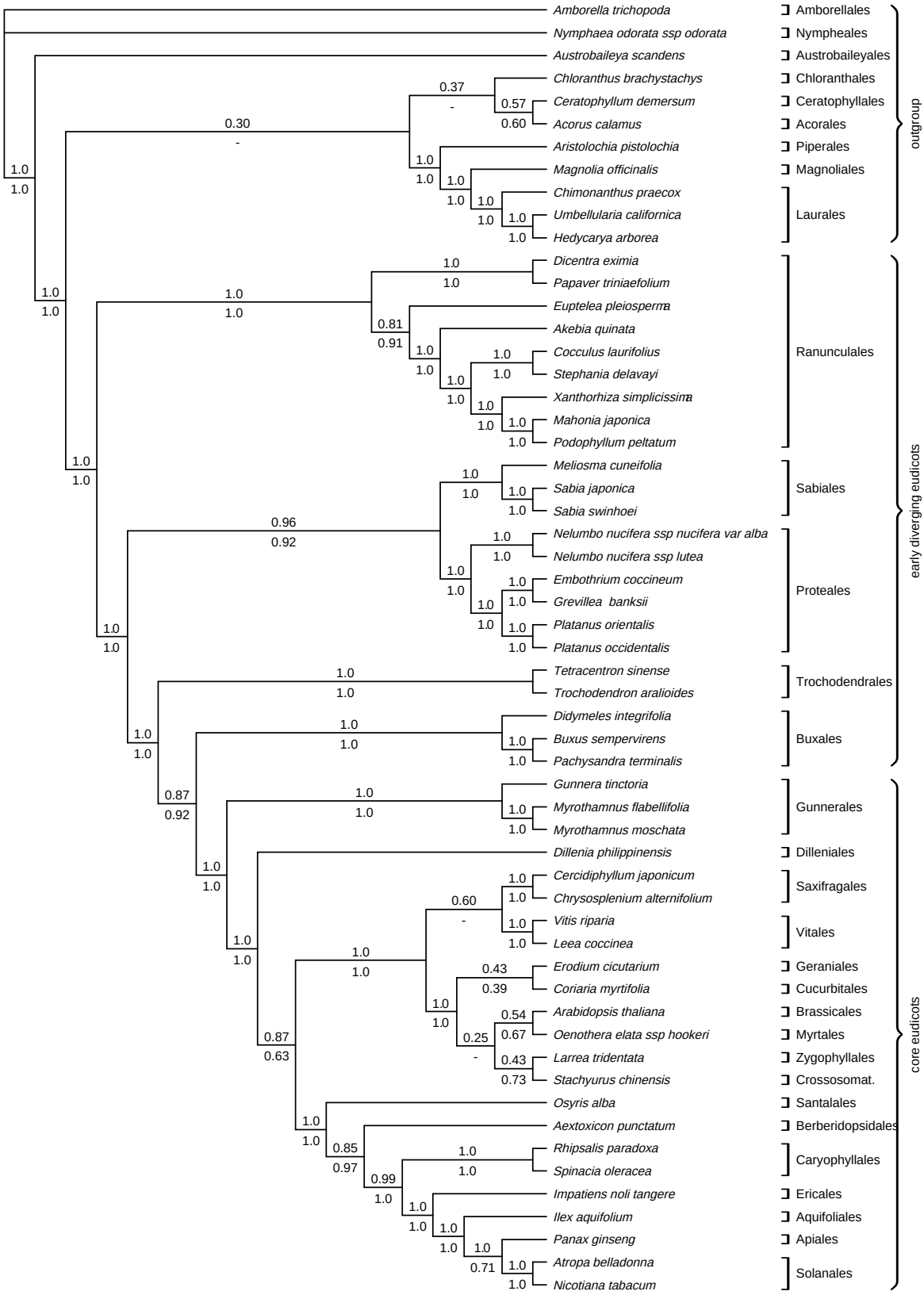


Figure 5: Bayesian tree based on the combined *trnK/matK-trnL-F-petD-atpB-rbcL-rpl16* matrix. Posterior Probabilities are depicted above (substitutions plus indels) and below (substitutions only) branches.

Table 4: Statistical values of selected nodes (A – J, see Figure 4), based on different combinations of the genomic regions under study, including substitutions as well as indels into calculations. Percentage of informative characters as well as number of characters and number of informative characters refer to substitutions plus indels, calculated with hotspots excluded. First values relate to bootstrap percentages from parsimony analyses, second values are posterior probabilities (BI). No. char.=number of characters, inf. char.=informative characters, "-"=node absent.

Combination	**Inf. char. [%]**	**No. char.**	**No. inf. char.**	**Node**									
				A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
trnK intron	23.3	2851	663	96/1.0	-/0.98	-/1.0	-/-	100/1.0	-/-	89/1.0	-/-	-/-	-/-
matK gene	49.5	1954	968	100/1.0	71/1.0	55/1.0	74/0.97	100/1.0	75/1.0	92/1.0	50/-	-/-	89/1.0
trnK/matK partition	34	4804	1631	100/1.0	84/1.0	94/1.0	53/0.84	100/1.0	81/1.0	100/1.0	<50/-	<50/-	84/1.0
petD partition	27.1	2051	556	98/1.0	68/1.0	-/-	-/-	100/1.0	62/1.00	86/1.0	-/-	-/-	-/-
atpB-rbcL partition	20.7	3085	640	95/1.0	-/-	95/1.0	76/0.89	90/1.0	<50/0.88	62/1.0	-/-	-/0.91	-/0.50
trnL-F partition	23.6	2744	648	100/1.0	66/1.0	96/1.0	64/-	99/1.0	<50/-	89/1.0	<50/-	-/-	87/1.0
rpl16 partition	19.1	4407	842	97/1.0	-/0.74	78/0.97	74/0.54	95/0.96	52/0.78	74/0.97	-/-	-/0.81	75/0.98
Spacers	19.6	5979	1171	100/1.0	65/0.79	96/1.0	64/0.69	100/1.0	57/0.94	96/1.0	<50/-	60/0.90	83/1.0
Group I intron	26.8	1153	309	90/1.0	-/0.99	<50/0.99	-/-	71/1.0	<50/-	<50/0.78	-/-	-/-	55/1.0
Group II introns	22.8	7596	1732	100/1.0	81/1.0	98/1.0	-/-	100/1.0	93/1.0	100/1.0	61/-	<50/-	-/0.86
Group I+II introns	23.3	8750	2041	100/1.0	91/1.0	100/1.0	-/-	100/1.0	96/1.0	100/1.0	56/-	72/-	73/1.0
Worberg & al. (2007)	32.2	6751	2172	100/1.0	98/1.0	99/1.0	75/0.81	100/1.0	97/1.0	100/1.0	74/-	-/-	99/1.0
Worberg & al. (2007) + *atpB-rbcL* partition	28.6	9837	2812	100/1.0	98/1.0	100/1.0	81/0.97	100/1.0	99/1.0	100/1.0	69/-	-/0.79	100/1.0
Worberg & al. (2007) +*trnK* partition	29.5	9601	2835	100/1.0	98/1.0	100/1.0	-/-	100/1.0	97/1.0	100/1.0	68/-	61/-	97/1.0
Worberg & al. (2007) + *rpl16* partition	27	11159	3014	100/1.0	98/1.0	100/1.0	87/0.75	100/1.0	99/1.0	100/1.0	71/-	54/0.53	100/1.0
This study	25.3	17095	4317	100/1.0	99/1.0	100/1.0	70/0.87	100/1.0	99/1.0	100/1.0	67/-	89/0.96	100/1.0

1.5 Discussion

1.5.1 Relationships among early-diverging eudicots

During the last decade a number of phylogenetic analyses mainly based on coding genes dealt with the relationships among the first-diverging eudicots and provided a profound framework (e.g. Hoot & al., 1999; Soltis & al., 2000; 2003; Hilu & al., 2003; Kim & al., 2004). By analyzing non-coding and rapidly evolving DNA from the large single-copy region of the chloroplast (*trnL-F*, *petD*, *matK*) and including indel information, Worberg & al. (2007) were able to present well supported phylogenetic hypotheses for the early-diverging eudicots, inferring Ranunculales as first branching lineage, followed by Sabiales, Protelaes, Trochodendrales, Buxales and the core eudicots. Even though statistical support obtained for the backbone nodes of the first diverging eudicots was generally high under parsimony, the second-branching position of Sabiales within the grade was only moderately supported (JK 83). In contrast, Bayesian inference resulted in a tree showing Sabiales branching after Proteales with no support (0.52 PP) while in Maximum likelihood no resolution for the respective positions of Sabiales and Proteales was gained. Therefore it was one of the central goals of the presented study to corroborate the branching order inside the basal eudicot grade, with an emphasis on the placement of Sabiales and Proteales. Assuming that the respective positions among the first-diverging eudicots could be confidently resolved by extending the data matrix of Worberg & al. (2007), complementary sequence data of two more group II introns (*trnK*, *rpl16*) as well as two spacers (*rps3-rpl16* and *atpB-rbcL*) was added. To consolidate the taxon-sampling a second species of *Sabia* (*Sabia swinhoei*) was included into analyses.

Ranunculales are sister to all other eudicots

As in most recent phylogenetic studies (e.g. Hoot & al., 1999; Soltis & al., 2000; 2003; Hilu & al., 2003; Worberg & al., 2007) Ranunculales were highly supported as being sister to the remainder of eudicots. Ever since the recognition of Eupteleaceae being a member of Ranuncuclales their true position within the order has been controversial. Initial studies using the plastid *rbcL* gene as a molecular marker revealed Eupteleaceae and Papaveraceae sensu lato (incl. Fumariaceae, *Hypecoum*, *Pteridophyllum*; Kadereit &

al., 1995) as being excluded from a core clade consisting of Circeasteraceae, Lardizabalaceae, Berberidaceae, Menispermaceae and Ranunculaceae (Chase & al., 1993; Savolainen & al., 2000b), although statistical support of the backbone nodes was lacking. An addition of the plastid *atpB* gene and nuclear ribosomal 18S sequences resulted in a topology which placed Papaveraceae basal to the remaining members of the Ranunculales (Hoot & al., 1999; Soltis & al., 2000), a scenario without support. The same phylogenetic hypothesis was presented by Doyle & Endress (2000) obtained through combined molecular and structural analyses, albeit the first branching position of Papaveraceae gained only weak bootstrap support (BS 65). This scenario was contradicted with the inclusion of 26S data in a four-gene analysis (*rbcL*, *atpB*, 18S, 26S) by Kim & al. (2004), which inferred Eupteleaceae as first branching lineage with moderate support under Parsimony (JK 70) while support was raised to significance using Bayesian inference. In the following a number of phylogenetic studies revealed Eupteleaceae as first-branching within Ranunculales (e.g. Hilu & al., 2003; Worberg & al., 2007; chapter 2; Wang & al., 2009 – combined molecular data of four genomic regions plus morphology) even though statistic values were never truly convincing. In contrast to the survey by Kim & al. (2004) and two comprehensive studies of relationships among Ranunculales enclosing both species of *Euptelea* (Wang & al., 2009; chapter 2 – including molecular data of seven non-coding and fast-evolving plastid regions), the combined analyses of Worberg & al. (2007) indicated that model based approaches might come to a different result, by placing *Euptelea* sister to Papaveraceae. However statistical support was lacking. In the present study MP and BI resulted in different topologies concerning the placement of *Euptelea* (see Figure 4 and 5). Neither the hypothesis assuming Eupteleaceae being the first branching lineage nor a sistergroup relationship between Papaveraceae and the remaining Ranunculales obtained through a model based approach gained significance in the combined analyses (BS 67/60 or PP 0.81/0.91, respectively). Moreover statistical values for the first-branching position of *Euptelea* clearly decreased in comparison to previous studies (e.g. Kim & al., 2004; Worberg & al., 2007 – JK 81/80; Wang & al., 2009 – BS 87, gained trough the addition of morphological characters). The application of the approximately unbiased test showed none of the two opposing scenarios as being significantly declinable, as indicated by the p-values in Figure 3. These results evidently pinpoint that the position of *Euptelea* has still to be considered as unclear. Eupteleaceae obviously are a very distinct lineage within Ranunculales, also indicated by morphological traits such as growth-form (big trees versus herbaceous and lianescent-

shrubby plants) and floral biology (e.g. wind-pollination syndrome; lack of perianth, long connective protrusion, pronounced dissymmetry of the floral base, long temporal gap between androecium and gynoecium initiation, small space for carpel initiation – Ren & al., 2007).

Proteales and Sabiales may be sister groups

As already shown in the combined analyses of Worberg & al. (2007) Sabiaceae were inferred as monophyletic within the present study, based on the inclusion of three species (*Sabia japonica*, *Sabia swinhoei*, *Meliosma cuneifolia*). However, the third genus of the family, *Ophiocaryon*, was not enclosed. While there is no doubt about the coherence of the family or rather order, its exact placement inside the early-diverging eudicots has remained an open question. Anatomical and morphological attributes like a wedge-shaped phloem and a nectary disk, rare traits inside first-diverging eudicots, point to a close relationship to Proteaceae (Kubitzki, 2007; for a review see Nandi & al., 1998). The three gene analysis of Hoot & al., (1999), based on molecular data of the *atpB*, *rbcL* and the 18S region and including one species of *Sabia*, revealed Sabiales as branching next after a clade consisting of Nelumbonaceae, Platanaceae and Proteaceae (classified as Proteales by APG II, 2003). This result was confirmed by the four gene analyses of Soltis & al. (2003), who added sequence data of the 26S gene and extended the taxon-sampling by including *Meliosma*. Though, statistical support was absent or stayed moderate (JK 76) for the respective nodes. In contrast parsimony analyses of Kim & al. (2004), carried on the basis of the same set of four molecular markers resulted in an inconclusive topology, indicating a close proximity to Trochodendraceae and Buxaceae, a scenario that seems rather unlikely. The broad-scale analysis of partial *matK* sequence data published by Hilu & al. (2003) resulted in a second prominent hypothesis by presenting *Meliosma* and thus Sabiales as being the second-branching lineage within early-diverging eudicots. Again statistical support was lacking in parsimony analyses whereas it was raised to the moderate level (PP 0.78) in Bayesian inference.

A major result of this study is that Sabiales emerged as sister to Proteales in both MP and BI. This result is in line with several comprehensive studies on angiosperm phylogeny (Qiu & al., 2006 – three-genome, eight-gene analyses; Moore & al., 2008 – 83-gene, 86-taxon plastid genome data set, Burleigh & al., 2009 – five gene, 567-taxon data matrix), but support was lacking or stayed on a moderate level in any case. As the only exception

PP-values were raised to significance through combining substitutions with indel information (PP 0.96, this study). It seems that this result is mainly due to the addition of the *atpB-rbcL* spacer as well as the *rpl16* partition (PP 0.91 or 0.81, respectively; Table 4). This inconclusive picture is summarized in the five-region analyses of Worberg & al. (2007) on the basis of a completed *matK* data matrix and additional data of the plastid *petD* and *trnL-F*. Their analyses resulted in moderate Jackknife support (JK 83) for the second-branching position of Sabiales within the early eudicot grade, with phylogenetic signal coming from complete *matK*, while the Bayesian tree showed *Sabia* and *Meliosma* branching after Proteales with no support (PP 0.52). Finally Maximum likelihood gave no resolution for the respective positions of Sabiales and Proteales.

For that reason the approximately unbiased test (Shimodaira, 2002) was applied for examining whether one of the three conclusive hypotheses is significantly more likely than the alternatives. Despite our evidence for a Sabiales/Proteales sister group relationship, these topological tests did not result in a significantly lower likelihood for the two alternative tree topologies (see p-values in Figure 3). It seems that the exact position of Sabiales continues to be an unanswered question, although the number of characters and parsimony informative sites was doubled in the current study in comparison to Worberg & al. (2007).

Trochodendrales and Buxales are successive sisters to core eudicots

The further topology of early-diverging eudicots is resolved as Trochodendrales branching off next, followed by Buxales and the core eudicots with Gunnerales being the first-diverging lineage in MP analysis as well as Bayesian inference. This result is in congruence with the findings of Worberg & al. (2007), albeit statistical support decreased to BS 70 in Parsimony analysis for the respective position of Trochodendrales. Previous studies were inconclusive about the exact placement of Buxales and the *Tetracentron-Trochodendron* lineage (e.g. Soltis & al., 2000; Kim & al., 2004 - MP), while partial *matK* sequences resulted in 91% JK for Buxales as sister to core eudicots in Maximum Parsimony and maximal statistical support in Bayesian analysis (Hilu & al., 2003). Complete *matK* as well as *trnL-F* provided congruent signal on Buxales as sister to core eudicots (JK 87 or 63, respectively), whereas *petD* resolved Trochodendrales as sister to the core eudicots with moderate support (JK 90, Worberg & al., 2007). In the current study the added *atpB-rbcL* as well as the *rpl16* partition supplied coincident phylogenetic

signal in parsimony analyses (BS 76 and 74, respectively), while the *trnK* partition (excluding *matK*) is incongruent by facilitating a sistergroup relationship of Trochodendrales and core eudicots (trees not shown). Nevertheless the combined analyses yielded moderate statistical support under parsimony for the *Tetracentron*+*Trochodendron* clade diverging before Buxales, which was increased within BI (PP 0.87/0.92).

1.5.2 Testing hypotheses of a unique genome history with parsimony, Bayesian and likelihood approaches

As noticed in a number of studies (e.g. Hilu & al., 2003; Quandt & al., 2007; Wanke & al., 2007; Olsson & al., 2009) Maximum Likelihood analyses as well as Bayesian Inference often resulted in more resolved and supported topologies in comparison to Maximum Parsimony. This is due to a better exploitation of the information provided by the underlying data. Additionally it is possible to accurately choose the "best" model, resulting in a profound evaluation of evolutionary scenarios in a statistic context. However, as pointed out by Kelchner & Thomas (2007) both, the conceptual as well as the formal model must represent the evolutionary process that resulted in the data under study, this being a fundamental requirement for a phylogenetic reconstruction to be accurate. As described by Kelchner & Thomas (2007), changing the hypothesis on the evolutionary process acting at a certain site can result in a differing valuation of branch lengths. This can lead to an altering conclusion concerning the portion of mutational change between two sequences. Therefore every method relying on a correct assessment of the amount of evolutionary changes among different lineages is dependent on an adequate model.

Comparing the combined analyses of Worberg & al. (2007) with the present study concerning resolution and statistical support gained through parsimony as well as model based approaches an interesting picture emerged, especially concerning difficult to resolve positions such as the branching order among Eupteleaceae, Papavaraceae and the remaining Ranunculales or the exact placement of Sabiales within early-diverging eudicots. By applying parsimony analyses both investigations resulted in the recognition of *Euptelea* being first-branching within Ranunculales. With the addition of four more non-coding regions bootstrap support clearly decreased. Topologies changed through the

application of model based methods. Within the five-region analyses of Worberg & al. (2007) Maximum Likelihood (ML) as well as Bayesian Inference revealed a sistergroup relationship of Eupteleaceae and Papaveraceae, while including nine genomic regions resulted in the recognition of Papaveraceae being the first-branching lineage through BI with increased PP-values. However, statistic support was either lacking or not truly convincing. It seems that model based methods, by being more "sensitive", pointed on difficult to clarify phylogenetic problems. This is also true considering the exact position of Sabiales inside early-branching eudicots. The study of Worberg & al. (2007) yielded an inconclusive picture in this respect. While model based approaches showed *Sabia* and *Meliosma* branching after Proteales without support (BI) or gave no resolution for the respective positions of Sabiales and Proteales (ML), Maximum Parsimony revealed Sabiales as second-branching lineage within early-diverging eudicots. Support stayed moderate for this scenario, just as for the alternative topology gained through parsimony analyses within the current study. Statistic values were raised for a possible sistergroup relationship between both orders within Bayesian calculations and reached significance by combining substitutions and indel information.

The reliability of statistical values has already been subject to empirical studies. Simmons & al. (2004) clearly demonstrated in an example that both, jackknife (bootstrap) and Bayesian methods, significantly differ from an ideal support index. While jackknifing (bootstrapping) underestimated statistic support values, they were clearly overestimated by Bayesian calculations. In addition the dimension of Bayesian values overestimating statistical support obviously exceeds the dimension of jackknife underestimating support. Therefore Simmons & al. (2004) stated that Posterior Probabilities gained through Bayesian Inference should not be taken as probabilities of clades being correctly resolved. This should be especially considered in the case of moderate support values, as noticed in the present study for the respective positions of Eupteleaceae, Sabiales or Trochodendrales. Therefore several topology tests were carried out in the present investigation. A number of trials have been used for evaluating the confidence of tree selection within phylogenetics, such as the bootstrap probability (BP; Felsenstein, 1985) and the Kishino-Hasegawa tests (KH; e.g. Kishino & Hasegawa, 1989). Probability values (=*P*-values) gained through both methods represent the possibility of the appropriate tree being the true tree. However, both, the BP test, as well as the KH test are biased by comparing a large number of trees at the same time, a fact leading to overconfidence in the wrong trees (for a review see Shimodaira, 2002). Several multiple comparisons

methods like the Shimodaira-Hasegawa (SH) or the weighted Shimodaira-Hasegawa (WSH) test where shown to adjust the selection bias of the KH test (e.g. Shimodaira & Hasegawa, 1999). Accessory to the SH test the approximately unbiased (AU) test as published by Shimodaira (2002) was chosen within the current study. This method is less conservative than the SH test and at the same time it squares the selection bias ignored in both the BP and KH test. The AU test is based on a multiscale bootstrap procedure, resulting in the approximately unbiased *P*-value calculated from the change in the bootstrap probabilities along the changing sequence length. Its application clearly revealed none of the tested topologies being significantly declinable. This result sustains the assumption that moderate statistic values should be handled with care.

1.5.3 Molecular evolution of genomic regions studied

Mutational hotspots in non-coding genomic regions have been subject to several studies on molecular evolution. It has been shown that these hotspots (H) are closely related to the secondary structure, thus corresponding to certain stem-loop elements where functional constraints are expected to be lowest (Borsch & al., 2003; Quandt & al., 2004; Löhne & Borsch, 2005). Worberg & al. (2007) introduced the question whether a similar pattern can be found in more derived eudicots as well. Therefore *petD* and *trnL-F* data of eudicots and basal angiosperms was compared in their study due to the position of hypervariable sequence parts referring to mutational hotspots. Several sequence stretches were identified as microsatellites not present in basal angiosperms (*petB-D* spacer, *trnL* intron, *trnL-F* spacer; compare Borsch & al., 2003; Löhne & Borsch, 2005). It was assumed that nucleotide substitutions must have resulted in longer A/T-stretches, which display an elevated probability for slipped-strand mispairing due to higher mutational rates, increasing with length (see Levinson & Gutman, 1987). Within the *petD* intron mutational hotspots clearly corresponded to stem-loops, with their position in largely accordance with the findings of Löhne & Borsch (2005) concerning basal angiosperms. Similarly, extremely variable sequence stretches within the *trnL* group I intron mainly corresponded to the terminal stem-loop parts of the usually least constrained P6 and P8 elements of the secondary structure (Borsch & al., 2003; Quandt & al., 2004; Quandt & Stech, 2005). In the current study two more spacers (*atpB-rbcL*, *rps3-rpl16*) as well as two additional group II introns (*trnK*, *rpl16*) were taken into consideration. All of these

genomic regions under survey displayed a large number of mutational hotspots as illustrated in Appendix B.

Altogether nine mutational hotspots were identified within the *atpB-rbcL* spacer and excluded from analyses. H1 to H8 are characterized by A/T-homonucleotide stretches of a different extend. Shorter regions such as H1, H2 and H3 contain single mononucleotide stretches that span up to 19 As in individual taxa, while larger mutational hotspots are composed of several microsatellites (e.g. H6). In addition an autapomorphic insert of 391 nucleotides was determined in the *atpB-rbcL* region of *Rhipsalis paradoxa* and excluded from calculations (H9). It comprises an inverted copy of 238 nucleotides from the neighbouring *rbcL* gene and is following a deletion, spanning about two third of the spacer. Since this seems to be an exceptional pattern within eudicots, a further investigation on the organisation of the chloroplast genome of *Rhipsalis* could lead to interesting findings. In the *rps3-rpl16* spacer five microsatellites (poly-A/Ts, H1-H5) were indentified, four of them being extremely short, and excluded due to the rules outlined in Olsson & al. (2009). Accordingly, poly-mononucleotide stretches spanning more than four nucleotides and displaying a length variation of at least two nucleotides should be excluded from analyses to prevent an involvement of spurious indel information.

By being the most length-variable region used in the current study the number of detected mutational hotspots was extremely high within the *rpl16* intron. Plotting these regions on the stylized secondary structure of a group II intron (Michel & al., 1989; Toor & al., 2001; Kelchner, 2002) on the basis of the annotation presented by Kelchner (2002) it became clear, that high variable sequence parts are corresponding to loops and bulges. H1 to H8 are located within the highly complex domain I (DI), mainly subdomains c and d. As in the *petD* intron of basal angiosperms (Löhne & al., 2005) and eudicots (Worberg & al., 2007) one mutational hotspot (H7) is found in the d2 stem-loop. It extends up to 99 nt in *Akebia quinata*. H8 is located in the d3 stem bulge, being a short poly-mononucleotide stretch of one to seven As. Since the *rpl16* intron is missing subdomains a and b in domain I and therefore the α tertiary interaction with the d3 stem bulge (Kelchner, 2002), it seems that this part of the intron is less constrained. Five mutational hotspots were detected in domain IV altogether covering up to 53.3% within *Dillenia philippinensis* (compare Table 6). A high degree of variability in size resulting in a raised percentage of excluded sites and an increased number of indels as compared to the remaining domains of the *rpl16* intron was already reported for Apioideae (Downie & al., 2000) as well as for

the Neckeraceae belonging to the pleurocarpous mosses (Olsson & al., unpubl.). Comparing domain IV of the *rpl16* intron with the corresponding structure of the *petD* intron and the *trnK* intron (including the maturase open reading frame) clearly revealed its much more higher variation in length and an increased percentage of excluded sequence information (Table 5 and 6). However, mutational hotspots occurred in all three exemplary domains studied. Additionally, the relative amount of coded indel characters stayed at almost the same high-grade level. Similar findings were made on the *petD* intron in basal angiosperms (Löhne & Borsch, 2005). These high levels of length-variability in domain IVs of chloroplast group II introns may be partly explained by their special conditions due to the maturase open reading frame (ORF) and its loss (for a review see Kelchner, 2002).

Table 5: Actual length, sequence length with hotspots excluded and percentage of sequence data excluded from analyses calculated for domain IV of three group II introns.

Taxon	sequence length– DOM IV [bp]			sequence length without hotspots [bp]			sequence length –hotspots [%]		
	trnK	***petD***	***rpl16***	***trnK***	***petD***	***rpl16***	***trnK***	***petD***	***rpl16***
Amborella trichopoda	1802	159	347	1705	152	273	5.4	4.4	21.3
Nymphaea odorata ssp tuberosa	1806	84	72	1731	77	52	4.2	8.3	27.8
Austrobaileya scandens	1779	161	226	1719	154	168	3.4	4.3	25.7
Ceratophyllum demersum	1695	124	265	1695	112	164	0	9.7	38.1
Acorus calamus	1772	167	233	1734	160	156	2.1	4.2	33
Chloranthus brachystachys	1765	160	271	1722	153	187	2.4	4.4	31
Aristolochia pistolochia	1781	152	307	1743	145	203	2.1	4.6	33.9
Magnolia officinalis	1758	154	233	1707	147	168	2.9	4.5	27.9
Umbellularia californica	1760	154	236	1717	147	171	2.4	4.5	27.5
Hedycarya arborea	1762	154	228	1720	147	168	2.4	4.5	26.3
Chimonanthus praecox	1751	153	230	1705	146	170	2.6	4.6	26.1
Euptelea pleiosperma	1761	154	237	1713	147	172	2.7	4.5	27.4
Akebia quinata	1743	166	246	1705	159	171	2.2	4.2	30.5
Dicentra eximia	1750	149	230	1717	142	146	1.9	4.7	36.5
Papaver triniaefolium	1753	160	220	1718	153	168	2	4.4	23.6
Cocculus laurifolius	1770	154	284	1741	147	209	1.6	4.5	26.4
Stephania delavayi	1801	155	286	1775	148	191	1.4	4.5	33.2
Xanthorhiza simplicissima	1749	172	202	1722	159	167	1.5	7.6	17.3
Mahonia japonica	1757	127	53	1722	120	43	2	5.5	18.9
Podophyllum peltatum	1741	176	261	1719	169	175	1.3	4	33
Sabia japonica	1762	161	232	1723	154	165	2.2	4.3	28.9
Sabia swinhoei	1762	160	232	1723	153	165	2.2	4.4	28.9
Meliosma cuneifolia	1753	166	233	1711	154	181	2.4	7.2	22.3
Nelumbo nucifera ssp nucifera	1750	169	248	1714	162	167	2.1	4.1	32.7
Nelumbo nucifera ssp lutea	1752	165	243	1716	158	167	2.1	4.2	31.3
Embothrium coccineum	1757	167	240	1719	160	183	2.2	4.2	23.8
Grevillea banksii	1759	167	239	1719	160	182	2.3	4.2	23.8
Platanus orientalis	1772	160	284	1734	153	184	2.1	4.4	35.2
Platanus occidentalis	1772	160	282	1734	153	184	2.1	4.4	34.8
Tetracentron sinense	1750	161	127	1704	154	79	2.6	4.3	37.8
Trochodendron aralioides	1754	160	237	1713	153	177	2.3	4.4	25.3
Didymeles integrifolia	1784	195	256	1736	157	187	2.7	19.5	27
Buxus sempervirens	1738	166	256	1703	159	190	2	4.2	25.8
Pachysandra terminalis	1739	156	232	1696	149	168	2.5	4.5	27.6
Gunnera tinctoria	1770	164	289	1719	157	191	2.9	4.3	33.9
Myrothamnus flabellifolia	1767	173	295	1717	150	192	2.8	13.3	34.9
Myrothamnus moschata	1766	173	295	1722	150	183	2.5	13.3	38
Cercidiphyllum japonicum	1743	187	260	1712	164	186	1.8	12.3	28.5
Chrysosplenium alternifolium	1781	162	222	1725	155	144	3.1	4.3	35.1
Vitis riparia	1739	176	316	1716	161	209	1.3	8.5	33.9
Leea coccinea	1727	177	330	1698	162	220	1.7	8.5	33.3
Dillenia philippinensis	1776	234	383	1735	166	179	2.3	29.1	53.3
Aextoxicon punctatum	1738	164	260	1701	157	186	2.1	4.3	28.5
Osyris alba	1743	168	298	1706	161	201	2.1	4.2	32.6
Rhipsalis paradoxa	1786	165	363	1745	152	279	2.3	7.9	23.1
Spinacia oleracea	1760	151	204	1719	146	170	2.3	3.3	16.7
Erodium cicutarium	1765	170	0	1701	160	0	3.6	5.9	0
Coriaria myrtifolia	1764	175	347	1731	166	205	1.9	5.1	40.9
Arabidopsis thaliana	1793	131	273	1753	122	211	2.2	6.9	22.7
Oenothera elata	1750	194	325	1715	164	158	2	15.5	51.4
Larrea tridentata	1766	151	175	1737	144	151	1.6	4.6	13.7
Stachyurus chinensis	1768	188	347	1727	177	234	2.3	5.9	32.6
Impatiens noli-tangere	1790	209	309	1701	191	193	5	8.6	37.5
Ilex aquifolium	1795	160	183	1726	148	114	3.8	7.5	37.7
Atropa belladonna	1739	192	273	1715	185	169	1.4	3.6	38.1
Nicotiana tabacum	1759	192	276	1731	185	169	1.6	3.6	38.8
Panax ginseng	1778	164	196	1719	157	127	3.3	4.3	35.2

Table 6: Variation and relative contribution of domain IV, calculated for the three group II introns under study. Length range, number and quality of characters as well as number and percentage of indels coded are calculated with mutational hotspots excluded; SD=Standard deviation, No.-char.=Number of characters, var.-char.=variable characters, inf.-char.=informative characters.

Region	**length range [bp]**	**mean sequence length [bp]**	**SD**	***CV* [%]**	**length range excl. hotspots [bp]**	**mean sequence length excl. hotspots [bp]**	**SD**	**No. char.**	**var. char. [%]**	**inf. char. [%]**	**No. of indels coded**	**Share of total number of indels coded in respective intron [%]**
trnK	1695 – 1806	1762	20	1.1	1695 – 1775	1720	15	2647	52.1	41	338	51%
petD	84 – 234	164	21	12.8	77 – 191	153	17	376	41.2	29.5	116	45.1%
rpl16	0 – 383	250	69	27.6	0 – 279	171	45	1042	16.4	12	273	43.1%

The *trnK* intron, the sole group II intron under study maintaining the ORF, is characterized by the possession of 11 mutational hotspots. Assigning these highly variable regions to the introns secondary structure presented by Hausner & al. (2006) it became clear that five of them are located in domain one, mainly subdomain d. H1 and H4 are due to poly-A/T stretches with a length up to 19 nucleotides, while H2 and H3 are resulting from big inserts within *Didymeles integrifolia* (161 nt) and *Aristolochia pistolochia* (82 nt), the latter corresponding to highly variable sequence parts within the d2 subdomain of the *petD* intron in basal angiosperms and eudicots (Löhne & Borsch, 2005; Worberg & al., 2007) as well as the *rpl16* intron (this study). Mutational hotspots were also determined within the domain 2 stem-loop (H6) and the less constrained terminal loop of domain 3 (H7, H8), all of them referring to microsatellites. These findings are in accordance with the results of Hausner & al. (2006), who compared eight representatives of the angiosperms in their study on the evolution of the *trnK* intron and clearly showed sequence deviations within the terminal part of domain 2 just like a variance of 52 nucleotides in the domain 3 stem-loop. As already mentioned above highly variable sequence parts were also found within domain IV of the *trnK* intron. H1 (*matK*) is a short length-variable satellite about 591 nt downstream the *matK*-gene, occurring in four members of the Ranunculales, *Amborella trichopoda* and *Gunnera tinctoria*. It ranges from three to nine nucleotides, thus maintaining the ORF. Hotspot H9 is following the maturase open reading frame and spans a highly divergent region up to 91 nucleotides at the beginning of the 3' *trnK*, which is just located in the stem-loop of domain IV. Sequence divergence was also noticed in the terminal loop of domain VI (H10), as already described by Hausner & al. (2006) and generalized for group two introns in Kelchner (2002). In accordance with the study of Hausner & al. (2006) hypervariable sequence parts referring to mutational hotspots in the *trnK* intron are rather found in loops than stems, a pattern, as mentioned above, already recognized in a number of surveys and for all the group two introns under study in eudicots. Nevertheless a more detailed and comparative examination on secondary structure, functional aspects and underlying mechanisms is needed to fully understand and to make generalizations on the pattern of molecular evolution within group II introns.

However, laboratory effort is affected by frequency and extend of poly A/T stretches within a genomic region. By displaying a number of mononucleotide stretches consisting of more than ten repeat units the sequencing of the *rpl16* partition required three or more sequencing primer reads in many cases. In contrast genomic regions such as the *atpB*-

rbcL spacer or the *petD* partition could be easily completely sequenced using universal primers (compare Worberg & al., 2007).

Comparing the partitions under survey, the *rpl16* intron is clearly shown to be the most variable region in length, followed be the *rps3-rpl16* spacer and the *trnL-F* partition (*trnL* group I intron and *trnL-F* spacer). The *rpl16* group II intron varies greatly in size within different land plants lineages, from 536 nucleotides in *Marchantia polymorpha* (Ohyama & al., 1986) to 1411 nt in *Spirodela oligorhiza* (Posno & al., 1986). Actually it is missing in several Geraniaceae, Plumbaginaceae and Goodeniaceae (Campagna & Downie, 1998), a finding confirmed for *Erodium cicutarium* in the current study. According to Campagna & Downie (1998) the intron possesses a size of about 1 kb in most angiosperms. This is in congruence with the present study were it ranged between 801 nucleotides in *Nymphaea odorata* and 1122 nt in *Dillenia philippinensis*. In addition a number of analyses carried out on different taxonomic levels clearly revealed the intron being more susceptible to length mutations in relation to other non-coding genomic regions used in phylogenetic reconstructions (Apioideae – Downie & al., 2000; Laurales – Renner & Chanderbali, 2000; Nymphaeales – Löhne & al., 2007), a result substantiated in this survey. The high variability in terms of length mutations occurring in the *trnL-F* spacer as well as in the *trnL* intron within eudicots was already noted by Worberg & al. (2007) and is in conformity with observations on basal angiosperms made by Borsch & al. (2003).

It is striking that despite a notable varying number of nucleotides and characters per region the proportion of coded indels as well as the relative amount of parsimony informative indel characters stayed at almost one level, a pattern already recognized within a study on Ranunculales on the basis of three non-coding and fast-evolving plastid regions (chapter 2). Besides it is noteworthy that coded length mutations were in large parts identified as being simple sequence repeats (compare Table 3), a finding in congruence with studies on the *trnT-F* as well as the *petD* region in basal angiosperms (Borsch & al., 2003; Löhne & Borsch, 2005) and the chloroplast inverted repeat (Graham & al., 2000). These observations lead to the suggestion that indels not assignable to certain SSRs were to a great extent concentrated in highly variable parts of the datasets and thus excluded from analyses. As in previous phylogenetic examinations (e.g. Löhne & Borsch, 2005; Worberg & al., 2007; Salomo, unpubl.; Borsch & Quandt, 2009), simple sequence repeats mostly range from five to six nucleotides in length (Figure 3). This pattern is difficult to explain since underlying molecular mechanisms are not fully understood yet.

1.5.4 Phylogentic structure

Phylogenetic structure in slowly-evolving versus rapidly-evolving DNA

The predominant view in molecular systematics favours the application of slowly-evolving or conservative DNA for inferring phylogenetic relationships at deeper taxonomic levels. This is due to the assumption of rapidly-evolving and non-coding being inappropriate based on putative high levels of homoplasy caused by multiple substitutions and frequent microstructural changes leading to non-alignability. However, recent phylogenetic studies using the fast-evolving *matK* gene from the large single copy region of the plastome and differing sets of plastid spacers and introns yielded well resolved and highly supported topologies for early-diverging angiosperms as well as for early-branching eudicots (e.g. Borsch & al., 2003; Löhne & Borsch, 2005; Worberg & al., 2007). Furthermore is was shown by Worberg & al. (2007) that small non-coding regions like *trnL-F* or *petD*, with a mean sequence length excluding mutational hotspots of 755 or 840 nt, respectively were resolving most of the eudicot tree. This result is comparable to the considerably longer *rbcL* gene, which comprises about 1400 nucleotides. These findings lead to the presumption of chloroplast introns and spacers having more phylogenetic signal than coding genes. Therefore the phylogenetic structure of five different data sets representing slowly-evolving protein-coding plastid genes (*atpB*, *rbcL*) and non-coding regions from the large single copy region of the chloroplast (spacers, the sole group I intron from the plastome and group II introns) was measured and compared applying the method developed by Müller & al. (2006). Beyond molecular data of the rapidly-evolving coding *matK* was subjected to analyses. The chosen approach uses resampling of identical numbers of parsimony-informative characters and evaluates various statistics of overall tree robustness and phylogenetic signal via a set of significance tests. A number of recent phylogenetic studies that were based on non-coding and fast-evolving DNA incorporated indel information into analyses and showed microstructural changes to be a reliable source of additional information (e.g. Worberg & al., 2007; Löhne & al., 2007; this study). Since there is no possibility to include this additional data into investigations on phylogenetic structure calculations refer to substitutions only. Analyses emerged on a result that considerably differs from the hypothesis presented above. Comparing all datasets it became clear that the slowly-evolving *atpB* gene provides one of the highest amounts of phylogenetic structure per informative character as well as per aligned position (see Table 8). Furthermore it is shown that it outstrips the rapidly-evolving *matK* in both respects. The *matK* gene is

known to exhibit different patterns of evolution in comparison to other plastid genes. It was shown that it has the highest rates of overall substitutions in comparison with other coding regions commonly applied in phylogenetic reconstructions, especially at non-synonymous sites, with substitution rates being not as strongly shifted towards the third codon position as normally described for other genes such as *atpB* and *rbcL* (Olmstead & Palmer, 1994; Johnson & Soltis, 1995; Müller & al., 2006). This pattern obviously results in a higher percentage of informative characters in comparison to the slowly-evolving protein-coding genes (compare Table 7). However, the calculation of bootstrap values for the individual topologies resulted in the recognition of *atpB* and *matK* performing at the same level.

In contrast the *rbcL* gene was found at the other end of the spectrum, a finding in line with the investigation of Müller & al. (2006) made on data of early-diverging angiosperms. Within their study the non-coding *trnT-F*, consisting of two spacers (*trnT-L* and *trnL-F*) and the *trnL* group I intron, and the rapidly-evolving *matK* clearly outperformed the more slowly-evolving *rbcL*. Both regions displayed a higher percentage of parsimony-informative characters as well as a significantly higher average phylogenetic signal per informative site. Additionally it was shown that phylogenetic structure per parsimony informative site is higher in the *trnT-F* region than in *matK*. Furthermore the non-coding region, being a combination of different partitions of non-coding DNA, displayed the highest amount of phylogenetic structure per sequenced nucleotide, followed by *matK* and *rbcL*. This pattern was considered to be correlated to different modes of molecular evolution of the genomic regions, since functional constraints are supposed to be lower in non-coding DNA than in coding regions (e.g. Kelchner & Clarke, 1997; Kelchner, 2002).

Within the present study the non-coding partitions were analysed separately, resulting in three different data matrices. Contrasting these partitions and *matK* revealed a new picture. Addressing the phylogenetic signal per parsimony informative position resulted in the recognition of the coding *matK* standing between the introns and the spacers, thus in this respect being comparable to the non-coding partitions under survey. A different pattern is revealed by considering the signal per aligned position. The non-coding partitions are shown to be ranking between *atpB* and *matK* at the one hand and the *rbcL* gene at the other hand. This placement is due to largely staggered alignments caused by frequent microstructural changes, mainly occurring in less constrained DNA. Therefore a methodical adjustment is needed to normalize this.

Table 7: Total number of characters, informative characters, referring to substitutions only, percentage of informative characters, as well as variation calculated for six different data partitions compared regarding their phylogenetic structure. No.-char.=Number of characters, inf.-char.=informative characters, var.-char.=variable characters, SD=Standard deviation. All calculations were carried out with mutational hotspots excluded.

Dataset	No. char.	No.inf.char.	inf. char. [%]	var. char. [%]	divergence [%]	length range [bp]	mean sequence length [bp]	SD
atpB gene	1497	422	28.2	39.7	6.832	1275-1497	1465	45
group I intron	915	243	26.6	37.4	12.266	295-487	450	33
group II introns	6137	1338	21.8	31.1	15.351	1588-2544	2356	122
matK gene	1856	942	50.8	64.7	16.427	1495-1548	1524	10
rbcL gene	1430	400	28.0	40.4	7.107	531-1428	1365	129
spacers	4825	831	17.2	24.5	16.48	963-1450	1318	67

Table 8: Differences in phylogenetic structure measured by sampling identical numbers exclusively from parsimony-informative characters from six different data matrices (group I intron, *rbcL*, *atpB*, spacers, *matK*, group II introns; top) or identical numbers of characters from all matrices (bottom). SE=standard error (calculated using 100 random sampling replicates), PI characters=parsimony informative characters, g I=group I intron, g II=group II introns. Measured differences are significant at α=0.05. Differences in phylogenetic structure are based on one statistic: R of first data matrix minus R of second matrix, from left to right. R was calculated on the basis of bootstrap proportions.

Comparison						Statistic R [mean]	SE	95% confidence interval		Higher in…
group I intron	***rbcL* gene**	***atpB* gene**	**spacers**	***matK* gene**	**group II introns**	**PI characters**		**ConLB**	**ConUB**	
						0.0227	0.0032	0.0165	0.0290	*atpB*
						0.0263	0.0030	0.0203	0.0322	*atpB*
						0.0503	0.0024	0.0455	0.0550	*atpB*
						0.0011	0.0021	-0.0031	0.0052	insignificant
						-0.0115	0.0029	-0.0173	-0.0057	g II
						-0.0184	0.0025	-0.0234	-0.0134	*matK*
						0.0660	0.0018	0.0625	0.0696	g I
						-0.0002	0.0024	-0.0048	0.0045	insignificant
						-0.0065	0.0019	-0.0102	-0.0027	g II
						-0.1210	0.0015	-0.1239	-0.1180	*atpB*
						-0.0939	0.0027	-0.0991	-0.0887	g II
						-0.0971	0.0028	-0.1024	-0.0917	*matK*
						-0.0715	0.0021	-0.0757	-0.0674	spacers
						-0.0465	0.0025	-0.0514	-0.0417	g II
						-0.0432	0.0016	-0.0463	-0.0400	*matK*
group I intron	***rbcL* gene**	***atpB* gene**	***matK* gene**	**spacers**	**group II introns**	**All characters**				
						0.0785	0.0032	0.0723	0.0847	*atpB*
						-0.0687	0.0019	-0.0724	-0.0649	*matK*
						0.1469	0.0027	0.1416	0.1521	*atpB*
						-0.0142	0.0025	-0.0192	-0.0093	*atpB*
						0.0201	0.0025	0.0152	0.0250	g I
						-0.1387	0.0025	-0.1436	-0.1338	*matK*
						0.0542	0.0022	0.0499	0.0584	g I
						0.0703	0.0029	0.0646	0.0760	g I
						0.1386	0.0029	0.1331	0.1442	*matK*
						0.1944	0.0024	0.1896	0.1992	*matK*
						-0.1266	0.0017	-0.1300	-0.1232	*atpB*
						-0.0591	0.0034	-0.0657	-0.0526	g II
						-0.1995	0.0023	-0.2040	-0.1951	*matK*
						0.0031	0.0028	-0.0023	0.0085	insignificant
						-0.0789	0.0018	-0.0824	-0.0755	g II

Phylogenetic structure in different non-coding partitions

Comparing three different data partitions (spacers, group I intron and group II introns) concerning resolution and statistical support obtained in parsimony analyses based on both, substitutions and coded indels, noticeable differences were realized. The application of the sole group I intron resulted in a topology without resolution or significant bootstrap support for the respective nodes of the backbone. This is also true concerning the branching order within Ranunculales, the placement of Sabiales and Proteales and the relationships among Proteales. In contrast MP calculations of all spacers and the group II introns under study yielded fully resolved backbone-topologies, differing in the placement of Trochodendrales or Buxales, respectively, with statistical support being basically highest when analyzing the combined *trnK-petD-rpl16* dataset (compare Table 4). In addition signal from the two datasets agrees on the monophyly of Ranunculales (BS 96 – spacers, BS 98 – group II introns) and the first-branching position of *Euptelea* within the order with statistical support being rather low. However, no clear statement regarding the phylogenetic utility of the different partitions can be made just on the basis of a few selected bootstrap values. Therefore the phylogenetic structure of the four data matrices was measured applying the method published by Müller & al. (2006). Considering the signal per aligned position it became obvious that the *trnL* group I intron provides the highest amount of phylogenetic structure, followed by the combination of three group II introns (Table 8). These findings correlate with the proportion of parsimony informative sites which is largest in the group I intron and smallest within the spacer partition (compare Table 7). Addressing the phylogenetic signal per parsimony informative position in a second analysis resulted in the recognition of a different pattern. Differences between the *trnL* intron and the group II introns became indistinct, while the spacer partition provided the least structure per parsimony informative position.

These findings clearly contradict the hypothesis of amount and quality of phylogenetic structure being highest in spacers, followed by the group I intron, while the group II introns were suggested to represent the other end of the spectrum. It was based on the assumption of mutational dynamics being similar among resembling structural elements. Thus helical elements are considered to be more constrained, displaying lower site-rates. In contrast unpaired segments, such as loops and bulges therefore should have higher site-rates, but also a higher quality of signal due to a lower proportion of parallel and convergent mutations, with alignability being required. However, spacers are clearly

shown to display the least amount of parsimony-informative characters as well as phylogenetic signal per parsimony informative position.

1.6 Conclusions

The utility of fast-evolving and non-coding genomic regions within deep-level phylogenetic reconstructions of angiosperm relationships has been proven in numerous studies during recent years. The advanced investigation on the phylogenetic structure of the different non-coding partitions in comparison to coding genes revealed a significantly higher average phylogenetic signal per informative site within spacers and introns than in the frequently applied *rbcL* gene. The rapidly-evolving well performing *matK* gene is shown to line up within the non-coding partitions in this respect. It is furthermore proven again that microstructural changes that frequently occur in less constrained non-coding DNA provide useful information within phylogenetic reconstructions. However, this study clearly demonstrates the opportunities and coevally the limitations of applying rapidly-evolving DNA. The analyses of an extended data matrix including complementary sequence data of two more group II introns as well as two spacers in comparison to the study of Worberg & al. (2007) resulted in almost the same well resolved and highly supported topology of the early-diverging eudicot grade. Nevertheless, difficult-to-resolve positions such as the exact placement of *Euptelea* within Ranunculales or the respective position of Sabiales could not be cleared up with confidence, albeit the number of parsimony informative characters was doubled in the current study. Thus these findings seem to corroborate the fact that the continuing addition of molecular markers to analyses may not be the most efficient solution in gaining robust hypothesis on phylogenetic relationships. Markers, mostly defined in practical terms as being a genomic region that can be easily amplified and sequenced, often representing compositions of different kinds of partitions, could be selected due to their phylogenetic structure and performance at a certain taxonomic level. However, as it was demonstrated for the *rpl16* region within the recent study, using a molecular marker can lead to compromising on high performance and high laboratory effort due to a raised number of sequencing primer reads needed.

By comparing a number of various spacers and introns concerning their molecular evolution it became obvious that it follows certain patterns in angiosperms, indicated by the occurrence of mutational hotspots, which are connected to structural and functional

constraints. This is especially demonstrated for the three group II introns under study where highly dynamic sequence parts were rather found in loops than stems. These mutational hotspots are usually well defined and can thus easily been excluded from tree inference. However, further work is needed to improve understanding of mechanisms underlying molecular evolution of genomic regions, being the basis for applying genomic regions to phylogenetics in a useful way.

1.7 Appendices

Appendix A

List of Primers used in this study

Primers used for the amplification of *petB-D* region along with their sequences and the taxa for which they were designed. References are given for primers that were not designed for this study.

Primer name	Sequence	Taxa	Reference
PIpetB1411F	GCC GTM TTT ATG TTA ATG C	angiosperms	Löhne & Borsch (2005)
PIpetD738R	AAT TTA GCY CTT AAT ACA GG	angiosperms	Löhne & Borsch (2005)

Primers used for the amplification of *trnK/matK* along with their sequences and the taxa for which they were designed. References are given for primers that were not designed for this study.

Primer name	Sequence	Taxa	Reference
trnKFbryo	GGG TTG CTA ACT CAA TGG TAG AG	land plants	Wicke & Quandt (in press)
psbARbryo	CGC TTT CGC GTC TTT CTA AAG	land plants	Wicke & Quandt (in press)
MG15	ATC TGG GTT GCT AAC TCA ATG	angiosperms	Liang & Hilu (1996)
MG1	AAC TAG TCG GAT GGA GTA GAT	angiosperms	Liang & Hilu (1996)
trnK2R	AAC TAG TCG GAT GGA GTA G	angiosperms	Johnson & Soltis (1995)
trnK-3R-angio1	CTC CCC AAG CCG TGC YTG C	angiosperms	Worberg & al. (2007)
psbA-R	CGC GTC TCT CTA AAA TTG CAG TCA T	angiosperms	Steele & Vilgalys (1994)
BEmatk3392F	CG(GC) ATT TGG TAT TTA GAT A	angiosperms	this study
EDtrnKF510angio1	GTA TCG CAC TAT GTA TCA T	eudicots	Worberg & al (2007)
EDtrnK600F	GTA GAA GAA RCA GTA TAT TG	eudicots	Worberg & al (2007)
EDtrnK882F	TTT GAC TGT ATC GCA CTA TGT ATC	eudicots	Worberg & al (2007)
trnK-3F-angio	GCA AGC ACG GTT TGG GGA G	eudicots	this study
LAUmatK1840R	AGT GAA CTG GAT TTA TTG TCA	Lauraceae	this study

CHLmatK1653R	CTG GAT TTA TTG TCA TAG CC	Chloranthaceae	this study
RANmatK641F	TTC YAA AGT CAA AAG AGC G	Ranunculales	this study
RAmatK2100R	TGA AAA TCA TTA ACA AAA ACT AC	Ranunculaceae	Worberg & al. (2007)
XANmatK1490F	TTC TTT CTC TAC GAG TAT CAT	Ranuculaceae	this study
BEtrnK1509F	GAC TGT ATC GCA CTA TGT A	Berberidaceae	this study
RANmatK2387R	AGG TCA TTG ATA CRA ATA ATA	Sabiaceae	this study
COCtrnKIF	TGG AGA TGA ATG TGT AGA AGA AAC	Menispermaceae	Worberg & al. (2007)
SABmatK2661F	GCT GCG ATT AGT ATC TTC C	Sabiaceae	this study
SABtrnK252F	CAC ATT TGG ATG AAG CAA C	Sabiaceae	this study
EUPTmatK1006F	GGC TAT CTT TCA AGT GTA CG	Sabiaceae	Worberg & al. (2007)
Le-7F	GGG TTG CTA ACT CAA CGG TAG	eudicots	Müller & al. (2004)
DIDYtrnK1316F	ACG AAT GTG TAG AAG AAG C	Didymelaceae	this study
MYRmatK3749F	CTT TGG CTC GTA AAC ATA AG	Myrothamnaceae	this study
LEmatK3391F	TTA TCC AAG CAT TCC CTC G	Leeaceae	this study
EDtrnK630F	GTA GGA GAA GCA GTA TAT TG	Leeaceae	Worberg & al. (2007)
SANtrnK1437F	TTC TAA TCA TCT TGT TAT CGC A	Santalaceae	this study

Primers used for the amplification of *trnL-F* region along with their sequences and the taxa for which they were designed. References are given for primers that were not designed for this study.

Primer name	Sequence	Taxa	Reference
trnTc	CGA AAT CGG TAG ACG CTA CG	angiosperms	Taberlet & al. (1991)
trnTf	ATT TGA ACT GGT GAC ACG AG	angiosperms	Taberlet & al. (1991)

Primers used for the amplification of *atpB-rbcL* region along with their sequences and the taxa for which they were designed. References are given for primers that were not designed for this study.

Primer name	Sequence	Taxa	Reference
atpB-rbcL-F1	CAC TCA TRC TAC RCT CTA ACT C	angiosperms	this study
atpB-rbcL-R	CAC CAG CTT TGA ATC CAA CAC C	angiosperms	this study
atpB-rbcL1869R	ATT GAA TRA GTA AAC RAT GGA	angiosperms	this study
atpB-rbcL379F	TGT CCG ATA GCA AGT TGA T	Austrobaileyaceae	this study
GREVatpB-rbcL1700F	ATA GCA AGT TGA TCG GTT	Proteaceae	this study
CA05ar1696F	AAT AAA TGT CCG ATA GCG G	Cactaceae	this study

Primers used fort he amplification of *rpl16* region along with their sequences and the taxa for which they were designed. References are given for primers that were not designed for this study.

Primer name	Sequence	Taxa	Reference
rps3Fa	CAA ATT GCR GGR CGT ATC G	angiosperms	this study
l16exon2	TCT TCC TCT ATG TTG TTT ACG	angiosperms	Downie & al. (2000)
rpl16_1900R	CGT TCC GCC ATC CCA CC	angiosperms	this study
rpl16_690F	GCT CAT TGC TTC GTA TTA TC	angiosperms	this study
rpl16_510F	TTA GTG TGT GAC TCG TTG GTT T	eudicots	this study
rpl16_688F	CCA ACT CAT CAC TTC GCA TT	core eudicots	this study
UMBrpl16_2382F	ATT TCT TCT GAT AGG TCA T	Lauraceae	this study
rpl16_1670R	CTY TCA YCC TTC CAT TTA TCC	Aristolochiaceae	this study
NYrpl16_1416R	TTG AGA ATA CGA AGC AAT GAG	Nymphaeaceae	this study
EUrpl16_1706F	TGA GAG AAA GAG AGA AGG A	Eupteleaceae	this study
DIDYrpl16_2228F	GGG TAG TGT AAT AAA GCA TCA	Didymelaceae	this study
SAXrpl16_1412R	AAT GCG AAG CAA TGA GTT GG	Saxifragaceae	this study
SAXrpl16_2380F	ATC TGT TCA TAG AGC AAA A	Saxifragaceae	this study
DILrpl16_2445F	GCG GAC TAA TCT GTA ATA A	Dilleniaceae	this study
CORrpl16_F2	AGA GAA GGT AGR GTT CCY	Coriariaceae	this study
IMPrpl16_1667R	CAC CCG TCC ATT TAT CCA CA	Balsaminaceae	this study

Taxon	*trnK intron*	*matK* gene	*trnL* intron	*trnL-F* spacer	*petB-D* spacer	*petD* intron	*atpB-rbcL spacer*	*rps3- rpl16 spacer*	*rpl16 intron*	*Position H1 trnK*	*Position H2 trnK*	*Position H3 trnK*	*Position H4 trnK*	*Position H5 trnK*
Amborella trichopoda	1088	1509	474	375	224	733	792	131	1110	-	-	-	416-434	449-462
Nymphaea odorata ssp tuberosa	1034	1530	520	380	204	639	794	156	801	-	-	-	394-404	419-432
Austrobaileya scandens	1002	1524	475	390	176	710	777	136	960	-	-	-	380-392	407-420
Ceratophyllum demersum	867	1548	528	442	190	694	740	186	1022	-	-	-	369-379	394-400
Acorus calamus	964	1536	520	377	190	726	813	160	939	-	-	-	386-393	408-414
Chloranthus brachystachys	894	1524	493	351	195	715	786	154	1014	-	-	-	292-306	321-334
Aristolochia pistolochia	1102	1530	510	372	200	699	769	167	1057	-	-	301-382	475-488	503-511
Magnolia officinalis	966	1524	490	356	198	701	758	136	957	-	236-236	-	369-383	398-415
Umbellularia californica	980	1524	482	363	197	716	744	151	970	-	-	-	375-388	403-409
Hedycarya arborea	975	1524	481	388	198	706	750	148	984	-	-	-	374-389	404-410
Chimonanthus praecox	942	1518	477	328	198	698	760	151	940	-	-	-	360-371	386-392
Euptelea pleiosperma	895	1524	500	380	197	702	720	146	1005	-	-	-	305-315	330-344
Akebia quinata	951	1521	503	371	213	709	765	193	1098	204-206	-	-	384-390	405-412
Dicentra eximia	963	1524	474	359	213	709	736	185	1003	-	-	-	375-382	397-418
Papaver triniaefolium	929	1525	519	363	191	718	734	152	1003	-	-	-	369-376	391-410
Cocculus laurifolius	880	1530	490	386	220	702	809	125	1071	-	-	-	316-324	339-346
Stephania delavaji	999	1546	501	378	223	704	832	157	1076	-	-	-	374-382	397-404
Xanthorhiza simplicissima	962	1527	501	345	193	728	790	160	958	182-184	-	-	386-393	417-422
Mahonia japonica	946	1531	479	746	197	690	852	155	824	174-176	-	-	363-371	393-400
Podophyllum peltatum	939	1539	466	387	219	737	757	174	1024	198-200	236-236	-	382-392	407-423
Sabia japonica	946	1536	503	367	189	706	756	84	1019	-	236-236	-	370-377	394-400
Sabia swinhoei	870	1536	480	374	190	705	738	163	992	-	160-160	-	294-301	318-324
Meliosma cuneifolia	922	1524	513	380	193	714	764	169	1033	-	-	-	346-354	369-382
Nelumbo nucif ssp nucif	968	1525	524	402	194	719	744	169	1024	-	-	-	367-375	390-406
Nelumbo nucif ssp lutea	954	1524	525	401	192	718	744	158	1016	-	-	-	354-362	377-388
Embothrium coccineum	922	1530	492	366	191	734	794	172	985	162-166	-	-	346-355	370-376
Grevillea banksii	977	1530	494	421	193	733	805	173	1011	198-202	-	-	380-389	404-409
Platanus orientalis	948	1539	500	366	200	708	754	180	1035	-	-	-	364-372	387-393
Platanus occidentalis	956	1539	523	366	200	709	755	181	1033	-	-	-	372-380	395-401
Tetracentron sinense	961	1516	442	397	200	704	758	142	867	-	-	-	369-378	393-400
Trochodendron aralioides	947	1516	439	369	204	709	766	142	969	-	-	-	353-362	377-383
Didymeles integrifolia	1150	1524	529	353	217	743	752	173	1033	-	233-393	-	532-543	558-571
Buxus sempervirens	938	1524	505	378	193	726	761	146	1025	-	-	-	360-370	385-398
Pachysandra terminalis	956	1524	507	370	193	704	753	160	995	-	-	-	383-393	408-421
Gunnera tinctoria	969	1536	511	359	196	721	741	137	1022	-	-	-	368-375	390-408
Myrothamnus flabellifolia	983	1530	498	349	202	725	772	143	1023	-	-	-	367-375	390-409
Myrothamnus moschata	976	1530	492	353	193	731	768	143	1013	-	-	-	373-381	396-421
Cercidiphyllum japonicum	965	1515	507	356	198	716	745	145	996	-	-	-	374-383	398-411
Chrysosplenium alternifolium	993	1530	464	186	193	696	760	171	962	-	-	-	372-380	395-411
Vitis riparia	982	1509	517	323	188	734	777	165	1040	-	241-241	-	381-388	403-422
Leea coccinea	997	1506	505	377	189	733	775	165	1063	-	238-238	-	394-401	416-429
Dillenia philippinensis	993	1527	495	412	191	799	755	186	1122	-	-	-	368-376	391-404
Aextoxicon punctatum	963	1509	509	355	193	716	761	151	974	195-199	-	-	377-385	400-413
Osyris alba	972	1520	528	376	193	726	687	164	1045	-	-	-	384-392	407-418
Rhipsalis paradoxa	962	1530	643	365	212	789	740	147	1109	-	-	-	361-367	382-395
Spinacia oleracea	978	1518	303	336	212	743	785	155	954	-	243-243	-	376-385	400-413
Erodium cicutarium	1091	1495	496	369	220	734	788	159	-	-	-	-	399-405	433-447
Coriaria myrtifolia	971	1521	570	377	191	743	792	166	1088	-	226-226	-	360-367	382-395

Arabidopsis thaliana	1044	1515	311	343	188	709	801	173	1056	192-195	-	-	392-399	414-430
Oenothera elata	931	1539	519	376	198	755	674	158	1104	167-173	-	-	349-360	375-388
Larrea tridentata	1057	1513	526	287	226	724	744	141	960	199-200	-	-	395-403	418-439
Stachyurus chinensis	943	1518	512	365	204	754	777	182	1082	-	-	-	330-341	356-374
Impatiens nolitangere	1025	1509	493	361	184	785	748	146	1019	187-190	-	-	380-381	396-409
Ilex aquifolium	1043	1515	491	361	194	720	753	163	916	-	-	-	397-405	420-432
Atropa belladonna	989	1530	496	362	190	742	814	146	1019	-	-	-	375-383	398-410
Nicotiana tabacum	996	1530	502	356	190	742	817	146	1020	-	-	-	379-387	402-414
Panax ginseng	1012	1512	506	361	174	751	780	155	944	199-201	-	-	396-403	418-430

Appendix B. Actual length of the genomic regions used in this study and the positions of mutational hotspots in the respective sequences. Sheet 1.

Taxon	*Position H6 trnK*	*Position H7 trnK*	*Position H8 trnK*	*Position H9 trnK*	*Position H10 trnK*	*Position H1 matK*	*Position H1 trnL*	*Position H2 trnL*	*Position H3 trnL*	*Position H1 trnL-F*	*Position H2 trnL-F*	*Position H3 trnL-F*	*Position H1 petB-D*	*Position H1 petD*
Amborella trichopoda	515-522	622-625	662-672	802-892	1049-1067	580-585	130-132	232-240	284-295	1-52	270-276	299-302	110-124	233-263
Nymphaea odorata ssp tuberosa	485-496	691-694	637-642	797-841	993-1013	-	139-145	240-244	281-336	1-63	269-276	311-318	101-109	223-262
Austrobaileya scandens	473-479	569-572	615-620	758-817	964-981	-	132-140	233-241	278-291	1-73	279-286	326-333	62-70	223-246
Ceratophyllum demersum	448-457	549-554	597-606	-	828-846	-	133-141	242-256	306-351	1-101	303-313	367-374	90-98	238-268
Acorus calamus	462-469	559-562	611-620	748-785	926-944	-	137-143	242-272	317-331	1-48	260-269	308-315	86-94	233-257
Chloranthus brachystachys	482-389	479-482	525-539	661-703	855-873	-	135-139	249-257	294-311	1-40	241-248	287-294	87-95	231-254
Aristolochia pistolochia	563-570	672-674	701-734	876-913	1060-1077	-	141-148	260-278	320-345	1-56	257-264	303-310	86-101	226-244
Magnolia officinalis	463-470	560-563	606-616	729-779	925-945	-	132-138	240-258	295-313	1-51	245-252	292-299	86-94	227-250
Umbellularia californica	469-476	566-567	608-618	740-782	930-959	-	132-143	241-254	291-310	1-47	252-259	299-306	86-94	242-260
Hedycarya arborea	458-465	561-564	607-617	737-778	930-954	-	132-140	238-251	288-306	1-55	277-284	324-331	86-94	236-254
Chimonanthus praecox	440-450	540-543	585-591	715-760	899-921	-	133-139	236-254	291-304	1-42	211-218	260-267	86-94	226-244
Euptelea pleiosperma	392-399	489-492	535-544	662-709	856-874	-	136-142	244-262	299-317	1-52	247-254	299-306	86-94	226-250
Akebia quinata	460-467	553-556	599-609	734-771	913-930	-	133-143	256-274	311-326	1-52	253-260	300-308	92-100	224-248
Dicentra eximia	466-473	563-566	612-622	745-777	924-942	-	136-142	237-249	288-307	1-44	231-238	283-290	91-113	222-251
Papaver triniaefolium	458-466	533-536	579-589	707-741	892-908	-	140-146	261-280	315-336	1-62	245-253	293-300	86-92	239-263
Cocculus laurifolius	394-397	494-496	538-548	666-691	863-877	586-588	135-141	243-265	301-313	1-61	248-253	298-305	99-107	226-250
Stephania delavaji	451-458	565-567	609-620	748-773	950-975	-	140-146	248-270	306-318	1-63	256-263	308-315	102-110	226-250
Xanthorhiza simplicissima	466-473	565-568	606-615	746-768	921-941	586-589	136-145	242-260	293-318	1-20	217-224	263-272	86-94	226-258
Mahonia japonica	448-455	546-549	592-604	727-758	905-925	589-591	115-126	222-240	275-301	1-52	297-324	474-501	90-98	231-256
Podophyllum peltatum	469-476	567-576	619-624	740-758	901-918	598-600	111-111	200-223	263-290	1-52	272-279	312-319	93-105	231-262
Sabia japonica	443-450	543-546	596-606	722-760	907-925	-	132-139	248-266	308-326	1-52	250-257	297-304	83-91	223-247
Sabia swinhoei	367-374	568-571	521-531	647-685	832-849	-	108-115	224-242	284-302	1-59	257-264	304-311	83-91	223-247
Meliosma cuneifolia	425-432	523-526	576-581	696-737	885-901	-	145-154	263-281	323-336	1-53	251-258	303-309	79-87	226-250
Nelumbo nucif ssp nucif	454-468	567-570	620-630	748-783	931-947	-	136-142	256-279	321-346	1-62	285-292	332-339	87-95	231-255
Nelumbo nucif ssp lutea	436-450	550-553	603-613	734-769	917-933	-	136-142	256-276	318-347	1-62	285-292	332-339	86-94	231-259
Embothrium coccineum	424-431	520-522	572-582	700-737	884-901	-	136-138	241-259	306-323	1-39	235-242	296-303	84-92	244-268
Grevillea banksii	457-464	565-567	617-627	754-793	940-956	-	141-143	243-261	308-325	1-48	304-312	352-359	86-94	242-267
Platanus orientalis	441-448	538-541	591-601	726-763	909-927	-	121-126	235-254	296-323	1-39	235-242	296-303	93-101	226-250
Platanus occidentalis	449-456	546-549	599-609	734-771	917-935	-	136-143	257-276	318-346	1-39	235-242	296-303	93-101	226-250
Tetracentron sinense	455-462	552-555	604-613	733-778	922-940	-	136-145	246-246	249-265	1-52	265-272	311-318	86-103	221-240
Trochodendron aralioides	438-445	535-538	587-596	722-762	909-926	-	135-143	244-244	247-262	1-57	269-276	315-322	93-105	226-250
Didymeles integrifolia	619-626	716-719	768-778	927-974	1113-1129	-	136-142	255-277	319-395	1-41	237-244	283-290	103-118	226-251
Buxus sempervirens	449-456	546-549	604-613	729-763	902-917	-	136-142	251-269	311-328	1-55	256-263	308-315	86-94	226-250
Pachysandra terminalis	469-476	566-569	619-629	747-789	919-935	-	136-142	251-269	312-330	1-51	248-255	300-307	86-94	226-250
Gunnera tinctoria	456-463	558-561	611-621	739-780	931-948	586-594	136-142	251-269	311-330	1-49	249-256	290-296	84-92	231-257
Myrothamnus flabellifolia	457-463	553-556	606-616	729-778	929-962	-	136-142	244-262	304-327	1-50	231-245	280-286	93-103	226-250
Myrothamnus moschata	469-475	565-567	617-622	739-782	933-955	-	136-142	244-256	298-321	1-50	231-246	284-290	86-94	226-250
Cercidiphyllum japonicum	459-471	561-564	614-624	747-777	928-944	-	138-144	253-271	313-330	1-52	239-246	286-292	86-94	208-232
Chrysosplenium alternifolium	459-469	561-564	614-624	750-805	957-972	-	139-147	242-242	270-287	1-44	84-87	127-130	86-94	215-239
Vitis riparia	470-477	567-570	620-630	754-776	942-961	-	136-142	263-275	317-335	1-50	191-199	244-255	85-93	224-253
Leea coccinea	477-484	590-593	643-653	777-805	956-976	-	136-142	251-263	305-323	1-51	260-264	304-310	86-94	226-255
Dillenia philippinensis	452-458	564-567	617-628	756-796	957-972	-	132-140	243-261	306-317	1-51	278-285	331-337	84-92	234-267
Aextoxicon punctatum	461-468	558-561	611-621	739-775	926-942	-	143-151	260-278	320-337	1-51	239-246	286-292	86-94	230-254
Osyris alba	466-473	558-559	607-618	733-769	917-951	-	151-157	259-276	321-339	1-60	262-269	309-311	88-96	225-249
Rhipsalis paradoxa	442-447	537-542	585-595	724-764	927-941	-	150-158	273-273	323-473	1-41	250-259	299-304	94-102	250-285
Spinacia oleracea	460-467	561-567	611-620	755-795	938-957	-	119-127	167-167	167-167	1-42	205-209	251-256	99-107	240-258

Erodium cicutarium	495-504	597-601	651-661	792-855	1002-1070	-	136-146	255-273	334-357	1-47	247-261	300-305	107-115	239-261
Coriaria myrtifolia	443-451	541-549	599-614	746-778	932-950	-	132-137	246-264	306-378	1-50	259-266	306-314	86-92	225-256
Arabidopsis thaliana	477-487	-	627-629	776-815	997-1023	-	130-130	195-195	195-195	1-82	225-232	257-279	82-89	225-278
Oenothera elata	436-442	532-539	589-599	685-719	887-910	-	139-147	258-271	332-348	1-58	259-278	318-324	85-93	241-265
Larrea tridentata	484-490	584-587	638-648	820-848	1016-1032	-	146-151	261-279	315-354	1-51	151-157	205-225	105-112	234-276
Stachyurus chinensis	422-430	525-528	578-588	713-753	914-922	-	136-142	251-269	311-335	1-58	232-247	287-297	91-99	224-263
Impatiens nolitangere	457-471	572-575	618-628	734-822	985-1004	-	145-151	255-267	300-316	1-49	254-260	302-308	76-84	234-277
Ilex aquifolium	480-495	585-593	643-652	775-843	1008-1022	-	125-129	238-256	292-314	1-46	243-250	292-298	92-100	222-253
Atropa belladonna	453-460	554-566	616-640	745-768	924-968	-	129-134	249-267	302-319	1-46	246-253	295-301	93-101	223-252
Nicotiana tabacum	457-464	558-570	620-625	745-772	930-975	-	129-134	249-261	308-325	1-46	241-248	289-295	93-101	223-252
Panax ginseng	478-479	574-576	626-635	759-817	977-991	-	139-140	249-267	309-326	1-47	240-249	291-297	81-89	241-272

Appendix B. Actual length of the genomic regions used in this study and the positions of mutational hotspots in the respective sequences. Sheet 2.

Taxon	*Position H2 petD*	*Position H3 petD*	*Position H1 atpB-rbcL*	*Position H2 atpB-rbcL*	*Position H3 atpB-rbcL*	*Position H4 atpB-rbcL*	*Position H5 atpB-rbcL*	*Position H6 atpB-rbcL*	*Position H7 atpB-rbcL*	*Position H8 atpB-rbcL*	*Position H9 atpB-rbcL*	*Position H1 rps3-rpl16*	*Position H2 rps3-rpl16*	*Position H3 rps3-rpl16*
Amborella trichopoda	389-419	623-630	3-7	-	41-42	182-195	273-276	383-397	495-524	559-564	-	27-28	33-34	54-56
Nymphaea odorata ssp tuberosa	386-414	528-535	3-6	25-28	49-50	189-202	277-296	398-402	497-538	565-570	-	50-50	55-57	77-80
Austrobaileya scandens	375-400	598-605	7-17	36-39	60-61	195-205	270-297	402-415	516-524	551-556	-	31-31	36-39	59-66
Ceratophyllum demersum	393-421	578-590	7-9	28-34	55-60	189-209	262-266	368-382	475-483	517-527	-	50-50	55-59	80-83
Acorus calamus	385-410	611-618	-	25-31	51-56	193-212	287-292	430-434	540-550	577-582	-	40-40	45-47	71-74
Chloranthus brachystachys	386-406	604-611	7-9	28-31	52-57	202-210	273-276	398-407	509-528	555-560	-	40-40	45-47	67-70
Aristolochia pistolochia	368-398	589-596	12-16	34-37	65-69	220-233	302-305	417-425	483-500	525-530	-	40-40	49-51	80-83
Magnolia officinalis	373-398	590-597	7-12	31-34	55-60	194-202	265-268	370-379	480-489	517-522	-	35-35	40-42	58-60
Umbellularia californica	384-409	601-608	7-10	29-32	53-58	196-204	276-279	381-403	499-509	531-536	-	40-40	45-46	62-64
Hedycarya arborea	378-403	595-602	7-12	31-34	55-60	203-211	279-282	386-395	496-514	541-546	-	40-40	45-47	63-65
Chimonanthus praecox	368-393	588-595	7-14	33-36	57-61	203-218	285-288	390-395	491-502	529-534	-	40-40	45-47	63-65
Euptelea pleiosperma	375-400	592-599	12-19	38-41	57-62	196-204	267-270	372-376	477-498	520-525	-	40-40	45-47	67-70
Akebia quinata	373-394	598-605	12-17	36-40	56-58	202-210	273-275	383-390	500-518	540-546	-	36-36	41-43	63-66
Dicentra eximia	376-401	598-605	7-16	35-38	47-54	189-199	266-269	371-378	479-489	511-516	-	41-41	46-48	68-70
Papaver triniaefolium	388-413	608-615	7-12	31-34	56-58	200-202	265-268	368-371	472-482	504-509	-	-	44-46	64-67
Cocculus laurifolius	376-401	592-599	7-12	31-36	52-54	182-195	253-257	427-433	540-553	581-590	-	-	-	49-51
Stephania delavaji	376-401	593-600	7-12	31-36	52-60	195-224	286-290	458-464	565-571	598-615	-	32-32	-	50-53
Xanthorhiza simplicissima	383-408	612-625	7-12	31-35	50-58	-	241-244	346-401	524-558	564-570	-	44-44	49-51	71-75
Mahonia japonica	385-414	579-586	7-12	21-24	-	172-187	250-253	352-359	464-488	510-523	-	40-40	45-47	67-70
Podophyllum peltatum	388-413	627-634	4-9	18-21	35-39	174-182	246-249	352-359	465-398	511-550	-	54-54	59-61	81-85
Sabia japonica	372-397	595-602	7-12	31-34	55-63	203-211	274-277	390-398	499-511	533-538	-	-	-	-
Sabia swinhoei	372-397	594-601	7-12	31-34	55-61	201-209	272-275	388-396	497-509	531-536	-	39-39	44-46	66-69
Meliosma cuneifolia	375-400	598-610	7-12	31-34	55-60	188-196	282-285	392-395	500-518	540-545	-	39-39	51-53	73-76
Nelumbo nucif ssp nucif	375-400	608-615	6-12	31-34	55-60	194-210	286-289	387-391	492-502	524-529	-	40-40	45-47	67-70
Nelumbo nucif ssp lutea	379-404	607-614	6-12	31-34	55-60	194-210	286-289	387-391	492-502	524-529	-	40-40	45-47	67-70
Embothrium coccineum	393-418	623-630	7-12	31-34	55-60	211-218	312-315	423-427	545-556	577-582	-	51-51	56-58	78-79
Grevillea banksii	392-417	622-629	4-9	28-31	52-57	213-220	314-317	427-431	556-567	588-593	-	51-51	56-58	78-79
Platanus orientalis	375-400	598-605	7-12	31-34	55-72	206-218	289-292	406-410	500-506	522-527	-	56-56	61-63	83-86
Platanus occidentalis	375-400	598-605	7-12	31-34	55-73	207-219	290-293	407-411	501-507	523-528	-	56-56	61-63	83-86
Tetracentron sinense	365-394	593-600	7-12	31-34	55-60	208-216	279-282	389-397	498-511	533-538	-	31-31	36-38	58-61
Trochodendron aralioides	375-400	598-605	7-12	31-34	55-60	208-216	277-280	387-395	496-509	531-536	-	31-31	36-38	58-61
Didymeles integrifolia	376-400	602-640	7-12	-	51-55	188-194	254-257	358-364	475-509	515-520	-	40-40	45-47	67-70
Buxus sempervirens	379-411	615-622	7-12	31-34	55-66	200-206	272-275	370-383	504-516	538-543	-	40-40	45-47	67-70
Pachysandra terminalis	375-400	594-601	7-12	31-34	55-60	200-206	266-269	370-391	496-518	524-535	-	40-40	45-47	68-71
Gunnera tinctoria	381-406	610-617	4-9	28-31	52-57	197-205	263-266	367-378	479-490	512-517	-	40-40	45-47	67-67
Myrothamnus flabellifolia	374-399	598-621	11-19	38-41	62-67	207-215	278-281	388-414	515-532	554-559	-	40-40	45-47	67-67
Myrothamnus moschata	374-406	605-628	11-24	43-46	67-72	212-220	283-286	387-403	504-521	543-548	-	40-40	45-47	67-67
Cercidiphyllum japonicum	356-381	590-613	4-6	25-28	49-55	189-198	262-265	380-393	494-506	528-533	-	40-40	45-47	67-70
Chrysosplenium alternifolium	359-378	585-592	7-12	31-35	56-61	206-214	281-284	401-404	525-530	536-541	-	54-54	59-61	81-92
Vitis riparia	381-409	615-630	7-15	33-36	57-58	190-198	250-253	371-390	509-526	548-553	-	49-49	54-56	76-79
Leea coccinea	383-408	615-630	12-17	42-45	72-73	213-221	270-273	391-399	506-524	546-551	-	40-40	45-47	67-70
Dillenia philippinensis	391-416	617-685	11-17	36-39	60-60	192-200	270-273	371-382	493-502	520-525	-	46-46	51-53	73-77
Aextoxicon punctatum	378-403	605-612	7-12	31-34	55-59	193-200	263-266	378-399	500-515	537-542	-	40-40	45-47	67-70
Osyris alba	384-409	615-622	-	22-24	45-50	184-192	256-259	331-335	448-457	479-483	-	45-45	-	69-71
Rhipsalis paradoxa	410-466	663-676	7-12	31-34	61-71	247-261	-	-	-	-	330-720	41-41	46-48	-

Spinacia oleracea	382-438	629-634	6-8	27-30	57-67	222-230	293-296	414-418	519-531	553-564	-	51-51	56-58	78-82
Erodium cicutarium	386-410	621-631	7-13	32-35	56-57	215-223	286-289	393-402	508-520	542-547	-	57-57	62-63	79-80
Coriaria myrtifolia	380-420	631-640	4-6	25-32	54-59	228-236	295-298	410-425	531-545	567-572	-	59-59	64-66	86-89
Arabidopsis thaliana	401-427	596-605	4-6	23-27	48-53	201-216	278-281	393-431	541-568	574-579	-	40-43	48-49	94-98
Oenothera elata	389-412	621-651	7-24	48-51	72-78	215-223	292-295	-	-	463-468	-	41-41	46-47	72-76
Larrea tridentata	400-419	614-621	4-9	28-38	59-66	127-135	-	366-372	496-507	521-527	-	49-49	54-55	70-75
Stachyurus chinensis	387-418	640-651	7-12	31-34	55-69	224-232	295-298	398-415	516-532	554-559	-	73-73	78-80	99-106
Impatiens nolitangere	405-433	653-671	4-13	32-35	56-61	187-195	258-282	376-389	497-507	513-519	-	44-44	49-51	-
Ilex aquifolium	379-406	605-617	4-13	32-35	56-61	185-192	255-258	368-401	503-518	524-529	-	45-45	50-52	72-75
Atropa belladonna	376-396	623-630	7-9	31-35	56-61	198-211	274-277	417-446	563-578	584-589	-	38-38	43-45	65-68
Nicotiana tabacum	376-396	623-630	7-9	31-35	56-61	198-211	274-277	417-446	563-581	587-592	-	38-38	43-45	65-68
Panax ginseng	395-433	640-647	7-16	35-40	66-71	206-214	271-274	394-415	527-538	544-549	-	40-40	45-47	66-67

Appendix B. Actual length of the genomic regions used in this study and the positions of mutational hotspots in the respective sequences. Sheet 3.

Taxon	Position H4 rps3-rpl16	Position H5 rps3-rpl16	Position H1 rpl16 intron	Position H2 rpl16 intron	Position H3 rpl16 intron	Position H4 rpl16 intron	Position H5 rpl16 intron	Position H6 rpl16 intron	Position H7 rpl16 intron	Position H8 rpl16 intron	Position H9-rpl16 intron	Position H10 rpl16 intron	Position H11 rpl16 intron	Position H12 rpl16 intron	Position H13 rpl16 intron
Amborella trichopoda	63-70	-	14-29	44-64	-	188-198	224-234	321-326	411-420	467-470	698-698	-	732-739	782-841	1012-1016
Nymphaea odorata ssp tuberosa	87-94	-	-	21-26	146-148	153-166	194-204	297-299	373-380	427-429	-	-	-	-	689-708
Austrobaileya scandens	73-80	-	14-23	38-46	163-168	178-195	216-226	319-321	380-395	442-444	672-672	-	700-719	727-750	851-863
Ceratophyllum demersum	90-97	-	14-29	44-55	-	199-212	232-242	337-343	406-421	468-470	704-704	-	732-762	774-831	930-940
Acorus calamus	87-93	-	14-23	41-45	175-180	185-198	219-224	317-319	354-369	-	641-641	-	667-702	710-738	833-843
Chloranthus brachystachys	77-84	-	14-23	38-44	183-183	188-201	222-233	330-332	391-406	453-455	684-684	-	721-735	754-809	915-926
Aristolochia pistolochia	90-95	-	14-24	39-54	180-185	192-205	226-235	328-330	394-409	456-458	700-700	-	728-736	755-839	961-969
Magnolia officinalis	67-67	-	14-23	38-43	169-170	175-188	210-219	315-317	376-391	438-441	671-671	-	699-707	721-764	865-875
Umbellularia californica	81-82	-	14-24	39-43	170-173	178-187	208-217	315-317	376-391	438-440	669-669	-	697-705	719-762	866-876
Hedycarya arborea	72-86	-	14-24	39-47	174-179	184-193	214-229	327-329	388-398	445-447	691-691	-	719-727	741-779	880-890
Chimonanthus praecox	72-83	-	14-24	39-47	173-174	179-188	189-198	291-294	353-373	420-422	651-651	-	679-687	701-739	842-852
Euptelea pleiosperma	77-84	-	14-19	-	170-172	207-220	241-251	344-347	411-431	478-480	709-709	-	737-750	764-802	907-917
Akebia quinata	87-102	-	14-29	44-51	183-185	190-203	229-239	-	423-521	568-570	792-792	-	820-833	847-895	999-1009
Dicentra eximia	77-84	-	14-35	50-69	222-222	227-240	261-282	375-377	423-433	480-485	714-715	-	743-757	771-820	899-915
Papaver triniaefolium	74-81	100-101	14-26	39-46	178-178	195-208	229-243	336-339	411-430	480-482	712-712	-	739-752	766-791	893-903
Cocculus laurifolius	-	-	14-30	-	194-200	205-222	243-253	352-354	437-445	491-493	722-722	-	769-792	806-838	961-977
Stephania delavaji	-	-	14-30	50-62	200-204	209-213	239-249	348-350	440-449	494-496	725-725	-	757-782	796-839	959-982
Xanthorhiza simplicissima	82-89	-	14-25	45-51	183-184	189-202	224-234	327-329	412-421	468-470	695-695	-	723-728	736-757	863-868
Mahonia japonica	77-84	104-104	14-39	54-61	193-194	199-217	245-255	-	419-437	484-486	-	-	-	-	727-736
Podophyllum peltatum	91-103	122-123	14-31	46-49	181-183	200-213	235-242	345-347	414-422	469-471	699-699	-	726-739	753-809	919-932
Sabia japonica	1-9	28-28	14-50	65-72	206-207	212-230	251-261	368-370	441-445	492-494	728-728	-	748-761	775-815	921-931
Sabia swinhoei	76-88	107-107	14-63	78-85	219-220	225-243	264-274	371-373	444-448	495-497	731-731	-	751-764	778-818	924-934
Meliosma cuneifolia	83-95	-	14-43	58-65	205-206	211-224	244-265	391-393	462-466	513-515	744-744	-	779-792	806-831	934-944
Nelumbo nucif ssp nucif	77-89	-	14-51	66-73	205-206	211-221	242-252	345-347	425-439	486-488	720-720	-	748-761	776-819	915-936
Nelumbo nucif ssp lutea	77-89	-	14-44	59-66	198-199	204-214	235-245	338-340	418-434	481-483	715-715	-	743-756	771-814	910-926
Embothrium coccineum	86-98	-	14-27	42-46	178-179	184-197	218-228	325-327	397-405	452-454	685-685	-	713-726	740-770	883-893
Grevillea banksii	86-99	-	14-25	40-44	176-177	182-195	216-226	328-330	400-418	465-467	717-717	-	745-758	772-802	914-924
Platanus orientalis	99-108	-	14-21	36-43	175-176	181-194	215-228	326-328	398-407	454-456	694-694	-	722-745	759-812	926-946
Platanus occidentalis	99-109	-	14-21	36-43	175-176	181-194	215-228	326-328	398-407	454-456	694-694	-	722-745	759-810	924-944
Tetracentron sinense	68-80	-	14-29	44-51	183-184	189-202	223-233	326-328	386-397	444-446	675-675	-	703-716	730-751	763-773
Trochodendron aralioides	68-80	-	14-29	44-51	183-184	189-201	222-232	325-327	385-395	442-444	673-673	-	701-714	728-761	871-881
Didymeles integrifolia	77-89	113-115	14-38	53-77	217-219	224-246	264-274	367-373	431-443	490-492	721-721	-	749-770	784-818	934-944
Buxus sempervirens	77-90	-	14-48	63-69	210-211	216-238	256-257	350-353	411-422	469-475	-	-	747-760	774-814	921-931
Pachysandra terminalis	78-91	-	14-48	63-64	205-206	211-234	252-253	346-351	409-420	467-469	698-698	-	735-748	768-805	891-901
Gunnera tinctoria	71-77	-	14-21	36-41	173-174	179-190	211-222	319-321	379-390	437-439	668-677	-	705-723	745-785	901-928
Myrothamnus flabellifolia	71-77	-	14-20	35-35	-	174-188	209-220	313-315	373-384	431-433	663-676	-	704-722	744-788	905-929
Myrothamnus moschata	71-77	-	14-20	35-35	-	172-181	202-213	306-308	366-378	425-427	657-672	-	700-718	740-791	899-923
Cercidiphyllum japonicum	77-89	-	14-22	30-37	169-171	176-193	214-225	323-325	385-395	442-444	679-684	-	717-735	757-790	896-910
Chrysosplenium alternifolium	99-108	-	14-21	27-31	159-165	170-179	200-216	333-335	385-394	439-441	675-680	695-699	708-729	764-791	852-868
Vitis riparia	86-98	-	14-22	37-46	182-183	188-211	233-244	345-347	404-413	-	682-687	702-703	722-739	753-818	955-969
Leea coccinea	77-89	113-114	14-21	36-45	182-183	188-201	222-233	332-334	397-407	454-454	677-682	697-698	717-737	751-789	937-978
Dillenia philippinensis	84-96	-	14-20	35-42	156-158	163-179	206-224	312-319	391-400	447-449	681-686	-	714-737	751-792	904-1035
Aextoxicon punctatum	77-89	-	14-21	36-41	173-175	180-198	219-230	323-325	383-395	442-444	673-678	-	711-729	751-784	890-904
Osyris alba	78-90	-	14-21	36-41	175-176	181-202	223-234	327-328	387-397	444-446	686-691	-	719-742	766-817	939-953
Rhipsalis paradoxa	76-82	-	14-22	37-42	180-182	187-196	215-244	333-336	400-413	460-462	695-699	-	731-761	-	980-1027

Spinacia oleracea	89-98	-	14-22	37-43	183-185	190-195	217-243	339-343	404-406	453-455	697-702	-	745-758	-	846-859
Erodium cicutarium	81-87	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Coriaria myrtifolia	93-98	-	14-20	35-41	184-185	190-200	221-228	328-329	389-399	446-448	694-694	-	723-756	778-884	-
Arabidopsis thaliana	105-113	-	14-17	31-49	-	195-201	222-233	332-333	394-405	452-454	715-722	-	786-800	828-866	-
Oenothera elata	83-92	-	15-25	42-47	164-165	170-187	251-268	363-365	428-435	482-484	714-715	-	743-796	810-920	-
Larrea tridentata	76-84	-	14-37	52-59	202-203	221-223	235-251	359-375	436-444	490-492	736-745	-	772-785	-	-
Stachyurus chinensis	113-126	-	14-21	36-41	177-178	183-201	220-235	328-330	388-394	441-443	672-677	-	705-715	737-817	976-990
Impatiens nolitangere	78-85	-	14-21	36-40	169-170	175-189	211-226	317-319	377-384	431-433	652-653	-	678-698	721-793	913-932
Ilex aquifolium	82-94	-	14-22	37-47	180-181	186-201	222-236	334-336	392-395	442-444	674-679	-	707-731	745-758	810-833
Atropa belladonna	71-77	-	15-21	36-37	-	175-189	210-226	319-321	380-395	442-444	681-682	-	707-720	734-775	880-925
Nicotiana tabacum	71-77	-	15-21	36-37	-	175-189	210-224	317-319	378-393	440-442	679-681	-	706-719	733-777	882-926
Panax ginseng	74-86	-	14-21	36-44	177-178	183-200	221-235	331-335	393-402	449-451	683-688	-	716-739	761-799	-

Appendix B. Actual length of the genomic regions used in this study and the positions of mutational hotspots in the respective sequences. Sheet 4.

Chapter 2

Resolving the backbone of the first diverging eudicot order: the Ranunculales

2.1 Abstract

The Ranunculales have previously been identified as the first diverging eudicot order, which include the woody Eupteleaceae as one of its early diverging lineages. Here, we present a phylogenetic analysis of the order based on molecular data of 50 taxa (including outgroup) and 7 regions from the large single copy region of the plastid genome (*petB-petD* spacer, *petD* group II intron, *trnL* group I intron, *trnL-F* spacer, *trnK* group II intron including *matK*, *trnK-psbA* spacer). Special emphasis is given on the evolution of growth forms within Ranunculales. The combined dataset comprised 7935 positions of aligned sequences plus 1272 binary indel characters. The extensive sampling resulted in fully resolved and highly supported phylogenies using maximum parsimony as well as Bayesian inference. Family relations within the core clade are identical in both approaches with the woody Eupteleaceae appearing as first branching lineage. However, the relationships among the early diverging Ranunculales could not be resolved with confidence. The branching order of Lardizabalaceae as being sister to the residual members of the order, followed by Menispermaceae, Berberidaceae and Ranunculaceae, the latter sharing a sistergroup relationship gained maximum statistical support. Inside the mainly lianescent Lardizabalaceae the shrubby genus *Decaisnea* was clearly depicted as first branching lineage. Additionally a monophyletic group containing the South American genera is shown as being sister to a clade consisting of three genera from East Asia. In Berberidaceae four chromosomal lineages (x = 6, x = 7, x = 8, x = 10) were identified with high confidence (96-100% BS). Five independent lineages were recognized within Ranunculaceae. *Glaucidium* and *Hydrastis* are shown to be early-diverging members of the family. The woody habit seems to have evolved 2 times independently within Ranunculaceae.

2.2 Introduction

The angiosperm order Ranunculales is characterized by an extremely diverse morphology. It comprises predominantly herbaceous groups as well as trees and lianescent/shrubby lineages. Currently, the order contains 7 families: Ranunculaceae, Berberidaceae, Menispermaceae, Lardizabalaceae, Circeasteraceae; Eupteleacae and Papaveraceae.

According to APG II (2003) 3 additional groups are discussed (Kingdoniaceae, Fumariaceae and Pteridophyllaceae). The order is characterized by its large and homogeneous form-Ss sieve-element plastids (Behnke, 1995), benzylisoquinoline alkaloids of the berberine and morphine type (Jensen, 1995) and epicuticular clustered wax tubules, similar to structures found in a number of non-ranunculid groups, including Cercidiphyllaceae, Winteraceae and Nelumbonaceae (Barthlott & Theisen, 1995), some Rosaceae as well as most gymnosperms (Barthlott & al., 2003). Within the Ranunculales shifts from di- to trimerous floral plans (or reverse) are observed (Drinnan & al., 1994; Damerval & Nadot, 2007).

Traditionally, Ranunculales have been placed within Magnoliidae (e.g., Cronquist, 1981; 1988), mostly due to their frequently polymerous flowers (Drinnan & al., 1994). Recent phylogenetic studies, however, identified them with high statistical support as the first branching order among eudicots (Chase & al., 1993; Hoot & al., 1999; Savolainen & al., 2000; Soltis & al., 2000; 2003; Hilu & al., 2003; Worberg & al., 2007), a clade first recognized by Donoghue & Doyle (1989) as well as Doyle & Hotton (1991) based on pollen grain morphology. Ranunculales mostly gain high support as being sister to the remainder of eudicots (Hoot & al., 1999; Soltis & al., 2000; 2003; Hilu & al., 2003; Worberg & al., 2007).

Another result that emerged from the increased use of sequence data for phylogenetic reconstructions and systematics is the inclusion of Eupteleaceae into Ranunculales. Because of its reduced floral morphology this monogeneric family has been placed next to Cercidiphyllaceae in Hamamelidales (Cronquist, 1981; 1988) or Hamamelididae (Takhtajan, 1997). Phylogenetic reconstructions from the 1990s indicated a position among the early diverging Ranunculales (Chase & al., 1993; Drinnan & al., 1994; Hoot & Crane, 1995; Hoot & al., 1999). Molecular studies revealed a core clade from which Eupteleaceae and Papaveraceae s.l. (incl. Fumariaceae, *Hypecoum* and *Pteridophyllum*, Kadereit & al., 1995) are excluded. A study by Hilu & al. (2003) based on partial *matK* data postulated Eupteleaceae as first branching lineage within Ranunculales, followed by Papaveraceae s.l. and the core clade including Lardizabalaceae, Menispermaceae, Berberidaceae and Ranunculaceae. Nevertheless statistical support for this hypothesis was lacking. A 4 gene analysis (Kim & al., 2004) and the combination of intron, spacer and *matK* sequences by Worberg & al. (2007) could increase confidence in the first branching position of Eupteleaceae. Alternative studies assumed a sistergroup relationship between Papaveraceae and the remaining Ranunculales (Hoot & al., 1999; Soltis & al., 2000) or as

a third variant both families as being sister to a core clade (Qiu & al., 2005). Increased knowledge on the placement of the respective families inside the early diverging Ranunculales is relevant for answering the question, whether the order (and eudicots) were ancestrally woody as inferred by Kim & al. (2004), or primitively herbaceous (Cronquist, 1981; 1988).

Inside the core clade of Ranunculales the positions of Lardizabalaceae and Menispermaceae are not fully clarified yet. Different studies show a Lardizabalaceae-Circeasteraceae clade as sister to the remainder of core Ranunculales (Hoot & al., 1999; Soltis & al., 2003; Kim & al., 2004), whereas Lardizabalaceae and Menispermaceae form a weakly supported monophyletic group by using *petD*+*trnL-F* +*matK* in a combined analysis (Worberg & al., 2007).

The sistergroup relationship between Berberidaceae and Ranunculaceae is clearly indicated and well supported in previous studies (e.g., Hoot & al., 1999; Soltis & al., 2000; Hilu & al., 2003), but the position of *Glaucidium* and *Hydrastis* is still controversial. Both genera have been controversial in the past in terms of their systematic positions. Some studies favoured an assignment to the family of Berberidaceae (e.g., Lotsy, 1911; Himmelbaur, 1913; Miyanji, 1930). In contrast several authors considered both to be related to the Ranunculaceae (e.g., Buchheim, 1964; Leppik, 1964; Cronquist, 1988). Some systematic treatments suggested a ditypic subfamily, Glaucidioideae or Hydrastidoideae (Buchheim, 1964; Thorne, 1974; 1976). Others proposed the separation from the Ranunculaceae and the establishment of two monotypic families (Glaucidiaceae – Tamura, 1962; 1972; Tobe, 1981; Hydrastidaceae – Lemesle, 1948; 1955; Tobe & Keating, 1985; Takhtajan, 1997). Using molecular data both genera were identified as closely related to the Ranunculaceae. Phylogenetic hypotheses assume a sistergroup relationship between *Hydrastis*+*Glaucidium* and the remaining Ranunculaceae (Hoot & al., 1999) or *Glaucidium* as first branching within the family (Soltis & al., 2003). Both scenarios gained only weak statistical support. Parsimony analyses based on molecular data of 4 genes carried out by Kim & al. (2004) depicted both species in a tritomy with the remaining Ranunculaceae.

Based on sequence data from 7 fast evolving molecular markers for 43 taxa (plus outgroups) representing the Ranunculales this paper presents the first thorough reconstruction of phylogenetic relationships within the order. Emphasis is given on the evolution of growth forms inside the group.

2.3 Material and methods

Plant material

Plant material was obtained from the Dresden University of Technology Botanical Garden (Germany), the Botanical Garden of the University of Ghent (Belgium) and the Botanic Garden of Talca University (Chile). Additionally samples were taken from collections of A. Stoll and D. Quandt. Vouchers are deposited in DR.

Taxon sampling and molecular markers

In total, 50 taxa from 13 families recognized by APG II (2003) were included in the analyses. Outgroup taxa were chosen to represent the first branching angiosperms, the magnoliids and monocots as well as early diverging eudicot lineages (Sabiales and Proteales). The taxon sampling comprises all major groups of the Ranunculales represented by 43 ingroup species. Among the supposedly early diverging Ranunculales the second species of *Euptelea* (*Euptelea polyandra*) is represented for the first time. Seven taxa of Papaveraceae were included in the analyses. *Pteridophyllum* and *Corydalis* were chosen to complement *Dicentra* (Papaveraceae s.l.) and *Eschscholzia* as well as *Stylophorum* to complete the Papaveraceae s.str. In addition, *Bocconia frutescens* was examined due to its woody habit. Seven genera of the Lardizabalaceae (except *Sargentodoxa*) were enclosed into the taxon sampling as well as *Glaucidium palmatum* and *Hydrastis canadensis* to receive more information on the placing of both genera. Seven species of the Berberidaceae were examined; inter alia *Nandina domestica*.

Molecular data for 7 plastid regions were generated: the *petB-petD* spacer, the *petD* group II intron, the *trnL* group I intron, the *trnL-F* spacer, the *trnK* group II intron (including *matK*) and the *trnK-psbA* spacer. For amplification and sequencing they were treated as 3 partitions ("*petD*" = *petB-petD* spacer and *petD* group II intron; "*trnL-F*" = *trnL* group I intron plus the *trnL-F* spacer; "*trnK*(*matK*)*-psbA*" = *trnK* group II intron (including *matK*) and the *trnK-psbA* spacer). The major part of the sequences was generated in this study. For *petD* 11 sequences were taken from Worberg & al. (2007), as well as for *trnL-F*. For *trnK(matK)-psbA* 30 completely new sequences were generated. 13 partial sequences originally produced for the analysis of basal eudicots presented in chapter 1 were completed by sequencing already existing products with additional primers. Sequences

from *Nandina domestica* and *Ranunculus macranthus* were obtained from the complete plastid chromosome sequence available in GenBank. Detailed information on all taxa included in this survey, respective vouchers and GenBank accession numbers is given in Appendix 1.

DNA isolation, amplification and sequencing

DNA was isolated from fresh or silica gel-dried plant material by using the CTAB-method described in Doyle & Doyle (1990). Three extractions were carried out to yield high amounts of genomic DNA (compare Borsch & al., 2003). In cases of suboptimal DNA quality extractions were cleaned using commercially available spin columns (Macherey-Nagel; Düren, Germany). Amplification and sequencing reactions were performed in a T3 Thermocycler or Gradient Thermocycler (Biometra; Göttingen, Germany). PCR protocols and reaction conditions followed Löhne & Borsch (2005) for *petD*, Borsch & al (2003) for *trnL-F*, Hilu & al. (2003) and Wicke & Quandt (in press) for *trnK(matK)-psbA*. Amplicons were purified using the NucleoSpin Extract II kit for cleanup of gel extraction (Macherey-Nagel; Düren, Germany) after running them out on a 1.2 % agarose gel for 2.5 h at 80 V.

Sequencing was performed using the PCR primers and specially designed internal primers in cases of long amplicons or problematic reads due to microsatellite areas. For *petD* the existing set of universal primers from Löhne & Borsch (2005) was used. The *trnL-F*-partition was amplified and sequenced by using primers trnL-C and trnL-F (Taberlet & al., 1991). Amplification of *trnK(matK)-psbA* was done with trnKFbryo (F, Wicke & Quandt, in press) and psbA-R (reverse, Steele & Vilgalys, 1994). For sequencing the whole fragment several additional internal primers were designed using SeqState v1.2 (Müller, 2005). All primers are listed in Table 1.

Table 1: Primers used for molecular work. D = direction.

Primer name	Sequence 5' – 3'	D	Reference	Region
PIpetB1411F	GCCGTMTTTATGTTAATGC	F	Löhne & Borsch (2005)	*petD*
PIpetD738R	AATTTAGCYCTTAATACAGG	R	Löhne & Borsch (2005)	*petD*
trnL-C	CGAAATCGGTAGACGCTACG	F	Taberlet & al. (1991)	*trnL-F*
trnL-F	ATTTGAACTGGTGACACGAG	R	Taberlet & al. (1991)	*trnL-F*
trnTFD	GGGGATAGAGGGACTTGAAC	R	Taberlet &al. (1991)	*trnL-F*
trnKFbryo	GGGTTGCTAACTCAATGGTAGAG	F	Wicke & Quandt (in press)	*trnK(matK)-psbA*
psbA-R	CGCGTCTCTCTAAAATTGCAGTCAT	R	Steele & Vilgalys (1994)	*trnK(matK)-psbA*
MG15F	ATCTGGGTTGCTAACTCAATG	F	Liang & Hilu (1996)	*trnK(matK)-psbA*
MG1	AACTAGTCGGATGGAGTAGAT	R	Liang & Hilu (1996)	*trnK(matK)-psbA*
BEtrnK1509F	GACTGTATCGCACTATGTA	F	This study	*trnK(matK)-psbA*
RANtrnK322F	GTGAATAAATGGATAGAGCC	F	This study	*trnK(matK)-psbA*
RANmatK641F	TTCYAAAGTCAAAAGAGCG	F	See chapter 1	*trnK(matK)-psbA*
RANmatK1265F	TTCCATTCTCACTGCGATTA	F	This study	*trnK(matK)-psbA*
RANmatK1414F	CCCATCCATCTKGAACTCTTGG	F	This study	*trnK(matK)-psbA*
XANmatK1490F	TTCTTTCTCTACGAGTATCAT	F	See chapter 1	*trnK(matK)-psbA*
RANmatK1797R	ATCTGAMATAATGYATGAAA	R	This study	*trnK(matK)-psbA*
LARmatK2353F	TCAACCTCTTCTACAGCCT	F	This study	*trnK(matK)-psbA*
RAmatK2100R	TGAAAATCATTAACAAAAACTAC	R	Worberg & al. (2007)	*trnK(matK)-psbA*
RANmatK2387R	AGGTCATTGATACRAATAATA	R	See chapter 1	*trnK(matK)-psbA*
EDmatKIF	CTCTGATTGGATCATTGGC	F	Worberg & al. (2007)	*trnK(matK)-psbA*

Direct sequencing was performed using the DTCS QuickStart Reaction Kit (BeckmannCoulter). Extension products were either run on a BeckmannCoulter CEQ 8000 automated sequencer in Dresden, or sequenced by Macrogen Inc., South Korea (www.macrogen.com). Sequences were edited manually with PhyDE v0.995 (Müller & al., 2005).

Alignment, indel coding and phylogentic analyses

Nucleotide sequences were aligned "by eye" using PhyDe v0.995, based on motif recognition as pointed out in Kelchner (2000) and Borsch & al. (2003). Sequence stretches with unclear primary homology were marked as "hotspots" (H) and excluded from the phylogenetic analyses. Inversions were inverted and thus included in the phylogenetic inferences as discussed by Quandt & al. (2003). For utilizing indel characters, the simple-indel coding method by Simmons & Ochoterena (2000) was applied via SeqState v1.2. Afterwards the indel matrix was combined with the nucleotide-sequence matrix and used for parsimony analyses and Bayesian Inference (BI). Most parsimonious trees (MPTs) were calculated by using the parsimony ratchet (Nixon, 1999) as implemented in PRAP (Müller, 2004). Ratchet settings were 20 random-addition cycles of 200 ratchet replicates, and upweighting 25% of the characters. In cases with multiple MPTs a strict consensus trees was drafted. Evaluation of nodes was done by

bootstrapping in PAUP* version 4.0b10 for Windows (Swofford, 2002) using 1000 replicates.

BI was done using MrBayes v3.1 published by Ronquist & Huelsenbeck (2003). The GTR + Γ + I model was applied for sequence data, and the restriction site model ("F81") for the indel matrix. Four runs (1,000,000 generations each) with 4 chains each were run simultaneously. Chains were sampled every 10th generation. The consensus tree and the posterior probability (PP) of clades were calculated based upon the trees sampled after the burn-in set at 250,000 generations. TreeGraph (Müller & Müller, 2004) was used for tree drawing. Datasets are deposited on the appended CD.

2.4 Results

Sequence variability

Each of the 3 partitions studied displayed considerable length variation as the individual spacers and introns do (Table 2). The *petB-petD* spacer extents from 176 to 223 nt, the *petD* intron from 673 to 737 nt, the *trnL* intron ranges from 397 to 533 nt, while the *trnL-F* spacer exhibits the greatest variation in length (142 to 467 nt). This is due to the fact that large parts of the *trnL-F* spacer are missing in the ranunculaceous genus *Clematis*. The *matK* gene is showing a length variation of 53 nucleotides. It ranges from 1503 to 1556 nt, whereas the entire *trnK* intron is displaying a length between 2351 and 2545 nt. Nucleotide counts deviating from the triplet code within *matK* exclusively occurred in downloaded sequenes from GenBank and are most likely artifical due to insufficient sequence editing. The *psbA* spacer is one of the most variable markers used in this study. It extends from 184 to 384 nt in length. High standard deviations of the mean sequence lengths as exposed by the *petD* and the *trnK(matK)-psbA* partition are due to partially missing sequence data (*Anemone acutiloba*, *Anemone transsilvanica*). The *petD* partition provided a set of 1620 characters, while *trnL-F* consists of 2147 characters. With 3618 positions the *trnK* intron including the *matK* gene displays the highest amount of aligned sequence characters. Comparing the *trnK* intron with the *matK* gene it is conspicuous that the coding part of the region provided about 38% more variable and informative characters than the non-coding pieces. In contrast the *trnK* provided about 86.7% of the indel information of *trnK(matK)*-region (356 of 421 coded indels). 127 (about 34.8%) of the indels coded within the non-coding part of the region are parsimony informative. In

comparison the *matK* gene contains only 13 (23.2%) parsimony informative indel characters.

The transition/transversion (Ti/Tv) ratio is lowest in the *petB-petD* spacer (0.844), and so does the GC content (29.4 %). Unlike the spacer the *petD* intron exhibits the highest values (1.305; 38.7%). The remaining non-coding genomic regions and the *matK* gene are characterized by Ti/Tv ratios ranging from 0.915 to 1.198 and GC contents between 31.0 and 35.9 %. Detailed information on sequence statistics of the several molecular markers studied is summarized in Table 2.

Table 2: Variation and relative contribution of the genomic regions studied. Number and quality of characters, indels coded and GC content, as well as transition/transversion ratio are calculated with mutational hotspots excluded; inversions were inverted. SD = Standard deviation, No.-char. = Number of characters, var.-char. = variable characters, inf.-char. = informative characters, PI indels = parsimony informative indels, Ti/Tv ratio = transition/transversion ratio.

Region	**mean sequence length (bp)**	**SD**	**mean sequence length excl. hotspots (bp)**	**SD**	**No. char.**	**var. char. [%]**	**inf. char. [%]**	**No. of indels coded**	**PI indels [%]**	**GC-content [%]**	**Ti/Tv ratio**
petB-petD spacer	195	41	194	41	496	24.6	15.5	85	34.1	29.4	0.844
petD-5'exon	8	2	8	2	8	0	0	0	0	50.0	-
petD intron	679	140	679	140	1116	38.3	26.0	170	34.1	38.7	1.305
trnL intron	484	22	459	18	844	32.7	21.6	165	30.9	35.3	1.049
trnL-3'exon	50	4	50	4	55	10.9	5.5	3	33.3	45.5	0.206
trnL-F spacer	366	54	352	53	1248	25.7	19.1	295	37.6	35.3	1.179
trnK(matK)	2436	218	2415	216	3618	45.7	32.0	421	33.3	34.3	1.151
trnK	923	149	901	146	1884	33.9	23.7	365	34.8	35.9	1.198
matK gene	1513	71	1513	71	1734	58.5	40.9	56	23.2	33.4	1.140
trnK-3'exon	33	8	33	8	35	14.3	5.7	0	0	64.6	0.413
trnk-psbA spacer	231	65	169	53	513	27.7	19.5	133	26.3	31.0	0.915

Mutational hotspots were identified in all 3 partitions. They were defined by length-variable poly A/T stretches (microsatellites) or showed difficulties in motif recognition due to frequent and overlapping microstructural changes comprising several nucleotides. Altogether 12 mutational hotspots were determined, generally ranging from 3 to 20 nt in length. In contrast to the other regions the *psbA*-spacer includes only one large mutational hotspot. It is characterised by poly A/T stretches and several unalignable sections covering large parts of the spacer. For further information on hotspot positions and extent see Table 3.

Table 3: Hotspot (H) positions in alignment and region.

No. hotspot	Position in alignment	Region
H1	133-170	*trnL* intron
H2	195-210	*trnL* intron
H3	594-690	*trnL* intron
H4	1051-1123	*trnL-F* spacer
H5	1341-1362	*trnL-F* spacer
H6	2057-2061	*trnL-F* spacer
H7	2630-2639	*petB-petD* spacer
H8	4463-4468	*trnK* intron 5’
H9	4772-4795	*trnK* intron 5’
H10	7350-7413	*trnK* intron 3’
H11	7696-7702	*trnK* intron 3’
H12	7813-8392	*trnk-psbA* spacer

The combined indel matrix provided a set of 1272 characters. Length mutations were mostly identified as simple sequence repeats (SSR) comprising 4-6 nucleotides. A number of indels is shared by specific clades. A prominent example is an inverted repeat of 6 bp, which is situated 153 nucleotides downstream the *matK* gene. It is unique to all species of the Lardizabalaceae studied and preceded by a synapomorphic deletion of 9 nucleotides (Fig. 1).

Several inversions were found in the molecular dataset. One autapomorphic inversion was identified in the *petD* intron of *Hydrastis canadensis* (alignment positions 2721-2735), 3 inversions occur in the *trnK(matK)-psbA* partition. Two of them are situated in the 5’ part of the *trnK* intron (651 and 708 nucleotides downstream the 5’ end of the *trnK* intron). Both are synapomorphic for the Ranunculaceae clade. The third one is found 1722 bp downstream the *matK* gene. It occurs in several taxa of the Berberidaceae, Papaveraceae and Ranunculaceae as well as in different species of the basal angiosperms (*Acorus calamus*, *Magnolia officinalis* and *Umbellularia californica*). In agreement with previous findings all 3 inversions are located in the terminal loop of a hairpin (e.g., Kelchner & Wendel, 1996; Quandt & Stech, 2004; Hernández-Maqueda & al., 2008).

Euptelea pleiosperma	ATAGATCAATTCTTTACGAACCCATGGAAAAT------TTAGG
Bocconia frutescens	ATGGATCGATTCTTTATGAACCCGTGAAAAAT------TTAGG
Stylophorum diphyllum	ATAGATCGATTCTTTATGAAACCACGAAAAGT------TTAGG
Eschscholzia californica	ATAGATCGATTCCTTATGAAACTGTGAAAAAT------GTAGG
Papaver triniifolium	ATGGATCAATTCTTTATGAATCCGTGAAAAAT------GTAGG
Pteridophyllum racemosum	ATAGATCGATTATTTACGAACCCGGTGAAATT------TTAGG
Dicentra eximia	AGGGATCAATTCCTTACGAACCTGTGGAAAAT------TTAGG
Corydalis nobilis	AAGGATCCAGTCCTTACGAACCCGGGGGAAAT------TTAGG
Stauntonia hexaphylla	ATAGATCGATTCTT---------GTGGAAAATTTCCATTTAGG
Stauntonia brachyanthera	ATAGATCGATTCTT---------GTGGAAAATTTCCATTTAGG
Stauntonia purpurea	ATAGATCGATTCTT---------GTGGAAAATTTCCATTTAGG
Akebia longeracemosa	ATAGATCGATTCTT---------GTGGAAAATTTCCATTTAGG
Akebia quinata	ATAGATCGATTCTT---------GTGGAAAATTTCCATTTAGG
Decaisnea fargesii	ATAGATCGATTCTT---------GTGGAAAATTGCAATTTAGG
Holboellia coriacea	ATAGATCGATTCTT---------GTGGAAAATTTCCATTTAGG
Sinofranchetia chinensis	ATAGATCGATTCTT---------GTGGAAAATTTCCATTTAGG
Lardizabala biternata	ATAGATCGATTCTT---------GTGGAAAATTTCCATTTAGG
Boquila trifoliolata	ATAGATCGATTCTT---------GTGGAAAATTTCCATTTAGG
Menispermum canadense	CTAGAGCGATTCTTTATGAACCTGTGGAAAAT------TTAGG
Cocculus laurifolius	CTAAAGCGATTATTTCGGAATCTGTGGAAAAT------TTAGG
Stephania delavayi	CTAAAGCGATTCTTTATGAATCTGTGAAAAAA------TTAGG
Nandina domestica	ATGGATCGATTCTTTATGAACCTATCGAAAAT------TTAGG
Mahonia japonica	ATGGATCGATTCTTTATGATTCTATCGATAAT------TTAGG

Figure 1: Illustration of an inverted repeat as well as a 9 nt deletion unique to Lardizabalaceae located 153 nt downstream of *matK*.

Phylogeny of Ranunculales

The combined data sets (*petD*, *trnL-F*, *trnK*(*matK)-psbA*) comprised 7935 characters. In total 2952 were variable and 2047 parsimony informative. The simple indel coding approach applied on the combined matrix yielded 1272 binary characters that were included in the analysis. The parsimony ratchet analysis resulted in 6 MPTs of 10112 steps (CI = 0.585, RI = 0.685) the strict consensus of which is shown in Fig 2. Bayesian inference resulted in a topology almost identical to the one obtained through parsimony (Fig. 3). The single most difference is the position of *Eschscholzia californica*. According to BI *Eschscholzia* is resolved as the first branching lineage within the Papaveraceae s.str. sharing a sistergroup relationship with the remaining members of the family, albeit support slightly misses significance at the 0.1 level (PP 0.87/0.81, as in the following the second value refers to support obtained with the binary indel matrix included in the analyses). In contrast, parsimony analyses depicted *Eschscholzia* as being sister to *Papaver* with moderate support (BS 83/84). Apart from single exceptions such as the first-branching position of *Euptelea* or the branching order within Berberidaceae indel coding generally increased the support of the clades.

Although both approaches (BI & MP) yielded fully resolved trees with high support for the individual clades, the resolution of the branching order among the first branching Ranunculales could not be solved with confidence. Both approaches resolved Eupteleaceae as first branching lineage, albeit lacking support. Within Papaveraceae s.l. the former Fumariaceae are clearly monophyletic with *Pteridophyllum* resolved as sister (BS 100/99, PP 1.0/0.97). The backbone of core Ranunculales (= Lardizabalaceae, Menispermaceae, Berberidaceae, Ranunculaceae) is well resolved and gained maximum support in all analyses. Lardizabalaceae are branching first, followed by Menispermaceae. Inside Lardizabalaceae the support of *Decaisnea fargesii* as sister to the remaining taxa is maximal. It is followed by *Sinofranchetia chinensis*, *Lardizabala biternata* + *Boquila trifoliolata* and a clade consisting of the remaining taxa. *Holboellia coriacea* is nested within the genus *Stauntonia* with high confidence (BS 81/83, PP 1.0/1.0). Within Berberidaceae the clade consisting of *Mahonia* and *Berberis* was depicted as being sister to *Gymnospermium* and *Nandina*, a scenario without bootstrap support (BS 54/<50). However, in model based analyses statistical support for this hypothesis gained significance (PP 0.97/0.94). Ranunculaceae include *Glaucidium* and *Hydrastis*. *Glaucidium* is identified as first lineage of the family with moderate bootstrap support (BS 66/77) while *Hydrastis* receives maximum support as the second branch in all approaches. *Xanthorhiza* is followed by a group containing *Semiaquilegia* and *Thalictrum* (BS 75/83, PP 1.0/1.0). *Anemone* is identified as non-monophyletic. The *Anemone-Clematis*- clade follows *Aconitum volubile* (BS 85/81, PP 1.0/1.0) and a monophyletic group consisting of *Helleborus* and *Ranunculus* (BS 100/100, PP 1.0/1.0).

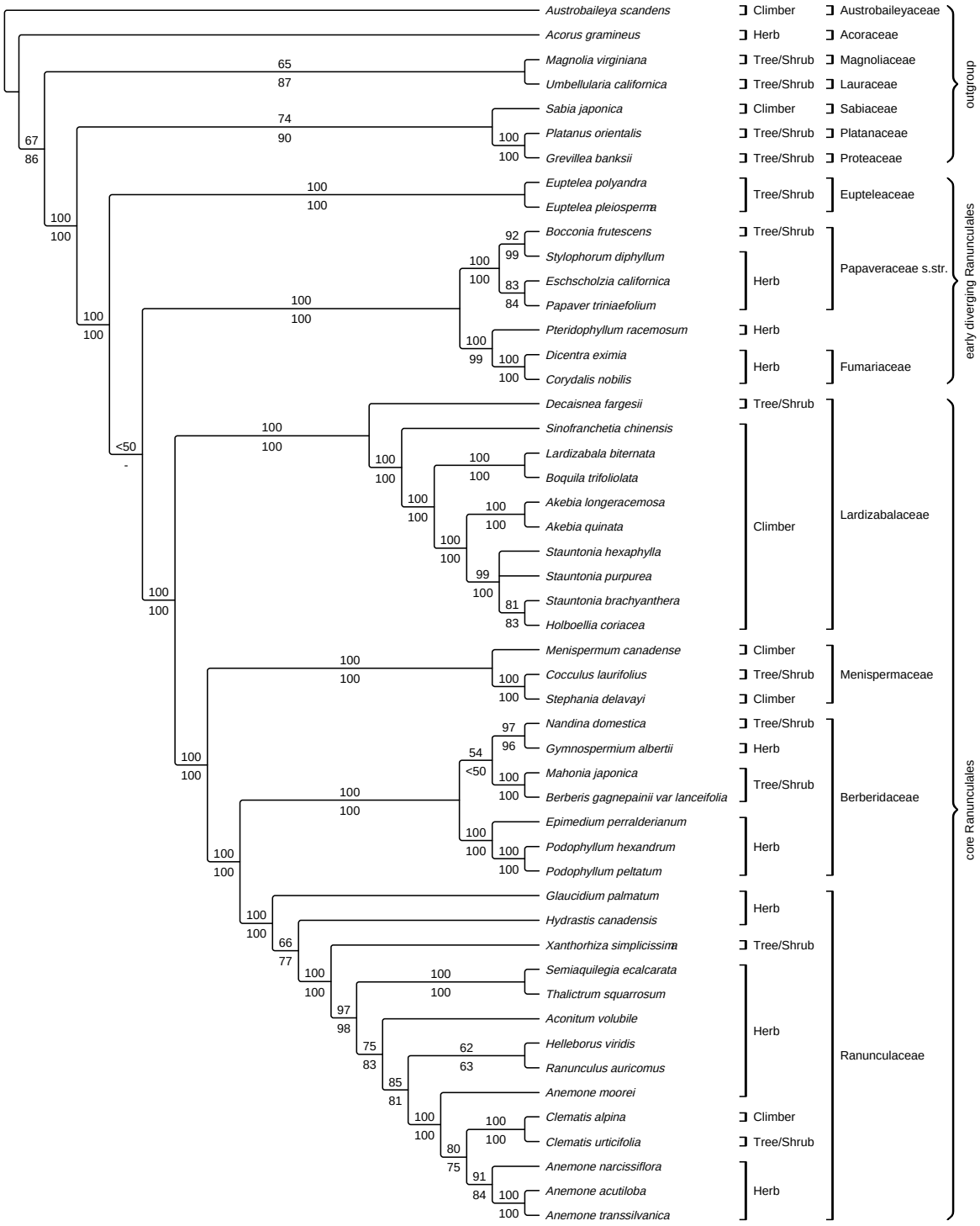


Figure 2: Strict consensus tree based on substitutions and indels of all 3 regions, inferred with MP. Values above and below branches are Bootstrap percentages, referring to substitutions only or substitutions plus indels, respectively. Growth forms are indicated behind taxon names.

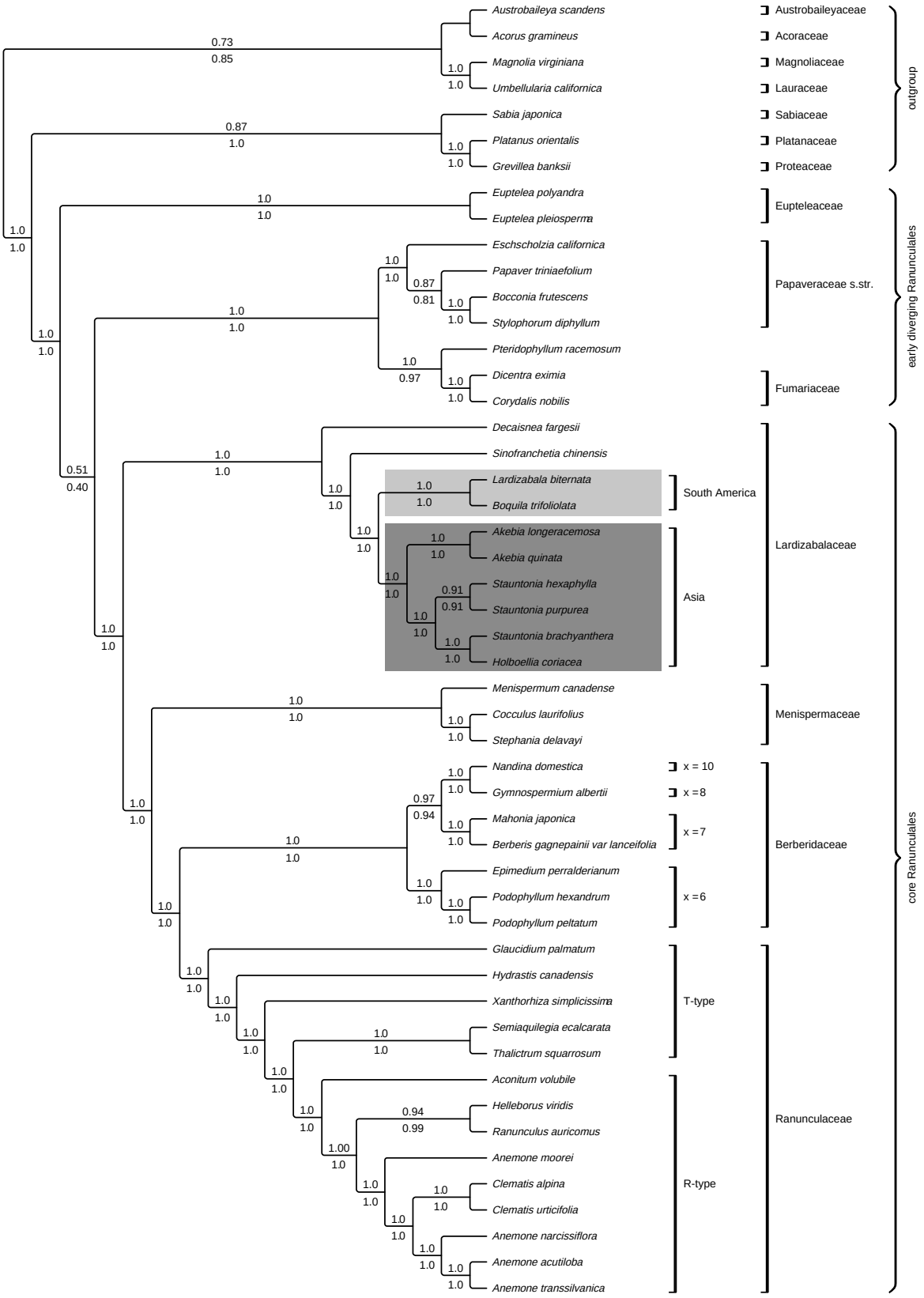


Figure 3: Bayesian tree based on the combined *petD+trnL-F+trnK(matK)-psbA* matrix. Posterior Probabilities are depicted above (substitutions only) and below (substitutions plus indels) branches. Base chromosome numbers within Berberidaceae and different chromosome types inside Ranuculaceae are indicated behind taxon names.

2.5 Discussion

A number of phylogenetic studies based on molecular data have supplied a framework of relationships among Ranunculales (Chase & al., 1993; Hoot & Crane, 1995; Hoot & al., 1999; Savolainen & al., 2000; Soltis & al., 2000; 2003; Hilu & al., 2003; Kim & al., 2004; Worberg & al., 2007). However, several questions concerning the branching order still remained to be answered.

Five coding genes have been used so far for reconstructing relationships inside early branching eudicots and thus Ranunculales as well. Initial analyses based on the plastid *rbcL* gene resulted in a topology showing Papaveraceae and Eupteleaceae outside a core clade consisting of Circeasteraceae, Lardizabalaceae, Berberidaceae, Menispermaceae and Ranunculaceae (Chase, 1993; Savolainen & al., 2000). Nevertheless support of the backbone-nodes was lacking. Hoot & al. (1999) and Soltis & al. (2000) added the plastid *atpB* gene and nuclear ribosomal 18S sequences. Furthermore, they studied an increased number of taxa. Their studies showed improved resolution and confidence, especially within the core clade of the order. Both depicted Ranunculaceae forming a clade with Menispermaceae and Berberidaceae (BS 88 or JK 70, respectively). Additionally, a clear sister group-relationship is shown between Berberidaceae and Ranunculaceae (BS 98, JK 92). Whereas, the succession of the aforementioned analyses postulated Papaveraceae s.l. as branching off first among early diverging Ranunculales, this scenario was contradicted with the inclusion of 26S data by Kim & al. (2004), This combined analysis of 4 coding regions (*rbcL*, *atpB*, 18S, 26S) inferred Eupteleaceae as first branching lineage, with moderate support under parsimony (BS 70) but significant support using Bayesian inference (Kim & al., 2004). By combining molecular data of the plastid *petD*, *trnL-F* and *matK* and indel information for the first time, Worberg & al. (2007) were able to enhance confidence in the first branching position of *Euptelea* (JK 80/81) through parsimony. However, their analyses indicated that model based approaches might reach a different conclusion, placing *Euptelea* sister to Papaveraceae s.l. in the Bayesian inferences, albeit lacking support.

The phylogenetic reconstructions using parsimony and Bayesian Inference resulted in well resolved topologies. As in most previous studies the Ranunculales gained maximum support, (e.g., Hoot & Crane, 1995; Soltis & al., 2003; Worberg & al., 2007). However, although the presented analyses are based on an increased and more representative sampling in terms of taxa and molecular markers, the position of the *Euptelea* could not

be resolved with confidence. Family relationships inside the core clade are identical in both, MP as well as BI and the branching order of Lardizabalaceae being sister to the remaining Ranunculales, followed by Menispermaceae, Berberidaceae and Ranunculaceae is in agreement with various multigene analyses (Hoot & al., 1999; Soltis & al., 2000; 2003; Kim & al., 2004) as well as the study of Hilu & al. (2003) based on *matK*.

Recent molecular studies already showed the high phylogenetic utility of rapidly evolving and non-coding genomic regions from the chloroplast genome in comparison to slowly evolving protein-coding markers, especially at high taxonomic levels (Borsch & al., 2003; Müller & al., 2006; Worberg & al., 2007). Statistical values of the sequence data used in this study show that the *matK* gene provided almost double the amount of informative characters in relation to the total number of characters compared to the introns and spacers and underlines the phylogenetic power of *matK* (see Table 2). As expected, indel information was mainly provided by introns and spacers. Whereas, both introns in *trnL* and *trnK* as well as the *petB-petD* spacer provided similar amounts of coded indels per mean sequence length, the number doubles in the *trnK-psbA* and *trnL-F* IGSs (see Table 2). This roughly corresponds with less evolutionary constrains in the spacers compared to introns that need to maintain a functional secondary and tertiary structure. With *matK* and the *trnk-psbA* spacer being the exception, generally about one third of the indels coded within each region were parsimony informative. By using the nucleotide-sequence matrix and the indel information of the 6 non-coding plastid markers + *matK* of an increased taxon sampling it was possible to raise statistical values of the respective nodes within the core clade from mainly moderate to maximum support, especially in MP analyses.

Early-diverging Ranunculales

A central goal of the study was to clarify the phylogenetic relationships within the early-diverging Ranunculales in order to gain insights into the ancestral conditions of growth forms at the base of the grade. Cronquist (1981; 1988) considered the order as originally herbaceous with all of its woody members being only secondarily woody. Phylogenetic studies by Hoot & al. (1999) and Soltis & al. (2000) which placed the primarily herbaceous Papaveraceae basal to the remainder of Ranunculales might increase confidence in this hypothesis although support for this scenario was lacking. Doyle & Endress (2000) presented the same topology obtained through combined molecular and

structural analyses. Again, the first branching position of Papaverales (= Papaveraceae) gained only weak bootstrap support (BS 65). Papaveraceae were followed by *Euptelea* (BS 97). In contrast, as mentioned above several other surveys showed the woody Eupteleaceae as first-branching among Ranunculales, although statistical support was never truly convincing (e.g., Hilu & al., 2003; Kim & al., 2004; Worberg & al., 2007). However, during the last 6 years the Eupteleaceae first hypothesis became more popular. Surprisingly, despite the increased taxon sampling and 2 additional markers in comparison to Worberg & al. (2007) support for Eupteleaceae as the first branching ranunculid family drops drastically in our study. This is extremely evident considering the results obtained via parsimony. Whereas, Worberg & al. (2007) reported moderate Jacknife support (81%) for the combined parsimony analysis, our analyses yielded no bootstrap support. However, in contrast to Worberg & al. (2007) Bayesian inference resolved *Euptelea* as branching off first. Nevertheless statistical support was absent (PP 0.51/0.40). In the light of previous studies (e.g., Soltis & al., 2000; Kim & al., 2004; Worberg & al., 2007) the decreasing support values are difficult to explain. One line of argumentation could be that Worberg & al., (2007) as well as this study is solemnly based on plastid markers, whereas Soltis & al. (2000) and Kim & al. (2004) rely on a combination of plastid and nuclear data. It seems that the Eupteleaceae first scenario in Kim & al. (2004) is mainly due to the addition of 26S data, as the same data matrix in Soltis & al. (2000) having only 18S as nuclear partition fails to resolve Eupteleaceae as first branching. This is rather surprising as the plastid markers used, especially *matK* or combinations thereof generally perform better compared to the rather peculiar 26S (e.g., Hilu & al., 2003 versus Kim & al., 2004, concerning resolution of the branching order within basal eudicots). Thus, 26S might have fixed a signal that could point towards an interesting phylogenetic problem at the first ranunculean dichotomy. However, the results clearly indicate that the almost accepted Eupteleaceae first hypothesis is far from being settled. Therefore, it remains difficult to reconstruct the ancestral condition of plant habit for Ranunculales as already stressed by Kim & al. (2004).

In agreement with the study of Hoot & al. (1999) the Papaveraceae s.l. are identified as monophyletic with maximum statistical support in all presented analyses. The Papaveraceae s.l. are united by the possession of a paracarpous gynoecium and the presence of secretory idioblasts or laticifers (except *Pteridophyllum*, Kadereit & al., 1994, 1995; Hoot & al., 1997). Four major subgroups were recognised in previous studies: *Pteridophyllum* a monotypic genus from Japan (= Pteridophyllaceae, Lidén, 1993a), the

Mediterranean and Asian genus *Hypecoum* (Fumariaceae subf. Hypecoideae; Dahl, 1989; 1990; 1992; Lidén, 1993b – genus not presented in this study), Fumariaceae subf. Fumariodeae (Lidén, 1993b) and Papaveraceae s.str. (Kadereit, 1993; Kadereit & al. 1994). Based on cladistic analyses of morphological characters 2 deviating phylogenetic scenarios were proposed, a result mainly based on rooting with different outgroups. Loconte & al. (1995) chose Ranunculaceae to represent the outgroup, based on the assumption of a sistergroup relationship between Ranunculaceae and Papaveraceae. As a result a grade of Papaveraceae s.str., *Pteridophyllum* and Fumariaceae appeared. In contrast analyses by Kadereit & al. (1994) assumed a sistergroup relationship of *Pteridophyllum* to Fumariaceae and Papaveraceae s.str, thus chosing *Pteridophyllum* as outgroup in their analyses addressing the evolution of the Papaveraceae (Kadereit & al. 1994). Recent phylogenetic analyses did not support the choice of outgroups. They identified Papaveraceae s.l. and Eupteleaceae as early diverging lineages within Ranunculales, being excluded from a core clade (e.g., Hoot & Crane, 1995; Hoot & al., 1999). However, molecular analyses of 2 plastid genes by Hoot & al. (1997) supported the scenario described by Kadereit & al. (1994). Nevertheless statistical support for the first-branching of *Pteridophyllum* among Papaveraceae s.l. was almost absent. Similarly, the sister group relation of Fumariaceae and Papavaraceae s.str. yielded no significant support. In contrast to Hoot & al. (1997), this study clearly places *Pteridophyllum* as sister to the Fumariaceae (BS 99/100, PP 0.9/1.0). Thus, Papaveraceae s.str. are sister to the *Pteridophyllum*-Fumariaceae clade (BS 100/100, PP 1.0/1.0). The conflicting position of *Pteridophyllum* between the presented analyses and Hoot & al. (1997) are most likely due to the choise of molecular markers in Hoot & al. (1997). The phylogenetic signal in the plastid (*atpB* plus *rbcL*) data set seems to be rather weak as indicated by the overall low support values. Since the model based as well as parsimony analyses converge to the same scenario with maximum support, differing earlier proposed scenarios seem to be unlikely.

Core Ranunculales

The predominantly twining woody Lardizabalaceae have been considered to be the most archaic member of the group most likely sharing a common origin with Menispermaceae (Takhtajan, 1997). Comparative studies of floral evolution and seed structure pointed at a close relationship of both families (Endress, 1995; Brückner, 1995). Recent molecular

studies placed both families within the core Ranunculales, despite the topological differences concerning the exact positions. Results of Soltis & al. (2000) showed Lardizabalaceae+*Sargentodoxa* in a tritomy with Circaeasteraceae and the other core Ranunculales. Using 3 or 4 genes respectively Hoot & Crane (1995), Hoot & al. (1999), Soltis & al. (2003) and Kim & al. (2004) presented a Lardizabalaceae-Circaeasteraceae clade as sister to the remaining members of the core Ranunculales, followed by Menispermaceae. Based on an extensive sampling and a combination of different fast evolving regions this study could increase confidence in a sistergroup relationship of the Lardizabalaceae to all residual families of the core group. The respective positions of Lardizabalaceae and Menispermaceae gained maximum statistical support in both approaches (MP & BI). Thus a second weakly supported hypothesis of both families forming a clade presented by Worberg & al. (2007) can be rejected. Since the respective positions of the herbaceous Circaeasteraceae was only weakly or moderately supported in previous studies and since the family was not included in the presented phylogenetic reconstructions the Lardizabalaceae-Circaeasteraceae clade can't be evaluated. Moreover, available *trnL-F* sequences in GenBank shared more similarities with Brassicales sequences than with Ranunculales and were therefore omitted from the phylogenetic analyses presented here. Thus the evolution of the woody climbing habit among the first branching lineages within the core Ranunculales remains to be solved. However, although comparative examinations of wood anatomy revealed similarities between Lardizabalaceae and Menispermaceae in many respects this seems to be attributed to the climbing habit rather than to systematic conditions (Carlquist, 1995).

For the first time 7 genera of Lardizabalaceae were included in a phylogenetic reconstruction among the order Ranunculales. However, the Asian *Sargentodoxa* a rarely cultivated taxon is not included here, due to the fact that no plant material was available. Two tribal classifications based on morphological data were published by Qin (1989) and Loconte & Estes (1989). Qin (1989) recognized 4 different tribes. Two of them were monotypic (Decaisneae, Sinofranchetieae), while the 2 South American genera *Boquila* and *Lardizabala* were placed into Lardizabaleae. The Asian genera *Akebia*, *Stauntonia* and *Holboellia* were treated as tribe Stauntonieae. *Sargentodoxa* was excluded from the family. The classification of Loconte & Estes (1989) differs by recognizing 2 subfamilies (Decaisneoideae, Lardizabaloideae). Lardizabaloideae were divided into Sinofranchetieae and 2 additional groups consisting of *Sargentodoxa* and *Boquila* (group 1) and *Lardizabala*, *Akebia*, *Holboellia* and *Stauntonia* (group 2). Lardizabalaceae clearly form a

monophyletic group which gains maximum statistical values. The shrubby *Decaisnea* was identified as first branching within the family (BS 100/100, PP 1.0/1.0). Takhtajan (1997) already suggested this genus to be the most ancestral member of the group based on its upright stem and polygamous flowers. Additionally this finding as well as the second branching position of *Sinofranchetia* is in accordance with previous phylogenetic studies based on traditional data and classification schemes (Qin, 1989; Loconte & Estes, 1989) and *atpB*+*rbcL*+18S (Hoot & al., 1995b). *Decaisnea* and *Sinofranchetia* are followed by a clade containing the South American genera, which received maximum support in all approaches. A cladistic analysis on RFLP data carried out by Kofuji & al. (1994) resulted in the same topology, although bootstrap support was only moderate (BS 70). The analysis of Kofuji & al. (1994) as well as the presented study showed a well supported sistergroup relationship between *Lardizabala*+*Boquila* and a clade of the remaining genera from East Asia, a scenario already indicated by analyses of Hoot & al. (1995a). This contradicts the results of Hoot & al. (1995b) which showed *Lardizabala* as being sister to the remainder of Lardizabalaceae. Since this scenario was basically not supported (BS 57) it seems to be unlikely. At the same time a hypothesis by Schuster (1976) concerning the present-day range of the family is supported. He proposed a possible origin in Gondwana and a subsequent spreading to Laurasia. Consequently the Chilean group represents a relict. Within the Asian clade of the family *Akebia* appears sister to the remaining 2 genera, as it is already indicated in previous studies (Kofuji & al., 1994; Hoot & al., 1995a; 1995b). *Holboellia coriacea* is confirmed as nested within *Stauntonia* with high confidence (BS 81/83, PP 1.0/1.0), a result already suggested by Kofuji & al. (1994). The topology presented in this study clearly corroborates the classification of Qin (1989), although *Sargentodoxa* was not included here. The analysis by Hoot & al. (1995b) based on 3 genes depicted *Sargentodoxa* as being sister to the remainder of the Lardizabalaceae. This is in agreement with other systematic treatments that placed the genus as a separate family with close relationship to Lardizabalaceae, mainly based on its differing gynoecium (Cronquist, 1988; Cheng-Yih & Kubitzki, 1993; Takhtajan, 1997).

Menispermaceae, a pantropical family usually found in tropical lowlands, contain about 71 genera with 450 species (Kessler, 1993). Only 3 genera of the family were included in the analysis, thus no real statement can be given on the phylogenetic relationships and the evolution of growth forms inside this group. Examining the plastid *ndhF* gene of 88 species and plotting the growth form (climber, tree/shrub, herb) on the strict consensus parsimony tree Ortiz & al. (2007) argue that the climbing habit is pleisiomorphic.

Similarly, tree habit seems to have evolved multiple times from the climbing growth habit (Ortiz & al.; 2007). Hoot & al. (2009) analysed newly generated *atpB* and *rbcL* data alone as well as in combination with the *ndhF* sequences already published. Their analyses resulted in the recognition of most of the lineages found by Ortiz & al. (2007). Nevertheless, additional data is needed, as the positions of several taxa displaying a tree habit in relation to each other are not fully resolved yet.

The consistently herbaceous Ranunculaceae have been regarded to be relatively primitive (Cronquist, 1988; Tamura, 1993) and to be closely related to Berberidaceae (Cronquist, 1988; Loconte, 1993; Takhtajan, 1997). Recent molecular studies clearly revealed both families as belonging to the core clade within Ranunculales, displaying a well supported sistergroup relationship (e.g., Hoot & al., 1999; Soltis & al., 2000; Hilu & al., 2003). This strongly coincides with the results of the presented survey, which gained maximum statistical support for the respective nodes in both, parsimony and Bayesian analyses.

Berberidaceae are one of the largest groups within Ranunculales containing more than 650 species with an extremely diverse morphology (Loconte, 1993). Modern classification schemes proposed for the family differ drastically from the traditional systematic treatments (Airy Shaw, 1973; Meacham, 1980; Terabayashi, 1985a; b; Loconte & Estes, 1989; for a summary see Kim & Jansen, 1998). Meacham (1980) recognized 4 groups which can be identified by fruit type and chromosome number. Loconte & Estes (1989) gave a similar classification by reanalyzing the morphological characters that includes a subfamily Nandinoideae into Berberidaceae. Recent phylogenetic studies based on restriction site data and the chloroplast gene *ndhF* (Kim & Jansen, 1998; Kim & al., 2004) revealed 4 chromosomal lineages (x = 6, x = 7, x = 8, x = 10). However, phylogenetic relationships among these groups were not resolved or statistical support was low. This is also reflected in the parsimony results of this study. However, both approaches, MP and MB confirmed Berberidaceae as monophyletic, all chromosomal lineages were identified with high confidence (BS 96-100, PP 1.0). *Nandina domestica* (x = 10; BS 97/96, PP 1.0/1.0) is clearly nested within Berberidaceae, forming a clade with *Gymnospermium albertii* (x = 8), a fact already recognized by Kim & al. (2004) who rejected the segregation of the genus as a distinct family or subfamily. Albeit relationships among the *Epimedium-Podophyllum* group (x = 6), a clade containing *Mahonia* and *Berberis* (x = 7) and *Nandina* and *Gymnospermium* are resolved in parsimony analyses, the possible sistergroup relationship of the latter lineages gained no statistical supported (BS 54/<50). In contrast to the indecisive parsimony analyses, BI seems to converge to

the scenario (Wang & al., 2007; this study), although PPs reported by Wang & al. (2007) for the respective node were not significant. However, in this study significant support in a Bayesian framework was gained (PP 0.97/0.94).

The Ranunculaceae, a family distributed throughout the world but preferentially in temperate or subcold climates, contains more than 50 genera and about 2500 species (Tamura, 1993; Takhtajan, 1997). Two different chromosome types were recognized by Langlet (1932). He recommended the division of the family into two subfamilies, Ranunculoideae exposing long chromosomes, that are curved several times (R-type) and Thalictroideae having small simply curved chromosomes (T-type). Cladistic studies using micromorphological characters (Hoot, 1991) and sequence data of 3 combined coding regions (Hoot & al., 1995) suggested the T-type to occur in the early diverging lineages of the group, whereas the R-type seemed to be more derived. This corresponds largely with the results of the presented analyses. As already indicated by studies based on restriction site analyses (Johansson & Jansen, 1993; Johansson, 1995) and sequence data including *atpB*, *rbcL*, *adh*, 18S and 26S rDNA sequences (Kosuge & al., 1995; Hoot, 1995; Ro & al., 1997; Wang & al., 2005) *Xanthorhiza* and *Semiaquilegia ecalcarata*+*Thalictrum* are shown to be well supported distinct lineages. This clearly contradicts the classification of Tamura (1993; 1995) who used chromosome-types and fruit morphology as most important characters, placing *Xanthorhiza* and *Semiaquilegia* in Isopyroideae, and *Thalictrum* in Thalictroideae, a monogeneric subfamily. The remaining members of the Ranunculaceae, possessing R-type chromosomes, constitute a monophyletic group. Statistical support was moderate in MP analyses (BS 75/83), but reached maximum in BI. The recognition of the 3 clades is in agreement with several molecular systematic studies (Jensen & al., 1995; Ro & al., 1997; Wang & al., 2005) that proposed the establishment of 3 independent subfamilies. An analysis of the ovule morphogenesis in Ranunculaceae carried out by Wang & Ren (2008) clearly sustained this treatment. Within the subfamily Ranunculoideae (sensu Ro & al., 1997) *Anemone* is shown to be non-monophyletic. *Anemone moorei* is depicted as first-branching within the *Anemone*-*Clematis*-clade, followed by *Clematis alpina*+*Clematis urticifolia* sharing a sistergroup relationship with a clade containing *A. narcissiflora*, *A. acutiloba* and *A. transsilvanica*. However, this branching order has to be treated with care, as sampling inside this clade is not representative of the species diversity. The close relationship of the 2 genera is reflected in the classification of Tamura (1995) as well as in the systematic treatment on the basis of molecular data by Jensen & al. (1995). Both authors placed the genera within the tribe

Anemoneae. Comprehensive phylogenetic surveys based on a dense taxon-sampling and using sequence data have been carried out either for *Clematis* (Miikeda & al., 2006) or *Anemone* (Ehrendorfer & Samuel, 2001; Schuettpelz & al., 2002). Thus a thorough study enclosing both genera is needed to resolve phylogenetic relations inside the tribe.

Glaucidium and *Hydrastis* have been problematic in the past in terms of their systematic position. Recent molecular studies revealed both genera as closely related to the Ranunculaceae. Analyses differ in showing a sistergroup relationship between *Hydrastis*+*Glaucidium* and Ranunculaceae (Hoot & al., 1999) or identifying *Glaucidium* as sister to *Hydrastis*+Ranunculaceae (Soltis & al., 2003). Both hypotheses were only weakly supported. The presented study clearly corroborates the results of the analysis using 4 genes by Soltis & al., (2003). *Glaucidium palmatum* is identified as first branching, followed by *Hydrastis canadensis* and the remaining Ranunculaceae. BS stays low--moderate in MP (BS 66/77) for the early-diverging position of *Glaucidium*, while PPs reached 1.0 in BI. Hoot (1995) advocates for keeping the genera in 2 monotypic families, not included into Ranunculaceae. This is in accordance with systematic treatments of Tamura (1972), Tobe (1981) and Tobe & Keating (1985) using morphology, anatomy (including embryology), palynology, chemistry and cytology. Nevertheless there are several features that point on two highly autapomorphic lineages within the Ranunculaceae. *Glaucidium* as well as *Hydrastis* share the presence of T-type chromosomes with the early-diverging members of the family (Gregory, 1941). Additionally *Hydrastis* is characterized by the possession of berberin and yellow rhizomes, common features of *Coptis* and *Xanthorhiza* (Hoot, 1995). The results of the presented phylogenetic reconstruction suggest the inclusion of both genera into Ranunculaceae, forming two distinct subfamilies, Glaucidioideae Loconte (Pl. Syst. Evol. [Suppl.] 9 104/105. 1995): *Glaucidium*; and Hydrastidoideae. This is (partly) congruent with the findings of Ro & al. (1997) who included *Hydrastis* in their molecular phylogenetic study of Ranunculaceae and recommended the subdivision of the family into Hydrastidoideae, Coptidoideae, Thalictriodeae and Ranunculoideae.

Most ranunculaceous genera are herbs or tuber/rhizome-forming perennial herbs holding annual shoots with primary growth. Woody stems are restricted to *Xanthorhiza* and *Clematis*, which may be described as only “weakly shrubby” (Isnard & al., 2003). Considering the results of the molecular phylogenetic reconstructions the woody habit seems to have evolved 2 times independently within the Ranunculaceae.

Appendix 1: Taxa analysed, voucher datails, GenBank accession numbers; family assignment according to APG II (2003). Taxa are listed in alphabetical order.

Family, species, origin/garden, voucher/herbarium or reference, *petD, trnL-F* and *trnK(matK)-psbA* EMBL accession numbers. A dash indicates missing data.

OUTGROUP: Acoraceae. *Acorus calamus* L.: Germany, BG Bonn, Löhne 51(BONN), AY590840, - ; This study update. *Acorus gramineus* [Soland.]: Germany, BG Bonn, Borsch 3458 (BONN), - , AY145336, -. **Austrobaileyaceae.** *Austrobaileya scandens* C.T. White: Germany, BG Bonn, Borsch 3464 (BONN), AY590867, AY145326, DQ185523. **Lauraceae.** *Umbellularia californica* (Hook. & Arn.) Nutt.: Germany, BG Bonn, Borsch 3471 (BONN), AY590850, AY145350, This study update. **Magnoliaceae.** *Magnolia officinalis* Rehder & E.H. Wilson: Germany, BG Bonn, Löhne 53 (BONN), AY590846, - , This study update. *Magnolia virginiana* L.: USA, Maryland, Borsch & Neinhuis 3280 (VPI, FR), - , AY145354, AB020988. **Platanaceae.** *Platanus orientalis* L.: Germany, BG Bonn, Worberg 005 (BONN), AM396538, AM397164, This study update. **Proteaceae.** *Grevillea banksii* R. Br.: Germany, BG Bonn, Borsch 3413 (BONN), AM396537, AM397163, This study update. **Sabiaceae.** *Sabia japonica* Maxim.: USA, NCU, Qiu 91025 (NCU), AM396533, AM397158, This study update. **INGROUP: Berberidaceae.** *Berberis gagnepainii* var. *lanceifolia* Ahrendt: Germany, BG Dresden, Living collection 3215-11, This study, This study, This study. *Epimedium perralderianum* Coss.: Germany, BG Dresden, Living collection 000663-15, This study, This study, This study. *Gymnospermium albertii* (Regel) Takht.: Germany, BG Dresden, Living collection 012081-12, This study, This study, This study. *Mahonia japonica* DC.: Germany, BG Bonn, Borsch 3405 (BONN), AM396531, AM397156, This study update. *Nandina domestica* Thunb.: GenBank, NC_008336, NC_008336, NC_008336. *Podophyllum hexandrum* Royle: Germany, BG Dresden, Living collection 68 Uppsala 208, This study, This study, This study. *Podophyllum peltatum* L.: Germany, BG Bonn, Borsch 3393 (BONN), AM396532, AM397157, This study update. **Eupteleaceae.** *Euptelea pleiosperma* Hook.f. & Thomson: Germany, BG Bonn, Worberg 003 (BONN), AM396525, AM397151, This study update. *Euptelea polyandra* Siebold & Zucc.: Germany, BG Dresden, Barniske 042 (DR), This study, This study, This study. **Lardizabalaceae.** *Akebia logeracemosa* Matsum.: Belgium, BG Ghent, Living collection 2004-1276, This study, This study, This study. *Akebia quinata* Decne.: Germany, BG Bonn, Borsch 3412 (BONN), AM396526, AM397152, This study update. *Boquila trifoliolata* (DC.) Decne.: Chile, Vilches Alto, Barniske 045 (DR), This study, -., This study. *Boquila trifoliolata* (DC.) Decne.: GenBank, -, AF335291, - . *Decaisnea fargesii* Franch.: Germany, BG Dresden, Barniske 053 (DR), This study, This study, This study. *Holboellia coriacea* Diels: Germany, BG Dresden, Barniske 046 (DR), This study, This study, This study. *Lardizabala biternata* Ruiz & Pav.: Chile, BG Talca, Barniske 044 (DR), This study, This study, This study. *Sinofranchetia chinensis* Hemsl.: Germany, BG Dresden, Barniske 048 (DR), This study, - , This study. *Sinofranchetia chinensis* Hemsl.: GenBank, - , AF335284, - . *Stauntonia brachyanthera* Hand.-Mazz.: Belgium, BG Ghent, Living collection 2001-2272, This study, This study, This study. *Stauntonia hexaphylla* Decne.: Germany, BG Dresden, Barniske 052 (DR), This study, This study, This study. *Stauntonia pupurea* Y.C. Lui & F.Y. Lu: Belgium, BG Ghent, Living collection 2005-1626, This study, This study, This study. **Menispermaceae.** *Cocculus laurifolius* DC.: Germany, BG Bonn, Borsch 3406 (BONN), AM396528, AM397159, This study update. *Menispermum canadense* L.: Germany, BG Dresden, Living collection 4088-20, This study, This study, This study. *Stephania delavayi* Diels.: Germany, BG Bonn, Borsch 3550 (BONN), AM396529, AM397154, This study update. **Papaveraceae.** *Bocconia frutescens* L.: Germany, BG Dresden, Living collection 012357-18, This study, This study, This study. *Corydalis nobilis* Pers.: Germany, BG Dresden, Barniske 060 (DR), This study, This study, This study, *Dicentra eximia* (Ker Gawl.) Torr.: Germany, BG Bonn, Borsch 3468 (BONN), AY590835, AY14536, This study update. *Eschscholzia californica* Cham.: Germany, BG Dresden ,Living collection 003892-22, This study, This study, This study. *Papaver triniaefolium* Boiss.: Germany, BG Bonn, Worberg 018 (BONN), AM396527, AM397153, This study update. *Pteridophyllum racemosum* Siebold & Zucc.: Belgium, BG Ghent, Living collection 2007-1447, This study, This study, This study. *Stylophorum diphyllum* Nutt.: Germany, BG Dresden, Barniske 062 (DR), This study, This study, This study. **Ranunculaceae.** *Aconitum volubile* Pall. ex Koelle: Germany, BG Dresden, Barniske 051 (DR), This study, This study, This study. *Anemone acutiloba* Laws.: GenBank, - , AM268056, DQ994677. *Anemone moorei* Espinosa: Chile, Vilches Alto, Herbarium of Universidad de Talca, This study, This study, This study. *Anemone narcissiflora* L.: Germany, BG Dresden, Living collection 006254-17, This study, This study, This study. *Anemone transsilvanica* (Fuss) Heuff.: GenBank, - , AM268059, DQ994670. *Clematis alpina* (L.) Mill.: Germany, BG Dresden, Living collection 10401-6, This study, This study, This study. *Clematis urticifolia* Nakai x Kitag.: Germany, BG Dresden, Living collection 007462-19, This study, This study, This study. *Glaucidium palmatum* Siebold & Zucc.: Germany, BG Dresden, Living collection 012121-07, This study, This study, This study. *Helleborus viridis* L.: Switzerland, Mt. Generoso; Barniske 049 (DR), This study, - , This study. *Helleborus viridis* L.: GenBank, - , AJ413301, - . *Hydrastis canadensis* L.: Germany, BG Dresden, Barniske 043 (DR), This study, This study, This study. *Ranunculus auricomus* L.: Germany, BG Dresden, Barniske 059 (DR), This study, This study, - . *Ranunculus macranthus* Scheele: GenBank, - , - , NC_008796. *Semiaquilegia ecalcarata* (Maxim.) Sprague & Hutch.: Germany, BG Dresden, Living collection 010374-15, This study, This study, This study. *Thalictrum squarrosum* Steph. ex Willd.: Germany, BG Dresden, Barniske 054 (DR), This study, This study, This study. *Xanthorhiza simplicissima*Marshall: Germany, BG Dresden, Barniske 061 (DR), - , - , This study update. *Xanthorhiza simplicissima* Marshall: Germany, BG Bonn, Borsch 3394 (BONN), AM396530, AM397155, - .

Chapter 3

Phylogenetic relationships among *Anemone*, *Pulsatilla*, *Hepatica* and *Clematis* (Ranunculaceae)

3.1 Abstract

The ranunculaceous tribe Anemoneae currently consists of the subtribes Anemoninae, including the members of the *Anemone*–complex (*Anemone*, *Hepatica*, *Pulsatilla* and *Knowltonia)*, and Clematidinae, consisiting of *Archiclematis*, *Clematis* and *Naravelia*. Recent comprehensive molecular-phylogenetic studies have been carried out either for the members of Clematidinae or Anemoninae. To test phylogentic relationships among the subtribes as well as position and taxonomic rank of several lineages inside the Anemoninae, a molecular study based on the nuclear ITS 1&2 region in combination with the plastid *atpB-rbcL* spacer was performed. Here we present a phylogenetic reconstruction enclosing members of all major groups of both subtribes. The combined data matrix comprised 2589 aligned sequence positions and provided a matrix of 422 binary indel characters. Phylogenetic reconstructions resulted in a sistergroup relationship between Clematidinae and Anemoninae which was highly supported in all approaches. *Pulsatilla* and *Hepatica* are confidently shown to be nested within the genus *Anemone*. The informal section *Hepatica* is lifted to subgeneric rank. Phytogeographical patterns inside the Anemoninae are shortly discussed.

3.2 Introduction

The ranunculaceous genera *Anemone*, *Pulsatilla*, *Hepatica* and *Clematis* have always been considered to be very closely related. This is reflected in the results of a cladistic study of Ranunculaceae based on morphology carried out by Hoot (1991) which placed *Anemone* in a clade with *Clematis*, *Hepatica*, *Pulsatilla* and *Ranunculus*. These genera were united by the presence of achenes and the chemical compound ranunculin. Tamura (1995) described the tribe Anemoneae within the subfamily Ranunculoideae, containing the subtribes Anemoninea (including *Anemone*, *Hepatica*, *Pulsatilla* and *Knowltonia*) and Clematidinae (consisting of *Archiclematis*, *Clematis* and *Naravelia*). Molecular studies based on restriction site variation of chloroplast DNA (Johansson & Jansen, 1993; Johansson, 1995) and a combined analysis of *atpB*, *rbcL* and 18S sequence data (Hoot, 1995a) strongly implied the monophyly of the tribe. Using restriction site variation of chloroplast DNA, Hoot & Palmer (1994) demonstrated a sistergroup relationship between a clade consisting of *Anemone*, *Pulsatilla*, *Knowltonia* and *Hepatica* and *Clematis*. These

findings are also reflected in the classification system of Jensen & al. (1995) which supported the formation of two independent subtribes within Anemoneae. In contrast phylogenetic studies based on molecular as well as morphological data (Wang & al., 2009; chapter 2) hint at the possible paraphyly of the subtribe Anemoninae, a hypothesis that needed to be tested comprehensively.

Within the subtribe Anemoninae positions and taxonomic rank of certain lineages of the group are not fully clarified yet. The genera *Hepatica*, *Pulsatilla*, *Knowltonia*, *Barneoudia* and *Oreithales* were often treated as members of *Anemone* by early classifications (Prantl, 1891; Janczewski, 1892), while most of these taxa were excluded from the genus by several authors (Ulbrich, 1905/06; Tamura, 1967; 1993) mainly on the basis of deviating fruit morphology. The systematic survey of Ranunculaceae published by Tamura (1995) using chromosomes and fruits as most important characters recognized seven distinct genera within Anemoninae, ignoring a study by Hoot & al. (1994) dealing with the tribe. On the basis of a plastid and nuclear ribosomal DNA restriction analysis and morphology the latter revealed *Hepatica*, *Pulsatilla* and *Knowltonia* to be nested within the genus *Anemone*, while various morphological characters pointed on a possible inclusion of *Barneoudia* and *Oreithales* as well. Thus Hoot & al. (1994) recommended the placement of all members of the subtribe Anemoninae into one single genus (*Anemone* s.l.). The third finding, which was supported by a second study by Hoot (1995b), was the formation of two distinct clades. One consisting of the majority of the *Anemone*-species, *Pulsatilla* and *Knowltonia*, exposing a chromosome base number of x = 8 and another with x = 7, including several *Anemone* groups and *Hepatica*. Therefore Hoot & al. (1994) presented a preliminary classification, dividing the genus *Anemone* s.l. into the subgenera *Anemone* (x=8) and *Anemonidium* (x=7).

Since comprehensive phylogenetic analyses based on a dense taxon-sampling and using plastid and nuclear sequence data have been carried out either for the members of the Clematidinae (Miikeda & al., 2006) or the Anemoninae (Ehrendorfer & Samuel, 2001; Schuettpelz & al., 2002) a thorough study enclosing all major groups of both subtribes and applying both, substitutions and coded indel characters to parsimony and model based methods is presented to increase confidence into phylogenetic relationships inside the Anemoneae. Position and taxonomic rank of certain lineages within the subtribe Anemoninae, especially *Pulsatilla* and *Hepatica* is tested. Phytogeographical patterns inside the subtribe are shortly discussed.

3.3 Material and methods

Plant material

Plant material was obtained from the Dresden University of Technology Botanical Garden (Germany). Additional samples were taken from collections of A. Stoll and D. Quandt. Vouchers are deposited in DR.

Taxon sampling and molecular markers

In total, 67 taxa were included in the analyses. Four outgroup taxa were chosen to represent members of the order of Ranunculales (Eupteleaceae, Lardizabalaceae, Menispermaceae and Berberidaceae). The taxon sampling comprises all major groups of the subtribe Clematidinae (sensu Tamura, 1995). They are represented by *Archiclematis* (1, = *Clematis alternata*), *Clematis* (28) and *Naravelia* (1). In addition, representatives of four genera of the subtribe Anemoninea (Tamura, 1995) were surveyed. 18 species of *Anemone* (sensu Tamura, 1995) were included into analyses as well as the traditional genera *Hepatica* (8, including *Anemone americana*), *Pulsatilla* (3, including *Anemone occidentalis*) and *Knowltonia* (= *Anemone knowltonia*).

Molecular data for two genomic regions were analyzed: the nuclear ribosomal ITS1&2 and the *atpB-rbcL* intergenic spacer (IGS) from the large single copy region of the chloroplast. Most sequences were downloaded from GenBank. For Clematidinae all molecular data was taken from Miikeda & al. (2006). Most sequences for *Anemone* were originally published by Schuettpelz & al. (2002). The *atpB-rbcL* sequence data of four outgroup taxa was picked from chapter 1. Six completely new sequences were generated. Three new sequences were produced for the ITS regions as well as for the *atpB-rbcL* spacer region. All taxa included in this analysis, voucher information and GenBank accession numbers are given in Table 1.

Table 1: Taxa analysed, voucher datails, GenBank accession numbers and references; family assignment according to APG II (2003). Taxa are listed in alphabetical order.

Taxon	Family	Voucher/Herbarium	Garden/Field Origin	ITS	*atpB-rbcL*
OUTGROUP					
Euptelea pleiosperma Siebold & Zucc	Eupteleaceae	A. Worberg 003 (BONN)	BG Bonn	-	chapter 1
Euptelea pleiosperma Siebold & Zucc	Eupteleaceae	GenBank	-	AF162214 Feng et al. (1999)	-
Akebia quinata Decne.	Lardizabalaceae	T. Borsch 3412 (BONN)	BG Bonn	-	chapter 1
Akebia quinata Decne.	Lardizabalaceae	GenBank	-	AY029791 Wang et al. (2001)	-
Cocculus laurifolius DC.	Menispermaceae	T. Borsch 3406 (BONN)	BG Bonn	-	chapter 1
Cocculus laurifolius DC.	Menispermaceae	GenBank	-	AY017392 Hong et al. (2000)	-
Podophyllum peltatum L.	Berberidaceae	T. Borsch 3393 (BONN)	BG Bonn	-	chapter 1
Podophyllum peltatum L.	Berberidaceae	GenBank	-	AF328964 Liu et al. (2000)	-
INGROUP					
Archiclematis alternata (Kitam. & Tamura) Tamura [=*Clematis alternata* Kitam. & Tamura]	Ranunculaceae	Genbank	-	AB120190 Miikeda et al. (2006)	AB115440 Miikeda et al. (2006)
Clematis afoliata J.Buch.	Ranunculaceae	GenBank	-	AB120193 Miikeda et al. (2006)	AB115443 Miikeda et al. (2006)
Clematis angustifolia Jacq.	Ranunculaceae	GenBank	-	AB120199 Miikeda et al. (2006)	AB115449 Miikeda et al. (2006)
Clematis apiifolia DC.	Ranunculaceae	GenBank	-	AB120180 Miikeda et al. (2006)	AB115430 Miikeda et al. (2006)
Clematis brachyura Maxim.	Ranunculaceae	GenBank	-	AB120204 Miikeda et al. (2006)	AB115454 Miikeda et al. (2006)
Clematis crassifolia Benth.	Ranunculaceae	GenBank	-	AB120194 Miikeda et al. (2006)	AB115444 Miikeda et al. (2006)
Clematis eichleri Tamura	Ranunculaceae	GenBank	-	AB120209 Miikeda et al. (2006)	AB115459 Miikeda et al. (2006)
Clematis fasciculiflora Franch.	Ranunculaceae	GenBank	-	AB120203 Miikeda et al. (2006)	AB115453 Miikeda et al. (2006)
Clematis florida Thunb.	Ranunculaceae	GenBank	-	AB120186 Miikeda et al. (2006)	AB115436 Miikeda et al. (2006)
Clematis fusca Turcz.	Ranunculaceae	GenBank	-	AB120179 Miikeda et al. (2006)	AB115429 Miikeda et al. (2006)
Clematis gentianoides DC.	Ranunculaceae	GenBank	-	AB120210 Miikeda et al. (2006)	AB115460 Miikeda et al. (2006)
Clematis	Ranunculaceae	GenBank	-	AB120187	AB115437

japonica Thunb.				Miikeda et al. (2006)	Miikeda et al. (2006)
Clematis lasiandra Maxim.	Ranunculaceae	GenBank	-	AB120185 Miikeda et al. (2006)	AB115435 Miikeda et al. (2006)
Clematis lasiantha Nutt. ex Torr. & A.Gray	Ranunculaceae	GenBank	-	AB120200 Miikeda et al. (2006)	AB115450 Miikeda et al. (2006)
Clematis ligusticifolia Nutt. ex Torr. & A.Gray	Ranunculaceae	GenBank	-	AB120201 Miikeda et al. (2006)	AB115451 Miikeda et al. (2006)
Clematis nobilis Nakai	Ranunculaceae	GenBank	-	AB120206 Miikeda et al. (2006)	AB115456 Miikeda et al. (2006)
Clematis ochotensis (Pall.) Poir.	Ranunculaceae	GenBank	-	AB120182 Miikeda et al. (2006)	AB115432 Miikeda et al. (2006)
Clematis orientalis L.	Ranunculaceae	GenBank	-	AB120196 Miikeda et al. (2006)	AB115446 Miikeda et al. (2006)
Clematis patens Morr. & Decne.	Ranunculaceae	GenBank	-	AB120184 Miikeda et al. (2006)	AB115434 Miikeda et al. (2006)
Clematis pierotii Miq.	Ranunculaceae	GenBank	-	AB120191 Miikeda et al. (2006)	AB115441 Miikeda et al. (20069
Clematis potaninii Maxim.	Ranunculaceae	GenBank	-	AB120198 Miikeda et al. (2006)	AB115448 Miikeda et al. (2006)
Clematis serratifolia Rehd.	Ranunculaceae	GenBank	-	AB120205 Miikeda et al. (2006)	AB115455 Miikeda et al. (2006)
Clematis stans Sieb. & Zucc.	Ranunculaceae	GenBank	-	AB120188 Miikeda et al. (2006)	AB115438 Miikeda et al. (2006)
Clematis tangutica Korsh.	Ranunculaceae	GenBank	-	AB120195 Miikeda et al. (2006)	AB115445 Miikeda et al. (2006)
Clematis tashiroi Maxim.	Ranunculaceae	GenBank	-	AB120192 Miikeda et al. (2006)	AB115442 Miikeda et al. (2006)
Clematis terniflora DC.	Ranunculaceae	GenBank	-	AB120183 Miikeda et al. (2006)	AB115433 Miikeda et al. (2006)
Clematis texensis Buckley	Ranunculaceae	GenBank	-	AB120197 Miikeda et al. (2006)	AB115447 Miikeda et al. (2006)
Clematis uncinata Champ. var *ovatifolia* (T.Ito) Ohwi	Ranunculaceae	GenBank	-	AB120189 Miikeda et al. (2006)	AB115439 Miikeda et al. (2006)
Clematis villosa DC.	Ranunculaceae	GenBank	-	AB120211 Miikeda et al. (2006)	AB115461 Miikeda et al. (2006)
Clematis vitalba L.	Ranunculaceae	GenBank	-	AB120207 Miikeda et al. (2006)	AB115457 Miikeda et al. (2006)
Clematis williamsii A.Gray	Ranunculaceae	GenBank	-	AB120181 Miikeda et al. (2006)	AB115431 Miikeda et al. (2006)
Naravelia laurifolia Wall. ex Hook.f. & Thomson	Ranunculaceae	GenBank	-	AB120208 Miikeda et al. (2006)	AB115458 Miikeda et al. (2006)
Anemone americana DC.	Ranunculaceae	GenBank	-	AY055386 Schuettpelz et	AY055407 Schuettpelz et

[=*Hepatica americana* (DC.) H. Hara]				al. (2002)	al. (2002)
Anemone antucensis Poeppig	Ranunculaceae	GenBank	-	AY056049 Schuettpelz et al. (2002)	AF311735 Schuetttpelz at al. (2002)
Anemone blanda Schott & Kotschy	Ranunculaceae	GenBank	-	AY055402 Schuettpelz et al. (2002)	AY055422 Schuettpelz et al. (2002)
Anemone caffra (Eckl. & Zeyh.)	Ranunculaceae	GenBank	-	AY055399 Schuettpelz et al. (2002)	AY055420 Schuettpelz et al. (2002)
Anemone canadensis L.	Ranunculaceae	GenBank	-	AY055387 Schuettpelz et al. (2002)	AY055408 Schuettpelz et al. (2002)
Anemone caroliniana Walter	Ranunculaceae	GenBank	-	AY055403 Schuettpelz et al. (2002)	AY055423 Schuettpelz et al. (2002)
Anemone crassifolia Hook.f.	Ranunculaceae	GenBank	-	AY055398 Schuettpelz et al. (2002)	AY055419 Schuettpelz et al. (2002)
Anemone demissa Hook.f. & Thomson	Ranunculaceae	GenBank	-	AY055392 Schuettpelz et al. (2002)	AY055413 Schuettpelz et al. (2002)
Anemone drummondii S.Watson	Ranunculaceae	GenBank	-	AY055404 Schuettpelz et al. (2002)	AY055424 Schuettpelz et al. (2002)
Anemone flaccida F.Schmidt	Ranunculaceae	GenBank	-	AY055391 Schuettpelz et al. (2002)	AY055412 Schuettpelz et al. (2002)
Anemone hupehensis Lemoine	Ranunculaceae	GenBank	-	AY055397 Schuettpelz et al. (2002)	AY055418 Schuettpelz et al. (2002)
Anemone keiskeana Ito	Ranunculaceae	GenBank	-	AY055390 Schuettpelz et al. (2002)	AY055411 Schuettpelz et al. (2002)
Anemone knowltonia Burtt-Davy [=*Knowltonia capensis* (L.) Huth]	Ranunculaceae	GenBank	-	AY055401 Schuettpelz et al. (2002)	AY055421 Schuettpelz et al. (2002)
Anemone moorei Esp.	Ranunculaceae	-	Vilches Alto, Chile	This study	This study
Anemone multifida Poir	Ranunculaceae	GenBank	-	AY055405 Schuettpelz et al. (2002)	AY055425 Schuettpelz et al. (2002)
Anemone narcissiflora L.	Ranunculaceae	GenBank	-	AY055393 Schuettpelz et al. (2002)	AY055414 Schuettpelz et al. (2002)
Anemone obtusiloba D.Don	Ranunculaceae	GenBank	-	AY055394 Schuettpelz et al. (2002)	AY055415 Schuettpelz et al. (2002)
Anemone occidentalis S.Watson [=*Pulsatilla occidentalis* (S.Watson) Freyn]	Ranunculaceae	GenBank	-	AY055400 Schuettpelz et al. (2002)	AY055426 Schuettpelz et al. (2002)
Anemone richardsonii Hook.f.	Ranunculaceae	GenBank	-	AY055388 Schuettpelz et al. (2002)	AY055409 Schuettpelz et al. (2002)
Anemone rivularis Buch.-Ham. ex DC.	Ranunculaceae	GenBank	-	AY055396 Schuettpelz et al. (2002)	AY055417 Schuettpelz et al. (2002)
Anemone tenuicaulis	Ranunculaceae	GenBank	-	AY055389 Schuettpelz et	AY055410 Schuettpelz et

(Cheeseman) Parkin & Sledge				al. (2002)	al. (2002)
Anemone trullifolia Hook.f. & Thomson	Ranunculaceae	GenBank	-	AY055395 Schuettpelz et al. (2002)	AY055416 Schuettpelz et al. (2002)
Hepatica acutiloba DC.	Ranunculaceae	GenBank	-	AM267285 Pfosser et al. (2006)	AM267300 Pfosser et al. (2006)
Hepatica asiatica Nakai	Ranunculaceae	GenBank	-	AM267289 Pfosser et al. (2006)	AM267296 Pfosser et al. (2006)
Hepatica henryi (Oliv.) Steward	Ranunculaceae	GenBank	-	AM267290 Pfosser et al. (2006)	AM267297 Pfosser et al. (2006)
Hepatica insularis Hanst.	Ranunculaceae	GenBank	-	AM267288 Pfosser et al. (2006)	AM267298 Pfosser et al. (2006)
Hepatica maxima Nakai	Ranunculaceae	GenBank	-	AM267282 Pfosser et al. (2006)	AM267295 Pfosser et al. (2006)
Hepatica nobilis Mill.	Ranunculaceae	GenBank	-	AM267286 Pfosser et al. (2006)	AM267294 Pfosser et al. (2006)
Hepatica transsilvanica Fuss.	Ranunculaceae	GenBank	-	AM267283 Pfosser et al. (2006)	AM267299 Pfosser et al. (2006)
Pulsatilla alpina subsp *apiifolia* Nyman	Ranunculaceae	-	Simplonpass, Switzerland	This study	This study
Pulsatilla vulgaris Mill.	Ranunculaceae	A.-M. Barniske 058 (DR)	BG Dresden	This study	This study

DNA isolation, amplification and sequencing

DNA was isolated from fresh or silica gel-dried plant material by using the CTAB-method described in Doyle & Doyle (1990). To yield high amounts of genomic DNA three extractions were carried out following the protocol outlined by Borsch & al. (2003). In cases of suboptimal DNA quality extractions were cleaned using commercially available spin columns (Macherey-Nagel; Düren, Germany). Amplification and sequencing reactions were performed in a T3 Thermocycler or Gradient Thermocycler (Biometra; Göttingen, Germany). The ITS regions were amplified using universal primers ITS4 and ITS5 published by White & al. (1990). PCR amplifications were performed in 50 µl-reactions containing 1U Taq DNA polymerase (SAWADY-*Taq*-DNA-Polymerase, Peqlab; Erlangen, Germany), 1 mM dNTP mix of each 0.25 mM, 1 x taq buffer (Peqlab), 1.25-2.5 mM $MgCL_2$ (Peqlab) and 20 pmol of each amplification primer. To prevent the formation of secondary structures of the nuclear ribosomal DNA Betain (Sigma-Aldrich; Taufkirchen, Germany) in a final concentration of 0.5 M was added. Amplifications were carried out as follows: 94°C for 5 min were followed by 40 cycles of denaturation (1 min, 94°C), primer annealing (1 min, 48°C), extension (45 s, 68°C) and a final extension at 68°C for 7 min. Amplification of the *atpB-rbcL* region was done using the universal primers atpB-rbcLF1 (forward) and atpB-rbcLR (reverse; both chapter 1, compare Table 2). PCR protocols and reaction conditions followed chapter 1. Purification of the amplicons was carried out using the NucleoSpin Extract II kit for cleanup of gel extraction (Macherey-Nagel; Düren, Germany) after running them out on a 1.2 % agarose gel for 2.5 h at 80 V. Direct sequencing was performed using the amplification primers (Table 2) and the DTCS QuickStart Reaction Kit (BeckmannCoulter). Extension products where either run on a BeckmannCoulter CEQ 8000 automated sequencer or sequenced by Macrogen Inc., South Korea (www. macrogen.com). Sequences were edited manually with PhyDE v0.995 (Müller & al., 2005).

Table 2: Primers used for molecular work

Primer name	Sequence	Direction	Reference	Region
ITS5	GGAAGTAAAAGTCGTAACAAGG	F	White et al. (1990)	ITS
ITS4	TCCTCCGCTTATTGATATGC	R	White et al. (1990)	ITS
atpB-rbcLF1	CACTCATRCTACRCTCTAACTC	F	See chapter 1	*atpB-rbcL*
atpB-rbcLR	CACCAGCTTTGAATCCAACACC	R	See chapter 1	*atpB-rbcL*

Alignment, indel coding and phylogenetic analyses

Nucleotide sequences were aligned "by eye" using PhyDE v0.995, based on the rules outlined in Kelchner (2000) and Borsch & al. (2003). Several sequence stretches with unclear primary homology were tagged as "hotspots" (H) and afterwards excluded from phylogenetic analyses. Indel characters were utilized by applying the simple-indel coding method pointed out in Simmons & Ochoterena (2000) via SeqState v1.2 (Müller, 2005). The resulting indel matrix was combined with the nucleotide-sequence matrix and used for parsimony analyses and Bayesian Inference (BI). Most parsimonious trees (MPT) were calculated by using the parsimony ratchet (Nixon, 1999) as implemented in PRAP (Müller, 2004). Ratchet settings were 20 random-addition cycles of 200 ratchet replicates, and upweighting 25 % of the characters. A strict consensus tree was created in cases with multiple MPTs. Nodes were evaluated by bootstrapping in PAUP* version 4.0b10 for Windows (Swofford, 2002) using 1000 replicates.

BI was performed using MrBayes v3.1 published by Ronquist & Huelsenbeck (2003), applying the GTR + Γ + I model for nucleotide sequence data, and the restriction site model ("F81") for the indel matrix. Four runs (1,000,000 generations each) with four chains each were run simultaneously. Chains were sampled every 10th generation. The consensus tree and the posterior probability (PP) of clades were calculated based upon the trees sampled after the burn-in set at 250,000 generations. TreeGraph (Müller & Müller, 2004) was used for drawing trees. Sequence statistics were calculated via SeqState v1.2. Datasets are deposited on the appended CD.

Molecular dating using BEAST

Molecular dating was performed using BEAST v1.4 as published by Drummond & Rambaut (2007), applying relaxed molecular clock models within Bayesian MCMC analyses. The GTR+Γ substitution model was chosen for the nucleotide sequence matrix as well as the uncorrelated longnormal relaxed clock (Drummond & al., 2006). Calibration was done using fossils as well as geological data (Table 3). As illustrated in Figure 4 two external calibration points were chosen, while two nodes within the subtribe Anemoninae were used to calibrate the chronogram. The Markov chain was run with 15,000,000 generations and sampled every 1000th generation. BEAST XML input files were generated via BEAUti, while Tracer was used for analyzing MCMC-log-files created in BEAST. A consensus tree based upon the trees sampled after the burn-in set at

1,500,000 generations was created using TreeAnnotator. FigTree was used for illustrating trees.

Table 3: Fossils and geological events used for calibration. Mya=Million years.

Assignment	Fossil/Event	Node	Structure	Locality	Stratigrafic zone	Age (Myr)	References
Order Ranunculales	*Teixeiraea lusitanica*	A	staminate flower	Vale de Agua locality, Portugal	late Albian-early Albian	112	von Balthazar et al. (2005)
Fam. Menispermaceae	*Prototinomiscium vangerowii*	B	fruit	Klikov-Schichtenfolge, Czeck Rep.	Turonian	91	Knobloch & Mai (1986)
Genus *Anemone* *A. antucensis-A. tenuicaulis*	Shift in magmatism preceding the separation of Marie Byrd Land and Tasmantia	C	-	-	-	100	McLoughlin (2001)
Genus *Anemone*, *A. antucensis-A.tenuicaulis*	Separation Antarctica-Australia (including Tasmania)	C	-	-	-	35	Sanmartin & Ronquist (2004)
Genus *Anemone*, *A. moorei-A.crassifolia*	Separation New Zealand-Antactica	D	-	-	-	84	Lawver et al. (1992); McLoughlin (2001; review)

3.4 Results & discussion

3.4.1 Sequence variability

Large parts of the *atpB-rbcL* IGS from the chloroplast as well as the nuclear ribosomal ITS region were analysed. The length of the plastid region under survey ranges from 556 to 778 nt, whereas ITS1 & 2 are clearly shorter. ITS1 extends from 151 to 289 nt, while ITS2 is displaying a length between 150 and 238 nt. Both regions differ considerably in transition/transversion ratio and GC-content (see Table 4). Individual parts of the ITS partition display a GC-content between 53.9 and 64.6 %, whereas the plastid region exhibits a lower value of only 27.5 %. In total 393 informative characters (due to substitutions only) were provided by the combined dataset. All non-coding regions supplied an equal amount of parsimony informative characters (Table 4). A comparison of the percentage of informative sites revealed that the *atpB-rbcL* partition, while displaying the largest amount of aligned sequence characters, only exhibits 8.3 % of parsimony informative positions, whereas ITS1 and ITS2 feature values of 32.2 or 24.9 %, respectively.

Table 4: Variation and relative contribution of the genomic regions studied. Number and quality of characters, indels coded and GC content, as well as transition/transversion ratio are calculated with mutational hotspots excluded. SD=Standard deviation, No. char.=Number of characters, var.-char.=variable characters, inf.-char.=informative characters, Ti/Tv ratio=transition/transversion ratio.

Region	mean sequence lenght (bp)	SD	mean sequence length excl. hotspots (bp)	SD	No. char.	var. char. [%]	inf. char. [%]	No. inf. char.	No. of indels coded	GC-content [%]	Ti/Tv ratio
ITS1	186	22	181	14	395	43.8	32.2	127	96	57.8	1.535
5.8S	163	1	163	1	166	18.7	11.4	19	9	53.9	2.818
ITS2	205	9	205	9	470	36.4	24.9	117	90	64.6	1.988
atpB-rbcL spacer	725	42	708	41	1558	16.6	8.3	129	227	27.5	1.003

In total four mutational hotspots were marked within the combined data matrix. One extended part of the nuclear ribosomal region was determined as hotspot and subsequently excluded from tree inference. It comprises a non-alignable part of ITS1 inside the outgroup species with a length range from 42 to 112 nt. Three mutational hotspots were identified within the *atpB-rbcL* IGS, all of them referring to microsatellites of different extent (compare Table 5). H2 consists of poly A/T stretches with a sequence length up to 34 nt in *Anemone knowltonia*. The remaining mutational hotspots are due to short poly-mononucleotide stretches of more than four nucleotides showing a length variation of at least two nts. According to the rules outlined in Olsson et al. (2009) these sequence parts should be excluded from analyses to prevent the involvement of spurious indel information.

Table 5: Hotspot (H) positions in alignment and region.

No. hotspot	Position in alignment	Region
H1	100 – 370	ITS1
H2	1489 – 1549	*atpB-rbcL* spacer
H3	2079 – 2087	*atpB-rbcL* spacer
H4	2680 – 2691	*atpB-rbcL* spacer

422 indels were coded and included into analysis. The *atpB-rbcL* partition provided a set of 227 coded characters, while ITS1 and ITS2 provided 96 or 90 coded insertions and deletions, respectively. Some of them were identified as autapomorphic, such as indel 261 within the *atpB-rbcL* IGS (alignment position 1476 – 1506) which is a deletion of 31 nucleotides unique to *Anemone richardsonii*. Other length mutations are synapomorphic for specific clades. One example is a simple sequence repeat of eight nucleotides (*atpB-rbcL*; alignment position 1819 – 1826) shared by all taxa belonging to the traditional genus *Hepatica*.

3.4.2 Phylogeny of the tribe Anemoneae

The combined data matrix (*atpB-rbcL* + ITS), excluding mutational hotspots supplied 2589 characters of which 634 were variable and 393 were parsimony informative. An indel matrix of 422 binary indel characters was added to the dataset. Maximum parsimony analysis resulted in 132 most parsimonious trees of 2108 steps (CI = 0.636, RI = 0.802).

MP as well as BI revealed a highly supported sister group relationship between Clematidinae and Anemoninae (Figure 1 and 2). The Clematidinae are shown as being monophyletic with moderate statistical support in parsimony analyses (BS 88/83, as in the following the first value refers to support obtained with the binary indel matrix included in the analyses) while in Bayesian Inference statistical support was raised to significance (PP 0.96/0.97). Topologies gained through maximum parsimony and Bayesian analyses differ in showing *Clematis ochotensis* being sister to a main clade within Clematidinae or a sister group relationship of two grades, respectively. However statistical support was absent or stayed at a weak to a moderate level in BI (PP 0.65/0.83 or 0.69/-, respectively). The Anemoninae are identified as forming a monophyletic group (BS 99/97, PP 1.0). They split up into two distinct highly supported branches (BS 100/95 or 100/97, respectively, PP 1.0 or 1.0/0.98, respectively). *Pulsatilla* is found within lineage I (BS 77/85, PP 0.99/1.0) as well as *Knowltonia capensis* (=*Anemone knowltonia*), being sister to *Anemone caffra* (BS 98/99, PP 1.0). Lineage II shows *Hepatica* as sister to the remaining species of *Anemone* with high statistic values in all approaches (BS 100/97, PP1.0/0.98).

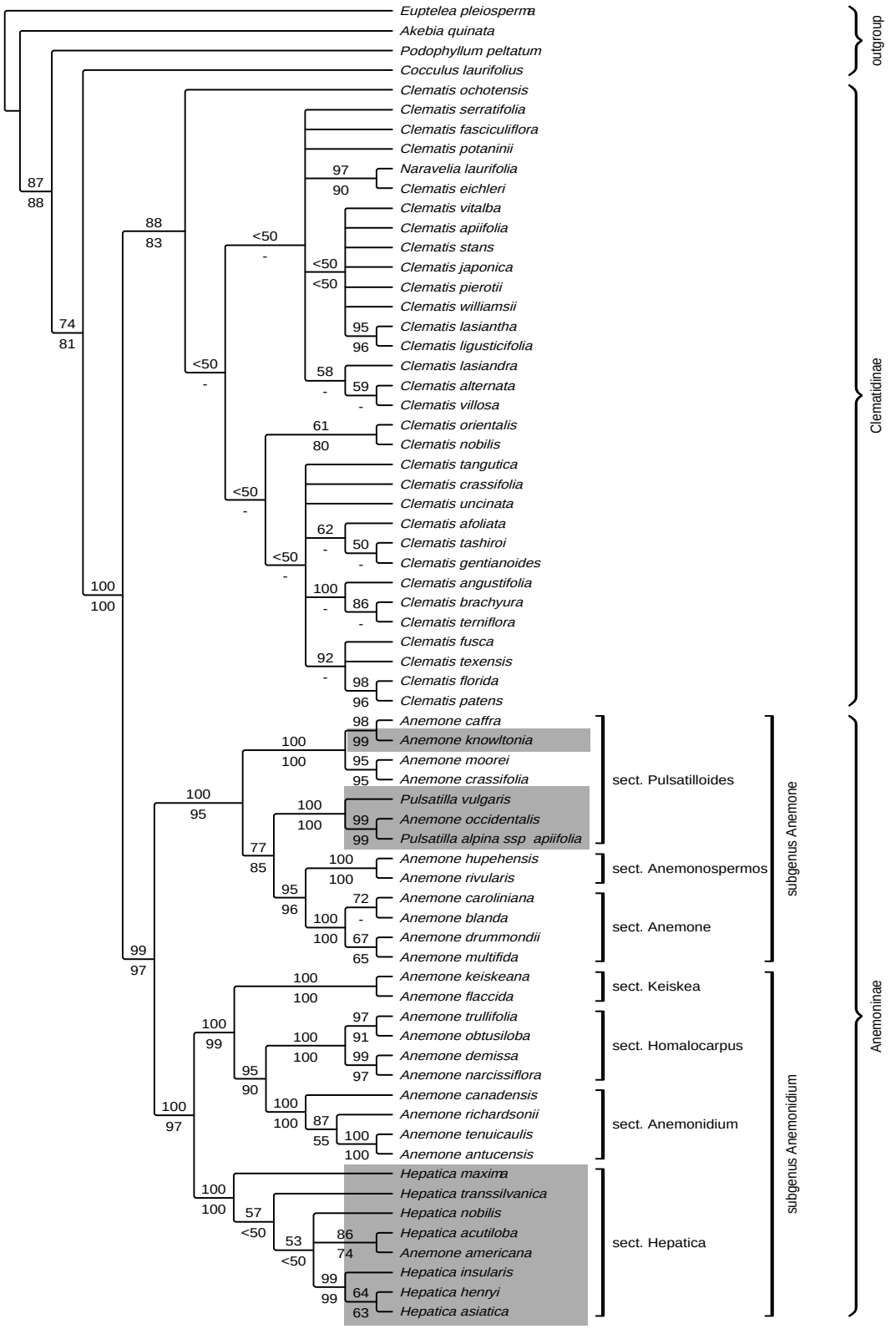


Figure 1: Maximum parsimony strict consensus tree based on the combined ITS+*atpB-rbcL* matrix, including substitutions and indel information. Values above and below branches are Bootstrap percentages, referring to substitions plus indels or substitutions only, respectively. Subgenera/sections given after brackets bear on the informal classification presented by Hoot et al. (1994). Members of the traditional genera *Knowltonia*, *Pulsatilla* and *Hepatica* are greyed out (top down). sect.=section.

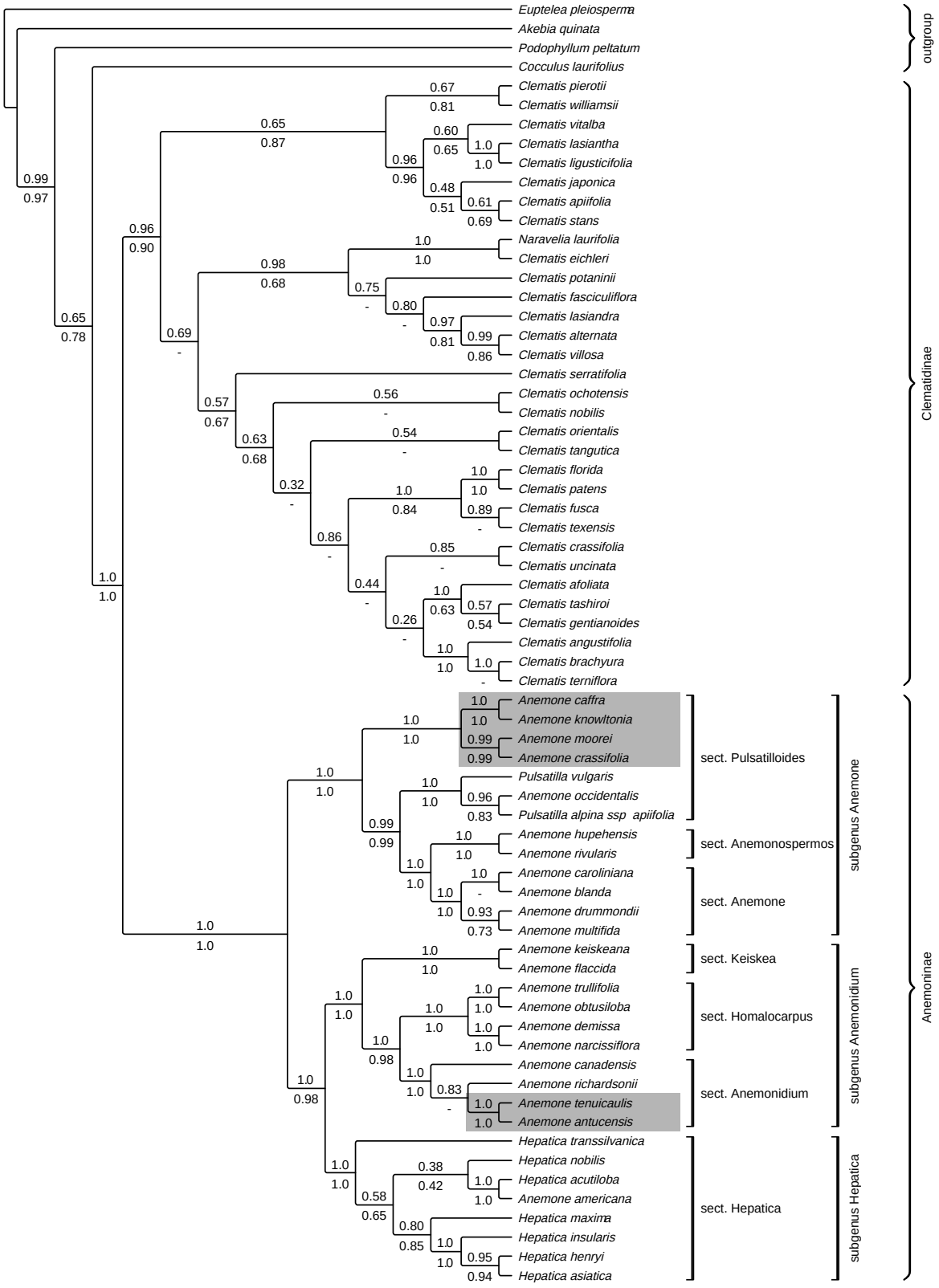


Figure 2: Bayesian tree on the basis of the combined data matrix of both genomic regions under study. Posterior Probabilities are depicted above (substitutions plus indels) and below (substitutions only) branches. Subgenera/sections after brackets refer to the revised informal classification. Clades showing a transoceanic disjunction are greyed out. sect=section.

***Phylogentic relationships among* Anemoninae *and* Clematidinae**

A central goal of the study was to clarify the phylogenetic relationships among the subtribes Anemoninea and Clematidinae. The genus *Clematis* has always been considered as a close relative of the members of the Anemoninae, from which it is mainly distinguished by its woody stems and its opposite pairs of leaves (Grey-Wilson, 2000). Investigations of the ovule morphogenesis of Wang & Ren (2007) supported the close affinity of the genera *Anemone* and *Clematis*. Phylogenetic studies of Johansson & Jansen (1993) and Johansson (1995) using restriction site variation of chloroplast DNA and including four representatives of each subtribe clearly showed both as being members of one highly supported clade. The monophyly of the Anemoneae was also demonstrated by studies on the basis of sequence data of the *adh* gene (Kosuge & al., 1995), a combined analysis of three genes (*atpB*, *rbcL*, 18S; Hoot, 1995a) and the survey presented in chapter 2 using 6 regions from the large single copy region of the chloroplast genome. These findings are in congruence with the results of the presented comprehensive study. Based on molecular sequence data of the nuclear ribosomal ITS1&2 and the plastid *atpB-rbcL* IGS the monophyly of the tribe gained maximum statistical support in all analyses.

In contrast several phylogenetic analyses differ in showing divergent branching orders within the clade. Johansson & Jansen (1993) as well as Johansson (1995) were not able to fully settle the phylogenetic relations among the two subtribes. A scenario showing a clade of *Anemone*, *Pulsatilla*, *Knowltonia* and *Hepatica*, sharing a sistergroup relationship with the Clematidinae gained no statistical support, while the hypothesis assuming Anemoninae to be paraphyletic in relation to the *Clematis*-group was only weakly supported in parsimony analysis. Based on sequence data of the nuclear *adh*-gene, Kosuge & al. (1995) showed a sister group relationship between the two subtribes, which obtained a moderate bootstrap value (BS 89). This result seems to be partly due to the limited taxon sampling. As shown by Hoot & al. (1994) and Hoot (1995b) the Anemoninae split up into two distinct lineages, one consisting of the majority of the *Anemone*-species, *Pulsatilla* and *Knowltonia* (chromosome base number x = 8) and another including *Hepatica* and different taxa of the genus *Anemone* (x = 7). No representative of the latter was included into the analyses of Kosuge & al. (1995). Inclosing members of both lineages Hoot & Palmer (1994) gained a topology presenting a sistergroup relationship between Anemoninae and Clematidinae. Nevertheless, the monophyly of the Anemoninae was only moderately supported (BS 70). Through a thorough taxon-sampling and the combination of plastid and nuclear sequence data it was possible to raise the statistic

values for this scenario to the significance level. Therefore the hypothesis of the Anemoninae being paraphyletic established in a recent comprehensive study of the Ranunculales carried out in chapter 2 on the basis of fast-evolving and non-coding plastid regions including indel information seems to be unlikely. These results seem to partly depend on the limited taxon sampling being not representative of the species diversity within the tribe. Anyway, individual parsimony analyses of the two datasets used in this presented study clearly demonstrated the phylogenetic signal to be mostly coming from the nuclear ribosomal partition. Topologies on the basis of sequence data of the *atpB-rbcL* spacer either showed Clematidinae in a polytomy with the two distinct lineages inside the *Anemone*-complex or, by the addition of indel information, displayed a sister group relationship of Clematidinae and Anemoninae without bootstrap support. However, the results of this recent survey obviously confirm the division of the tribe Anemoneae into the subtribes Anemoninae and Clematidinae. This classification is further corroborated by the strikingly differing molecular rates and estimated node ages that are generally lower within Clematidinae (compare Figure 3 and Figure 4).

Figure 3: Bayesian phylogram based on the combined ITS+*atpB-rbcL* data matrix, with substitutions and coded indel characters included into analyses.

***Phylogenetic relationships inside the subtribe* Anemoninae**

A number of studies based on molecular data have provided important information on the phylogeny and evolution of the Anemoninae. Investigations by Hoot & al. (1994) and Hoot (1995b) on the basis of three independent datasets (chloroplast DNA restriction sites, nuclear ribosomal DNA restriction fragments and morphological/cytological variation) already indicated the subsumption of *Pulsatilla*, *Knowltonia* and *Hepatica* within *Anemone* s.l. as well as the formation of two distinct clades inside the *Anemone*-complex. These results were clearly corroborated by analyses using sequence data (Ehrendorfer & Samuel, 2001; Schuettpelz & al., 2002). Ehrendorfer & Samuel (2001) featured a tree containing 21 taxa and one hybrid of the subtribe Anemoninae using sequence information from the plastid *atpB-rbcL* spacer-region only. Despite clade I and II were recognized in parsimony analysis, bootstrap support was lacking. *Pulsatilla grandis* was depicted as sister to the remainder of clade I, whereas the two species of *Hepatica* were deeply nested within clade II. Schuettpelz & al. (2002) presented a deviating taxon-sampling and included *Knowltonia capensis* (= *Anemone knowltonia*) into their analysis of the combined *atpB-rbcL* IGS and ITS data. The genus *Pulsatilla* as well as *Knowltonia* were demonstrated to have their origin within the well supported lineage I (= subgenus *Anemone* sensu Hoot & al., 1994). Lineage II, which was wrapped up as subgenus *Anemonidium* (Hoot & al., 1994), gained no statistical support. Nevertheless the genus *Hepatica* was clearly shown to branch first inside this clade (BS 99). The present study is mainly based on the molecular data generated by Schuettpelz & al. (2002). By combining the molecular data of both partitions with an indel matrix and including an increased taxon-sampling it was possible to raise statistical support of both lineages inside the subtribe to an almost maximum in maximum parsimony as well as in Bayesian Inference. In lineage I a clade containing four species from the southern hemisphere is branching first, followed by a monophyletic group consisting of three members of the genus *Pulsatilla*, while Schuettpelz & al. (2002) presented *Pulsatilla (occidentalis)*+the southern hemisphere species as being sister to the remaining representatives of the lineage. However, the genus *Pulsatilla* obviously belongs to clade I (chromosome base number of x = 8), just as *Knowltonia*. The topology of lineage II (x = 7) is identical to that derived by Schuettpelz & al. (2002). Eight species of *Hepatica* were included into the survey. They are shown to form a highly supported clade in all approaches, being sister to the residual species of the subgenus *Anemonidium*.

Interestingly, branch-lengths differ considerably inside the genus, as indicated by the Bayesian phylogram (Figure 3). Branches within the *Hepatica*-clade (= section *Hepatica* sensu Hoot & al., 1994) are shown to be distinctly shorter than those of the remainder of the genus *Anemone*, while molecular rates within the *Pulsatilla*-clade (= *Pulsatilla* group sensu Hoot & al., 1994) do not significantly differ. A similar picture emerged by applying the data matrix to a molecular dating approach. As shown in Figure 4 the *Hepatica*-clade represents a very distinct and young lineage as compared to the remaining Anemoninae, while displaying a rapid radiation. The species belonging to section *Hepatica* are restricted to the temperate zone of the northern hemisphere with a great variation in Eastern Asia (Hultén & Fries, 1986; Tamura, 1995) mainly growing in deciduous forests or bushes. Furthermore they differ in floral as well as fruit morphology from the remaining members of the subgenus *Anemonidium* because of the involucre close to sepals, their bract-like involucral leaves and long stalked achenes (compare Hoot & al., 1994). Recapitulatory, the *Hepatica*-clade seems to be a very distinct lineage within the subgenus *Anemonidium*. This fact should be taken into consideration when thinking about a classification of *Anemone* at the subgenus level, as done by Hoot & al. (1994) on the basis of an anlysis combining morphological with molecular data for the first time. Therefore their preliminary classification of the *Anemone*-complex, presenting the two subgenera *Anemone* and *Anemonidium* mainly on the basis of their base chromosome number, should be complemented by lifting the section *Hepatica* to the subgenus level. This leads to the following revised informal classification of the genus *Anemone*:

Subgen. *Anemone*
Sect. *Anemonospermos* DC.
Rivularis group
Vitifolia group
Sect. *Pulsatilloides* DC.
Crassifolia group
Caffra group
Knowltonia group
Pulsatilla group
Sect. *Anemone*
Coronaria group
Baldensis group
Nemorosa group
Multifida group
Subgen. *Anemonidium* (Spach) Juz.
Sect. *Anemonidium* Spach
Sect. *Keiskea* Tamura
Sect. *Homalocarpus* DC.
Narcissiflora group
Obtusiloba group
Subgen. *Hepatica* (Miller) Peterm.
Sect. *Hepatica* Spreng.

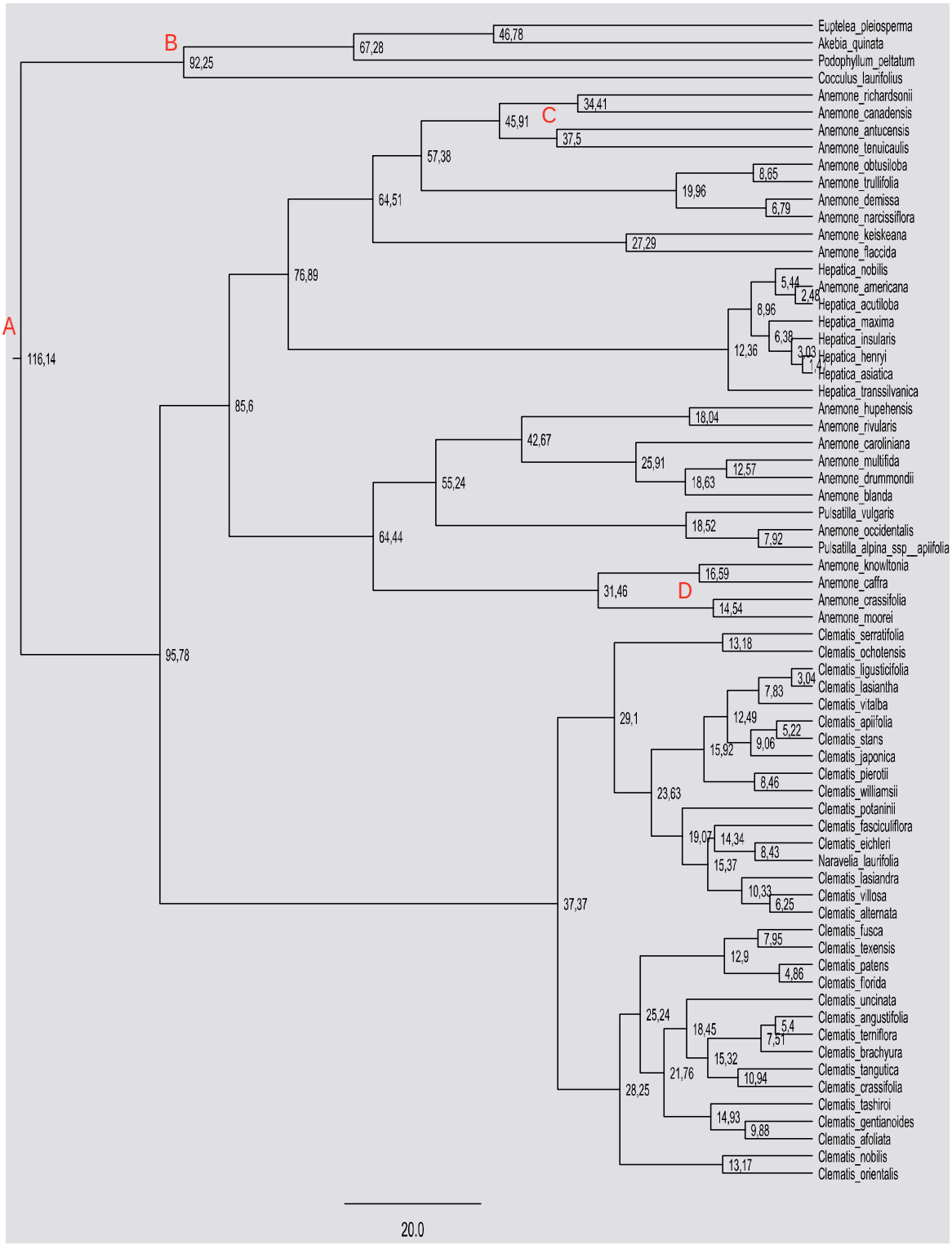


Figure 4: Consensus tree generated through BEAST analyses. Estimated ages are given on the right, letters (A-C) indicating calibration points on the left of the respective nodes.

3.4.3 Phytogeographical aspects within the subtribe Anemoninae

One of the remarkable results of the study by Schuettpelz & al. (2002) was the close relationship of the South American *Anemone antucensis* and *A. tenuicaulis* from New Zealand, since the latter was assumed to be closely related to *A. crassifolia*, a Tasmanian species. Furthermore the aforementioned *A. crassifolia* was shown as being sister to a clade consisting of two taxa originating from South Africa (*A. knowltonia* = Knowltonia capensis and *A. caffra*). Adding sequence data of *A. moorei* resulted in the recognition of a new phylogenetic hypothesis within subgenus Anemone, assuming a close affinity of this South American species and *A. crassifolia* (Figure 1 & 2), while being sister to the South African species. Thus, for an explanation of the present day distribution of this clade, possible links between South America and Tasmania have to be considered similar to the *A. antucensis*+*A. tenuicaulis* clade. Furthermore a linkage to South Africa has to be explored. Schuettpelz & al. (2002) discussed a vicariance model as being a more likely explanation for the distribution pattern of the genus than long-distance dispersal. The latter seemed to be rather unlikely due to fruit morphology and restricted geographic ranges of the most species in question. This could be also true concerning *A. moorei*, which is characterized by glabrous achenes with a short style. Furthermore it is endemic to the Chilean province of Talca (Ruiz, 2001; A. Stoll pers. comm.). Moreover the distribution pattern of the *Anemone* clades in question is congruent with the sequential break-up of Africa, southern South America and Australia from the Gondwanan landmass as illustrated in the literature (McLoughlin, 2001; Sanmartín & Ronquist, 2004). Biogeographic patterns that are consistent with the break-up history of Gondwana are known from several plant groups, such as the genus *Gunnera* (Wanntorp & Wanntorp, 2003) or certain lineages inside the core monocots (Bremer & Janssen, 2006). To test a possible vicariance scenario within the genus *Anemone* a molecular dating approach was carried out using relaxed molecular clock models (Drummond & al., 2006) as implemented in BEAST (Drummond & Rambaut, 2007). The clade consisting of *A. moori* and *A. crassifolia* as well as the *A. antucensis*+ *A. tenuicaulis* clade belonging to subgenus *Anemonidium* of the genus were included into the calibration (compare Table 3 and Figure 4). As illustrated in Figure 4, no one of the three nodes associated with transoceanic disjunctions gained an age estimate consistent with a Gondwanan vicariance model. The node, representing the clade consisting of *A. crassifolia* and *A. moori*, was dated 14.5 Myr, post-dating the ultimate isolation of Australia (including Tasmania) from

Antarctica (and thus South America) during the late Eocene (35 Myr, McLoughlin, 2001). Furthermore, the split between the African species and the well supported Tasmanian-Chilean-clade, dated at about 31.5 Myr, clearly post-date the separation of Africa from the rest of Gondwana at about 105 Myr (McLoughlin, 2001). Similarly the node representing the split between *A. tenuicaulis* originating from New Zealand and the South American *A. antucensis* (within subgenus *Anemonidium*) is post-dating the well documented separation of New Zealand from Antarctica at 84 Myr (McLoughlin, 2001). These findings suggest that long-distance dispersal may be the cause of the disjuct distribution in the genus *Anemone*. Similar findings were obtained concerning Antherospermataceae (Renner & al., 2000) and Myristicaceae (Doyle & al., 2004), as well as Proteaceae (Baker & al., 2007), leading to the conclusion of the distribution patterns within the families to be (partly) originating from transoceanic dispersal. Summarizing all considerations and results, future work is needed for comprehensively understanding underlying processes leading to this biogeographical pattern.

References

Airy Shaw HK: *J.C. Willis' a dictionary of the flowering plants and ferns.* Cambridge: Cambridge University Press; 1973.

APG II [Anderberg AA, Bremer B, Bremer K, Chase MW, Fay MF, Goldblatt P, Judd WS, Källersjö M, Kårehed J, Kron KA, Lundberg J, Nickrent DL, Olmstead RG, Oxelman B, Pires JC, Reveal JL, Rodman JE, Rudall PJ, Soltis DE, Soltis PS, Stevens PF, Savolainen V, Sytsma KJ, van der Bank M, Wurdack K, Xiang JQ-Y, Zmarzty S]: **An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II.** *Bot J Linn Soc* 2003, **141:**399-436.

Baker NP, Weston PH, Rutschmann F, Sauquet H: **Molecular dating of the "Gondwanan" plant family Proteaceae is only partially congruent with the timing of the break-up of Gondwana.** *J Biogeogr* 2007, **34:**2012–2027.

Barthlott W, Theisen I: **Epicuticular wax ultrastructure and classification of Ranunculiflorae.** *Pl Syst Evol* 1995, (Suppl 9):39-45.

Barthlott W, Theisen I, Borsch T, Neinhuis C: **Epicuticular waxes and vascular plant systematics: integrating micromorphological and chemical data.** In *Deep Morphology: Toward a Renaissance of morphology in Plant Systematics*. Edited by Stuessy TF, Mayer V, Hörandl E. Ruggell, Liechtenstein: A.R.G. Gantner Verlag; 2003:189-206.

Behnke H-D: **Sieve-element plastids, phloem proteins, and evolution of the Ranunculanae.** *Pl Syst Evol* 1995, (Suppl 9):25-37.

Borsch T, Hilu W, Quandt D, Wilde V, Neinhuis C, Barthlott W: **Non-coding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms.** *J Evol Biol* 2003, **16:**558-576.

Borsch T, Löhne C, Müller K, Hilu KW, Wanke S, Worberg A, Barthlott W, Neinhuis C, Quandt D: **Towards understanding basal angiosperm diversification: recent insights using rapidly evolving genomic regions.** *Nova Acta Leopoldina NF* 2005, **92:**85-110.

Borsch T, Quandt, D**: Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA.** *Plant Syst Evol* 2009, **282:**169-199.

Bremer K, Janssen T: **Gondwanan origin of major monocot groups inferred from-dispersal-vicariance analysis.** *Aliso* 2006, **22:**22–27.

Brückner C: **Comparative seed structure in the Ranunculiflorae.** *Pl Syst Evol* 1995, (Suppl 9): 83-84.

Buchheim G: **Reihe Ranunculales**. In *A. Engler's Syllabus der Pflanzenfamilien. Volume 2.* Edited by Melchior H. Berlin: Gebrüder Borntraeger; 1964:131-147.

Burleigh JG, Hilu KW, Soltis DE: **Inferring phylogenies with incomplete data sets: a 5-gene 567-taxon analysis of angiosperms.** *BMC Evol Biol* 2009, **9:**61.

Campagna ML, Downie SR: **The intron in chloroplast gene *rpl16* is missing from the flowering plant families Geraniaceae, Goodeniaceae, and Plumbaginaceae.** *Transactions of the Illinois State Academy of Science* 1998, **91:**1-11.

Carlquist S: **Wood anatomy of Ranunculiflorae: a summary.** *Pl Syst Evol* 1995, (Suppl 9):11-24.

Cech TR, Damberger SH, Gutell RR: **Representation of the secondary and tertiary structure of group I introns.** *Struct Biol* 1994, **1:** 273–280.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hillis HG, Qui GYY, Kron KA, Rettig JH, Conti E, Palmer J, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GHJ,

Graham SW, Barret SCH, Dayanandan S, Albert VA: **Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*.** *Ann Mo Bot Gard* 1993, **80:**528-580.

Chen-Yih W, Kubitzki K: **Lardizabalaceae.** In *The families and genera of vascular plants*. *Volume 2.* Edited by Kubitzki K, Rohwer JG, Bittrich V. Heidelberg: Springer-Verlag; 1993:361-365.

Cronquist A: *An integrated system of classification of flowering plants.* New York: Columbia University Press; 1981.

Cronquist A: *The evolution and classification of flowering plants.* Bronx, New York: The New York Botanical Garden; 1988.

Dahl AE: **Taxonomic and morphological studies in *Hypecoum* sect. *Hypecoum* (Papaveraceae).** *Pl Syst Evol* 1989, **163:**227-280.

Dahl AE: **Infrageneric devision of the genus *Hypecoum* L. (Papaveraceae).** *Nordic J Bot* 1990, **10:**129-140.

Dahl AE: **Artificial crossing experiments within *Hypecoum* sect. *Hypecoum* (Papaveraceae).** *Nordic J Bot* 1992, **12:**13-29.

Damerval C, Nadot S: **Evolution of perianth and stamen characteristics with respect to floral symmetry in Ranunculales.** *Ann Bot* 2007, **99:**631-640.

Donoghue MJ, Doyle JA: 1989. **Phylogenetic analysis of angiosperms and the relationships of the Hamamelidae.** In *Evolution, Systematics and Phylogeny of the Hamamelidae.* Edited by Crane PR, Blackmore S. Oxford: Clarendon Press; 1989:17-45.

Downie SR, Katz-Downie D, Watson MF: **A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rpl16* and *rpoc1* intron sequences: towards a suprageneric classification of subfamily Apioideae.** *Am J Bot* 2000, **87:**273-292.

Doyle JJ, Doyle JL: **Isolation of plant DNA from fresh tissue.** *Focus* 1990, **12:**13-15.

Doyle JA, Endress PK: **Morphological phylogenetic analysis of basal angiosperms: comparison and combination with molecular data.** *Int J Plant Sci* 2000, **161** (Suppl. 6):121-153.

Doyle JA, Sauquet H, Scharaschkin T, Le Thomas A: **Phylogeny, molecular and fossil dating, and biogeographic history of Annonaceae and Myristicaceae (Magnoliales).** *Int J Plant Sci* 2004, **165:**55-67.

Doyle J, Hotton CL: **Diversification of early angiosperm pollen in a cladistic context.** In *Pollen and spores*. Edited by Blackmore S, Barnes RW. Oxford: Clarendon Press; 1991: 169-195.

Drinnan AN, Crane PR, Hoot SB: **Patterns of the floral evolution in the early diversification of non-magnoliid dicotyledons (eudicots).** *Pl Syst Evol* 1994, (Suppl 8):93-122.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A**: Relaxed Phylogenetics and Dating with Confidence.** *PLoS Biology* 2006, **4:**e88.

Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evolutionary Biology* 2007, **7**:214.

Ehrendorfer F, Samuel R: **Contributions to a molecular phylogeny and systematics of *Anemone* and related genera (Ranunculaceae – Anemoninae).** *Acta Phytotax Sin* 2001, **39:** 293-307.

Endress PK: **Floral structure and evolution in Ranunculanae.** *Pl Syst Evol* 1995, (Suppl 9):47-61.

Felsenstein J: **Confidence limits on phylogenies: An approach using the bootstrap.** *Evolution* 1985, **39:**783-791.

Golenberg EM, Clegg MT, Durbin ML, Doebly J, Ma DP: **Evolution of a non-coding region of the chloroplast genome.** *Mol Phylogenet Evol* 1993, **2:**52-64.

Graham SW, Reeves PA, Burns ACE, Olmstead RG: **Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference.** *Int J Plant Sci* 2000, **161:**83-96.

Gregory WC: **Phylogenetic and cytological studies in Ranunculaceae. Juss.** *Trans Amer Phil Soc* 1941, **31:**441-520.

Grey-Wilson C: *Clematis. The genus.* Portland, Oregon: Timber Press; 2000.

Hausner G, Olson R, Simon D, Johnson I, Sanders ER, Karol KG, McCourt RM, Zimmerly S: **Origin an d evolution of the chloroplast *trnK* (*matK*) intron: a model for evolution of group II intron RNA structures.** *Mol Biol Evol* 2006, **23:**380-391.

Hernández-Maqueda R, Quandt D, Werner O, Muñoz J: **Phylogenetic relationships and generic classification of the Grimmiaceae.** *Molec Phyl Evol* 2008, **46:**863-877.

Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW: **Angiosperm phylogeny based on *matK* sequence information.** *Am J Bot* 2003, **90:**1758-1776.

Himmelbaur W: **Die Berberidaceen und ihre Stellung im System.** *Denkschr Kaiserl Akad Wiss Math-Naturwiss Kl* 1913, **89:**733-796.

Hoot SB: **Phylogeny of the Ranunculaceae based on epidermal microcharacters and macromorphology.** *Syst Bot* 1991, **16:**741-755.

Hoot SB, Palmer JD: **Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera.** *J Mol Evol* 1994, **38:**274-281.

Hoot SB, Reznicek AA, Palmer JD: **Phylogenetic relationships in Anemone (Ranunculaceae) based on morphology and chloroplast DNA.** *Syst Bot* 1994, **19:**169-200.

Hoot SB: **Phylogeny of the Ranunculaceae based on preliminary *atpB, rbcL* and 18S nuclear ribosomal DNA sequence data.** *Pl Syst Evol* 1995a, (Suppl 9):241-251.

Hoot SB: **Phylogenetic relationships in *Anemone* (Ranunculaceae) based on DNA restriction site variation and morphology.** *Pl Syst Evol* 1995b, (Suppl 9):295-300.

Hoot SB, Crane PR: **Inter-familial relationships in the Ranunculidae based on molecular systematics.** *Pl Syst Evol* 1995, (Suppl 9):119-131.

Hoot SB, Culham A, Crane PR: **The utility of *atpB* gene sequences in resolving phylogenetic relationships: comparison with *rbcL* and 18S ribosomal DNA sequences in the Lardizabalaceae.** *Ann Missouri Bot Gard* 1995a, **82:**194-207.

Hoot SB, Culham A, Crane PR: **Phylogenetic relationships of the Lardizabalaceae and Sargentodoxaceae: chloroplast and nuclear DNA sequence evidence.** *Pl Syst Evol* 1995b, (Suppl 9):195-199.

Hoot SB, Kadereit JW, Blattner FR, Jork KB, Schwarzbach AE, Crane PR: **Data congruence and phylogeny of the Papaveraceae s.l. based on four data sets: *atpB* and *rbcL* Sequences, *trnK* restriction sites, and morphological characters.** *Syst Bot* 1997, **22:**575-590.

Hoot SB, Douglas AW: **Phylogeny of the Proteaceae based on *atpB* and *atpB-rbcL* intergenic spacer region sequences.** *Aust J Syst Bot* 1998, **11:**301-320.

Hoot SB, Magallón S, Crane PR: **Phylogeny of basal eudicots based on three molecular data sets: *atpB, rbcL,* and 18S nuclear ribosomal DNA sequences.** *Ann Mo Bot Gard* 1999, **86:**1-32.

Hoot SB, Zautke H, Harris DJ, Crane PR, Neves SS: **Phylogenenetic patterns in Menispermaceae based on multiple chloroplast sequence data.** *Syst Bot* 2009, **34:**44-56.

Hultén E, Fries M, 1986: *Atlas of North European vascular plants. North of the tropic of cancer. Volume 1.* Königstein, Germany: Koeltz Scientific Books; 1986.

Isnard S, Speck T, Rowe NP: **Mechanical architecture and development in *Clematis*: implications for canalised evolution of growth forms.** *New Phytologist* 2003, **158:**543-559.

Janczewski ME de: **Études morphologique sur le genre *Anemone* L.** *Revue Génerale de Botanique* 1892, 4:241-258.

Jensen U: **Secondary compounds of the Ranunculiflorae.** *Pl Syst Evol* 1995, (Suppl 9):85-97.

Jensen U, Hoot SB, Johansson JT, Kosuge K: **Systematics and phylogeny of the Ranunculaceae – a revised family concept on the basis of molecular data.** *Pl Syst Evol* 1995, (Suppl 9):273-280.

Johansson JT, Jansen RK: **Chloroplast DNA variation and phylogeny of the Ranunculaceae.** *Pl Syst Evol* 1993, **187:**29-49.

Johansson JT: **A revised chloroplast DNA phylogeny of the Ranunculaceae.** *Pl Syst Evol* 1995, (Suppl 9):253-261.

Johnson LA, Soltis DE: **Phylogenetic inference in Saxifragaceae s.str. and *Gilia* (Polemoniaceae) using *matK* sequences.** *Ann Mo Bot Gard* 1995, **82:**149-175.

Kadereit JW: **Papaveraceae.** In *The families and genera of vascular plants. Volume 2.* Edited by Kubitzki K, Rohwer JG, Bittrich V. Heidelberg : Springer-Verlag; 1993:494-506.

Kadereit JW, Blattner FR, Jork KB, Schwarzbach A: **Phylogenetic analysis of the Papaveraceae s.l. (incl. Fumariaceae, Hypecoaceae, and *Pteridophyllum*) based on morphological characters.** *Bot Jahrb Syst* 1994, **116:**361-390.

Kadereit JW, Blattner FR, Jork KB, Schwarzbach A: **The phylogeny of the Papaveraceae sensu lato: morphological, geographical and ecological implications.** *Pl Syst Evol* 1995, (Suppl 9):133-145.

Kanno A, Hirai A: **A transcription map of the chloroplast genome from rice (*Oryza sativa*).** *Curr Genet* 1993, **23:**166-174.

Kelchner SA, Wendel JF: **Hairpins create minute inversions in non-coding regions of chloroplast DNA.** *Curr Genet* 1996, **30:**259-262.

Kelchner SA, Clark, LG: **Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in Chusquea and the Bambusoideae (Poaceae).** *Mol Phylogenet Evol* 1997, **8:**385-397.

Kelchner SA: **The evolution of non-coding chloroplast DNA and its application in plant systematics.** *Ann Mo Bot Gard* 2000, **87:**482-498.

Kelchner SA: **Group II introns as phylogenetic tools: structure, function, and evolutionary constraints.** *Am J Bot* 2002, **89:**1651-1669.

Kelchner SA, Thomas MA: **Model use in phylogenetics: nine key questions.** *Trends Ecol Evol* 2007, **22:**87-94.

Kessler PJA: **Menispermaceae.** In *The families and genera of vascular plants. Volume 2.* Edited by Kubitzki K, Rohwer JG, Bittrich V. Heidelberg: Springer-Verlag; 1993:402-418.

Kim S, Soltis DE, Soltis PS, Zanis MJ, Suh Y: **Phylogenetic relationships among early-diverging eudicots based on four genes: were the eudicots ancestrally woody?** *Mol Phylogenet Evol* 2004, **31:**16-30.

Kim Y-D, Jansen RK: **Chloroplast DNA restriction site variation and phylogeny of the Berberidaceae.** *Am J Bot* 1998, **85:**1766-1778.

Kim Y-D, Kim S-H, Kim CH, Jansen RK: **Phylogeny of Berberidaceae based on sequences of the chloroplast gene *ndhF*.** *Biochemical Systematics and Ecology* 2004, **32:**291-301.

Kishino H, Hasegawa M: **Evaluation of the maximum-likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea.** *J Mol Evol* 1989, **29:**170-179.

Knobloch E, Mai DH: **Monographie der Früchte und Samen in der Kreide von Mitteleuropa.** *Rozpravy ústredního ústavu geologickénho, Praha* 1986, **47:**1-219.

Kofuji R, Ueda K, Yamaguchi K, Shimizu T: **Molecular Phylogeny in the Lardizabalaceae.** *J Plant Res* 1994, **107:**339-348.

Kosuge K, Sawada K, Denda T, Adachi J, Watanabe K: **Phylogenetic relationships of some genera in the Ranunculaceae based on alcohol dehydrogenase genes.** *Pl Syst Evol* 1995, (Suppl 9):263-271.

Kubitzki K: **Sabiaceae.** In *The families and genera of vascular plants*. *Volume 9.* Edited by Kubitzki K. Heidelberg: Springer-Verlag; 2007:413-417.

Langlet O: **Über Chromosomenverhältnisse und Systematik der Ranunculaceae.** *Svensk Bot Tidskr* 1932, **26:**381-400.

Lawver LA; Gahagan LM, Coffin MF: **The development of paleoseaways around Antarctica.** *Antarctic Research Series* 1992, **56:**7–30.

Lemesle R: **Position phylogénétique de l'*Hydrastis canadensis* L. et du *Crossosoma californicum* Nutt., d'aprés les particularités histologiques du xylème.** *Compt Rend Hebd Seances Acad Sci* 1948, **227:**221-223.

Lemesle R: **Contribution a l'etude de quelque familles de dicotyledones consideree comme primitives.** *Phytomorphology* 1955, **5:**11-45.

Leppik E: **Floral evolution in the Ranunculaceae.** *Iowa St J Sci* 1964, **39:**1-101.

Levinson G, Gutman A: **Slipped-strand mispairing: a major mechanism for DNA sequence evolution.** *Mol Biol Evol* 1987, **4:**203-221.

Lidén M: **Pteridophyllaceae.** In *The families and genera of vascular plants. Volume 2.* Edited by Kubitzki K, Rohwer JG, Bittrich V. Heidelberg: Springer-Verlag; 1993a:556-557.

Liang H, Hilu KW: **Application of the *matK* gene sequences to grass systematics.** *Can J Bot* 1996, **74:**125-134.

Lidén M: **Fumariaceae.** In *The families and genera of vascular plants. Volume 2.* Edited by Kubitzki K, Rohwer JG, Bittrich, V. Heidelberg: Springer-Verlag; 1993b:310-318.

Loconte H, Estes JR: **Phylogenetic systematics of Berberidaceae and Ranunculales (Magnoliidae).** *Syst Bot* 1989, **14:**565-579.

Loconte H: **Berberidaceae.** In *The families and genera of vascular plants. Volume 2.* Edited by Kubitzki K, Rohwer JG, Bittrich V. Heidelberg: Springer-Verlag; 1993:147-152.

Loconte H, Campbell LM, Stevenson DW: **Ordinal and familial relationships of ranunculid genera.** *Pl Syst Evol* 1995, (Suppl 9):99-118.

Löhne C, Borsch T: **Phylogenetic utility and molecular evolution of the *petD* group II intron in basal angiosperms.** *Mol Biol Evol* 2005, **22:**317-332.

Löhne C, Borsch T, Wiersema JH: **Phylogenetic analysis of Nymphaeales using fast-evolving and noncoding chloroplast markers.** *Bot J Linn Soc* 2007, **154:**141-163.

Lotsy JP: *Vorträge über botanische Stammesgeschichte*1911, **3:**578-594.

Maier RM, Neckermann K, Igloi GL, Kossel H: **Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing.** *J Mol Biol* 1995, **251:**614-628.

McLoughlin S: The **breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism.** *Aust J Bot* 2001, **49:**271–300.

Meacham CA: **Phylogeny of the Berberidaceae with an evaluation of classifications.** *Syst Bot* 1980, **5:**149-172.

Michel F, Umesono K, Ozeki H.: **Comparative and functional anatomy of group II catalytic introns – a review.** *Gene* 1989, **82:**5-30.

Miikeda O, Kita K, Handa T, Yukawa T: **Phylogenetic relationships of *Clematis* (Ranunculaceae) based on chloroplast and nuclear DNA sequences.** *Bot J Linn Soc* 2006, **152:**153-168.

Miyaji Y: **Beiträge zur Chromosomenphylogenie der Berberidaceen.** *Planta* 1930, **11:**650-659.

Moore M, Bell C, Soltis PS, Soltis DE: **Analysis of an 83-gene, 86-taxon plastid genome data set resolves relationships among several recalcitrant deep-level eudicot lineages [abstract].** *Botany* 2008, **203.**

Morgenstern B: **DIALIGN 2: improvement of the segment-to-segment approach to mutliple sequence alignment.** *Bioinformatics* 1999, **15:**211-218.

Müller K**: PRAP – computation of Bremer support for large data sets.** *Mol Phylogenet Evol* 2004, **31:**780-782.

Müller K, Borsch T, Legendre L, Porembski S, Theisen I, Barthlott W: **Evolution of carnivory in Lentibulariaceae and the Lamiales.** *Plant Biology*, 2004, **6:**477-490.

Müller J, Müller K: **Treegraph: automated drawing of complex tree figures using an extensible tree description format.** *Mol Ecol Notes* 2004, **4:**786-788.

Müller J, Müller K, Quandt D, Neinhuis C: **PhyDE – Phylogenetic Data Editor,** 2005 version 0.995.[http://www.phyde.de].

Müller K, Bosch T: **Phylogenetics of Amaranthaceae based on *matK/trnK* sequence data – evidence from parsimony, likelihood and Bayesian analyses.** *Ann Missouri Bot Gard* 2005a, **92:**66-102.

Müller K, Borsch T: **Phylogenetics of *Utricularia* (Lentibulariaceae) and molecular evolution of the *trnK* intron in a lineage with high substitutional rates.** *Plat Syst Evol* 2005b, **250:**39-67.

Müller K: 2005. **SeqState – primer design and sequence statistics for phylogenetic DNA data sets.** *Appl Bioinform* 2005, **4:**65-69.

Müller KF, Borsch T, Hilu KW: **Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: Contrasting *matK, trnT-F,* and *rbcL* in basal angiosperms.** *Mol Phylogenet Evol* 2006, **41:**99-117.

Nandi OI, Chase MW, Endress PK: **Combined cladistic analysis of angiosperms using *rbcL* and non-molecular datasets.** *Ann Mo Bot Gard* 1998, **85:**137-212.

Nixon KC: **The parsimony ratched: a rapid means for analyzing large data sets.** *Cladistics* 1999, **15:**407-414.

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H: **Complete nucleotide sequence of liverwort *Marchantia polymorpha* chloroplast DNA.** *Plant Molecular Biology Reporter* 1986, **4:**148-175.

Olmstead RG, Palmer J: **Chloroplast DNA systematics: a review of methods and data analysis.** *Am J Bot* 1994, **81:**1205-1224.

Olsson S, Buchbender V, Enroth J, Hedenäs L, Huttonen S, Quandt, D: **Phylogenetic alnalyses reveal high levels of polyphyly among pleurocarpous lineages as well as novel clades.** *The Bryologist* 2009, **112:** 447-466.

Ortiz RDC, Kellogg EA, Werff HVD: **Molecular phylogeny of the moonseed family (Menispermaceae): implications for morphological diversification.** *Am J Bot* 2007, **94:**1425-1438.

Posno M, Van Vliet A, Groot GSP: **The gene for *Spirodela oligorhiza* chloroplast ribosomal protein homologous to E. coli ribosomal protein l16 is split by a large intron near its 5' end: structure and expression.** *Nucleic Acid Research* 1986, **14:**3181-3195.

Prantl K: **Ranunculaceae.** In *Die natürlichen Pflanzenfamilien. Volume 3 II.* Edited by Engler A, Prantl K. Leipzig: W. Engelmann; 1891:43-66.

Qin HN: **An investigation on carpels of Lardizabalaceae in relation to taxonomy and phylogeny.** *Cathaya* 1989, **1:**61-82.

Qiu Y-L, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chase MW: **The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes.** *Nature* 1999, **402:**404-407.

Qiu Y-L, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chase MW: **Phylogeny of basal angiosperms: analyses of five genes from three genomes.** *Int J Plant Sci* 2000, **161:**3-27.

Qiu Y-L, Dombrowska O, Lee J, Li L, Whitlock BA, Bernasconi-Quadroni F, Rest JS, Davis CC, Borsch T, Hilu KW, Renner SS, Soltis DE, Soltis PS, Zanis MJ, Cannone JJ, Gutell RR, Powell M, Savolainen V, Chatrou LW, Chase MW: **Phylogenetic analyses of**

basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *Int J Plant Sci* 2005, **166:**815-842.

Qiu Y-L, Li L, Hendry TA, Li R, Taylor DW, Issa MJ, Ronen AJ, Vekaria ML, White AM: **Reconstructing the basal angiosperm phylogeny: evaluating information content of mitochondrial genes.** *Taxon* 2006, **55:**837-856.

Quandt D, Müller K, Huttunen S: **Characterisation of the chloroplast DNA *psbT-H* region and the influence of dyad symmetrical elements on phylogenetic reconstructions.** *Plant Biology* 2003, **5:**400-410.

Quandt D, Müller K, Stech M, Hilu KW, Frey W, Frahm JP, Borsch T: **Molecular evolution of the chloroplast *trnL-F* region in land plants.** *Monographs in Systematic Botany from the Missouri Botanical Garden* 2004, **98:**13-37.

Quandt D, Stech M: **Molecular evolution and phylogenetic utility of the chloroplast *trnT-trnF* region in bryophytes.** *Pl Biol* 2004, **6:**545-554.

Quandt D, Stech M: **Molecular evolution of the *trnL*(UAA) intron in bryophytes.** *Mol Phylogenet Evol* 2005, **36:**429-443.

Quandt D, Bell N, Stech M: **Unravelling the knot: the Pulchrinodaceae fam. nov. (Bryales).** *Nova Hedwigia Beiheft* 2007, **131:** 21-39.

Ren Y, Li H-F, Zhao L, Endress PK: **Floral Morphogenesis in Euptelea (Eupteleaceae, Ranuncuclales).** *Ann Bot* 2007, **100:**185-193.

Renner SS: **Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data.** *Am J Bot* 1999, **86:**1301-1315.

Renner SS, Chanderbali AS: **What is the relationship among Hernandiaceae, Lauraceae, and Monimiaceae, and why is this question so difficult to answer?** *Int J Plant Sci* 2000, **161** (6 Suppl)**:**109-119.

Renner SS, Foreman DB, Murray D: **Timing transantarctic disjunctions in the Antherospermataceae (Laurales): evidence from coding and noncoding chloroplast sequences.** *Syst Biol* 2000, **49:**579-591.

Ro K-E, Keener CS, McPheron BA: **Molecular phylogenetic study of the Ranunculaceae: utility of the nuclear 26S ribosomal DNA in inferring intrafamilial relationships.** *Mol Phylogenet Evol* 1997, **8:**117-127.

Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19:**1572-1574.

Ruiz E: **Ranunculaceae**. In *Flora de Chile*. *Volume 2*. Edited by Marticorena C, Rodríguez R. Concepción: Universidad de Concepción; 2001:40–84.

Sanmartín I, Ronquist F: **Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns.** *Syst Biol* 2004, **53:**216–243.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S: **Complete structure of the chloroplast genome of *Arabidopsis thaliana*.** *DNA Res* 1999, **6:**283-290.

Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, Bayer C, Fay MF, De Bruijn AY, Sullivan S, Qui Y-L: **Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences.** *Syst Biol* 2000a, **49:**306-362.

Savolainen V, Fay MF, Albach DC, Backlund M, Van der Bank M, Cameron KM, Johnson SA, Lledo L, Pintaud JC, Powell M, Sheanan MC, Soltis DE, Soltis PS, Weston P, Whitten WM, Wurdack KJ, Chase MW: **Phylogeny of the eudicots: a nearly complete familial analysis of the *rbcL* gene sequences.** *Kew Bull* 2000b, **55:**257-309.

Schmitz-Linneweber C, Maier RM, Alcaraz JP, Cottet A, Herrmann RG, Mache R: **The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization.** *Plant Mol Biol* 2001, **45:**307-315.

Schuettpelz E, Hoot SB, Samuel R, Ehrendorfer F: **Multiple origins of Southern Hemisphere *Anemone* (Ranunculaceae) based on plastid and nuclear sequence data.** *Plant Syst Evol* 2002, **231:**143-151.

Schütze P, Freitag H, Weising K: **An integrated molecular and morphological study of the subfamily Suaeoideae Ulbr. (Chenopodiaceae)**. *Plant Syst Evol* 2003, **239:**257-286.

Schuster RM: **Plate tectonics and its bearing on the geographical origin and dispersal of angiosperms.** In *Origin and early evolution of angiosperms*. Edited by Beck CB. New York: Columbia University Press; 1976:48-138.

Shimodaira H, Hasegawa M: **Multiple comparisons of log-likelihoods with applications to phylogenetic inference.** *Mol Biol Evol* 1999, **16:**1114-1116.

Shimodaira H, Hasegawa, M: **CONSEL: for assessing the confidence of phylogenetic tree selection.** *Bioinformatics* 2001, **17:** 1246-1247.

Shimodaira H: **An approximately unbiased test of phylogenetic tree selection.** *Syst Biol* 2002, **51:**492-508.

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M: **The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression.** *EMBO J* 1986, **5:**2043-2049.

Simmons MP, Ochoterena H: **Gaps as characters in sequence-based phylogenetic analyses.** *Syst Biol* 2000, **49:**369-381.

Simmons MP, Pickett KM, Miya M: **How meaningful are Bayesian support values?** *Mol Biol Evol* 2004, **21:**188-199.

Simões AO, Endress ME, Niet T Van Der, Kinoshita LS, Conti E: **Tribal and intergeneric relationships of Mesechiteae (Apocynoideae, Apocynaceae): evidence from three noncoding plastid DNA regions and morphology.** *Am J Bot* 2004, **91:**1409-1418.

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS: **Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences.** *Bot J Linn Soc* 2000, **133:**381-461.

Soltis DE, Senters AE, Zanis MJ, Kim S, Thompson JD, Soltis PS, Ronse De Craene LP, Endress PK, Farris JS: **Gunnerales are sister to other core eudicots: implications for the evolution of pentamery.** *Am J Bot* 2003, **90:**461-470.

Stech M, Quandt D: **Molecular evolution and phylogenetic utility of the chloroplast *atpB-rbcL* spacer in bryophytes.** In *Plant Genome: Biodiversity and Evolution. Volume 2B.* Edited by Sharma AK, Sharma A. Enfield: Science Publishers; 2006:409-431.

Steele KP, Vilgalys R: **Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*.** *Syst Bot* 1994, **19:**126-142.

Swofford DL: *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4.* Sunderland, Massachusetts: Sinauer Associates; 2002.

Taberlet P, Gielly L, Pautou G, Bouvet J: **Universal primers for amplification of three non-coding regions of the chloroplast DNA.** *Plant Mol Biol* 1991, **17:**1105-1109.

Takhtajan A: *Diversity and Classification of Flowering Plants.* New York: Columbia University Press; 1997.

Tamura M: **Taxonomical and phylogenetical consideration of the Ranunculaceae.** *Acta Phytotax Geobot* 1962, **20:**71-81 (in Japanese).

Tamura M: **Morphology and phyletic relationship of the Glaucidiaceae.** *Bot Mag Tokyo* 1972, **85:**29-41.

Tamura M: **Morphology, ecology and phylogeny of the Ranunculaceae VII.** *Sci Reports Osaka Univ* 1967, **16:**21-43.

Tamura M: **Ranunculaceae.** In *The families and genera of vascular plants*. *Volume 2.* Edited by Kubitzki K, Rohwer JG, Bittrich V. Heidelberg: Springer-Verlag; 1993:563-583.

Tamura M: **Ranunculaceae.** In *Die natürlichen Pflanzenfamilien*. *Volume 17a (4).* Edited by Hiepko P. Berlin: Duncker & Humblot; 1995:223-555.

Terabayashi S: **The comparative floral anatomy and systematics of the Berberidaceae. I. Morphology.** *Memoirs of the Faculty of Science, Kyoto University, Series of Biology* 1985a, **10:**73-90.

Terabayashi S: **The comparative floral anatomy and systematics of the Berberidaceae. II. Systematic considerations.** *Acta Phytotax Geobot* 1985b, **36:**1-13.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucl Acids Res* 1997, **24:**4876-4882.

Thorne RF: **A phylogenetic classification of the Annoniflorae.** *Aliso* 1974, **8:**147-209.

Thorne RF: **A phylogenetic classification of the angiosperms.** *Evol Biol* 1976, **9:**35-106.

Tobe H: **Embryological studies in *Glaucidium palmatum* Sieb. et Zucc. with a discussion on the taxonomy of the genus.** *Bot Mag Tokyo* 1981, **94:**207-224.

Tobe H, Keating RC: **The morphology and anatomy of *Hydrastis* (Ranunculales): systematic reevaluation of the genus.** *Bot Mag Tokyo* 1985, **98:**291-316.

Toor N, Hausner G, Zimmerly S: **Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases.** *RNA* 2001, **7:**1142-1152.

Ulbrich E: **Über die systematische Gliederung und geographische Verbreitung der Gattung *Anemone* L.** *Bot Jahrb Syst, Pflanzengeschichte Pflanzengeographie* 1905/06, **37:**172-334.

Von Balthazar M, Pedersen KR, Friis EM: ***Teixeiraea lusitanica* gen. et sp. nov., a ranunculalean flower from the Early Cretaceous of Portugal.** *Pl Syst Evol* 2005, **255:**55–75.

Wang W, Li R-Q, Chen Z-D: **Systematic position of *Asteropyrum* (Ranunculaceae) inferred from chloroplast and nuclear sequences.** *Pl Syst Evol* 2005, **255:**41-54.

Wang W, Chen Z-D, Liu Y, Li R-Q, Li J-H: **Phylogenetic and biogeographic diversification of Berberidaceae in the northern hemisphere.** *Syst Bot* 2007, **32:**731-742.

Wang W, Lu A-M, Endress ME, Chen ZD: **Phylogeny and classification of Ranunculales: evidence from four molecular loci and morphological data.** *Perspectives in Plant Ecology, Evolution and Systematics* 2009, **11:**81-110.

Wang Z-F, Ren Y: **Ovule Morphogenesis in Ranunculaceae and its systematic significance.** *Ann Bot* 2008, **101:**447-462.

Wanke S, Jaramillo MA, Borsch T, Samain MS, Quandt D, Neinhuis C: **Evolution of Piperales – *matK* gene and *trnK* intron sequence data reveal lineage specific resolution contrast.** *Mol Phylogenet Evol* 2007, **42:**477-497.

Wanntorp L, Wanntorp HE: **The biogeography of *Gunnera* L.: vicariance and dispersal.** *J Biogeogr* 2003, **30:**979–987.

Wheeler WC, Gladstein DS, De Laet J: *POY. Version 3.0.*, New York; 1996–2003.

White TJ, Bruns T, Lee S, Taylor J: **Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics.** In *PCR Protocols. A guide to methods and*

applications. Edited by Innis MA, Gelfand DH, Snisky JJ, White TJ. San Diego: Academic Press; 1990:315-322.

Wicke S, Quandt D: **Universal primers for amplification of the *trnK/matK-psbA* region in land plants.** *An Jard Bot Madrid*, in press.

Worberg A, Quandt D, Barniske A-M, Löhne C, Hilu KW, Borsch T: **Phylogeny of basal eudicots: Insights from non-coding and rapidly evolving DNA.** *ODE* 2007, **7:**55-77.

Zhang W: **Phylogeny of the grass family (Poaceae) from *rpl16* intron sequence data.** *Mol Phylogenet Evol* 2000, **15:**135-146.

Curriculum Vitae

Identity

Name:	Anna-Magdalena Barniske
Date of Birth:	19.03.1976 in Halle/Saale
Nationality:	German
Marital status:	unmarried
Working address:	Zellescher Weg 20b
	01217 Dresden, Germany
Telephone:	+49 351 46335834
Email:	anna.magdalena.barniske@tu-dresden.de

Professional Experience

02/2005 – 10/2009	Research assistant at the "Technische Universität Dresden", Dresden/Germany Molecular evolution and phylogenetic signal of chloroplast introns; Evolution of growth forms within Ranunculales
09/2003 – 12/2004	Research assistant for the "Martin-Luther-Universität Halle-Wittenberg", Halle/Saale/Germany Project: On preservation of the endemic saxonian species *Calamagrostis rivalis* H.Scholz at the river Mulde
02/1998 – 07/1999	Student assistant at the „Martin-Luther-Universität Halle-Wittenberg", Germany Working group: Chorology

University studies

02/2005 – 10/2009	Dissertation at the "Technische Universiät Dresden", Dresden/Germany Thesis: Mutational dynamics and phylogenetic utility of plastid introns and spacers in early branching eudicots Supervisor: Prof. Dr. C. Neinhuis, Prof. Dr. D. Quandt
04/2002	Diplom at the "Martin-Luther-Universität Halle-Wittenberg", Halle/Saale/Germany Degree: Diplom-Biologin. Rate: 1,9 Thesis: Contributions to the population biology, morphology and sociology of *Armoracia rusticana* GARTN., MEY. & SCHERB. (*Brassicaceae*) Supervisor: Prof. Dr. E. J. Jäger
10/1995 – 04/2002	Martin-Luther-Universität Halle-Wittenberg", Halle/Saale/Germany Major: Geobotany, Minors: Zoology, Geology/Palaeontology and Genetics

Publication List

Publications

Worberg A., Quandt D., **Barniske A.-M.**, Löhne C., Hilu K.W., Borsch T., (2007). Towards understanding early Eudicot diversification: insights from rapidly evolving and non-coding DNA. Organisms, Diversity and Evolution 7, 55 – 77.

Symposium abstracts

Worberg A., Quandt D., **Barniske A.-M.**, Löhne C., Hilu K.W., Borsch T.: Towards understanding early Eudicot diversification: insights from rapidly evolving and non-coding DNA. Botany 2006, Chico.

Worberg A., Quandt D., **Barniske A.-M.**, Löhne C., Hilu K.W., Borsch T.: Towards understanding early Eudicot diversification: insights from rapidly evolving and non-coding DNA. 17th International Symposium "Biodiversity and Evolutionary Biology", Bonn.

Barniske A.-M., Borsch T., Worberg A:, Müller K., Quandt D.: Corroborating the branching order among basal Eudicots: More insights from more rapidly evolving and non-coding DNA. Botany 2007, Chico.

Posters

Barniske A.-M., Borsch T., Worberg A., Müller K., Quandt D.: Corroborating the branching order among basal Eudicots: More insights from more rapidly evolving and non-coding DNA. Botanikertagung der Deutschen Botanischen Gesellschaft 2007, Hamburg.

Salomo, K., Borsch, T., **Barniske, A.-M.**, Neinhuis, C., Quandt, D.: Indel evolution and secondary structures of group I & II introns in Asterids, Systematics 2008, Göttingen.

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Die Bestimmungen der Promotionsordnung sind mir bekannt.