



**Unravelling the impact of substitution rates and
time on branch lengths in phylogenetic trees**

MSc Biology Thesis Biosystematics (BIS-80430)
Wageningen University of Research
By Isolde van Riemsdijk
St.nr: 910104694070
Supervisor: Lars W. Chatrou
2nd Supervisor: Freek Bakker
May 2014

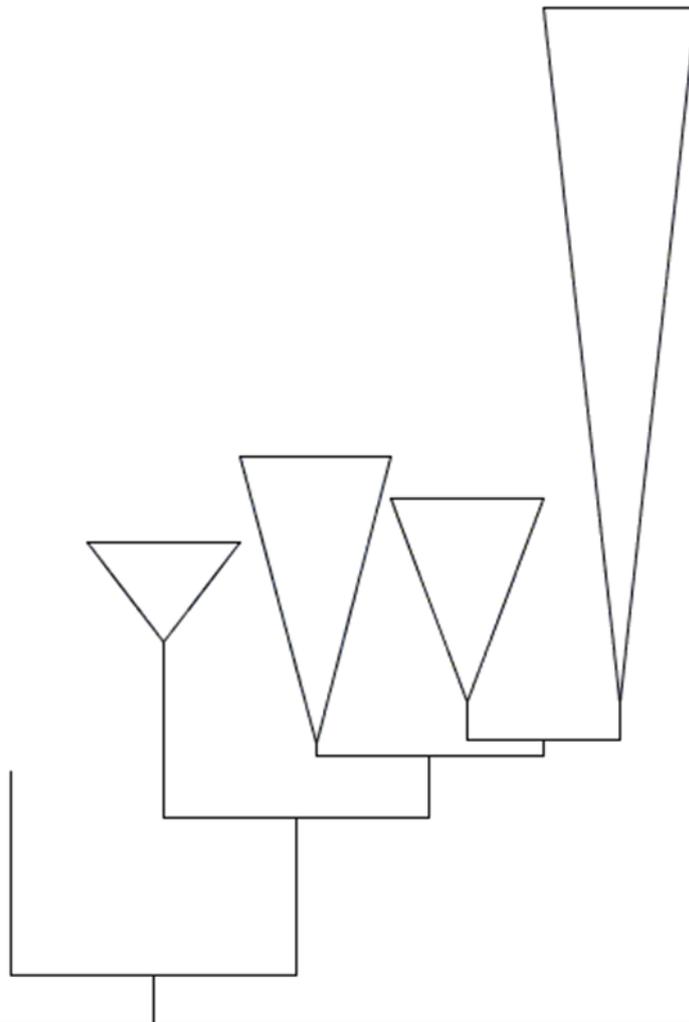


Figure 1: Phylogeny of the Annonaceae constructed with RAxML from the total dataset with clades from left to right: *Eupomatia bennettii*, *Anaxagorea*, Ambavioideae, the short branch clade (SBC) alias the Malmeoideae and the long branch clade (LBC) alias the Annonoideae

Contents

| | |
|--|----|
| 1. Preface | vi |
| 2. Summary | vi |
| 3. Introduction | 1 |
| 3.1 <i>Genomics and the assumption of a molecular clock</i> | 2 |
| 3.2 <i>Evolutionary divergence times</i> | 4 |
| 3.3 <i>Using fossil characteristics to construct a calibration prior distribution</i> | 4 |
| 4. Research questions | 8 |
| 5. Materials and methods | 9 |
| 5.1 <i>Construction of the DNA sequence supermatrix</i> | 9 |
| 5.2 <i>Finding the best possible partition scheme and models</i> | 10 |
| 5.2.1 <i>Different analyses for LBC and SBC datasets and comparison</i> | 10 |
| 5.2.2 <i>PartitionFinder settings and results</i> | 11 |
| 5.3 <i>Phylogenetic tree construction</i> | 11 |
| 5.4 <i>Restriction of nodes with fossils</i> | 11 |
| 5.4.2 <i>Fossils providing deep node ages in the Annonaceae phylogeny</i> | 12 |
| 5.4.3 <i>Fossils providing shallow node ages in the Annonaceae phylogeny</i> | 14 |
| 5.5 <i>Dating analyses with the 4M dataset: BEAST</i> | 18 |
| 6. Results | 20 |
| 6.1 <i>PartitionFinder: models for the LBC and SBC</i> | 20 |
| 6.2 <i>RAXML bootstrap phylogenetic trees</i> | 20 |
| 6.3 <i>BEAST analyses results</i> | 20 |
| 6.3.1 <i>Node age estimates of the fossil exclusion and LBC:SBC ratio analyses</i> | 21 |
| 6.3.2 <i>Mean substitution rate estimates for different interesting analyses</i> | 27 |
| 7. Discussion | 30 |
| 7.1 <i>First analyses with the data</i> | 30 |
| 7.2 <i>Comparing published ages to results</i> | 31 |
| 7.3 <i>Changing effects of calibration priors in different combinations and data selection</i> | 32 |
| 8. Conclusion | 35 |
| 9. Future research | 36 |
| 10. Acknowledgements | 38 |
| 11. References | 38 |
| Appendix 1: Species table with GenBank numbers and references | 43 |
| Appendix 2: Starting partition schemes and resulting best schemes for each dataset..... | 62 |
| Appendix 3: RAXML bootstrap trees for the 4M and total dataset | 63 |
| Appendix 4: Record of BEAST analysis runs in Tracer v. 1.5..... | 70 |
| Appendix 5: Maximum clade credibility trees resulting from the BEAST analyses..... | 74 |

1. Preface

I have always had a special interest in evolution. I find it amazing every time to see how simple the processes underlying evolution can be, while the variety of species is so large. I came to work on the problem of the presence of a long branch and short branch clade in the Lentibulariaceae. I learned that the programmes used to investigate phylogenetic trees have become very complicated, but still cannot discern between elapsed time since divergence and substitution rate. To know how, when and in the future maybe even why the divergence of extant plant groups has taken place will shed new light on plant evolution. To solve the problem of the fusion of time since divergence and substitution rate in phylogenetic programs, first it has to be investigated together with the already existing ways of accounting for rate heterogeneity in phylogenetic research.

2. Summary

This thesis explores the possibilities to make a distinction between the influences of sequence data in which rate heterogeneity is shown and calibration priors on age and rate estimates. Rate heterogeneity impedes the assumption of a strict molecular clock and causes demand for more complicated models. Annonaceae data have been proved in previous publications to show rate heterogeneity.

A dataset of 252 species from the Annonaceae family and four plastid markers (*matK*, *ndhF*, *psbA-trnH* and *trnL-trnF*) was constructed by combination of already existing datasets. BEAST is used to perform dating analysis taking rate heterogeneity into account.

Former dating analyses only used (older) deep node fossils for the Annonaceae clade. This thesis also includes four (younger) fossils assigned to clades situated more shallow in the phylogeny of Annonaceae. Including more fossils increases support of a hypothesis of ages. To identify the influences of the shallow fossil calibrations separately and in combination, all possible younger fossil calibration combinations are used in dating analyses in BEAST.

To be able to identify the influence of the sequence data on age and rate estimates also analyses were done including all fossil calibrations but only part of the data. Two extreme situations are created by reducing the short branch clade (SBC) sequences to a minimum of five species in one dataset, remaining all long branch clade (LBC) sequences and vice versa. Mean age estimates and 95 % HPDs are recorded for eight nodes of interest and rate distributions for five analyses are obtained.

The results show that the influence of calibration priors is not easy to interpret. Age estimates probably are very much influenced by the amount of sequence data which is constrained by a calibration prior and the properties of those sequences. The age estimates of the total dataset and total prior set are in general older than ages published before.

3. Introduction

The phylogenetics of Annonaceae have a long history of changes and uncertainties (for a recent summary read Chatrou *et al.* (2012)). The phylogenetic tree of the Annonaceae is not resolved yet at the species level. Using a large dataset, Chatrou *et al.* (2012) constructed a tree which showed improvement in generic representation and resolution. Still, despite the high amount of sequence data, there are some issues related to resolving the Annonaceae tree. One of them is the observation that there are symptoms of rate heterogeneity within the Annonaceae (Doyle *et al.*, 2004; Pirie & Doyle, 2012). This causes the topology of the tree to have a long branch clade (LBC) and short branch clade (SBC) (Richardson *et al.*, 2004), see for an example figure 1. Rate heterogeneity makes it harder to estimate divergence dates in a reliable manner. This effect of rate heterogeneity on divergence time estimate is shown by the difference in the result of two studies on dating of the LBC and SBC of the Annonaceae family. Richardson *et al.* (2004) shows the divergence dates of the two clades to be quite similar. The crown node of the SBC was estimated at 62.5 ± 3.6 Mya and the LBC crown node at 60.2 ± 3.6 Mya with *Archaeanthus* as calibration point and using NPRS (figure 2). The analysis of Couvreur *et al.* (2011), done with more data (including the same data as Richardson) but with BEAST resulted in quite different ages. The SBC crown node was estimated at 40.0-25.8 Mya and the LBC crown node age at 72.4-59.2 Mya. Is it true there is a difference in age between the LBC and SBC clade or are they approximately the same age? Although an interesting question, the answer will always be a new hypothesis, since we will not be able to know for certain what the real evolutionary past looked like. All we can do is take these results as a warning. Different tools tend to give different results.

When rate heterogeneity is present in a data set, the use of relaxed molecular clocks

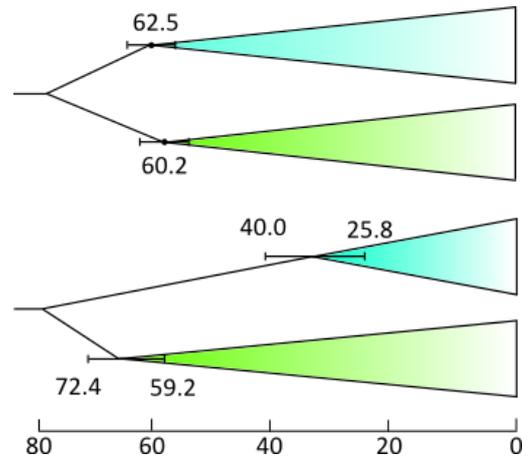


Figure 2: Annonaceae LBC (green) and SBC (blue) dated in the articles of Richardson *et al.* (2004) (above, ages ± 3.6 My) and Couvreur *et al.* (2011) (below), age axis in My

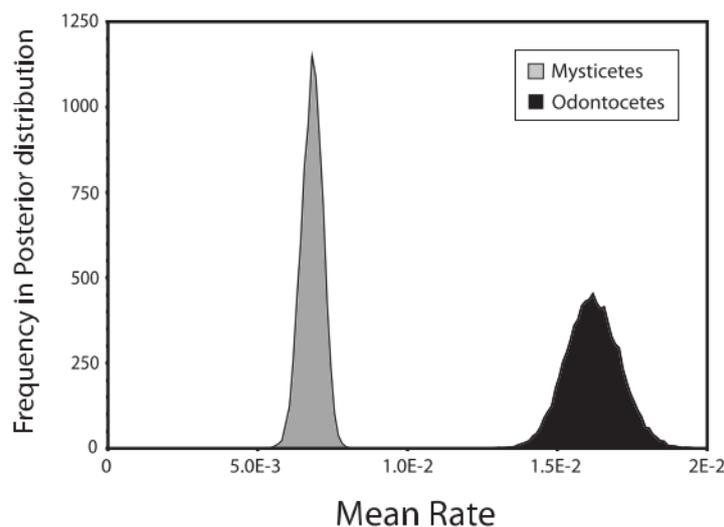


Figure 3: Molecular rate distribution with mean substitution rate on x-axis for crown Odontoceti and Mysticeti (clades of Cetacea) in the RLC framework mitochondrial data set (Dornburg *et al.*, 2012)

may help to estimate branch lengths more accurately (Drummond *et al.*, 2006; Magallón *et al.*, 2013). On the other hand, when the substitution rate changes fast and locally between lineages, drawing (auto-correlated) substitution rates from a single Poisson distribution as is done in relaxed clock methods (Drummond *et al.*, 2006) might not handle the differences in rate within the data successfully. The substitution rates can then be better described by two (or more) distributions. For Cetacea, it was found that substitution rates differ significantly across lineages (Dornburg *et al.*, 2012), resulting in two distinct posterior distributions of substitution rates representing two sister clades (figure 3). It is not yet possible to test for rate differences between lineages and models without the researcher assigning data partitions in advance.

The Annonaceae consist of about 2500 species to date (Rainer & Chatrou, 2011) which are mainly found in tropical rain forests. Annonaceae species can be trees, shrubs or lianas (Encyclopaedia Britannica Online Academic Edition, 2014). The fruits of many species are locally used as food, medicine, spices or as tools (Encyclopaedia Britannica Online Academic Edition, 2014). A recent publication of a special issue of the Botanical Journal of the Linnean Society dedicated to the Annonaceae, contained the current state of systematic, ecological and evolutionary research on the plant family (The Linnean Society of London, 2012).

3.1 Genomics and the assumption of a molecular clock

Calculating the divergence time of a species with a strict molecular clock can be troublesome (box 1). Not only within a species' genome, but also between species the substitution rate may differ. Plenty of tools are available which account for the effect of rate heterogeneity in different ways yet the use of a tool and its accompanying assumptions will depend on the characteristics of the data.

Box 1: Differences in substitution rates

First, substitution rates are not constant within a genome (Morton & Clegg, 1995; Thorne & Kishino, 2002). When using slow (highly conserved genes with low substitution rates) and faster (*matK*) data from land plants it appears the substitution rate is changing over lineages in similar patterns for different genes (Magallón *et al.*, 2013).

Second, synonymous and nonsynonymous rates can differ for chloroplast DNA, which also has to be taken into account when calculating divergence times (Bousquet, *et al.*, 1992). One would expect an increase in the nonsynonymous substitution rate for genes involved in niche adaptation when populations split into different species and significantly affect the substitution rate, either by selective pressure or a change in effective population size, though it might be hard to prove (Duchene & Bromham, 2013). On the other hand, synonymous rates are also related to speciation in flowering plants (Barraclough *et al.*, 2001). Effective population size is suggested as a possible cause, as well as the influence of generation time on the rate of synonymous substitutions over real time and variation of substitution rates over lineages. In the case of Annonaceae, the difference in branch lengths over the phylogenetic tree are not caused by the influence of different synonymous and nonsynonymous rates (Chatrou *et al.*, Unpublished).

Third, there can be a difference in the rate of substitution between the three codon positions. By using only synonymous substitutions, one could possibly avoid this difference in the rate of substitution in codon position, since synonymous substitutions do not change the codon. A solution to many of these problems is using data partitions (paragraph 5.2).

The difference in branch lengths between the Annonaceae long and short branch clades indicates that some estimates in the phylogenetic tree are not accurate. The tools used for dating the Annonaceae phylogeny should be able to model rate heterogeneity in a way that fits the data best. BEAST (v1.7, Drummond *et al.*, 2012) allows for drawing substitution rates from one normal distribution, while (as stated above) this might not be the right assumption given the data.

PAML has BASEML and CODEML to estimate species divergence times under local-clock models (Yang, 2007). In PAML it is possible to manually select parts of the phylogenetic tree where the clock model changes. This means the researcher decides where the rate changes, after which the resulting branch models can be tested with a likelihood analysis. r8s (v1.5, Sanderson, 2003) is a tool which allows to use parametric, nonparametric and semiparametric methods, in order to better estimate rates and times by relaxing the assumption of constant rates. It also allows the use of multiple age constraints and fits smoothing models. The level of smoothing can be determined by means of cross validation from severe (molecular clock) to highly unconstrained, allowing very rapid changes in rate across the tree. Also branch ends can occur at any time in this model, which allows for investigation of rate variation.

When analysing the divergence between two species using only molecular data the estimated divergence may be too far back in time. When a polymorphism is present in a population, the species arising from that population often contain a fixed alternative allele (figure 4). This effect may be larger in chloroplast and mitochondrial DNA than in nuclear DNA (Benton & Ayala, 2003).

Because of the abovementioned differences in evolutionary rates between species, species sampling is of great importance when handling rate heterogeneity. Estimation of variables in maximum-likelihood models is influenced by species sampling (Jack Sullivan *et al.*, 1999). Increased taxon sampling improves parameter estimation and when more species are added, long branches become shorter because less unobserved changes need to be modelled (Heath *et al.*, 2008).

Finally, gene trees are not species trees as can also be concluded from many of the problems of phylogenetic reconstruction mentioned above (figure 4). This also means that, when using geographic and/or fossil data and molecular data, two different species concepts are used. Since the time scale on which this research concentrates ranges about 100 My and fossil ages can be estimated at best with 1 My accuracy, this problem is probably not encountered.

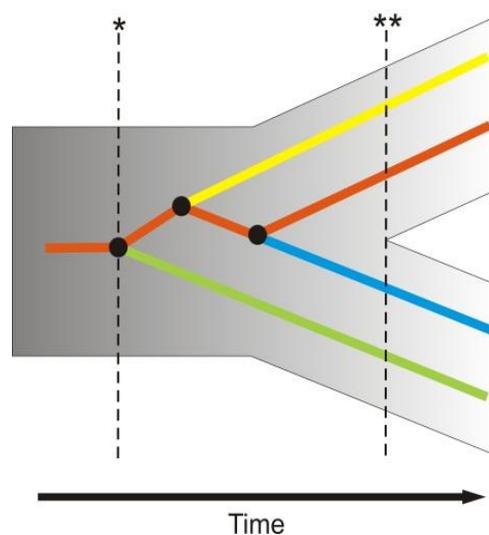


Figure 4: Gene tree within species tree, colours indicate different gene types, *tMRCA of the red and green gene variant, **morphological separation of the species

3.2 Evolutionary divergence times

To be able to turn relative substitution rates into absolute substitution rates along branches, one can estimate divergence times by forcing real time events onto the phylogenetic tree. Real time events can be paleogeographic events (box 2) as well as fossils. Fossils are more probable to provide realistic age estimates for Annonaceae. Using fossils to date clades has become popular within phylogenetic studies, although sometimes the limitations of the fossil record are not taken into account (Gandolfo *et al.*, 2008; Heads, 2005; Nixon, 1996).

Because of the difficulties with the fossil record mentioned in box 3, it is important to get expert opinion on available fossils before they can be properly used in research. When characters of fossils only occur in one clade in the angiosperms, placement of the fossils will not be problematic (Crepet, 2008). Using multiple fossil calibrations with soft bounds in phylogenies with heterogeneity reduces the possibility of errors in estimating the species divergence times (Benton & Donoghue, 2007; Reisz & Müller, 2004; Yang & Rannala, 2006). Especially with variable diversification rates the use of multiple fossil dates can shed light on the patterns and degree of rate variation (Benton *et al.*, 2003).

Box 2: Paleogeographic events to date phylogenetic trees and Annonaceae

Phylogenetic age estimate models can be calibrated using paleogeographic events as dates of nodes containing species with different geographical dispersal, explained by the event. In order to use dates from geographical dispersal, the sampling of species should be done sufficiently dispersed and intensive, in a geological sense (Heads, 2005). For some lineages it might be the case that no well determined geographical event can be found.

Annonaceae are present in tropical forests around the world. This dispersal pattern may be explained by plate tectonic theory in combination with climatic change (Richardson *et al.*, 2004). But the dispersal pattern of Annonaceae clades is amongst others influenced by long-distance dispersal (Erkens *et al.*, 2009; Richardson *et al.*, 2004). The best alternative seems the use of fossils to date nodes.

3.3 Using fossil characteristics to construct a calibration prior distribution

So far, ways of determining minimal ages of nodes have been discussed. The maximum age of nodes cannot so easily be argued by the use of fossils and is often done using some prior estimate (Clarke *et al.*, 2011; Donoghue & Benton, 2007; Ho & Phillips, 2009). Nowak *et al.* (2013) use a model to estimate, with the help of the known fossils, the real FAD of a species. They attempt to estimate when a species actually came to exist. There are many factors influencing the time between a speciation event from which it descends and the first appearance of a fossil from that species. Such a corrected calibration prior should be dependent on how fast a species evolves, the factors influencing the fossilisation 'ability' of a species (box 3) and the factors influencing whether a fossil is found and correctly identified. The magnitude of the influence these factors have on the first appearance of a fossil in the record is hard to estimate and chance always plays a major role (Reisz *et al.*, 2004). An attempt to estimate the distance between FAD and the real splitting event in the phylogeny is maybe not the most correct way of accounting for these uncertainties.

Ho & Phillips (2009) give an overview of different types of priors for age estimate of clades by the use of fossil ages, of which a short summary is presented in box 4.

Box 3: Fossilisation processes and first appearance dates

The process of fossilisation depends on the materials an organism is composed of and in what environment it lives. In the case of plants some of the hardest tissues are lignin, cellulose and sporopollenin, but compared to bone and shell these tissues are rather soft. After fossilisation, the fossil can be altered into unrecognizable shapes due to geological processes such as plate uplift, mountain formation or erosion. If the fossil survives until it reaches the earth's crust again, the fossil has to be found, taken up in a collection and identified correctly. In the case of plant fossils, the fossil gatherers have to have some botanical knowledge to be able to recognize a fossil as being a (part of) a plant. This short account of plant fossilisation is based on the classic paper of Raup (1972) describing nine filters influencing the composition of the fossil record. The fossil record is thus at least very patchy in three dimensions; time, space and species covering (Gandolfo *et al.*, 2008).

To be able to use a fossil for calibrating a clade divergence, it is important to be sure about the species or clade to which the fossil is assigned. Some individual fossils may be hard to relate correctly to extant species. Since additional (new) information may place them elsewhere entirely, it may be best not to use these fossils (Nixon, 1996). Due to taphonomy (missing parts of an organism due to processes around and after time of death) it may be hard to place a fossil accurately (Donoghue & Purnell, 2009; Nixon, 1996). The application of fossil calibrations on nodes in phylogenies is often done without proper research on phylogenetic relationships, placing fossils as direct ancestors or sister taxa of extant species (Gandolfo *et al.*, 2008).

The oldest specimen of a fossil species gives the first appearance date (FAD) of a fossil species. The age of a fossil is estimated by means of the stratigraphic layer it was found in. Stratigraphic layers are defined by a clear upper and lower boundary. This usually reduces the age to the 'relative age', which may be accurate with less than 1 My. When the fossil was not found too long ago an even more accurate date could be determined by radiometric methods (Benton *et al.*, 2007). There are examples of new fossil finds which predate the FAD of a fossil species with up to 100 My which, as could be expected of such finds, change the previous views of radiation dates on the involved groups radically (Heads, 2005). As Heads indicates, only fossils dated earlier than former used fossils should be regarded as relevant for calibration of phylogenetic dating analyses.

Box 4: Prior types for age estimates: a summary of Ho & Phillips (2009)

- 1) The point calibration (Figure 5A) is the type of prior in which the age of divergence is set to one point in time. This can only be done when the fossil species used is the real common ancestor of two extant lineages, which is very unlikely.
- 2) The hard minimum bound (Figure 5B) is the prior most often used for fossil data, since the information provided by fossil evidence is only a statement of minimal divergence age of the clade the fossil belongs to.
- 3) The hard maximum bound (Figure 5C) is not often used, because it needs the determination of a clade being absent at a certain time, and yet to evolve. Evidence for absence is hard to find and this kind of reasoning is very easily disproved by discovery of a fossil that is dated earlier than the maximum hard bound (Heads, 2005). To be able to include the uncertainty about the fossil age and placement in the analysis, one can use soft bounds (Figure 5D,E). These soft bounds add a tail of exponential decreasing probability.
- 4) The normal distribution (Figure 5F) has soft bounds for both maximum and minimum age of divergence and is a good prior when calibrating on the basis of geological events. If a fossil shows characters which make it an intermediate between being ancestral and derived a normal prior distribution can be used. It is only suitable for those cases in which there is no justification for a bias towards the fossil age as there is in the lognormal distribution.
- 5) The lognormal distribution (Figure 5G) has a wide variety of shapes depending on the parameter values used for the mean, the standard deviation and the hard minimum bound. The probability of divergence is biased towards the age of the fossil. An example of a reason for a bias towards the fossil age is when the fossil shows much apomorphic characters and little plesiomorphies. The beauty of using this prior distribution is that it gives a probability of zero for the node age to be the same as the age of the fossil.
- 6) As opposed to the lognormal distribution, the exponential distribution (Figure 5H) only has two parameters, the mean and the hard minimum bound. A reason to apply this prior distribution is when there is evidence that the fossil species is situated close to the divergence of the clade.

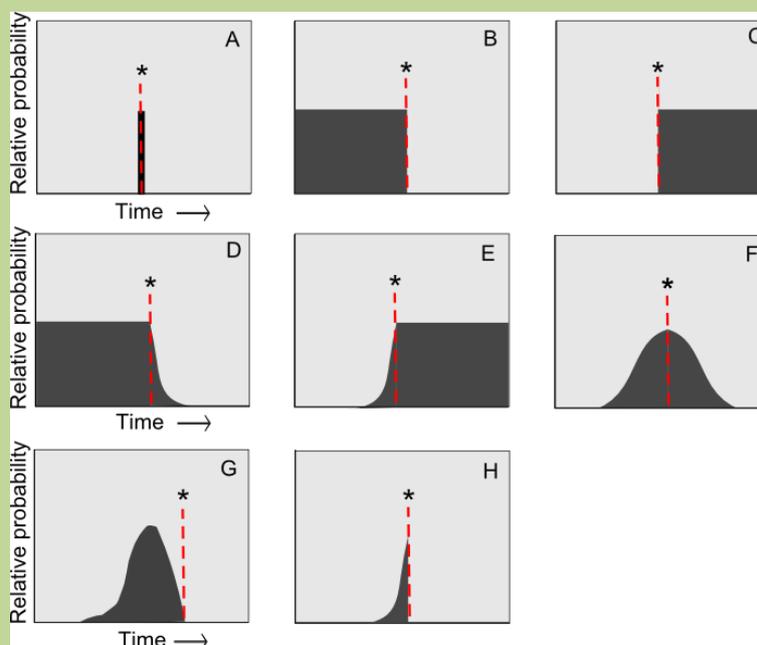


Figure 5: prior shapes with relative probability of node age on the y-axis and time on the x axis, age of the fossil indicated by dotted line with asterisk (based on Ho & Phillips (2009))

When a fossil shows morphological similarities with multiple extant species it can be placed on the crown node of the clade of extant species showing similar characteristics. In the case of figure 6 the hatched filling and round shape of the fossil (F) indicates it is related to the "hatched squared" species and the "round with a dot" species. The corresponding characters should have arisen somewhere between the crown node to which the fossil is assigned because of its synapomorphic characters (2) and the stem node of that same clade (1). In this way, the fossil provides a minimum (youngest) age for the stem node (1).

Because of the arguments mentioned above, molecular data often overestimates divergence times and paleontological data often underestimates divergence times (Rodríguez-Trelles *et al.*, 2002). Shaul & Graur (2002) even suggest not to use molecular data at all to estimate divergence time. According to them the use of two different kinds of data causes general incongruence in estimates of divergence dates of clades. But when further investigating the phylogenetic relationship, working with more and more data, Benton and Ayala (2003) show for the mammal phylogenetic tree that the results of both kinds of data approach each other. And the inclusion of multiple fossils to provide calibration priors in a phylogenetic analysis lower the possible erroneous effect of a single fossil (Conroy & van Tuinen, 2003; Near & Sanderson, 2004; Smith & Peterson, 2002).

Concluding, the shape of the prior distribution depends on the (expected) error in age determination of the fossil itself as well as the certainty of the placement in the phylogenetic tree (Wheat & Wahlberg, 2013). The most reliable seems to be to use multiple fossils for minimum bounds and only use one hard maximum bound on the deepest fossil prior of the clade. Only when there is clear indication for other prior distributions to be more appropriate (box 4) they should be used. By using only minimum hard bounds there are as little assumptions made as possible (Clarke *et al.*, 2011).

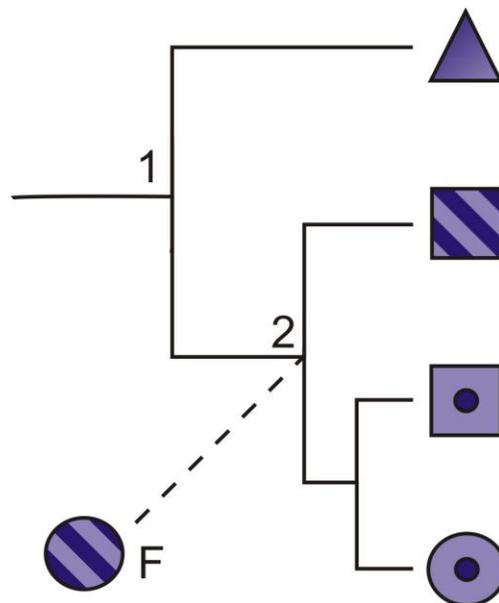


Figure 6: placement (dashed line) of fossil (F) in a crown group according to morphological characteristics (node 2) with the stem node (1)

4. Research questions

In this thesis I discuss some fossils which can be placed in the Annonaceae phylogenetic tree in a reliable manner. When using different fossil combinations the branch lengths will change. Using multiple fossils for age calibrations will decrease the influence of erroneous use of fossil calibrations. The use of multiple fossil constraints in trees with rate heterogeneity is likely to reduce the difference in branch length between the long branch and short branch clade.

The research questions for this thesis are:

- 1) What fossils can be used and how can they be used in a reliable way to constrain a phylogenetic tree when rate heterogeneity is present and date the tree?
- 2) Do fossil calibrations influence the branch length in a phylogenetic tree when rate heterogeneity is present?
- 3) Do different selections of LBC and SBC species sequences influence the estimated ages when calibration priors are present?

To investigate the influence of sequence data on the age estimates, the amount of sequences of the long branch and the short branch clade are varied. When rate heterogeneity is actually present this might result in different ages for the clades and therefore result in different branch lengths. In an analysis already performed by my supervisor, L.W. Chatrou, the same method was used to investigate the influence of the sequence data on the age of the LBC and SBC. As a result, the depicted the age estimates for the different ratios of LBC/SBC sequences changes the estimated ages as shown in figure 7.

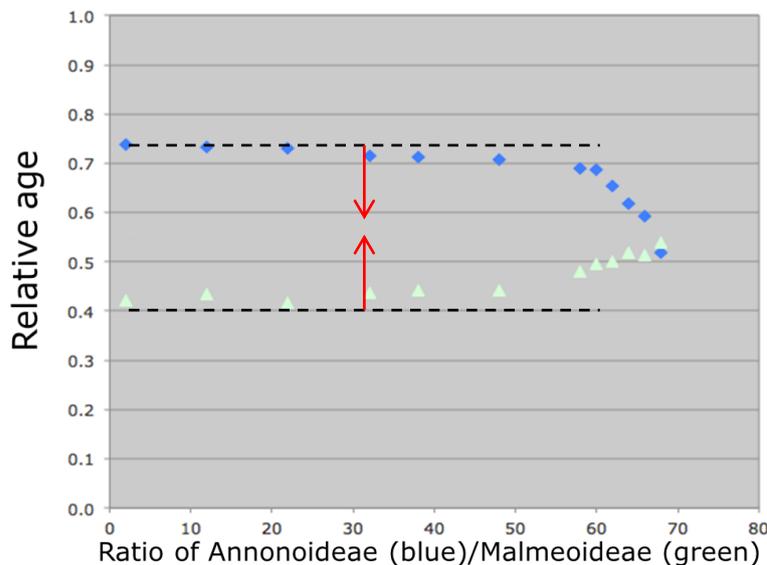


Figure 7: Relative age against ratio of Annonoideae (blue, LBC) and Malmeoideae (green, SBC), graph provided by L.W. Chatrou

5. Materials and methods

A data set was constructed combining datasets from different publications. Some trees were constructed with RAXML to check the resolution and topology. These trees were also used to determine the species belonging to the clades of which calibration fossils are available for the BEAST analyses. A reduced dataset with four markers was used for different dating analyses with different fossil calibration priors, leaving out one at a time, to see how the priors influenced the node ages and branch lengths. Also, multiple dating analyses with different in LBC/SBC species ratio were done to see how the sequence data influences the node ages and branch lengths.

5.1 Construction of the DNA sequence supermatrix

As a basis, the dataset of Chatrou *et al.*, (2012) was used. New sequences were added mainly for species from the genera *Guatteria*, *Annona*, *Duguetia* and some species from other genera. These sequences were provided by the authors of the following articles: Chaowasku *et al.*, (2012); Chatrou *et al.*, (2009); Erkens *et al.*, (2007); Mols *et al.*, (2004); Pirie *et al.*, (2005). Sequences were also added directly from GenBank. The GenBank codes and references can be found in appendix 1.

The main goal during the construction of the total dataset was to get as much sequence data for the different markers and different species as possible without too much data missing. Not all species had all markers, species with only one or two sequences are not included in the dataset.

The markers used in this thesis are seven of the eight markers originally used in Chatrou *et al.*, (2012). *trnT-trnL* was excluded since the alignment of this marker contains many indels. Alignments with much indels are less informative for branch length estimation. The markers used for this thesis thus are *matK*, *ndhF*, *rbcl*, *trnL-trnF*, *psbA-trnH*, *atpB-rbcl* and *trnS-trnG*. These are coding as well as non-coding genes, respectively the first three and last four markers. This should avoid the appearance of short branches and long branches due to single mechanisms of evolution in genes.

The sequences were manually aligned to the already existing alignments of Chatrou *et al.*, (2012) in Mesquite v. 2.74 (Maddison & Maddison, 2011). Indels were not coded as separate characters because their evolution cannot be described by parametric models (Couvreur *et al.*, 2011; Sjödin *et al.*, 2010). Ambiguous indels, microsatellites and characters were removed. This resulted in the following lengths of alignments: *matK*; 788 bp, *ndhF*; 1832 bp, *rbcl*; 1350 bp, *trnL-trnF*; 1047 bp, *psbA-trnH*; 386 bp, *atpB-rbcl*; 823 bp and *trnS-trnG*; 768 bp. Finally, the marker alignments were concatenated with SequenceMatrix (Vaidya *et al.*, 2014), resulting in a supermatrix of 6994 bp and a total of 457 species. The estimated proportion of invariant sites of the total database is 0.2930 (SplitsTree; Huson & Bryant, 2006).

From this total dataset (referred to as 'Full') a smaller dataset was made in order to reduce computational hours and increase data coverage. This smaller dataset (referred to as '4M') included the four markers *matK*, *ndhF*, *psbA-trnH* and *trnL-trnF*, and included only the species with all four markers present in the dataset. The 4M dataset consisted of 252 species and had 3449 bp sequence length.

Table 1 shows the numbers of species and genera currently described in Annonaceae (Chatrou *et al.*, 2012), the number of species and genera in

Table 1: Numbers of species and genera of Annonaceae sampled in the 4M and full dataset per clade and the total of described and accepted species and genera (Chatrou *et al.*, 2012).

| | Species | | | Genera | | |
|-------------------|---------|------|-------|--------|------|-------|
| | 4M | Full | Total | 4M | Full | Total |
| <i>Anaxagorea</i> | 2 | 4 | 30 | 1 | 1 | 1 |
| Ambavioidae | 5 | 10 | 52 | 5 | 9 | 9 |
| Annonoideae (LBC) | 170 | 225 | 1.201 | 31 | 38 | 40 |
| Malmeoideae (SBC) | 75 | 191 | 732 | 37 | 43 | 47 |
| Total | 252 | 457 | 2.015 | 74 | 91 | 97 |

the reduced 4M dataset which is used for the dating analyses and the full dataset.

5.2 Finding the best possible partition scheme and models

To find the best partition scheme, I used PartitionFinder (Lanfear *et al.*, 2012). For the PartitionFinder analyses only those species with the complete set of markers were used. For species lists for each partition scheme, see the species table with GenBank numbers in appendix 1. First, the species with a complete set of all seven markers (referred to as dataset '7M'; consisting of 55 species and 6805 bp) were used to perform an overall analysis. Second, an analysis for the 4M dataset was run to be able to run analyses with more species.

The different schemes were compared by their Bayesian Information Criterion (BIC), as communicated by L.W. Chatrou with Robert Lanfear on 11-12-2013. The log likelihood cannot be used for comparing the fit of different models, since it will always go up when more parameters are used. This would cause over-fitting; the model would not anymore describe an actual process, but just adapt to the data points until it has found a perfect fit. Therefore the BIC, AIC or AICc should be used to compare the 'relative goodness' of a model. These criteria are calculating the goodness of a model using the lnL and using a penalty for extra parameters (amongst others: J. Sullivan & Joyce, (2005)). The BIC is in this case preferred because of a severe penalty on extra parameters and the BIC works better on large concatenated genomic datasets than the AIC (Groussin *et al.*, 2013).

5.2.1 Different analyses for LBC and SBC datasets and comparison

Five different partition schemes for each dataset were used (see also table 2 for the different analyses), to make sure that the most partitioned starting scheme results in the best partition scheme found by PartitionFinder. The different partitions I tested were partitioned by only the genes (for example: *matK*(1+2+3) (1)), all three substitution sites for all coding markers separately (*matK*(1,2,3) (2)), the first two substitution sites together and the third substitution site separately for each marker (*matK*(1+2,3) (3)), all three substitution sites together for all coding markers (*matK* + *rbcL*(1,2,3) (4)) and the first two substitution sites together for all markers together (for example: *matK* + *rbcL*(1,2+3) (5)). Noncoding markers were tested for nucleotide position partitions. In table 1 of appendix 2 the different starting schemes can be found. To find the codon positions of the coding genes, I used the Mesquite options 'set codon positions' and 'least stop codons'. Indeed, the most partitioned scheme always gave the lowest BIC (table 2).

I also ran PartitionFinder for four markers for the long branch clade and the short branch clade separately (definition of the clades according to Richardson *et al.*, (2004)), datasets referred to as dataset LBC and SBC, respectively). Since rate heterogeneity is expected, the dataset might need different models for each clade (the SBC and LBC). Although it is not possible to use a partition of sequences based on species, the result might give us a better understanding of how our data is behaving.

Table 2: partitions (rows) tested for all datasets (columns) with BIC compared per dataset; best BIC underlined, starting partitions and best resulting partitions with their models can be found in appendix 2

| | 7M | 4M | 4M: LBC | 4M: SBC |
|-------------------------------------|--------------|--------------|----------------|----------------|
| Whole genes (1) | 85247 | 69779 | 39823 | 25551 |
| All substitution sites (2) | <u>84677</u> | <u>69217</u> | <u>39498</u> | <u>25387</u> |
| Two substitution sites (3) | 84681 | 69788 | 39518 | 25404 |
| All substitution sites combined (4) | 85144 | 69749 | 39750 | 25664 |
| Two substitution sites combined (5) | 85146 | 69276 | 39763 | 25664 |

5.2.2 PartitionFinder settings and results

The outgroup taxa were removed from the datasets for these analyses since the outgroup is not as densely sampled as the ingroup and the outgroup taxa are only present because of their information about phylogenetic relations and ages. Including them would compromise the choice of the right model for the rest of the data.

Branch lengths were linked since the objective is a species tree based on the evolution of different markers. We therefore assume that the markers all evolve according to an underlying species tree (Lanfear *et al.*, 2012). BIC was used as model selection criterion. The search algorithm was set to 'greedy', since the partitioning often resulted in more than 12 data blocks. All analyses were first done for all models, later for only the models used in BEAST. The BEAST models did not fit the data best. For an overview of all partitions tested for each dataset see table 1 of appendix 2, the resulting best partitions of each analysis and their log likelihood (lnL) and BIC (both datasets for all models and for BEAST models) can be found in tables 2A-3B in appendix 2. The two best partitions for the 4M dataset of LBC and SBC can be found in table 5 in the results chapter.

5.3 Phylogenetic tree construction

To run a dating analysis in BEAST, a prior of age of the MRCA for at least one group is needed. In order to know what species are in a (monophyletic) group, one needs a phylogenetic tree. RAxML (Stamatakis, 2006; Stamatakis *et al.*, 2008), available through the Cipres Science Gateway (Miller *et al.*, 2010), was used to construct the initial topology onto which the fossils can be placed. It was run with an un-partitioned dataset. No PartitionFinder analysis was run for the evolutionary models used within RAxML. With RAxML a maximum likelihood search was performed, with estimation of the proportion of invariable sites and a rapid bootstrap analysis was conducted for the best-scoring ML tree in a single programme run, using the same method as Chatrou *et al.*, (2012). This was done for both the total and 4M dataset including the outgroup. For the resulting topologies see figures 1 and 2 of appendix 3.

5.4 Restriction of nodes with fossils

Fossils are placed within their crown group, providing an age for the stem node of the clade they are placed in (Doyle & Donoghue, 1993). As explained in the introduction, the fossils dating the lower nodes in the phylogeny will only be used as hard minimum bounds. But we also need a maximum bound for the fossil near the root in order to prevent the root age to become too old. In this case only hard bounds were used, because no sufficient information nor phylogenetic arguments could be found to construct priors otherwise. Table 3 shows a compilation of fossils used in this thesis. Two of those are fossils which have not been used in any dating analysis before. These fossils were identified by L.W. Chatrou, S. Manchester and G. Stull.

Table 3: Overview of fossils used in this thesis, with a short description, age, node on which a prior based on the age of the fossil will be placed and the publications which describe and place the fossils

| Fossil | Description | Age | Prior node | Publications |
|---------------------------------|----------------------------------|---|--|-------------------------------------|
| <i>Archaeanthus</i> (1) | multifollicular angiosperm fruit | 98 Mya (Late Albian) | Stem node <i>Magnoliaceae</i> | Dilcher & Crane, 1984 |
| <i>Futabanthus</i> (2) | flower | 89 Mya (Early Coniacian, Late Cretaceous) | Crown node <i>Annonaceae</i> | Takahashi, Friis <i>et al.</i> 2008 |
| <i>Anonaspermum commune</i> (3) | seed | 65.0-47.8 Mya (Early Eocene, London Clay) | Crown node <i>Annonoideae</i> | Reid & Chandler, 1933 |
| <i>Duquetia</i> (4) | leaf | 41.2-47.8 Mya (Middle Eocene, Claiborne clay pits) | Crown node <i>Duquetia</i> | Roth, 1981 |
| African Malmeoideae (5) | seed | 33.9-28.1 Mya (Early Oligocene, Fayoum Egypt) | Crown node African <i>Malmeoideae</i> clade | - |
| <i>Mosannonna/Oxandra</i> (6) | seed | 23.0-15.97 Mya (Early Miocene) | Stem node <i>Mosanonna</i> and <i>Oxandra</i> clades | - |

5.4.1 Arguments excluding some often used fossils from this research

The fossil *Lethomasites* is excluded from this research, although it provides an older estimate for the minimum age of the same node as does *Futabanthus*. The phylogenetic relation of *Lethomasites* to the Annonaceae is debated. The fossil pollen of *Lethomasites* (monosulcate pollen type with granular exine structure) cannot be placed within the Magnoliales with great confidence. The fossil is thought to be positioned near Annonaceae and the three Australian genera. Around the time of divergence of this lineage exine structure changed from columellar to granular (Doyle & Endress, 2000; Sauquet *et al.*, 2003). To be able to place *Lethomasites* with more confidence, other organs should be available (Doyle *et al.*, 2004).

A fossil leaf believed to be from the genus *Alphonsea* is left out because the argumentation of the researchers on placing this fossil as *Alphonsea* is not convincing. Their argument is based on leaf shape, venation and the geography of the site the fossil was found (Srivastava & Mehrotra, 2013). On the same characteristics this fossil could be placed in a different group altogether (personal communication with L.W. Chatrou, 2013).

5.4.2 Fossils providing deep node ages in the Annonaceae phylogeny

The fossil *Archaeanthus linnenbergeri* (see for a reconstruction figure 8) has many features in common with extant Magnoliidae. The flower of this fossil shows distinctive derived stipules and an elongated receptacle. The fruits have numerous well-spaced follicles. These are synapomorphies within the *Magnoliaceae*. Therefore, this fossil can be placed among the descendants of the most recent common ancestor (MRCA) of the *Magnoliaceae* and *Annonaceae* (Dilcher & Crane, 1984). It could be placed differently (and more accurately), were it not that the outgroup species from Himantandraceae and Degeneriaceae were excluded from the database after the first RAxML tree construction, since their placement was debatable and their inclusion could possibly compromise the dating analysis. *Archaeanthus* can be used as a minimum age for the stem node of *Magnoliaceae* (see for fossil placements and priors for the RAxML 4G phylogenetic tree (figure 14). *Endressinia brasiliiana* (Mohr & Bernardes-de-Oliveira, 2004) was used as a maximum bound for this same node. The method with which *Endressinia* is assigned to *Eupomatiaceae* has been found unreliable by some authors, since the characters of this

magnolialean fossil flower are not restricted to Eupomatiaceae (Crepet, 2008). But along with it, there is quite some evidence that around the age of *Endressinia* there was indeed a split off in the Magnoliaceae, which provides us with a maximum bound for this node.

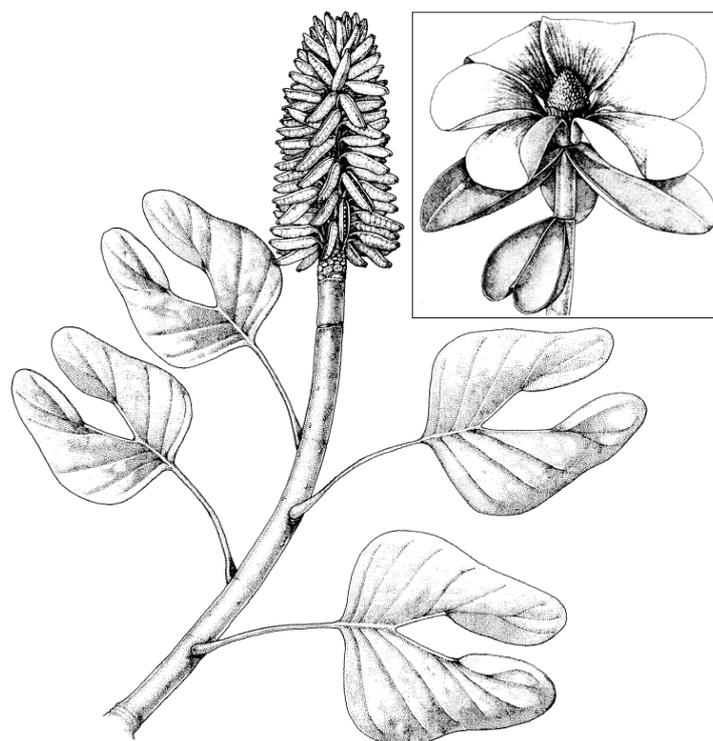


Figure 8: *Archaeanthus linnenbergi*; reconstruction of a flower (inset) and a leafy twig bearing a multifollicular axis (from: Dilcher *et al.*, 1984)

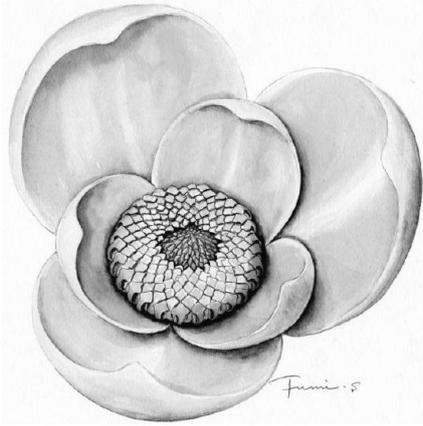


Figure 9: *Futabanthus asamigawaensis*; reconstruction of a flower with opened tepals to show androecium and gynoecium (from: Takahashi *et al.*, 2008)

The *Futabanthus asamigawaensis* fossil (see for a reconstruction figure 9) can be used to date a second deep node. This fossil is added to the magnoliid clade because of the general shape of the flower (numerous stamens and carpels, the receptacle is flattened and disk-like and bears a small number of tepals around the rims). The typical androecium and stamens proves a close affinity of *Futabanthus* to Annonaceae. The stamens of *Futabanthus* and the shape of the connective protrusion place the fossil near the base of the Annonaceae (Takahashi *et al.*, 2008). Stamminoides are present in all Annonaceae except in *Anaxagorea*. Stamminoides also occur elsewhere in the Magnoliales; in Eupomatiaceae, Degeneriaceae and Himantandraceae. Adding these characteristics up, *Futabanthus* is related to the phylogenetic backbone between the stem and crown node of the extant Annonaceae, providing a prior on the crown node of the Annonaceae.

5.4.3 Fossils providing shallow node ages in the Annonaceae phylogeny

The *Anonaspermum commune* (figure 10) seed from the Early Eocene of the London Clay can be used as a constraint on the *Annonoideae* group. This seed shows a typical lamellar structure on the seed skin. It has a raphe (r) around the seed which continues around the hylum (h). This is a feature present throughout the Annonoideae with the exception of Bocageae, which justifies a minimum constraint on the Annonoideae crown node.

A *Duguetia* leaf (Roth, 1981) was studied by L.W. Chatrou in February 2013 in California. When studying the organic remnants of this fossil microscopically, he found stellate and peltate trichomes (figure 11). This, together with the elliptic form of the lamina and pinnate venation (Dilcher & Lott, 2005), are characteristics indicating that this is a fossil leaf of a *Duguetia* spp. African *Duguetia* species (for example *D. staudtii* and *D. confinis*) have stellate hairs on their leaves. Neotropical *Duguetia* species have stellate or scale-like hairs on their leaves. Leaf hair type, size and density are the most valuable characteristics for species grouping within *Duguetia* (Chatrou, 1998). Since these two geographical clades make up the total of *Duguetia* species (Erkens *et al.*, 2009) this fossil can be used as a prior on the crown node of *Duguetia*.

Two undescribed fossils which were investigated by L.W. Chatrou will be used in this research. The first is a *Mosannonna* or *Oxandra* seed (figure 12) from the Early Oligocene. It has a typical seed skin with a rippled structure and there is indication of the flower being androdioecious. Since these features are present in both clades (Chatrou, 1998), the seed should be placed somewhere in between *Mosannonna* and *Oxandra*. Because of the two genera it could be assigned to, the fossil is referred to as '*Mosoxandra*', which name will be used further in this thesis. Since the placement cannot be made

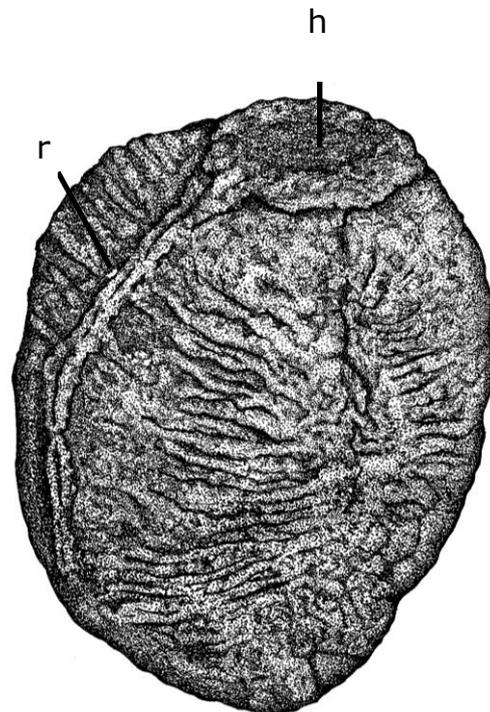


Figure 10: *Anonaspermum commune* (Reid *et al.*, 1933) from Early Eocene of the London Clay, with raphe (r) and hylum (h). Drawing by I. van Riemsdijk based on picture from (Reid *et al.*, 1933)

more exact, this fossil will be used as a minimum constraint on the stem node of the two sister clades.

The second undescribed fossil is an African Malmeoideae seed (figure 13) from the Early Oligocene of Fayoum in Egypt. A very typical feature of this fossil is its wart-like seed skin, which makes it suitable for a minimum constraint for the crown node of the African Malmeoideae. This clade is equal to the tribe *Piptostigmateteae* with exception of the genera *Annickia*, of which all species have seeds with these characteristics.

Concluding (figure 14), two fossils are used for calibrations for deep nodes (*Archaeanthus* and *Futabanthus*) and one to constrain the ages of these nodes of being estimated too old (*Endressinia*). There are two fossil priors placed within the SBC (the minimum constraints of the African Malmeoideae seed and the '*Mosoxandra*' seed), a fossil prior placed on the crown node of the LBC (a minimum constraint provided by an *Anonaspernum* seed) and a fossil prior within the LBC (a minimum constraint of a *Duguetia* fossil leaf). To obtain also maximum constraints for the priors, the age of *Archaeanthus* was used.

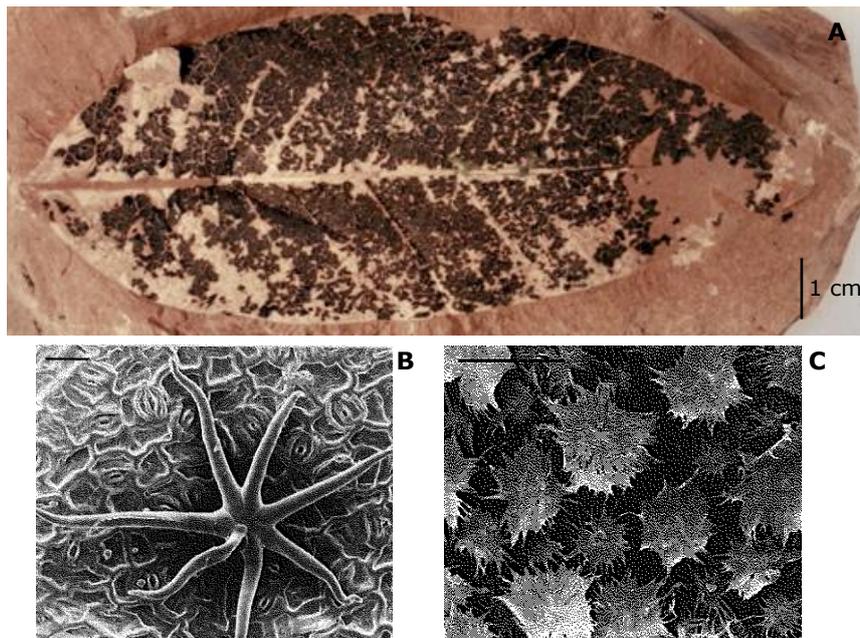


Figure 11: *Duguetia leei* fossil leaf (Roth, 1981) (A) with examples of *Duguetia steltate* (B, bar = 20 μm) and scale-like (C, bar = 100 μm) leaf trichomes from *D. riberensis* and *D. furfuracea* respectively (pictures from Chatrou, 1998)

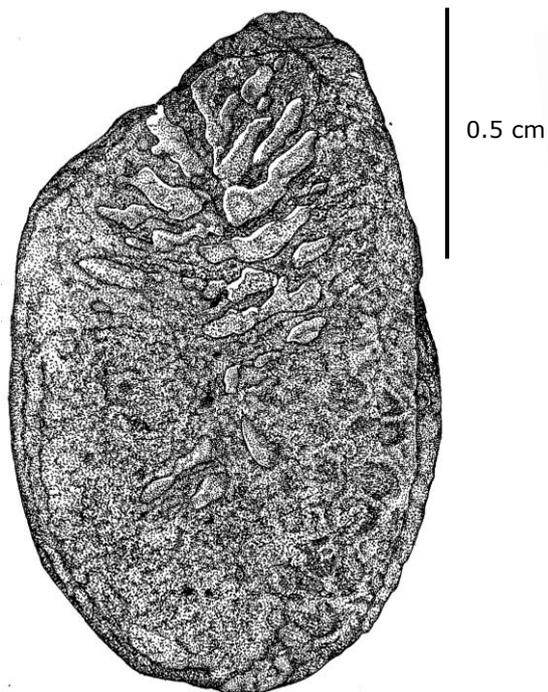


Figure 12: '*Mosoxandra*' fossil seed (identified as *Mosannonna* or *Oxandra* seed by L.W. Chatrou). Drawing by I. van Riemsdijk from picture provided by L.W. Chatrou

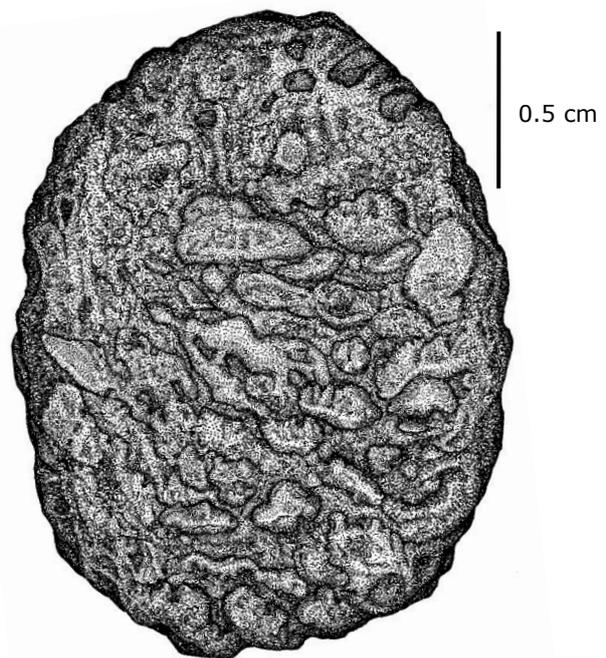


Figure 13: African Malmeoideae fossil seed (identified by L.W. Chatrou) Drawing by I. van Riemsdijk from picture provided by L.W. Chatrou

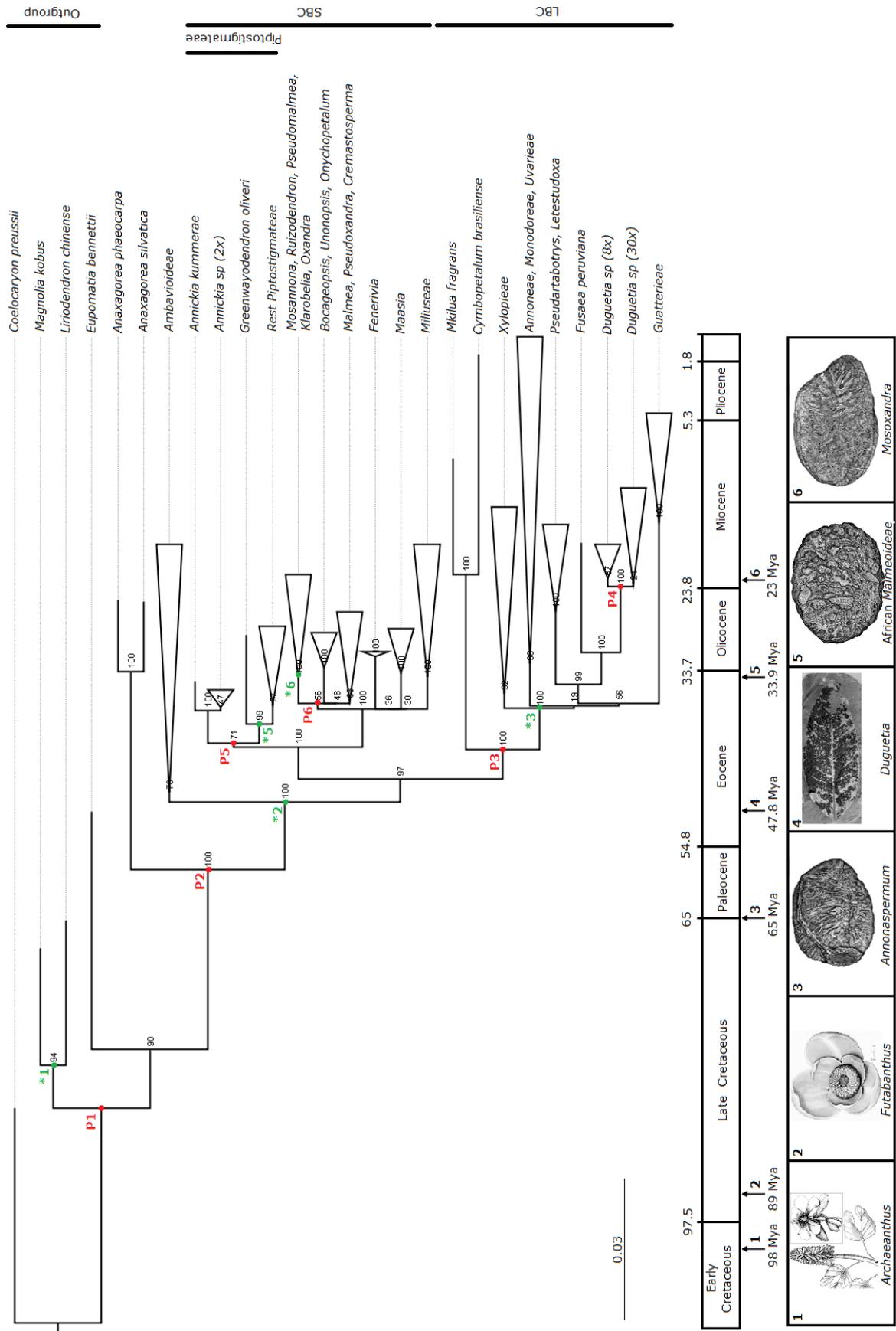


Figure 14: RAXML tree with bootstrap values on the nodes, phylogenetic fossil placements (green and *) and calibration prior placement (red and P), scalebar indicating substitutions along the branchlength, with timescale indicating ages and fossils below

5.5 Dating analyses with the 4M dataset: BEAST

Two types of dating analyses with BEAST were performed (see for theoretical background the introduction). Choice of model within BEAST is based on the PartitionFinder results.

1. The first type of analysis is focussed on the influence of inclusion of age constraints on branch length age estimate in the surrounding phylogeny. The 4M dataset phylogeny will be dated with different combinations of the fossil calibration priors to estimate divergence ages. Since rate heterogeneity is expected, some form of a relaxed clock model should be used. First, all constraints will be included. Subsequently, the same analysis will be performed with the oldest two calibrations, and different combinations of the younger constraints to see whether the branch lengths are influenced by the use of the fossil constraints very much (see table 4). For all analysis the estimated ages of the (constrained or unconstrained) prior nodes are recorded, the age of the deepest node in the tree, the age of the MRCA of the LBC and SBC and the age of the crown node of the SBC were recorded, together with their 95 % HPDs.

2. The second type of analysis is focussed on the influence of the data of the SBC and LBC on the age estimate. In order to see the effect of the different substitution rates in the clades, the amount of species in the dataset from the SBC and LBC respectively will be decreased in two extreme ratios (5:all and all:5). Decreasing the amount of information available for the SBC species respectively the LBC species, reduces the influence of the (presumably) aberrant data on node ages of the surrounding phylogeny.

The xml file for BEAST first was generated in BEAUti v. 1.7.5 (Drummond *et al.*, 2013). Newick starting trees were manually added, because BEAST was not able to immediately fit a tree within the multiple age constraints. These trees were obtained from BEAST runs with soft fossil bounds and the node ages were manually adapted to start within the bounds of the age constraints of this analysis. For each different set of fossils, three different starting trees were used to get different starting points for searching the parameter space.

The substitution models and clock models were unlinked, while the gene trees were linked to obtain an approximation of a species tree. The relaxed lognormal clock model was used for all partitions, in which rates for each branch are drawn from a log-normal distribution.

The taxon sets were as following: ingroup, which is dated by *Archaeanthus*, a group dated by *Futabanthus*, a group dated by the African Malmeoideae and a group dated by *Duquetia*, which were all set to be monophyletic, their bootstrap values from the RAxML analysis were > 70. The groups dated by '*Mosoxandra*' and *Anonaspermum* were set to polyphyletic. The '*Mosoxandra*' node had a bootstrap value of 56. The clade dated by *Anonaspermum* was not monophyletic in the trees that were constructed as starting trees. For an overview of species composition of the groups, see figure 1 in appendix 3. Substitution models were set according to results from PartitionFinder (see table 1, appendix 2). The tree speciation model was set to speciation Yule process (Gernhard, 2008; Yule, 1925). The ages used for the fossil prior nodes are shown in bold in table 3.

Table 4: Fossil combinations with dataset A, the two deep node fossil constraints are always used, indicated by 'basic', the first three letters of the fossil names indicate the calibration priors attached to be included

| Analysis | Fossils | | | | |
|----------|---------|------------|------------|------------|------------|
| | Basic | <i>Mos</i> | <i>Mal</i> | <i>Dug</i> | <i>Ann</i> |
| 1 | X | X | X | X | X |
| 2 | X | | | | |
| 3 | X | | X | X | X |
| 4 | X | X | | X | X |
| 5 | X | X | X | | X |
| 6 | X | X | X | X | |
| 7 | X | X | X | | |
| 8 | X | X | | X | |
| 9 | X | X | | | X |
| 10 | X | | X | X | |
| 11 | X | | X | | X |
| 12 | X | | | X | X |
| 13 | X | X | | | |
| 14 | X | | X | | |
| 15 | X | | | X | |
| 16 | X | | | | X |

The age of *Archaeanthus* was used as a maximum boundary for the calibration priors based on other fossils. Compared to the ages of the other fossils, *Archaeanthus* is much older and therefore not influencing the estimate of the prior nodes. BEAST was set to run 30,000,000 generations and save parameters and trees every 3,000. The Beagle library (Ayres *et al.*, 2012) was not used.

The analyses were run on the Cipres Science Gateway. The analyses with different sets of fossil priors, 16 in total (see table 4) were run two times for each starting tree. This adds up to six runs per fossil prior set, resulting in a total of 96 runs. The analyses all had the same settings as indicated above. The resulting logfiles and time trees were combined with LogCombiner from the BEAST package. The burn-in for the logfiles was estimated by eye in Tracer v. 1.5.0. By combination of the runs in Tracer it appeared that some had not converged to the same stationary distribution as the other runs. When a run had not converged it was removed to obtain high enough effective sample size values (all ESS values > 200) before further analysis. For a record of the different analyses run, whether they converged or not, resulting burn in and other specifics about the Tracer output and combination of different runs see appendix 4.

The trees in the tree files resulting from 30,000,000 generations long runs with trees saved every 3,000 generations, were combined in LogCombiner v. 1.7.5 (with a burn-in of 30,001 and resampling every 15,000) resulting in about 2000 trees for each run. These trees were used to construct a maximum clade credibility tree in TreeAnnotator. The resulting tree was viewed in FigTree and the mean and 95 % HPDs were collected.

For a subset of the analyses, the mean rates of the different markers were obtained from a combined log file. From the fossil exclusion analyses the analysis with only the LBC calibration priors and the analysis with only the SBC calibration priors was used. The two analyses with different LBC:SBC ratios were included and the analysis with the total dataset and the total set of calibration priors was used. The latter was included to be able to compare the different analyses to. This log file was constructed in LogCombiner by removing the burn in as recorded in appendix 4. The frequency of the mean rate of the different markers was then obtained by defining bins in Excel.

6. Results

6.1 PartitionFinder: models for the LBC and SBC

When analysing the different possibilities of data partitioning it should, in theory, be possible to find a difference between the models and/or partitions which are the most fit for the LBC and SBC. In table 5 the data partitions are shown with the corresponding models found by PartitionFinder. Most models were found to be the same for LBC and SBC. Three partitions were found to have (slightly) different best models, the third codon position of *matK*, *rbcl* and the second codon position of *rbcl*. Apparently, *matK* and *rbcl* (third and first codon position) are best described by (slightly) different models (grey rows). One should keep in mind that, in this setup, when two partitions behave according to the same model it does not necessarily mean they have the same parameter values and thus can be described by identical models with the same parameters.

Table 5: the best partitions as found by PartitionFinder with best fitting models, for the 4M analysis of the LBC and SBC, $\ln L(\text{LBC}) = -18382.88$, $\text{BIC}(\text{LBC}) = 39497.96$, $\ln L(\text{SBC}) = -12018.98$, and $\text{BIC}(\text{SBC}) = 25387.45$, differences in model choice highlighted in grey

| LBC | | SBC | |
|--------------------------|---------|--------------------------|-----------|
| Partition | Model | Partition | Model |
| matK_3, rbcl_3 | K81uf+G | matK_3, rbcl_3 | K81uf+I+G |
| matK_1, matK_2, trnLF | K81uf+G | matK_1, matK_2, trnLF | K81uf+G |
| psbAtrnH | HKY+G | psbAtrnH | HKY+G |
| rbcl_1 | TVM+I+G | rbcl_1 | HKY+I+G |
| rbcl_2 | JC+I+G | rbcl_2 | JC+I+G |

6.2 RAxML bootstrap phylogenetic trees

In the first phylogenetic analyses in RAxML the relationships among the outgroup species were not resolved very well (bootstrap values of 24-39). This was problematic since the fossils *Archaeanthus* and *Futabanthus* needed to be placed on these nodes. Therefore we decided to exclude some species which were not critical for the placement of the *Archaeanthus* or *Futabanthus* fossil from the outgroup, namely *Degeneria vitiensis*, *Galbulimima belgraveana* and *Persea americana*, see also appendix 1. Again RAxML bootstrap analyses were run for both datasets, resulting in a better resolved outgroup. The new phylogeny of the RAxML bootstrap analysis of the 4M dataset as well as the total dataset resulted in a tree resembling the main topology of (Chatrou, *et al.*, 2012). The nodes that were calibrated had high confidences (56-100). In figure 1 of appendix 3 the total phylogenetic tree of the 4M dataset can be found, with the LBC and SBC indicated and the phylogenetic fossil placements (indicated with *), the prior placements (indicated with P) and the numbers corresponding to table 4 in Chapter 5. A simplified version of this tree can be found in figure 14. The 7M dataset RAxML tree can be found in figure 2 of appendix 3.

6.3 BEAST analyses results

The BEAST analyses were performed on the basis of two approaches. One was leaving out calibration priors in order to investigate their effect on the node ages elsewhere in the tree and the mean substitution rates for each marker. The second was leaving out part of the sequence data (either the majority of the LBC sequences or the majority of the SBC data).

Most analysis converged after 30,000,000 generations. Sometimes one run from the same starting tree (so starting from the same spot in the parameter space) did and one did not converge to the same parameter level as the other run. This indicates that when

an analysis would have been allowed to run even longer, the unconverged ones would also have reached the same level of likelihood as the other runs did. Unconverged runs (with lower likelihoods) were excluded. Combining the converged runs nearly always resulted in Tracer combinations with ESS values > 200 , as is recorded in appendix 4. The parameters which did not converge are also recorded in appendix 4. The most common unconverged parameters are the covariance of *psbA-trnH*, the *rbcl* and *trnL-trnF* tree likelihoods, *rbcl* CP 1 tree likelihood and the *rbcl* standard deviation of the uncorrelated log-normal relaxed clock (*uclid.stdev*).

Analysis 12 was the only analysis with which no combination of the different runs resulted in Tracer results with ESS values > 200 . A combination was chosen where the prior trace ESS was still below 200, but the traces of the parameters did not show big jumps to other likelihood levels. This specific analysis contains the two LBC fossil priors, but not the ones from the SBC. This raises the question whether this might be a case of conflict between the calibration priors and sequence data.

6.3.1 Node age estimates of the fossil exclusion and LBC:SBC ratio analyses

In table 6 the estimated node ages as indicated in figure 15 are shown, together with the values of the borders of their 95 % HPDs. Figure 15 first shows the phylogenetic tree (A) with the locations of the nodes of table 6 with graphs of the age estimates as red dots with their 95 % HPDs as a black line. The values at the nodes are the posterior values and the x axis shows time (My). Figure 15B shows the different analysis with the constrained nodes in blue and the analysis with the LBC:SBC ratio data sets striped; horizontally for the dataset with nearly only SBC sequences and vertically for nearly only the LBC sequences. Some interesting observations can be made with the help of this figure.

Figure 15C shows the estimated ages of the root height (node A) and the *Archeanthus* prior node (B). The estimates can be divided into three age levels. The first level (± 119 Mya) corresponds to those analysis excluding at least the *Duguetia* calibration prior. The second level (± 122 Mya) corresponds to those analysis including *Duguetia* or including *Duguetia* and using the reduced LBC dataset. The outlier at 127 Mya is the result of the reduced SBC dataset. When the LBC sequences are overrepresented the age is overestimated. The age estimates of the *Archeanthus* prior node are constrained only by their upper bound and also seem to move to younger ages when *Duguetia* is excluded as a calibration prior, synchronous to the age estimates of the root height.

The *Futabanthus* fossil prior node, as opposed to the *Archeanthus* prior node, is most constrained by the lower bound of the calibration prior (figure 15D). There are two levels of estimated ages. The first lies around 91 Mya and corresponds to those analyses in which the *Duguetia* prior is excluded. The second level of estimated ages lies around 95 Mya and is the result of the inclusion of the *Duguetia* prior. The two analysis with the different LBC:SBC ratios now appear to be estimating the age rather old compared to the other analyses with the *Duguetia* prior, the reduced SBC shows a nearly as big difference as do the analyses with the *Duguetia* prior included.

The next node in the tree, the MRCA of the SBC and LBC (D, figure 15E) had no age constraint. It is therefore not surprising it shows larger differences in age estimate (± 15 My). For the analyses with different calibration prior combinations the graph shows three different levels of age estimate. The first level can be seen around 80 Mya, corresponding to those analyses including the *Duguetia* calibration prior, but not all of them, and the ratio analysis including only the LBC sequences. The second level between 75 and 70 Mya corresponds to those analyses excluding the *Duguetia* prior. The third level includes the two older estimates of only *Anonaspermum* excluded (6) and *Anonaspermum* and the African *Malmeoidaea* excluded (8). Also these two nodes show a highly skewed 95% HPD around the estimate, indicating there is some unidentified constraint on that node which prevents it to be estimated even older.

The SBC crown node (E, figure 15F) neither had calibration priors, but might be of interest. The graph seems to show roughly two bands of ages. The first band of

divergence age estimates lies between 50-45 Mya, the result of the inclusion of the *Duguetia* calibration prior, and the second 43-40 Mya. In this case, the exclusion of either the sequences from the SBC or the LBC seem to make no big difference in the age estimate of this node.

The graphs in figure 15(G-I) show the resulting age estimates of the nodes which in some cases are and in some cases are not constrained by their 'own' calibration priors. This means in some cases the dotted line is surpassed by either one of the ends of the 95 % HPD, or even the mean age estimate. First the two calibration prior nodes from the SBC will be discussed, then those of the LBC.

The age estimate of the African Malmeoideae node (figure 15G) shows little change when released from its prior (35-41 Mya). The only case in which the mean age estimate is estimated younger (27 Mya) than the prior age is when all other shallow fossil priors are excluded and only the deep nodes are constrained. The estimates of the mean age when part of the data is missing (either excluding LBC or SBC sequences) are the oldest of all estimates (± 44 Mya), but do not differ much from each other. Both latter estimates are slightly older than the estimate of the first (including all sequences and all fossil priors) analysis.

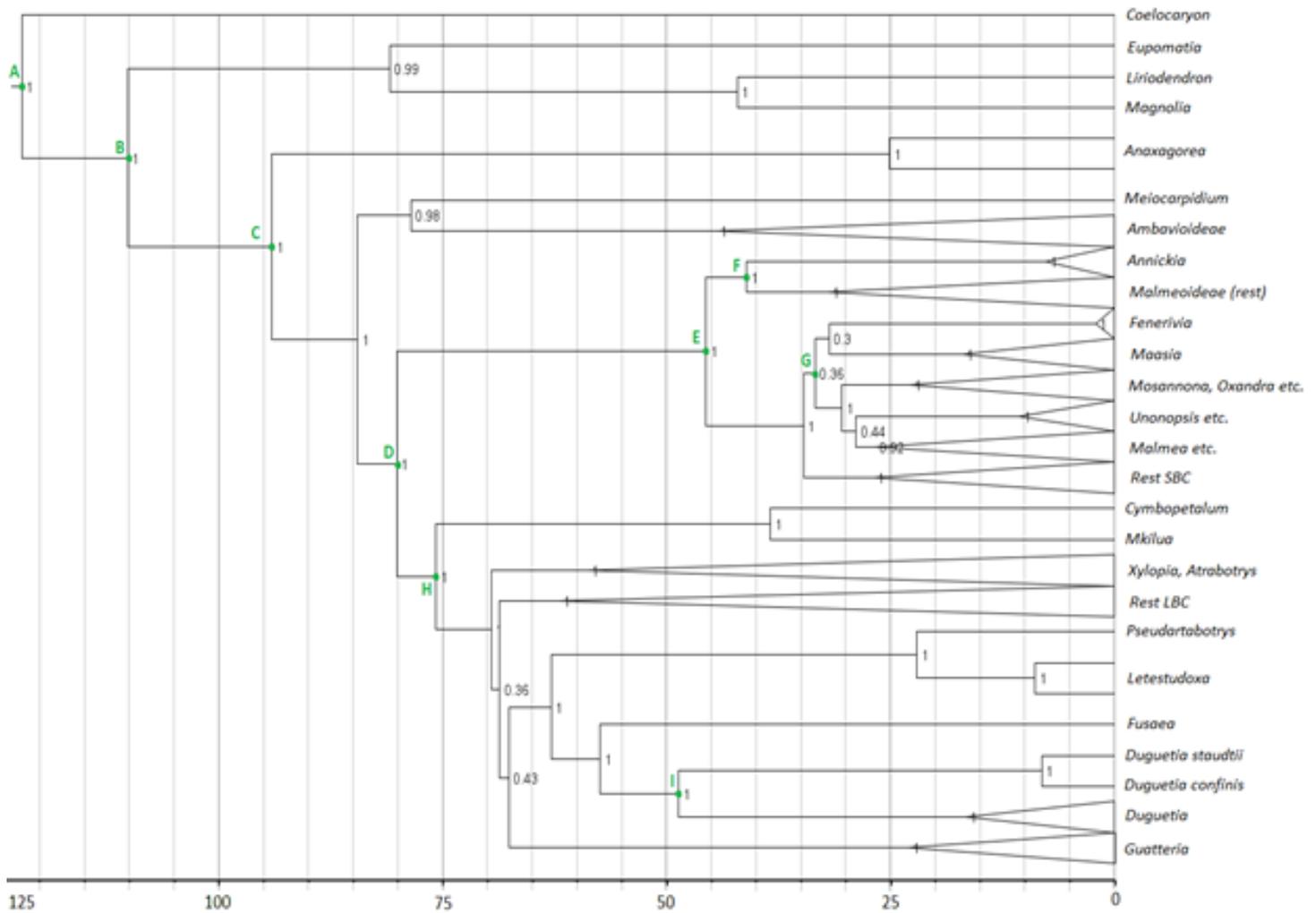
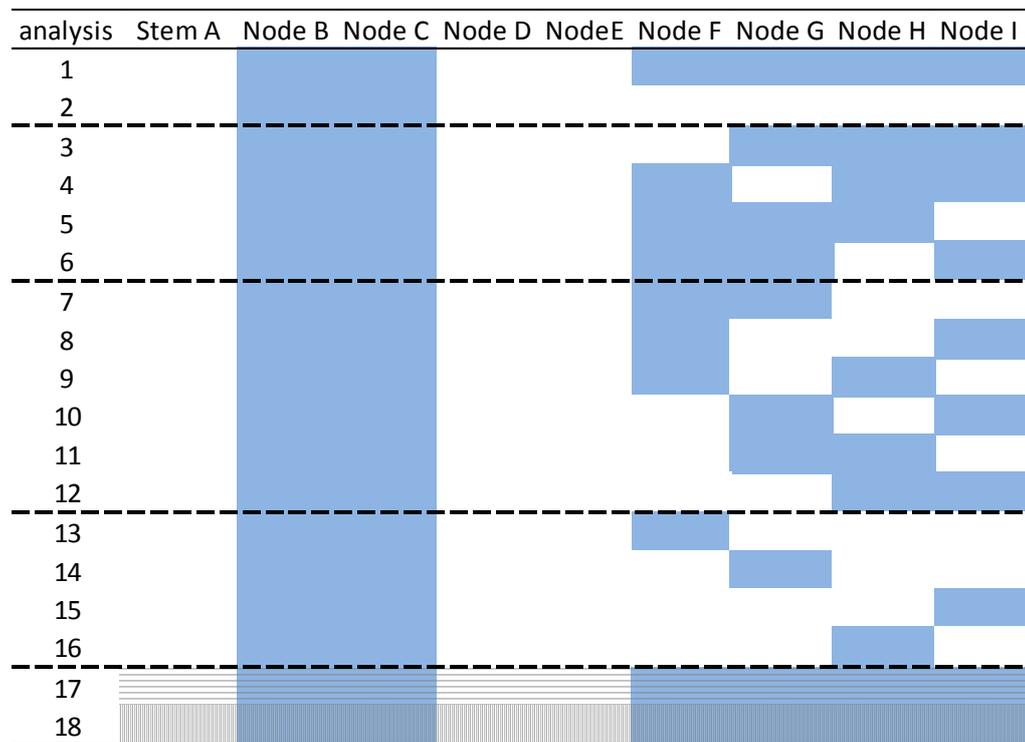
The age estimate of the '*Mosoxandra*' prior node (figure 15H) seems to have been rather little influenced by its' constraint. The 95% HPDs do not touch the prior boundary, except for analyses seven and eighteen. The 95 % HPD intervals of analyses nine and thirteen did also touch the boundary, but did not include the calibration prior in their analyses. These age estimates and 95% HPDs suggest that the prior of '*Mosoxandra*' is uninformative and/or estimating the age of the prior node to be too young. Therefore it probably does not influence the estimate of other ages very much either. This is confirmed by the fact that the other nodes did not show much difference in age estimates when only this calibration prior was removed (f.e. analysis 3). Outliers in the mean age estimate are those of analyses two and twelve. The estimate of the second analysis (including only the deep calibration priors) is youngest compared to the other estimates (25 Mya). When any other calibration prior is included the age of this node is estimated to be older. The outlier of analysis twelve (35 Mya) is caused by the inclusion of only the LBC fossils. The two last analyses with the reduced datasets show that when the LBC sequences are reduced, the age of the '*Mosoxandra*' prior node is estimated to be relatively old (34 Mya). When the SBC sequences are reduced, the age remains comparable to the other analyses (30 Mya).

The LBC crown node (or *Anonaspermum* prior node, figure 15I) deviates from the other results. In most cases where the *Duguetia* fossil node was included, the calibration prior of *Anonaspermum* did not influence its' own mean age estimate (± 76 Mya). The lower bound of the 95% HPD in those cases does not touch the age boundary (65 Mya). Except in the case of analysis seventeen, where only the SBC sequence data and nearly none of the LBC data was included. When the *Duguetia* calibration prior was not included, the mean age of the *Anonaspermum* prior node (70-65 Mya) was being constrained by the lower bound of its' prior. Two of the largest outliers are the age estimates of analyses two and thirteen.

The *Duguetia* crown node with the calibration prior of 47.8 My (dotted line; figure 15J) is the last node discussed here. Throughout the other results, the large influence of this calibration prior already became clear. There is a large difference (± 31 My) between the age estimates of this node for the different analysis. The divergence ages of the *Duguetia* crown node are estimated around 17 Mya for the unconstrained analyses. When constrained, the estimates are all around 49 Mya. Effects of including only SBC or LBC are not visible in this graph. A swift conclusion would be to exclude this prior because it is too influential on the age estimates to remain included in further analyses. Chapter 7 will discuss the implications of this result further.

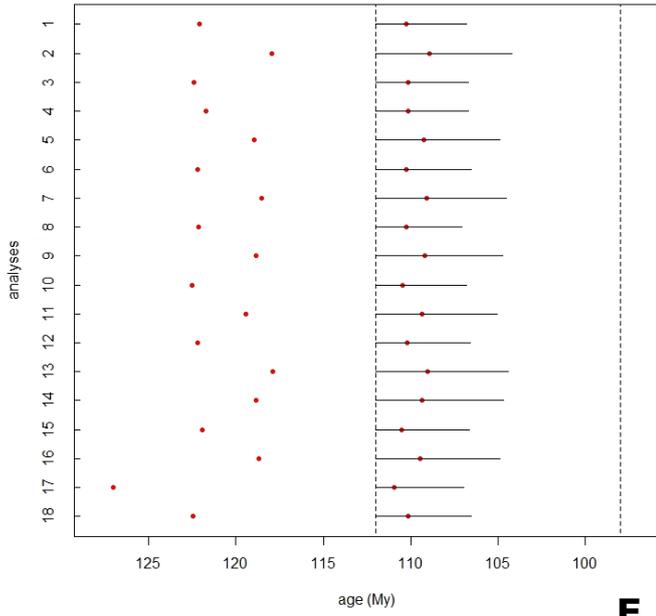
Table 6: The estimated mean node ages (A) of the different BEAST analyses (# corresponding to table 4 in Chapter 5) with the lower (L) and upper (U) boundary ages of the 95% HPDs, ages which are constricted by fossil calibration priors indicated in blue

| # | Estimated nodes | | | | | | | Fossil Prior nodes | | | | | | | | | | | | | | | | | |
|----|-----------------|------------------|-------|-------|---------|-------|-------|-------------------------|--------|--------|------------------------|-------|-------|---------------------------|-------|-------|-----------------------|-------|-------|-------------------------|-------|-------|---------------------|-------|-------|
| | Stemnode (A) | MRCA SBC&LBC (B) | | | SBC (C) | | | <i>Archaeanthus</i> (D) | | | <i>Futabanthus</i> (E) | | | ' <i>Mosoxandra</i> ' (F) | | | <i>Malmeoidae</i> (G) | | | <i>Anonaspermum</i> (H) | | | <i>Duguetia</i> (I) | | |
| | A | A | L | U | A | L | U | A | L | U | A | L | U | A | L | U | A | L | U | A | L | U | A | L | U |
| 1 | 122.06 | 80.17 | 73.93 | 86.33 | 45.65 | 37.12 | 54.61 | 110.24 | 106.78 | 112.00 | 94.13 | 89.01 | 98.79 | 30.47 | 24.95 | 35.98 | 41.12 | 33.91 | 49.10 | 75.76 | 69.82 | 81.91 | 48.71 | 47.80 | 50.40 |
| 2 | 117.97 | 70.51 | 62.03 | 78.23 | 38.85 | 30.78 | 47.21 | 108.94 | 104.21 | 112.00 | 91.30 | 89.00 | 95.12 | 25.37 | 20.34 | 30.65 | 27.04 | 19.74 | 35.21 | 55.72 | 47.80 | 63.65 | 16.14 | 11.34 | 21.68 |
| 3 | 122.40 | 80.07 | 74.07 | 86.39 | 46.04 | 37.31 | 55.36 | 110.15 | 106.70 | 112.00 | 94.17 | 89.02 | 98.93 | 30.79 | 25.14 | 36.80 | 41.42 | 33.92 | 49.64 | 75.69 | 69.63 | 81.63 | 48.65 | 47.80 | 50.28 |
| 4 | 121.73 | 79.91 | 74.00 | 86.52 | 45.19 | 35.49 | 55.46 | 110.14 | 106.72 | 112.00 | 94.04 | 89.02 | 98.71 | 30.29 | 24.36 | 35.85 | 40.19 | 30.60 | 51.49 | 68.12 | 62.46 | 73.65 | 48.65 | 47.80 | 50.34 |
| 5 | 118.97 | 73.70 | 68.56 | 79.24 | 41.75 | 35.33 | 48.86 | 109.26 | 104.88 | 112.00 | 91.68 | 89.00 | 95.70 | 30.07 | 25.42 | 34.61 | 38.13 | 33.90 | 44.30 | 68.84 | 65.00 | 73.62 | 16.88 | 11.84 | 22.64 |
| 6 | 122.20 | 85.04 | 74.41 | 86.77 | 46.28 | 37.74 | 55.32 | 110.27 | 106.54 | 112.00 | 94.75 | 89.01 | 99.25 | 31.14 | 25.50 | 36.75 | 41.71 | 33.90 | 49.53 | 76.29 | 70.12 | 82.26 | 48.84 | 47.80 | 50.41 |
| 7 | 118.55 | 72.05 | 64.86 | 79.78 | 41.47 | 35.18 | 48.87 | 109.11 | 104.55 | 112.00 | 91.43 | 89.00 | 95.35 | 27.39 | 23.02 | 31.54 | 37.96 | 33.90 | 44.17 | 66.95 | 59.52 | 74.32 | 16.81 | 11.59 | 22.28 |
| 8 | 122.16 | 84.28 | 74.11 | 86.05 | 45.82 | 36.11 | 55.67 | 110.25 | 107.07 | 112.00 | 94.12 | 89.07 | 98.64 | 29.69 | 24.37 | 35.41 | 40.70 | 30.99 | 51.10 | 75.60 | 69.49 | 81.37 | 48.66 | 47.80 | 50.22 |
| 9 | 118.83 | 73.34 | 68.11 | 78.93 | 40.58 | 33.05 | 48.65 | 109.22 | 104.72 | 112.00 | 91.60 | 89.00 | 95.61 | 27.14 | 23.04 | 31.30 | 36.14 | 28.19 | 44.67 | 68.50 | 65.00 | 73.38 | 16.97 | 12.03 | 22.39 |
| 10 | 122.51 | 80.17 | 73.48 | 85.88 | 45.80 | 36.49 | 54.50 | 110.45 | 106.78 | 112.00 | 94.24 | 89.02 | 98.59 | 30.73 | 24.49 | 36.61 | 41.25 | 33.91 | 49.07 | 75.74 | 69.15 | 81.41 | 48.90 | 47.80 | 50.26 |
| 11 | 119.46 | 73.50 | 68.11 | 79.12 | 41.99 | 34.79 | 49.17 | 109.35 | 105.08 | 112.00 | 91.74 | 89.00 | 95.82 | 27.60 | 22.29 | 32.34 | 38.30 | 33.90 | 45.06 | 68.74 | 65.00 | 73.63 | 16.89 | 12.32 | 22.75 |
| 12 | 122.18 | 80.44 | 73.58 | 86.24 | 46.31 | 35.93 | 55.77 | 110.18 | 106.59 | 112.00 | 94.51 | 89.03 | 99.37 | 34.39 | 28.33 | 40.85 | 41.13 | 31.06 | 51.50 | 75.96 | 69.88 | 82.41 | 48.62 | 47.80 | 50.25 |
| 13 | 117.87 | 71.30 | 63.21 | 79.02 | 39.19 | 31.32 | 47.17 | 109.03 | 104.43 | 112.00 | 91.37 | 89.00 | 95.09 | 26.54 | 23.02 | 30.64 | 34.94 | 26.41 | 43.50 | 58.79 | 50.92 | 65.90 | 16.46 | 11.02 | 21.86 |
| 14 | 118.87 | 72.08 | 64.02 | 78.91 | 41.65 | 35.47 | 48.62 | 109.35 | 104.67 | 112.00 | 91.59 | 89.00 | 95.04 | 27.26 | 21.54 | 32.12 | 38.08 | 33.90 | 43.74 | 66.98 | 59.17 | 74.65 | 17.01 | 11.51 | 21.93 |
| 15 | 121.91 | 80.61 | 73.90 | 86.74 | 46.17 | 36.28 | 54.88 | 110.51 | 106.62 | 112.00 | 94.87 | 89.04 | 99.23 | 30.87 | 25.51 | 35.89 | 41.15 | 31.35 | 51.00 | 76.26 | 69.29 | 81.91 | 49.04 | 47.80 | 50.33 |
| 16 | 118.69 | 73.55 | 68.39 | 79.29 | 40.45 | 31.49 | 48.79 | 109.45 | 104.90 | 111.99 | 91.89 | 89.00 | 95.82 | 27.35 | 21.94 | 32.16 | 35.96 | 26.82 | 44.52 | 68.73 | 65.00 | 73.36 | 17.40 | 12.35 | 23.06 |
| 17 | 127.02 | 76.62 | 70.69 | 82.51 | 49.59 | 39.89 | 59.96 | 110.94 | 106.94 | 112.00 | 92.90 | 89.00 | 97.48 | 33.53 | 26.60 | 40.08 | 44.40 | 34.30 | 54.10 | 70.55 | 65.02 | 75.83 | 48.76 | 47.80 | 50.65 |
| 18 | 122.48 | 80.08 | 73.84 | 86.29 | 47.31 | 35.74 | 61.11 | 110.14 | 106.56 | 112.00 | 94.25 | 89.01 | 99.00 | 29.33 | 23.00 | 38.12 | 44.20 | 33.90 | 56.83 | 75.78 | 69.59 | 81.84 | 48.71 | 47.80 | 50.50 |

A**B**

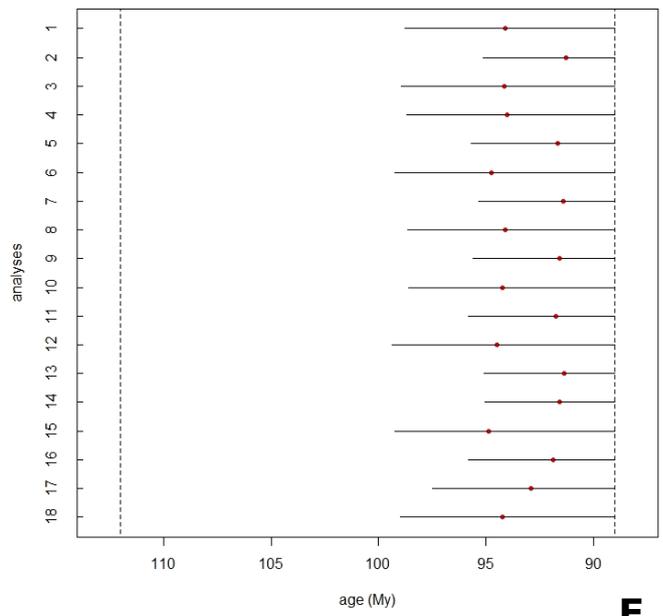
Rootheight (node A) and Archaeanthus node B

C



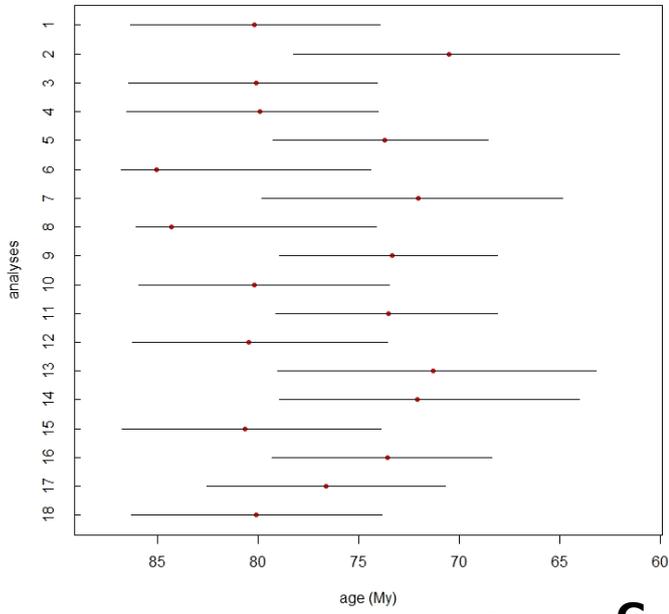
Futabanthus node C

D



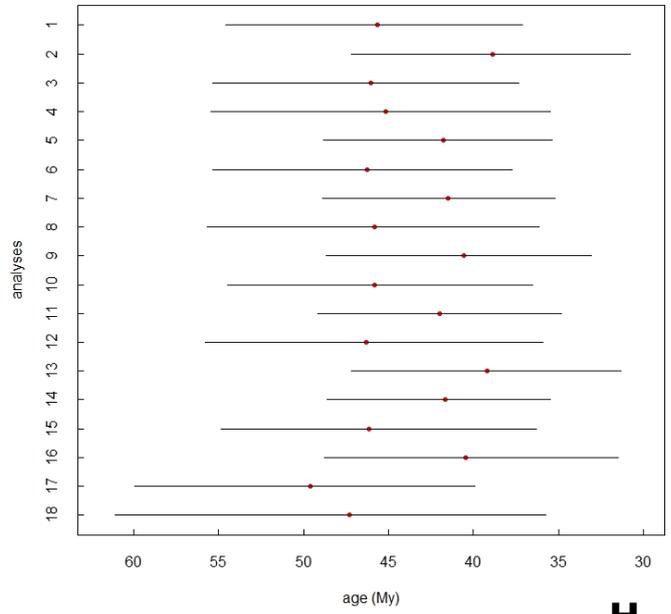
MRCA SBC and LBC node D

E



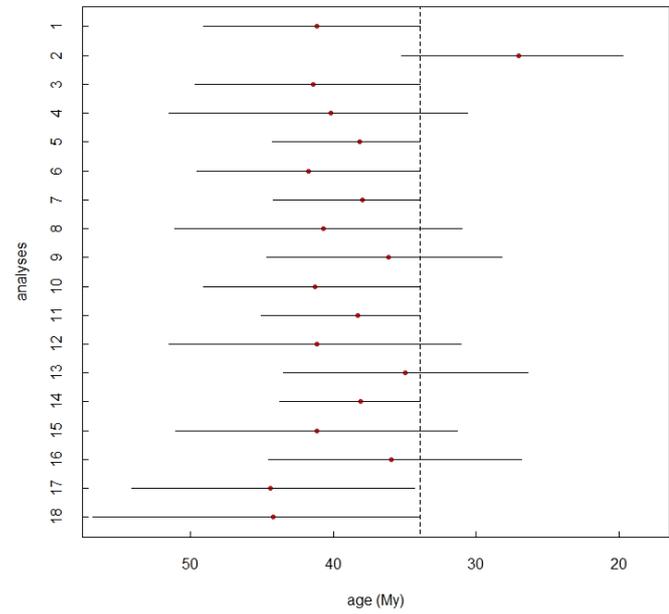
SBC crownnode E

F



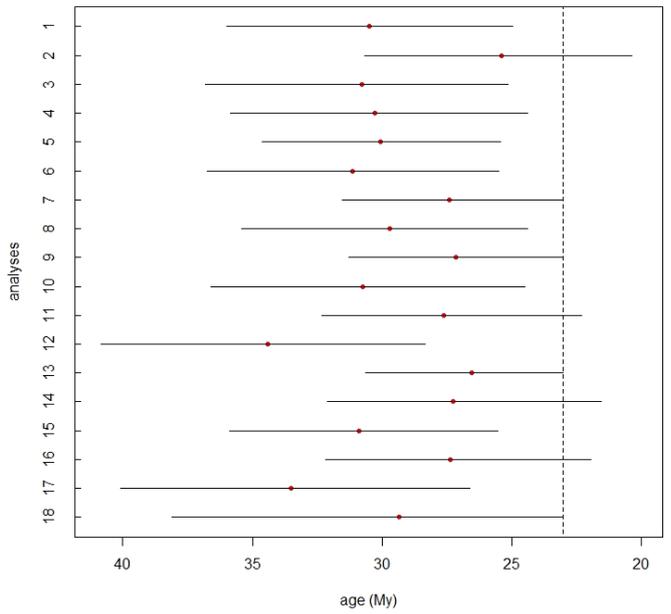
African Malmeoideae node F

G



Mosoxandra node G

H



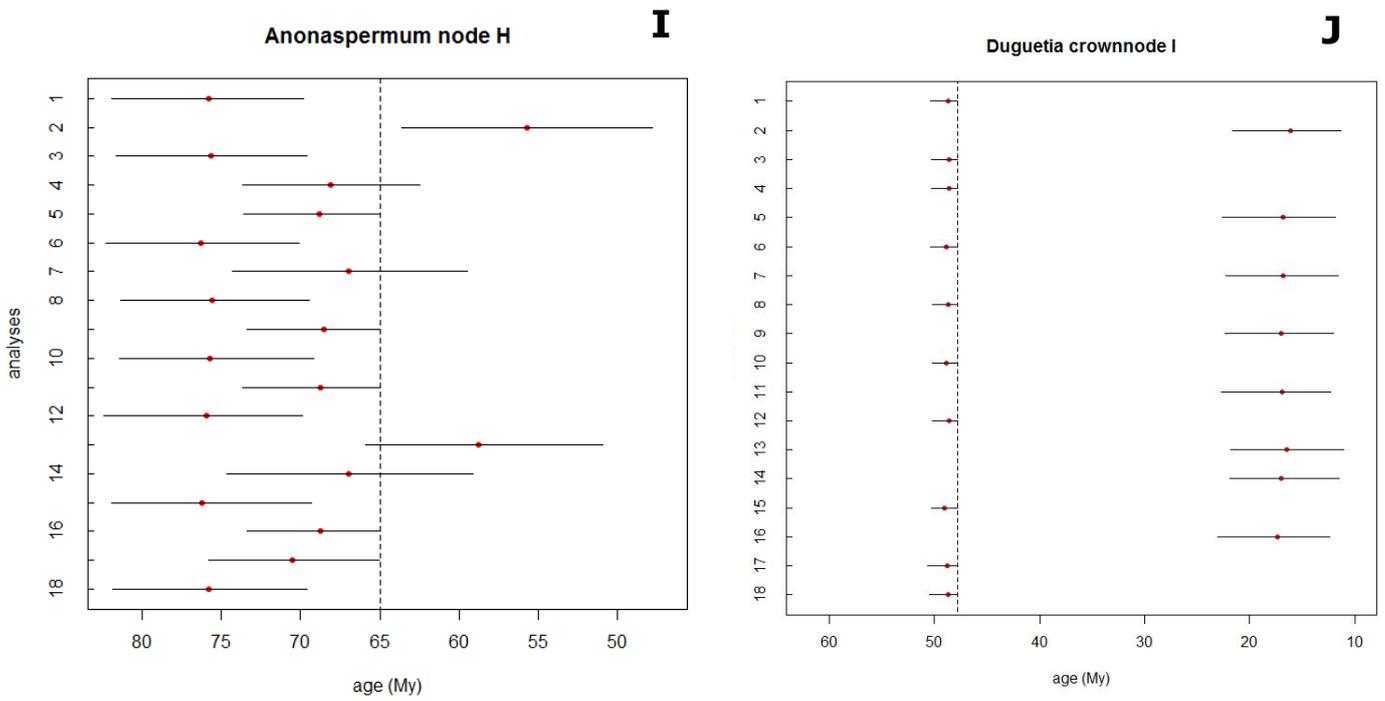


Figure 15: phylogenetic tree (**A**) obtained from analysis one with time on the x axis and posterior values on shown nodes with green dots with letters corresponding to the nodes of the mean age estimate graphs (C-F), with graph titles corresponding to the fossils on which the priors are based; **C:** *Archaeanthus*, **D:** *Futabanthus*, **G:** African *Malmeoideae*, **H:** *Mosoxandra*, **I:** *Anonaspermum* and **J:** *Duguetia*, of which **G, H, I, J** sometimes had no prior. Nodes on which no prior was set are the root node (**B**), the MRCA of the LBC and SBC (**E**) and the crown node of the SBC (**F**). Figure **B** shows an overview of table 6 with prior nodes with constraints per analysis, constrained nodes in blue, horizontal lines (analysis 17) indicate a dataset with a reduced amount of LBC data, vertical lines indicate a dataset with a reduced amount of SBC data. The mean age estimate graphs (C-J) show red dots as estimates of the mean divergence ages of the nodes from the graph title with black lines indicating the range of the 95 % HPD and dotted lines as boundaries of the calibration priors (which are not applied in some of the analyses)

6.3.2 Mean substitution rate estimates for different interesting analyses

For five analyses of special interest the mean substitution rates for the markers were obtained from the log files (figure 17). The five analyses included are three of the analysis with the full dataset and different calibration prior combinations: all fossil priors (1), only the SBC fossils included (7) and only the LBC fossils included (12) and two analyses with all the calibration priors but with reduced LBC data (17) and with the reduced SBC data (18). Numbers refer to the numbering of the analyses in paragraph 6.3.1 and table 4 in Chapter 5. It is important to keep in mind that in the total 4M dataset the species ratios were maintained as they are in real life (see also table 1).

The analysis with the full dataset and all fossil priors (black line, referred to as: 'total analysis') will be used to compare the other analyses to (figure 17). The overall pattern of rate distribution compared to the total analysis is the same for all markers. Depending on the marker this pattern is sometimes more clearly visible. The *rbcL* mean substitution rate differs most from the rates of the concatenated data (figure 17B). The analyses with the different LBC:SBC ratios result in a lower mean substitution rate estimate than the total analysis. The analysis with the small amount of LBC sequence seems to have the lowest estimated mean substitution rate, while the analysis with the small amount of SBC sequence appears to be on the low side, but is more comparable to the mean rate of the total analysis. The analyses with only the SBC and LBC fossils estimate the mean substitution rate for *rbcL* high compared to the total analysis. The analysis with only the SBC fossils has the highest estimated mean rate, while the analysis with only the LBC fossils estimates the mean rate higher than the total analysis does, but lower than the SBC fossils analysis.

For *matK*, *psbA-trnH* and *trnL-trnF* the analysis with the low amount of SBC species as well as the analysis with the LBC fossils appear to estimate the mean rate quite similar to the total analysis. The SBC calibration prior analysis estimate the rates only slightly higher, while the reduced LBC data analysis estimates the mean rates slightly lower than does the total analysis.

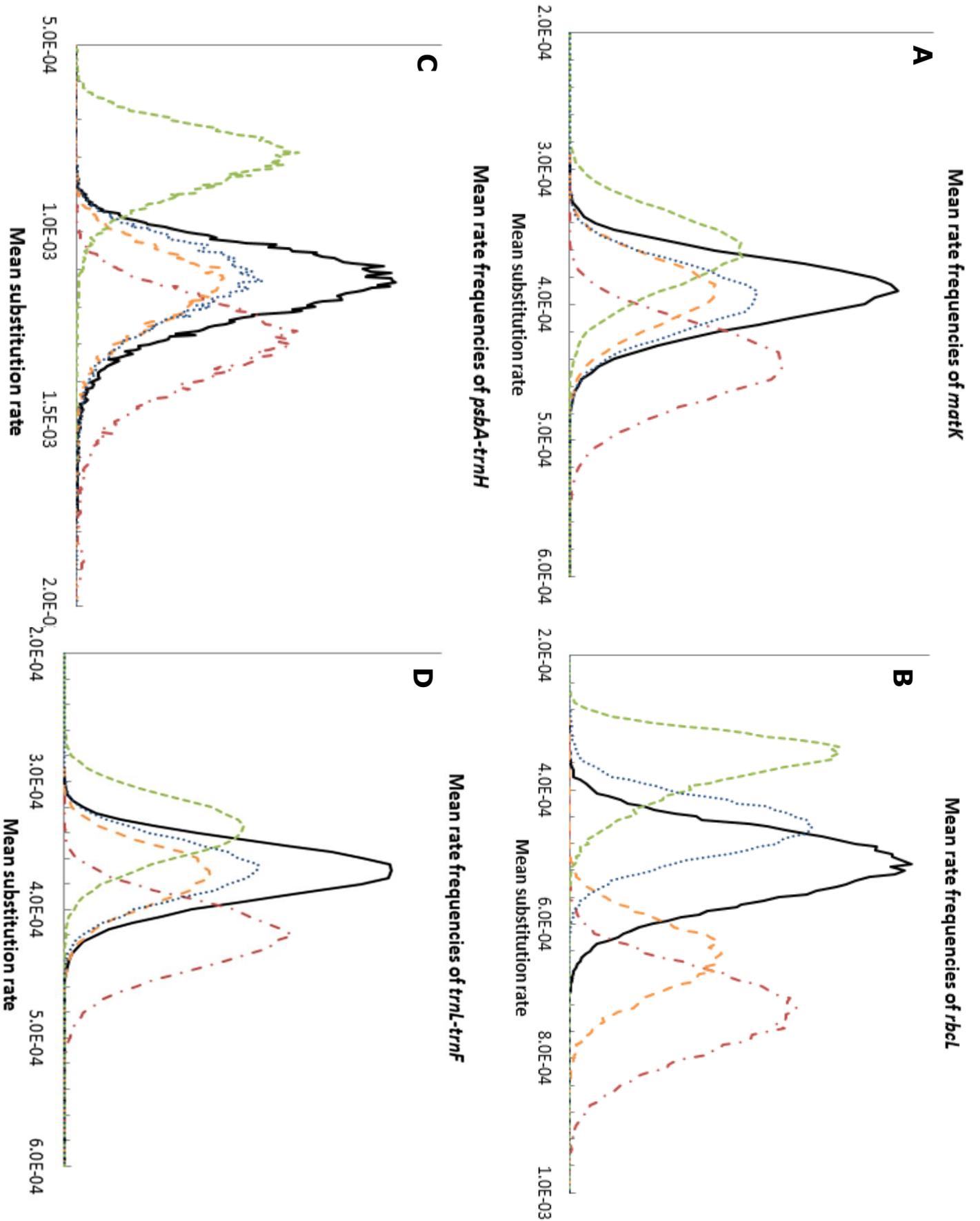


Figure 16: Mean rate frequencies for the four markers: *matK* (A), *rbcL* (B), *psbA-trnH* (C) and *trnL-trnF* (D), for different analyses (all fossils, only LBC fossils, only the SBC fossils, reduced LBC data and reduced SBC data) frequencies are not comparable between analyses, thus y axis is left empty

Something should be said about tree topologies as well; the changes within a tree can influence the ages as swapping of large groups forces the ages of the nodes around it to change and vice versa. While the different prior sets and datasets were included, sometimes tree topology has changed between analyses. This influences the resulting ages, therefore the 18 maximum clade credibility trees resulting from the analyses were included in appendix 5. The analyses with the most differing tree topologies were analysis 3, 7, 13 and 14. An example of a topological change which might influence the estimate of node ages is the order of divergence of the *Duguetia* clade, the rest of the LBC and the *Guatteria* clade in analysis 3.

Concluding, table 7 was made to compare the influences of the LBC and SBC calibration priors and the LBC and SBC data on ages and rates. The influence of *Duguetia* resulted in older ages for all analysis including the LBC fossils. The rates of the LBC calibration prior analysis were not much changed relative to the total analysis. For *rbcl* the substitution rates were only slightly higher, especially compared to the effect of the SBC calibration priors, which raised the rates much more than did the LBC calibration priors. Including only LBC data the rate does not become much lower, only in the case of *rbcl* and less severe than does the SBC data.

Table 7: overview of results of the influence of LBC and SBC calibration priors and LBC and SBC data on the mean age estimate of some major nodes and the mean substitution rate estimates of the four markers, ↑ means the age becomes older or the substitution rate faster, ↓ means the age estimate becomes younger or the rate slower, - means the effect was ambiguous

| | LBC fossils | SBC fossils | LBC data | SBC data |
|-----------------------|-------------|-------------|----------|----------|
| LBC age | ↑ | - | - | ↓ |
| SBC age | ↑ | - | - | - |
| MRCA age | ↑ | - | - | ↑ |
| Root age | ↑ | - | - | ↑ |
| <i>matK</i> rate | - | ↑ | - | ↓ |
| <i>rbcl</i> rate | ↑ | ↑ | ↓ | ↓ |
| <i>psbA-trnH</i> rate | - | ↑ | - | ↓ |
| <i>trnL-trnF</i> rate | - | ↑ | - | ↓ |

7. Discussion

Choosing a model to interpret data poses a chicken-egg dilemma, because you cannot possibly know how data behaves before you use models to analyse it. In former research, the appearance of the long branch and short branch clades in the Annonaceae was noticed. Research indeed proved rate heterogeneity to be present within a Annonaceae dataset with likelihood ratio tests and tried to implement this knowledge by choice of model (Richardson *et al.*, 2004). Two articles mentioned earlier dated the Annonaceae phylogenetic tree, but their results differed (Couvreur *et al.*, 2011; Richardson *et al.*, 2004).

7.1 First analyses with the data

A first programme which can be used to examine sequence data and find the appropriate evolutionary model is PartitionFinder. To see whether the LBC and SBC behave according to different models, runs were done with only the LBC sequences and only the SBC sequences. Though the result can in no way give certainty whether the two clades behave according to the same parameter values, they behave according to approximately the same models. The use of the same model on all data is justified, as long as the parameters are allowed to behave differently when necessary.

The first trees constructed in RAxML on the Cipres Science Gateway indeed show long branches and short branches (appendix 3). Bootstrap values are satisfying for most nodes (> 70) and only shallow nodes have lower bootstrap values (0-70), with some polytomies. The topology of this tree was used to determine the prior node (see also the chapter Materials and Methods). Different analyses were constructed by including and excluding the calibration priors and including and excluding the maximum amount of LBC and SBC species sequences without losing the calibration prior nodes. By recording the resulting mean age estimates, their 95% HPDs and some of the resulting substitution rate distributions for the different markers, the effects of priors and data became more clear.

Reaching convergence to the same level of most runs seemed to give problems even when runs were starting with the same tree topology. Apparently the parameter space is so complicated it would sometimes take over 30,000,000 generations before the run reached the same likelihood level as the other analyses. It is not guaranteed that the levels to which I 'decided' the analysis had converged, are the actual best likelihoods in the entire parameter space. When a parameter space is complicated, like in this case, it is very possible a BEAST run will not reach the point with the highest posterior probability (Lakner *et al.*, 2008).

The differences between tree topologies were not accounted for when analysing the results. The results from the analyses with the differing topologies are not discarded because we are interested in the effects of the different priors and data and the topology is closely related to these.

The *rbcl* marker shows a higher degree of divergence in rate than the other markers. Interestingly, the *rbcl* standard deviation of the uncorrelated log-normal relaxed clock (ucl.d.stdev) was also not converged for several analyses, indicating the high variability of the estimate of the rate within *rbcl*. *rbcl* has been shown before to show rate heterogeneity within at least seed plants as a whole and Betulaceae (Bousquet *et al.*, 1992). Possible solutions would be either excluding *rbcl* for further analyses in BEAST, use a programme which can use multiple distributions for rate estimates or design one that can.

7.2 Comparing published ages to results

Two publications dating the Annonaceae, Couvreur *et al.*, (2011) and Richardson *et al.*, (2004) are used to compare the dates obtained in this thesis to. Couvreur *et al.*, (2011) published a table with results from two other publications (Pirie *et al.*, 2012; Su & Saunders, 2009). Pirie *et al.*, (2012) used the penalized likelihood method assuming rate autocorrelation. Couvreur *et al.*, (2011) and Su & Saunders, (2009) used BEAST to estimate mean ages, as was done in this thesis. Richardson *et al.*, (2004) used nonparametric rate smoothing (NPRS; Sanderson, 1997) and penalized likelihood (Sanderson, 2002), with two different calibration points. From these, the analysis with *Archaeanthus* (98 Mya) was used to compare to this study, because this study also uses *Archaeanthus* as a calibration point. Table 8 was made to give an overview of dates from these publications.

Compared to previous age estimates, the 'all including' analysis estimates the divergence ages of the Annonaceae relatively high. The publication of Pirie *et al.*, (2012) shows much younger age estimates, which may be caused by the use of different methods, since the calibration priors used are comparable.

The age of the LBC crown node is estimated to be older than the SBC crown node age in this thesis, which is consistent with most publications. Although this thesis found the age of the LBC to be much older compared to the age of the LBC the other studies. Only the publication of Richardson *et al.*, (2004) estimates the LBC crown to be younger than the SBC crown. The results of this thesis support the hypothesis that the LBC indeed is older than the SBC.

The method of Richardson *et al.*, (2004) differs a lot from the other methods. The species sampling was comparable and the calibration prior of the analysis shown in table 8 is provided by the fossil *Archaeanthus*. In order to adjust for rate heterogeneity, NPRS and penalized likelihood were used to determine the amount of rate smoothing. Using the branch lengths instead of sequences in itself might be enough to account for the difference in the ages of the LBC and SBC compared to other studies.

When using only the deep node calibration priors (*Archaeanthus* and *Futabanthus*), the Annonaceae crown node is estimated to be 91.3 Mya (95% HPD: 95.12-89.00), the LBC/SBC split is estimated to be 70.51 Mya (78.23-62.03), the LBC crown node 65.43 Mya (73.53-57.40) and the SBC crown node is estimated to be 38.85 Mya (47.21-30.78). These ages are congruent with the ages found by other publications, especially those calculated by Couvreur *et al.*, (2011). Couvreur *et al.*, (2011) used a different species sampling with fewer LBC sequences than SBC sequences than in the species sampling of this thesis. The dataset in this study was constructed with a different approach, keeping the amount of species approximate to the distribution as in (taxonomic) reality. Two age constraints were used by Couvreur *et al.*, (2011). The first was a uniform prior with minimal bounds (115.1-114.9 Mya) based on the *Endressinia* fossil, fixing the age of the Magnoliineae crown node. The second is a minimum constraint on the Annonaceae crown node based on the *Futabanthus* fossil with an exponential prior distribution with a hard bound offset of 89 Mya. The tail of the distribution is allowed to reach 115 Mya, resulting in a prior which indicates a divergence of Annonaceae somewhere between 115 Mya and 89 Mya, with a higher chance of divergence towards the younger age.

The *Duguetia* prior age estimated in this thesis is older than the ages estimated in the other studies. This possibly explains the older estimated ages of the analysis with all priors and the 4M dataset compared to other publications. It is not necessarily wrong to use this calibration prior, as it might help BEAST correct for the signal of the data. Substitution rates being drawn from a normal distribution as a way of accounting for rate heterogeneity might in this case not be enough compensation for the rate differences within the sequence data.

Table 8: A summary of estimated ages of some Annonaceae nodes from different studies with different methods (NPRS, penalized likelihood and Bayesian relaxed-clock, with ages in Mya, SD in My, or 95 % HPD (highest priority density) or 95 % CI (confidence interval) in My

| Node | NPRS | Penalized likelihood | BEAST | | |
|------------------|---------------------------------|----------------------|-----------------------|---------------------------------|---------------------|
| | Age in Mya (SD) | | Age in Mya (95% HPD) | | Age in Mya (95% CI) |
| | Richardson <i>et al.</i> , 2004 | Pirie & Doyle (2012) | Su & Saunders, (2009) | Couvreur <i>et al.</i> , (2011) | This thesis |
| Annonaceae crown | 90.6 (1.3) | 75.5 (1.7) | 89.4 (90.4-89.0) | 90.44 (92.98-89.00) | 94.13 (98.79-89.00) |
| LBC/SBC split | 66.7 (2.3) | 62.8 (1.9) | 67.3 (78.1-55.2) | 71.71 (78.26-64.77) | 80.17 (86.33-73.93) |
| LBC crown | 60.2 (2.3) | 57.6 (2.1) | 59.6 (70.5-48.1) | 65.85 (72.42-59.16) | 75.76 (81.91-69.82) |
| SBC crown | 62.5 (3.6) | 55.3 (2.4) | 39.8 (55.1-26.8) | 32.77 (40.00-25.80) | 45.65 (54.61-37.12) |

7.3 Changing effects of calibration priors in different combinations and data selection

Different calibration prior combinations and data resulted in different node ages. This makes interpreting these results very difficult. For the *Archaeanthus* fossil prior node the inclusion of only LBC resulted in the age being estimated older compared to the other analyses. The Annonaceae crown node became older when only LBC sequences were included. The change towards older ages in these deeper nodes can be caused by the inclusion of only the LBC sequences, because of its own presumed faster substitution rate or by the absence of the slower substitution rates within the SBC, or any combination between these two extremes.

The exclusion of only the *Anonasperrum* calibration prior and the exclusion of the combination of the *Anonasperrum* and African Malmeoideae calibration prior both resulted in relatively old age estimates. The *Anonasperrum* fossil is situated somewhere in between the deep node priors and the *Duguetia* prior. By taking out the *Anonasperrum* calibration prior, the effects of the other priors become stronger. Especially in the case of the already noted effect of the *Duguetia* prior. This result indicates that not only the effect of a singular fossil prior is important, but also the combination of fossils may be very influential in the estimate of node ages.

The African Malmeoideae calibration prior seems realistic and informative, since it in plenty cases is estimated somewhere within the lower part of the 95 % HPD. In comparison the '*Mosoxandra*' calibration prior seems more uninformative and might be underestimating the age of the divergence date it was intended to constrain.

The *Anonasperrum* calibration prior was constrained by its boundaries when only the SBC data was present. It appears the age was estimated relatively young (more towards the boundary of the prior) whereas it did not become as young in the other analysis. This might be caused by the absence of the LBC data; when the LBC data is no longer present to push the age of the LBC crown node back in time the node becomes much younger.

Also should be noted here that the *maximum* age constraints used in this study only in the case of the *Archaeanthus/Endressinia* node were constraining the estimated ages. This gives reason to once again look at this prior before further use, but as far as this thesis goes, these deeper priors are not proven erroneous. The '*Mosoxandra*' calibration prior is the only one to be excluded for further analyses, because of its highly uninformative nature. Nonetheless it could still be included as a safeguard for when other priors fail to estimate the *Mosannonna/Oxandra* split in an appropriate manner.

Near & Sanderson (2004) used an approach comparable to this study, but tried to validate the influence of different fossils by including them one by one and calculating the difference between the estimated age with and without the calibration prior. The authors

compared the fit of the calibration priors (after discarding the worst performing priors) and estimated the amount of rate smoothing their model performed best with.

An interesting result of this thesis is that it appears the priors are all influencing the age estimates to a different degree, congruent with the results of Near & Sanderson (2004). In the very obvious example of the *Duguetia* prior we can conclude that this specific calibration prior is very influential throughout the whole tree. In some cases, a prior all of a sudden had a big influence, but only on a selection of the node ages. For example because of the exclusion of the calibration priors of *Anonasperrimum* on its own node and in combination with the African Malmeoideae, the age of the split between the SBC and LBC became relatively much older.

As noted by Ho & Phillips, (2009), the method of Near & Sanderson (2004) has three theoretical shortcomings. First the method uses node estimates as point calibrations, which, when proven inappropriate, is no indication that the node estimate cannot be used as a minimum constraint. Secondly, the method of Near & Sanderson (2004) will exclude those calibration points which are inconsistent with the other calibration points (Marshall, 2008). For example discarding the *Duguetia* prior in the case of this study, would maybe be equal to discarding the most informative prior of all. Finally, it is difficult to make a distinction between errors related to branch length estimation and rate shifts and the quality of the estimate (Hugall *et al.*, 2007). It was expected that the substitution rate within the Annonaceae would be changing, therefore these results may be a reflection of that phenomenon instead of an indication of poor calibration use.

A possible difference between calibration prior influences can be caused by the amount of sequence data and its properties, which is constrained by the calibration prior. The more aberrant sequence data is present within the range of a calibration prior, the more pressure it will exert on the models' parameters. If the *Duguetia* species' sequences are all the results of high substitution rates and you constrain them to be of a certain age, this will overrule the effects of the other priors. Would we have used a smaller amount of *Duguetia* sequence data, the influence probably would have been less extreme.

This raises the question whether it might be possible to find an optimal combination of calibration priors and data inclusion. As Benton *et al.*, (2003) conclude so beautifully; "In the quest for the tree of life, it is arid to claim that either fossils or molecules are the sole arbiter of dating or of tree shape. It is more reasonable to accept that both data sets have their strengths and weaknesses and that each can then be used to assess the other." And it might be the best to judge combinations of fossils and data per model, because some will perform best with a different combination of data and priors than others, depending on the assumptions the models make and how the models handle rate heterogeneity.

For BEAST the assumption is that over the entire range of data, the rate can be drawn from a single normal distribution. The results from this thesis suggest this assumption is violated, since for all markers including only the SBC priors results in a lower rate estimate while including only the SBC data leads to a higher rate estimate. This means that part of the data is not behaving according to the assumption made by BEAST and therefore it may be better to choose a different model. Another possibility is to adjust the dataset by removing those sequences which violate the assumption of the equality of rates. This approach is used by Ayala & Rzhetsky, (1998). They used amongst others a statistical method to reduce their dataset to those sequences which behaved according to the overall molecular clock. More about this interesting method to choose data can be found in Chapter 9. There is a risk attached to this approach, namely the exclusion of all sequences with differing rates (Bromham & Hendy, 2000), in this case either (nearly) the entire SBC or the entire LBC.

Bromham & Hendy (2000) proposed a solution by using multiple fossils to constrain different nodes and estimate the rates from those nodes and extrapolating these rates to other branches. With PAML (Yang, 2007) it might be possible to follow this procedure without extrapolating rates by hand. Assigning different rates to different parts of the tree where the calibration priors are placed would also be a relatively objective manner to assign different rates. But a major disadvantage of this approach is that an additional

assumption will be that the substitution rate between the calibration priors remains (approximately) the same. For the data of this thesis, that assumption might not be violated, since the calibration priors were neatly covering the two clades in which the rate is suspected to be different. On the other hand, there is no way to account for rate changes within the prior clades.

Another way of using clocklike methods despite rate heterogeneity is provided by r8s (Sanderson, 2003), which combines NPRS (Sanderson, 1997) and penalized likelihood (Sanderson, 2002). NPRS only performs well when there is a large enough dataset available, the substitution rate is heterogeneous and there are moderate to high levels of rate autocorrelation. The programme does not take the sequences into account but uses the branch length estimates provided by some other method (for example penalized likelihood as used by Richardson *et al.*, (2004)). As Sanderson himself states in his 1997 publication, the more direct use of sequences to estimate ages should be possible. This being said in combination with our desire to investigate where and how much the rates change, the use of an intermediate estimate like branch length prohibits the more direct investigation of rate changes. R8s can be used to obtain better estimates of ages, but it does not provide much more insight in the underlying processes.

The Annonaceae once again proves to be very complicated to work with, but nonetheless a great case study to investigate model performance (Pirie & Doyle, 2012). The approach and findings of this thesis are not just applicable to the Annonaceae family, but may also be of help for the investigation of other phylogenetic groups with a history of extreme substitution rate differences, for example Betulaceae (Plantae)(Bousquet *et al.*, 1992), Solenogastres (Mollusca)(Meyer *et al.*, 2010) and primates (Perelman *et al.*, 2011).

8. Conclusion

Compared to former age estimates for the Annonaceae the age estimates calculated in this thesis show a as large difference between the ages of the LBC and SBC as Couvreur *et al.*, (2011). Though by accounting for rate heterogeneity in a whole different way, Richardson *et al.*, (2004) found the ages to be nearly the same. The use of BEAST as a model to correctly handle the rate heterogeneity in Annonaceae is not convincing. By reducing the rate variance between the two clades by excluding sequences or by choosing a different phylogenetic dating model the ages may approach the dates of Richardson *et al.*, (2004) more closely.

All fossils used in this thesis except the '*Mosoxandra*' fossil can be used to calibrate the Annonaceae phylogenetic tree. The *Duguetia* fossil does influence age estimates clearly, but not in an unexpected manner. Considering how much *Duguetia* sequences should be constrained by it is advisable.

When investigating the performance of a phylogenetic model with the influences of calibration priors and sequence data the alteration of including calibration priors is helpful. It may be easier to use a quantitative method to value the influence of different calibration priors (Near *et al.*, 2004), but investigating the modelling process by comparing some internal node ages and comparing substitution rate estimate distributions can be a more qualitative approach. The approach of this thesis has as an advance that it can show when a model is wrong, instead of only showing a calibration prior does not work, while in fact the model cannot handle the combination of the prior with the data appropriately. The same applies to the selection of sequence data. Reducing rate heterogeneity by excluding the sequences which cause the rate to become heterogenic one can adapt to the model, but should also consider that the rate heterogeneity tells us something about a groups' phylogenetic history. It has become apparent that indeed both the choice of sequence data as the choice of fossil calibration prior influence the branch length estimation.

The method used in this thesis helps understanding the influences of different priors and selections of sequence data better and can provide insight into the behaviour of the model. Especially when the suspicion is present that a models' assumptions might not all be met, the method of this thesis is helpful to find ways to fit the input better to the model or provide reason to choose for a different model.

9. Future research

This thesis' approach is in general useful for investigating phylogenetic groups in which rate heterogeneity has been found. It is also a good starting point for further dating analyses for the Annonaceae family itself.

Before starting the data analysis, it is necessary to construct a dataset which is as informative as possible. The dataset of this thesis can be used as a starting point, but further on, if there is indication that too much sequence data is constrained by one calibration prior, it might be necessary to adapt species covering in order to fit the model best. The dataset can then be divided into partitions of which the best fitting models for the used programme can be found using software like PartitionFinder.

Some models mentioned before in this thesis, which are designed to handle rate heterogeneity are r8s, BEAST and PAML. All have their advantages and disadvantages, and one should not be hesitant to switch model when it appears the assumptions are not met. Three approaches seem reasonable from this point onwards.

The first step in choosing priors and models is to choose the best calibration priors (the two deep priors, *Archaeanthus* and *Futabanthus*, and for the more shallow nodes African Malmeoideae, *Duguetia* and *Anonaspermum*) and start with different selections of sequence data. It can be informative to see what the results are when different calibration priors get more or less sequence data underneath their nodes, so different amounts of *Duguetia* species, different amounts of other LBC species and different amounts of Malmeoideae species. This could show for example that when less *Duguetia* species are present, the nodes further away from the *Duguetia* crown node will be less influenced (dated younger) as they did when the *Duguetia* prior was absent.

The second method could be somewhat like that of Ayala & Rzhetsky (1998). They calculated the genetic distance for each lineage from the root to the tips of a neighbour joining phylogenetic tree. For all taxa is determined whether they diverge from the average rate of the total tree, followed by a X^2 test after which the sequences most diverging were removed ($P < 0.05$), resulting in a dataset with only sequences evolving in the same rate. For using the dataset in BEAST the allowance for sequences to diverge from the mean substitution rate can be set to a less severe level of for example 0.01, since BEAST can handle rate heterogeneity up to a certain level. The downside of this approach is that such a method could entirely remove one of the clades with differing substitution rates.

The third method to continue with within BEAST is to simultaneously change the amount of sequences and add and remove calibration priors. This would be comparable to analyses 17 and 18 (two extreme inclusions of LBC and SBC ratios in the sequence data), but then also removing the calibration priors and also taking some ratios in between the two extreme ratios used in this thesis.

This thesis shows BEAST may not be a good model to use for dating the Annonaceae clade. PAML has some properties which make it an interesting alternative. When PAML also cannot deal with the rate heterogeneity in the Annonaceae, the rates of the sequences beneath a calibration prior node can be forced to become more alike through a similar method as that of Ayala & Rzhetsky (1998).

It could be interesting to construct an artificial sequence dataset with two groups with different substitution rates, which could not come from the same distribution but are close. This dataset can be used to perform the same analyses as were done in this thesis with some calibration priors in the one clade and some in the other. And it could also be used to experiment with the inclusion of calibration priors and the inclusion of data underneath those calibration priors. This could provide more insight into whether it is possible to adjust a dataset to fit the models' assumptions and still estimate the ages close to the 'true' divergence ages (which would be known in this experiment).

Some problems encountered during this thesis are the large dataset, which costs a lot of time to assemble, a lot of storage space and a lot of computational hours when running

analyses. My advise is when trying to date large phylogenetic trees to use methods like congruification as described by Eastman *et al.*, (2013). When trying to find the appropriate model, a smaller dataset than used in this thesis could be used. Because of the large amount of computational hours the 4M dataset took to be analysed I could not do all the intended analyses and did not succeed to date the phylogenetic tree of the total dataset.

10. Acknowledgements

Coming to know my interest of fossils, my supervisor Lars W. Chatrou provided me with an excellent subject for my master thesis. I am very grateful for this opportunity, I've never had so much fun with a project before. Lars helped me through some difficult stages by giving helpful advice. I admire his scientific approach of problems in phylogenetics and am very grateful for his time and effort. Thank you very much, Lars! Freek Bakker always was ready to advise on the matter of phylogenetic tree construction and choice of software. I want to thank some other members of the Biosystematics group for their help, Sara van der Kerke, Setareh Mohammedin, Rens Holmer and Jesse Erens for their suggestions and help when I got stuck with technical problems. Also the members of the Biosystematics Island, thank you very much for your always uplifting spirits and phylogenetic jokes. I want to thank Hans de Vries for advising me on the drawings of the fossils. Special note should be made to professor Miller and the Cipres Science Gateway, without whom this research could not have taken place. The effort of giving scientists all over the world access to their computational resources is a valuable asset for many scientists, my gratitude for that. Last, but definitely not least, I would like to thank my parents and friends for their support throughout the project.

11. References

- Ayala, F. J., & Rzhetsky, A. (1998). Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. *Proceedings of the National Academy of Sciences of the United States of America*, 95(2), 606–611.
- Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., ... Suchard, M. A. (2012). BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Systematic Biology*, 61(1), 170–3. doi:10.1093/sysbio/syr100
- Barraclough, T. G., & Savolainen, V. (2001). Evolutionary Rates and Species Diversity in Flowering Plants. *Evolution*, 55(4), 677–683.
- Benton, M. J., & Ayala, F. J. (2003). Dating the Tree of Life. *Science (New York, N.Y.)*, 300(5626), 1698–700. doi:10.1126/science.1077795
- Benton, M. J., & Donoghue, P. C. J. (2007). Paleontological evidence to date the tree of life. *Molecular Biology and Evolution*, 24(1), 26–53. doi:10.1093/molbev/msl150
- Bousquet, J., Strauss, S. H., Doerksen, A. H., & Price, R. A. (1992). Extensive variation in evolutionary rate of rbcL gene sequences among seed plants. *Proceedings of the National Academy of Sciences*, 89(16), 7844–7848. doi:10.1073/pnas.89.16.7844
- Bousquet, J., Strauss, S., & Li, P. (1992). Complete congruence between morphological and rbcL-based molecular phylogenies in birches and related species (Betulaceae). *Molecular Biology and Evolution*, 9(6), 1076–1088. Retrieved from <http://mbe.oxfordjournals.org/cgi/content/long/9/6/1076>
- Bromham, L. D., & Hendy, M. D. (2000). Can fast early rates reconcile molecular dates with the Cambrian explosion? *Proceedings of the Royal Society Biological Sciences*, 267(1447), 1041–7. doi:10.1098/rspb.2000.1108
- Chaowasku, T., Johnson, D. M., van der Ham, R. W. J. M., & Chatrou, L. W. (2012). Characterization of *Hubera* (Annonaceae), a new genus segregated from *Polyalthia* and allied to *Miliusa*. *Phytotaxa*, 69, 33–56.
- Chatrou, L. W. (1998). Revision of the Malmea Alliance: Malmea and three new, Neotropical genera: *Klarobelia*, *Mosannonna*, and *Pseudomalmea*. In *Changing Genera: Systematic studies in Neotropical and West African Annonaceae* (pp. 103–192). Utrecht: Herbarium Division, Utrecht University.
- Chatrou, L. W., Erkens, R. H. J., Richardson, J. E., Saunders, R. M. K., & Fay, M. F. (2012). The natural history of Annonaceae. *Botanical Journal of the Linnean Society*, 169(1), 1–4. doi:10.1111/j.1095-8339.2012.01242.x
- Chatrou, L. W., Escribano, M. P., Viruel, M. a, Maas, J. W., Richardson, J. E., & Hormaza, J. I. (2009). Flanking regions of monomorphic microsatellite loci provide a new source of data for plant species-level phylogenetics. *Molecular Phylogenetics and Evolution*, 53(3), 726–33. doi:10.1016/j.ympev.2009.07.024

- Chatrou, L. W., Pirie, M. D., Erkens, R. H. J., Couvreur, T. L. P., Neubig, K. M., Abbott, J. R., ... Chase, M. W. (2012). A new subfamilial and tribal classification of the pantropical flowering plant family Annonaceae informed by molecular phylogenetics. *Botanical Journal of the Linnean Society*, *169*(1), 5–40. doi:10.1111/j.1095-8339.2012.01235.x
- Chatrou, L. W., Pirie, M. D., van Velzen, R., & Bakker, F. T. (2013). Annonaceae Substitution Rates - A Codon Model Perspective. *unpublished*.
- Clarke, J. T., Warnock, R. C. M., & Donoghue, P. C. J. (2011). Establishing a time-scale for plant evolution. *The New Phytologist*, *192*, 266–301. doi:10.1111/j.1469-8137.2011.03794.x
- Conroy, C. J., & van Tuinen, M. (2003). Extracting Time From Phylogenies: Positive Interplay between Fossil and Genetic Data. *Journal of Mammalogy*, *84*(2), 444–455. doi:10.1644/1545-1542(2003)084<0444:ETFPPI>2.0.CO;2
- Couvreur, T. L. P., Pirie, M. D., Chatrou, L. W., Saunders, R. M. K., Su, Y. C. F., Richardson, J. E., & Erkens, R. H. J. (2011). Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. *Journal of Biogeography*, *38*, 664–680. doi:10.1111/j.1365-2699.2010.02434.x
- Crepet, W. L. (2008). The Fossil Record of Angiosperms: Requiem or Renaissance? *Annals of the Missouri Botanical Garden*, *95*(1), 3–33. doi:10.3417/2007065
- Dilcher, D. L., & Crane, P. R. (1984). *Archaenthus*: An Early Angiosperm From the Cenomanian of the Western Interior of North America. *Annals of the Missouri Botanical Garden*, *71*(2), 351–383. Retrieved from <http://www.jstor.org/stable/2399030>
- Dilcher, D. L., & Lott, T. A. (2005). A Middle Eocene Fossil Plant Assemblage (Powers Clay Pit) From Western Tennessee. *Bulletin Florida Museum Natural History*, *45*(1), 1–43.
- Donoghue, P. C. J., & Benton, M. J. (2007). Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution*, *22*(8), 424–431. doi:10.1016/j.tree.2007.05.005
- Donoghue, P. C. J., & Purnell, M. A. (2009). Distinguishing heat from light in debate over controversial fossils. *BioEssays*, *31*, 178–189. doi:10.1002/bies.200800128
- Dornburg, A., Brandley, M. C., McGowen, M. R., & Near, T. J. (2012). Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution*, *29*(2), 721–36. doi:10.1093/molbev/msr228
- Doyle, J. A., & Donoghue, M. J. (1993). Phylogenies and angiosperm diversification. *Paleobiology*, *19*(2), 141–167. Retrieved from <http://paleobiol.geoscienceworld.org/cgi/content/long/19/2/141>
- Doyle, J. A., & Endress, P. K. (2000). Morphological Phylogenetic Analysis of Basal Angiosperms: Comparison and Combination with Molecular Data. *International Journal of Plant Sciences*, *161*(S6), S121–S153. doi:10.1086/317578
- Doyle, J. A., Sauquet, H., Scharaschkin, T., & Le Thomas, A. (2004). Phylogeny, Molecular and Fossil Dating, and Biogeographic History of Annonaceae and Myristicaceae (Magnoliales). *International Journal of Plant Sciences*, *165*(S4), S55–S67.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, *4*(5), e88. doi:10.1371/journal.pbio.0040088
- Drummond, A. J., Rambaut, A., Suchard, M. A., & Xie, W. (2013). BEAUti. Retrieved from <http://beast.bio.ed.ac.uk>
- Drummond, A. J., Suchard, M. a, Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–73. doi:10.1093/molbev/mss075
- Duchene, D., & Bromham, L. (2013). Rates of molecular evolution and diversification in plants: chloroplast substitution rates correlate with species-richness in the Proteaceae. *BMC Evolutionary Biology*, *13*(1), 65. doi:10.1186/1471-2148-13-65
- Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: support for time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, *4*(4), 688–691. doi:10.1111/2041-210X.12051
- Encyclopaedia Britannica Online Academic Edition. (2014). "Annonaceae." *Encyclopaedia Britannica Inc., 2014*. Retrieved from <http://www.britannica.com/EBchecked/topic/26391/Annonaceae>
- Erkens, R. H. J., Chatrou, L. W., Koek-Noorman, J., Maas, J. W., & Maas, P. J. M. (2007). Classification of a large and widespread genus of Neotropical trees, Guatteria (Annonaceae) and its three satellite genera Guatterrella, Guatterriopsis and Heteropetalum. *Taxon*, *56*(3), 757–774.
- Erkens, R. H. J., Maas, J. W., & Couvreur, T. L. P. (2009). From Africa via Europe to South America: migrational route of a species-rich genus of Neotropical lowland rain forest trees (Guatteria, Annonaceae). *Journal of Biogeography*, *36*, 2338–2352. doi:10.1111/j.1365-2699.2009.02162.x

- Gandolfo, M. a., Nixon, K. C., & Crepet, W. L. (2008). Selection of Fossils for Calibration of Molecular Dating Models. *Annals of the Missouri Botanical Garden*, 95(1), 34–42. doi:10.3417/2007064
- Gernhard, T. (2008). The conditioned reconstructed process. *Journal of Theoretical Biology*, 253(4), 769–78. doi:10.1016/j.jtbi.2008.04.005
- Groussin, M., Boussau, B., & Gouy, M. (2013). A branch-heterogeneous model of protein evolution for efficient inference of ancestral sequences. *Systematic Biology*, 62(4), 523–538. doi:10.1093/sysbio/syt016
- Heads, M. (2005). Dating nodes on molecular phylogenies: a critique of molecular biogeography. *Cladistics*, 21, 62–78. doi:10.1111/j.1096-0031.2005.00052.x
- Heath, T. A., Hedtke, S. M., & Hillis, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, 46(3), 239–257. doi:10.3724/SP.J.1002.2008.08016
- Ho, S. Y. W., & Phillips, M. J. (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, 58(3), 367–80. doi:10.1093/sysbio/syp035
- Hugall, A. F., Foster, R., & Lee, M. S. Y. (2007). Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic Biology*, 56(4), 543–63. doi:10.1080/10635150701477825
- Huson, D. H., & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23(2), 254–267.
- Lakner, C., van der Mark, P., Huelsenbeck, J. P., Larget, B., & Ronquist, F. (2008). Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. *Systematic biology*, 57(1), 86–103. doi:10.1080/10635150801886156
- Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 26(6), 1695–1701. Retrieved from <http://dx.doi.org/10.1093/molbev/mss020>
- Maddison, W. P., & Maddison, D. R. (2011). Mesquite: a modular system for evolutionary analysis. Retrieved from <http://mesquiteproject.org>
- Magallón, S., Hilu, K. W., & Quandt, D. (2013). Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany*, 100(3), 556–73. doi:10.3732/ajb.1200416
- Marshall, C. R. (2008). A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *The American Naturalist*, 171(6), 726–42. doi:10.1086/587523
- Meyer, A., Todt, C., Mikkelsen, N. T., & Lieb, B. (2010). Fast evolving 18S rRNA sequences from Solenogastres (Mollusca) resist standard PCR amplification and give new insights into mollusk substitution rate heterogeneity. *BMC Evolutionary Biology*, 10(1), 70. doi:10.1186/1471-2148-10-70
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Cipres Science Gateway. *Creating the CIPRES Science Gateway for inference of large phylogenetic trees*.
- Mohr, B. A. R., & Bernardes-de-Oliveira, M. E. C. (2004). Endressinia brasiliana, a Magnolialean Angiosperm from the Lower Cretaceous Crato Formation (Brazil). *International Journal of Plant Sciences*, 165(6), 1121–1133. Retrieved from <http://www.jstor.org/stable/10.1086/423879>
- Mols, J. B., Gravendeel, B., Chatrou, L. W., Pirie, M. D., Bygrave, P. C., Chase, M. W., & Kessler, P. J. (2004). Clades in Asian Annonaceae: Monophyletic Genera in the Polyphyletic Miliuseae. *American Journal of Botany*, 91(4), 590–600.
- Morton, B. R., & Clegg, M. T. (1995). Neighboring Base Composition Is Strongly Correlated with Base Substitution Bias in a Region of the Chloroplast Genome. *Journal of Molecular evolution*, (41), 597–603.
- Near, T. J., & Sanderson, M. J. (2004). Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359(1450), 1477–83. doi:10.1098/rstb.2004.1523
- Nixon, K. C. (1996). Paleobotany in cladistics and cladistics in paleobotany: enlightenment and uncertainty. *Review of Palaeobotany and Palynology*, 90(3-4), 361–373. doi:10.1016/0034-6667(95)00092-5
- Nowak, M. D., Smith, A. B., Simpson, C., & Zwickl, D. J. (2013). A Simple Method for Estimating Informative Node Age Priors for the Fossil Calibration of Molecular Divergence Time Analyses. *PLoS One*, 8(6), e66245. doi:10.1371/journal.pone.0066245
- Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A. M., ... Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, 7(3), e1001342. doi:10.1371/journal.pgen.1001342

- Pirie, M. D., Chatrou, L. W., Erkens, R. H. J., Jan, W., van der Niet, T., Mols, J. B., & James, E. (2005). Chapter 8: Phylogeny reconstruction and molecular dating in four Neotropical genera of Annonaceae: the effect of taxon sampling in age estimations. In *Regnum Vegetabile* (pp. 149–174).
- Pirie, M. D., & Doyle, J. A. (2012). Dating clades with fossils and molecules: the case of Annonaceae. *Botanical Journal of the Linnean Society*, *169*(1), 84–116. doi:10.1111/j.1095-8339.2012.01234.x
- Rainer, H., & Chatrou, L. W. (2011). Catalogue of Life. Retrieved June 18, 2013, from <http://www.catalogueoflife.org/col/details/database/id/40>
- Raup, D. M. (1972). Taxonomic Diversity during the Phanerozoic. *American Association for the Advancement of Science*, *177*(4054), 1065–1071.
- Reid, E. M., & Chandler, M. E. J. (1933). *The London Clay Flora*. London: British Museum (Natural History).
- Reisz, R. R., & Müller, J. (2004). Molecular timescales and the fossil record: a paleontological perspective. *Trends in Genetics*, *20*(5), 237–41. doi:10.1016/j.tig.2004.03.007
- Richardson, J. E., Chatrou, L. W., Mols, J. B., Erkens, R. H. J., & Pirie, M. D. (2004). Historical biogeography of Two Cosmopolitan Families of Flowering Plants: Annonaceae and Rhamnaceae. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1450), 1495–1508. doi:10.1098/rstb.2004.1537
- Rodríguez-Trelles, F., Tarrío, R., & Ayala, F. J. (2002). A methodological bias toward molecular evolutionary time and overestimation of scales. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 8112–8115.
- Roth, J. L. (1981). Epidermal Studies in the Annonaceae and Related Families. *Dissertation, Indiana University*.
- Sanderson, M. J. (1997). A Nonparametric Approach of Rate Constancy to Estimating Divergence Times in the Absence. *Molecular Biology and Evolution*, *14*(12), 1218–1231.
- Sanderson, M. J. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, *19*(1), 101–109. doi:10.1093/oxfordjournals.molbev.a003974
- Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics Applications Note*, *19*(2), 301–302.
- Sauquet, H., Doyle, J. A., Scharaschkin, T., Borsch, T., Hilu, K. W., Chatrou, L. W., & Le Thomas, A. (2003). Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society*, *142*(2), 125–186. doi:10.1046/j.1095-8339.2003.00171.x
- Shaul, S., & Graur, D. (2002). Playing chicken (*Gallus gallus*): methodological inconsistencies of molecular divergence date estimates due to secondary calibration points. *Gene*, *300*(1-2), 59–61. doi:10.1016/S0378-1119(02)00851-X
- Sjödin, P., Bataillon, T., & Schierup, M. H. (2010). Insertion and Deletion Processes in Recent Human History. *PloS One*, *5*(1), 1–10. doi:10.1371/Citation
- Smith, A. B., & Peterson, K. J. (2002). Dating the Time of Origin of Major Clades: Molecular Clocks and the Fossil Record. *Annual Review of Earth and Planetary Sciences*, *30*(1), 65–88. doi:10.1146/annurev.earth.30.091201.140057
- Srivastava, G., & Mehrotra, R. C. (2013). First fossil record of *Alphonsea* Hk. f. & T. (Annonaceae) from the Late Oligocene sediments of Assam, India and comments on its phylogeography. *PloS One*, *8*(1), e53177. doi:10.1371/journal.pone.0053177
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–90. doi:10.1093/bioinformatics/btl446
- Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology*, *57*(5), 758–71. doi:10.1080/10635150802429642
- Su, Y. C. F., & Saunders, R. M. K. (2009). Evolutionary divergence times in the Annonaceae: evidence of a late Miocene origin of *Pseuduvaria* in Sundaland with subsequent diversification in New Guinea. *BMC Evolutionary Biology*, *9*(153), 1–19. doi:10.1186/1471-2148-9-153
- Sullivan, J., & Joyce, P. (2005). Model Selection in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *36*, 445–466. Retrieved from <http://www.jstor.org/stable/30033812>
- Sullivan, J., Swofford, D. L., & Naylor, G. J. P. (1999). The Effect of Taxon Sampling on Estimating Rate Heterogeneity Parameters of Maximum-Likelihood Models. *Molecular Biology and Evolution*, *16*(10), 1347–1356.
- Takahashi, M., Friis, E. M., Uesugi, K., Suzuki, Y., & Crane, P. R. (2008). Floral Evidence of Annonaceae from the late Cretaceous of Japan. *International Journal of Plant Sciences*, *169*(7), 908–917.

- Takahashi, M., Friis, E. M., Uesugi, K., Suzuki, Y., & Crane, P. R. (2008). Floral Evidence of Annonaceae from the Late Cretaceous of Japan. *International Journal of Plant Sciences*, 169(7), 908–917. doi:10.1086/589693
- Thorne, J. L., & Kishino, H. (2002). Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology*, 51(5), 689–702. doi:10.1080/10635150290102456
- Vaidya, G., Lohman, D. J., & Meier, R. (2014). SequenceMatrix: concatenation software for the fast assembly of multigene datasets with character set and codon information.
- Wheat, C. W., & Wahlberg, N. (2013). Critiquing blind dating: the dangers of over-confident date estimates in comparative genomics. *Trends in Ecology & Evolution*, 1–7. doi:10.1016/j.tree.2013.07.007
- Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. doi:10.1093/molbev/msm088
- Yang, Z., & Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, 23(1), 212–26. doi:10.1093/molbev/msj024
- Yule, G. U. (1925). A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London Series B, Containing Papers of a Biological Character*, 213, 21–87. Retrieved from <http://www.jstor.org/stable/92117>

Appendix 1: Species table with GenBank numbers and references

Table 1: species table with references, vouchers, country of origin/cultivation and GenBank numbers for seven markers and record of inclusion in different datasets used in this thesis; a dataset with seven markers (7M), with four markers (4M) and the LBC and SBC datasets for PartitionFinder. In the first two datasets, the outgroup species (top 6 species) were first excluded for the PartitionFinder analyses, but included for the RAxML analyses. After the RAxML analyses, it was decided to remove the ingroup species indicated with a *. UP instead of a GenBank number indicates the sequence was not published yet on GenBank at 21/11/2013 but was provided with one of the datasets mentioned in Chapter 5

| Identification | | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|--|---------------------------------------|---|--------------------------------|-------------|-------------|--|------------------|------------------|------------------|----------------------|----|-----|-----|
| Species | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Persea americana</i> Mill. * | UUBG 87GR00058 | Cultivated in UUBG, of Neotropical origin | AY841592 | — | JQ437545 | AY841669 | JQ513882 | JQ513883 | — | | | | |
| <i>Coelocaryon preussii</i> Warb. | Wieringa, J.J. 3640 (WAG) | Gabon | AY743437 | AY743475 | JQ437546 | AY743456 | — | — | — | | | | |
| <i>Degeneria vitiensis</i> I.W.Bailey & A.C.Sm. / <i>D. roseiflora</i> J.M.Mill. * | Mixed origin | - | L12643 | AB055549 | AY394736 | AY220414 (intron) AY220361 (spacer) | — | — | — | | | | |
| <i>Eupomatia bennettii</i> F.Muell. | Chatrou, L.W. s.n. (U) | Cultivated in UUBG, origin | DQ861790 | JQ437547 | AY218175 | DQ861842 | — | — | JQ513885 | | | | |
| <i>Galbulimima belgraveana</i> (F.Muell.) Sprague * | Mixed origin | - | L12646 | AF465294 | AY218176 | AY220415 (intron) AY220362 (spacer) | — | — | — | | | | |
| <i>Liriodendron chinense</i> Sargent | Chatrou, L.W. 279 (U) | Cultivated in UUBG, origin China | AY841593 | — | — | AY841670 | AY841424 | — | — | | | | |
| <i>Magnolia kobus</i> DC. | Chatrou, L.W. 278 (U) | Cultivated in UUBG, origin Japan | AY743438 | AY743476 | — | AY743457 | AY841425 | — | — | | | | |
| <i>Alphonsea boniana</i> Finet & Gagnep. | Kessler, P.J.A. 3116 (L) | Vietnam | AY318965 | AY518809 | — | AY319077 | — | — | — | | | | |
| <i>Alphonsea elliptica</i> Hook.F. & Thomson | Van Balgooy, M. 5141 (L) | Indonesia | AY318966 | AY518807 | JQ690401 | AY319078 | JQ690402 | — | — | | x | | x |
| <i>Alphonsea javanica</i> Scheff. | Chase, M.W. 2072 (K) | Indonesia | AY318967 | AY518810 | — | AY319079 | — | — | — | | | | |
| <i>Alphonsea kinabaluensis</i> J. Sinclair | Risdale, DV-S-3048 (L) | Malaysia | AY318968 | AY518811 | — | AY319080 | — | — | — | | | | |
| <i>Alphonsea</i> sp PK3186 | Kessler, P.J.A. 3186 (TISTR, Bangkok) | Thailand | — | AY518808 | JQ690404 | AY319082 | JQ690405 | — | — | | | | |
| <i>Ambavia gerrardii</i> (Baill.) Le Thomas | Sauquet, H. 23 (P) | Madagascar | — | AY220435 | AY218168 | AY220358 | — | — | — | | | | |
| <i>Anaxagorea javanica</i> (Craib, R.E.Fr) Maas & Westra | Kessler, P.J.A. 3112 (L) | Singapore | AY319075 | AY518882 | — | AY319189 | — | — | — | | | | |
| <i>Anaxagorea luzonensis</i> A. Gray | Kessler, P.J.A. 3231 (L) | Thailand | AY319074 | AY518883 | — | AY319188 | — | — | — | | | | |
| <i>Anaxagorea phaeocarpa</i> Mart. | Maas, P.J.M. 8592 (U) | Ecuador | AY238952 | AY238960 | EF179279 | AY231284 (intron) | AY841426 | EF179244 | EF179321 | x | x | | |
| <i>Anaxagorea silvatica</i> R.E.Fr. | Maas, P.J.M. 8836 (U) | Brazil | AY743439 | AY743477 | EF179280 | AY743458 | AY841427 | AY578140 | EF179322 | x | x | | |
| <i>Annickia chlorantha</i> (Oliv.) Setten & Maas | Sosef, M.S.M. 1877 (WAG) | Gabon | AY841594 | AY841393 | AY841401 | AY841671 | AY841442 | AY841370 | AY841550 | x | x | | x |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|------------------------------|---|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Annickia kummerae</i> (Engl. & Diels) Setten & Maas | Johnson, D.M., 1942 (OWU) | Tanzania | AY319057 | AY518877 | — | AY319171 | AY841443 | — | — | | x | | x |
| <i>Annickia pilosa</i> (Exell) Setten & Maas | Sosef, M.S.M. 1803 (WAG) | Gabon | AY743450 | AY743488 | AY841402 | AY743469 | AY841444 | AY841371 | AY841551 | x | x | | x |
| <i>Annona amazonica</i> R.E.Fr. | Chatrou, L.W. 462 (L) | Bolivia | EU420853 | — | — | EU420836 | — | — | — | | | | |
| <i>Annona bicolor</i> Urb. | Maas, P.J.M. 8381 (U) | Mexico | EU420854 | — | — | EU420837 | — | — | — | | | | |
| <i>Annona cordifolia</i> (SzyszzyÅ.,) R.E.FR. | Chatrou, L.W. 343 (U) | Bolivia | EU420855 | — | — | UP | — | — | — | | | | |
| <i>Annona cuspidata</i> (Mart.) H. Rainer | Jansen-Jacobs, M.J. 5957 (U) | Guyana | EU420869 | — | — | EU420851 | — | — | — | | | | |
| <i>Annona deceptrix</i> (Westra) H. Rainer | Maas, P.J.M et al. 8564 (U) | Ecuador | AY841595 | — | — | AY841672 | — | — | — | | | | |
| <i>Annona dumetorum</i> syn. <i>rosei</i> Saff. | Maas, P.J.M. 8374 (U) | Dominican Republic | EU420856 | — | — | EU420838 | — | — | — | | | | |
| <i>Annona edulis</i> (Triana & Planch.) H. Rainer | Chatrou, L.W. et al. 198 (U) | Peru | AY841655 | — | — | AY841733 | — | — | — | | | | |
| <i>Annona flava</i> | Unpublished | - | UP | — | — | UP | — | — | — | | | | |
| <i>Annona foetida</i> Mart. | Unpublished | - | UP | — | — | UP | — | — | — | | | | |
| <i>Annona glabra</i> L. | Chatrou, L.W. 467 (U) | Cultivated in UUBG, origin Florida | AY841596 | DQ125050 | EF179281 | AY841673 | DQ125116 | EF179246 | EF179323 | x | x | x | |
| <i>Annona herzogii</i> (R.E.Fr.) H. Rainer syn. <i>rollinia</i> | Chatrou, L.W. 162 (U) | Peru | AY841656 | DQ125062 | EF179308 | AY841734 | DQ125132 | EF179273 | EF179350 | x | x | x | |
| <i>Annona holosericea</i> Saff. | Maas, P.J.M. 8445 (U) | Honduras | EU420858 | — | — | EU420840 | — | — | — | | | | |
| <i>Annona hypoglauca</i> Mart. | Chatrou, L.W. 444 (U) | Bolivia | EU420859 | — | — | EU420841 | — | — | — | | | | |
| <i>Annona macrophyllata</i> Donn. Sm. | Unpublished | - | UP | — | — | UP | — | — | — | | | | |
| <i>Annona montana</i> Macfad. | Chatrou, L.W. 484 (U) | Tree cultivated in GGBG | EU420860 | — | — | EU420842 | — | — | — | | | | |
| <i>Annona mucosa</i> Jacq. | Chatrou, L.W. 247 (U) | Peru | EU420870 | — | — | EU420852 | — | — | — | | | | |
| <i>Annona muricata</i> L. | Chatrou, L.W. 468 (U) | Cultivated in UUBG, of Neotropical origin | AY743440 | AY743478 | EF179282 | AY743459 5 | AY841428 | EF179247 | EF179324 | x | x | x | |
| <i>Annona neochrysoarpa</i> H. Rainer | Pirie, M.D. 43 (U) | Peru | EU420868 | — | — | EU420850 | — | — | — | | | | |
| <i>Annona oligocarpa</i> syn. <i>neglecta</i> R.E.Fr | Maas, P.J.M. 8522 (U) | Ecuador | EU420861 | — | — | EU420843 | — | — | — | | | | |
| <i>Annona pruinosa</i> G. E. Schatz | Chatrou, L.W. 77 (U) | Costa Rica | EU420862 | — | — | EU420844 | — | — | — | | | | |
| <i>Annona reticulata</i> L. | Chatrou, L.W. et al. 290 (U) | Bolivia | EU420863 | — | — | EU420845 | — | — | — | | | | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|--|-------------------------------|---|--------------------------------|-------------|-------------|--|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Annona senegalensis</i> Pers. | Chatrou, L.W. 469 (U) | ? | AY841597 | — | — | AY841674 | — | — | — | | | | |
| <i>Annona squamosa</i> | Van Proosdij, A.S.J. 1133 (U) | Netherlands Antilles | EU420865 | — | — | EU420847 | — | — | — | | | | |
| <i>Annona symphyocarpa</i> Sandwith | Ek, R.C. 1270 (U) | Guyana | EU420866 | — | — | EU420848 | — | — | — | | | | |
| <i>Annona urbaniana</i> R.E.Fr | Maas, P.J.M. 8392 (U) | Dominican Republic | EU420867 | — | — | EU420849 | — | — | — | | | | |
| <i>Anonidium</i> sp. | Cheek, M. 7896 (K) | Cameroon | AY841598 | DQ125051 | EF179283 | AY841675 | DQ125117 | EF179248 | EF179325 | x | x | x | |
| <i>Artabotrys hexapetalus</i> (L.f.) Bhandari | UUBG 94GR01614 (U) | Cultivated in UUBG, origin India | AY238953 | AY238962 | EF179284 | AY231286 (intron) AY238946 (spacer) | AY841429 | EF179249 | EF179326 | x | x | x | |
| <i>Artabotrys</i> sp. | Wieringa, J.J. 4018 (WAG) | Gabon | AY841599 | DQ125052 | EF179285 | AY841676 | DQ125118 | EF179250 | EF179327 | x | x | x | |
| <i>Asimina longifolia</i> A.Gray | Weerasooriya, A. s.n. (U) | USA | DQ124939 | DQ125053 | EF179286 | AY841677 | DQ125119 | EF179251 | EF179328 | x | x | x | |
| <i>Asimina rugelii</i> B.L.Rob. | Abbott, J.R., 22361 (FLAS) | USA | JQ513887 | — | — | GQ139881 | — | — | — | | | | |
| <i>Asimina triloba</i> (L.) Dunal | Chatrou, L.W. 276 (U) | Cultivated in UUBG, origin USA | AY743441 | AY743479 | EF179287 | AY743460 | AY841430 | EF179252 | EF179329 | x | x | x | |
| <i>Asteranthe asterias</i> (S. Moore) Engl. & Diels | Robertson, A. 7548 (WAG) | Kenya | EU169757 | — | EU169711 | EU169757 | EU169734 | — | EU169801 | | | | |
| <i>Bocageopsis canescens</i> (Spruce ex Benth.) R.E.Fr. | Maas et al. 9243 (U) | Brazil | JQ690407 | JQ690409 | JQ690410 | JQ690408 | JQ690411 | — | — | | x | | x |
| <i>Bocageopsis multiflora</i> (Mart.) R.E.Fr. | Jansen-Jacobs, M.J. 5789 (U) | Guyana | AY841600 | DQ018262 | — | AY841678 | AY841445 | — | — | | x | | x |
| <i>Bocageopsis pleiosperma</i> Maas | Miralha, J.M.S. 300 (U) | Brazil | AY841601 | — | — | AY841679 | — | — | — | | | | |
| <i>Cananga odorata</i> (Lam.) Hook.f & Thomson | Chatrou, L.W. 93 (U) | Costa Rica | AY841602 | AY841394 | AY841403 | AY841680 | AY841431 | AY841372 | AY841548 | x | x | | |
| <i>Cleistopholis glauca</i> Pierre ex Engl. & Diels | Wieringa, J.J. 3278 (WAG) | Gabon | AY841603 | AY841395 | AY841404 | AY841681 | AY841432 | AY841373 | AY841549 | x | x | | |
| <i>Crematosperma brevipes</i> (DC.) R.E.Fr. | Scharf, U. 76 (U) | French Guiana | AY743527 | AY743550 | AY841405 | AY743573 | AY841447 | AY841374 | AY841552 | x | x | | x |
| <i>Crematosperma cauliflorum</i> R.E.Fr. | Chatrou, L.W. 224 (U) | Peru | AY743519 | AY743542 | AY841406 | AY743565 | AY841448 | AY841375 | AY841553 | x | x | | x |
| <i>Crematosperma leiophyllum</i> (Diels) R.E.Fr. | Pirie, M.D. 2 (U) | Bolivia | AY743523 | AY743546 | DQ018123 | AY743569 | AY841449 | — | — | | x | | x |
| <i>Crematosperma megalophyllum</i> R.E.Fr. | Chatrou L.W. 259 (U) | Ecuador | AY743522 | AY743545 | DQ018122 | AY743568 | AY841451 | — | — | | x | | x |
| <i>Crematosperma microcarpum</i> R.E.Fr. | Chatrou L.W. 208 (U) | Peru | AY743518 | AY743541 | DQ018120 | AY743564 | AY841452 | — | — | | x | | x |
| <i>Cyathocalyx martabanicus</i> Hook.f. & Thomson | Mols, J.B. 11 (L) | Cult. in Kebun Raya Bogor, Indonesia | AY841605 | DQ125054 | EF179288 | AY841683 | DQ125120 | EF179253 | EF179330 | x | x | | |
| <i>Cymbopetalum brasiliense</i> (Vell.) Benth. ex Baill. | UUBG 84GR00275 | Cultivated in UUBG, originating from Brazil | AY841608 | DQ125055 | EF179289 | AY841686 | DQ125121 | EF179254 | EF179331 | x | x | x | |

| Identification | | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|------------------------------|------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| Species | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Cymbopetalum</i> sp. | Chatrou, L.W. et al. 44 (U) | Costa Rica | AY841523 | DQ018258 | — | AY841537 | — | — | — | | | | |
| <i>Cymbopetalum torulosum</i> G.E.Schatz | Chatrou, L.W. 54 (U) | Costa Rica | AY743442 | AY743480 | — | AY743461 | — | — | — | | | | |
| <i>Dasyaschalon sootepense</i> Craib. | Kessler, P.J.A. 3201 (L) | Thailand | AY743443 | AY743481 | JQ768600 | AY743462 | JQ768642 | — | — | | x | x | |
| <i>Dendrokingstonia gardneri</i> | Unpublished | - | UP | UP | UP | UP | UP | — | — | | x | | |
| <i>Dendrokingstonia nervosa</i> (Hook. F. & Thomson) Rauschert | Unpublished | - | — | — | UP | UP | UP | — | — | | | | |
| <i>Desmopsis microcarpa</i> R.E.Fr. | Chatrou, L.W. 85 (U) | Costa Rica | AY319059 | AY518804 | JX544771 | AY319173 | AY841461 | — | — | | x | | x |
| <i>Desmopsis schippii</i> Standl. | Chatrou, L.W. 94 (U) | Costa Rica | AY319060 | AY518805 | — | AY319174 | — | — | — | | | | |
| <i>Desmos chinensis</i> Lour. | C.-C. Pang N2 (HKU) | Hong Kong | JQ762414 | JQ768567 | JQ768603 | JQ762415 | JQ768646 | — | — | | x | | |
| <i>Desmos elegans</i> (Thwaites) Saff. | Kostermans 24761 (L) | Sri Lanka | HQ214067 | JQ768571 | — | HQ214069 | JQ768650 | — | — | | x | | |
| <i>Desmos macrocalyx</i> Finet & Gagnep. | Kessler, P.J.A. 3199 (L) | Thailand | AY841610 | EF179277 | EF179290 | AY841688 | EF179313 | EF179255 | EF179332 | x | x | | |
| <i>Diclinanona calycina</i> (Diels) R.E.Fr. | Pirie, M.D. 116 (U) | Peru | KC196270 | KC196271 | — | KC196272 | KC196269 | — | — | | x | x | |
| <i>Diclinanona tessmannii</i> Diels | Maas, P.J.M. et al. 8198 (U) | Peru | AY841611 | DQ125056 | EF179291 | AY841689 | EF179314 | — | EF179333 | | x | x | |
| <i>Dielsiothamnus divaricatus</i> (Diels) R.E.Fr. | Johnson, D.M. 1903 (OWU) | Tanzania | EU169781 | EU169692 | — | EU169759 | EU169736 | — | EU169803 | | x | | |
| <i>Disepalum pulchrum</i> (King) J.Sinclair | Chan, R. 192 (FLAS) | Malaysia | JQ513888 | — | — | GQ139909 | — | — | — | | | | |
| <i>Disepalum platipetalum</i> Merr. | Takeuchi & Sombas 18201 (L) | Indonesia | AY841612 | DQ125057 | EF179292 | AY841690 | DQ125122 | EF179257 | EF179334 | x | x | x | |
| <i>Drepananthus biovulatus</i> (Boerl.) Survesw. & R.M.K.Saunders | Wong 46009 (L) | Indonesia | HM173779 | — | — | HM173751 | HM173693 | — | — | | | | |
| <i>Duckeanthus grandiflorus</i> R.E.Fr | Unpublished | - | UP | — | — | — | — | — | — | | | | |
| <i>Duguetia bahiensis</i> Maas | Amorim, A.M. 800 (U) | Brazil | AY738152 | AY740532 | — | AY740564 | UP | — | — | | x | x | |
| <i>Duguetia cadaverica</i> Huber | Jansen-Jacobs, M.J. 5868 (U) | Guyana | AY738153 | AY740533 | — | AY740565 | UP | — | — | | x | x | |
| <i>Duguetia calycina</i> Benoist | Jansen-Jacobs, M.J. 5661 (U) | Guyana | AY738154 | AY740534 | — | AY740566 | UP | — | — | | x | x | |
| <i>Duguetia cauliflora</i> R.E.Fr | Jansen-Jacobs, M.J. 5687 (U) | Guyana | AY738155 | AY740535 | — | AY740567 | UP | — | — | | x | x | |
| <i>Duguetia chrysea</i> Maas | Maas, P.J.M. 8053 (U) | Brazil | AY841613 | AY740536 | — | AY740568 | AY841435 | — | — | | x | x | |
| <i>Duguetia confinis</i> (Engl. & Diels) Chatrou | Wieringa, J.J. 3290 (WAG) | Gabon | AY738157 | AY740537 | — | AY740569 | UP | — | — | | x | x | |
| <i>Duguetia confusa</i> Maas | Chatrou, L.W. 42 (U) | Costa Rica | AY738158 | AY740538 | — | AY740570 | UP | — | — | | x | x | |
| <i>Duguetia echinophora</i> R.E.Fr. | Maas, P.J.M. 8046 (U) | Brazil | AY738159 | AY740539 | — | AY740571 | UP | — | — | | x | x | |
| <i>Duguetia flagellaris</i> 0378 Huber | Unpublished | - | UP | UP | — | UP | UP | — | — | | x | x | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|--|------------------------------|------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Duguetia flagellaris</i> 0377 Huber | Unpublished | - | UP | — | — | — | UP | — | — | | | | |
| <i>Duguetia furfuracea</i> (A. St.-Hil) Saff. | Unpublished | - | UP | UP | — | UP | UP | — | — | | x | x | |
| <i>Duguetia guianensis</i> R.E.Fr | Chatrou, L.W., UG-NB 33 (U) | Guyana | AY738160 | AY740540 | — | AY740572 | UP | — | — | | x | x | |
| <i>Duguetia hadrantha</i> (Diels) R.E.Fr. | Chatrou, L.W. 181 (U) | Peru | AY738161 | AY740541 | EF179293 | AY740573 | DQ125123 | EF179258 | EF179335 | x | x | x | |
| <i>Duguetia inconspicua</i> Sagot | Unpublished | - | UP | UP | — | UP | UP | — | — | | x | x | |
| <i>Duguetia lepidota</i> (Miq.) Pulle | Unpublished | - | UP | — | — | UP | UP | — | — | | | | |
| <i>Duguetia lucida</i> Urb. | Chatrou, L.W. 367 (U) | Bolivia | AY738162 | AY740542 | — | AY740574 | UP | — | — | | x | x | |
| <i>Duguetia macrocalyx</i> R.E.Fr. | Jansen-Jacobs, M.J. 3011 (U) | Guyana | AY738163 | AY740543 | — | AY740575 | UP | — | — | | x | x | |
| <i>Duguetia macrophylla</i> R.E.Fr. | Maas, P.J.M. 8242 (U) | Peru | AY738164 | AY740544 | — | AY740576 | UP | — | — | | x | x | |
| <i>Duguetia marcgraviana</i> Mart. | Chatrou, L.W. 327 (U) | Bolivia | AY738165 | AY740545 | — | AY740577 | UP | — | — | | x | x | |
| <i>Duguetia megalocarpa</i> R.E.Fr. | Maas, P.J.M. 8045 (U) | Brazil | AY738166 | AY740546 | — | AY740578 | UP | — | — | | x | x | |
| <i>Duguetia moricandiana</i> Mart. | De Carvalho, A.M.V. 3322 (U) | Brazil | AY738167 | AY740547 | — | AY740579 | UP | — | — | | x | x | |
| <i>Duguetia neglecta</i> Sandwith | Jansen-Jacobs, M.J. 5655 (U) | Guyana | AY738168 | AY740548 | — | AY740580 | UP | — | — | | x | x | |
| <i>Duguetia odorata</i> (Diels) J. F. Macbr. | Chatrou, L.W. 207 (U) | Peru | AY738169 | AY740549 | — | AY740581 | UP | — | — | | x | x | |
| <i>Duguetia panamensis</i> Sandl. | Chatrou, L.W. 97 (U) | Costa Rica | AY738170 | AY740550 | — | AY740582 | UP | — | — | | x | x | |
| <i>Duguetia peruviana</i> (R.E.Fr.) J. F. Macbr. | Maas, P.J.M. et al. 8571 (U) | Ecuador | AY738171 | AY740551 | — | AY740583 | UP | — | — | | x | x | |
| <i>Duguetia pycnastera</i> Sandwith | Miralha, J.M.S. 241 (U) | Brazil | AY738172 | AY740552 | — | AY740584 | UP | — | — | | x | x | |
| <i>Duguetia quitarensis</i> Benth. | Chatrou, L.W. 261 (U) | Peru | AY738173 | AY740553 | — | AY740585 | UP | — | — | | x | x | |
| <i>Duguetia riedeliana</i> R.E.Fr. | Maas, P.J.M. 8891 (U) | Brazil | AY738174 | AY740554 | — | AY740586 | UP | — | — | | x | x | |
| <i>Duguetia riparia</i> Huber | Unpublished | - | UP | UP | — | UP | UP | — | — | | x | x | |
| <i>Duguetia rotundifolia</i> R.E.Fr. | Unpublished | - | UP | UP | — | UP | UP | — | — | | x | x | |
| <i>Duguetia salicifolia</i> R.E.Fr. | Cordeiro, I. 915 (U) | Brazil | AY738175 | AY740555 | — | AY740587 | UP | — | — | | x | x | |
| <i>Duguetia sessilis</i> (Vell.) Maas | Maas, P.J.M. 8838 (U) | Brazil | AY738176 | AY740556 | — | AY740588 | UP | — | — | | x | x | |
| <i>Duguetia sooretamae</i> Maas | Maas, P.J.M. 8827 (U) | Brazil | AY738177 | AY740557 | — | AY740589 | DQ861746 | — | — | | x | x | |
| <i>Duguetia spixiana</i> Mart. | Unpublished | - | UP | UP | — | UP | UP | — | — | | x | x | |
| <i>Duguetia staudtii</i> (Engl. & Diels) Chatrou | van Andel, T.R. 3290 (U) | Cameroon | AY738178 | AY740558 | EF179294 | AY740590 | DQ125124 | EF179259 | EF179336 | x | x | x | |
| <i>Duguetia stelechantha</i> (Diels) R.E.Fr. | Maas, P.J.M. 8058 (U) | Brazil | AY738179 | AY740559 | — | AY740591 | UP | — | — | | x | x | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|---------------------------------|------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Duguetia surinamensis</i> R.E.Fr. | Maas, P.J.M. 8057 (U) | Brazil | AY738180 | AY740560 | — | AY740592 | UP | — | — | | x | x | |
| <i>Duguetia ulei</i> (Diels) R.E.Fr. | Miralha, J.M.S. 243 (U) | Brazil | AY738181 | AY740561 | — | AY740593 | UP | — | — | | x | x | |
| <i>Duguetia uniflora</i> (DC.) Mart. | Coelho, D. INPA-3711 (U) | Brazil | AY738182 | AY740562 | — | AY740594 | UP | — | — | | x | x | |
| <i>Duguetia yeshidan</i> Sandwith | Jansen-Jacobs, M.J. 6129 (U) | Guyana | AY738183 | AY740563 | — | AY740595 | UP | — | — | | x | x | |
| <i>Encisanthum fuscum</i> (King) Airy Shaw | Kessler, P.J.A. PK 3222 (L) | Thailand | AY318973 | AY518787 | JX544779 | AY319085 | JX544792 | — | — | | x | | x |
| <i>Encisanthum membranifolium</i> J.Sinclair | Kessler, P.J.A. 3198 (L) | Thailand | AY318974 | AY518788 | — | AY319086 | — | — | — | | | | |
| <i>Encisanthum paradoxum</i> Becc. | Kessler, P.J.A. 2746 (L) | Indonesia | AY318975 | AY518789 | — | AY319087 | — | — | — | | | | |
| <i>Ephedranthus boliviensis</i> Chatrou & Pirie | Chatrou, L.W. 301 (U) | Bolivia | AY841614 | — | — | AY841692 | — | — | — | | | | |
| <i>Ephedranthus parviflorus</i> S. Moore | Prance, G.T. et al. 19246 (U) | Brazil | AY841615 | UP | — | AY841693 | AY841462 | — | — | | x | | x |
| <i>Ephedranthus</i> sp 0284 | Maas, P.J.M. 8826 (U) | Brazil | AY841616 | AY841396 | AY841407 | AY841694 | AY841463 | AY841376 | AY841554 | x | x | | x |
| <i>Ephedranthus</i> sp 0105 | Chatrou, L.W. et al. 173 (U) | Peru | AY319061 | — | — | AY319175 | AY841464 | — | — | | | | |
| <i>Fenerivia chapelieri</i> (Baill.) R.M.K. Saunders | Ludovic & Ralotoarivony 221 (P) | Madagascar | JF810387 | JF810375 | JQ723788 | JF810399 | UP | — | — | | x | | |
| <i>Fenerivia ghesquiereana</i> (Cavaco&Keraudren) R.M.K. Saunders | Randrianaivo, R., 282 (WAG) | Madagascar | JF810389 | JF810377 | UP | JF810401 | UP | — | — | | x | | |
| <i>Fissistigma glaucescens</i> (Hance) Merr. | Law, C.L. 00/07b (L) | Hong Kong | AY743444 | AY743482 | — | AY743463 | — | — | — | | | | |
| <i>Fissistigma uonicum</i> (Dunn.) Merr. | Law, C.L. 00/05 (L) | Hong Kong | AY841617 | — | — | AY841695 | — | — | — | | | | |
| <i>Fitzalania bidwilli</i> I (Benth.) Jessup | Sankowsky BRI:4139 | Australia | JQ723851 | JQ723764 | — | JQ723904 | — | — | — | | | | |
| <i>Fitzalania bidwilli</i> II (Benth.) Jessup | Sankowsky BRI:3179 | Australia | JQ723852 | JQ723765 | JQ723789 | JQ723905 | — | — | — | | | | |
| <i>Fitzalania heteropetala</i> (F. Muell.) F. Muell. | Sankowsky BRI:4140 | Australia | JQ723853 | JQ723766 | JQ723790 | JQ723906 | — | — | — | | | | |
| <i>Friesodielsia desmoides</i> (Craib) Steenis | Kessler, P.J.A. 3189 (L) | Thailand | AY841618 | JQ768577 | JQ768612 | AY841696 | JQ768656 | — | — | | x | x | |
| <i>Friesodielsia</i> sp. | Wieringa, J.J. 3605 (WAG) | Gabon | AY841619 | JQ768580 | JQ768615 | AY841697 | JQ768659 | — | — | | x | x | |
| <i>Fusaea longifolia</i> (Aubl.) Saff. | Chatrou, L.W. 175 (U) | Peru | AY841620 | — | — | AY841698 | — | — | — | | | | |
| <i>Fusaea peruviana</i> R.E.Fr. | Chatrou, L.W. 179 (U) | Peru | AY743445 | AY743483 | EF179295 | AY743464 | AY841436 | EF179260 | EF179337 | x | x | x | |
| <i>Goniothalamus griffithii</i> Hook.f. & Thomson | Kessler, P.J.A. 3188 (L) | Thailand | AY743446 | AY743484 | EF179296 | AY743465 | DQ125125 | EF179261 | EF179338 | x | x | x | |
| <i>Goniothalamus laoticus</i> (Finet & Gagnep.) BÅçn | Kessler, P.J.A. 3203 (L) | Thailand | AY841621 | — | — | AY841699 | — | — | — | | | | |
| <i>Goniothalamus tapis</i> Miq. | Kessler, P.J.A. 3193 (L) | Thailand | AY841622 | DQ125058 | EF179297 | AY841700 | DQ125126 | EF179262 | EF179339 | x | x | x | |
| <i>Greenwayodendron oliveri</i> (Engl.) Verdc. | Jongkind, C.C.H. 1795 (WAG) | Ghana | AY743451 | AY743489 | AY841408 | AY743470 | AY841465 | AY841377 | AY841555 | x | x | | x |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|--|--------------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Greenwayodendron suaveolens</i> (Engl. & Diels) Verdc. | <i>Semsei 2376</i> (K) | Kenya | AY841524 | — | — | AY841538 | — | — | — | | | | |
| <i>Guamia</i> sp. <i>mexico</i> | <i>Rainer, H. 1593</i> (WU) | Mexico | AY841623 | UP | — | AY841701 | UP | — | — | | | | x |
| <i>Guatteria aeruginosa</i> | <i>Chatrou, L.W. 66</i> (U) | Costa Rica | AY740958 | AY740909 | EF179299 | AY741007 | DQ125136 | EF179264 | EF179341 | x | x | x | |
| <i>Guatteria alata</i> Maas & van Setten | <i>Mori, S.A. 2894</i> (U) | Panama | AY740959 | AY740910 | — | AY741008 | DQ125137 | — | — | | x | x | |
| <i>Guatteria allenii</i> R.E.Fr. | <i>Mori, S.A. 2952</i> (U) | Panama | AY740960 | AY740911 | — | AY741009 | DQ125138 | — | — | | x | x | |
| <i>Guatteria alta</i> R.E.Fr. | <i>Gentry, A.H. & Monsalve, M. 48250</i> (U) | Colombia | DQ124941 | DQ125065 | — | DQ124999 | DQ125139 | — | — | | x | x | |
| <i>Guatteria alutacea</i> Diels | <i>Chatrou, L.W. et al. 339</i> (U) | Bolivia | AY740961 | AY740912 | — | AY741010 | DQ125140 | — | — | | x | x | |
| <i>Guatteria amplifolia</i> Triana & Planch | <i>Chatrou, L.W. et al. 111</i> (U) | Costa Rica | DQ124942 | DQ125066 | — | DQ125000 | DQ125141 | — | — | | x | x | |
| <i>Guatteria anomala</i> R.E.Fr. | <i>Ishiki, M. 2233</i> (U) | Mexico | AY740962 | AY740913 | EF179298 | AY741011 | AY841437 | EF179263 | EF179340 | x | x | x | |
| <i>Guatteria atra</i> Sandwith | <i>Gopaul, D. & Maas, P.J.M. 2791</i> (U) | Guyana | AY740963 | AY740914 | — | AY741012 | DQ125142 | — | — | | x | x | |
| <i>Guatteria australis</i> A. St.-Hil. | <i>Lobao, A.Q. & Fiaschi, P. 499</i> (U) | Brazil | AY740964 | AY740915 | — | AY741013 | AY841438 | — | — | | x | x | |
| <i>Guatteria blainii</i> (Griseb.) Urb. | <i>Maas, P.J.M. 6443</i> (U) | Dominican Republic | AY740965 | AY740916 | — | AY741014 | DQ125143 | — | — | | x | x | |
| <i>Guatteria boliviana</i> H.J.P. Winkl. | <i>Solomon, J.C. 10789</i> (U) | Bolivia | DQ124943 | DQ125067 | — | DQ125001 | DQ125144 | — | — | | x | x | |
| <i>Guatteria brevicuspis</i> syn. <i>blepharophylla</i> Mart. | <i>Prance, G.T. 16328</i> (U) | Brazil | AY740966 | AY740917 | — | AY741015 | DQ125145 | — | — | | x | x | |
| <i>Guatteria buchtienii</i> R.E.Fr. | Unpublished | - | | | — | | UP | — | — | | | | |
| <i>Guatteria candolleana</i> Schtdl. | <i>Harley, R.M. et al. 17360</i> (U) | Brazil | DQ124946 | DQ125070 | — | DQ125004 | DQ125148 | — | — | | x | x | |
| <i>Guatteria caribaea</i> Urb. | <i>Tuxill, J. 89</i> (U) | Dominican Republic | AY740967 | AY740918 | — | AY741016 | DQ125149 | — | — | | x | x | |
| <i>Guatteria chiriquiensis</i> R.E.Fr. | <i>Chatrou, L.W. et al. 43</i> (U) | Costa Rica | AY740968 | AY740919 | — | AY741017 | DQ125150 | — | — | | x | x | |
| <i>Guatteria diospyroides</i> Baill. | <i>Chatrou, L.W. 84</i> (U) | Costa Rica | AY740969 | AY740920 | — | AY741018 | DQ125152 | — | — | | x | x | |
| <i>Guatteria discolor</i> R.E.Fr. | <i>Maas, P.J.M. et al. 9030</i> (U) | Brazil | AY740970 | AY740921 | — | AY741019 | DQ125153 | — | — | | x | x | |
| <i>Guatteria dumetorum</i> R.E.Fr. | <i>FLORPAN 2497</i> (U) | Panama | AY740971 | AY740922 | — | AY741020 | DQ125154 | — | — | | x | x | |
| <i>Guatteria dusenii</i> R.E.Fr. syn. <i>australis</i> | <i>Dusén, P. 13752</i> (S) | Brazil | DQ124948 | DQ125072 | — | DQ125006 | DQ125155 | — | — | | x | x | |
| <i>Guatteria elata</i> R.E.Fr. | <i>Chatrou, L.W. et al. 252</i> (U) | Peru | AY740972 | AY740923 | — | AY741021 | DQ125156 | — | — | | | | |
| <i>Guatteria elegantissima</i> R.E.Fr. | <i>Gentry, A.H. 56948</i> (U) | Colombia | AY740973 | AY740924 | — | AY741022 | DQ125157 | — | — | | x | x | |
| <i>Guatteria ferruginea</i> A. St.-Hil. | <i>Lobao, A.Q. 643</i> (U) | Brazil | DQ124949 | DQ125073 | — | DQ125007 | DQ125158 | — | — | | x | x | |
| <i>Guatteria foliosa</i> Benth. | <i>Chatrou, L.W. 325</i> (U) | Bolivia | AY740974 | AY740925 | — | AY741023 | DQ125159 | — | — | | x | x | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|--|---|---------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Guatteria galeottiana</i> Baill. | Beaman, J.H. 6121 (U) | Mexico | DQ124950 | DQ125074 | — | DQ125008 | DQ125160 | — | — | | x | x | |
| <i>Guatteria glabrescens</i> R.E.Fr. syn. <i>australis</i> | Maas, P.J.M. et al. 8816 (U) | Brazil | AY740975 | AY740926 | — | AY741024 | DQ125161 | — | — | | x | x | |
| <i>Guatteria guianensis</i> (Aubl.) R.E.Fr. | Webber, A.C. et al. 1884 (U) | Brazil | AY740976 | AY740927 | — | AY741025 | DQ125163 | — | — | | x | x | |
| <i>Guatteria heteropetala</i> Benth. | Unpublished | - | UP | UP | — | — | UP | — | — | | | | |
| <i>Guatteria heterotricha</i> R.E.Fr. | Monsalve, B.M. 1262 (U) | Colombia | AY740977 | AY740928 | — | AY741026 | DQ125164 | — | — | | x | x | |
| <i>Guatteria hyposericea</i> Diels | Chatrou, L.W. et al. 375 (U) | Bolivia | AY740978 | AY740929 | — | AY741027 | DQ125166 | — | — | | x | x | |
| <i>Guatteria inuncta</i> R.E.Fr. syn. <i>diospyroides</i> | Liesner, R.L. 14631 (U) | Costa Rica | AY740979 | AY740930 | — | AY741028 | DQ125167 | — | — | | x | x | |
| <i>Guatteria inundata</i> Mart. | Chatrou, L.W. et al. 191 (U) | Peru | AY740980 | AY740931 | — | AY741029 | DQ125168 | — | — | | x | x | |
| <i>Guatteria jefensis</i> Barringer | Valdespino, I.A. 685 (U) | Panama | AY740981 | AY740932 | — | AY741030 | DQ125169 | — | — | | x | x | |
| <i>Guatteria latifolia</i> (Mart.) R.E.Fr. | Lobao, A.Q. 544 (U) | Brazil | AY740982 | AY740933 | — | AY741031 | DQ125170 | — | — | | x | x | |
| <i>Guatteria latisepala</i> R.E.Fr. | Sánchez, D. et al. 404 (U) | Colombia | DQ124953 | DQ125077 | — | DQ125011 | DQ125171 | — | — | | x | x | |
| <i>Guatteria liesneri</i> D.M. Johnson & N.A. Murray | Cid F., C.A. 8403 (U) | Brazil | AY740983 | AY740934 | — | AY741032 | DQ125172 | — | — | | x | x | |
| <i>Guatteria macropus</i> Mart. | Pirani, J.R. 2725 (U) | Brazil | AY740984 | AY740935 | — | AY741033 | DQ125174 | — | — | | x | x | |
| <i>Guatteria maypurensis</i> Kunth | Jansen-Jacobs, M..J. et al. 5416 (U) | Guyana | AY740985 | AY740936 | — | AY741034 | DQ125175 | — | — | | x | x | |
| <i>Guatteria megalophylla</i> Diels | Chatrou, L.W. et al. 216 (U) | Pery | AY740986 | AY740937 | — | AY741035 | DQ125176 | — | — | | x | x | |
| <i>Guatteria</i> cf. <i>meliodora</i> R.E.Fr. | Maas, P.J.M. et al. 9231 (U) | Brazil | DQ124955 | DQ125079 | — | DQ125013 | DQ125177 | — | — | | x | x | |
| <i>Guatteria mexiae</i> R.E.Fr. | Mori, S.A. et al. 9722 (U) | Brazil | DQ124956 | DQ125080 | — | DQ125014 | DQ125178 | — | — | | x | x | |
| <i>Guatteria modesta</i> Diels | Dulmen, A. van et al. 183 (U) | Colombia | DQ124957 | DQ125081 | — | DQ125015 | DQ125179 | — | — | | x | x | |
| <i>Guatteria multivenia</i> Diels syn. <i>guianensis</i> | Maas, P.J.M. 8511 (U) | Ecuador | AY740987 | AY740938 | — | AY741036 | DQ125180 | — | — | | x | x | |
| <i>Guatteria notabilis</i> Mello-Silva & Pirani | Lobao, A.Q. 623 (U) | Brazil | DQ124958 | DQ125082 | — | DQ125016 | DQ125181 | — | — | | x | x | |
| <i>Guatteria oligocarpa</i> Mart. | Maas, P.J.M. 7006 (U) | Brazil | AY740988 | AY740939 | — | AY741037 | DQ125182 | — | — | | x | x | |
| <i>Guatteria olivacea</i> R.E.Fr. | Chatrou, L.W. et al. 209 (U) | Peru | AY740989 | AY740940 | — | AY741038 | DQ125183 | — | — | | x | x | |
| <i>Guatteria oliviformis</i> Donn. Sm. | Chatrou, L.W. et al. 80 (U) | Costa Rica | AY740990 | AY740941 | — | AY741039 | DQ125184 | — | — | | x | x | |
| <i>Guatteria ouregou</i> (Aubl.) Dunal | Scharf, U. 85 (U) | French Guiana | AY740991 | AY740942 | — | AY741040 | AY741040 | — | — | | x | x | |
| <i>Guatteria pacifica</i> 821 R.E.Fr. | Gentry, A.H. & Faber-Langendoen, D. 62881 (U) | Colombia | DQ124959 | DQ125083 | — | DQ125017 | DQ125186 | — | — | | x | x | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|--|---|---------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Guatteria pacifica</i> 887 R.E.Fr. | <i>Cuatrecasas, J. 17150</i> (U) | Colombia | DQ124960 | DQ125084 | — | DQ125018 | DQ125187 | — | — | | x | x | |
| <i>Guatteria paraensis</i> R.E.Fr. | <i>Froes, R. 1753</i> (U) | Brazil | DQ124961 | DQ125085 | — | DQ125019 | DQ125188 | — | — | | x | x | |
| <i>Guatteria parvifolia</i> R.E.Fr. syn. <i>australis</i> | <i>Gottsberger, G.K. 573007</i> (U) | Brazil | AY740992 | AY740943 | — | AY741041 | DQ125189 | — | — | | x | x | |
| <i>Guatteria pittieri</i> R.E.Fr. | <i>Werff, H.H. van der 9767</i> (U) | Colombia | AY740993 | AY740944 | — | AY741042 | DQ125190 | — | — | | x | x | |
| <i>Guatteria poeppigiana</i> Mart. | <i>Prance, G.T. & Pennington, T.D. 1775</i> (U) | colombia | AY740993 | AY740944 | — | AY741042 | DQ125190 | — | — | | x | x | |
| <i>Guatteria pogonopus</i> Mart. | <i>Kollmann, L. et al. 202</i> (U) | Brazil | DQ124963 | DQ125087 | — | DQ125021 | DQ125192 | — | — | | x | x | |
| <i>Guatteria pohliana</i> Schldt. | <i>Anderson, W.R. et al. 35703</i> (U) | Brazil | DQ124964 | DQ125088 | — | DQ125022 | DQ125193 | — | — | | x | x | |
| <i>Guatteria polyantha</i> R.E.Fr. | <i>Costa, R.C Monteiro da 284</i> (S) | Brazil | DQ124965 | DQ125089 | — | DQ125023 | DQ125194 | — | — | | x | x | |
| <i>Guatteria polycarpa</i> R.E.Fr. syn. <i>australis</i> | <i>Dusén, P. 7414</i> (S) | Brazil | DQ124966 | DQ125090 | — | DQ125024 | DQ125195 | — | — | | x | x | |
| <i>Guatteria pudica</i> N.Zamora & Maas | <i>Chatrou, L.W. 107</i> (U) | Costa Rica | AY740994 | AY740945 | JQ769093 | AY741043 | DQ125197 | JQ513884 | FJ842397 | x | x | x | |
| <i>Guatteria punctata</i> (Aubl.) R.A. Howard | <i>Mohlino, J.F. 1593</i> (U) | French Guiana | AY740995 | AY740946 | — | AY741044 | DQ125198 | — | — | | x | x | |
| <i>Guatteria puncticulata</i> R.E.Fr. syn. <i>modesta</i> | <i>Chatrou, L.W. et al. 172</i> (U) | Peru | AY740996 | AY740947 | — | AY741045 | DQ125199 | — | — | | x | x | |
| <i>Guatteria ramiflora</i> (D.R. Simpson) Erkens & Maas | Unpublished | - | UP | UP | — | | UP | — | — | | | | |
| <i>Guatteria recurvisepala</i> R.E.Fr. | <i>Chatrou, L.W. et al. 61</i> (U) | Costa Rica | AY740997 | AY740948 | — | AY741046 | DQ125200 | — | — | | x | x | |
| <i>Guatteria reflexa</i> R.E.Fr. syn. <i>australis</i> | <i>Glaziou, L.A. 5725</i> (S) | Brazil | DQ124968 | DQ125092 | — | DQ125026 | DQ125201 | — | — | | x | x | |
| <i>Guatteria rhamnoides</i> R.E.Fr. syn. <i>glauca</i> | Unpublished | - | UP | UP | — | — | UP | — | — | | | | |
| <i>Guatteria rotundata</i> Maas & van Setten | <i>Mori, S.A. 5531</i> (U) | Panama | AY740998 | AY740949 | — | AY741047 | DQ125204 | — | — | | x | x | |
| <i>Guatteria rupestris</i> Mello-Silva & Pirani | <i>CFCR 4116</i> (U) | Brazil | AY740999 | AY740950 | — | AY741048 | DQ125205 | — | — | | x | x | |
| <i>Guatteria salicifolia</i> R.E.Fr. syn. <i>australis</i> | Unpublished | - | UP | UP | — | | UP | — | — | | | | |
| <i>Guatteria scandens</i> Ducke | <i>Jansen-Jacobs, M.J. et al. 5494</i> (U) | Guyana | DQ124971 | DQ125095 | — | DQ125029 | DQ125207 | — | — | | x | x | |
| <i>Guatteria schlechtendaliana</i> Mart. | <i>Kollmann, L. et al. 871</i> (U) | Brazil | DQ124972 | DQ125096 | — | DQ125030 | DQ125208 | — | — | | x | x | |
| <i>Guatteria schomburgkiana</i> Mart. | <i>Scharf, U. 60</i> (U) | Guyana | AY741001 | AY740952 | — | AY741050 | DQ125209 | — | — | | x | x | |
| <i>Guatteria schunkevigoi</i> D.R. Simpson | <i>Schunke, V.J. 3551</i> (S) | Peru | DQ124973 | DQ125097 | — | DQ125031 | DQ125210 | — | — | | x | x | |
| <i>Guatteria scytophylla</i> Diels | <i>Maas, P.J.M. et al. 6956</i> (U) | Brazil | AY741002 | AY740953 | — | AY741051 | DQ125211 | — | — | | x | x | |
| <i>Guatteria sellowiana</i> 0557 Schldt. | <i>Lobao, A.Q., 557</i> (U) | Brazil | AY741003 | AY740954 | — | AY741052 | DQ125212 | — | — | | x | x | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|---|-------------------|--------------------------------|-------------|-------------|-------------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Guatteria sellowiana</i> 511 Schldl. | Unpublished | - | UP | UP | — | — | UP | — | — | | | | |
| <i>Guatteria sessilicarpa</i> Maas & van Setten | McPherson, G. 12599 (U) | Panama | AY741004 | AY740955 | — | AY741053 | DQ125213 | — | — | | x | x | |
| <i>Guatteria sessilis</i> R.E.Fr. syn. <i>schomburgkiana</i> | Liesner, R. 8546 (U) | Venezuela | DQ124975 | DQ125099 | — | DQ125033 | DQ125215 | — | — | | x | x | |
| <i>Guatteria sordida</i> var. <i>ovalis</i> R.E.Fr. syn. <i>australis</i> | Riedel, L. 1689 (S) | Brazil | DQ124976 | DQ125100 | — | DQ125034 | DQ125216 | — | — | | x | x | |
| <i>Guatteria stipitata</i> R.E.Fr. | Jaramillo, N. & Chamik, D. 820 (U), | Peru | DQ124978 | DQ125102 | — | DQ125036 | DQ125218 | — | — | | x | x | |
| <i>Guatteria subsessilis</i> Mart. | Maas, P.J.M. et al. 8684 (U) | Bolivia | DQ124979 | DQ125103 | — | DQ125037 | DQ125037 | — | — | | x | x | |
| <i>Guatteria tonduzii</i> Diels | Chatrou, L.W. et al. 121 (U) | Costa Rica | AY741005 | AY740956 | — | AY741054 | DQ125228 | — | — | | x | x | |
| <i>Guatteria trichoclonia</i> Diels | Schunke, V.J. 14061 (U) | Peru | DQ124988 | DQ125112 | — | DQ125046 | DQ125229 | — | — | | x | x | |
| <i>Guatteria venezuelana</i> R.E.Fr. | Wagfield, R. & van der Werff, H.H. 6688 (U) | Venezuela | DQ124989 | DQ125113 | — | DQ125047 | DQ125230 | — | — | | x | x | |
| <i>Guatteria verruculosa</i> R.E.Fr. | Fosberg, F.R. 19126 (S) | Colombia | DQ124990 | DQ125114 | — | DQ125048 | DQ125231 | — | — | | x | x | |
| <i>Guatteria villosissima</i> A. St.-Hil. | Lobao, A.Q. 630 (U) | Brazil | AY741006 | AY740957 | — | AY741055 | DQ125232 | — | — | | x | x | |
| <i>Guatteria wachenheimii</i> Benoist | Scharf, U. 43 (U) | Guyana | DQ124991 | DQ125115 | — | DQ125049 | DQ125233 | — | — | | x | x | |
| <i>Haplostichanthus longirostris</i> (Scheff.) Heusden | Takeuchi 15656 (L) | Papua New Guinea | AY318979 | AY518826 | — | AY319091 | — | — | — | | | | |
| <i>Hexalobus crispiflorus</i> A. Rich. | Sosef, M.S.M. 2287 (WAG) | Gabon | EU169782 | EU169693 | EU169713 | EU169760 | EU169737 | — | EU169804 | | x | | |
| <i>Hexalobus salicifolius</i> Engl. | Sosef, M.S.M. 2376 (WAG) | Gabon | EU169783 | EU169694 | EU169714 | EU169761 | EU169738 | — | EU169805 | | x | | |
| <i>Hornsuchia citriodora</i> D.M.Johnson | Maas, P.J.M. 8828 (U) | Brazil | AY841625 | — | — | AY841703 | — | — | — | | | | |
| <i>Hubera henricii</i> | Dorr & Koenders 3033 (WAG) | Madagascar | — | — | JX544880 | JX544870 | JX544860 | — | — | | | | |
| <i>Hubera jenkinsii</i> | Chaowasku, T. DS (L) | Thailand | — | — | JX544842 | JX544803 | JX544812 | — | — | | | | |
| <i>Hubera korinti yvonne</i> | Ratnayake, R.M.C.S. 2/03 (HKU) | Sri Lanka | EU522289 | EU522234 | JX544877 | EU522179 | EU522124 | EU522345 | — | | x | | |
| <i>Hubera nitidissima</i> | Ford AF 4967 | Australia | — | JQ889989 | JQ889986 | JQ889988 | JQ889981 | — | — | | | | |
| <i>Hubera pendula</i> Capuron ex G.E.Schatz & Le Thomas | Rabehohitra 2386 (K) | Madagascar | AY319030 | AY518852 | JQ889987 | AY319144 | JQ889982 | — | — | | x | | |
| <i>Hubera perrieri</i> | Capuron 20.977-SF (K) | Madagascar | — | — | JX544881 | JX544871 | JX544861 | — | — | | | | |
| <i>Hubera rumphii</i> (Blume ex Hensch.) Merr. | Van Balgooy, M. 5654 (L) | Indonesia | AY319031 | AY518791 | JX544841 | AY319145 | JX544811 | — | — | | x | | |
| <i>Hubera stuhlmannii</i> (Engl.) Verdc. | Luke 1424 (K) | Kenya | AY319035 | AY518853 | JX544882 | AY319149 | JX544862 | — | — | | x | | |
| <i>Hubera tanganyikensis</i> | Couvreur 66 (WAG) | Tanzania | — | — | JX544883 | JX544872 | JX544863 | — | — | | | | |
| <i>Isolona campanulata</i> Engl. & Diels | UUBG 86GR00240 | UUBG, of tropical | AY238954 | AY238963 | EU169715 | AY231287 (intron) | DQ125127 | EF179266 | EU169806 | x | x | x | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | | |
|--|------------------------------|--|--------------------------------|-------------|-------------|-------------------------------|------------------|------------------|------------------|----------------------|----|-----|-----|---|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC | |
| <i>Isolona cooperi</i> Hutch. & Dalziel ex G.P.Cooper & Record | UUBG 84GR00382 | African origin UUBG, originating from Ivory Coast | AY841626 | — | EU216636 | AY238947 (spacer) AY841704 | — | EU216657 | — | EU216612 | | | | |
| <i>Klarobelia inundata</i> Chatrou | Chatrou, L.W. 205 (U) | Peru | AY743452 | AY743490 | AY841409 | AY743471 | AY841469 | AY841378 | AY841556 | x | x | | | x |
| <i>Klarobelia megalocarpa</i> Chatrou | Maas et al. 8521 (U) | Ecuador | AY319062 | AY518866 | — | AY319176 | AY841470 | — | — | | x | | | x |
| <i>Klarobelia stipitata</i> Chatrou | Chatrou, L.W. 113 (U) | Costa Rica | AY841628 | UP | — | AY841706 | AY841472 | — | — | | x | | | x |
| <i>Letestudoxa bella</i> Pellegr. | Wieringa, J.J. 2797 (WAG) | Gabon | AY841629 | DQ125059 | EF179302 | AY841707 | DQ125128 | EF179267 | EF179344 | x | x | x | | |
| <i>Letestudoxa glabrifolia</i> Chatrou & Repetur | Breteler, F.J. 12858 (WAG) | Gabon | AY841630 | UP | — | AY841708 | UP | — | — | | x | x | | |
| <i>Lettowianthus stellatus</i> Diels | Robertson, A. 7505 (WAG) | Kenya | EU169775 | EU169686 | — | EU169753 | EU169730 | — | EU169797 | | x | | | |
| <i>Maasia discolor</i> (Diels) Mols, Kessler & Rogstad | Takeuchi & Ama 16394 (L) | Papua New Guinea | AY319021 | AY518872 | AY841416 | AY319135 | AY841500 | AY841385 | AY841563 | x | x | | | |
| <i>Maasia glauca</i> (Hassk.) Mols, Kessler & Rogstad | Mols, J.B. 20 (L) | Indonesia | AY319023 | AY518871 | — | AY319137 | AY841501 | — | — | | x | | | |
| <i>Maasia sumatrana</i> (Miq.) Mols, Kessler & Rogstad | SAN 143918 (SAN) | Malaysia | AY319039 | AY518873 | AY841418 | AY319153 | AY841503 | AY841387 | AY841565 | x | x | | | |
| <i>Malmea dielsiana</i> R.E.Fr. | Chatrou, L.W. 122 (U) | Peru | AY238955 | AY238964 | AY841410 | AY231288 (intron) | AY841473 | AY841379 | AY841557 | x | x | | | x |
| <i>Malmea</i> sp. | Chatrou, L.W. 8 (U) | Peru | AY841527 | AY841397 | AY841411 | AY841541 | AY841475 | AY841380 | AY841558 | x | x | | | x |
| <i>Malmea surinamensis</i> Chatrou | Jansen-Jacobs, M.J. 6207 (U) | Suriname | AY743453 | AY743491 | — | AY743472 | AY841476 | — | — | | x | | | x |
| <i>Marsypopetalum crassum</i> (R. Parker) B. Xue & R.M.K. Saunders | Chalermglin 521212-1 (HKU) | Thailand | HQ286577 | HQ286571 | JQ723792 | HQ286583 | — | — | — | | | | | |
| <i>Marsypopetalum heteropetalum</i> | Unpublished | - | — | — | UP | UP | UP | — | — | | | | | |
| <i>Marsypopetalum kraburianum</i> | Unpublished | - | — | — | UP | UP | UP | — | — | | | | | |
| <i>Marsypopetalum littorale</i> (Blume) B.Xue & R.M.K.Saunders | Rastini 153 (L) | Indonesia | AY319026 | AY518835 | JX544827 | AY319140 | JX544804 | — | — | | x | | | x |
| <i>Marsypopetalum lucidum</i> (Merr.) B. Xue & R.M.K. Saunders | Kanehira 2606 (NY) | Philippines | HQ286578 | HQ286572 | — | HQ286584 | — | — | — | | | | | |
| <i>Marsypopetalum modestum</i> (Pierre) B. Xue & R.M.K. Saunders | Unpublished | - | UP | UP | UP | UP | UP | — | — | | x | | | x |
| <i>Marsypopetalum pallidum</i> (Blume) Kurz | Kessler, P.J.A. 3192 (L) | Thailand | AY318980 | AY518834 | — | AY319092 | — | — | — | | | | | |
| <i>Marsypopetalum triste</i> (Pierre) B. Xue & R.M.K. Saunders | Poillane 19622 (NY) | Vietnam | HQ286579 | HQ286573 | — | HQ286585 | — | — | — | | | | | |
| <i>Meiocarpidium lepidotum</i> (Oliv.) Engl. & Diels | Breteler, F. 13947 (WAG) | Gabon | EU169776 | EU169687 | UP | EU169754 | EU169731 | — | EU169798 | | x | | | |
| <i>Meiogyne cylindrocarpa</i> (Burck) Heusden | Sankowsky BRI:3175 | Australia | JQ723856 | JQ723769 | JQ723795 | JQ723909 | — | — | — | | | | | |
| <i>Meiogyne cylindrocarpa</i> subsp. <i>trichocarpa</i> I Jessup | Sankowsky BRI:3190 | Australia | JQ723857 | JQ723770 | JQ723796 | JQ723910 | — | — | — | | | | | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|---------------------------------------|------------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Meiogyne cylindrocarpa</i> subsp. <i>trichocarpa</i> II Jessup | Sankowsky BRI: 4141 | Australia | JQ723858 | JQ723771 | JQ723797 | JQ723911 | — | — | — | | | | |
| <i>Meiogyne glabra</i> Heusden | Barker & Vinas 66735 (L) | Papua New Guinea | JQ723859 | JQ723772 | — | JQ723912 | — | — | — | | | | |
| <i>Meiogyne hainanensis</i> (Merr.) BÅçn | How P: 70628 | China | JQ723860 | JQ723773 | — | JQ723913 | — | — | — | | | | |
| <i>Meiogyne hirsuta</i> (Jessup) Jessup | Sankowsky BRI:3151 | Australia | JQ723861 | JQ723774 | JQ723798 | JQ723914 | — | — | — | | | | |
| <i>Meiogyne lecardii</i> (Guillaumin) Heusden | MacKee L:16292 | New Caledonia | JQ723862 | JQ723775 | JQ723799 | JQ723915 | — | — | — | | | | |
| <i>Meiogyne mindorensis</i> (Merr.) Heusden | Podzorski L:SMHI 76 | Philippines | JQ723863 | JQ723776 | JQ723800 | JQ723916 | — | — | — | | | | |
| <i>Meiogyne monosperma</i> (Hook. f. & Thomson) Heusden | Rogstad L:920 | Malaysia | JQ723864 | JQ723777 | — | JQ723917 | — | — | — | | | | |
| <i>Meiogyne pannosa</i> (Dalzell) J. Sinclair | Indu L:2457 | India | JQ723865 | JQ723778 | JQ723801 | JQ723918 | — | — | — | | | | |
| <i>Meiogyne stenopetala</i> (F. Muell.) Heusden | Sankowsky BRI:3193 | Australia | JQ723866 | JQ723779 | JQ723803 | JQ723919 | — | — | — | | | | |
| <i>Meiogyne verrucosa</i> Jessup | Sankowsky BRI:3188 | Australia | JQ723867 | JQ723780 | JQ723804 | JQ723920 | — | — | — | | | | |
| <i>Melodorum fruticosum</i> Lour. | Chalermglin 440214-2 (L) | Thailand | AY319071 | AY518878 | — | — | — | — | — | | | | |
| <i>Melodorum</i> cf. <i>fruticosum</i> Mols 2 | Mols, J.B. 2 (L) | Indonesia | AY319072 | AY518879 | — | — | — | — | — | | | | |
| <i>Mezzettia parviflora</i> Becc. | Okada 3388 (L) | Indonesia | AY318983 | AY518881 | — | AY319095 | — | — | — | | | | |
| <i>Miliusa amplexicaulis</i> Ridl. | Chaowasku, T. 54 (L) | Thailand | — | — | JQ690479 | JQ690478 | JQ690480 | — | — | | | | |
| <i>Miliusa balansae</i> Finet & Gagnep | Harder et al. 7233 (MO) | Vietnam | — | — | JQ690483 | JQ690482 | JQ690484 | — | — | | | | |
| <i>Miliusa brahei</i> (F. Muell.) Jessup | Ford AF 5005 | Australia | — | — | JQ690431 | JQ690430 | JQ690432 | — | — | | | | |
| <i>Miliusa butonensis</i> | Coode 6279 (L) | Indonesia | — | — | JQ690435 | JQ690434 | JQ690436 | — | — | | | | |
| <i>Miliusa campanulata</i> Pierre | Chalermglin 44047-11 (TISTR, Bangkok) | Thailand | AY318984 | AY518842 | — | AY319096 | — | — | — | | | | |
| <i>Miliusa cuneata</i> Craib | Chalermglin 440214-7 (L) | Thailand | AY318985 | AY518844 | — | AY319097 | — | — | — | | | | |
| <i>Miliusa dioeca</i> | Unpublished | - | UP | — | UP | UP | UP | — | — | | | | |
| <i>Miliusa fusca</i> Pierre | Chaowasku, T. 46 (L) | Thailand | — | — | JQ690443 | JQ690442 | JQ690444 | — | — | | | | |
| <i>Miliusa horsfieldii</i> (Benn.) Pierre | Mols, J.B. 1 (L) | Indonesia | AY318986 | AY518849 | JQ690447 | AY319098 | JQ690448 | — | — | | x | | x |
| <i>Miliusa intermedia</i> | Unpublished | - | — | — | UP | UP | UP | — | — | | | | |
| <i>Miliusa koolsii</i> (Kosterm.) J. Sinclair | Hoogland 4927 (CANB) | Papua New Guinea | — | — | JQ690455 | JQ690454 | JQ690456 | — | — | | | | |
| <i>Miliusa lanceolata</i> Chaowasku & Keßler | Brass 28198 (L) | Papua New Guinea | — | — | JQ690459 | JQ690458 | JQ690460 | — | — | | | | |
| <i>Miliusa lineata</i> (Craib) Ast syn. <i>horsfieldii</i> | Kessler, P.J.A. PK 3202 (P) | Thailand | AY318987 | AY518848 | — | AY319099 | — | — | — | | | | |

| Identification | | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|--|--------------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| Species | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Miliusa longipes</i> King | <i>Pholsena 2651</i> (L) | Thailand | AY318988 | AY518843 | — | AY319100 | — | — | — | | | | |
| <i>Miliusa macrocarpa</i> Hook. F. & Thomson | <i>Grierson & Long 4095</i> (E) | Bhutan | — | JQ690499 | JQ690500 | JQ690498 | JQ690501 | — | — | | | | |
| <i>Miliusa macropoda</i> Miq. | <i>Ambriansyah & Arifin AA 272</i> (L) | Indonesia | — | — | JQ690463 | JQ690462 | JQ690464 | — | — | | | | |
| <i>Miliusa mollis</i> 1756 Pierre | <i>Pholsena 1756</i> (L) | Thailand | AY318989 | AY518850 | — | AY319101 | — | — | — | | | | |
| <i>Miliusa mollis</i> 3207 | <i>Kessler, P.J.A. 3207</i> | Thailand | AY318990 | AY518851 | JQ690503 | AY319102 | JQ690504 | — | — | | x | | x |
| <i>Miliusa montana</i> Gardner ex Hook. F. & Thomson | <i>Hladik 1039</i> (US) | Sri Lanka | — | JQ690507 | JQ690508 | JQ690506 | JQ690509 | — | — | | | | |
| <i>Miliusa novoguineensis</i> Mols & Keßler | <i>Womersley NGF. 24845</i> (NSW) | - | — | — | JQ690467 | JQ690466 | JQ690468 | — | — | | | | |
| <i>Miliusa oropheoides</i> | Unpublished | - | — | — | UP | UP | UP | — | — | | | | |
| <i>Miliusa parviflora</i> Ridl. | <i>Chaowasku, T. 98</i> (L) | Thailand | — | — | JQ690471 | JQ690470 | JQ690472 | — | — | | | | |
| <i>Miliusa sclerocarpa</i> (A. DC.) Kurz | <i>Chaowasku, T. 19</i> (L) | Thailand | — | — | JQ690475 | JQ690474 | JQ690476 | — | — | | | | |
| <i>Miliusa thorelii</i> Finet & Gagnep | <i>Kessler P.J.A. PK 3184</i> (L) | Thailand | AY318992 | AY518846 | JQ690519 | AY319104 | JQ690520 | — | — | | x | | x |
| <i>Miliusa traceyi</i> Jessup | <i>Ford AF 4778</i> | Australia | — | JQ690532 | JQ690533 | JQ690531 | JQ690534 | — | — | | | | |
| <i>Miliusa velutina</i> (Dunal) Hook. F. & Thomson | <i>Pholsena 2842</i> (L) | Thailand | AY318993 | AY518847 | JQ690536 | AY319105 | JQ690537 | — | — | | x | | x |
| <i>Mischogyne michelioides</i> Exell | <i>Bamps, P. 4459</i> (WAG) | Angola | EU169764 | EU169697 | EU169718 | EU169764 | EU169741 | — | EU169809 | | x | | |
| <i>Mitrella kentii</i> (Blume) Miq. | <i>Gardette, E. 2239</i> (K) | Malaysia | AY841633 | — | — | AY841711 | — | — | — | | | | |
| <i>Mitrephora alba</i> Ridl. | <i>Chalermglin 440304-1</i> (TISTR, Bangkok) | Thailand | AY318994 | AY518855 | JQ889983 | AY319106 | JQ889978 | — | — | | x | | x |
| <i>Mitrephora keithii</i> Ridl. | <i>Kessler, P.J.A. PK 3190</i> (L) | Thailand | AY318995 | AY518857 | — | AY319108 | EU522122 | EU522343 | — | | x | | x |
| <i>Mitrephora macrocarpa</i> (Miq.) Weeras. & R.M.K. Saunders | <i>Mols, J.B. 8</i> (L) | Indonesia | — | AY518859 | UP | AY319107 | UP | — | — | | | | |
| <i>Mitrephora polypyrena</i> (Blume) Miq. | <i>Mols, J.B. 7</i> (L) | Indonesia | AY318997 | AY518858 | — | AY319110 | — | — | — | | | | |
| <i>Mitrephora teysmannii</i> Scheff. | <i>Kessler, P.J.A. 3226</i> (L) | Thailand | AY318996 | — | — | AY319109 | — | — | — | | | | |
| <i>Mkilua fragrans</i> Verdc. | <i>Chatrou, L.W. 474</i> (U) | Cultivated in UUBG | AY841634 | DQ125060 | EF179303 | AY841712 | DQ861696 | EF179268 | EF179345 | x | x | | |
| <i>Monanthotaxis whytei</i> (Stapf) Verdc. | <i>UUBG 84GR00388</i> | Cultivated in UUBG | AY841635 | EF179278 | EF179304 | AY841713 | EF179315 | EF179269 | EF179346 | x | x | x | |
| <i>Monanthotaxis</i> sp. | <i>Wieringa, J.J. 3833</i> (WAG) | Gabon | AY841636 | — | — | AY841713 | — | — | — | | | | |
| <i>Monocarpia euneura</i> Miq. | <i>Slik, J.W.F. 2002-2931</i> (L) | Indonesia | AY318998 | AY518865 | AY841412 | AY319111 | AY841477 | AY841381 | AY841559 | x | x | | x |
| <i>Monocarpia marginalis</i> (Scheff.) J. Sinclair | <i>Kaewruang 1</i> (L) | Thailand | JQ690395 | JQ690397 | JQ690398 | JQ690396 | JQ690399 | — | — | | x | | x |
| <i>Monocyclanthus vegnei</i> Keay | <i>Jongkind, C.C.H. 6992</i> (WAG) | Liberia | EU169765 | EU169698 | — | EU169787 | EU169742 | — | EU169810 | | | | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|---|--|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Monodora crispata</i> Engl. | UUBG E64GR0066 | Cultivated in UUBG, origin Ivory Coast | AY841637 | EU169699 | EU169720 | AY841715 | EU169743 | — | EU169811 | | x | x | |
| <i>Monodora myristica</i> (Gaertn.) Dunal | UUBG E84GR00389 | Cultivated in UUBG, origin Ivory Coast | AY743447 | EU169700 | EU169721 | AY743466 | DQ125129 | EF179270 | EU169812 | x | x | x | |
| <i>Monodora tenuifolia</i> Benth. | Chatrou, L.W. 478 (U) | ? | AY841638 | — | — | AY841716 | EU216671 | — | EU216626 | | | | |
| <i>Monoon longipes</i> (Miq.) Koord. & Valet. | Ridsdale, C.E. DV-M2-11443 (L) | Malaysia | AY319028 | AY518829 | — | AY319142 | — | — | — | | | | |
| <i>Mosannona costaricensis</i> (R.E.Fr.) Chatrou | Chatrou, L.W. 90 (U) | Costa Rica | AY743510 | AY743503 | AY841413 | AY743496 | AY841479 | AY841382 | AY841560 | x | x | | x |
| <i>Mosannona pacifica</i> Chatrou | Maas, P.J.M. 8531 (U) | Ecuador | AY743513 | AY743506 | — | AY743499 | AY841482 | — | — | | x | | x |
| <i>Mosannona papillosa</i> Chatrou | Pitman, N. s.n. (U) | Ecuador | AY743514 | AY743507 | — | AY743500 | AY841483 | — | — | | x | | x |
| <i>Mosannona vasquezii</i> Chatrou | Chatrou, L.W. 226 (U) | Peru | AY743515 | AY743508 | — | AY319178 | AY841484 | — | — | | x | | x |
| <i>Mwasumbia alba</i> Couvreur & D.M.Johnson | Couvreur, T.L.P. 85 (WAG) | Tanzania | EU747680 | UP | UP | EU747674 | UP | — | — | | x | | |
| <i>Neostenanthera myristicifolia</i> (Oliv.) Exell | Wieringa, J.J. 3566 (WAG) | Gabon | AY743448 | AY743486 | EF179306 | AY743467 | DQ125130 | EF179271 | EF179348 | x | x | x | |
| <i>Neo-uvaria acuminatissima</i> (Miq.) Airy Shaw | Ridsdale, C.E. DV-SR-4671 (L) | Malaysia | AY318999 | AY518793 | — | AY319112 | — | — | — | | | | |
| <i>Neo-uvaria parallelivenia</i> (Boerl.) H.Okada & K.Ueda | Kessler, P.J.A. sub IV-H-73 (L) | Indonesia | AY319000 | AY518794 | UP | AY319113 | UP | — | — | | x | | x |
| <i>Neo-uvaria telopea</i> | Chaowasku, T. 77 (L) | Thailand | JX544755 | JX544751 | JX544778 | JX544783 | JX544791 | — | — | | x | | x |
| <i>Onychopetalum periquino</i> (Rusby) D.M.Johnson & N.A.Murray | Chatrou, L.W. 425 (U) | Bolivia | AY319065 | AY518876 | AY841414 | AY319179 | AY841485 | AY841383 | AY841561 | x | x | | x |
| <i>Ophrypetalum odoratum</i> Diels | Robertson, A. 7547 (WAG) | Kenya | EU169789 | EU169702 | EU169723 | EU169767 | EU169745 | — | EU169814 | | x | | |
| <i>Orophea brandisii</i> Hook. f. & Thomson | Kessler, P.J.A. 3180 (L) | Thailand | AY319003 | AY518813 | — | AY319116 | — | — | — | | | | |
| <i>Orophea celebica</i> (Blume) Miq. | Kessler, P.J.A. 2953 (L) | Indonesia | AY319004 | AY518814 | — | AY319117 | — | — | — | | | | |
| <i>Orophea cf malayana</i> Kessler | Hoffmann 11 (K) | Malaysia | AY319045 | AY518820 | — | AY319159 | — | — | — | | | | |
| <i>Orophea creaghii</i> (Ridl.) Leonardia & Kessler | Kessler, P.J.A. 1605 (L) | Indonesia | AY841632 | AY518817 | — | AY841710 | — | — | — | | | | |
| <i>Orophea enneandra</i> Blume | Kessler, P.J.A. sub XX-D-179 | Indonesia | AY319007 | AY518816 | — | AY319120 | — | — | — | | | | |
| <i>Orophea enterocarpa</i> Maingay ex Hook.f. & Thomson | Chalermglin 440403 (TISTR Bangkok) | Thailand | AY319006 | AY518815 | JQ690416 | AY319119 | — | — | — | | | | |
| <i>Orophea kerrii</i> Kessler | Chalermglin 440416-1 (TISTR Bangkok) | Thailand | AY319008 | AY518818 | JQ690419 | AY319121 | JQ690420 | — | — | | x | | x |
| <i>Orophea polycarpa</i> A.DC. | Kessler, P.J.A. 3234 (L) | Thailand | AY319010 | AY518819 | — | AY319123 | — | — | — | | | | |
| <i>Oxandra asbeckii</i> (Pulle) R.E.Fr. | University of Guyana, course Neotrop. Botany UG-NB-55 (U) | Guyana | AY841639 | — | — | AY841717 | AY841486 | — | — | | | | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|--|---|------------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Oxandra espiptana</i> (Spruce ex Benth.) Baill. | <i>Chatrou, L.W. et al. 133 (U)</i> | Peru | AY319066 | DQ018260 | UP | AY319180 | AY841487 | — | — | | x | | x |
| <i>Oxandra euneura</i> Diels | <i>Chatrou, L.W. et al 249 (U)</i> | Peru | AY841640 | — | — | AY841718 | AY841488 | — | — | | | | |
| <i>Oxandra longipetala</i> R.E.Fr. | <i>Chatrou, L.W. 114 (U)</i> | Costa Rica | AY841641 | — | — | AY841719 | AY841490 | — | — | | | | |
| <i>Oxandra macrophylla</i> R.E.Fr. | <i>Chatrou, L.W. 204 (U)</i> | Peru | AY841642 | UP | UP | AY841720 | AY841491 | — | — | | x | | x |
| <i>Oxandra polyantha</i> R.E.Fr. | <i>Chatrou, L.W. 215 (U)</i> | Peru | AY841643 | — | — | AY841721 | AY841493 | — | — | | | | |
| <i>Oxandra sphaerocarpa</i> R.E.Fr. | <i>Maas, P.J.M. et al 8226 (U)</i> | Peru | AY841644 | — | — | AY841722 | AY841494 | — | — | | | | |
| <i>Oxandra venezuelana</i> R.E.Fr. | <i>Chatrou, L.W. 120 (U)</i> | Costa Rica | AY841645 | JQ690413 | JQ690414 | AY841723 | AY841495 | — | — | | x | | x |
| <i>Oxandra xylopioides</i> Diels | <i>Chatrou, L.W. 165 (U)</i> | Peru | AY841646 | — | — | AY841724 | AY841496 | — | — | | | | |
| <i>Phaeanthus ebracteolatus</i> (C.Presl.) Merr. | <i>Utteridge, T. 17 (KL)</i> | Papua New Guinea | AY319012 | AY518863 | — | AY319125 | — | — | — | | | | |
| <i>Phaeanthus splendens</i> Miq. | <i>Kessler, P.J.A. B 1564 (L)</i> | Borneo | JX544754 | AY518864 | JX544777 | AY319126 | JX544790 | — | — | | x | | x |
| <i>Phaeanthus</i> sp** | <i>Takeuchi, XX 18407 (L)</i> | Sumatra | — | KC857574 | — | KC857573 | KC857576 | — | — | | | | |
| <i>Piptostigma fasciculatum</i> (De Wild.) Boutique ex R.E.Fr. | <i>Jongkind, C.C.H. et al. 1862 (WAG)</i> | Ghana | AY841647 | UP | UP | AY841725 | AY841497 | — | — | | x | | x |
| <i>Piptostigma mortehani</i> De Wild. | <i>Wieringa, J.J. 2779 (WAG)</i> | Gabon | AY743454 | AY743492 | AY841415 | AY743473 | AY841498 | AY841384 | AY841562 | x | x | | x |
| <i>Piptostigma pilosum</i> Oliv. | <i>Wieringa, J.J. 2030 (WAG)</i> | Cameroon | AY841648 | — | — | AY841726 | AY841499 | — | — | | | | |
| <i>Platymitra macrocarpa</i> Boerl. | <i>Okada 3457 (L)</i> | Indonesia | AY319013 | AY518812 | JQ690422 | AY319127 | JQ690423 | — | — | | x | | x |
| <i>Platymitra</i> sp | <i>Chaowasku, T. 100 (L)</i> | Thailand | — | JQ690426 | UP | JQ690425 | JQ690428 | — | — | | | | |
| <i>Polyalthia borneensis</i> Merr. | <i>Ridsdale, C.E. DV-SR-7921 (L)</i> | Malaysia | AY319014 | AY518821 | — | AY319128 | — | — | — | | | | |
| <i>Polyalthia bullata</i> King | <i>Chaowasku, T. 34 (L)</i> | Thailand | — | JX544825 | JX544839 | JX544800 | JX544809 | — | — | | | | |
| <i>Polyalthia cauliflora</i> Hook.f. & Thomson | <i>Kessler, P.J.A. 3114 (L)</i> | Singapore | AY319015 | AY518823 | JX544837 | AY319129 | — | — | — | | | | |
| <i>Polyalthia celebica</i> Miq. | <i>Mols, J.B. 9 (L)</i> | Indonesia | AY319016 | AY518827 | JX544838 | AY319130 | JX544808 | — | — | | x | | x |
| <i>Polyalthia cerasoides</i> (Roxb.) Benth. & Hook.f. ex Beddome | <i>Chalermglin 440214-4 (L)</i> | Thailand | AY319017 | AY518854 | JQ889985 | AY319131 | JQ889980 | — | — | | x | | x |
| <i>Polyalthia cinnamomea</i> Hook.f. & Thomson | <i>Ridsdale, C.E. DV-M1-347 (L)</i> | Malaysia | AY319018 | AY518828 | — | AY319132 | — | — | — | | | | |
| <i>Polyalthia congesta</i> (Ridl.) J.Sinclair | <i>Ridsdale, C.E. DV-S-5105 (L)</i> | Malaysia | AY319019 | AY518790 | — | AY319133 | — | — | — | | | | |
| <i>Polyalthia debilis</i> (Pierre) Finet & Gagnep. | <i>Kessler, P.J.A. 3228 (L)</i> | Thailand | AY319020 | AY518832 | — | AY319134 | — | — | — | | | | |
| <i>Polyalthia flagellaris</i> (Becc.) Airy Shaw | <i>Duling 38 (K)</i> | Brunei | AY319022 | AY518824 | — | AY319136 | — | — | — | | | | |
| <i>Polyalthia</i> cf. <i>glabra</i> (Hook.f. & Thomson) J.Sinclair | <i>Rastini 224 (L)</i> | Indonesia | AY319032 | AY518782 | — | AY319146 | — | — | — | | | | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|---|------------------------------------|--------------------------------|-------------|-------------|--|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Polyalthia lateriflora</i> (Blume) King | <i>Hort. Bot. Bog. XII-B-VII-37</i> (L) | Indonesia | AY319024 | AY518781 | — | AY319138 | — | — | — | | | | |
| <i>Polyalthia longifolia</i> (Sonn.) Thwaites | <i>Johnson, D.M. 1965</i> (OWU) | Tanzania | AY319027 | AY518786 | — | AY319141 | — | — | — | | | | |
| <i>Polyalthia</i> cf. <i>longifolia</i> (Sonn.) Thwaites | <i>Mols, J.B. 14</i> (L) | Indonesia | AY319025 | AY518785 | — | AY319139 | — | — | — | | | | |
| <i>Polyalthia obliqua</i> Hook.f. & Thomson | <i>Ambriansyah 1694</i> (L) | Indonesia | AY319029 | AY518822 | — | AY319143 | — | — | — | | | | |
| <i>Polyalthia sclerophylla</i> Hook.f. & Thomson | <i>Hort. Bot. Bog. XX-D-82</i> (L) | Indonesia | AY319033 | AY518783 | — | AY319147 | — | — | — | | | | |
| <i>Polyalthia</i> sp. Borneo | <i>Ridsdale DV-M1-12314</i> (L) | Malaysia | UP | UP | — | AY319084 | — | — | — | | | | |
| <i>Polyalthia stenopetala</i> (Hook.f. & Thomson) Finet & Gagnep. | <i>Chalermglin 440302</i> (TISTR Bangkok) | Thailand | AY319034 | — | — | AY319148 | — | — | — | | | | |
| <i>Polyalthia subcordata</i> I (Blume) Blume | <i>Gravendeel, B. et al. 549</i> (L) | Indonesia | AY319036 | AY518830 | — | AY319150 | — | — | — | | | | |
| <i>Polyalthia subcordata</i> II (Blume) Blume | <i>Gravendeel, B. 678</i> (L) | Indonesia | AY319037 | AY518831 | — | AY319151 | — | — | — | | | | |
| <i>Polyalthia suberosa</i> (Roxb.) Thwaites Blume (Blume) | <i>UUBG 83GR00317</i> | Cultivated in UUBG, origin India | AY238956 | AY238965 | AY841417 | AY231289 (intron) AY238949 (spacer) | AY841502 | AY841386 | AY841564 | x | x | | x |
| <i>Polyalthia viridis</i> Craib | <i>Chalermglin 440214-3</i> (L) | Thailand | AY319040 | AY518784 | JX544780 | AY319154 | JX544793 | — | — | | x | | x |
| <i>Polyceratocarpus microtrichus</i> (Engl. & Diels) Ghesq. ex Pellegr. | <i>Bos, J.J. 6684</i> (WAG) | Cameroon | EU747683 | — | — | EU747677 | — | — | — | | | | |
| <i>Polyceratocarpus pellegrini</i> Le Thomas | <i>de Wilde J.J.E. 8718</i> (WAG) | Cameroon | EU747684 | — | — | EU747678 | — | — | — | | | | |
| <i>Polyceratocarpus</i> sp YB2 | <i>Couvreur, T.L.P. 101</i> (WAG) | ? | EU747681 | UP | UP | EU747675 | UP | — | — | | x | | |
| <i>Popowia hirta</i> Miq. | <i>Kessler, P.J.A. B 1628</i> (L) | Indonesia | AY319042 | AY518860 | JX544830 | AY319156 | JX544806 | — | — | | x | | x |
| <i>Popowia odoardi</i> Diels | <i>Ridsdale, C.E. DV-SR-7422</i> (L) | Malaysia | AY319043 | AY518861 | — | AY319157 | — | — | — | | | | |
| <i>Popowia pisocarpa</i> (Blume) Endl. | <i>Van Balgooy, M. 5683</i> (L) | Indonesia | AY319044 | AY518862 | JQ723812 | AY319158 | UP | — | — | | x | | x |
| <i>Porcellia steinbachii</i> (Diels) R.E.Fr. | <i>UUBG 99GR00210</i> | Cultivated in UUBG, origin Bolivia | AY841649 | — | — | AY841727 | — | — | — | | | | |
| <i>Pseudartabotrys letestui</i> Pellegr. | <i>Wieringa, J.J. 3273</i> (WAG) | Gabon | AY841650 | DQ125061 | EF179307 | AY841728 | DQ125131 | EF179272 | EF179349 | x | x | x | |
| <i>Pseudephedranthus fragrans</i> (R.E.Fr.) Aristeg. | <i>Maas, P.J.M. 6878</i> (U) | Venezuela | AY841651 | — | — | AY841729 | — | — | — | | | | |
| <i>Pseudomalmea diclina</i> (R.E.Fr.) Chatrou | <i>Chatrou, L.W. 211</i> (U) | Peru | AY319068 | AY518867 | AY841419 | AY319128 | AY841506 | AY841388 | AY841566 | x | x | | x |
| <i>Pseudomalmea</i> sp. | <i>Idarraga, A. 13</i> (U) | Colombia | AY841652 | — | — | AY841730 | AY841507 | — | — | | | | |
| <i>Pseudoxandra lucida</i> R.E.Fr. | <i>Chatrou, L.W. et al, 212</i> (U) | Peru | AY319076 | AY518870 | AY841420 | AY319190 | AY841510 | AY841389 | AY841567 | x | x | | x |
| <i>Pseudoxandra polyphleba</i> (Diels) R.E.Fr. | <i>Maas, P.J.M. 8227</i> (U) | Peru | AY841654 | JQ769091 | JQ769092 | AY841732 | AY841512 | — | — | | x | | x |

| Identification | | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|-------------------------------|------------------|--------------------------------|-------------|-------------|--------------|------------------|------------------|------------------|----------------------|----|-----|-----|
| Species | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Pseudoxandra spiritus-sancti</i> Maas | Maas, P.J.M. 8833 (U) | Brazil | AY841533 | AY841399 | AY841421 | AY841547 | AY841513 | AY841390 | AY841568 | x | x | | x |
| <i>Pseuduvaria brachyantha</i> Y.C.F. Su & R.M.K. Saunders | Takeuchi & Ama 15677 (L) | ? | AY319046 | AY518837 | — | AY319160 | — | — | — | | | | |
| <i>Pseuduvaria coriacea</i> Y.C.F. Su & R.M.K. Saunders | Takeuchi & Ama 16314 (L) | Papua New Guinea | AY319047 | AY518838 | — | AY319161 | — | — | — | | | | |
| <i>Pseuduvaria fragrans</i> Y.C.F. Su & R.M.K. Saunders | Chaowasku, T. 27 (L) | Thailand | JQ723871 | EU522286 | JX544829 | EU522231 | EU522176 | EU522397 | — | | x | | x |
| <i>Pseuduvaria megalopus</i> (K.Schum.) Y.C.F.Su & Mols | Takeuchi 15599 (L) | Papua New Guinea | AY319011 | EU522263 | — | AY319124 | — | — | — | | | | |
| <i>Pseuduvaria pamattonis</i> (Miq.) Y.C.F.Su & R.M.K.Saunders | Slik, J.W.F. 2002-2911 (L) | Indonesia | AY319049 | AY518840 | — | AY319163 | — | — | — | | | | |
| <i>Pseuduvaria phuyensis</i> (R.M.K.Saunders, Y.C.F.Su & Chalermglin) Y.C.F.Su & R.M.K.Saunders | Kessler, P.J.A. 3221 (L) | Thailand | AY319001 | AY518841 | — | AY319114 | — | — | — | | | | |
| <i>Pseuduvaria rugosa</i> (Blume) Merr. | Kessler, P.J.A. 3209 (L) | Thailand | AY319048 | AY518839 | — | AY319162 | — | — | — | | | | |
| <i>Pseuduvaria setosa</i> (King) J. Sinclair | Maxwell, J.F. 86-208 (L) | Thailand | EU522334 | EU522279 | UP | EU522224 | EU522169 | EU522390 | — | | x | | x |
| <i>Ruizodendron ovale</i> (Ruiz & Pav.) R.E.Fr. | Maas, P.J.M. 8600 (U) | Ecuador | AY841657 | HQ214070 | UP | AY841735 | AY841514 | — | — | | x | | x |
| <i>Sageraea elliptica</i> (A.DC.) Hook. f. & Thomson | Chaowasku, T. 45 (L) | Thailand | — | UP | UP | UP | UP | — | — | | | | |
| <i>Sageraea lanceolata</i> Miq. | Ridsdale, C.E. DV-M2-1692 (L) | Malaysia | AY319050 | AY518799 | JX544774 | AY319164 | JX544787 | — | — | | x | | x |
| <i>Sanrafaelia ruffonammari</i> Verdc. | Kayombo 3027 (MO) | Tanzania | EU169790 | EU169703 | EU169724 | EU169768 | EU169746 | — | EU169815 | | x | | |
| <i>Sapranthus microcarpus</i> (Donn.Sm.) R.E.Fr. | Maas, P.J.M. 8457 (U) | Honduras | AY319052 | AY518806 | — | AY319166 | — | — | — | | | | |
| <i>Sapranthus viridiflorus</i> G.E.Schatz | Chatrou, L.W. 55 (U) | Costa Rica | AY319051 | AY743493 | AY841422 | AY319165 | AY841515 | AY841391 | AY841569 | x | x | | x |
| <i>Siamocananga</i> gen nov | Unpublished | - | — | UP | UP | UP | UP | — | — | | | | |
| <i>Sphaerocoryne gracilis</i> (Oliv. ex Engl. & Diels) Verdc. | Robertson, A. 7554 (WAG) | Kenya | EU169755 | EU169688 | JQ768623 | EU169777 | EU169732 | — | EU169799 | | x | | |
| <i>Stelechocarpus burahol</i> (Blume) Hook.f. & Thomson | Mols, J.B. 13 (L) | Indonesia | AY319053 | AY518803 | JX544775 | AY319167 | JX544788 | — | — | | x | | x |
| <i>Stelechocarpus cauliflorus</i> (Scheff.) J.Sinclair | Hort. Bot. Bog. XV-A-196 (L) | Indonesia | AY319054 | AY518800 | JX544776 | AY319168 | JX544789 | — | — | | x | | x |
| <i>Stelechocarpus expansus</i> | Unpublished | - | — | UP | UP | UP | UP | — | — | | | | |
| <i>Stenanona costaricensis</i> R.E.Fr. | Chatrou, L.W. 67 (U) | Costa Rica | AY319069 | AY518801 | JX544772 | AY319183 | AY841516 | — | — | | x | | x |
| <i>Stenanona panamensis</i> Standl. | Chatrou, L.W. 100 (U) | Costa Rica | AY319070 | AY518802 | — | AY319184 | — | — | — | | | | |
| <i>Tetrameranthus duckei</i> R.E.Fr. | Stevenson, D.W. 1002 (U) | Brazil | AY841658 | — | — | AY841736 | AY841439 | — | — | | | | |
| <i>Tetrameranthus laomae</i> D.R.Simpson | Pipoly, J. 13407 (U) | Peru | AY841659 | — | — | AY841737 | — | — | — | | | | |
| <i>Toussaintia orientalis</i> Verdc. | Johnson, D.M. 1957 (OWU) | Tanzania | EU169778 | EU169689 | EU169710 | EU169756 | EU169733 | — | EU169800 | | x | | |

| Identification | | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|------------------------------|---|--------------------------------|-------------|-------------|--|------------------|------------------|------------------|----------------------|----|-----|-----|
| Species | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| <i>Tridimeris</i> sp. | Schatz, G.E. 1198 (K) | Mexico | AY319055 | JX544750 | JX544773 | AY319169 | JX544786 | — | — | | x | | x |
| <i>Trigynaea duckei</i> (R.E.Fr.) R.E.Fr. | Chatrou, L.W. 129 (U) | Peru | AY841660 | — | — | AY841738 | — | — | — | | | | |
| <i>Trigynaea lanceipetala</i> D.M.Johnson & N.A.Murray | Chatrou, L.W. 234 (U) | Peru | AY743449 | AY743487 | EF179309 | AY743468 | UP | EF179274 | EF179351 | x | | x | |
| <i>Trivalvaria macrophylla</i> borneo (1) | Unpublished | - | — | — | UP | UP | UP | — | — | | | | |
| <i>Trivalvaria macrophylla</i> malayp (2) | Unpublished | - | — | — | UP | - | UP | — | — | | | | |
| <i>Trivalvaria macrophylla</i> (Blume) Miq. | Chase, M.W. 1207 (K) | Indonesia | AY319056 | — | — | AY319170 | — | — | — | | | | |
| <i>Trivalvaria mollis</i> | Unpublished | - | — | UP | UP | UP | UP | — | — | | | | |
| <i>Trivalvaria pumilla</i> | Unpublished | - | — | — | UP | UP | UP | — | — | | | | |
| <i>Trivalvaria</i> sp. <i>phangnga</i> | Unpublished | - | — | — | UP | UP | UP | — | — | | | | |
| <i>Trivalvaria</i> sp. <i>umpang</i> | Unpublished | - | — | — | UP | UP | UP | — | — | | | | |
| <i>Trivalvaria</i> sp. 1 | Chaowasku, T. 35 (L) | Thailand | JX544822 | JX544824 | JX544828 | JX544794 | — | — | — | | | | |
| <i>Trivalvaria</i> sp. 2 | Chaowasku, T. 56 (L) | Thailand | — | KC857602* | KC857603* | KC857601* | KC857604* | — | — | | | | |
| <i>Unonopsis pittieri</i> Saff. | Chatrou, L.W. 68 (U) | Costa Rica | AY841661 | DQ018264 | — | AY841739 | AY841517 | — | — | | x | | x |
| <i>Unonopsis rufescens</i> (Baill.) R.E.Fr. | Orava, C. 9 (U) | French Guiana | AY743455 | AY743494 | — | AY743474 | AY841518 | — | — | | x | | x |
| <i>Unonopsis stipitata</i> Diels | Chatrou, L.W. 253 (U) | Peru | AY841662 | AY841400 | AY841423 | AY841740 | AY841519 | AY841392 | AY841570 | x | x | | x |
| <i>Uvaria chamae</i> P.Beauv. | Chatrou, L.W. 482 (U) | Cultivated in UUBG, origin Togo | AY841663 | — | — | AY841741 | — | — | — | | | | |
| <i>Uvaria cherrevensis</i> (Pierre ex Finet & Gagnep.) L.L. Zhou, Y.C.F. Su & R.M.K. Saunders | Maxwell 90-625 (L) | Thailand | FJ743823 | FJ743750 | — | FJ743858 | FJ743787 | — | — | | x | x | |
| <i>Uvaria clementis</i> (Merr.) Attanayake, I.M.Turner & R.M.K.Saunders | Kessler, P.J.A. 3211 (L) | Thailand | AY841606 | — | — | FJ743853 | — | — | — | | | | |
| <i>Uvaria cuneifolia</i> (Hook.f. & Thomson) L.L. Zhou, Y.C.F.Su & R.M.K. Saunders | Mohtar S48169 (L) | Indonesia | FJ743822 | FJ743749 | — | FJ743857 | FJ743786 | — | — | | x | x | |
| <i>Uvaria dulcis</i> Dunal | Maxwell, J.F. 88-509 (L) | Thailand | FJ743815 | FJ743740 | — | FJ743849 | FJ743777 | — | — | | x | x | |
| <i>Uvaria grandiflora</i> Roxb. ex Hornem. | Saunders 05/1 (HKU) | Thailand | FJ743836 | FJ743764 | — | FJ743870 | FJ743802 | — | — | | x | x | |
| <i>Uvaria griffithii</i> L.L.Zhou, Y.C.F.Su & R.M.K.Saunders | Chalermglin 440402-2 (TISTR) | Thailand | FJ743820 | FJ743746 | — | FJ743855 | FJ743783 | — | — | | x | x | |
| <i>Uvaria lucida</i> Benth. subsp. <i>virens</i> (N.E.Br.) Verdc. | UUBG 84GR00334 | Cultivated in UUBG, origin West African | AY238957 | AY238966 | EF179310 | AY231290 (intron) AY238950 (spacer) | AY841440 | EF179275 | EF179352 | x | x | x | |
| <i>Uvaria siamensis</i> (Scheff.) L.L.Zhou, Y.C.F.Su & R.M.K.Saunders | Saunders 07/3 (HKU) | Cultivated in Hong Kong Botanic | FJ743824 | FJ743752 | — | FJ743859 | FJ743790 | — | — | | x | x | |

| Species | Identification | | GenBank identification numbers | | | | | | | Included in datasets | | | |
|---|-------------------------------------|---------------------------------|--------------------------------|-------------|-------------|--|------------------|------------------|------------------|----------------------|----|-----|-----|
| | Voucher | Country | <i>rbcL</i> | <i>matK</i> | <i>ndhF</i> | <i>trnLF</i> | <i>psbA-trnH</i> | <i>atpB-rbcL</i> | <i>trnS-trnG</i> | 7M | 4M | LBC | SBC |
| | | Gardens | | | | | | | | | | | |
| <i>Uvariastrum pierreanum</i> (Engl. & Diels) Sprague & Hutch. | <i>Jongkind, C.C.H. 4707</i> (WAG) | Ivory Coast | EU169791 | EU169705 | EU169725 | EU169769 | — | — | — | | | | |
| <i>Uvariastrum pynaertii</i> De Wild. | <i>Wieringa, J.J. 2620</i> (WAG) | Gabon | EU169770 | EU169705 | — | EU169792 | EU169748 | — | EU169816 | | x | | |
| <i>Uvariodendron kirkii</i> Verdc. | <i>Robertson, A. 7550</i> (WAG) | Kenya | EU169771 | EU169706 | EU169726 | EU169793 | EU169749 | — | EU169817 | | x | | |
| <i>Uvariodendron molundense</i> (Diels) R.E.Fr. | <i>Sosef, M.S.M. 2219</i> (WAG) | Gabon | EU169772 | EU169707 | EU169727 | EU169794 | EU169750 | — | EU169818 | | x | | |
| <i>Uvariopsis korupensis</i> Gereau & Kenfack | <i>Richardson, J.E. 212</i> (WAG) | Gabon | EU169774 | EU169709 | EU169729 | EU169796 | EU169752 | — | EU169820 | | x | | |
| <i>Uvariopsis vanderystii</i> Robyns & Ghesq. | <i>Sosef, M.S.M. 2241</i> (WAG) | Gabon | EU169773 | EU169708 | EU169728 | EU169795 | — EU169751 | — | EU169819 | | x | | |
| <i>Uvariopsis tripetala</i> (Baker.f.) G.E.Schatz | <i>Jongkind, C.C.H. 4356</i> (WAG) | Ivory Coast | EU169758 | — | EU169712 | EU169758 | EU169735 | — | EU169802 | | | | |
| <i>Woodiellantha</i> sp. | <i>Lugas 311</i> (K) | Malaysia | AY841665 | — | — | AY841743 | — | — | — | | | | |
| <i>Xylopia ferruginea</i> (Hook.f. & Thomson) Hook.f. & Thomson | <i>Silk, J.W.F. 2002-S 558</i> (L) | Indonesia | AY841666 | DQ125063 | EF179311 | AY841744 | DQ125133 | — | — | x | x | x | |
| <i>Xylopia frutescens</i> Aubl. | <i>Chatrou, L.W. et al. 106</i> (U) | Costa Rica | AY841667 | — | — | AY841745 | AY841441 | — | — | | | | |
| <i>Xylopia hypolampra</i> Mildbr. & Diels | <i>Wieringa, J.J. 3748</i> (WAG) | Gabon | AY841668 | — | — | AY841746 | — | — | — | | | | |
| <i>Xylopia peruviana</i> R.E.Fr. | <i>Chatrou, L.W. 483</i> (U) | Cultivated in UUBG, origin Peru | AY238958 | AY238967 | EF179312 | AY231291 (intron) AY238951 (spacer) | DQ125134 | EF179276 | EF179353 | | x | x | |

Appendix 2: Starting partition schemes and resulting best schemes for each dataset

Table 1: Different starting schemes for different datasets (7M and 4M), starting partition schemes of 4M are identical to the starting schemes for 4M: LBC and 4M: SBC

| | 7M | 4M |
|--|--|--|
| <i>Only genes</i> | matK; ndhF; rbcL; trnL-trnF; psbA-trnH; atpB-rbcL; trnS-trnG | matK; ndhF; trnL-trnF; psbA-trnH |
| <i>All substitution sites</i> | matK_1; matK_2; matK_3; ndhF_1; ndhF_2; ndhF_3; rbcL_1; rbcL_2; rbcL_3; trnL-trnF; psbA-trnH; atpB-rbcL; trnS-trnG | matK_1; matK_2; matK_3; ndhF_1; ndhF_2; ndhF_3; trnL-trnF; psbA-trnH |
| <i>Two substitution sites</i> | matK_12; matK_3; ndhF_12; ndhF_3; rbcL_12; rbcL_3; trnL-trnF; psbA-trnH; atpB-rbcL; trnS-trnG | matK_12; matK_3; ndhF_12; ndhF_3; trnL-trnF; psbA-trnH |
| <i>All substitution sites together</i> | matK_ndhF_rbcL_1; matK_ndhF_rbcL_2; matK_ndhF_rbcL_3; trnL-trnF; psbA-trnH; atpB-rbcL; trnS-trnG | matK_ndhF_1; matK_ndhF_2; matK_ndhF_3; trnL-trnF; psbA-trnH; |
| <i>Two substitution sites together</i> | matK_ndhF_rbcL_12; matK_ndhF_rbcL_3; trnL-trnF; psbA-trnH; atpB-rbcL; trnS-trnG | matK_ndhF_12; matK_ndhF_3; trnL-trnF; psbA-trnH |

Table 2A: Resulting best partition scheme and models for the 7M dataset for all models. InL = -41628.04. BIC = 84676.97

| Partition | Model |
|---------------------------------|---------|
| matK_3, trnSG | K81uf+G |
| atpBrbcL, matK_1, matK_2, trnLF | K81uf+G |
| ndhF_1, ndhF_2, rbcL_3 | GTR+I+G |
| ndhF_3 | TVM+I+G |
| psbA_trnH | K81uf+G |
| rbcL_1 | TVM+I+G |
| rbcL_2 | JC+I+G |

Table 2B: Resulting best partition scheme and models for the 7M dataset for BEAST models. InL = -41634.45. BIC = 84689.78

| Partition | Model |
|---|---------|
| matK_3, rbcL_3 | HKY+G |
| atpBrbcL, matK_1, matK_2, trnLF, trnStrnG | GTR+G |
| ndhF_1, ndhF_2 | GTR+I+G |
| ndhF_3 | GTR+I+G |
| psbAtrnH | HKY+G |
| rbcL_1 | HKY+I+G |
| rbcL_2 | K80+I+G |

Table 3A: Resulting best partition scheme and models for the 4M dataset for all models. InL = -32360.47. BIC = 69217.44

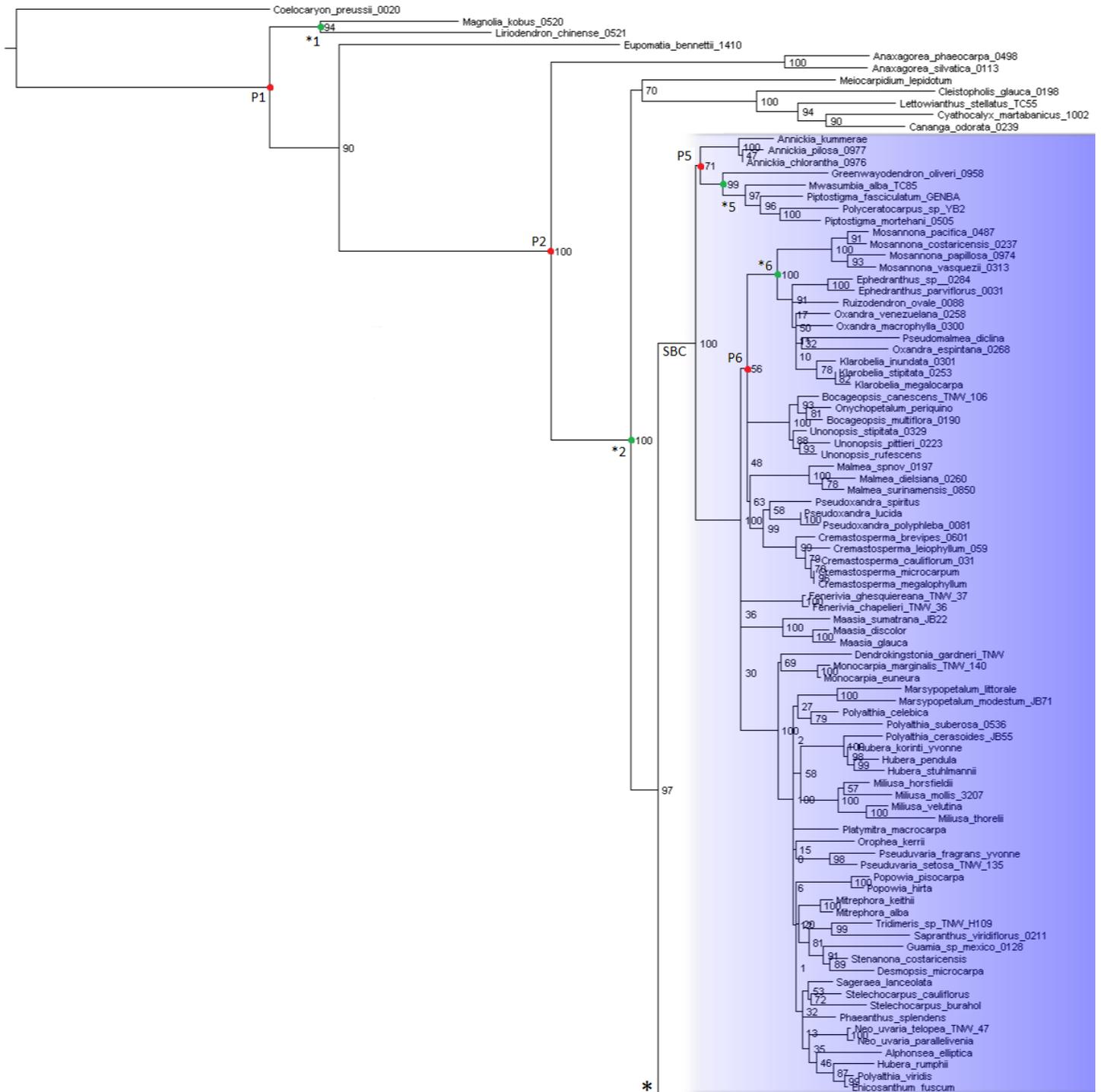
| Partitions | Model |
|-----------------------|---------|
| matK_3, rbcL_3 | TVM+G |
| matK_1, matK_2, trnLF | K81uf+G |
| psbAtrnH | HKY+G |
| rbcL_1 | GTR+I+G |
| rbcL_2 | K81+I+G |

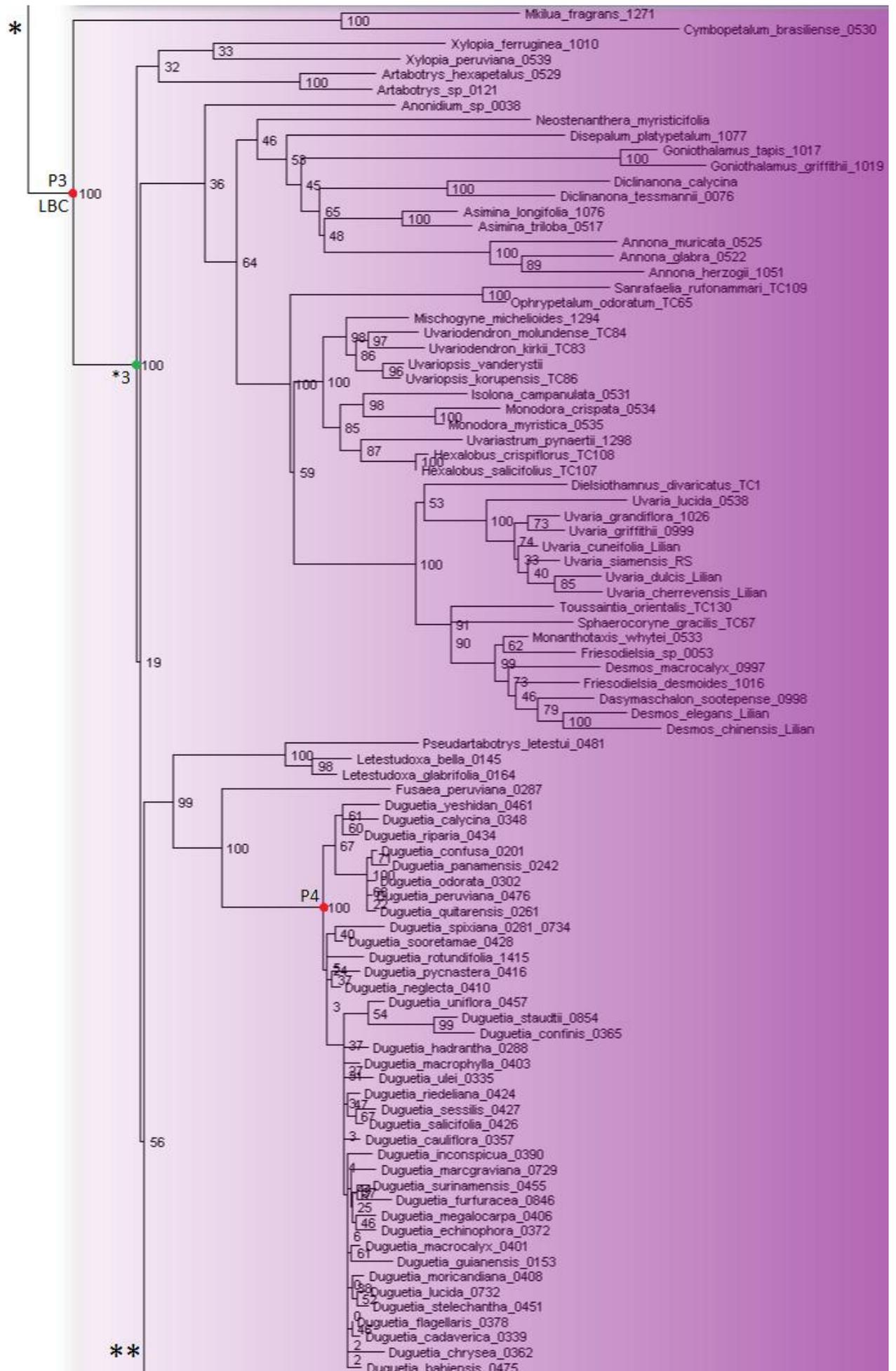
Table 3B: Resulting best partition scheme and models for the 4M dataset for BEAST models. InL = -41714.77. BIC = 84744.52

| Partition | Model |
|---|---------|
| atpBrbcL, matK_1, matK_2, matK_3, trnLF, trnStrnG | GTR+G |
| ndhF_1, ndhF_2, rbcL_3 | GTR+I+G |
| ndhF_3, psbAtrnH | GTR+I+G |
| rbcL_1, rbcL_2 | GTR+I+G |

Appendix 3: RAxML bootstrap trees for the 4M and total dataset

Figure 1: 4M dataset RAxML bootstrap tree with bootstrap values on the nodes, the crown node of the clade including the fossil (green dots and * with number indicating the fossil) and prior placement (red dots and P with number indicating the corresponding fossil). Fossils used are: *Archaeanthus/Endressinia* (1), *Futabanthus* (2), *Anonaspermum* (3), *Duguetia* (4) with no indication of the crown node of the clade including the fossil, since this node cannot be identified with high certainty, African Malmeoideae (5) and '*Mosoxandra*' (6). The SBC is indicated with blue, the LBC with red.





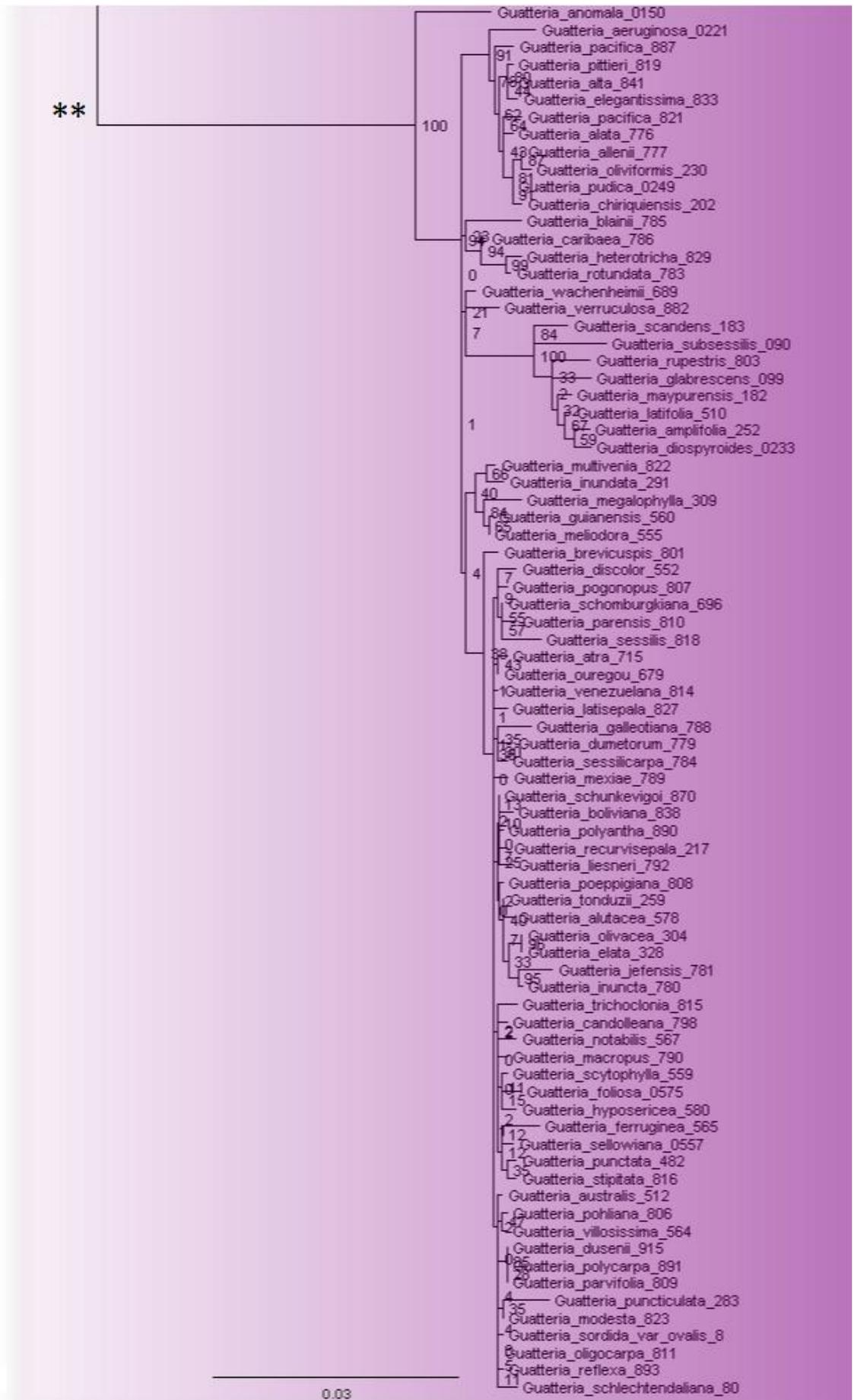
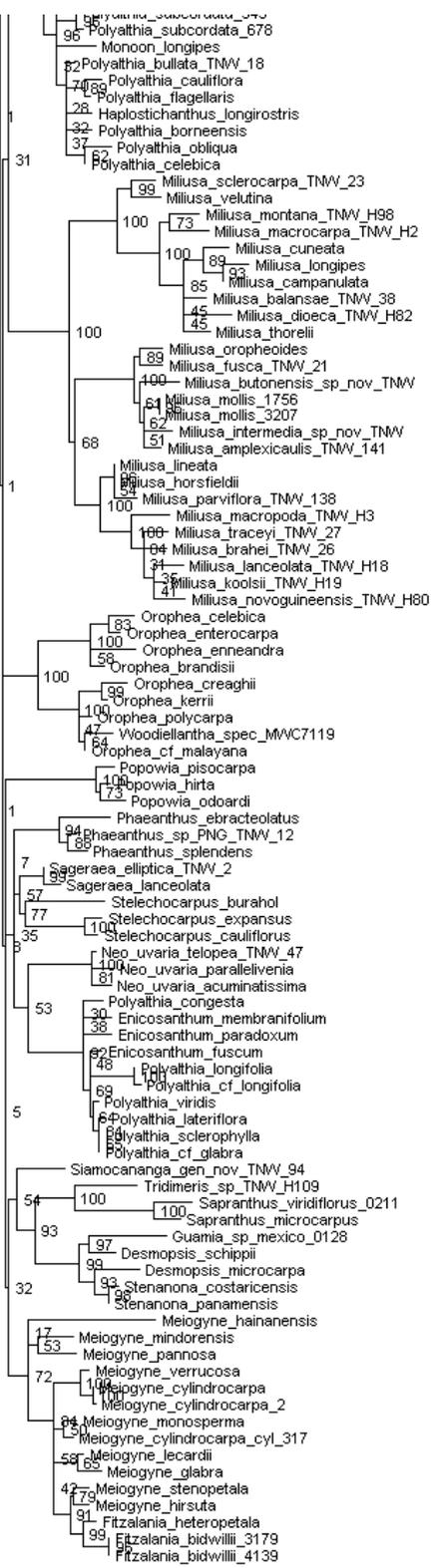


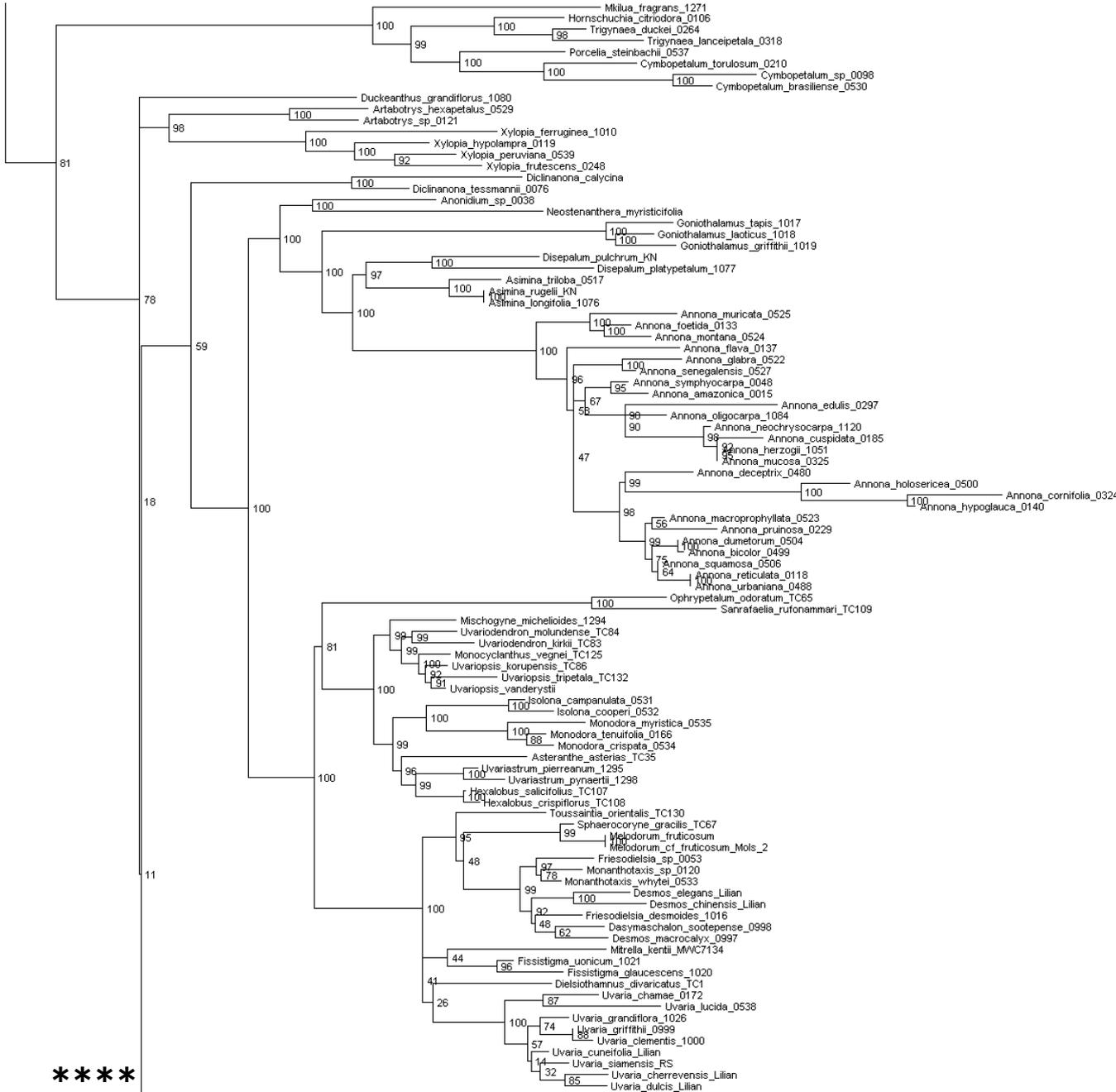
Figure 2: RAxML bootstrap tree from the total dataset with bootstrap values on the nodes



*
97

**





Appendix 4: Record of BEAST analysis runs in Tracer v. 1.5

Table 9: Record of BEAST analysis runs with information about the combinations of runs made in Tracer, numbers of analysis congruent to the numbers in table 4 in chapter 5, runs are from starting trees (C, J, M) with two runs per starting tree for the analyses containing all data, and one per starting tree for the analyses containing the different ratios for the LBC and SBC clades (17 and 18), the burn-in was taken as 1,000,000, except in those cases where a larger burn-in was necessary, then burn-in was chosen by eye, the numbers of parameters with an $ESS < 100$ and with an $100 > ESS < 200$ were recorded and identified, when further treatment of the data was different from the material and method, this was recorded in the notes.

| Analysis | Run | Converged* | Burn-in | # ESS<100 | # 100>ESS<200 | Identification unconverged parameters & notes |
|----------|-----|------------|---------|-----------|---------------|---|
| 1 | CA | yes | 1000000 | 0 | 1 | 100>ESS<200: psbatrnH.covariance |
| | CB | yes | 1000000 | | | |
| | JA | yes | 1000000 | | | |
| | JB | no | - | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 2 | CA | yes | 1000000 | 0 | 2 | 100>ESS<200: rbcl.CP1.treelikelihood |
| | CB | yes | 1000000 | | | 100>ESS<200: trnLF.treelikelihood |
| | JA | no | - | | | |
| | JB | no | - | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 3 | CA | yes | 1000000 | 1 | 1 | ESS<100: psbatrnH.covariance |
| | CB | no | - | | | 100>ESS<200: rbcl.CP1.treelikelihood |
| | JA | yes | 1000000 | | | |
| | JB | no | - | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 4 | CA | yes | 1000000 | 0 | 1 | 100>ESS<200: trnLF.treelikelihood |
| | CB | yes | 1000000 | | | Note 1 |
| | JA | yes | 1000000 | | | |
| | JB | yes | 1000000 | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 5 | CA | no | - | 1 | 3 | ESS<100: trnLF.treelikelihood |
| | CB | yes | 1000000 | | | 100>ESS<200: rbcl.uclid.stdev |
| | JA | yes | 1000000 | | | 100>ESS<200: psbatrnH.covariance |
| | JB | no | - | | | 100>ESS<200: rbcl.CP1.treelikelihood |
| | MA | yes | 1000000 | | | |
| | MB | no | - | | | |
| 6 | CA | yes | 1000000 | 1 | 1 | ESS<100: rbcl.CP1.treelikelihood |
| | CB | yes | 1000000 | | | 100>ESS<200: trnLF.treelikelihood |
| | JA | no | - | | | |

| Analysis | Run | Converged* | Burn-in | # ESS<100 | # 100>ESS<200 | Identification unconverged parameters & notes |
|----------|-----|------------|----------|-----------|---------------|---|
| | JB | no | - | | | |
| | MA | yes | 1000000 | | | |
| 6 | MB | yes | 1000000 | | | |
| 7 | CA | yes | 1000000 | 0 | 2 | 100>ESS<200: trnH.covariance |
| | CB | yes | 1000000 | | | 100>ESS<200: rbcl.CP1.treelikelihood |
| | JA | yes | 1000000 | | | |
| | JB | yes | 1000000 | | | |
| | MA | no | - | | | |
| | MB | no | - | | | |
| 8 | CA | yes | 1000000 | 0 | 1 | 100>ESS<200: trnLF.treelikelihood |
| | CB | yes | 1000000 | | | |
| | JA | no | - | | | |
| | JB | no | - | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 9 | CA | yes | 1000000 | 0 | 1 | 100>ESS<200: psbatrnH.covariance |
| | CB | yes | 1000000 | | | |
| | JA | no | - | | | |
| | JB | no | - | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 10 | CA | yes | 1000000 | 0 | 2 | 100>ESS<200: psbatrnH.covariance |
| | CB | yes | 1000000 | | | 100>ESS<200: trnlf.treelikelihood |
| | JA | no | - | | | |
| | JB | no | - | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 11 | CA | yes | 1000000 | 0 | 4 | 100>ESS<200: psbatrnH.covariance |
| | CB | yes | 1000000 | | | 100>ESS<200:matK.CP1+2.treelikelihood |
| | JA | no | - | | | 100>ESS<200: trnH.treelikelihood |
| | JB | no | - | | | 100>ESS<200:rbcl.CP1.treelikelihood |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 12 | CA | yes | 20000000 | 1 | 2 | ESS<100: psbatrnH.covariance |
| | CB | no | - | | | 100>ESS<200: prior |
| | JA | no | - | | | 100>ESS<200: speciation |
| | JB | no | - | | | Note 2 |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |

| Analysis | Run | Converged* | Burn-in | # ESS<100 | # 100>ESS<200 | Identification unconverged parameters & notes |
|----------|-----|------------|----------|-----------|---------------|--|
| 13 | CA | yes | | 1 | 1 | ESS<100: psbatrnH.covariance |
| | CB | yes | 1000000 | | | 100>ESS<200: rbcl.CP1.treelikelihood |
| | JA | no | - | | | |
| | JB | yes | 1000000 | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 14 | CA | yes | 1000000 | 1 | 0 | ESS<100: psbatrnH.covariance |
| | CB | yes | 1000000 | | | |
| | JA | no | - | | | |
| | JB | x | - | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 15 | CA | yes | 1000000 | 0 | 2 | 100>ESS<200: psbatrnH.covariance |
| | CB | yes | 5000000 | | | 100>ESS<200: rbcl.CP1.treelikelihood |
| | JA | yes | 10000000 | | | Note 3 |
| | JB | yes | 1000000 | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 16 | CA | yes | 1000000 | 0 | 2 | 100>ESS<200: psbatrnH.covariance |
| | CB | no | - | | | 100>ESS<200: rbcl.CP1.treelikelihood |
| | JA | no | - | | | |
| | JB | yes | 1000000 | | | |
| | MA | yes | 1000000 | | | |
| | MB | yes | 1000000 | | | |
| 17** | CA | yes | 1000000 | 0 | 0 | - |
| | JA | yes | 1000000 | | | Note 4 |
| | MA | yes | 1000000 | | | |
| 18*** | CA | yes | 1000000 | 2 | 2 | ESS<100: trnIF.treelikelihood |
| | JA | yes | 1000000 | | | ESS<100: rbcl.CP1.treelikelihood |
| | MA | yes | 1000000 | | | 100>ESS<200: psbatrnH.covariance 100>ESS<200: matK.CP1+2.treelikelihood Note 5 |

* Converged to the same parameter level; no big jumps in parameter traces were observed anymore. Where possible was chosen for the runs with the highest likelihoods.

** LBC reduced

*** SBC reduced

Note 1: results in too large treesfile for treeannotator: resample every 21000 (remain 1/7 samples)

Note 2: tree files: burn-in: CA: 600001, MA:30001, MB:30001

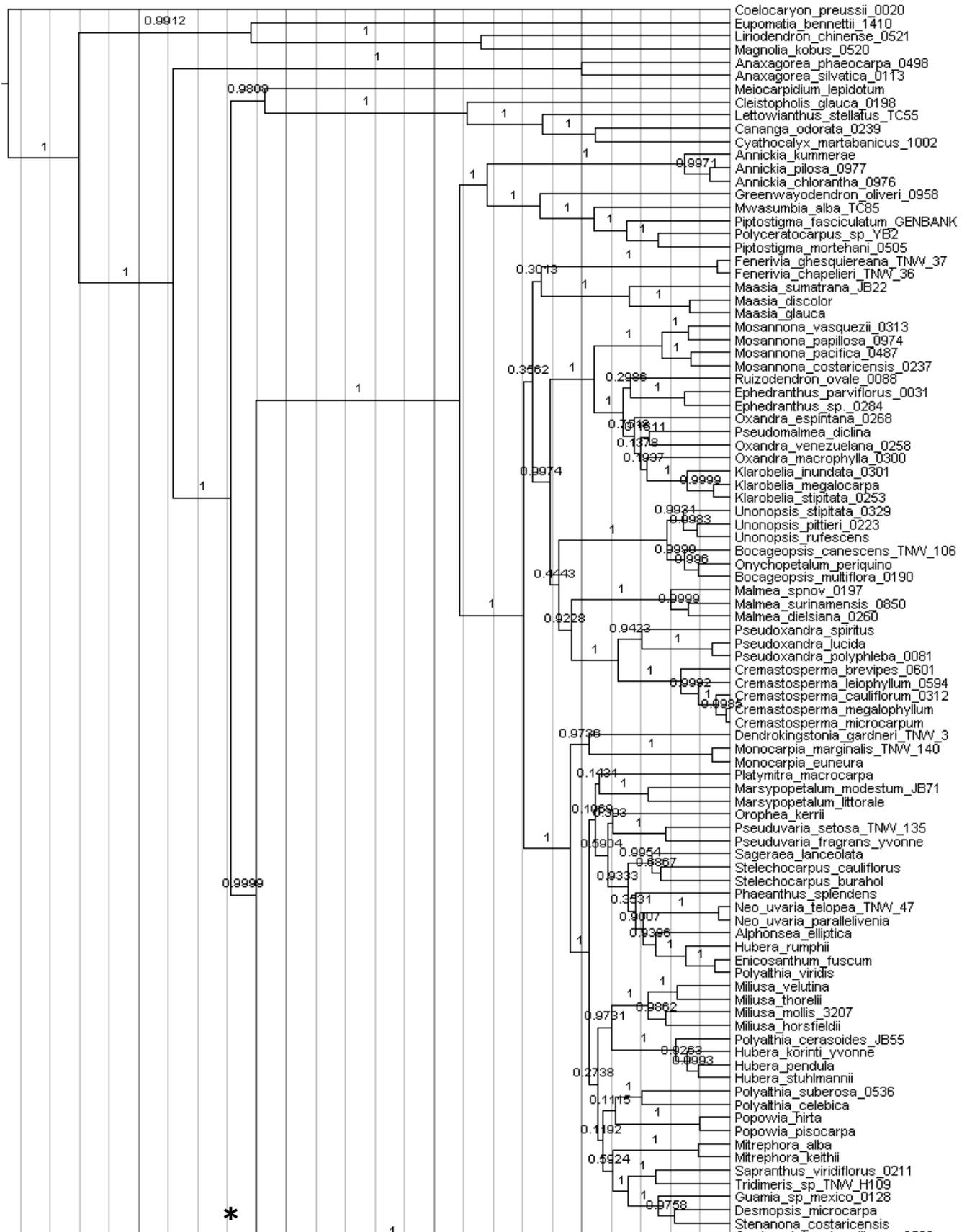
Note 3: tree files combined in log combiner: 30001,150001,300001,30001,30001 and 30001 burn-in respectively: because of higher burn-in for CB and JA.

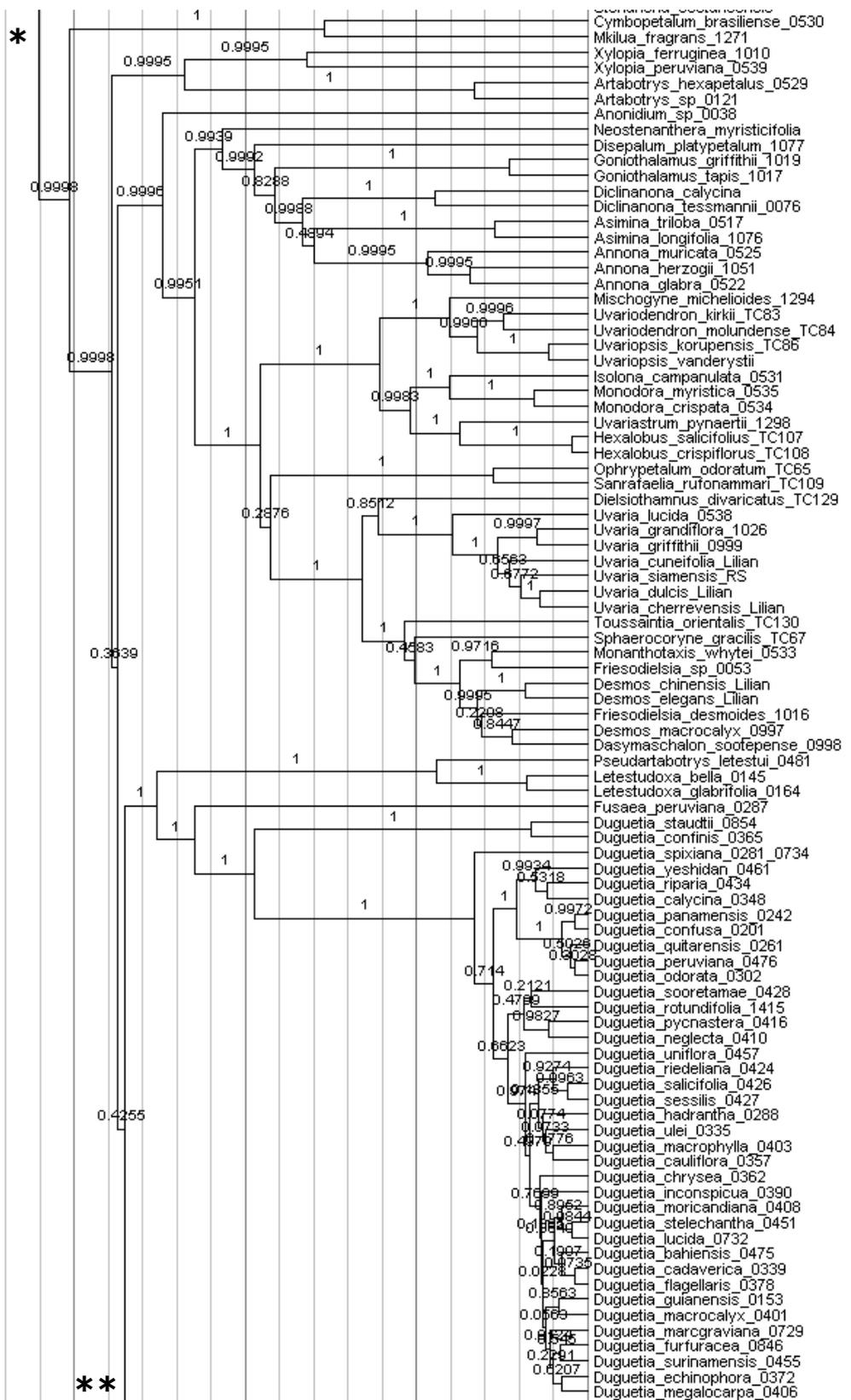
Note 4: resample state at low freq. 9000 instead of 15000 to obtain enough trees.

Note 5: resample state at low freq. 9000 ipv 15000 to obtain enough trees.

Appendix 5: Maximum clade credibility trees resulting from the BEAST analyses

Figure 1: Maximum clade credibility (MCC) tree for analysis 1 (all fossils included) with posterior values on the branches





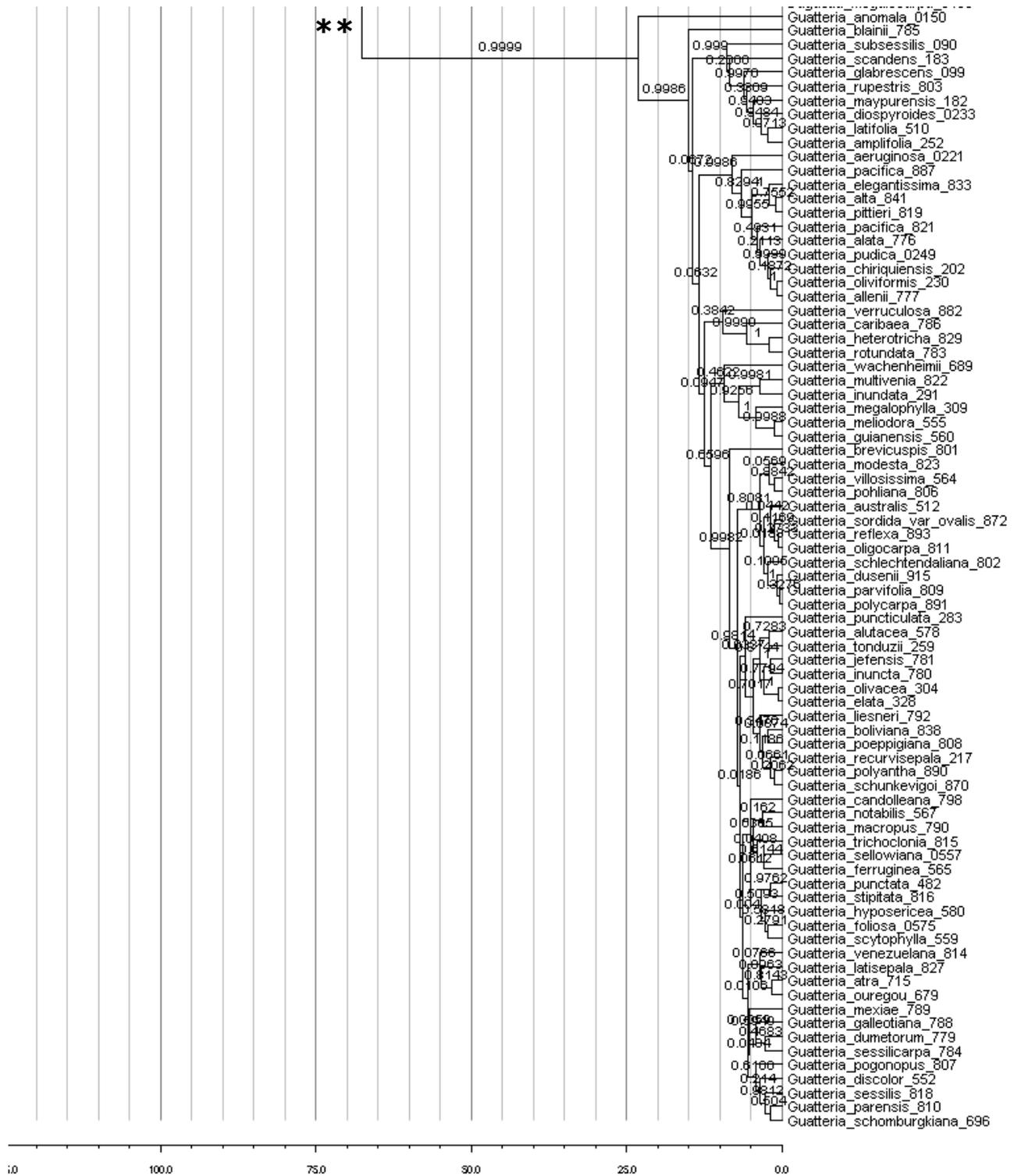
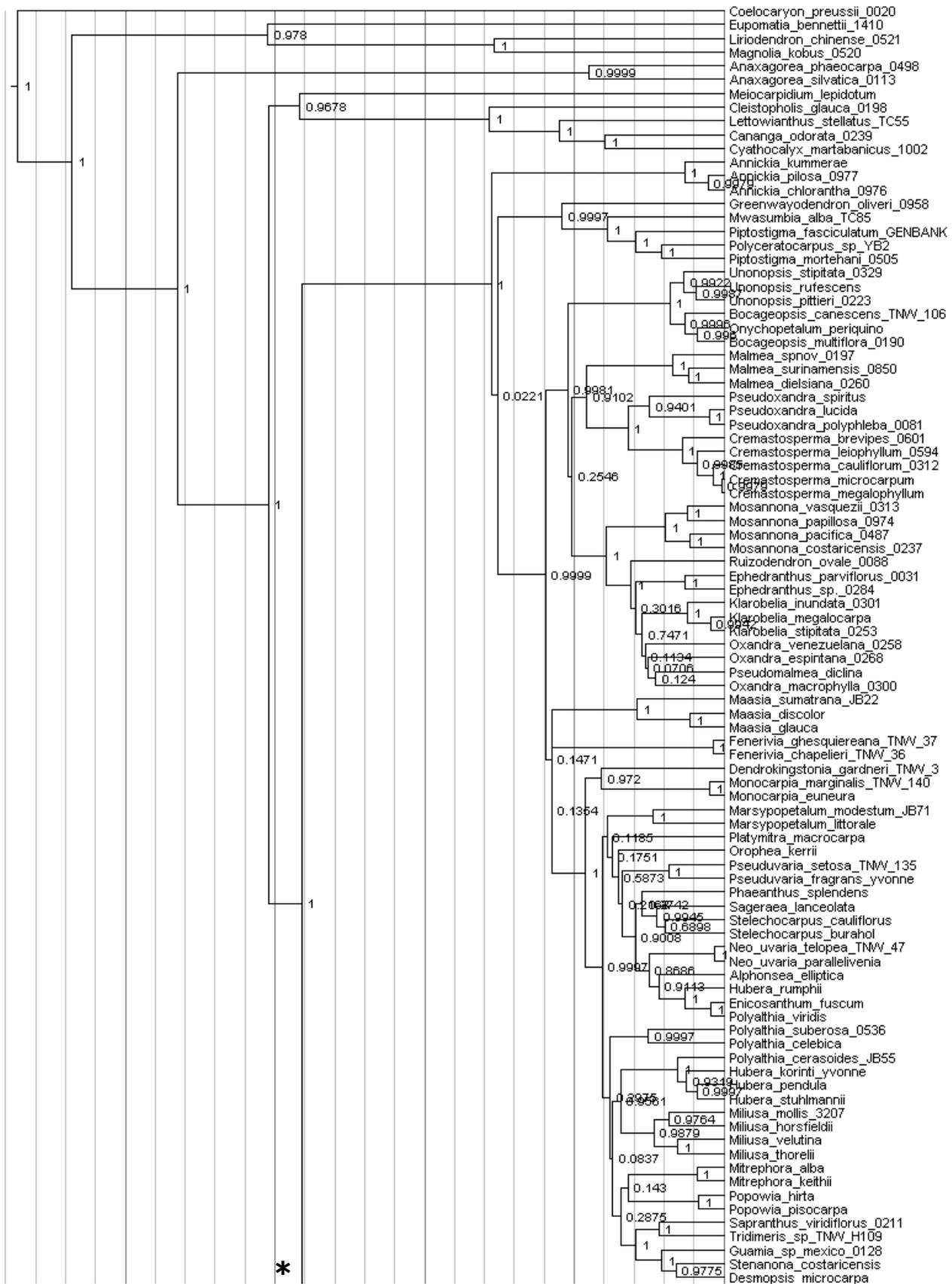
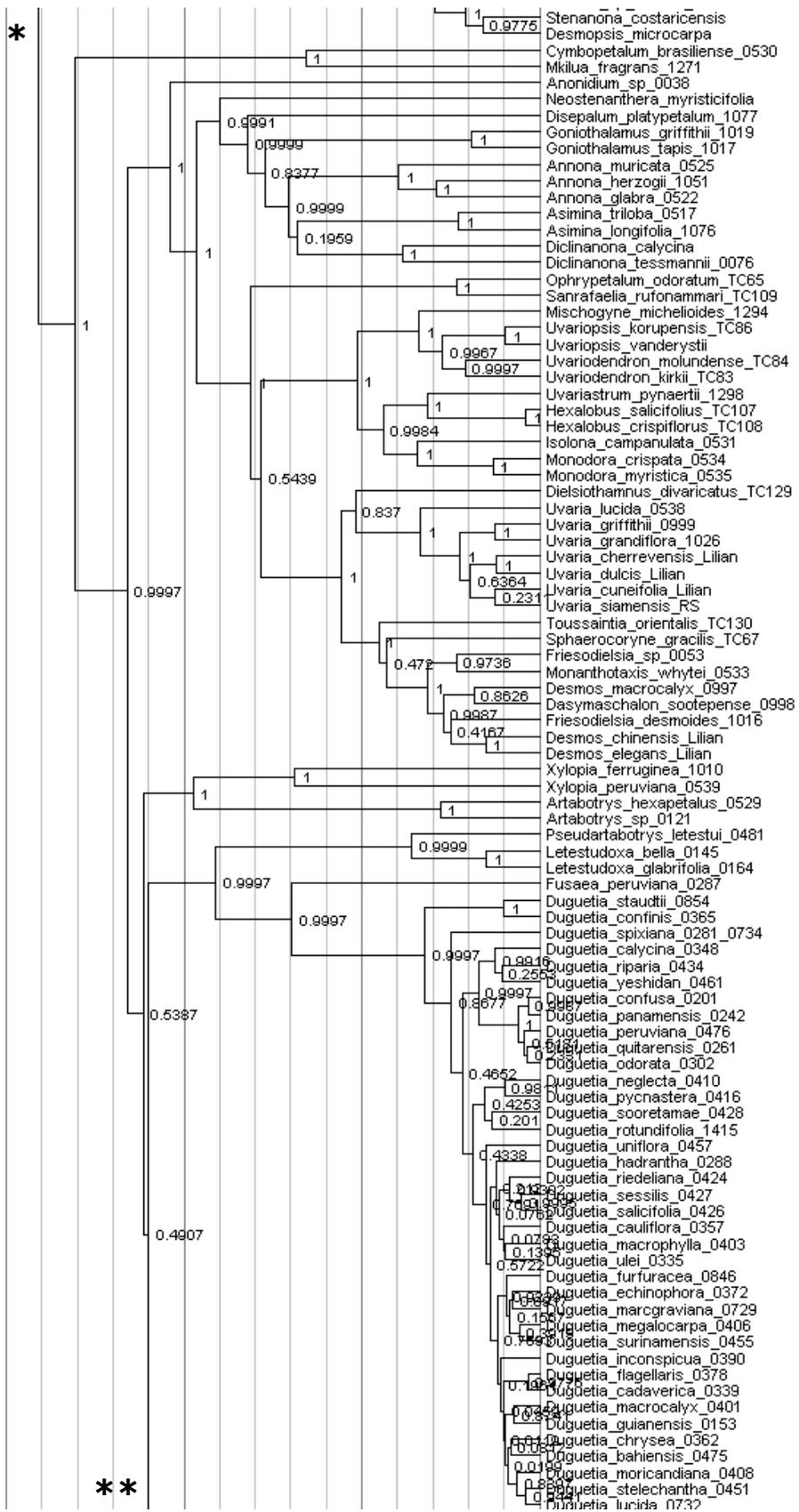


Figure 2: MCC tree for analysis 2 (only deep priors included) with posterior values on the branches





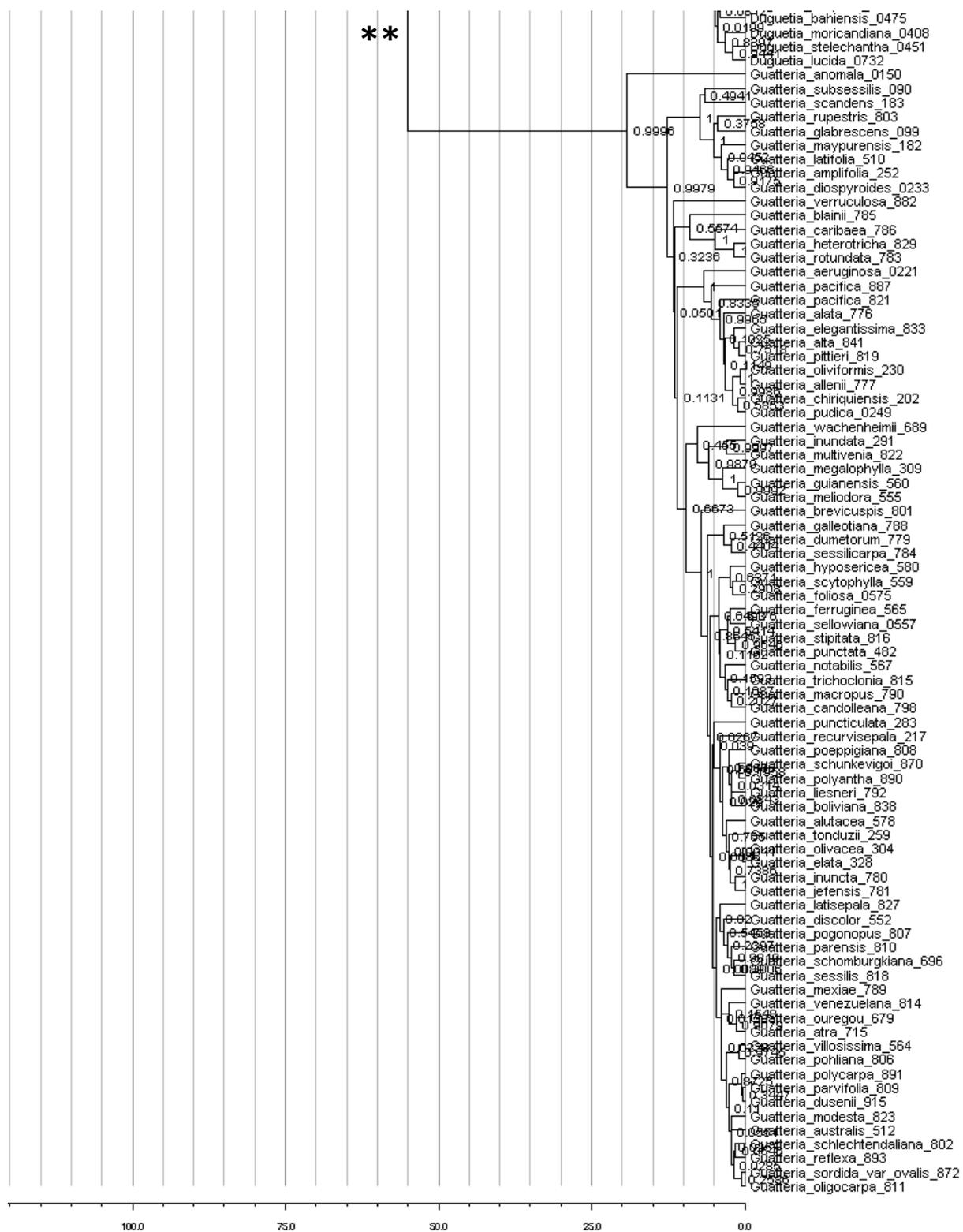
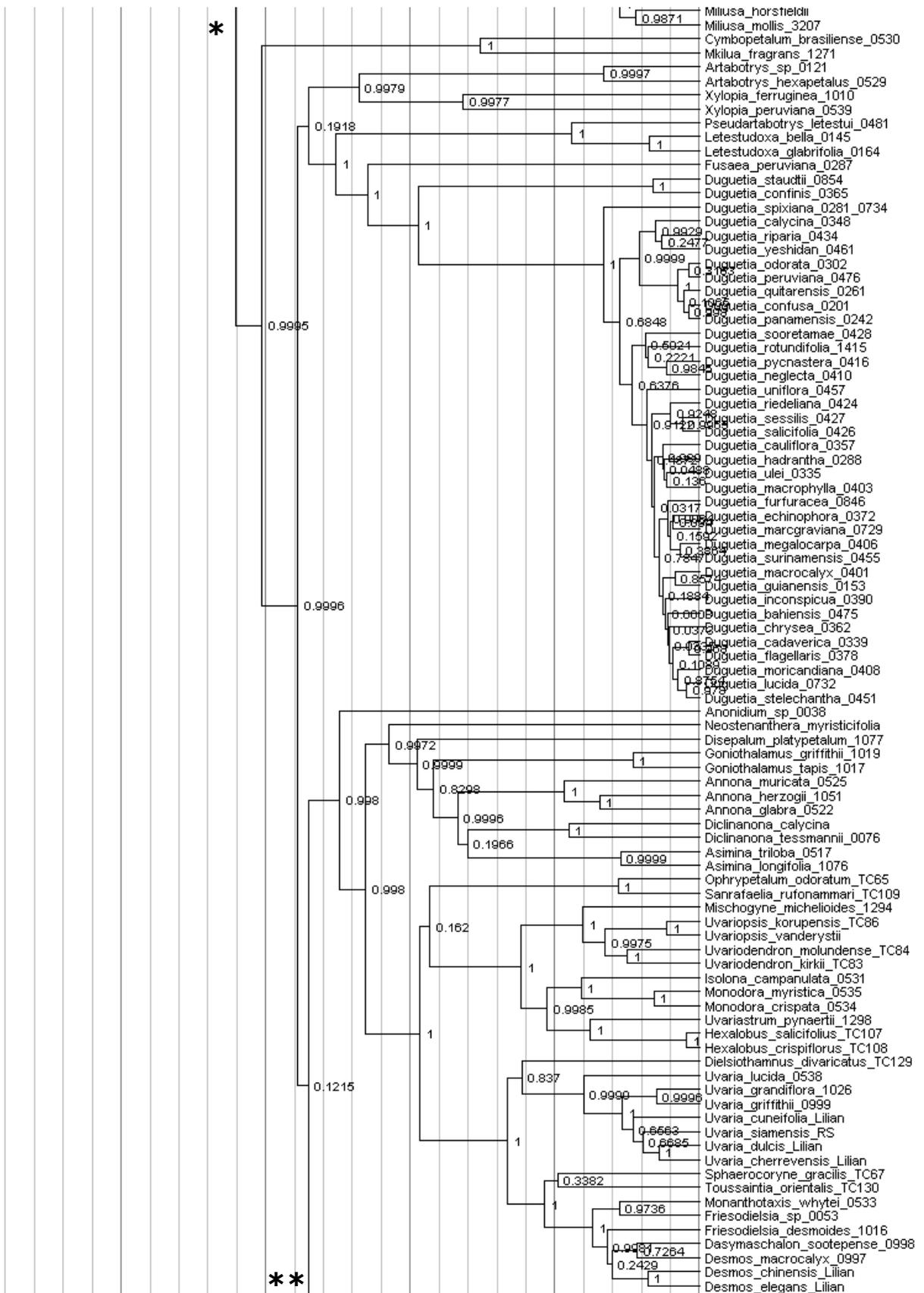


Figure 3: MCC tree for analysis 3 ('*Mosoxandra*' prior excluded) with posterior values on the branches





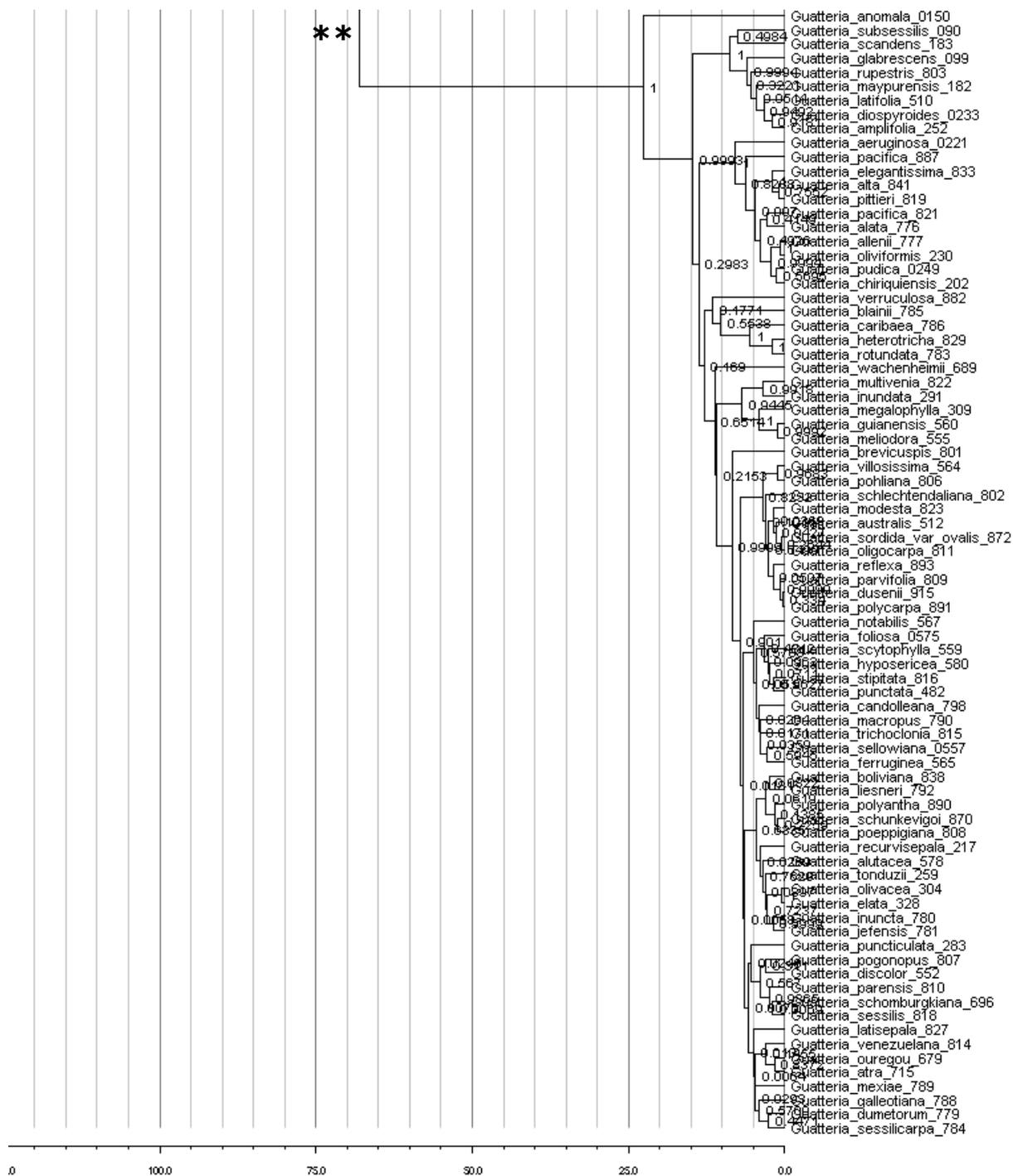
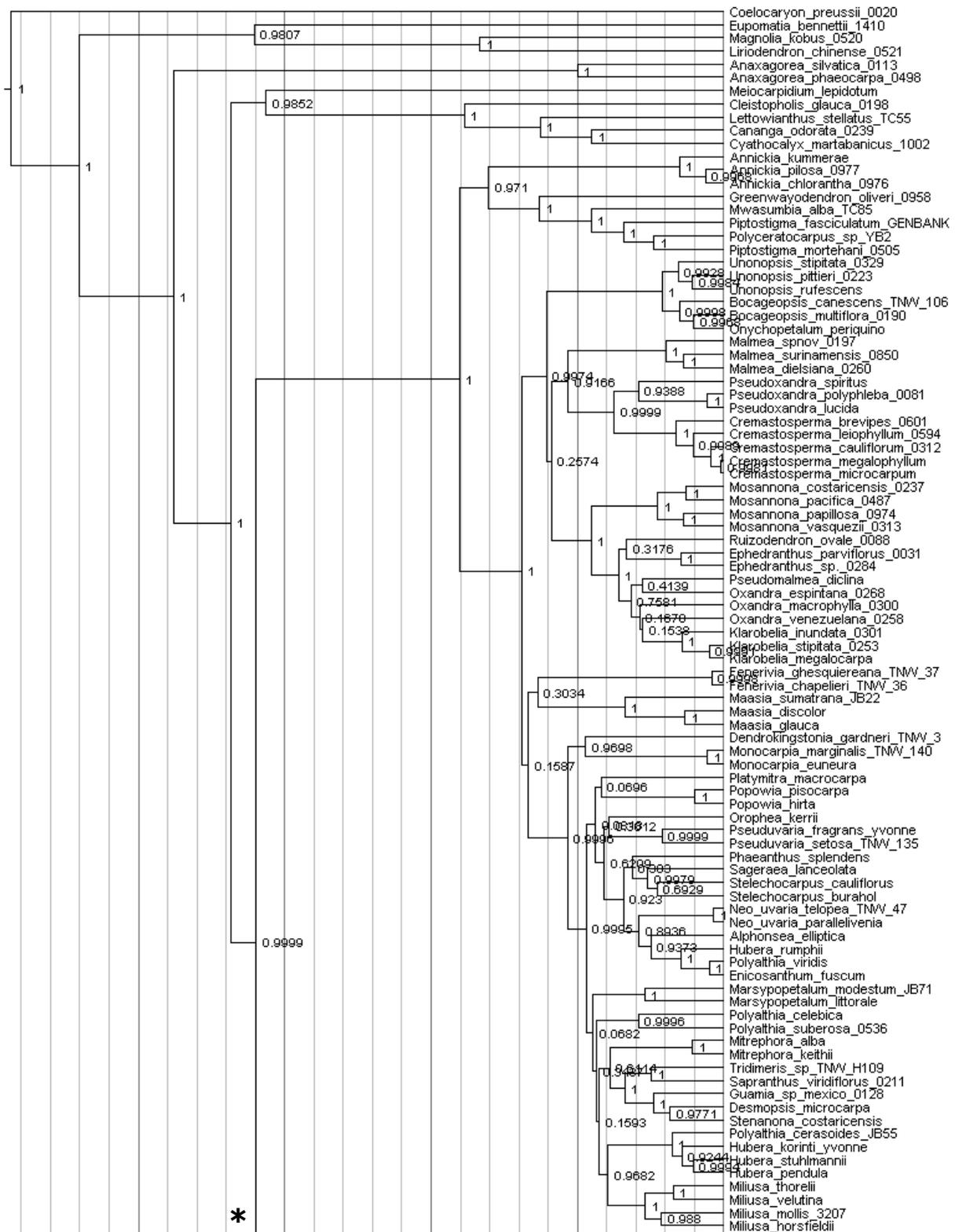
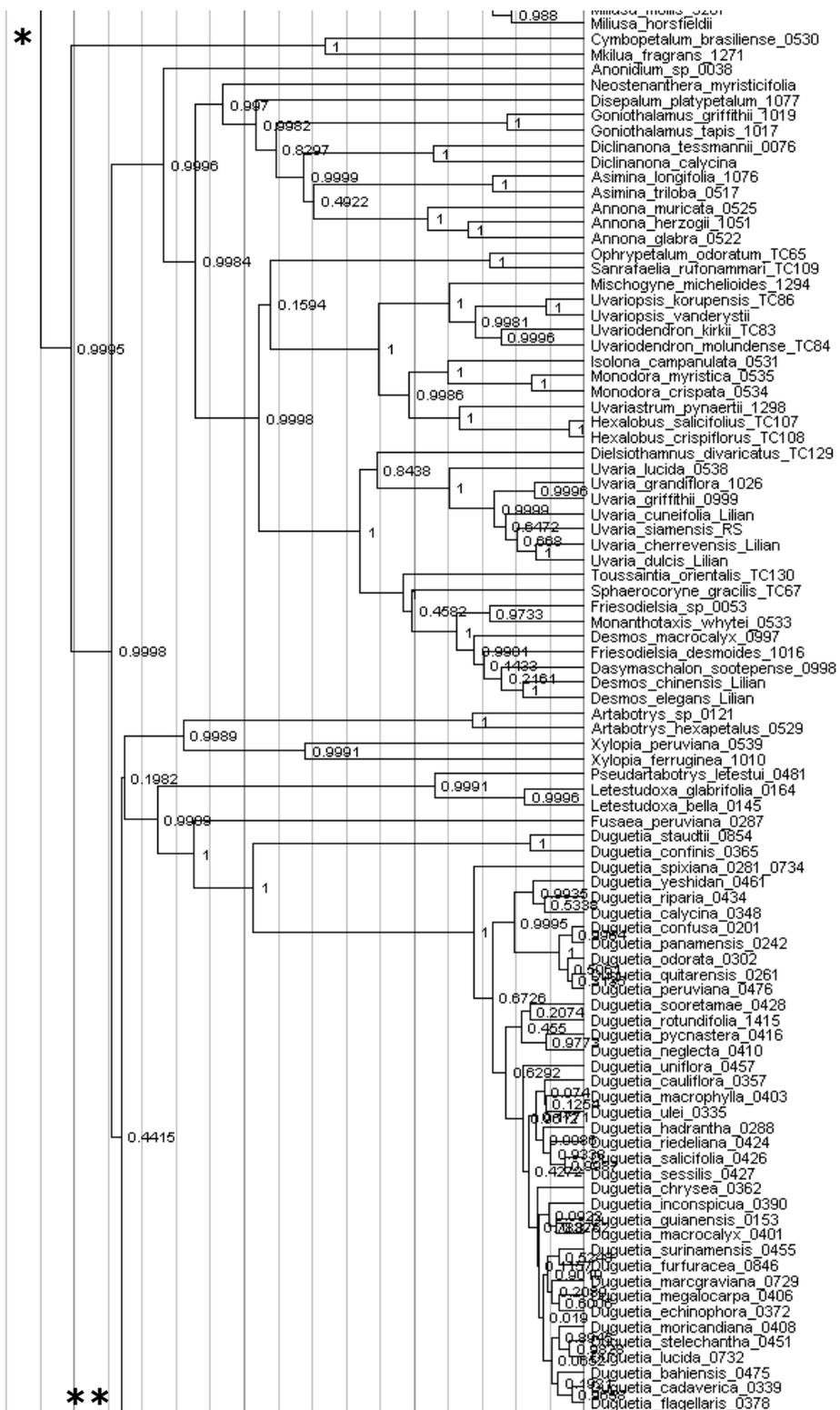


Figure 4: MCC tree for analysis 4 (African *Malmeoidea* prior excluded) with posterior values on the branches





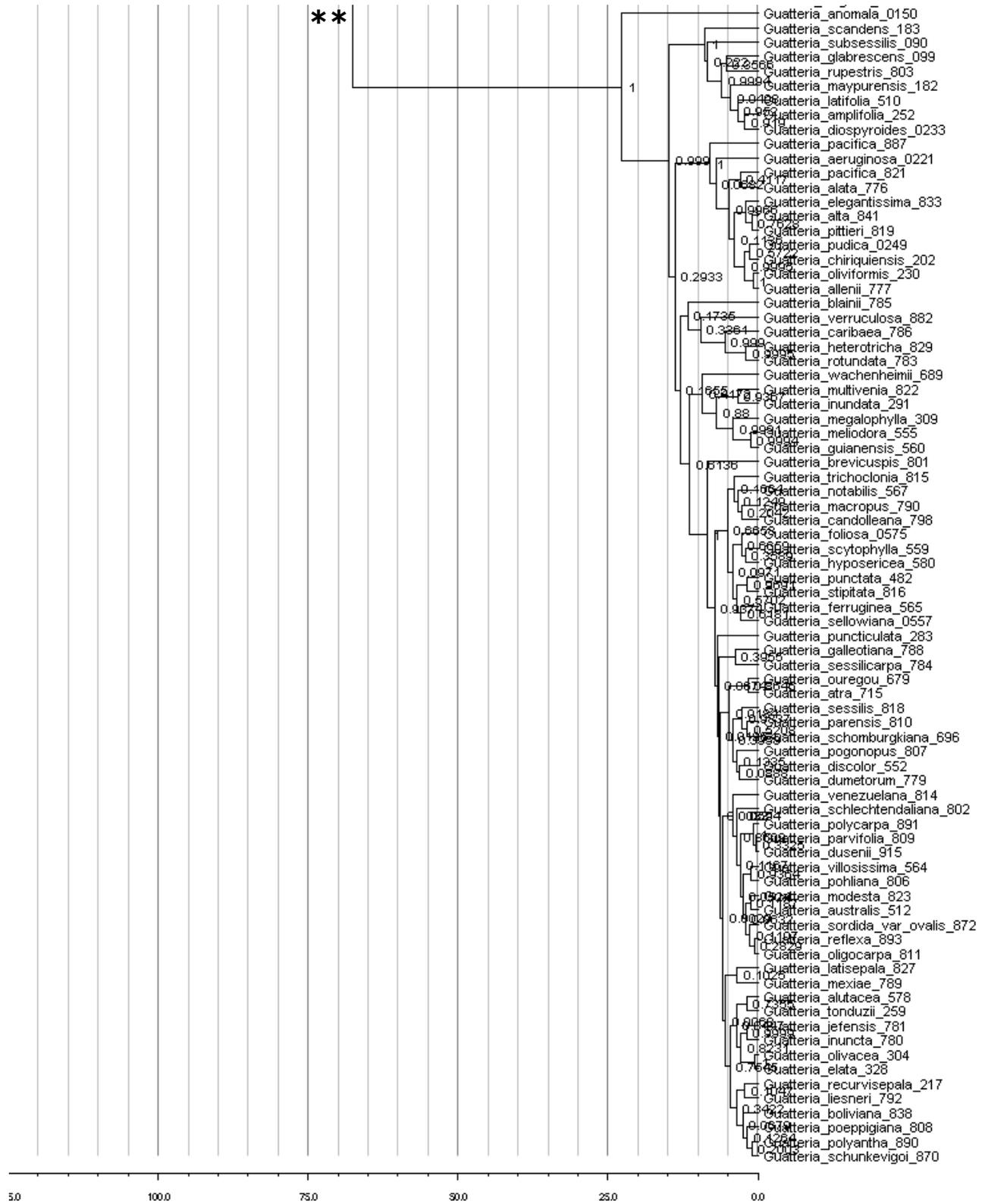
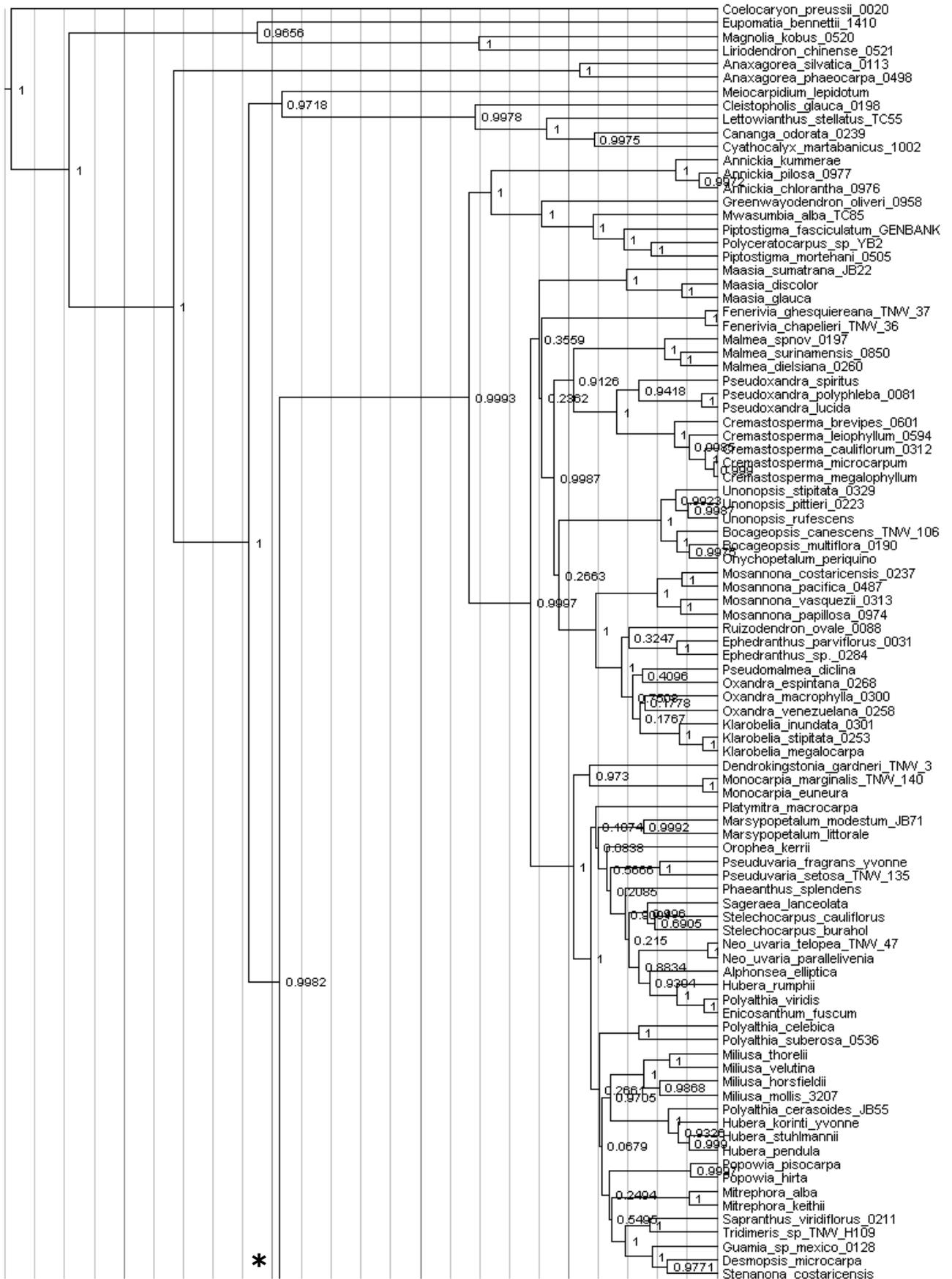


Figure 5: MCC tree for analysis 5 (*Duguetia* prior excluded) with posterior values on the branches



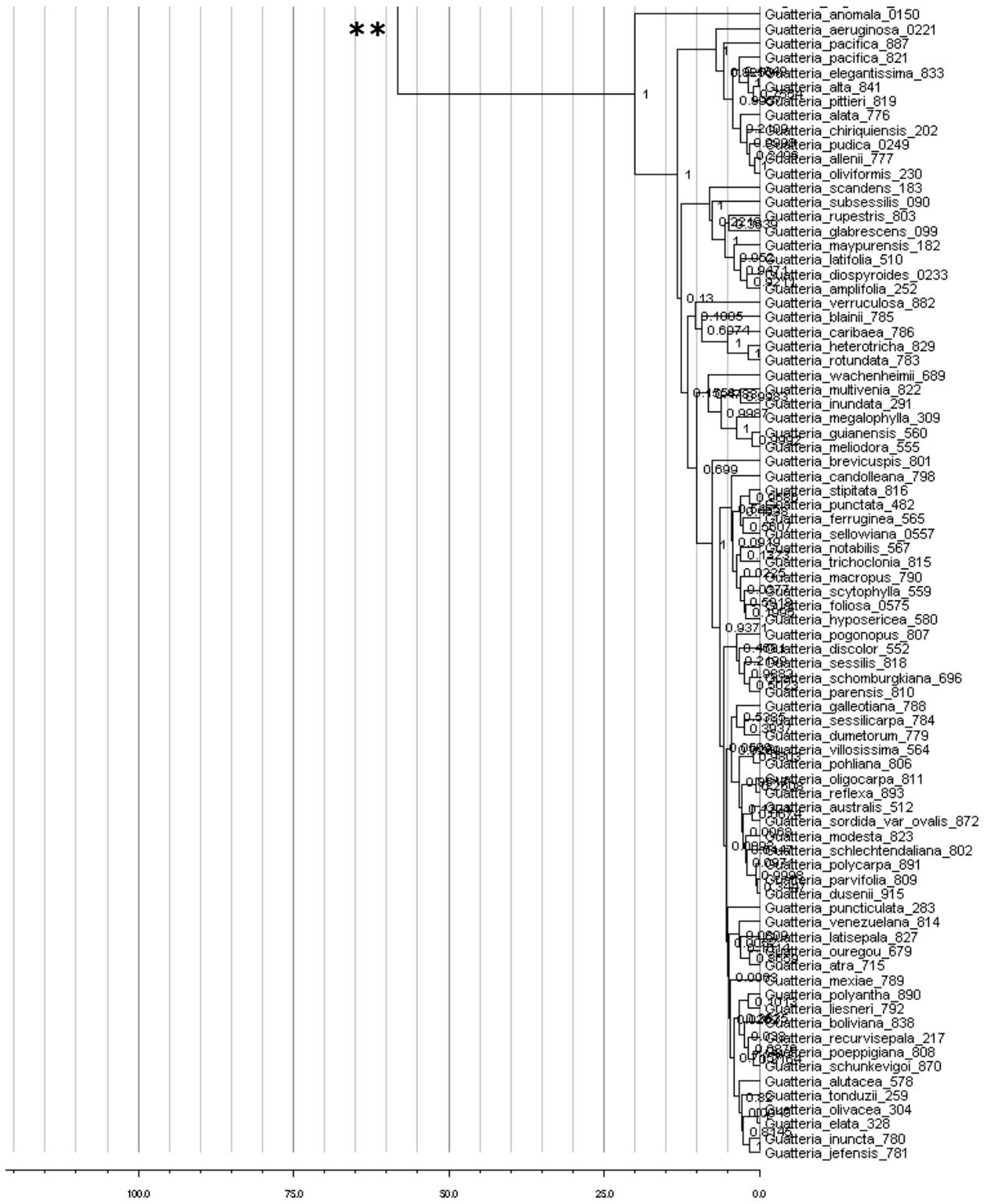
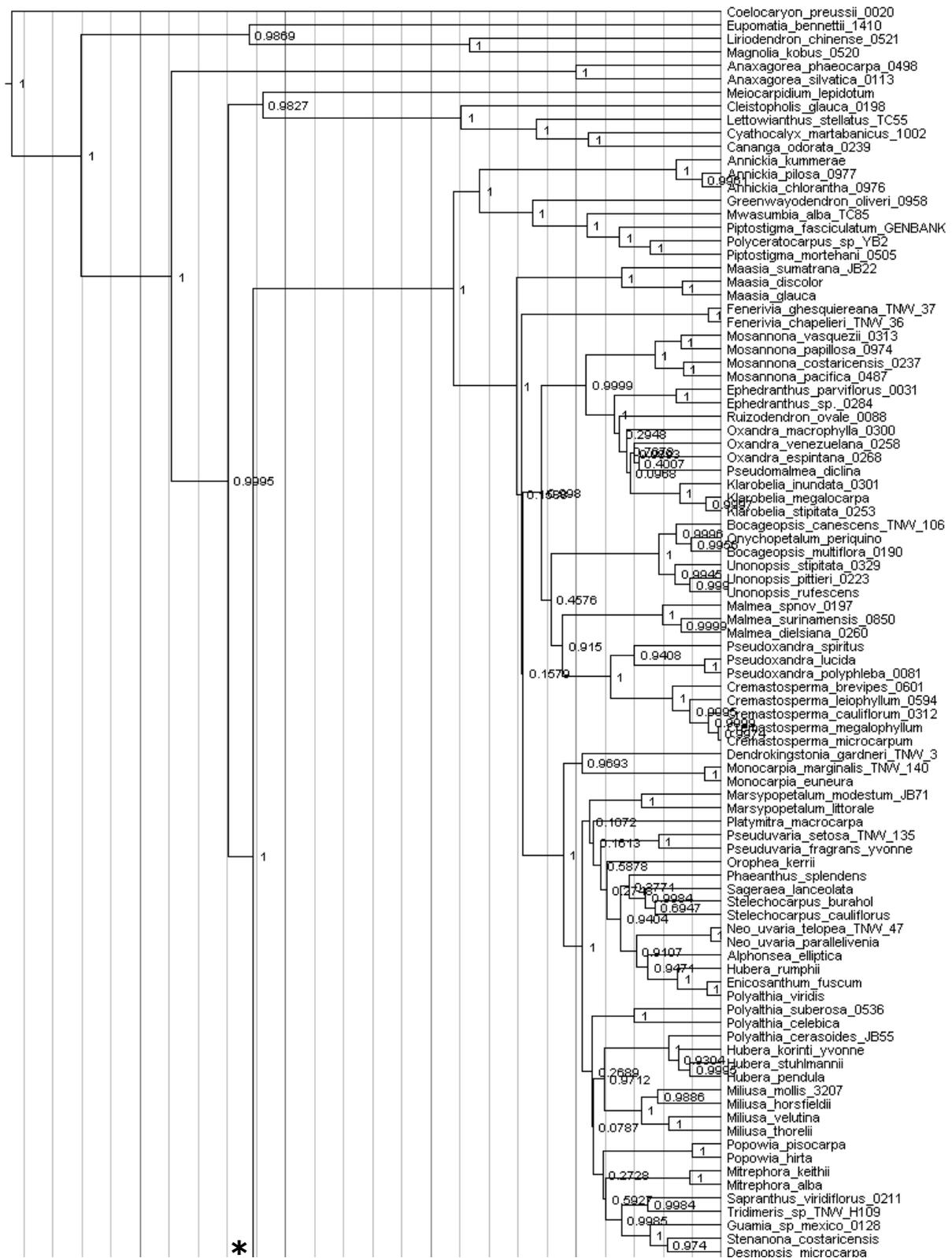
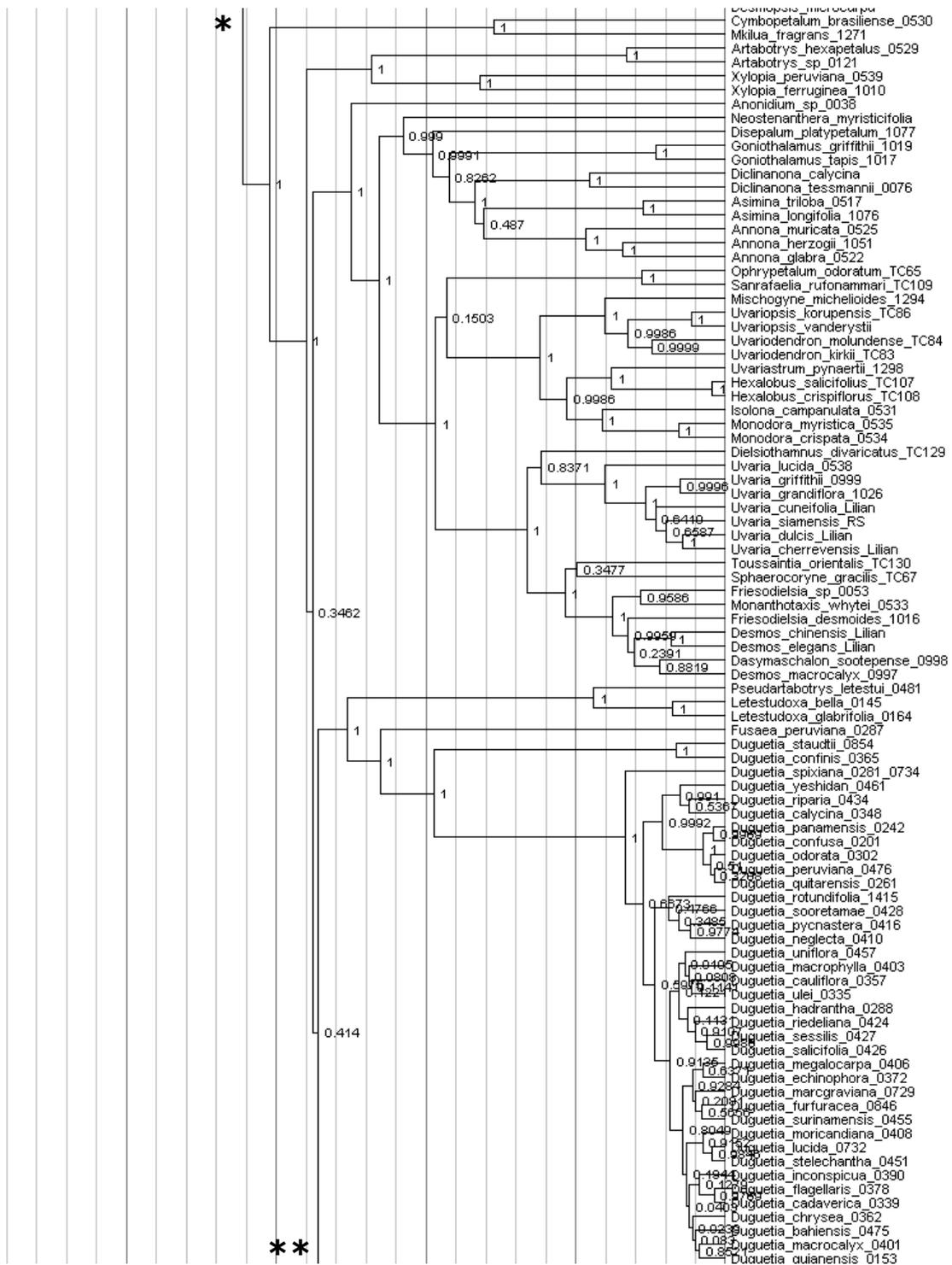


Figure 6: MCC tree for analysis 6 (*Anonasperrum* prior excluded) with posterior values on the branches





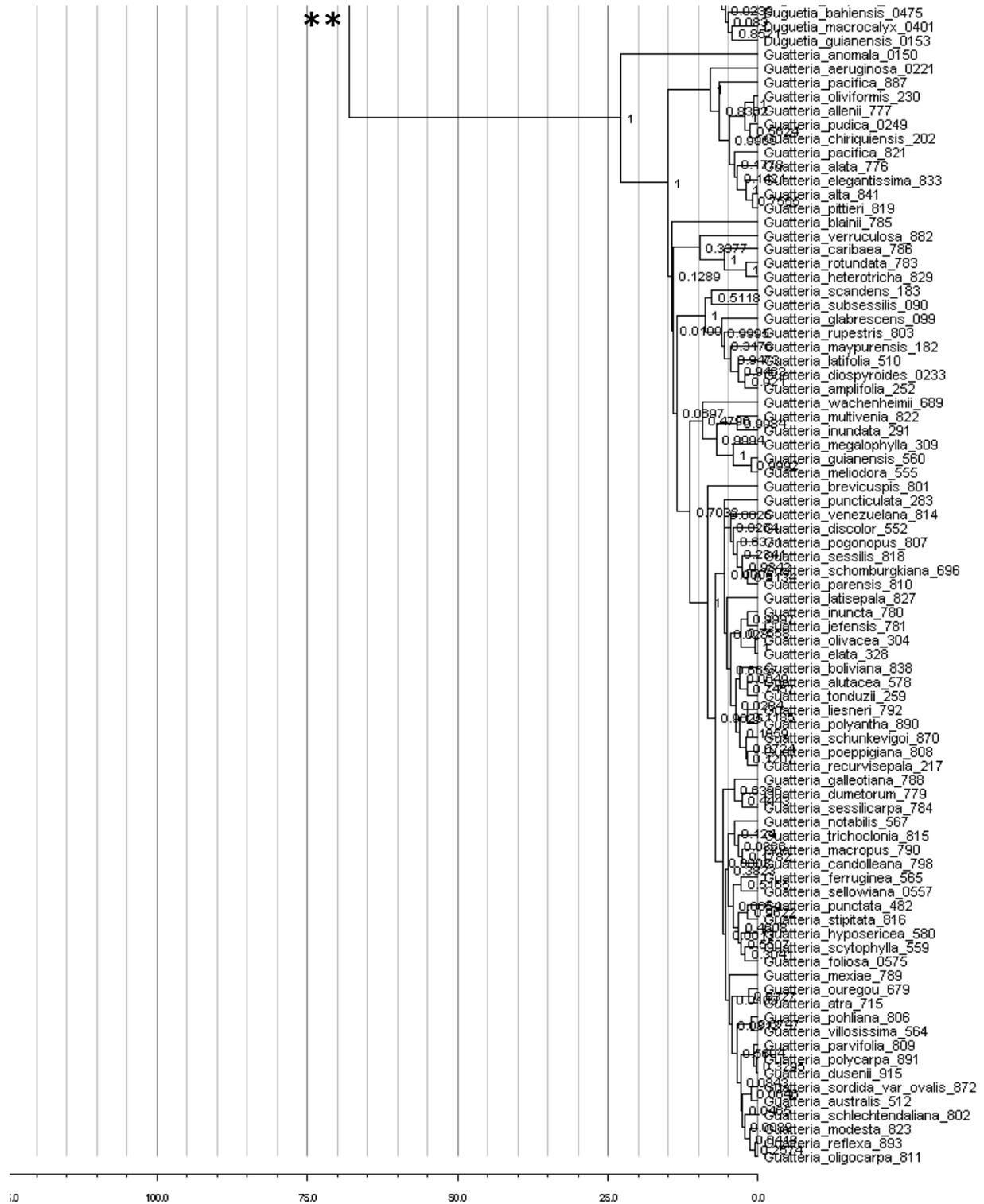
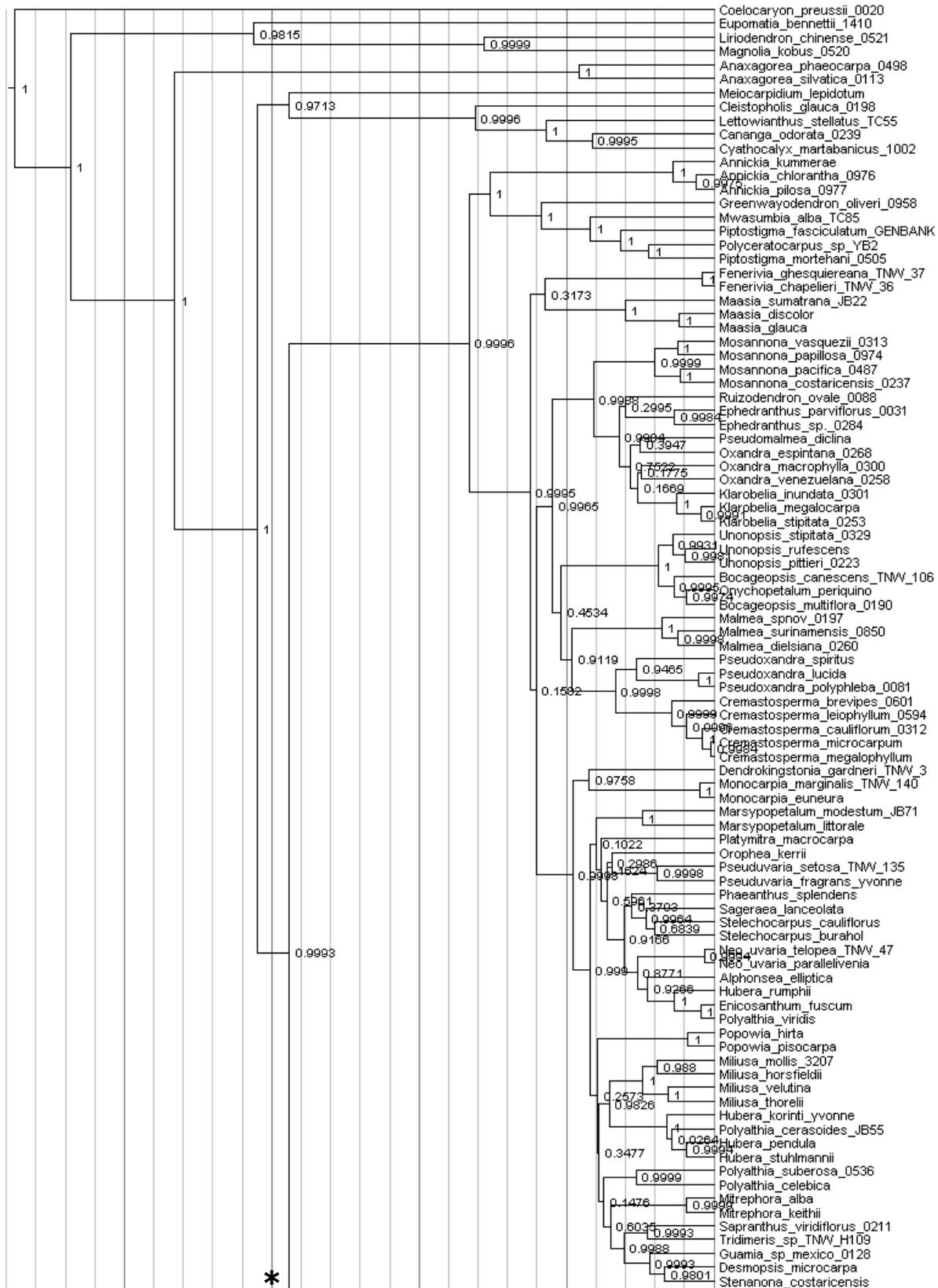
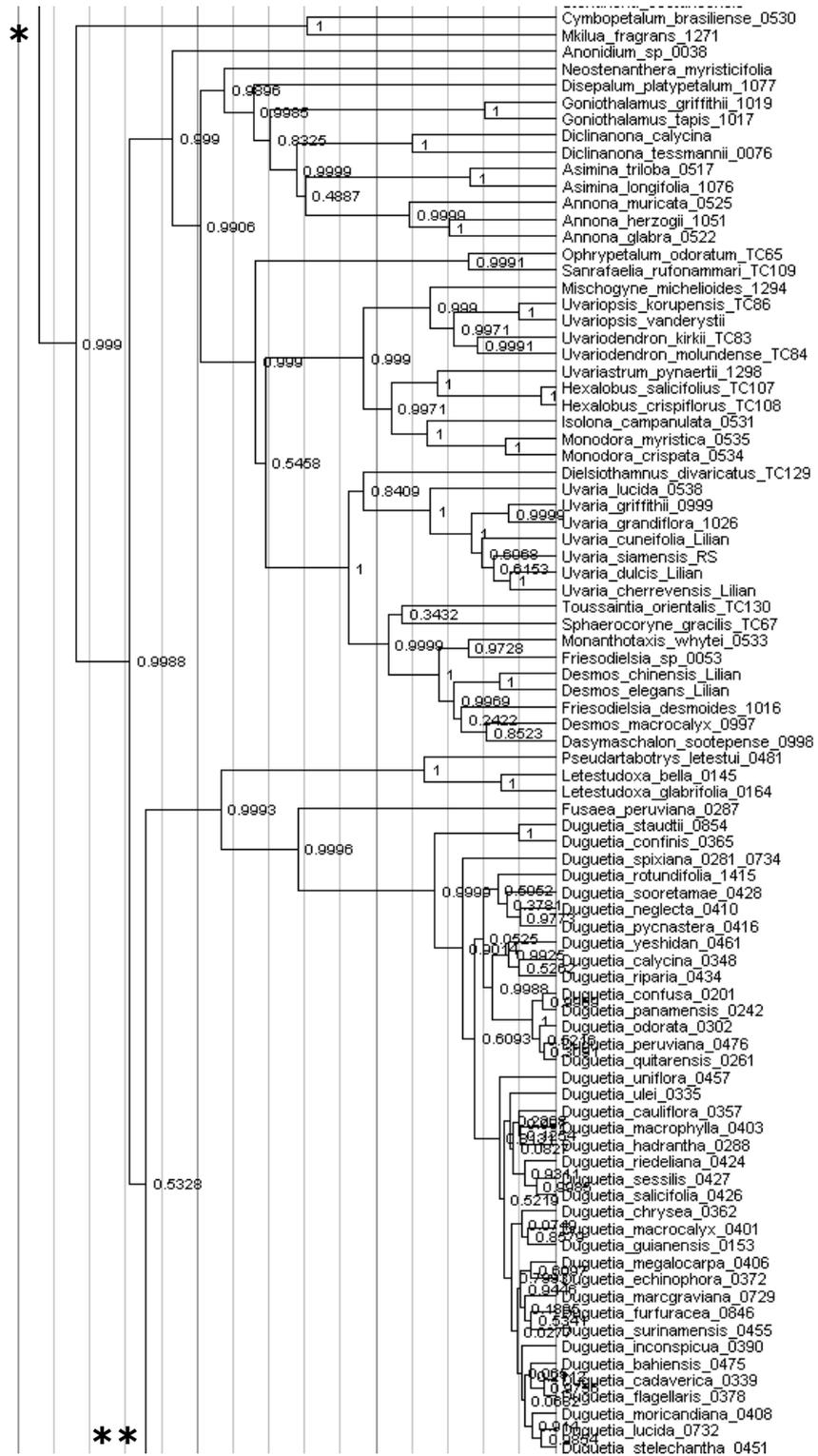
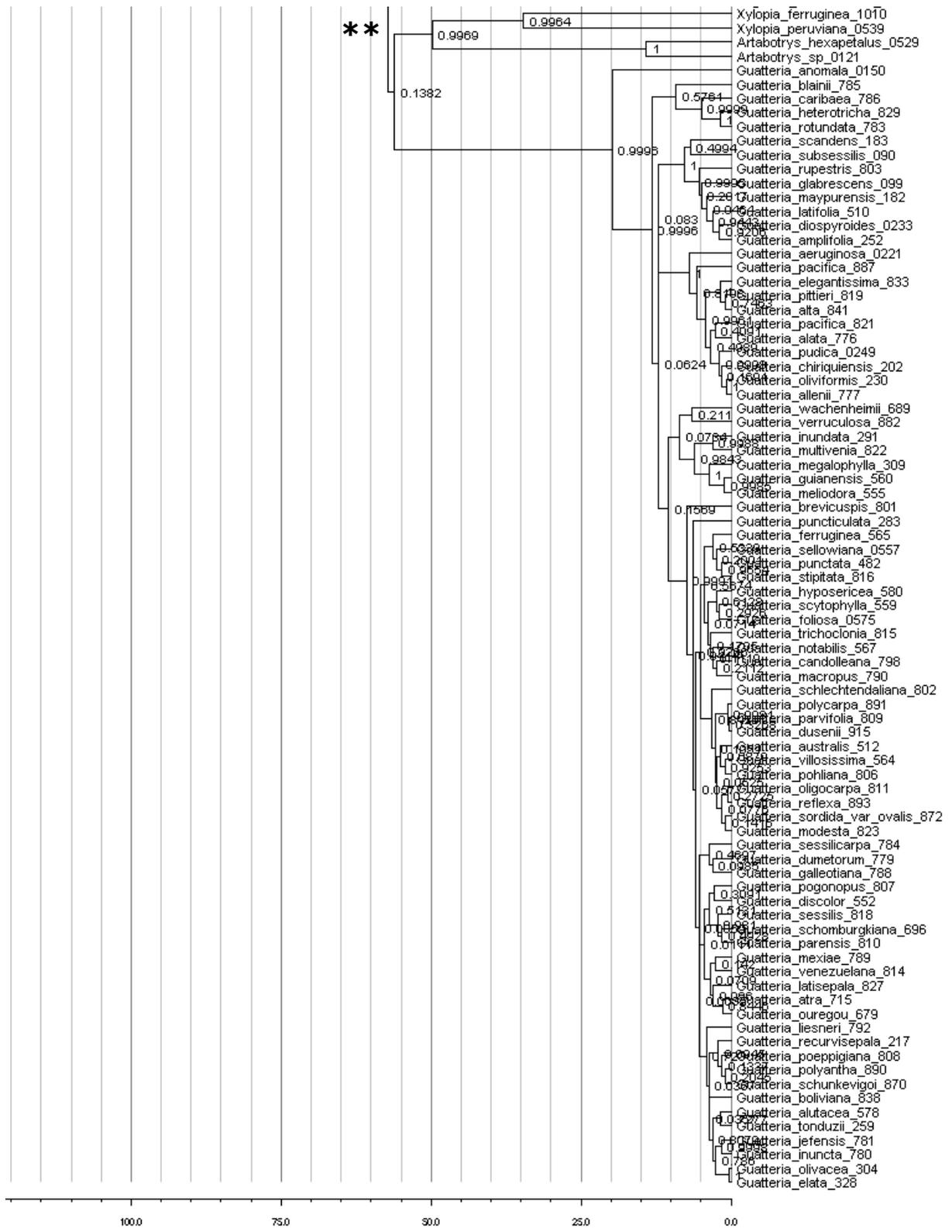
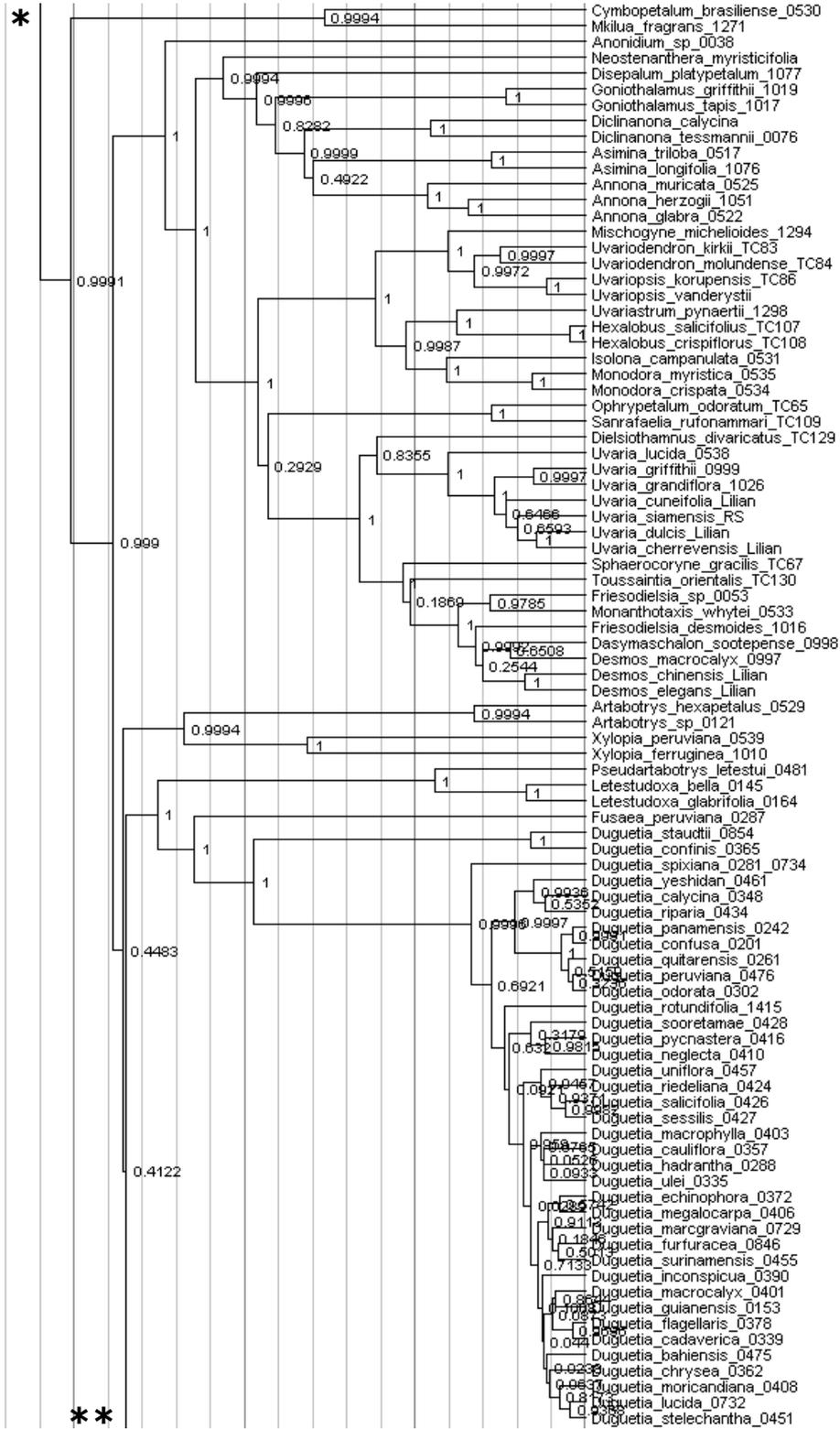


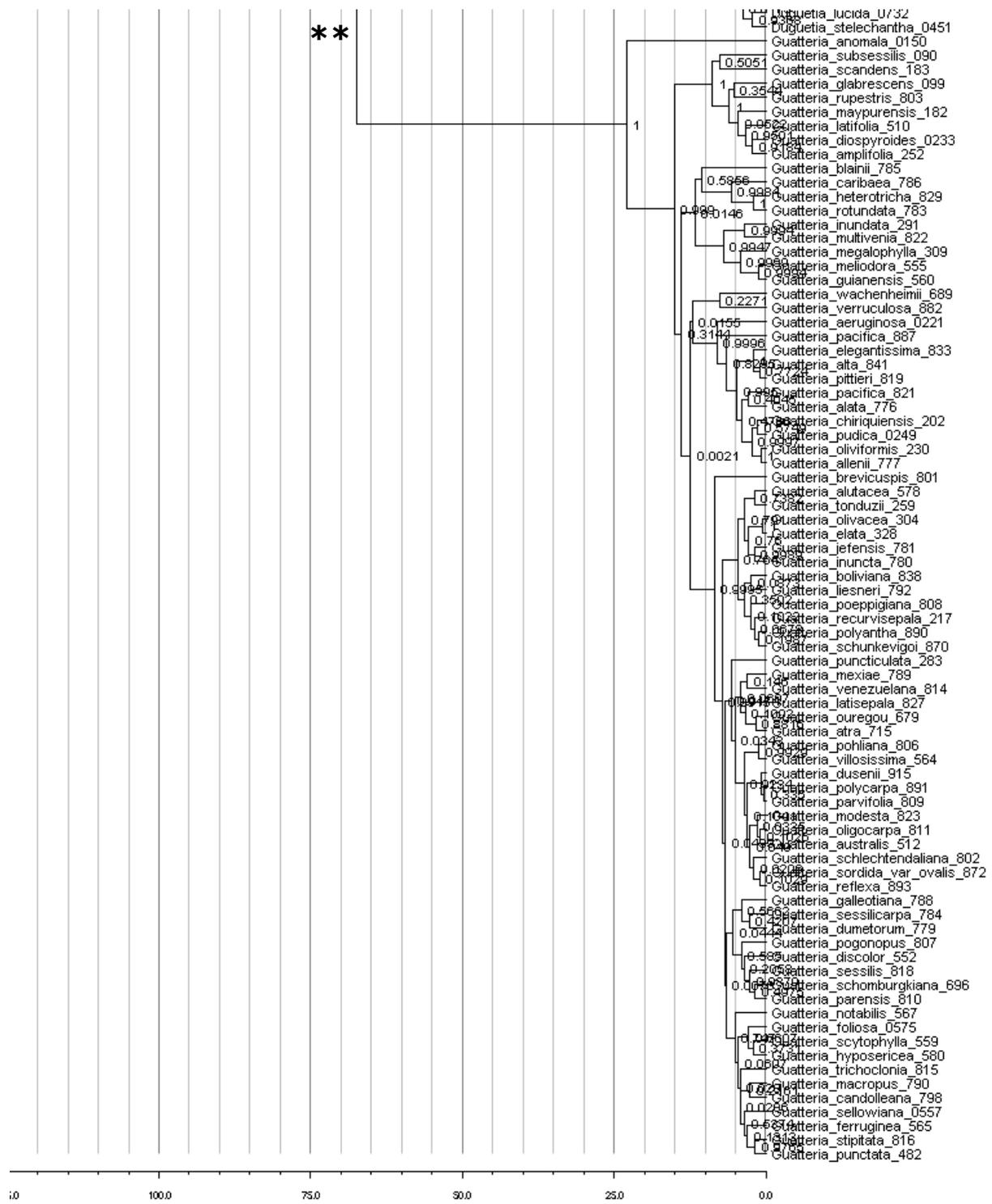
Figure 7: MCC tree for analysis 7 (LBC fossils excluded) with posterior values on the branches

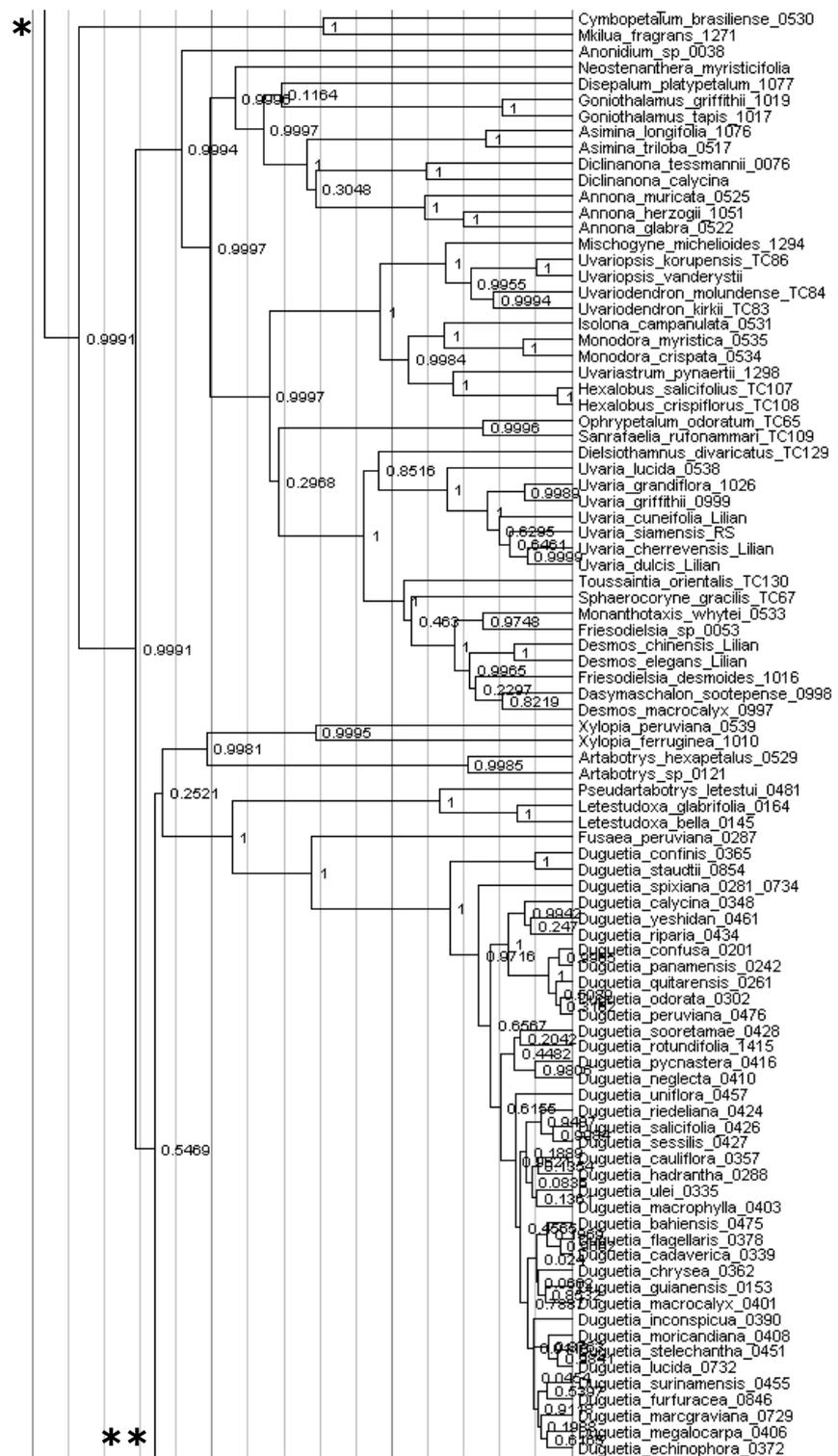












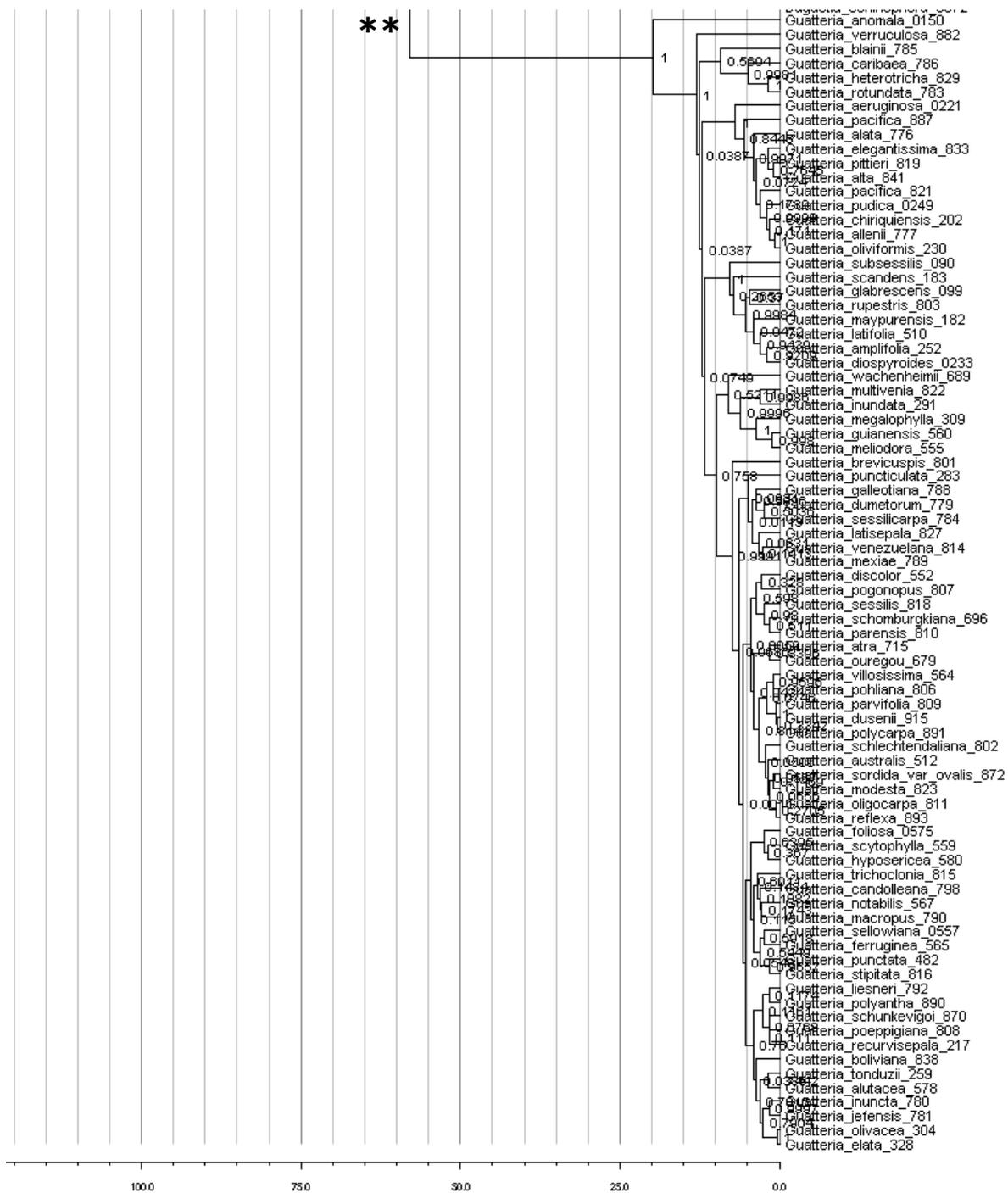
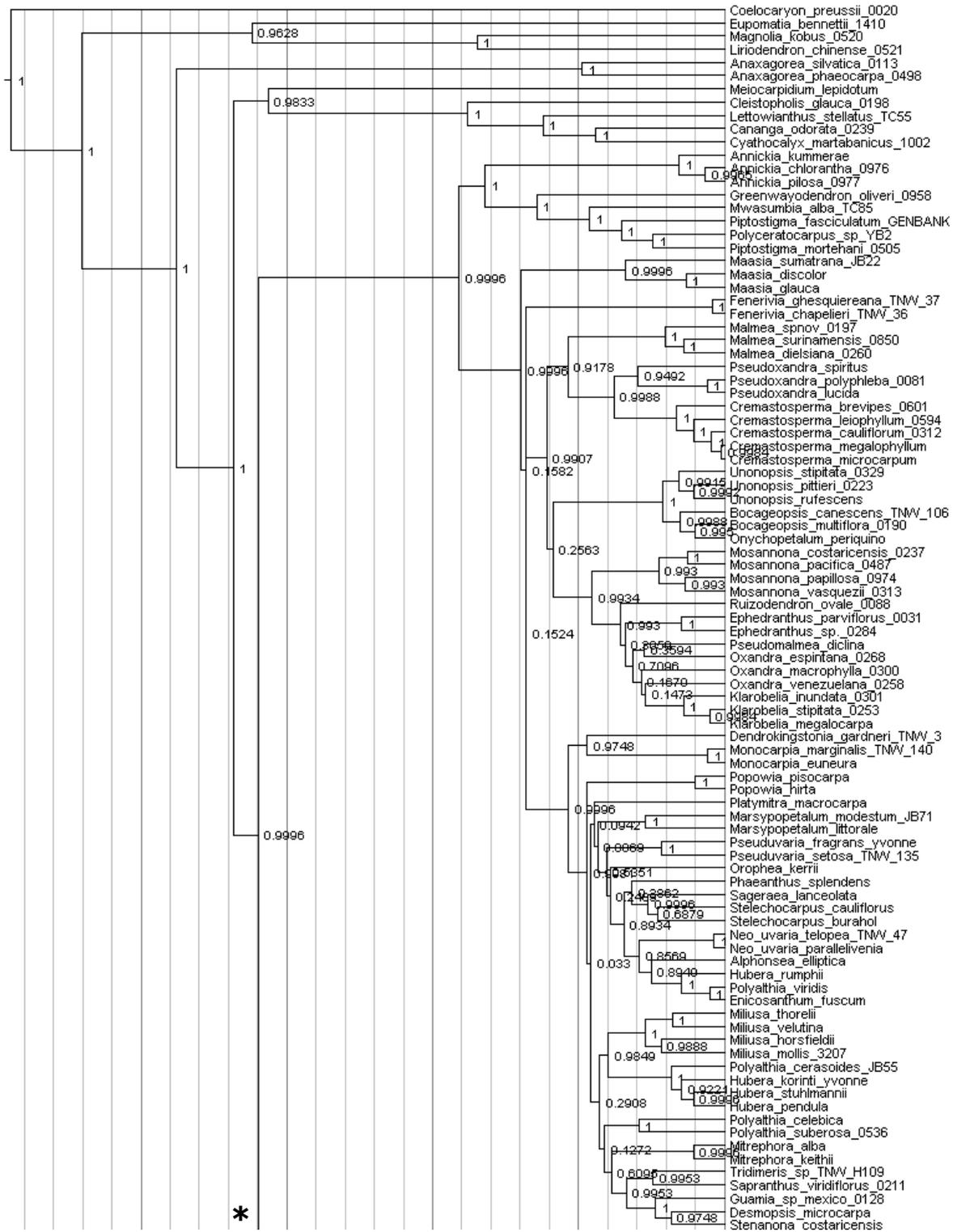


Figure 10: MCC tree for analysis 10 ('*Mosoxandra*' and *Anonasperrum* priors excluded) with posterior values on the branches



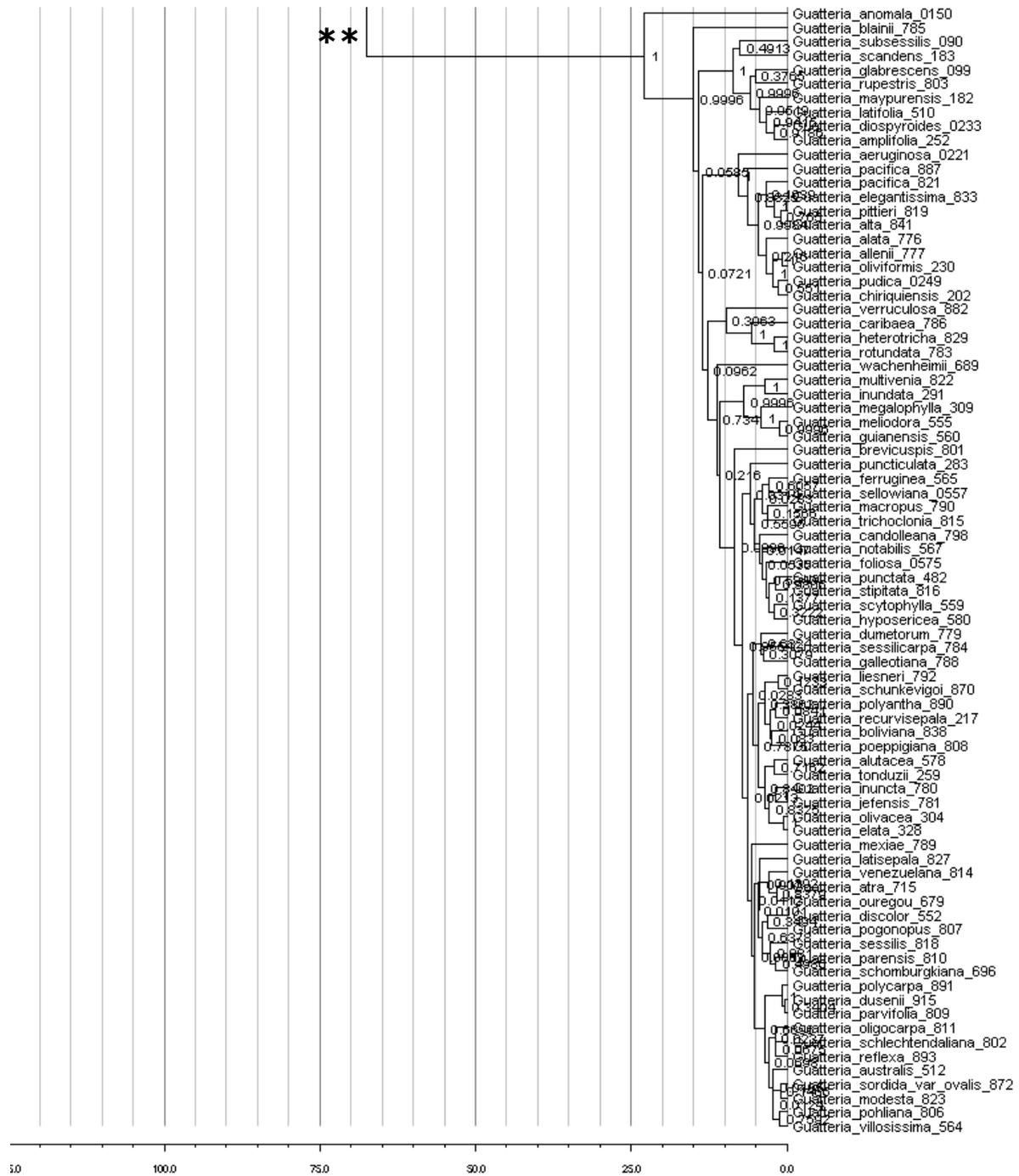
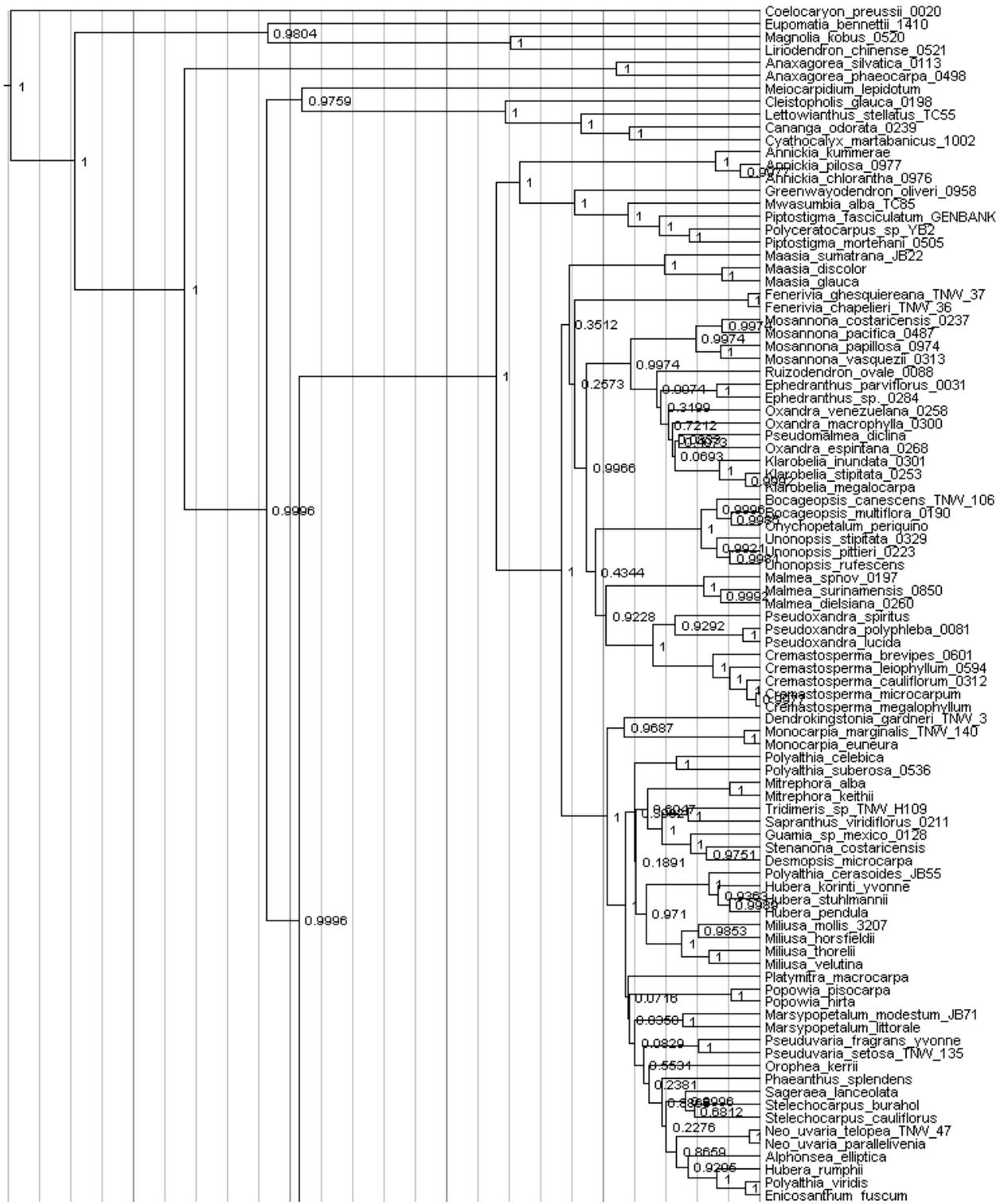
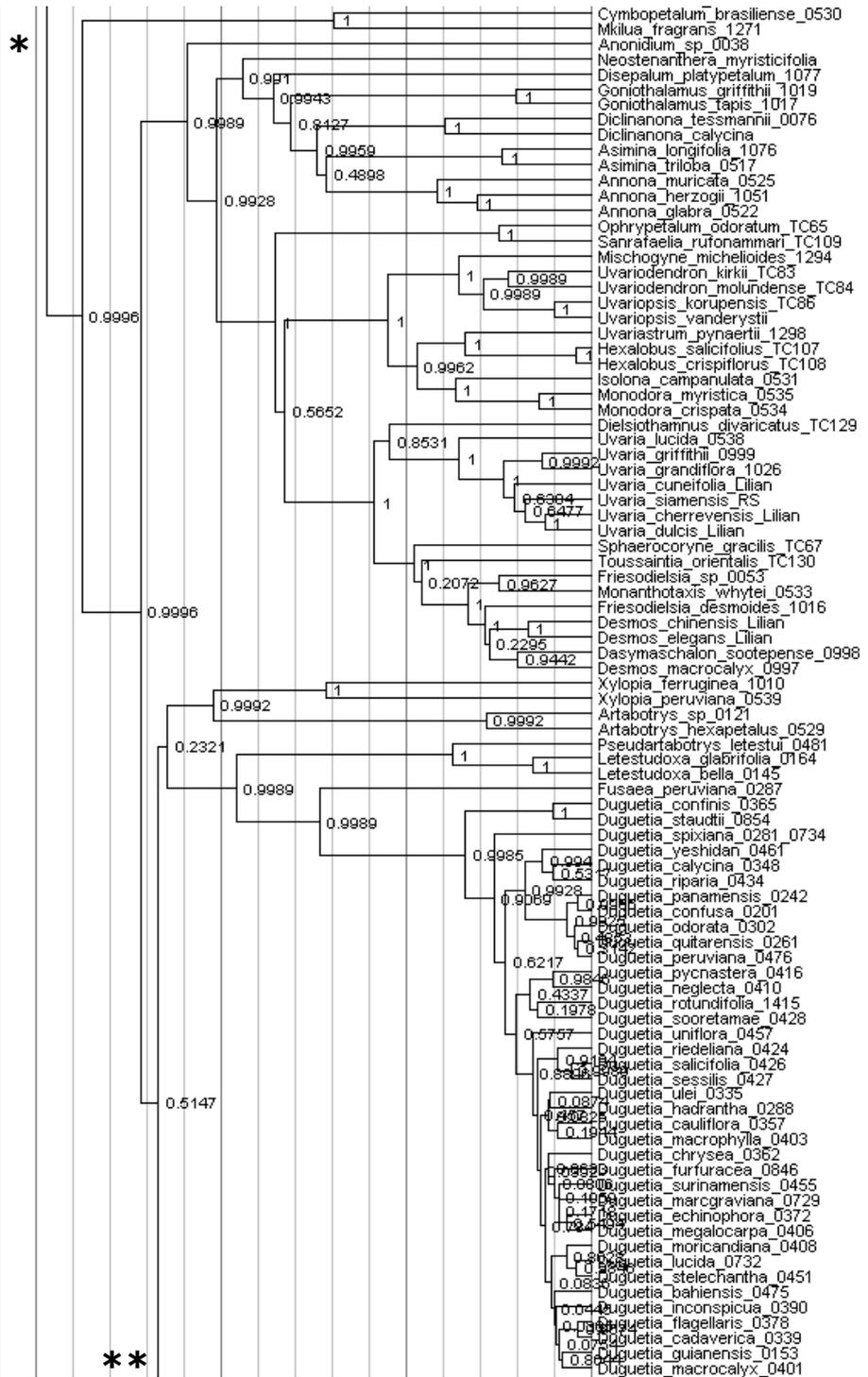


Figure 11: MCC tree for analysis 11 ('*Mosoxandra*' and *Anonaspermum* priors excluded) with posterior values





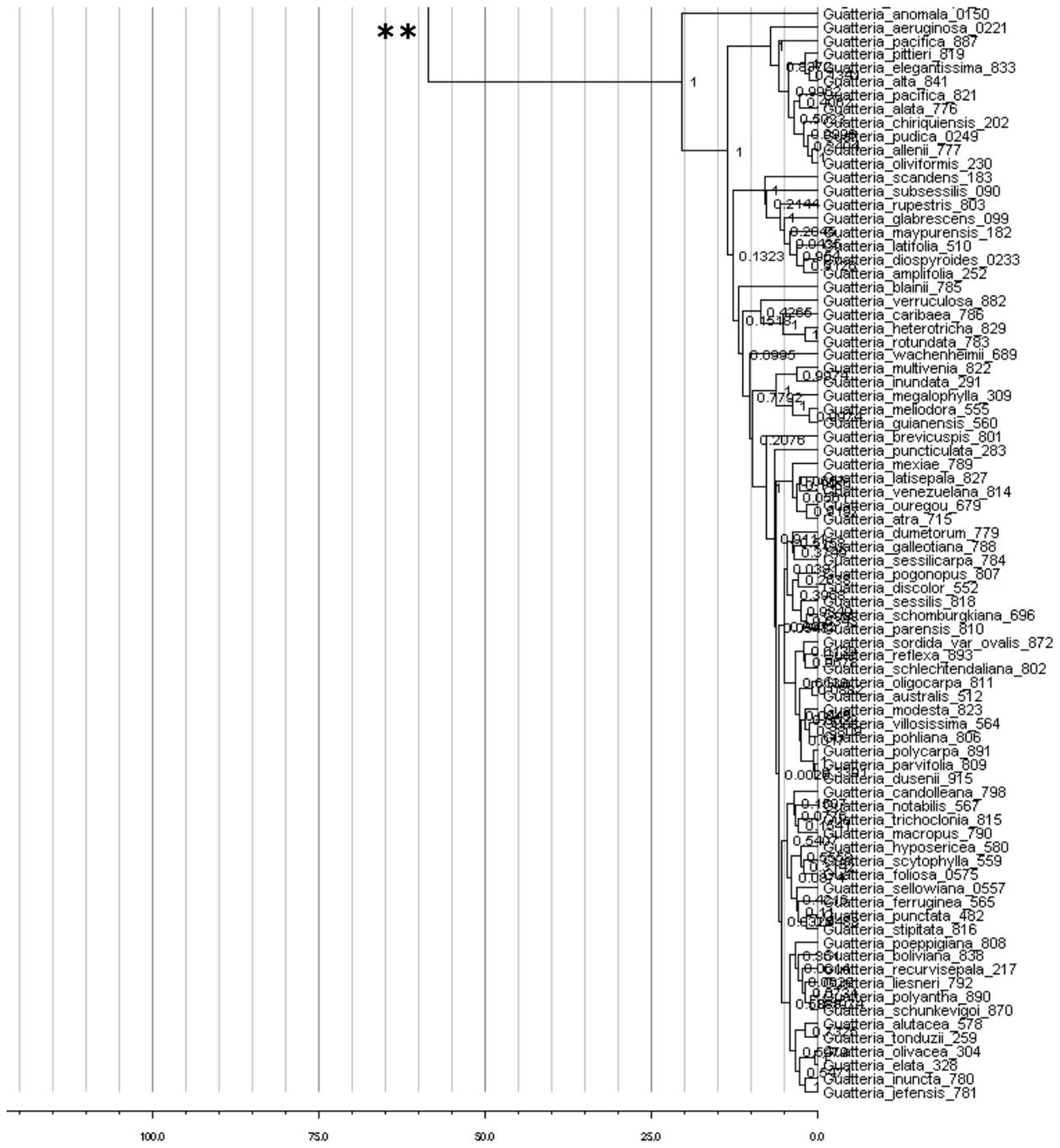
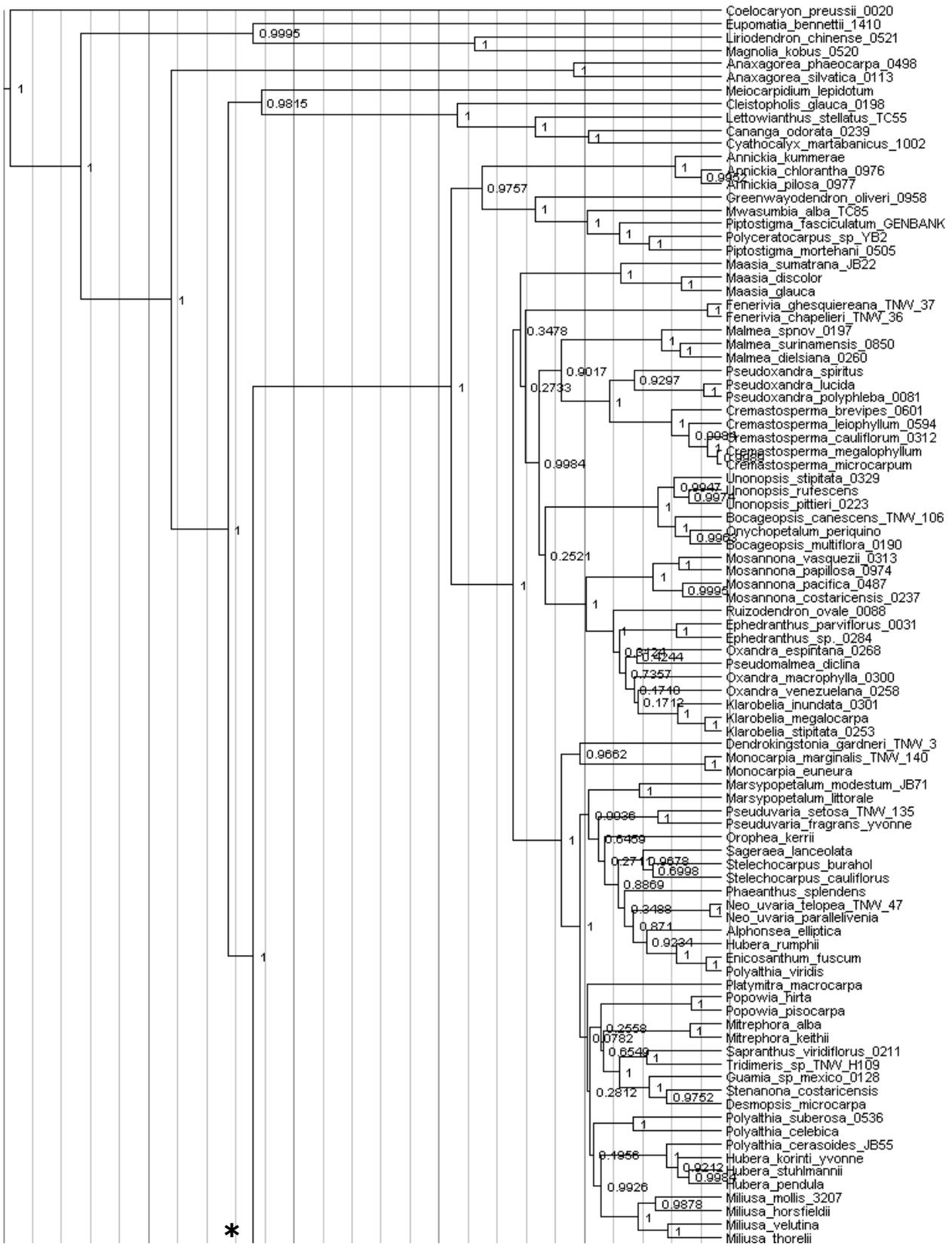
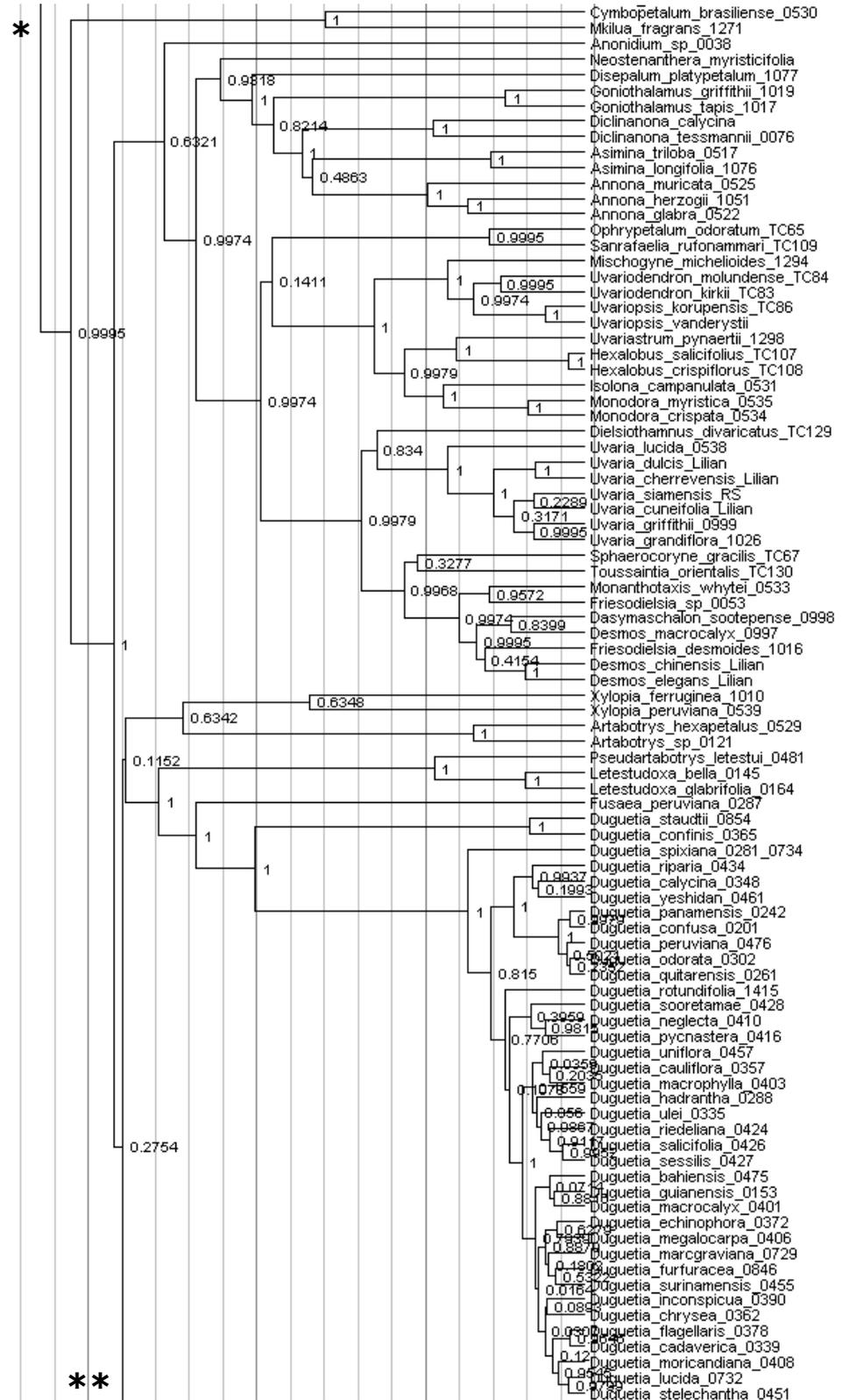
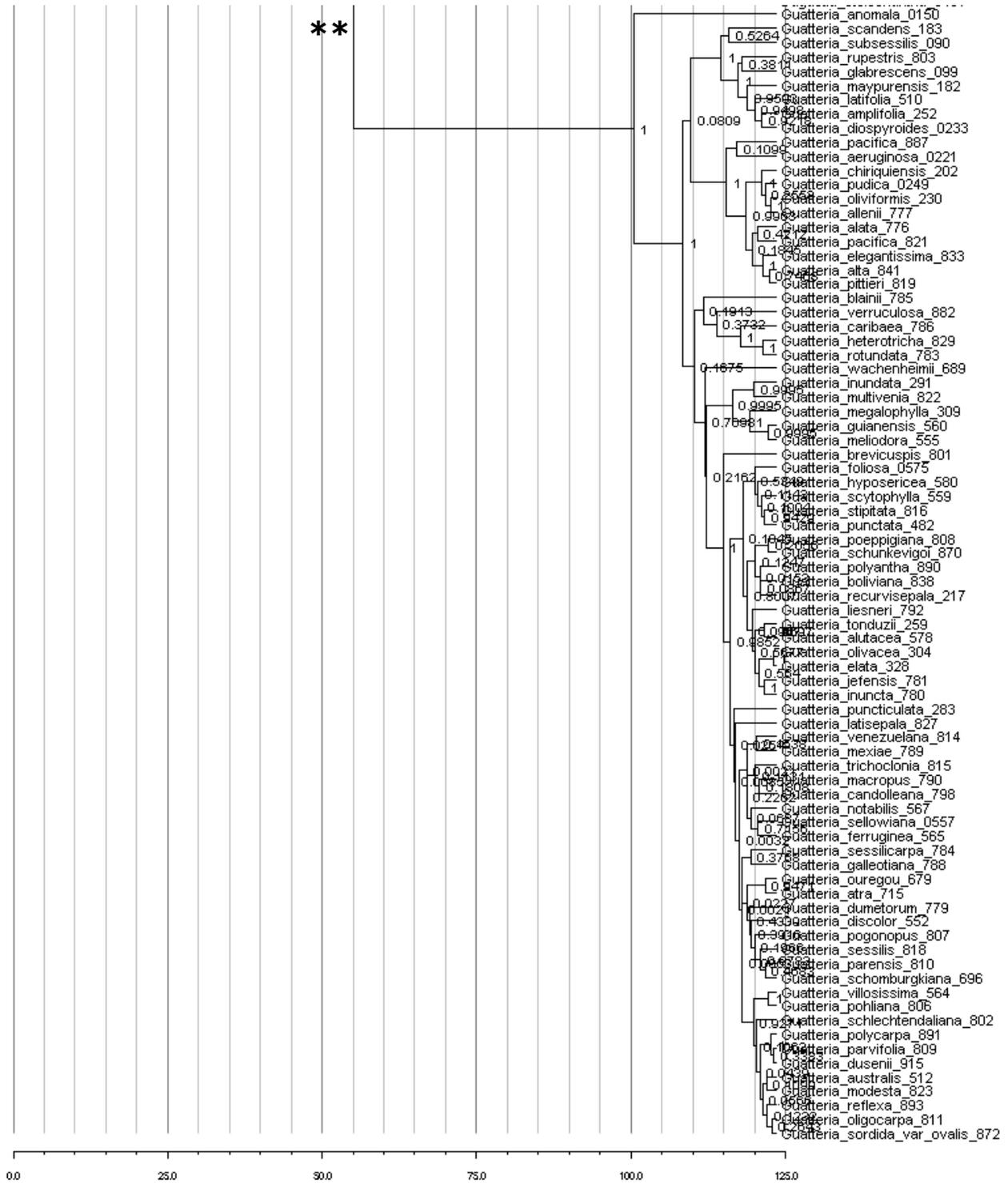


Figure 12: MCC tree for analysis 12 (SBC priors excluded) with posterior values on the branches







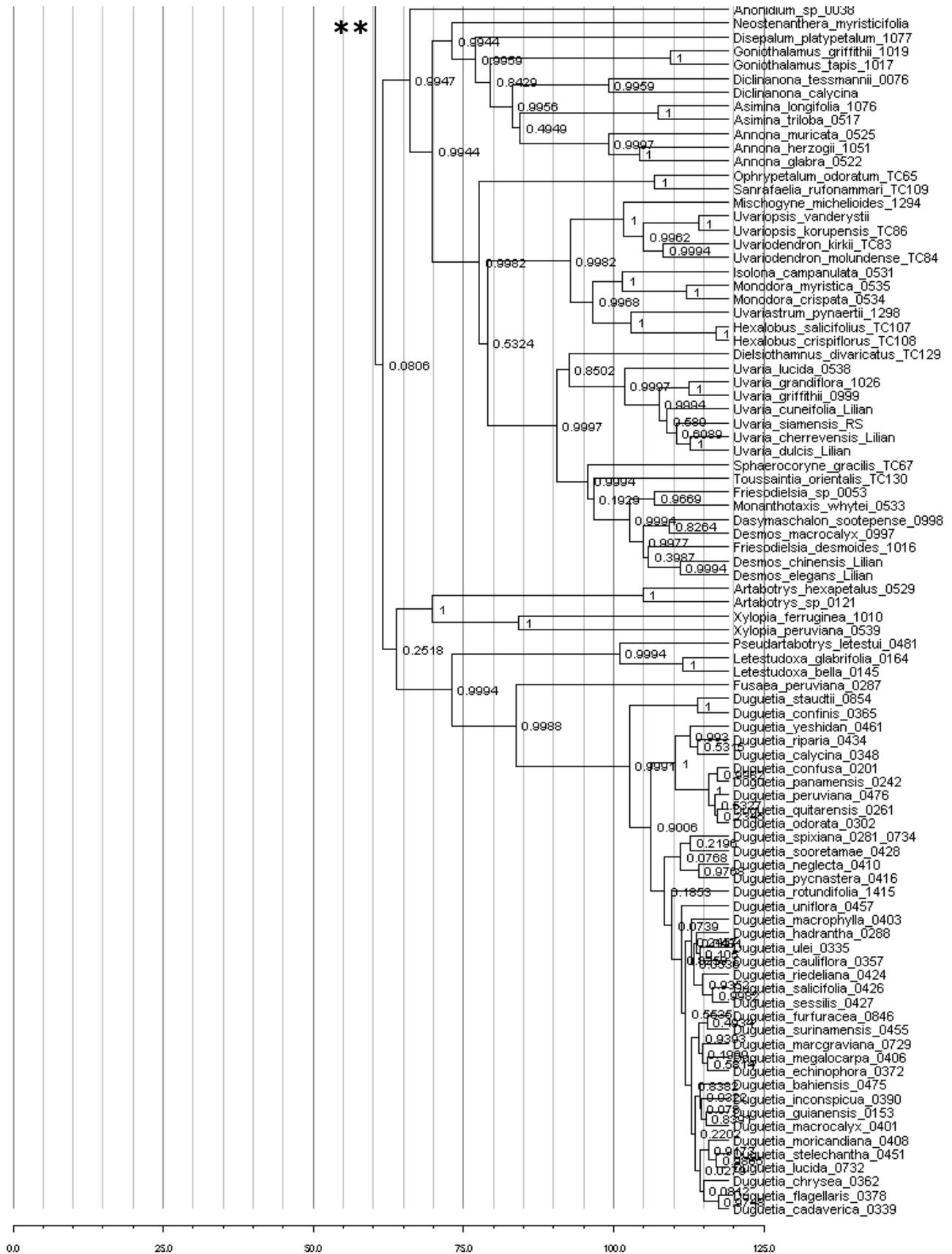
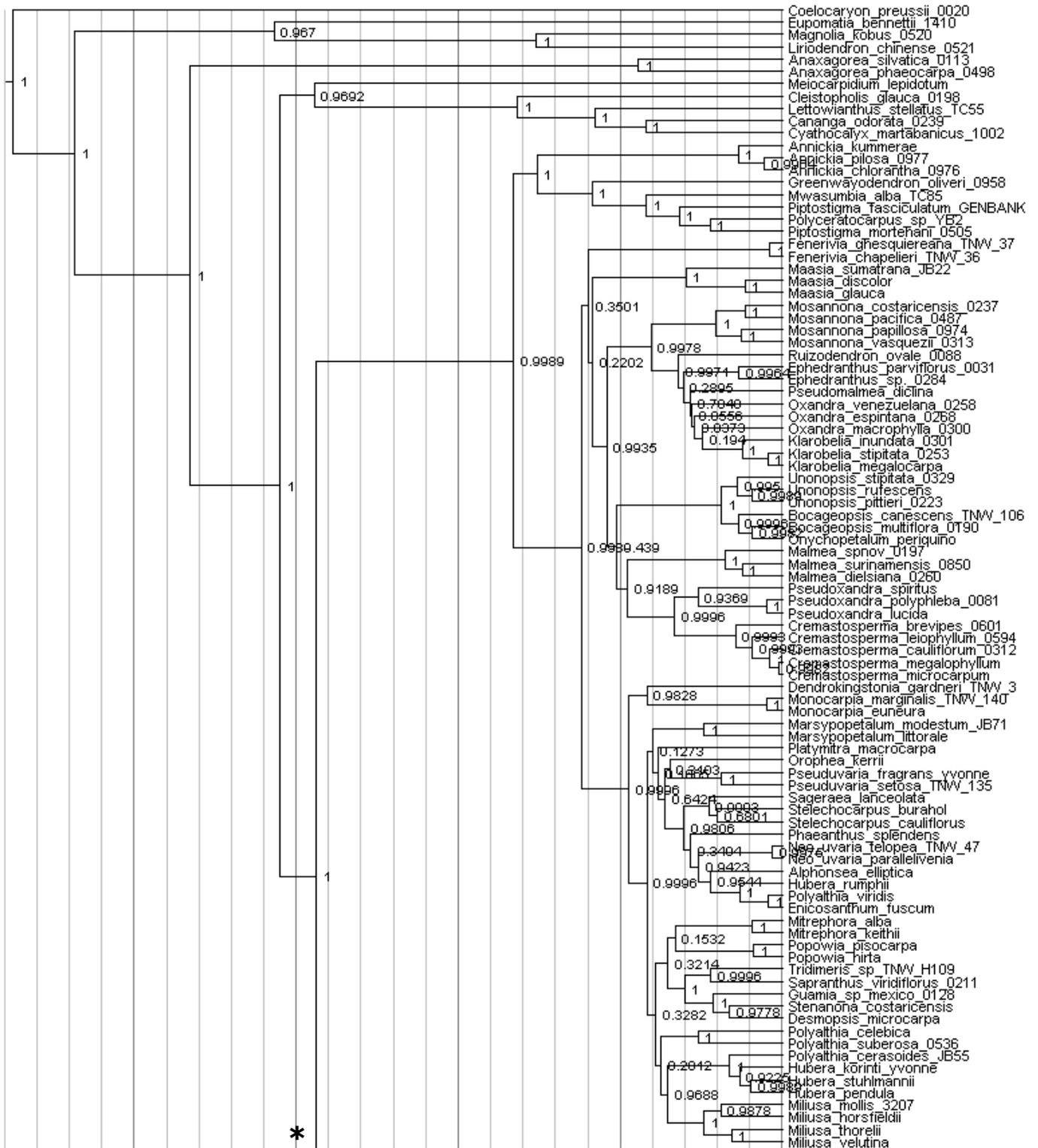
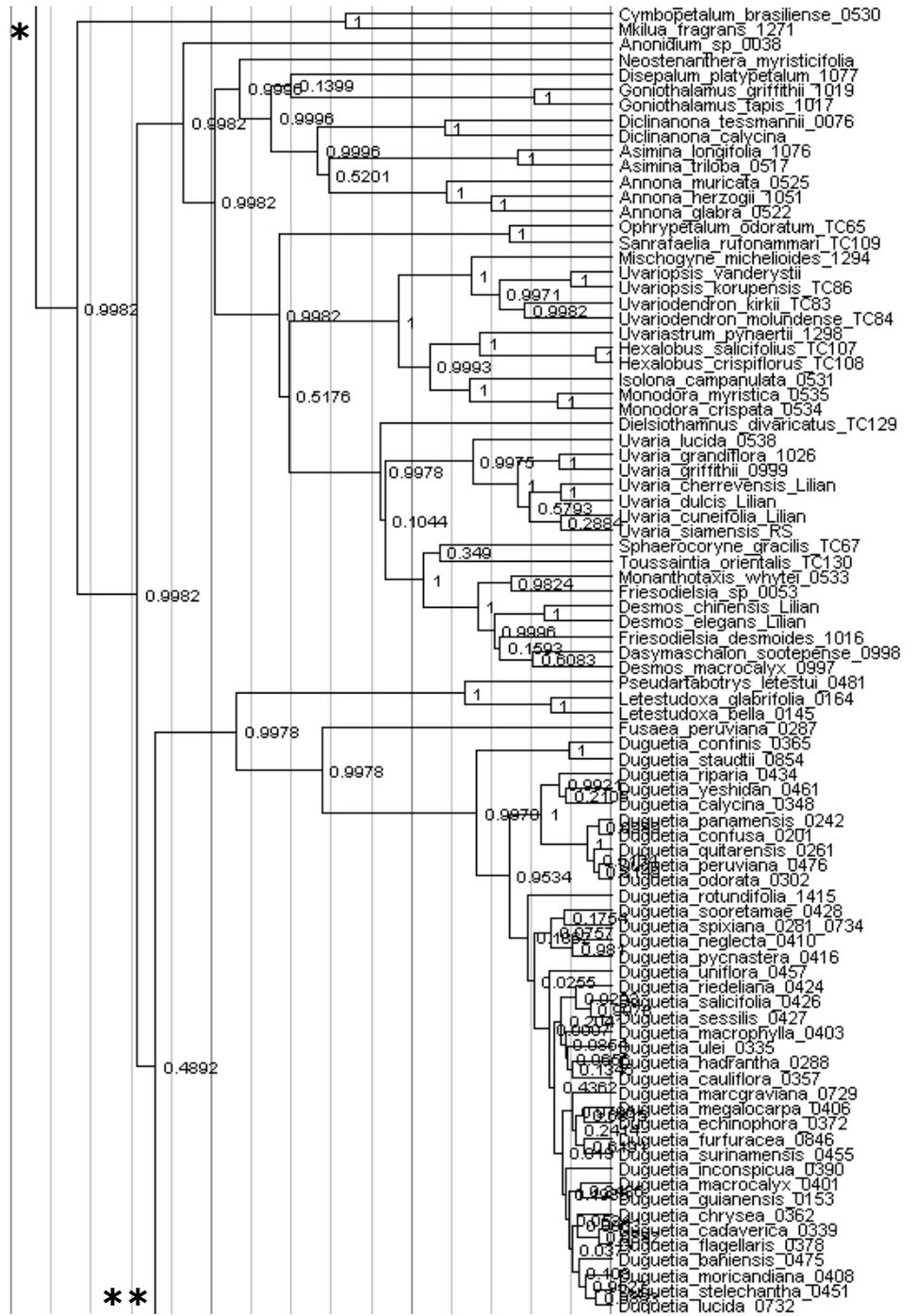


Figure 14: MCC tree for analysis 14 (only African Malmeoideae included) with posterior values on the branches





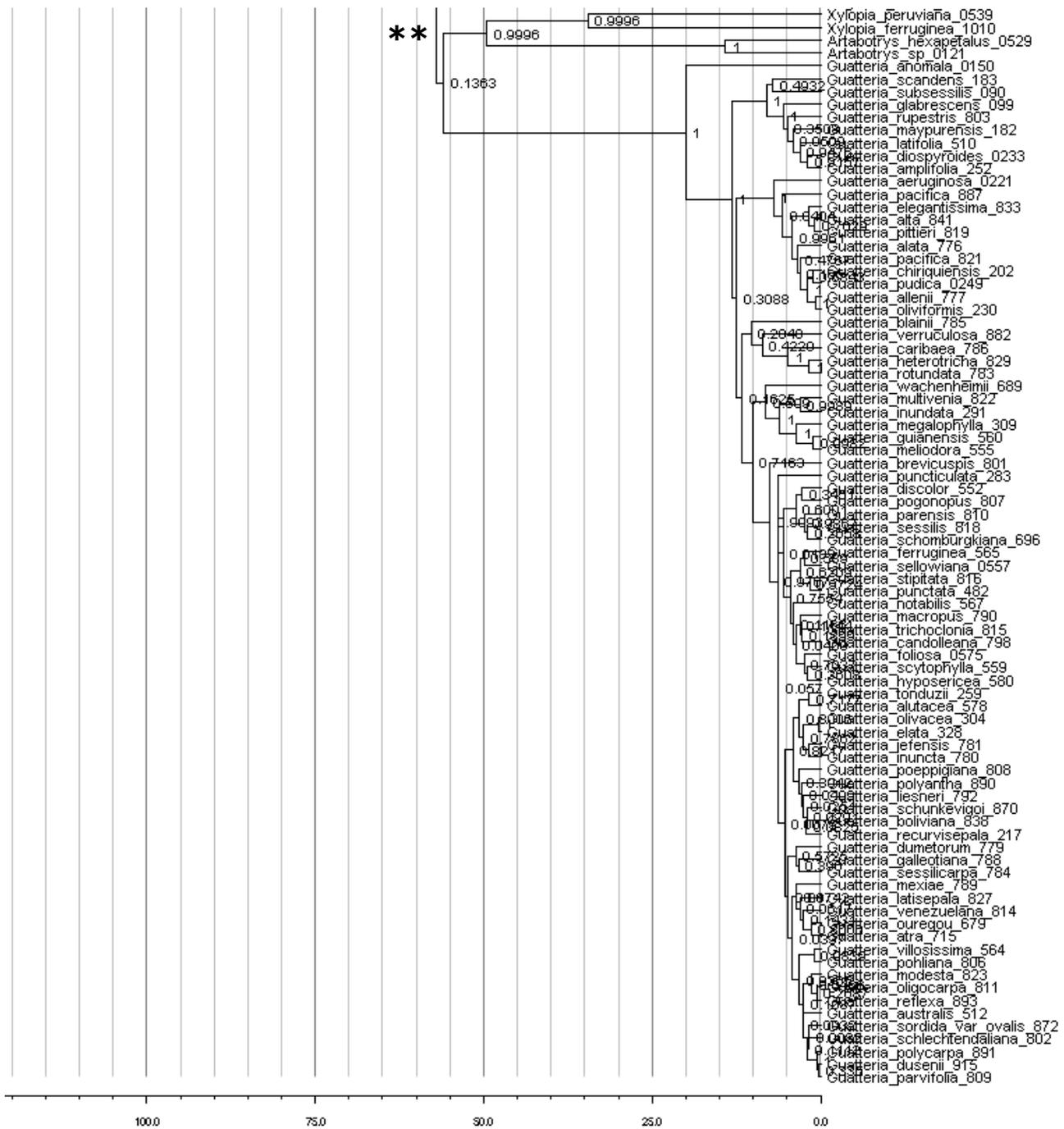
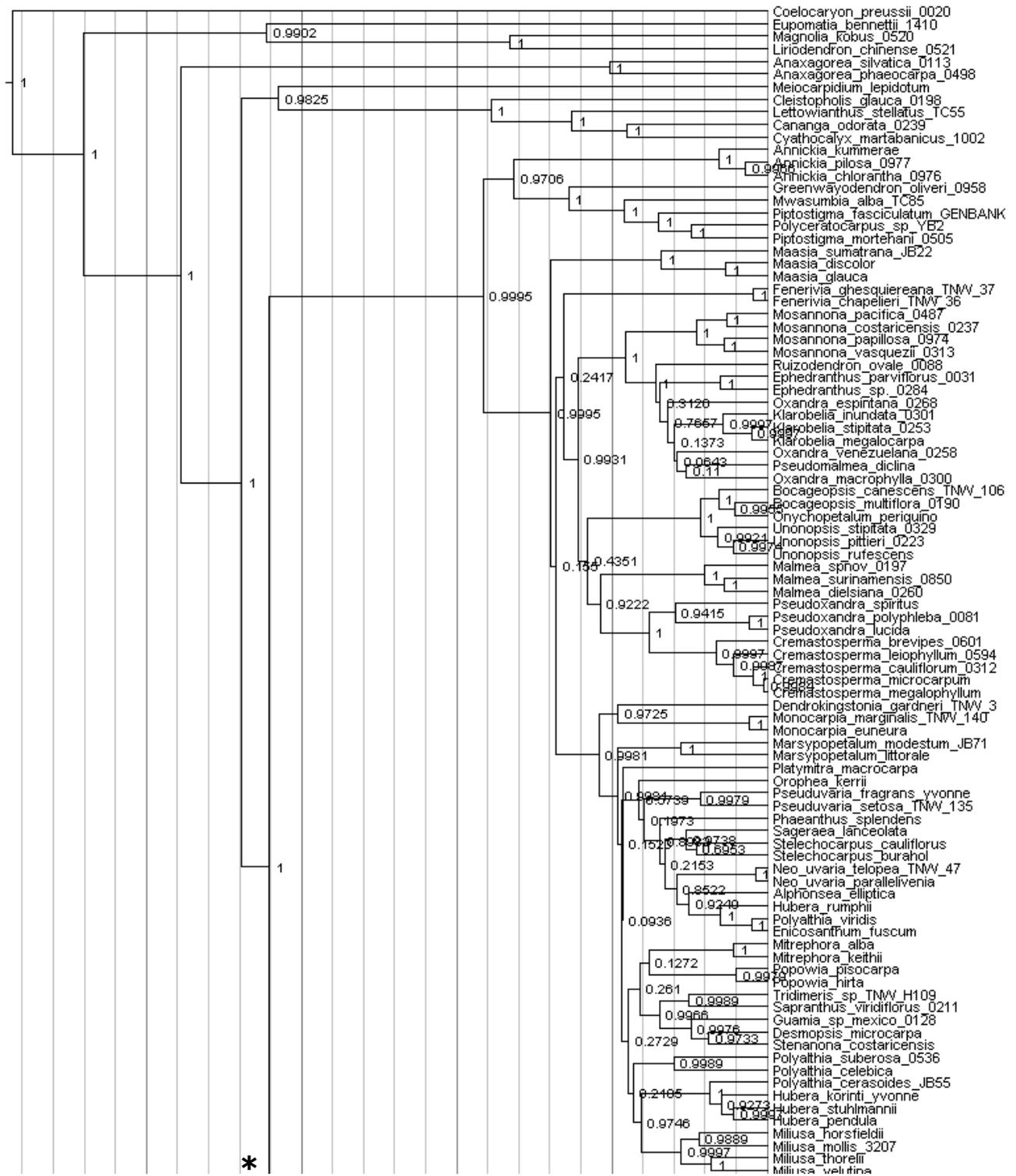


Figure 15: MCC tree for analysis 15 (only *Duguetia* prior included) with posterior values on the branches



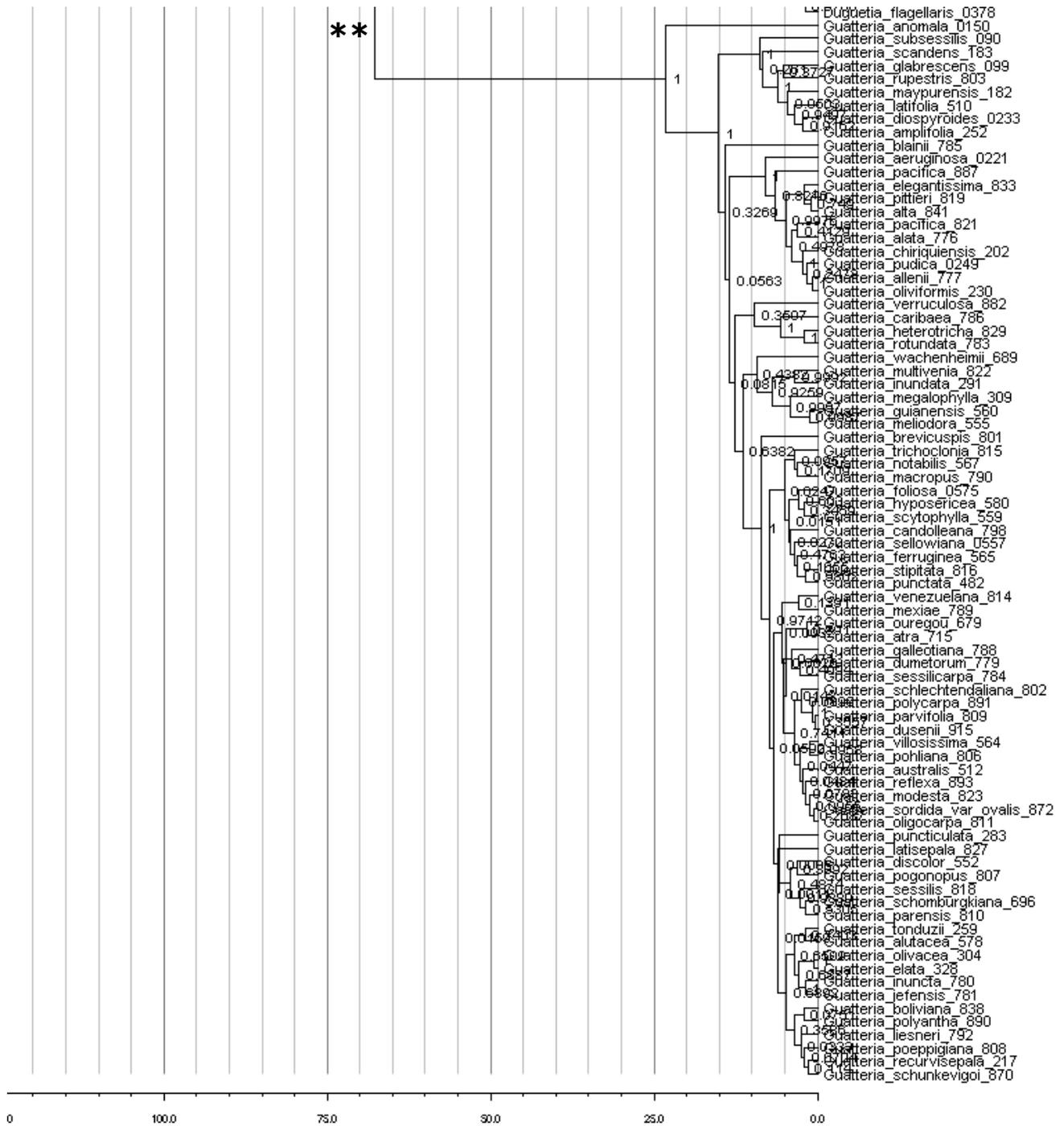
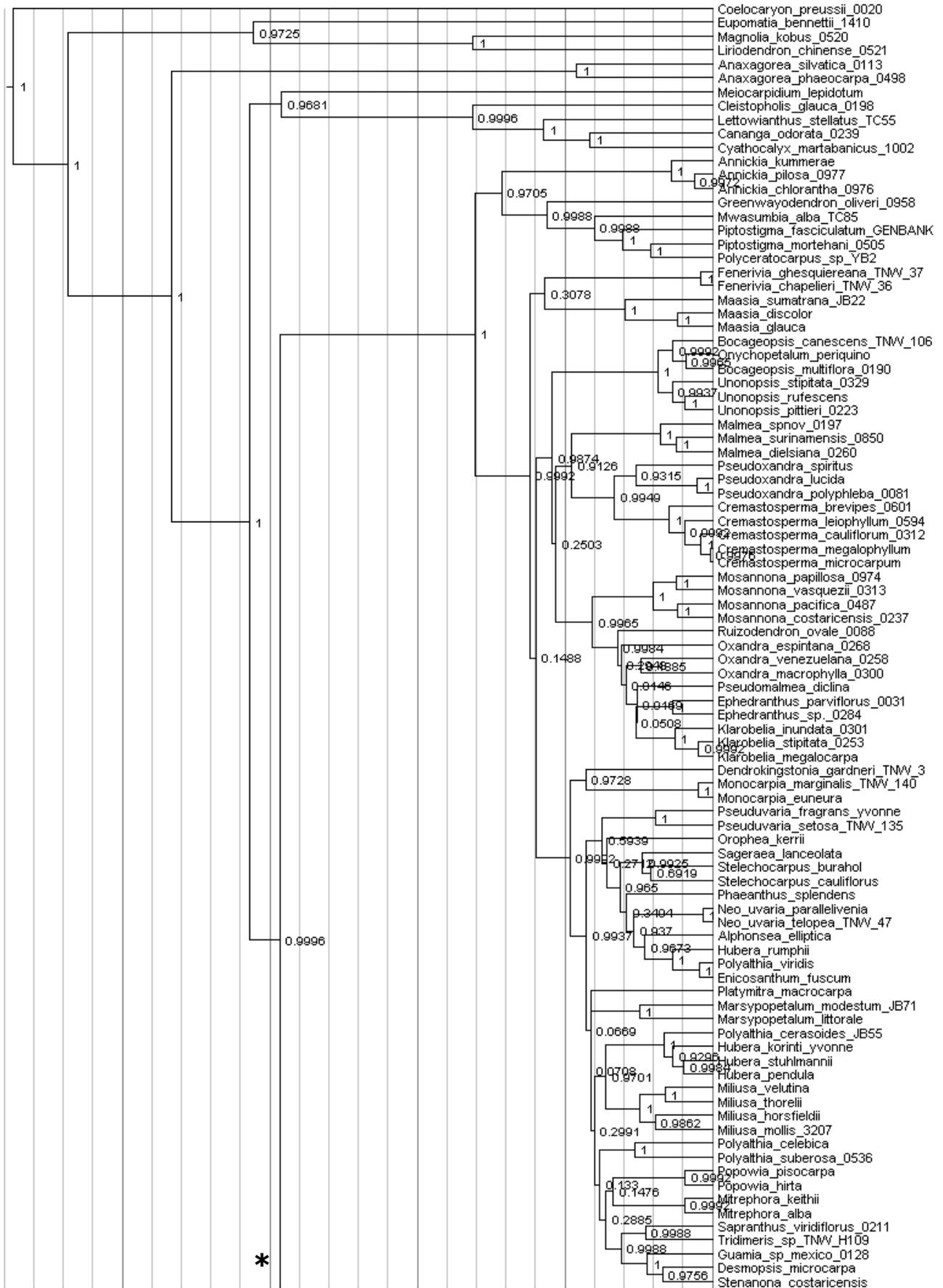
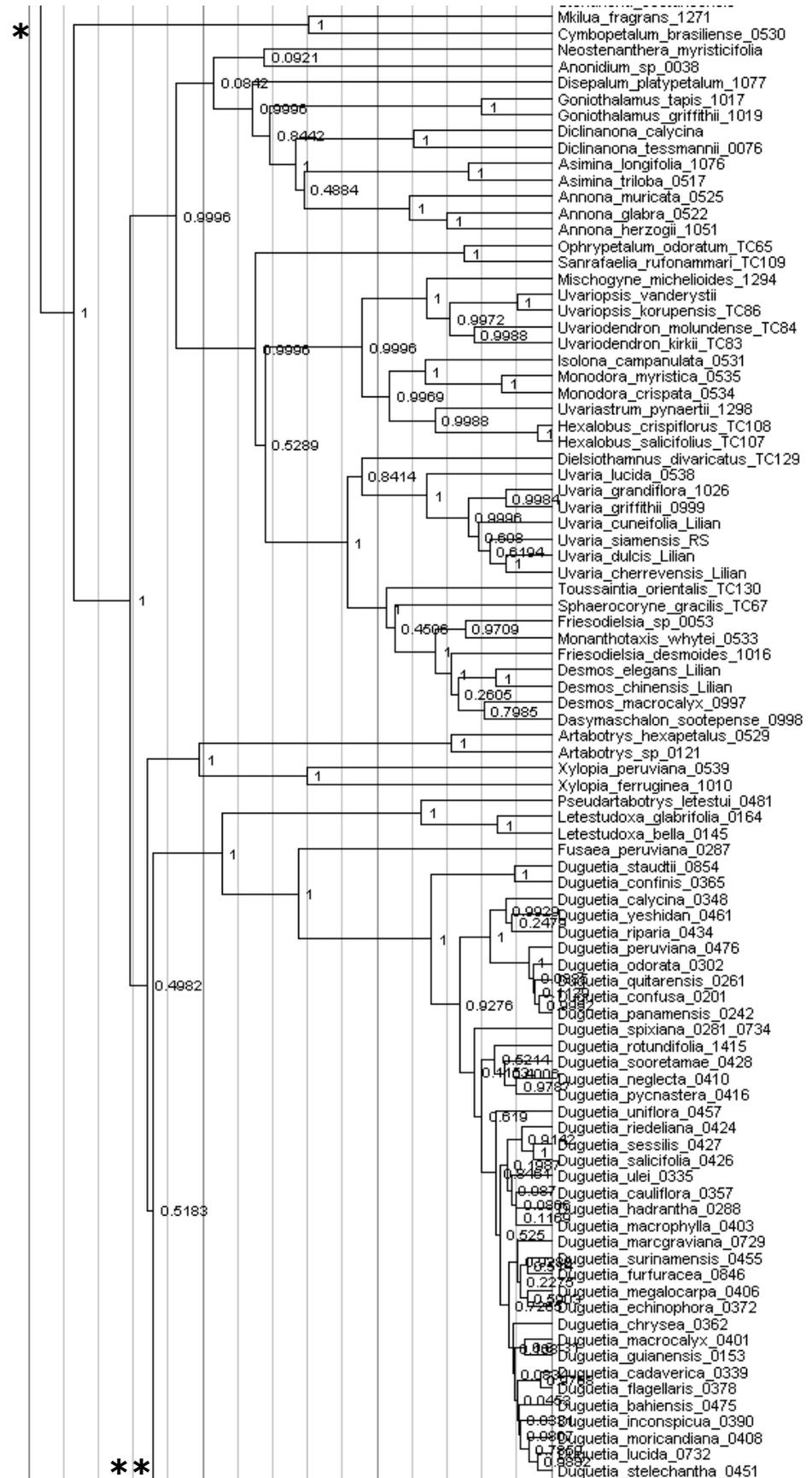


Figure 16: MCC tree for analysis 16 (only *Anonasperrum* prior included) with posterior values on the branches





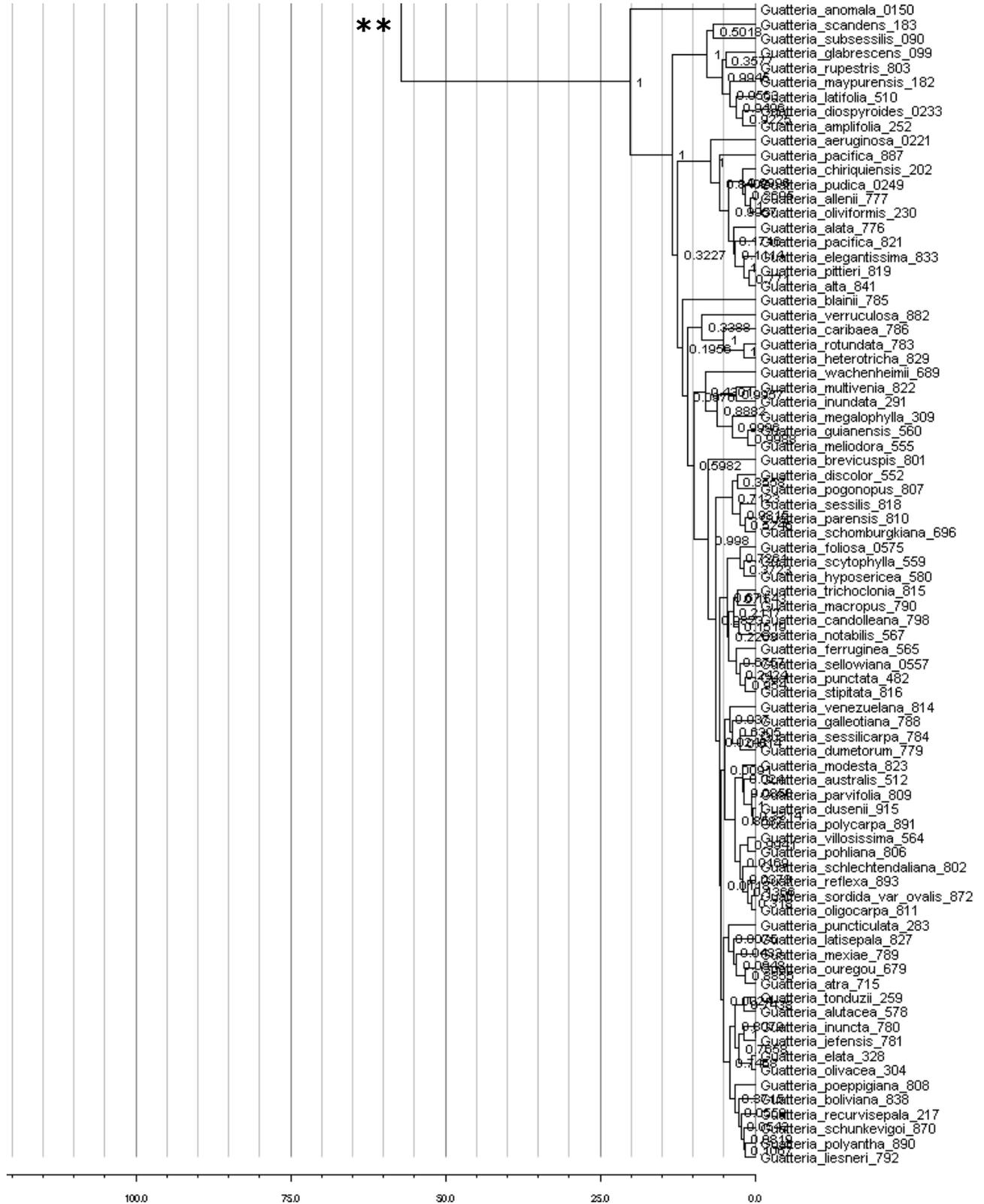


Figure 17: MCC tree for analysis 17 (all priors, reduced LBC sequences) with posterior values on the branches

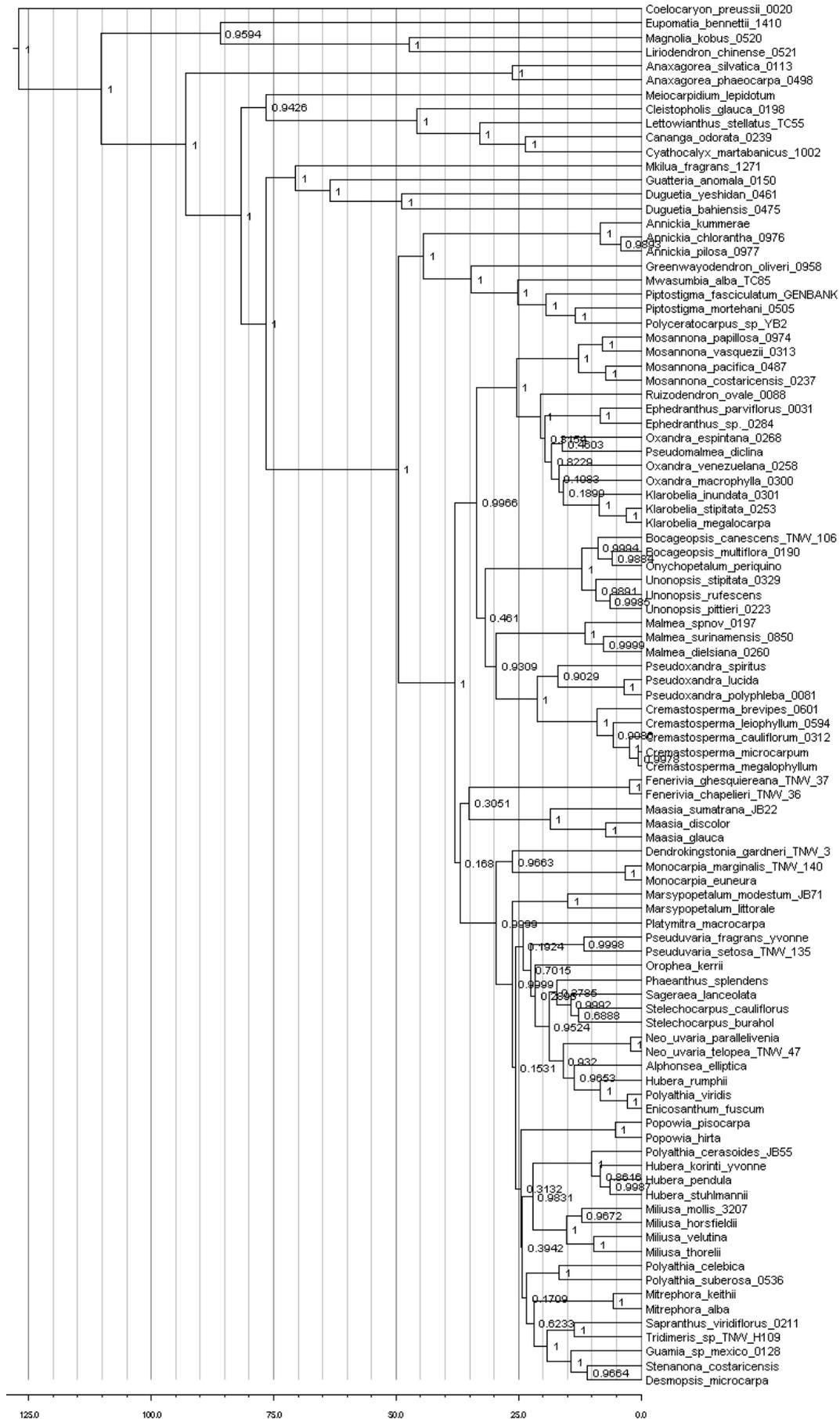


Figure 18: MCC tree for analysis 18 (all priors, SBC sequences reduced) with posterior values on the branches

