PHYLOGENETIC ANALYSES OF COMBINED MORPHOLOGICAL AND MOLECULAR DATA SETS ON THE *APHANOCALYX-BIKINIA-TETRABERLINIA* GROUP (LEGUMINOSAE, CAESALPINIOIDEAE, DETARIEAE *S.L.*)

JAN J. WIERINGA[1*] & G.Y. FANNIE GERVAIS[2]

[1]National Herbarium of The Netherlands, Wageningen University branch, Wageningen University, Biosystematics group, Gen. Foulkesweg 37, 6703 BL Wageningen, The Netherlands
[2]Institut de Recherche en Biologie Végétale, Université de Montréal, 4101 Sherbrooke est, Montréal, Québec, Canada H1X 2B2

Abstract

Three molecular and one morphological data sets of a group of African tree genera (Leguminosae, Detarieae *s.l.*) are combined in simultaneous phylogenetic analyses. A few taxa with long branches show an unstable placement, while several other clades remain intact in all analyses. In the genus *Aphanocalyx* two separate copies of ITS appear to be present. We provide a coding method to circumvent the problems this gives for combining data. Our study strongly supports the monophyly of *Aphanocalyx* subg. *Aphanocalyx*, with subg. *Antherodontus* as its most probable sister group. *Bikinia* and *Tetraberlinia* together form a monophyletic group. *Bikinia* is most likely also a monophyletic group, probably sister to *Tetraberlinia*, but possibly arising from within it. The internal and terminal branches of this group are relatively short, indicating this group may have radiated fairly recently. *Julbernardia* is probably monophyletic, and may be sister to *Bikinia* and *Tetraberlinia*. The position of the monotypic genus *Icuria* is still not clear. In the molecular phylogenies this taxon has a long branch with an unstable position. *Icuria* either has a long separate history or it originates from an anomalous event such as hybridisation. We found strong support for a clade containing *Brachystegia*, *Aphanocalyx*, *Bikinia*, *Icuria*, *Julbernardia* and *Tetraberlinia*.

Introduction

In many woody vegetation types in tropical Africa caesalpinioid legumes play a key role in terms of both species numbers and biomass. A fair number of species are known for their gregarious occurrence. Some species from this group can form mono- or co-dominant stands in wet tropical forests, while others are the major woody component of savanna-woodlands (Gérard, 1960; Pierlot, 1966; Wieringa, 1999; Lubke *et al.*, in press). The vast majority of these gregarious species are contained in the "Macrolobieae" clade of Bruneau *et al.* (2000, 2001). Nineteen genera belong to the "Macrolobieae" clade, with approximately 165 species. Despite its name this clade

---

* author for correspondence: Jan.Wieringa@wur.nl

is not identical to the tribe Macrolobieae Breteler (Breteler, 1995) because the genus *Macrolobium* is excluded. The species-rich and abundant occurrence of this successful group may be the result of several relatively recent radiations (Wieringa, 1999; Mackinder, 2000).

Recent phylogenetic studies of the "Macrolobieae" clade or parts of it (Wieringa, 1999; Bruneau *et al.*, 2000, 2001; Gervais, 2000; Gervais and Bruneau, 2002) show many polytomies, indicating that relationships between the genera are poorly resolved. In some cases not even the monophyly of the genera could be ascertained. A morphological analysis of the former genus *Monopetalanthus* with several other related genera from the "Macrolobieae" clade (principally *Aphanocalyx*, *Julbernardia*, *Michelsonia* and *Tetraberlinia*) resulted in major nomenclatural changes for the species included in this genus (Wieringa, 1999). *Monopetalanthus* was merged with *Aphanocalyx*, while one of its species was moved to *Tetraberlinia* and another 6 to the new genus *Bikinia*. Subsequent studies on chloroplast and nuclear DNA of this clade focussing on the same genera (Gervais, 2000; Gervais and Bruneau, 2002) confirmed most morphological results. In some cases molecular data supported some relationships that were only very weakly supported in morphological data. The best example is *Bikinia* and *Tetraberlinia* being more closely related to each other than they are to *Aphanocalyx* and *Julbernardia*. On the other hand, the analyses of Gervais (2000) and Gervais and Bruneau (2002) cast doubt on some of the morphological relationships. For example *Aphanocalyx heitzii* (Pellegr.) Wieringa, a representative of *Aphanocalyx* subg. *Antherodontus* Wieringa, seems to be more related to *Julbernardia pellegriniana* Troupin than it is to other *Aphanocalyx* species. Another difference is that *Bikinia* appears to be derived from within *Tetraberlinia* in these studies. The last result also contradicts the results of a small AFLP study (Wieringa and Zevenbergen, 1999) that contained species from *Aphanocalyx*, *Bikinia*, *Julbernardia* and *Tetraberlinia* and was completely congruent with the initial morphological results (Wieringa, 1999).

The two data sources differed in an interesting way. Morphological data were quite strong around the generic level; the genera came out as fairly well to strongly supported clades, while there was also a reasonable amout of resolution within genera, but the relationships between genera remained a mystery (Wieringa, 1999). The DNA analyses on the other hand had far more problems achieving resolution at the species level. At generic level the resolution was in some cases not strong enough to prove monophyly of the genera, but this data did show far more resolution between the genera (Gervais and Bruneau, 2002). Similar results, where morphological data has resolution at a lower taxonomic level (between species), while molecular data provided resolution at a higher level (between species groups), have been found by Pennington (1996).

The present study has two main objectives. First we would like to test whether the combination of DNA data with morphological data results in stronger and more resolved phylogenies and what happens to those instances where the separate data sets give conflicting results. Our second objective is this phylogeny itself. We want to evaluate the monophyly of *Aphanocalyx* and its two subgenera, and of *Bikinia*, *Julbernardia* and *Tetraberlinia*. Moreover we are interested in the relationships among these genera.

Materials and methods

Data sets

For the combined analysis we have four data sets available. The first is a rDNA internal transcribed spacer (ITS) data set containing 55 sequences from 50 specimens (cloning revealed that five specimens showed two different sequences) belonging to 42 taxa. Ambiguously aligned parts of the alignment (287 characters) were excluded from the analysis. This resulted in 371 informative characters (Gervais, 2000; Gervais and Bruneau, 2002).

The second data set is based on sequences of the chloroplast *psb*A-*trn*H spacer. It consists of 57 specimens belonging to 50 taxa. It contains 76 informative characters, including one inversion of 15 bp, and 23 insertions, deletions or duplications (Gervais, 2000; Gervais and Bruneau, 2002). The third data set is a large set of chloroplast *trn*L intron sequences of which 42 taxa belonged to the "Macrolobieae" clade in which we are interested (Bruneau *et al.*, 2001).

The fourth data set consists of 104 morphological characters, of which 102 are informative, scored for 45 taxa. This final set is that of Wieringa (1999) augmented by one newly discovered species (*Tetraberlinia apiphila* Wieringa ined.). Further, a few additions based on new material have been made, the most important being that the floral characters of *Icuria dunensis* Wieringa have been added (most of the floral characters, character 47–93, were coded as missing data in Wieringa, 1999).

The insertions /deletions and inversion in the chloroplast DNA data sets have been coded as presence/absence characters. For more details on the accessions, the DNA extraction methods, the ITS and *psb*A-*trn*H sequences and how they were aligned see Gervais (2000) and Gervais and Bruneau (2002); for those on the *trn*L sequences see Bruneau *et al.* (2001). Apart from some binary and multistate characters the morphological data set also contains some characters with an intermediate state. Such characters have been treated as ordered and received $\frac{1}{2}$ weight to ensure that the distance between the two real states remains one (similar to Sosef, 1994). Another 10 continuous characters were coded using Thiele's (1993) gap weighting method. How and why these characters were coded as such is discussed at length by Wieringa (1999). The improved morphological matrix is available on request from the first author.

In order to be able to combine these data sets, all sequences of individual specimens were combined in a single matrix line and the morphological data of that taxon were added. Since we wanted to compare the results of combining data sets with the results of the individual data sets, we needed all characters within a data set to play the same role in the combined data set as they were doing in the individual data set. Hence we had to respect original weighting and ordered/unordered settings. Since some of the morphological characters were coded as $\frac{1}{2}$, $\frac{1}{4}$ or $\frac{1}{8}$ steps (a result of the Gap weighting), all characters received weight 8 and the relevant morphological characters were set back to subsequently weight 4, 2 and 1.

Coding of different copies of ITS

The ITS analyses of Gervais and Bruneau (2002) revealed that in *Aphanocalyx* two different paralogous copies of ITS are present. The gene-tree resulting from this analysis shows a partly duplicated *Aphanocalyx* branch, sister to the other *Aphanocalyx* branch. Since both ITS copies have not been found in all *Aphanocalyx* species, the two branches are not identical. However, the presence of several species in both branches, in one case even based on ITS genes originating from the same specimen, clearly points to a duplication of the ITS gene which took place in the common ancestor of all *Aphanocalyx* species. Only in one specimen were both copies present, in all others only a single copy was found, but in two species both gene types were at least present in different specimens. In our present study we are interested in the phylogenetic relationships in this group and the effect of combining data, not in the gene tree of ITS. The presence of a copy of ITS seriously hinders the possibility of combining ITS data with other data sources in such a way that clade resolutions will be strengthened. While combining data sets, Gervais (2000) and Gervais and Bruneau (2002) used all ITS sequences in combination with the relevant chloroplast gene. In such an analysis, within *Aphanocalyx* the ITS data will be "pushing" for trees having the two duplicated branches of the ITS gene tree, while the other data are (ideally) "pushing" for trees representing the taxon phylogeny. The conflicting data will decrease the resolution in the *Aphanocalyx* clade instead of enhancing it. Since the information content of ITS data is larger than that of the other data sources, the

resulting trees will probably show the duplicate branch pattern, but weakly supported because of the conflicting data present. This is indeed what happened in such analyses conducted by Gervais (2000), Wieringa and Gervais (2001) and Gervais and Bruneau (2002).

To be able to use the ITS data in a constructive manner we coded the two paralogous ITS copies in *Aphanocalyx* as two different genes (Fig. 1). The decision as to which of the two copies a sequence belonged was based on the initial ITS gene-tree. For each of these paralogous copies the ITS outside *Aphanocalyx* can be considered homologous. So, we could add the ITS codes of taxa outside *Aphanocalyx* to both genes in the matrix (Fig. 1C). However, during an analysis this would result in the ITS being weighted double over branches outside *Aphanocalyx*. Ideally the analysis program would be able to back-weigh such characters over these branches, but
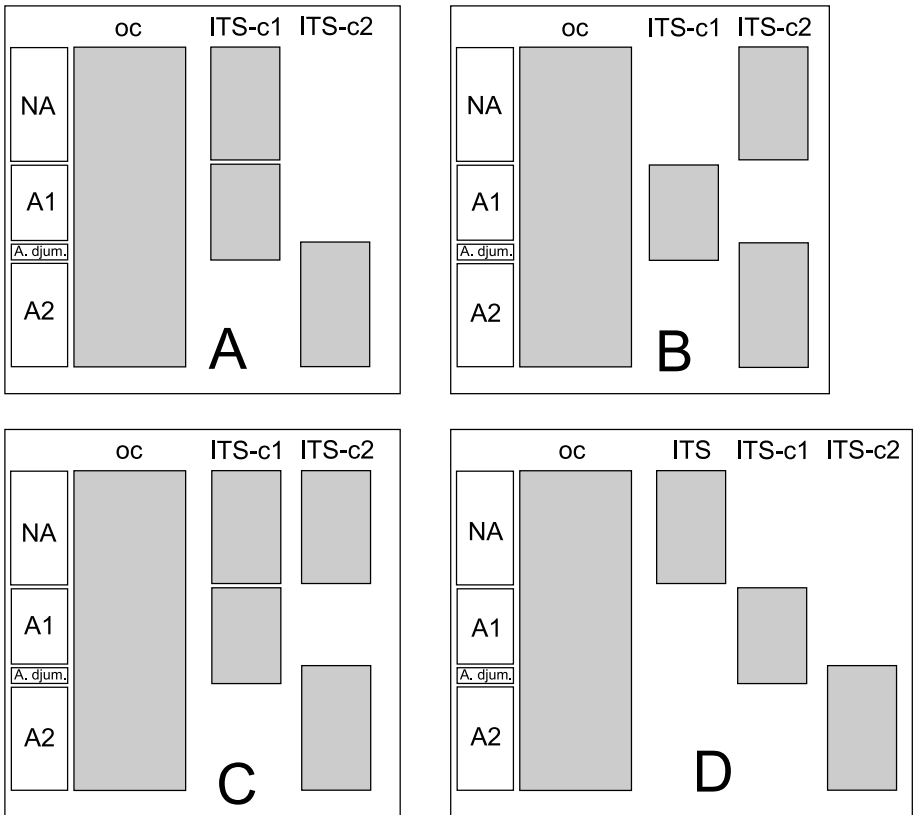
FIG. 1. Diagram of the four different combined data matrices used for the analyses (version A–D). The **columns** represent characters blocks: oc = other characters, ITS = ITS outside *Aphanocalyx* when seen as a separate character, ITS-c1 = paralogous ITS copy A, ITS-c2 = paralogous ITS copy B. The **rows** represent accessions: NA = non-*Aphanocalyx* accessions, A1 = *Aphanocalyx* accessions possessing ITS copy A, A2 = *Aphanocalyx* accesssions possessing ITS copy B, A.djum. = *Aphanocalyx djumaensis*, which accession contained both ITS copies.

present programs are not able to do so. To deal with this problem we have performed three analyses for each data set including ITS data: (1) one in which the non-*Aphanocalyx* ITS sequences were homologised to the first copy (version A, Fig. 1A). This is the copy in which *A. djumaensis* clone A and the *A. heitzii* sample belong in the paper of Gervais and Bruneau (2002); (2) one in which they were homologised to the other copy (version B, Fig. 1B; this is the copy where the two *A. ledermannii* samples and *A. djumaensis* clone B belong); and (3) one where they were added to both copies (version C, Fig. 1C). In this last version all ITS characters received half weight, implying that outside *Aphanocalyx* they counted as full (twice with half weight), but inside *Aphanocalyx* only for half weight. In version A and B the copy where no other sequences were added to was coded as missing for the other accessions. Since in the same *A. djumaensis* sample both copies were found, this accession was the only accession that had both character sets coded in all analyses.

Phylogenetic analyses

The first analysis (see Table 1, analysis 1) was performed with the improved morphological data set. Based on Bruneau *et al.* (2000) and Gervais and Bruneau (2002), *Berlinia* is the least related genus of the included genera, and hence *Berlinia bracteosa* Benth. was chosen as outgroup. The second series of analyses (Table 1, analyses 2A–C) performed used all molecular data sets for specimens of which at least two of the three molecular sequences were available. The third series of analyses (Table 1, analyses 3A–C) only used specimens for which all three molecular data sets were available. Since *Cryptosepalum tetraphyllum* (Hook.f.) Benth. is the least related of the taxa included in the analyses of series 2 and 3 (Bruneau *et al.*, 2000) it was used as outgroup. The fourth series of analyses (Table 1, analyses 4A–C) was run using all four data sets on all specimens where all four datasets were available. Since *Icuria dunensis* has a fairly long branch and it appears in quite different places in the different analyses, a fifth series of analyses (Table 1, analyses 5A–C) was performed similar to series 4, but excluding *Icuria*, in order to explore the influence of long branch-attraction as a phenomenon. Based on the results from the second and third analyses *Microberlinia brazzavillensis* A.Chev. was chosen as the most appropriate outgroup for the analyses series four and five.

Some taxa, such as *Aphanocalyx heitzii, Icuria dunensis, Julbernardia pellegriniana* and *Tetraberlinia bifoliolata* appear to have long terminal branches, possibly causing long branch attraction artefact (Felsenstein, 1978). Maximum likelihood analysis has been claimed to be immune to this problem (Felsenstein, 1978; Huelsenbeck, 1995). We therefore performed two maximum likelihood analyses to test whether the placement of these taxa might be influenced by long branch attraction. One analysis was performed on the ITS data set, in which only the sequences of the first copy were included for *Aphanocalyx*. The second was performed on the chloroplast data sets. These maximum likelihood analyses were performed using a 8-parameter model (GTR + I + Γ) for which the parameters were estimated on some of the most parsimonious trees for the same data set, and subsequently fixed a priori.

To be able to see the effects of the way we coded the two copies of the ITS region in *Aphanocalyx*, we conducted two more analyses. One analysis included the combined morphological and molecular data set. The two different ITS copies inside *Aphanocalyx* were coded as two separate character series, apart from the third series consisting of ITS from outside *Aphanocalyx* (Fig. 1D). Here none of the three sets is considered homologous to one of the other sets. The second analysis used the same combined data set but without all *Aphanocalyx* specimens.

All analyses were performed with PAUP 4.0b8a (Swofford, 2002) on a PowerMac G4. Heuristic searches were performed with 100 random addition sequences replicates and tree bisection-reconnection branch swapping. Branches of zero length were collapsed. All heuristic analyses were followed by a Jackknife analysis (36% deletion, fast stepwise addition) with 10,000 replicates.

Results

The analysis of the improved morphological data set (Table 1, analysis 1) produced two shortest trees (Fig. 2). Compared with the results of Wieringa (1999) there is one striking difference: the new species *Tetraberlinia apiphila* seems to link *Tetraberlinia* to *Bikinia*. Based on its general morphology, the new species seems very closely related to *T. bifoliolata* (sterile samples cannot be told apart), although it shares some characters with *Bikinia* as well. In the morphological tree this species is placed at the apex of the old *Tetraberlinia* phylogeny, indeed closest to *T. bifoliolata*, but subsequently the entire *Bikinia* clade is added as its sister. Although these are the shortest trees, we find this pectinate solution very unlikely. Our doubt is strengthened by the fact that in the Jackknife analysis the clade *Tetraberlinia apiphila* + *T. bifoliolata* receives a value of 56%, which opposes the clade *T. apiphila* + *Bikinia*, which receives no support whatsoever.

Contrary to what was predicted by Wieringa (1999) the coding of floral characters of *Icuria* did not result in *Icuria* becoming separate from the *Bikinia* clade. Since most floral organs are very much reduced or absent in *Icuria*, the also fairly reduced floral elements of *Bikinia* apparently provide a good position for linkage. The flowers are in fact quite distinct, but they differ in characters that were not coded or were unique for *Icuria*.

The series of analyses of combined molecular data sets for taxa with at least two of the three sequences available (Table 1, analyses 2A–C) resolve a clade containing *Icuria*, *Tetraberlinia* and *Bikinia* in analyses 2A and 2C, and in a part of the shortest trees of analysis 2B. In these cases *Icuria* becomes sister to two clones of the same sample of *Bikinia pellegrinii*, where both species have long branches. This clade is resolved as sister to a clade consisting of a monophyletic *Tetraberlinia* which is sister to the rest of *Bikinia* in analyses 2A and 2C, or it becomes part of the *Bikinia* clade which then is sister of *Tetraberlinia* minus *T. bifoliolata* in part of the trees of analysis 2B. The latter species, with a very long branch, becomes sister to this combined clade. All three jackknife analyses are similar in that they strongly support a *Tetraberlinia* (minus *T. bifoliolata*) clade and several internal parts of *Bikinia* are supported. *Aphanocalyx* entirely collapses, although this genus is resolved as monophyletic in analyses 2A and 2C and as such in part of the 1441 trees in analysis 2B. In all three analyses of series 2 there is strong support (81–97% jackknife; JN) for *Brachystegia, Julbernardia, Icuria, Bikinia, Tetraberlinia* and *Aphanocalyx* forming a monophyletic group.

The analyses of combined molecular data sets for taxa with all data available (series 3) are fairly similar; they only differ in the position of *Aphanocalyx heitzii* and *Icuria dunensis*. When one is placed as sister to *Aphanocalyx* subg. *Aphanocalyx*, the other becomes sister to *Julbernardia pellegriniana*, while in some other trees they are exchanged. In analysis 3C *A. heitzii* is sister to subg. *Aphanocalyx* and *Icuria* is sister to *J. pellegriniana* (Fig. 3), in analysis 3A their position is exchanged, while in analysis 3B both topologies are found. The subgenus *Aphanocalyx* has moderate support (57–68% JN) for its monophyly in all three analyses of series 3. In all cases *Bikinia* is also monophyletic with high (80–92%) jackknife support. Usually it is sister to *Tetraberlinia* minus *T. bifoliolata*. *Tetraberlinia bifoliolata* is then resolved as sister to this *Bikinia* + *Tetraberlinia* clade, but this last clade has little support (51–58% JN). In some of the trees of analysis 3A *T. bifoliolata* becomes part of *Tetraberlinia*, rendering both genera monophyletic. In all three analyses of series 3, there is a high (83–91%) jackknife support for a *Tetraberlinia* + *Bikinia* clade. Again there is strong support (85–98% JN) for a clade containing *Brachystegia, Julbernardia, Icuria, Bikinia, Tetraberlinia* and *Aphanocalyx*.

In the combined molecular and morphological analyses (series 4) the three analyses result in three different topologies. In analysis 4A, *Tetraberlinia* and *Bikinia* are both monophyletic and sisters, while *Icuria* is sister to *Aphanocalyx*. In analysis 4B, *Icuria* becomes part of *Bikinia*, although this is contested by a jackknife support of 66% for a monophyletic *Bikinia* in this analysis, and *T. bifoliolata* becomes sister to the

TABLE 1. Number of included taxa, characters, informative characters (inf.c), number of resulting trees, their length and some tree statistics: consistency index (c.i.), retention index (r.i.) and rescaled consistency index (r.c.) for the parsimony analyses performed. Lengths should be divided by 8 to get proper unweighed lengths.

Analysis character set	char.	taxa	A			B			C		
			inf.c. c.i.	trees r.i.	length r.c.	inf.c. c.i.	trees r.i.	length r.c.	inf.c c.i.	trees r.i.	length r.c.
1 morphology	104	45	102 0.25	2 0.63	3853 0.16						
2 2 out of 3 molecular sets	2197	51	457 0.52	235 0.55	16160 0.29	552 0.62	1441 0.61	16176 0.38	724 0.6	59 0.59	13996 0.32
3 all molecular sets available	2197	37	385 0.57	41 0.53	13464 0.30	419 0.67	451 0.58	13456 0.39	616 0.59	15 0.56	11936 0.33
4 all 4 sets available	2301	25	391 0.60	4 0.57	11271 0.35	390 0.70	1 0.63	11244 0.44	528 0.62	2 0.61	9709 0.38
5 all sets available, minus *Icuria*	2301	24	372 0.63	2 0.61	10213 0.39	361 0.72	3 0.67	10128 0.48	482 0.65	2 0.65	8604 0.42

rest of *Tetraberlinia* + *Bikinia* + *Icuria*. In analysis 4C *Icuria* becomes part of *Bikinia*, but *Tetraberlinia* remains monophyletic in this analysis. In all analyses of series 4 *Aphanocalyx* subg. *Aphanocalyx* is monophyletic (with 98–100% JN), in analyses 4B and 4C with *A. heitzii* as its sister, but in analysis 4B *A. heitzii* is sister to *Julbernardia pellegriniana*. The jackknife trees of all analyses of series 4 are similar. In all three there is moderate support (61–66% JN) for a monophyletic *Tetraberlinia*, although the shortest tree in analysis 4B resolved *Tetraberlinia* as paraphyletic because *T. bifoliolata* was placed sister to the clade containing the remainder of *Tetraberlinia* on one hand, and *Bikinia* + *Icuria* on the other. In all three analyses there is moderate to strong support (66–84% JN) for a monophyletic *Bikinia*, even though in some trees *Icuria* was part of *Bikinia* in the shortest tree.

The results from analysis 4D, where none of the three ITS copies is considered homologous, are nearly identical to those of 4C, the topology is only different within *Aphanocalyx* and most jackknife values are lower. The exclusion of *Aphanocalyx* (analysis 4E) does not have much effect on the topology of the other taxa either.

Since quite some instability in previous analyses seems derived from the unstable position of *Icuria* in these analyses, this taxon was deleted from the combined data set (analyses series 5). The phylogenies resulting from the different versions (e.g., Fig. 4) differ only in some internal branches in *Aphanocalyx*. There is a high support (97–99% JN) for a monophyletic *Bikinia*, a fairly low support (52–61% JN) for a monophyletic *Tetraberlinia*, and a high support again for a *Bikinia* + *Tetraberlinia* clade (86–91% JN). Another result is a high support (79–87% JN) for a monophyletic *Aphanocalyx*, where *A. heitzii* becomes sister to the very highly supported (99–100% JN) subg. *Aphanocalyx*. Moderate support (69–86% JN) exists for the *Julbernardia* clade, which in all three versions is placed as sister to the *Bikinia* + *Tetraberlinia* clade, although only once slightly supported (52% JN).

The maximum likelihood analysis of the ITS data resulted in a single most likely tree. *Icuria dunensis, Tetraberlinia bifoliolata* and the two clones of *Bikinia pellegrinii* B13305 together still have long branches, but now *Bikinia* is monophyletic, with *Icuria* sister to *Bikinia* + *Tetraberlinia*, with the exception that *T. bifoliolata* is sister to *Aphanocalyx* + *Bikinia* + *Tetraberlinia*. *Aphanocalyx heitzii* is sister to all other species of *Aphanocalyx* (subg. *Aphanocalyx*). Most of the internal and terminal branches in the *Tetraberlinia* and *Bikinia* clades are very short compared to other genera (*Aphanocalyx*, *Brachystegia* and *Julbernardia*).
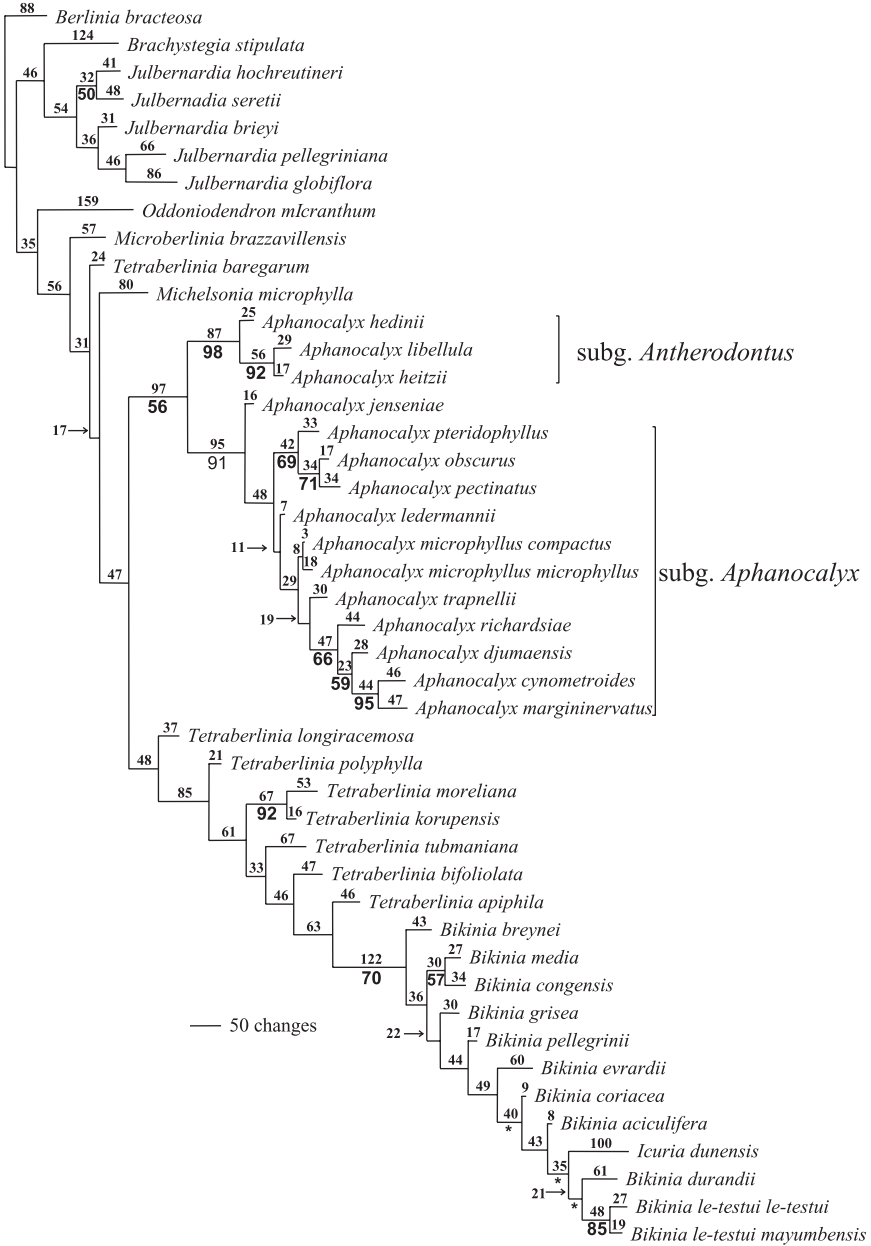
88 *Berlinia bracteosa*
124 *Brachystegia stipulata*
46
32 41 *Julbernardia hochreutineri*
50 48 *Julbernadia seretii*
54
31 *Julbernardia brieyi*
36
46 66 *Julbernardia pellegriniana*
86 *Julbernardia globiflora*
159 *Oddoniodendron mIcranthum*
35
57 *Microberlinia brazzavillensis*
24 *Tetraberlinia baregarum*
56
80 *Michelsonia microphylla*
31
87 25 *Aphanocalyx hedinii*
98 56 29 *Aphanocalyx libellula*
92 17 *Aphanocalyx heitzii*
} subg. *Antherodontus*
97 16 *Aphanocalyx jenseniae*
56
17→ 33 *Aphanocalyx pteridophyllus*
95 42 17 *Aphanocalyx obscurus*
91 69 34 34 *Aphanocalyx pectinatus*
71
48 7 *Aphanocalyx ledermannii*
3 *Aphanocalyx microphyllus compactus*
11→ 8 18 *Aphanocalyx microphyllus microphyllus*
29
30 *Aphanocalyx trapnellii*
19→ 44 *Aphanocalyx richardsiae*
47 28 *Aphanocalyx djumaensis*
47 66 23 46 *Aphanocalyx cynometroides*
59 44 *Aphanocalyx margininervatus*
95 47
subg. *Aphanocalyx*

37 *Tetraberlinia longiracemosa*
48
21 *Tetraberlinia polyphylla*
85
67 53 *Tetraberlinia moreliana*
92 16 *Tetraberlinia korupensis*
61
67 *Tetraberlinia tubmaniana*
33
47 *Tetraberlinia bifoliolata*
46
46 *Tetraberlinia apiphila*
63
43 *Bikinia breynei*
122 27 *Bikinia media*
70 30 34 *Bikinia congensis*
57
36 30 *Bikinia grisea*
17 *Bikinia pellegrinii*
22→
44 60 *Bikinia evrardii*
49 9 *Bikinia coriacea*
40 8 *Bikinia aciculifera*
* 43 100 *Icuria dunensis*
35 61 *Bikinia durandii*
21→ *
48 27 *Bikinia le-testui le-testui*
* 85 19 *Bikinia le-testui mayumbensis*

—— 50 changes

FIG. 2. One of the 2 shortest trees resulting from the parsimony analysis of morphological data (analysis 1). Values above branches indicate branch length, whilst values below branches indicate jackknife support. A * below a branch indicates that the branch collapses in the strict consensus tree.
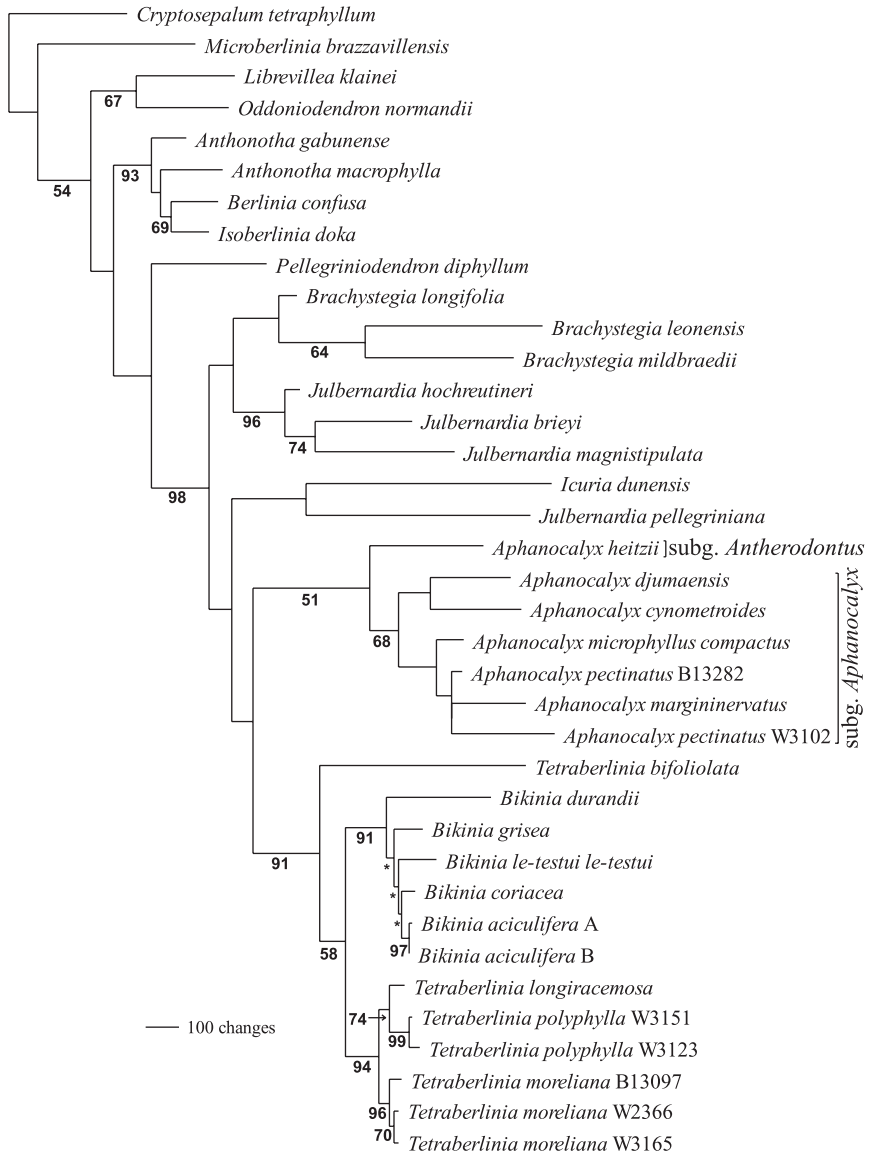
FIG. 3. One of 15 most parsimonious trees resulting from the analysis of all molecular data for all taxa with all 3 sequences available where outside-*Aphanocalyx* ITS is homologised to both paralogous copies in *Aphanocalyx* (analysis 3C). Values below branches indicate jackknife values. A * indicates the branch collapses in the strict consensus tree.

The maximum likelihood analysis of the chloroplast data produced 16 most likely trees. These all show a clade containing all *Bikinia* and *Tetraberlinia* species, although internally intermingled, which is sister to three *Julbernardia* species. A fourth *Julbernardia* (*J. pellegriniana*) is placed together with *Aphanocalyx heitzii* in a clade with all *Brachystegia* species. *Aphanocalyx* subg. *Aphanocalyx* is a monophyletic group, which comes out of a polytomy together with *Icuria*, the *Brachystegia* clade and the *Julbernardia* + *Tetraberlinia* + *Bikinia* clade.

In both ML analyses *Icuria* is placed outside other generic clades, while *Bikinia pellegrinii* B13305 is always placed together with other *Bikinia* species. Other taxa with long branches, like *Tetraberlinia bifoliolata* and *Aphanocalyx heitzii* come out as expected on morphological grounds in one analysis, while they are placed anomalously in the other analysis.

Discussion

Effect of combining data

Our first objective was to evaluate whether combining morphological and molecular data sets in a simultaneous analysis results in more resolved and well-supported phylogenies, than when analysing the datasets separately. The answer to this question is yes. Compared with the results from the old (Wieringa, 1999) and improved morphological data set, all supports for major clades (*Aphanocalyx, Bikinia, Tetraberlinia, Julbernardia*) have become higher in the combined analyses. An initially high support for a combined *Icuria* + *Bikinia* clade in the morphological analysis is refuted, even though such a clade was present in both the morphological and the ITS analyses. As will be discussed below, the placement of *Icuria* in *Bikinia* is highly questionable, so the fact that this initial support now collapses should be seen as an advantage due to the combination of sets. Compared to separate molecular analyses the resolution in the combined analyses is clearly higher, although, due to the not identical taxon sampling, the number of taxa has become lower. Compared to the combined molecular analysis for taxa with all three sequences available (analyses series 3), support for some clades has become stronger (*Aphanocalyx* and subg. *Aphanocalyx*), for some it has become weaker (*Bikinia* + *Tetraberlinia* clade), while for others it depends on the particular analysis (e.g., *Bikinia*). However, while most of the molecular analyses resolve *Tetraberlinia* as paraphyletic without real support (51–58% JN), the combined analysis resolves it monophyletic with weak support (61–66% JN). Again, the paraphyletic situation was due to a questionable placement, in this case of *T. bifoliolata* (see below), so the weak support for a monophyletic *Tetraberlinia* can be seen as a better result.

It is interesting to note that both the molecular analyses and the improved morphological analysis placed *Bikinia* as derived from *Tetraberlinia*, although both in their own way, while the combined analyses result in two monophyletic sister genera. The results of the series of analyses 4 and 5 (Fig. 4) are congruent with the results of the pilot AFLP analysis of Wieringa and Zevenbergen (1999), which included the critical taxon *Tetraberlinia bifoliolata*. Regrettably, there are no molecular data yet available for the new species *Tetraberlinia apiphila*. Since this species combines some morphological characters of *Tetraberlinia* and *Bikinia*, and hence plays a connecting role between these two genera in a morphological phylogenetic analysis, its incorporation may provide further clues to the real phylogenetic relationships in this group.

Long branch attraction

In all separate analyses (morphology, ITS, chloroplast DNA), *Icuria* ends up at the end of a long to very long branch (e.g., Fig. 3). That this monotypic genus jumps around from one genus to another in separate analyses, or even in between shortest trees within one analysis, suggests that its placement suffers from long branch attraction. Indeed, we see that in most cases it is linked to another long branch, such
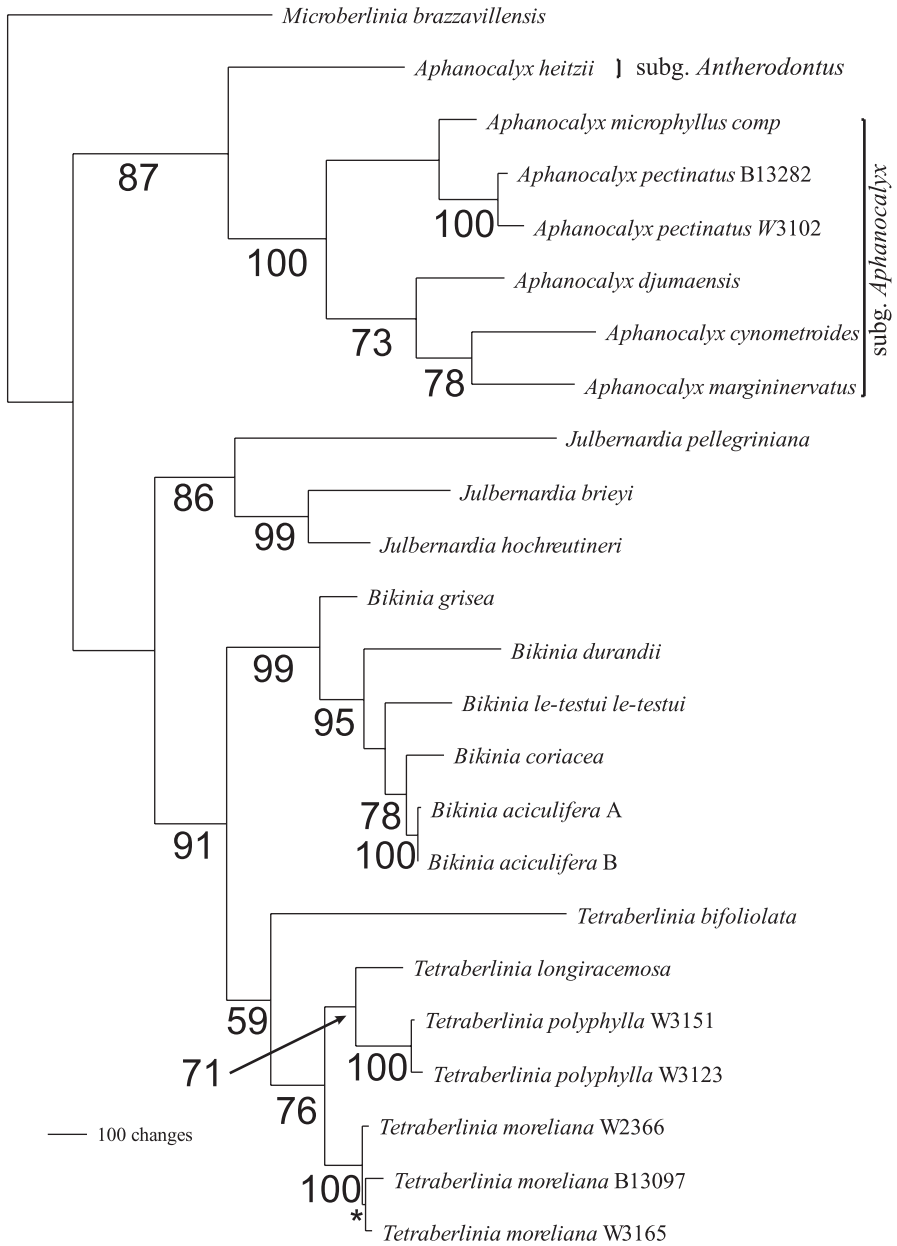
FIG. 4. One of the two shortest trees resulting from the parsimony analysis of the combined molecular and morphological data for all taxa with all 4 data sets available but without *Icuria*, where outside-*Aphanocalyx* ITS is homologised to both of the paralogous ITS copies in *Aphanocalyx* (analysis 5C). Values below branches indicate jackknife values. A * indicates the branch collapses in the strict consensus tree.

191

as the terminal branches of *Julbernardia pellegriniana*, *Aphanocalyx heitzii* and *Bikinia durandii* (F.Halle & Normand) Wieringa, or internal long branches as that to the two nearly identical ITS clones of *Bikinia pellegrinii* (B13305) or that to well supported clades like *Aphanocalyx* subg. *Aphanocalyx*.

In the maximum likelihood analyses, which are less sensitive to long branch attraction, the position of *Icuria* is still unclear. However, in both analyses the most likely trees showed *Icuria* as a separate lineage basal to at least *Bikinia* and *Tetraberlinia* or as part of a polytomy with these and other genera. In analyses of series 5, where *Icuria* was deleted, JN support values for both the *Bikinia* clade and the *Bikinia* + *Tetraberlinia* were considerably higher than they were in analyses of series 4 that included *Icuria*. Apparently the contradicting evidence that existed in analyses of series 4 was mainly caused by the presence of *Icuria*. These results suggest that the placement of *Icuria* within one of these genera is an erroneous case of long branch attraction. Similar, but less severe, long branches with subsequent unstable placement occur in *Tetraberlinia bifoliolata*, *Aphanocalyx heitzii* and *Julbernardia pellegriniana*. It is the placement of these long-branched species which results in these genera becoming para- or polyphyletic in some of the analyses.

Future research should focus on the origin of these long branches, especially whether increased substitution rates could be involved. In the case of *Aphanocalyx heitzii* the long branches may become shorter if the two related species from its subgenus are added. Subdivision of long branches may reduce the errors caused by long branch attraction (e.g. Hendy & Penny, 1989), however, it may also introduce new errors (Poe & Swofford, 1999). The case of *Tetraberlinia bifoliolata* might be more complicated. In morphological terms this species is a typical member of *Tetraberlinia*, and its placement on a relatively long branch apart from other *Tetraberlinia* species is suspect. Adding *T. tubmaniana* J. Léonard and the new species *T. apiphila* might help here, but if this does not help, we should be aware there might be something special about this taxon, like introgression from another genus. Since hybrids will share apomorphologies with different clades, their placement in a parsimony analysis will always be relatively costly, and hence give a long terminal branch of the hybrid taxon. Actually, hybrids should not be placed in a cladogram at all, because their actual phylogenetic pattern is reticulate instead of only branching.

A completely different case is that of *Icuria* and *Julbernardia pellegriniana* (also classified in the monotypic genus *Paraberlinia*). Both these taxa are more or less morphologically distinct, and the long branch might either point to a real long separate history whose origin is difficult to reconstruct, or these taxa find their origin in a hybridisation event.

Duplication of ITS region

A special problem with this data set arose by the duplication of the ITS region in the *Aphanocalyx* clade. In combined analyses where the different copies were treated as the same sequence in separate taxa (e.g., Gervais and Bruneau, 2002), the resulting trees always showed the duplicated tree. This was to be expected since the number of ITS characters is larger than all other characters together, which are subsequently overruled. Because of the conflicting evidence, in such combined analyses the *Aphanocalyx* clade does not have much resolution or support. To circumvent this problem several options are available. A very conservative approach was performed in analysis 4E, where all *Aphanocalyx* species are left out of the analyses. Since nothing changed in the topology outside *Aphanocalyx* compared to analyses where these taxa were included, it seems reasonable to include *Aphanocalyx* taxa in some way. This way we can see where *Aphanocalyx* fits in the phylogeny, test its monophyly and get some ideas about its internal structure. A conservative approach including *Aphanocalyx*, is to code the two copies as separate sequences without considering one of them homologous to the sequences of the other taxa, as was performed in analysis 4D. The topology of this analysis is hardly different from that

of 4C, but the support for the *Aphanocalyx* clade is fairly low in analysis 4D, probably due to the relatively small number of shared characters between taxa inside and outside *Aphanocalyx*. Whether the different ITS copies are made homologous in some way or not, apparently does not make much difference, but in the latter case, the placement and its support of *Aphanocalyx* relative to other genera will be solely based on the morphological and chloroplast data sets. Since the original support for the *Aphanocalyx* clade in the morphological data set was fairly strong, the support was expected to remain more or less intact, and it did.

In analysis 4C of series 4 where both paralogous ITS copies inside *Aphanocalyx* were homologised to the external ITS, the Jackknife values, retention index and consistency index usually lay in between the values resulting from the other two versions where only one of the copies was homologised. This suggests that the double homologisation is a good way to get a general idea of which support is present in the data. Placement of the clade containing the duplicate ITS region will be based on the patterns in both copies, and it was to be expected that support for a given solution would be intermediate between the support present in the individual copies. Regrettably the double homologisation weighs the ITS data inside *Aphanocalyx* for only half weight.

Phylogenetic results

Our second objective was to gain knowledge on the phylogeny of this group. Part of our question is answered: *Aphanocalyx* subg. *Aphanocalyx* is a monophyletic group, and *Aphanocalyx* subg. *Antherodontus* most probably is its sister clade, which would resolve the genus as monophyletic as well. If we consider the anomalous behaviour of *Icuria* as the result of long branch attraction, then we can conclude that *Bikinia* is a monophyletic group as well. It should be noted that in the ITS phylogeny (Gervais, 2000) *Bikinia congensis* Wieringa ended up next to *Anthonotha*; this species was not part of this study because all other molecular data are missing for this taxon. Its placement is so anomalous, given its strong morphological affinity to other *Bikinia* species, that the ITS sequence should be reproduced and other DNA sequences of this taxon should be added.

The morphological studies have so far been unable to support the monophyly of *Bikinia* + *Tetraberlinia*. Again, if *Icuria* is excluded, the evidence for these two genera forming a clade is quite strong (91% JN in analysis 5C). Our next question is whether *Tetraberlinia* is a monophyletic group or not. Our evidence suggests that at least *T. longiracemosa* (A.Chev.) Wieringa, *T. polyphylla* (Harms) J.Leonard ex Voorh. and *T. moreliana* Aubrev. form a monophyletic group, where the first two are more closely related to each other. The combined data sets analyses suggest that they indeed form a clade with the fourth species in the analysis, *T. bifoliolata*, but the support is not very strong. The single sample included of *T. bifoliolata* has a very long branch in the ITS data set, which may affect its placement. Incorporating another sequence of this species and molecular data of *T. apiphila* will probably shed some light on this issue.

The internal and terminal branches of *Bikinia* and *Tetraberlinia* are very short, especially for molecular data. Possibly the species in these genera, especially *Bikinia*, are relatively young. Since the two sister species *B. coriacea* (J.Morel ex Aubrev.) Wieringa and *B. aciculifera* Wieringa occur in two adjacent proposed Pleistocene glacial forest refuges (Maley, 1987; Sosef, 1994), it is possible that they only speciated during one of the last glacial periods. A similar vicariant species pair is also present in *Tetraberlinia* (*T. moreliana* and *T. korupensis* Wieringa, Fig. 1), but this needs further phylogenetic testing before conclusions may be drawn. If both genera radiated recently, new evidence should not only come from more molecular data and added taxa, but also additional morphological characters should be sought (Bateman, 1999).

Julbernardia appears to be a monophyletic genus. In the final analyses (analyses series 5) there is considerable support (69–91% JN) that it also includes *J. pellegriniana*, which species previously has been classified as a separate genus. Still,

the long branch leading to this species is suspect and warrants further study (see above). *Julbernardia* may be sister to the *Bikinia + Tetraberlinia* clade, but so far without real support.

The internal topology of *Aphanocalyx* subg. *Aphanocalyx* was reasonably well resolved in the morphological data, but the chloroplast DNA data (Gervais, 2000) are very inconclusive, while the ITS data only add confusion due to the presence of the two copies. As a result resolution has become lower inside the clade of this subgenus, only the once strong link between *A. cynometroides* Oliv. and *A. margininervatus* J.Leonard, although weakened, seems to hold.

Another significant result from the combined molecular data sets is that *Brachystegia*, *Aphanocalyx*, *Bikinia*, *Icuria*, *Julbernardia* and *Tetraberlinia* form a clade ("babijt" clade). This clade was also present in the analyses of Gervais and Bruneau (2002), but, due to our different coding of the duplicate ITS region, the support in our analysis has increased to 97% JN in analysis 2C and 98% JN in 3C. In our analyses this clade is always sister to *Pellegriniodendron diphyllum* (Harms) J.Leonard, but without jackknife support. Since *Pellegriniodendron* is probably derived from within *Gilbertiodendron* (see Bruneau *et al.*, 2000), these two genera together should be included in future studies. Based on its morphology, the genus *Michelsonia*, which has not been analysed for DNA so far, might be part of this "babijt" clade, which would then become the "bambijt" clade. With over 70 species, this clade contains nearly half of the species of the "Macrolobieae" clade, and includes the genera *Julbernardia* and *Brachystegia*, which both radiated into the African savannas.

Conclusions

Combing morphological and molecular data results in trees with a higher resolution and support compared to separate analyses. Also some new topologies, not present in any of the separate analyses were found. It was difficult to assess the placement of several taxa due to long branch attraction. *Aphanocalyx heitzii, Icuria dunensis, Tetraberlinia bifoliolata* and *Julbernardia pellegriniana,* in particular suffer from this phenomenon. For cases where paralogous copies of genes are present, we introduced a coding which homologises one or both copies with the homologous DNA of taxa outside the clade that includes the copy. The method where both copies are homologised is preferred, although it is even better to evaluate all three options.

Aphanocalyx subg. *Aphanocalyx* is a monophyletic group, the genus as a whole is probably monophyletic as well. There is strong evidence that *Bikinia* is a monophyletic genus: it is most probably is sister to *Tetraberlinia*, but it may arise from within that genus. Together *Bikinia* and *Tetraberlinia* form a monophyletic group, possibly sister to a monophyletic *Julbernardia*. *Icuria* is probably a separate lineage, together with the genera *Aphanocalyx, Bikinia, Brachystegia, Julbernardia* and *Tetraberlinia* it forms the "babijt" clade.

Acknowledgements

Literature cited

Bateman, R.M. (1999). Integrating molecular and morphological evidence of evolutionary radiations. In: P.M. Hollingsworth, R.M. Bateman and R.J. Gornall (editors). Molecular systematics and plant evolution, pp. 432–471. Taylor & Francis, London.

Breteler, F.J. (1995). The boundary between Amherstieae and Detarieae (Caesalpinioideae). In: M.D. Crisp and J.J. Doyle (editors). Advances in legume systematics, part 7, Phylogeny, pp. 53–61. Royal Botanic Gardens, Kew.

Bruneau, A., Breteler, F.J., Wieringa, J.J, Gervais, G.Y.F. and Forest, F. (2000). Phylogenetic relationships in tribes Macrolobieae and Detarieae as inferred from chloroplast *trn*L intron sequences. In: P.S. Herendeen and A. Bruneau (editors). Advances in legume systematics 9, pp. 121–149. Royal Botanic Gardens, Kew.

Bruneau, A., Forest, F., Herendeen, P.S., Klitgaard, B.B. and Lewis, G.P. (2001). Phylogenetic relationships in the Caesalpinioideae (Leguminosae) as inferred from chloroplast *trn*L intron sequences. *Systematic Botany* 26: 487–514.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

Gérard, Ph. (1960). Étude écologique de la forêt dense à *Gilbertiodendron dewevrei* dans la région de l'Uele. *Publications de l'institut national pour l'étude agronomique du Congo Belge, série scientifique 87.*

Gervais, G.Y.F. (2000). Phylogénie moléculaire de *Monopetalanthus* Harms (Fabaceae: Caesalpinioideae) et des genres affiliés. M.Sc. thesis, Université de Montréal, Canada.

Gervais, G.Y.F. and Bruneau, A. (2002). Phylogenetic analysis of a polyphyletic African genus of Caesalpinioideae (Leguminosae): *Monopetalanthus* Harms. *Plant Systematics and Evolution* 235 (1–4): 19–34.

Hendy, M.D. and Penny, D. (1989). A framework for the quantitative study of evolutionary trees. *Systematic Zoology* 38: 297–309.

Huelsenbeck, J.P. (1995). Performance of phylogenetic methods in simulation. *Systematic Biology* 44: 17–48.

Lubke, R.A., Wieringa, J.J., Avis, A.M., Dold, A., Brink, E. and Bandeira, S. (in press). An undescribed unique coastal forest from north-eastern Mozambique. *Trends in Ecology and Evolution.*

Mackinder, B.A. (2000). The monophyly and taxonomic position of *Berlinia* Hook.f. (Leguminosae-Caesalpinioideae): a preliminary study. In: P.S. Herendeen and A. Bruneau (editors). Advances in legume systematics 9, pp. 151–163. Royal Botanic Gardens, Kew.

Maley, J. (1987). Fragmentation de la forêt dense humide africaine et extensions des biotopes montagnards au Quaternaire récent: nouvelles données polliniques et chronologiques. Implications paléoclimatiques et biogéographiques. In: J.A. Coetzee (editor). Palaeoecology of Africa 18, pp. 307–334. A.A. Balkema, Rotterdam.

Pennington, R.T. (1996). Molecular and morphological data provide phylogenetic resolution at different hierarchical levels in *Andira*. *Systematic Biology* 45: 496–515.

Pierlot, R. (1966). Structure et composition de forêts denses d'Afrique Centrale, spécialement celles du Kivu. *Memoires de l'academie royale des sciences d'outre-mer. Class des sciences naturelles et médicales,* N.S. 16(4): 1–367.

Poe, S. and Swofford, D.L. (1999). Taxon sampling revisited. *Nature* 398: 299–300.

Sosef, M.S.M. (1994). Refuge Begonias: Taxonomy, phylogeny and historical biogeography of *Begonia* sect. *Loasibegonia* and sect. *Scutobegonia* in relation to glacial rain forest refuges in Africa. *Wageningen Agricultural University Papers* 94(1): I–XV, 1–320.

Swofford, D.L. (2002). PAUP*: Phylogenetic analysis using parsimony, version 4.0b8a. Sinauer Associates, Sunderland, Massachusetts, USA.

Thiele, K. (1993). The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics* 9: 275–304.

Wieringa, J.J. (1999). *Monopetalanthus* exit. A systematic study of *Aphanocalyx, Bikinia, Icuria, Michelsonia* and *Tetraberlinia* (Leguminosae, Caesalpinioideae). *Wageningen Agricultural University Papers* 99(4): I–XVI, 1–320.

Wieringa, J.J. and Zevenbergen, M.J. (1999). 7 AFLP analyses. *Wageningen Agricultural University Papers* 99(4): 95–98.

Wieringa, J.J. and Gervais, G.Y.F. (2001). The *Aphanocalyx-Tetraberlinia-Bikinia* group: analysis based on improved and combined morphological and DNA data. Program and abstracts. Legumes Down Under 2001. Canberra, Australia. p. 96.