CHAPTER 2

MOLECULAR PHYLOGENY OF GENUS ARTABOTRYS (ANNONACEAE)

2.1 INTRODUCTION

Annonaceae is classified within the order Magnoliales (APG II, 2003), in which it is highly supported as sister group to Eupomatiaceae, based on cladistic analysis of morphological and molecular data (Sauquet *et al*., 2003), with *Anaxagorea* resolved as sister to the rest of the family (Doyle *et. al*., 2000; Sauquet *et al*., 2003). Results of phylogenetic research on the Annonaceae based on *rbc*L and *trn*L- F sequence data (Mols, 2004; Richardson *et al*., 2004) divided the rest of Annonaceae between a small clade including *Cananga* and *Cleistopholis,* and a large clade including the majority of species of the family. This large clade is further divided into two sub- clades that have been termed the Long Branch Clade (LBC) and Short Branch Clade (SBC). (Richardson *et al*., 2004). These are pan-tropically distributed clades comprising the majority of genera and species in Annonaceae. The LBC and the SBC consist of approximately the same number of genera, but the number of species is different. Within the SBC one can find c. 50 genera comprising 700 species in total. The LBC contains c. 50 genera and includes 7 species-rich genera of the family (*Annona,* 175 species; *Artabotrys*, 100 species; *Duguetia*, 95 species; *Goniothalamus*, 120 species; *Guatteria,* 280 species; *Uvaria,* 150 species and *Xylopia*, 145 species). The remaining c. 500 species of this clade are divided in c. 40 genera. The relationships between the genera within SCB have been

studied. Mols (2004) and Pirie *et al*. (2006) used molecular data (*trn*L-F, *rbc*L and *mat*K sequences) to establish relationships within and between the selected genera distributed in Southeast Asia (*Miliusa*) and Neotropics (*Cremastosperma*, *Guatteria, Dugetia* and *Mosanona*). Subsequently, Erkens (2007) have further investigated relationships within the LBC, using the *ndh*F in combination with the *trn*L-F region to produce the trees which completely resolved all unresolved nodes found by Richardson *et al*. (2004) apart from the *Xylopia-Artabotrys* clade, the position of which was still weakly supported. This means that the phylogenetic position of this clade should be regarded as ambiguous. Increasing the sequence data of taxa is needed.

This study focuses on the genus *Artabotrys* one of the species-rich genera within the LBC which is distributed in tropical Africa and Asia. This genus is easily recognized by the flattened hooks at the base of peduncle and its climbing habit. There are three sepals which are valvate, free or united at the base. The six petals are valvate, in two whorls of three, usually subequal, spoon-shaped, connivent around the reproductive organs. *Artabotrys* will be the first genus with an African/ Asian distribution to be studied in more detail.

The aim of this study is to clarify the phylogenetic relationships within the genus and to investigate the biogeography of the genus in a phylogenetic framework.

Two research questions will be addressed:

(1) Is the genus *Artabotrys* monophyletic?

(2) What are the phylogenetic relationships among African and Asian species?

To answer these questions, parsimony analyses were conducted from DNA sequence data. The plastids DNA; *trn*L-F spacer, *psb*A- *trn*H and *ndh*F were expected to be useful in reconstructing phylogenetic relationships within the genus *Artabotrys* based on previous studies in Family Annonaceae (Bygrave, 2000; Chatrou *et al*., 2002; Richardson *et al*. 2004; Pirie *et al*. 2006; Erken(2007). The low-copy nuclear gene *APETALA3* (*AP*3) were chosen to improve the robustness and resolution of phylogenetic relationships at low taxonomic level, in view of the applicability of *AP3* in *Impatiens* (Balsaminaceae) Janssens *et al*, (2007).

2.2 MATERIAL AND METHODS

Plant samples

We sampled a total number of 40 accessions, covering 21 species of *Artabotrys* and 12 outgroups were sampled. The sampling was based on geographical distribution in the Palaeotropics, using herbarium specimens and silica- dried leaves sampled from Asia and Africa. Outgroups were selected from genus *Asimina, Duguetia, Fusaea*, *Xylopia*, and *Goniothalamus*, Richardson et al. (2004) found a strongly supported sister group relationship between *Artabotrys* and *Xylopia* using *rbc*L and *trn*L-F sequence data. The remaining outgroup taxa are in the LBC, as are *Artabotrys* and *Xylopia*. For a list of all taxa are showed in Table 2.1.

Table 2.1 List of species, collector and collection number, herbarium or silica- gel materials (H/ S), and origin.

Species	Collector and collection number	H/S material	Origin
A. aurantiacus Engl.	J.de Bruijn 1835 (WAG)	H	Africa
A. boonei De Wild.	(WAG)	H	Africa
A. brachypetalus Benth.	Calane da Silva, P.Jansen, Marimee, Amanhica 137 (WAG)	H	Africa
A. burmanicus A.DC.	U. Thongpairoj 210 (CMU)	S	Asia
A. harmandii Finet et Gagnep	U. Thongpairoj 201 (CMU)	S	Asia
*A. harmandii*_FER	U. Thongpairoj 222 (CMU)	S	Asia
*A. harmandii*_WY	U. Thongpairoj 216 (CMU)	S	Asia
A. hexapetalus(L.f.) Bhandari	U. Thongpairoj 233 (CMU)	S	Asia
A. insignis Engl.& Diels	(WAG)	H	Africa
A. madagascariensis Miq.	G. Mc Pherson 18836(WAG)	H	Africa
A.modestus Diels *ssp*. *macranthus* Verdc.	S.A. Robertson 4090 (WAG)	H	Africa

Table 2.1 List of species, collector number, herbarium or silica-gel materials (H/S), and origin (continued).

Species	Collector and collection number	H/ S material	Origin
A. monteiroae Oliv.	R.E.S.Tanner 4384 (WAG)	H	Africa
A. multiflorus C.E.C. Fischer	U. Thongpairoj 230 (CMU)	S	Asia
A. rosea Boerl.	(L)	H	Asia
A. sauveolens Blume	U. Thongpairoj 214 (CMU)	S	Asia
*A. spinosus*_DFH	U. Thongpairoj 227 (CMU)	S	Asia
A. thomsonii Oliv.	G. Mc Pherson 16305 (WAG)	H	Africa
A. venustus King	(L)	H	Asia
A. velutinus Scot- Elliot	J.de Koning 5939 (WAG)	H	Africa
A. sp_KB	U. Thongpairoj 211 (CMU)	S	Asia
A. sp_NHN	U. Thongpairoj 220 (CMU)	S	Asia
A. sp_PK3	U. Thongpairoj 206 (CMU)	S	Asia
A. sp_UT226	U. Thongpairoj 226 (CMU)	S	Asia
A. sp_WP2	U. Thongpairoj 209 (CMU)	S	Asia
A. sp_WP1	U. Thongpairoj 208 (CMU)	S	Asia

DNA extractions, PCR amplification and DNA sequencing

Total genomic DNA was extracted from herbarium material or silica-dried leaves (10-50 mg) using 2x CTAB method described by Doyle and Doyle (1987), following the DNA extraction protocol of the Wageningen Biosystematics Group (Vrielink *et al.,* 2005). Concentration and quality of the extracted DNA was measured by using the Nanodrop ND-1000 spectrophotometer. Leaf material was taken individually from each species. The *trn*L-F region was amplified using the primer combinations c-d and e-f (Taberlet *et al.*, 1991). The thermal cycling protocol: 1 min. denaturation at 94 ˚C, 1 min. annealing at 53 ˚C, an extension of 2 min. at 72 ˚C, following by one cycle of 7 min. at 72 ˚C. The *psb*A-*trn*H intergenic spacer was amplified and sequenced using primer *psb*A and *trn*H (GUG) (Hamilton, 1999). The *ndh*F gene was amplified and sequenced in two overlapping pieces using primers 1, 972 and 2110R (Olmstead& Sweere, 1994) in combination 1/ 972R and 972/ 2110R. The *AP*3 gene was amplified with the following two primers: *AP3*F and *AP3*R. All primers were designed by Thomas Couvreur (unpublished) at the Wageningen University. The used primer sequences are listed in table 2.2.

All PCR products were purified using a QIAquick PCR Purification Kit (QIAGEN). The cleaned fragments were sequenced using a fluorescent dye- labeled sequencing reaction (DYEnamic tm ET Terminator Cycle Sequencing Kit, Amersham Biosciences), and the high throughput ABI sequencing facilities at Greenomics, Wageningen University, The Netherlands.

Phylogenetic analyses

Data matrices and alignment: sequences were assembled and edited using the Staden Package version 1.5.3 (Staden *et al*., 2003) and aligned using the ClustalW multiple alignment option in the BioEdit sequence alignment editor 7. 0. 4. 1 (Hall, 1999), subsequently adjusted manually and added to the existing *trn*L-F, *ndh*F, *psb*A-*trn*H and *AP*3 data set. Gaps in the alignments were coded as missing data. Doubtful regions were excluded. The combined partitions were analyzed.

Maximum parsimony (MP) analysis: Data were analyzed using the parsimony algorithm of the software package PAUP* 4.0b10 (Swofford, 2001). Heuristic searches were conducted with tree-bisection reconnection (TBR) branch swapping, 10,000 random addition replicates, saving 200 trees per addition replicate. Characters were equally weighted and character states were specified to be unordered. Bootstrap analysis was carried out to calculate the relative support for individual clades found in the parsimony analysis (Felsenstein, 1985).

Table 2.2 Primer sequences used for amplification and sequencing of the markers *trn*L-F, *ndh*F, *psb*A-*trn*H and *AP*3.

Marker	Primer	Primer sequence
*trn*L- F	c	5'-CGAAATCGGTAGACGCTACG-3'
	d	5'-GGGGATAGAGGGACTTGAAC-3'
	e	5'-GGTTCAAGTCCCTCTATCCC-3'
	f	5'-ATTTGAACTGGTGACACGAG-3'
*ndh*F	1	5'-ATGGAACAKACATATSAATATGCGTGG 3'
	972	5' –GTCTCAATTGGGTTATATGATG 3'
	972R	5 '-CATAATATAACCCAATTGAGAC-3'
	2110R	5'-CCCCCTAYATATTTGATACCTTCTCC-3'
*psb*A-*trn*H	*psb*A	5'- GTTATGCATGAACGTAATGCTC -3'
	*trn*H	5'- CGCGCATGGTGGATTCACAATCC -3'
AP3	AP3F	5' CTGCGCGGTCTTGAGC 3'
	AP3R	5' TTGCGTGTGCTTCATGCAGG 3'

For each of the 500 bootstrap replicates, a heuristic search was conducted with identical setting as in the original heuristic analysis. Bootstrap percentages were interpreted following Richardson *et al*. (2004); 50-74% represents weak support, 75-84% moderate support and 85-100% strong support.

Selecting the best fitting DNA substitution model: ModelTest 2.2 (Posada and Crandall, 1998) was used to select the substitution model best fitting each sequence data partition.

Bayesian analysis: The combined datasets were also analyzed using Bayesian inference, as implemented in MrBayes version 2.2 (Huelsenbeck, 2000). The data was partitioned according to the separate markers used and both rates and substitution models were allowed to vary across the partitions. Prior values for the DNA substitution models were applied to each partition. Prior probabilities for all trees were equal. *Asimina angustifolia* was chosen as the single outgroup taxon permitted by MrBayes. MCMC analyses were run for 5,000,000 generations with four simultaneous MCMC chains (one cold, three heated) to calculate posterior probabilities (PP), saving one tree per 100 generations. The burn-in values were determined from likelihood values. PAUP* 4.0b10a was used to calculate a 50% majority-rule consensus trees and report the posterior probabilities (PP) for each clade.

2.3 RESULTS

For all accessions the *trn*L-F, *psb*A-*trn*H and *ndh*F plastid DNA regions and *AP*3 nuclear gene were sequenced, with the exception of some accessions for which no or only partial sequences of certain regions were obtained. The final alignment has a total length of 4591 characters including, 1062 characters of *trn*L-F, 2129 characters of *ndh*F, 558 characters of *psb*A-*trn*H and 841 characters of *AP*3 respectively.

Maximum Parsimony analysis (MP)

Analyses of the separate data sets results in most parsimonious trees that were congruent but always less resolved than the most parsimonious trees from the combined analysis (data not shown). The combined molecular data set consisted of a total of 2711 characters. Of these characters, 292 were variable, 390 of which were potentially phylogenetically informative. The MP analysis yielded 792 trees with a tree length of 955, with a consistency index (CI) of 0.81 and a retention index (RI) of 0.84. Statistics of the four markers separately are given in Table 2.3. Support estimated using a 500 replicate bootstrap analysis is represented in Fig.2.1. The genus *Artabotrys* is a strongly supported monophyletic group (BS 100 %). Two sister clades within the genus are strongly supported. The first clade (Clade A) consists of *A. velutinus*, *A. madagascariensis, A. modestus* and *A. insignis* (BS 98%) and the African species are sister to all other species within the genus (BS 100%). The second major subclade (Clade B) consists of all Asian species of *Artabotrys* (BS 100%). Overall the topology of Asian Clade is strongly supported except one node. The node subtending the B4, B3 and B5 clade is only weakly supported (BS 36%). The internal nodes of B4 clade also received weak bootstrap support (BS < 50%).

Best fitting models for each marker as estimated by ModelTest are presented in Table 2.3.

Bayesian analysis (BA)

After 5,000,000 generations, standard deviation of the split frequencies in the simultaneous two MCMC run in MrBayes was 0.002636, which means that they had

converged on the same tree topologies. Of the 27, 6004 trees obtained from BA the burn-in estimated, resulting in 1,000 trees being discarded. The remaining trees yielded a consensus tree with a similar to the bootstrap consensus tree (50% majority rule). Two clades with bootstrap support 100% did not receive high posterior probability values (PP): internal clade of D (PP= 0.69) and an internal of African clade (PP= 0.64).

Table 2.3 Details of maximum parsimony search.

Markers	Characters	Parsimony informative characters including outgroups	Parsimony informative characters excluding outgroups	Model
*trn*L-F	856	112 (13.08%)	8 (0.94%)	HKY+G
*psb*A-*trn*H	441	70 (15.87%)	17 (3.85%)	HKY+G
*ndh*F	676	148 (21.89%)	20 (2.95%)	GTR+G
*AP*3	737	59 (8.00%)	59 (8.00%)	HKY+G

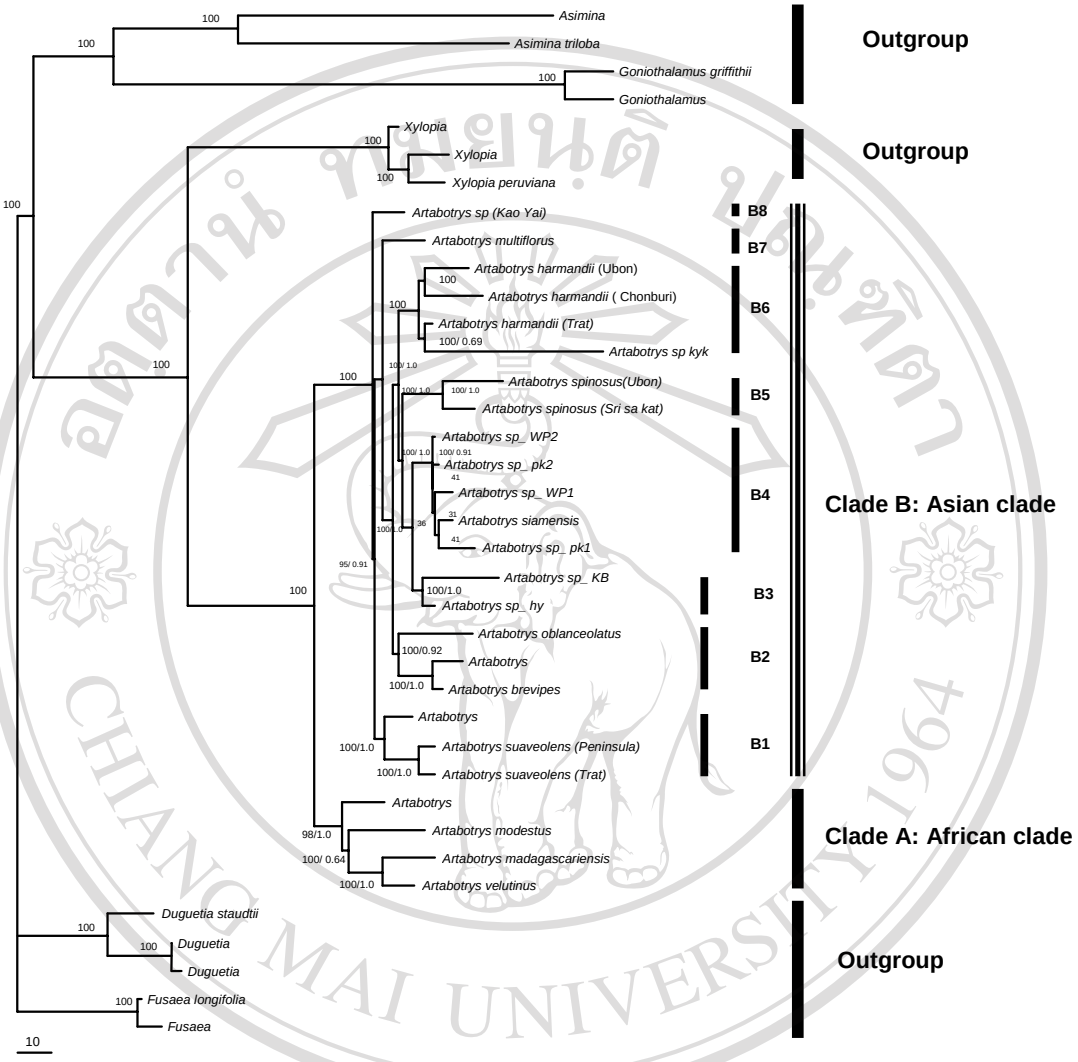

Figure 2.1 One of the shortest phylograms after maximum parsimony analysis of the combined data set. Bootstrap support value (left of slash) and posterior possibility (right of slash) are indicated above the branches.

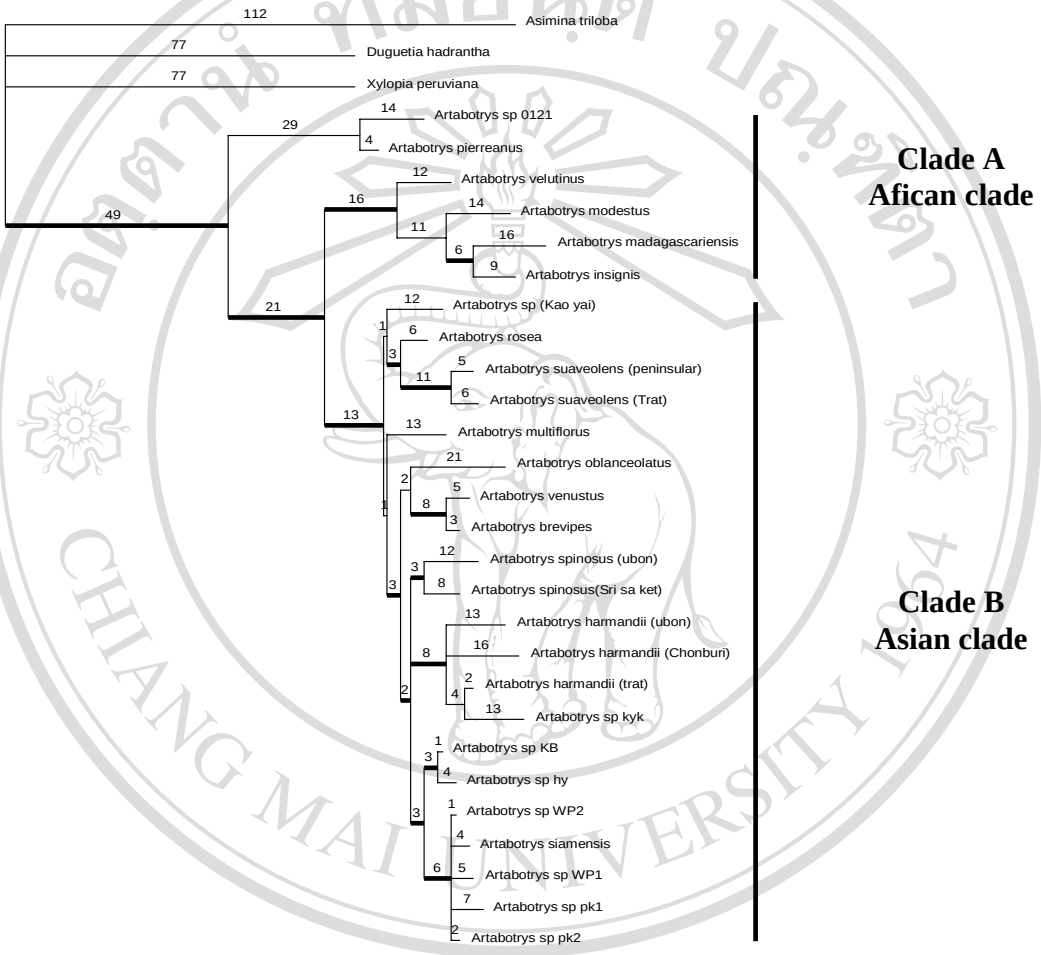

Figure 2.2 50% Majority rule consensus tree of the Bayesian inference analysis of combined *trn*L-F, *ndh*F, *psb*A-*trn*H and *AP*3 data set. Branch lengths are indicated above the branches

2.4 DISCUSSIONS

In total, 25 samples of *Artabotrys* were included in this study, among which were 4 from Africa and 21 from Asia. Two large subclades, containing eight small subclades were found. The phylogeny of the genus is described below.

Phylogenetic utility of markers

In this result several markers from different regions of the chloroplast and the low-copy nuclear genes were used (Figure 2.1-2.2). For coding plastid region, only *ndh*F region was used. It contained 2.95 % of parsimony informative characters. This region is located in the small single-copy region of the chloroplast genome, close to the junction of the inverted repeat and encodes a subunit of chloroplast NADH dehydrogenase. Most of the main clades are fairly well resolved, but still a polytomy appears in the Asian clade with this marker.

Two non-coding regions have been used in this study. The least variable, in terms of number of variable characters, and shortest was the *psb*A-*trn*H spacer region, contains 3.85 % of parsimony informative characters. This region is easily amplified. The percentage of variable characters is slightly higher than of the *trn*L-F region. However, separate phylogenetic analysis of this region resulted a consensus tree that was only partially resolved.

The *trn*L-F region (*trn*L intron and *trn*L-F spacer) is one of the most widely used chloroplast regions in plant systematics (Shaw *et al*., 2005) and has been used in Annonaceae research before (Mols *et al.,* 2004; Richardson *et al*., 2004; Pirie *et al*., 2006). This region is located in the large single copy region of the chloroplast and

consists of three tandemly arranged transfer RNA genes separated by non-coding spacer regions. It contains 0.94 % of parsimony informative characters. Separate phylogenetic analysis of this region resulted a consensus tree that was little resolved. This indicates that *trn*L-F region is less suitable for elucidating infrageneric relationships in *Artabotrys*.

The most variable in this study was the nuclear gene *AP*3. This region contains 8 % of parsimony informative characters. It has for instance been shown that the trees based on *AP*3 are more fully resolved than on basis of either *trn*L-F, *ndh*F or *psb*A-*trn*H. This marker might therefore contribute substantially to species-level phylogenetic work in *Artabotrys*.

Phylogeny and Monophyly of *Artabotrys*

The monophyly of the genus as a whole was confirmed with strongly supported (BS 100%). The relationship between African species and Asian species are clarified. Two major subclades can be recognized. Clade A (African) consists of four species which distributed in tropical Africa and situated as sister to the Asian clade (Clade B) with high percentage in both supported (BS 100% and PP= 1.0). See Fig. 2.1.

Clade B ends up in eight small subclades (Fig. 2.1- 2.2). Clade B1 consists of three specimens covering two species (*A. suaveolens, A. rosea*) which differ from other by having terete petals. Clade B8, containing only one specimen from Eastern of Thailand (*A. sp*. Kao yai). It resembles the species from clade B1, *A. suaveolens* and clade B7, *A. multiflorus* in the number of flowers, but differs in the shape and size of petals. However, the placement of clade B8 is situated as sister to the rest of Thai species.

In Clade B2, including two species (*A. oblanceolatus* and *A. brevipes*) of Thai *Artabotrys* and one species from Malaysia (*A. venustus*). All species of this clade have flat petals, leaf glabrous and round stamen connective apex which an exception in *A. brevipes*. This species much more resembles the species from clade B6 (*A. harmandii*) with glabrous plant and mucronate-apiculate connective apex. However, clade B6 differs from others by having very coriaceous leaves and glossy above. Clade B5, contains two specimens of *A. spinosus* from Thailand which differ in it scandent shrub habit and having spines. The clade B3 contains two specimens of *A. burmanicus* differ from the other members of clade B4 in its tawny tomentose. Lastly, clade B4 including all specimens found in Thailand which have rusty hairs but less than the species in clade B3. However, bootstrap support within this clade is low.

2.5 CONCLUSIONS

Results of phylogeny reconstruction presented here support monophyly of the genus *Artabotrys*. The combined DNA sequence data set of *Artabotrys* shows in a completely resolved and overall strongly supported phylogenetic tree. The monophyly African clade is sister to Asian clade. Therefore, the geographic origin of the common ancestor of *Artabotrys* is ambiguous. Further taxon sampling is needed resolve this.

The plastid *trn*L-F, *ndh*F and *psb*A-*trn*H and low-copy nuclear gene (*AP*3) were investigated with respect to usefulness in phylogenetic reconstruction. No single plastid marker so far has the power to resolve relationships within the genus as well as *AP*3 does. Further investigation will be required to increase sampling of the nuclear gene *AP*3

sequence data, to elucidate the relationships within the Asian clade, especially among Thai species.