# CHAPTER III

# ANALYSIS OF MOLECULAR DATA IN THE GENUS *GONIOTHALAMUS* USING NUCLEAR DNA AND CHLOROPLAST DNA MARKERS

## 3.1 INTRODUCTION

In the past, taxonomic data was generally collected from morphological variation. They are useful for the delimitation of genera and species but have overlap at higher taxonomic level (e.g. tribal level). More or less formal classification based on intuition or phenetic analyses of morphological characters do not accurately predict relationships between genera (Koek-Noorman et al., 1997). Then, the sources of taxonomic evidence such as anatomy, embryology, chromosomes, palynology, secondary plant compounds, proteins and DNA have been incorporated in classification and phylogenetic studies. Recently, DNA sequence data provide a source of characters suitable for building a phylogenetic classification particularly an independent scheme for classification.

### 3.1.1 The chloroplast encoded *trnL-F* region

Chloroplast (cp-) DNA genome is a circular molecule and subdivided into two single copy regions, the large single copy (LSC) region and the small single copy (SSC) region, which are separated by inverted repeats. The LSC region is slightly less conserved in sequence than the rest of the chloroplast genome (Clegg, Learn and Golenberg, 1991); hence it is potentially more useful for studies at lower taxonomic levels. Interspecific cpDNA polymorphisms are predominant in the form of length variation due to insertion/deletion mutations (McCauley, 1995) and variation in copy number of mononucleotide microsatellite repeats (Powell, Morgante and Andre, 1995). Sequences from noncoding regions of the cpDNA genome are often used in systematics because such regions tend to evolve relatively rapidly, for example the *trnL-F* intergenic spacer (Gielly and Taberlet, 1994), the *atpB-rbcL* intergenic spacer (Hodges and Arnold, 1994), the *rbcL-psaI* intergenic region (Morton and Clegg, 1993).

The cpDNA *trnT-L-F* region in land plants consists of the transfer RNA genes $trnT^{UGU}$, $trnL^{UAA}$ and $trnF^{GAA}$ arranged in tandem, separated by non-coding spacer regions. It is positioned in the large single copy region, approximately 8 Kbp downstream of *rbcL*. The *trnL* gene of cyanobacteria and a number of chloroplast genomes, including that of all land plants, contain a group-I intron positioned between the U and the A of the UAA anticodon loop. This intron is inferred by phylogenetic analysis to have been present in the cyanobacterial ancestor of the plastid lineages of Rhodophyta, Chlorophyta and Gluacocystophyta (Besendahl et al., 2000).

The succession of conserved *trn* genes and the apparent absence of gene rearrangements in the *trnL-F* region facilitated the design of plant universal primers by Taberlet et al., (1991); in particular the *trnL* intron and *trmL-F* spacer (collectively, the *trnL-F* region) has become one of the most widely used chloroplast markers for phylogenetic analysis in plants (Quandt et al., 2004).

### 3.1.2 Internal Transcribed Spacer (ITS) Region

Nuclear ribosomal DNA (nrDNA) is organized as individual chromosomal units that are repeated thousands of times in most higher plant genomes. Each of these units contains the three genes that encode the 18S, 5.8S and 26S ribosomal RNA subunits, as well as several different spacer DNA regions. The nucleotide sequence variation found in both of the internal transcribed spacer regions (ITS-1 and ITS-2) is used extensively for the systematic analysis of closely related taxa, at least in part due to the rapid rate of evolutionary change characterizing these DNA regions (Baldwin et al., 1995). The ITS sequences have proven to be a valuable source of characters to address phylogenetic relationships among closely related species in different plant families (Francisco-Ortega et al., 2001). The ITS sequences have also proven powerful in revealing hybridization and reticulate evolution (e.g. Sang, Crawford and Stuessy, 1997)

### 3.1.3 Systematic Studies of Annonaceae

Presently, there are a number of molecular phylogenetic investigations in Annonaceae. Meade (2000) used wide range of molecular data, including RAPDs,

RFLP, the *trnL-F* intergenic spacer and ITS sequences, to establish relationships within and between a small numbers of selected genera in Annonaceae. The early comprehensive molecular phylogeny of the family was carried out by Bygrave (2000; also partly published in Doyle et al. 2000) on 130 Annonacean taxa collected worldwide using *rbcL* gene.

Later, Erken (2002) studied the phylogenetic relationships among 47 taxa of *Guatteria* using DNA sequences from 6 regions in the plastid genome, including *rbcL* region, *matK* region, *trnL-F* spacer, *trnL-F* spacer and *trnL* intron and *psbA-trnH* spacer. The results suggested that *Guatteria* is a very well supported monophyletic group and the genus seems to have its origin in Central-America, rather than in South-America. However, these data are still insufficient to resolve the relationships between all closely related species.

Mols, Keßler and Gravendeel (2002) performed the phylogenetic investigation in *Miliusa* using cpDNA markers, including *rbcL*, *trnL* intron and *trnL-F* intergenic spacer, from more than 100 taxa occurring in Asia, Africa and America. The results showed that *Miliusa* proved to be monophyletic. *Polyalthia*, on the other hand, was highly polyphyletic.

Chatrou et al. (2002) presented the generic phylogenies of Neotropical Annonaceae based on cpDNA markers (*rbcL*, *trnT-L* intergenic spacer, *trnL* intron, *trnL-F* intergenic spacer and *psbA-trnH* spacer and partial *matK*) and the result indicated that *Guatteria* has the highest levels of divergence.

Annonaceae is included in the order Magnoliales (APG II, 2003). It has been suggested to be the sister taxon to Eupomatiaceae based on cladistic analysis of morphological and molecular data (Sauquet et al., 2003). Within the family, *Anaxagorea* resolved as the basal group to the rest of the family (Doyle, Bygrave and Le Thomas, 2000; Sauquet et al., 2003). The further use of morphology for phylogeny reconstruction in Annonaceae has been problematic due to difficulties in homology assessment and high levels of homoplasy (Doyle and Le Thomas, 1996). Results from phylogenetic

analyses based on *rbcL* and *trnL-F* sequence data (Mols et al., 2004; Richardson et al., 2004) supported the position of *Anaxagorea* and further divided the rest of Annonaceae between the small clades including *Cananga* and *Cleistopholis*, sister group to the larger clade including the majority of species of the family.

Later, sequences from the *trnL-F* region (not including the *trnT-L* region and *trnL* 5' exon) had been used in combination with those from further chloroplast markers (*rbcL* and *matK*) in phylogenetic reconstruction of the family Annonaceae (Sauquet et al., 2003; Mols et al., 2004; Richardson et al., 2004). The majority of species of Annonaceae fall within two large clades. The informally named 'long branch clade' (LBC) represents around 1,500 of the total 2,500 species, and is characterized by an inaperturate pollen condition. The LBC are further divided into 7 species-rich genera, i.e. *Annona*, *Artabotrys*, *Duguetia*, *Goniothalamus*, *Guatteria*, *Uvaria* and *Xylopia*. The sister group to the LBC is so-named the 'short branch clade' (SBC), representing yet another 700 species. The remaining species formed the basal grade.

Scharaschkin and Doyle (2005) investigated phylogeny and historical biogeography of the genus *Anaxagorea* using 75 morphological characters and molecular sequences from the *atpB-rbcL*, *psbA-trnH* and *trnL-trnF* spacers and the *trnL* intron. Molecular analyses alone did not support the monophyly of the Asian species, but the morphological and combined molecular and morphological analyses did.

## 3.2 MATERIALS AND METHODS

### 3.2.1 Taxon Sampling

Forty-one species of *Goniothalamus* were included and representing the geographical distribution of the genus (Table 3.1). Selected species from the LBC, *Annona squamosa* L., *Dasymaschalon lomentaceum* Finet & Gagnep., *Rollinia herzogii* R.E. Fries (AY841734), *Asimina triloba* (L.) Dunal (AY220359), *Disepalum* sp. (AY841690), *Anonidium* sp. (AY841675), *Neostenanthera myristicifolia* (Oliv.) Exell (AY743467) and *Friesodielsia desmoides* (Craib) Steenis, and the SBC, *Mitrephora keithii* Ridl. and *Polyalthia viridis* Craib, were selected as outgroup taxa. Seventeen

specimens from fresh leaves, nine specimens from silica-dried specimens and seventeen specimens from herbarium specimens were used for DNA extraction. All voucher specimens were deposited at Department of Botany Herbarium, Faculty of Science, Chulalongkorn University (BCU); The University of Hong Kong Herbarium (HKU); Nationaal Herbarium Nederland, Leiden branch (L) and The Arnold Arboretum (A).

### 3.2.2 DNA Extraction

Total genomic DNA was extracted from fresh leaf materials followed a modified cetyl trimethyl ammonium bromide (CTAB) method (Agrawal et al., 1992). For silica-gel dried materials or herbarium materials, the genomic DNA extraction followed the DNeasy Plant Mini Kit (QIAGEN, Leusden, Netherlands) according to the manufacturer's instructions or a modified CTAB method together with QIAquick™ PCR purification Kit (QIAGEN, Leusden, Netherlands) (Appendix A).

### 3.2.3 PCR amplification and sequencing

PCR reactions were set up in 50 µl reactions, usually containing 1X PCR buffer, 0.2mM dNTPs in equimolar ratio, 0.2pmol primer, 1U of *Taq* DNA polymerase (QIAGEN, Leusden, Netherlands), 50-100ng of genomic DNA and 2.5mM $MgCl_2$ for amplifying the *trnL-F* intergenic spacer or 1X Q-solution for amplifying the ITS fragment. Other modifications included increasing the amount of $MgCl_2$ or substituting 0.5% BSA for $MgCl_2$. Amplification was carried out using the PTC-100 MJ Research

The *trnL-F* intergenic spacer was amplified using the primer combinations e/f, 'e' (5'-GGTTCAAGTCCCTCTATCCC-3') and 'f' (5'-ATTTGAACTGGTGACACGAG-3') (Taberlet et al., 1991). The thermal cycling protocol comprised of 30 cycles, each with 1 minute 30 second of denaturation at 94 °C, 2 minutes of annealing at 49-50 °C, and 3 minutes of extension at 72 °C, concluding with an additional extension of 10 min at 72 °C after the final cycle.

**Table 3.1** List of taxa used for phylogenetic analyses. Herbarium abbreviation was followed Index Herbariorum I (Holmgren, Keuken and Schofield, 1981)

| No. | Species | Origin | Voucher specimen (Herbarium) |
|---|---|---|---|
| 1 | *G. aruensis* Scheff. | New Guinea | J. Regalado and W. Takcuchi 1409 (HKU) |
| 2 | *G. aurantiacus* sp. nov. | Kanchanaburi, Thailand | Saunders and Chalermglin 04/30 (HKU) |
| 3 | G. *australis* Jessup | Australia | A. Ford 4758 (HKU) |
| 4 | *G. borneensis* Mat-Salleh | Borneo | K. Sidiyasa et al. 2637 (HKU) |
| 5 | *G. cheliensis* Hu | Nan, Thailand | P. Chalermglin 470228 (BCU) |
| 6 | *G. clemensii* Bân | Borneo | A.C. Church 436 (A) |
| 7 | *G. costulatus* Miq. | Sumatra | Martati 169 (HKU) |
| 8 | *G. curtisii* King | Peninsular Malaysia | V. Balgooy 2122 (HKU) |
| 9 | *G. dewildei* R.M.K. Saunders | Sumatra | De Wilde and de Wilde Duyfies 201229 (L) |
| 10 | *G. dolichopetalus* Merr. | Borneo | A.C. Church 258 (A) |
| 11 | *G. elegans* Ast | Ubon Ratchathani, Thailand | M. Nakkuntod 40 (BCU) |
| 12 | *G. gardneri* Hook. f. & Thomson | Sri Lanka | H. Tillekeratne s.n. (HKU) |
| 13 | *G. giganteus* Hook. f. & Thomson | Bangkok, Thailand (cultivated) | M. Nakkuntod 60 (BCU) |
| 14 | *G. holttumii* J. Sinclair | Peninsular Malaysia | K.M. Kochummen FRI 16631 (HKU) |
| 15 | *G. hookeri* Thwaites | Sri Lanka | R.M.K. Saunders and A.D. Weerasooriya 00/09 (HKU) |
| 16 | G. *lanceolatus* (Bân) Mat-Salleh | Borneo | S.H. Rogstad 689 (A) |
| 17 | *G. laoticus*1 (Finet & Gagnep.) Bân | Nakhon Phanom, Thailand | P. Chalermglin 470425/1 (BCU) |

Table 3.1 (Continued)

| No. | Species | Origin | Voucher specimen (Herbarium) |
|---|---|---|---|
| 18 | *G. laoticus2* (Finet & Gagnep.) Bân | Nakhon Ratchasima , Thailand | Saunders et al. 04/1 (HKU) |
| 19 | *G. macrophyllus* (Blume) Hook. f. & Thomson var. *macrophyllus* | Narathiwat, Thailand | S. Kitamura MN7 (BCU) |
| 20 | *G. macrophyllus* (Blume) Hook. f. & Thomson var. *siamensis* J. Sinclair | Ranong, Thailand | P. Chalermglin 461220 (BCU) |
| 21 | *G. majestatis* P. Kessler | Sulawesi | McDonald and Ismail 3896 (A) |
| 22 | *G. malayanus* Hook. f. & Thomson | Narathiwat, Thailand | M. Nakkuntod 16 (BCU) |
| 23 | *G. montanus* J. Sinclair | Peninsular Malaysia | E. Soepadmo and M. Suhaimi 43 (L) |
| 24 | *G. parallelivenius* Ridl. | Sumatra, Borneo | Z. Arifin AA3014 (HKU) |
| 25 | *G. repevensis* Pierre | Thailand | Saunders et al. 04/8 (HKU) |
| 26 | *G. rongklanus* sp. nov. | Kampangphet, Thailand | P. Chalermglin s.n. (BCU) |
| 27 | *G. rotundisepalus* Henderson | Peninsular Malaysia | T.C. Whitmore FRI 4208 (L) |
| 28 | *G. salicinus* Hook. f. & Thomson | Sri Lanka | Saunders et al. 01/3 (HKU) |
| 29 | *G. sawtehii* C.E.C. Fisch. | Petchburi, Thailand | Saunders et al. 04/14 (HKU) |
| 30 | *G. scortechinii* King | Narathiwat, Thailand | S. Kitamura MN21 (BCU) |
| 31 | *G. stenomitra* (a new sp. in prep) | Pathumthani, Thailand | M. Nakkuntod 1 (BCU) |
| 32 | *G. subevenius* King | Narathiwat, Thailand | S. Kitamura MN22 (BCU) |
| 33 | *G. tapis* Miq. | Pathumthani, Thailand (cultivated) | M. Nakkuntod 2 (BCU) |
| 34 | *G. tamirensis* Pierre ex Finet & Gagnap. | Mukdahan, Thailand | P. Chalermglin 470425/2 (BCU) |
| 35 | *G. thwaitesii* Hook. f. & Thomson | Sri Lanka | Saunders et al. 01/5 (HKU) |

Table 3.1 (Continued)

| No. | Species | Origin | Voucher specimen (Herbarium) |
|---|---|---|---|
| 36 | *G. tomentosus* R.M.K. Saunders | Peninsular Malaysia | K. Ogata KEP110361 (HKU) |
| 37 | *G. tortilipetalus* Henderson | Narathiwat, Thailand | S. Nakkuntod MN58 (BCU) |
| 38 | *G. umbrosus* J. Sinclair | Narathiwat, Thailand | S. Kitamura MN23 (BCU) |
| 39 | *G. undulatus* Ridl. | Narathiwat, Thailand | M. Nakkuntod 11 (BCU) |
| 40 | *G. uvaroides* King | Peninsular Malaysia | Z. Sohadi FRI14716 (L) |
| 41 | *G. wrayi* King | Peninsular Malaysia | T.C. Whitmore FRI810 (L) |
| 42 | *Goniothalamus* sp. (Panan Sroi) | Pathumthani, Thailand (cultivated) | M. Nakkuntod 4 (BCU) |
| 43 | *Goniothalamus* sp. (Phu Soi Dao2) | Utharadit, Thailand | Saunders et al. 04/40 (HKU) |
| 44 | *Annona squamosa* L. | Bangkok, Thailand | M. Nakkuntod 45 (BCU) |
| 45 | *Dasymaschalon lomentaceum* Finet & Gagnep. | Pathumthani, Thailand (cultivated) | M. Nakkuntod 5 (BCU) |
| 46 | *Anonidium* sp. | Africa | AY 841675* |
| 47 | *Neostenanthera myristicifolia* (Oliv.) Exell | Africa | AY 743467* |
| 48 | *Rollinia herzogii* R.E. Fries | America | AY 841734* |
| 49 | *Asimina triloba* (L.) Dunal | America | AY 220359* |
| 50 | *Disepalum* sp. | Asia | AY 841690* |
| 51 | *Frisodielsia desmoides* Craib | Pathumthani, Thailand (cultivated) | M. Nakkuntod 3 (BCU) |
| 52 | *Mitrephora keitii* Ridl. | Pathumthani, Thailand (cultivated) | M. Nakkuntod 8 (BCU) |
| 53 | *Polyalthia viridis* Craib | Bangkok, Thailand | M. Nakkuntod 62 (BCU) |

* = DNA sequences from Genebank

The ITS fragment was amplified using the primer combinations ITS1/ITS4, 'ITS1' (5'-TCCGTAGGTGAACCTGCGG-3') and 'ITS4' (5'-TCCTCCGCTTATTGATATGC-3') (White et al., 1990). The thermal cycling protocol comprised of 30 cycles, each with 1 minute 30 second of denaturation at 94 °C, 2 minutes of annealing at 49-50 °C, and 3 minutes of extension at 72 °C, concluding with an additional extension of 10 min at 72 °C after the final cycle.

Primers used for amplification and sequencing of the *trnL-F* intergenic spacer and the ITS region were similar. Cloned products were sequenced using T7 and SP6 pGEM-T vector primers.

Double stranded PCR products were purified using the QIAquick™ PCR purification Kit (QIAGEN, Leusden, Netherlands) or PCR Purification Kit (MOBIO Laboratories, Inc., California, USA) according to manufacturer's instructions, and the purified products were eluted in 30 µl or 14 µl of autoclaved $H_2O$, respectively. Some samples that posted difficulties in DNA sequencing were cloned into pGEM-T vector (Promega, Madison, USA). Cloned products were purified using small-scale preparation of plasmid DNA lysis by alkali (Sambrook, Fritch and Maniatis, 1989). Automated DNA sequencing was done by Macrogen (Korea), Bio Service Unit (BSU, Thailand) and Genome Research Centre (Hong Kong).

### 3.2.4 Data Analysis

Upon completion of DNA sequencing, base calling was verified by examining fluorographs in Chromas version 1.45 (McCarthy, 1997), and corrections were made as necessary. Sequence alignment was done manually using the GeneDoc version 2.6.002 (Nicholas and Nicholas, 1997). Nucleotide frequencies, corrected distances and inferred numbers of transitions (ts) and transversions (tv) were determined by Mega 3.1 (Kumar, Tamura and Nei, 2004). Phylogenetic analyses were conducted using the parsimony method in PAUP* version 4.0b10 (Swofford, 1998). To search for the most-parsimonious (MP) tree, heuristic search with hundred random sequence additions, tree bisection reconnection (TBR) branch swapping and save all best trees were used. Characters were equally weighted. Bootstrap (Felsenstein, 1985) was used to obtain a

measure of support for each branch. Thousand bootstrap replications were carried out using full heuristic search. The strict consensus trees were presented for *trnL-F* and ITS data sets. Trees for *trnL-F* data set were shown in two consensus trees which as representative of each topology, whilst ITS tree was generated only one strict consensus tree.

Combined data analyses were performed in reduced *trnL-F* spacer data set and ITS data set. Partition homogeneity test was used to check incongruence between these two data sets. Phylogenetic analyses and bootstrap analyses were conducted in the same manner as described above.

## 3.3 RESULTS

### 3.3.1 *trnL-F* intergenic spacer analysis (full data: 44 ingroup + 10 outgroup)

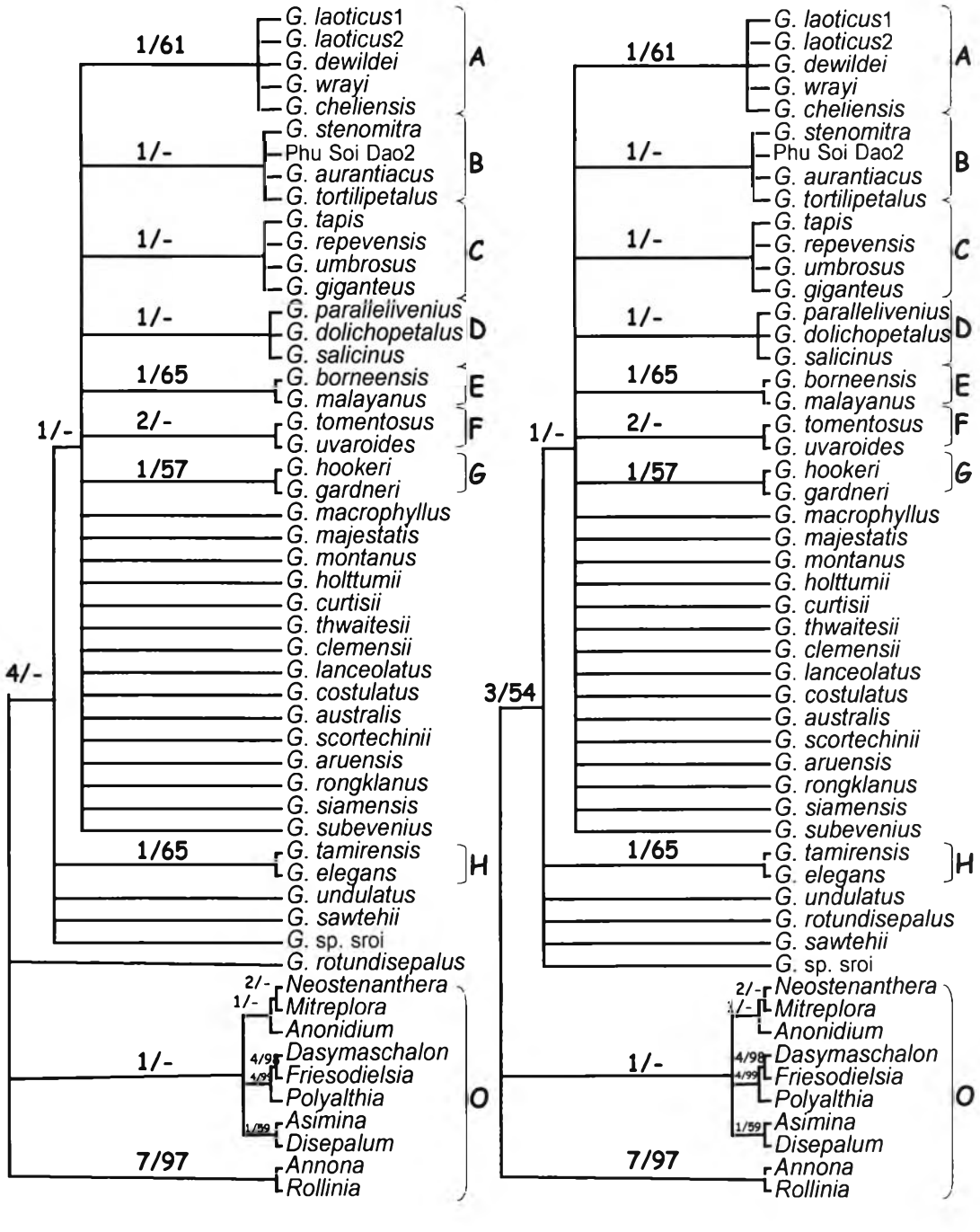
The length of the *trnL-F* spacer region ranged from 430 (*G. uvaroides*) to 476 bp (*Asimina triloba*). Within *Goniothalamus* the length ranged from 430 (*G. uvaroides*) to 465 bp (*G. majestatis*). The aligned *trnL-F* spacer sequence of 53 taxa was consisted of 617 bp. This non-coding region within *Goniothalamus* was quite variable and 18 indels ranging in size from 1 to 20 bp had to be inserted, three of which were phylogenetically informative. Of all characters used, 100 were variable, of which 34 (11.1%) were potentially phylogenetically informative while 75 (21.6%) were singleton. GC content was approximately 40%. Nucleotide pairs frequencies shows transition/transversion ratio to be about 1.2.

The parsimony analysis resulted in 38 equally most parsimonious trees, with a length of 137 steps, CI=0.8832 and RI=0.8431 (Table 3.2). Two topologies could be recognized (Figure 3.1). The difference between these two topologies was the position of *G. rotundisepalus*, either formed the sister clade to the larger clade or was a part of it. Regardless of the position of *G. rotundisepalus*, there was no difference between these two topologies. Although bootstrap values were low to moderate in the large clade, a few clades were recognized. In the large clade, 7 subclades formed a polytomy. However, there were only 3 recognizable subclades. The first clade composed of five

species, namely *G. wrayi*, *G. dewildei*, *G. laoticus* No.1, *G. laoticus* No.2 and *G. cheliensis*, with 61% bootstrap support. The second clade (65% bootstrap support) was consisted of two species, i.e. *G. borneensis* and *G. malayamus*. The last clade (57% bootstrap support) was of *G. hookeri* and *G. gardneri*. The rest subclades were shown with low bootstrap support, i.e. the *G. stenomitra*-Phu Soi Dao2-*G. aurantiacus*-*G. tortilipetalus* clade, the *G. tapis*-*G. repevensis*-*G. umbrosus*-*G. giganteus* clade, the *G. parallelivenius*-*G. dolichopetalus*-*G. salicinus* clade, the *G. tomentosus*-*G. uvaroides* clade. Overall branch supports were relatively low; few were more than 50% and none was more than 70%.

**Table 3.2** Results from the maximum parsimony analyses of *trnL-F* intergenic spacer, ITS and combined data.

| | *trn*L-F spacer (full data) | *trn*L-F spacer (reduced data) | ITS data | Combined data |
|---|---|---|---|---|
| No. of taxa | 53 | 23 | 23 | 23 |
| No. of characters | 617 | 617 | 884 | 1374 |
| Characters excluded | 312 | 0 | 181 | 181 |
| ts/tv ratio | 1.2 | 1.2 | 1.1 | - |
| No. of constant sites | 205 | 559 | 500 | 932 |
| No. of variable sites | 100 | 58 | 203 | 261 |
| No. of phylogenetically informative site | 34 | 9 | 75 | 85 |
| Tree length (steps) | 137 | 65 | 304 | 373 |
| No. of most parsimonious trees | 38 | 1 | 6 | 1 |
| Consistency index (CI) | 0.8832 | 0.9538 | 0.7796 | 0.8016 |
| Retention index (RI) | 0.8431 | 0.8500 | 0.7100 | 0.7075 |




A B

**Figure 3.1** The two topologies of the consensus trees of *Goniothalamus* from the full *trnL-F* data set. The clade indicated by "O" was an outgroup. A, B, C, D, E, F, G and H were groups for discussion. Number in front of slash was branch length and number after slash was bootstrap value from 100 replicates. Hyphen indicated bootstrap value below 50%.

### 3.3.2 ITS analysis

The length of ITS region ranged from 748 (*G. macrophyllus*, *G. tortilipetalus* and *G. repevensis*) to 792 bp (*G. giganteus*). The aligned sequence of 23 taxa was consisted of 884 bp. Twenty-three indels ranging in size from 1 to 26 bp had to be inserted. Of the 884 characters, 203 were variable, of which 75 (10.7%) were potentially phylogenetically informative while 128 (18.2%) were singleton. GC content was about 69%. Nucleotide pairs frequencies shows transition/transversion ratio to be about 1.1.

The parsimony analysis yielded 6 equally most parsimonious trees of 304 steps with CI=0.7796 and RI=0.7100 (Table 3.2). Three topologies could be recognized. The difference between these three topologies was the early branching clade. First topology showed *G. scortechinii* was the early branching clade, whilst clade H formed as a early branching clade in second topology. However, clade C that excluded *G. giganteus* was placed as a early branching clade in the last topology. Then the strict consensus tree was built and shown as Figure 3.2. The strict consensus tree formed polytomy with 4 recognizable clades, although generally low to moderate bootstrap support. The first clade (53% bootstrap support), *G. subevenius* was the first taxon to diverse, followed by two large subclades. The first subclades (89% bootstrap support) comprised of *G. malayanus* was sister to *G. cheliensis* with 99% bootstrap support and *G. laoticus* No.1 was sister to *G.laoticus* No.2 with 100% bootstrap support. The latter subclade, *Goniothalamus* sp. (sroi) was sister to the *G. undulatus*-*G. sawtehii* clade. The second clade (53% bootstrap support) consisted of four independent taxa and the *G. tapis*-*G. repevensis*-*G. umbrosus* subclade with 100% bootstrap support. Moreover, *G. tapis* was sister to *G. repevensis* with 66% bootstrap support. The third clade (88% bootstrap support) comprised four taxa, namely Phu Soi Dao2, *G. tortilipetalus*, *G. aurantiacus* and *G. stenomitra*. *G. tortilipetalus* was sister to *G. aurantiacus* with 82% bootstrap support. The last clade (79% bootstrap support) was of *G. tamirensis* and *G. elegens*. *G. scortechinii* didn't form the clade. The results from ITS locus yielded better resolved trees than that from the *trnL-F* data set, but still with generally low to moderate bootstrap supports on many branches.

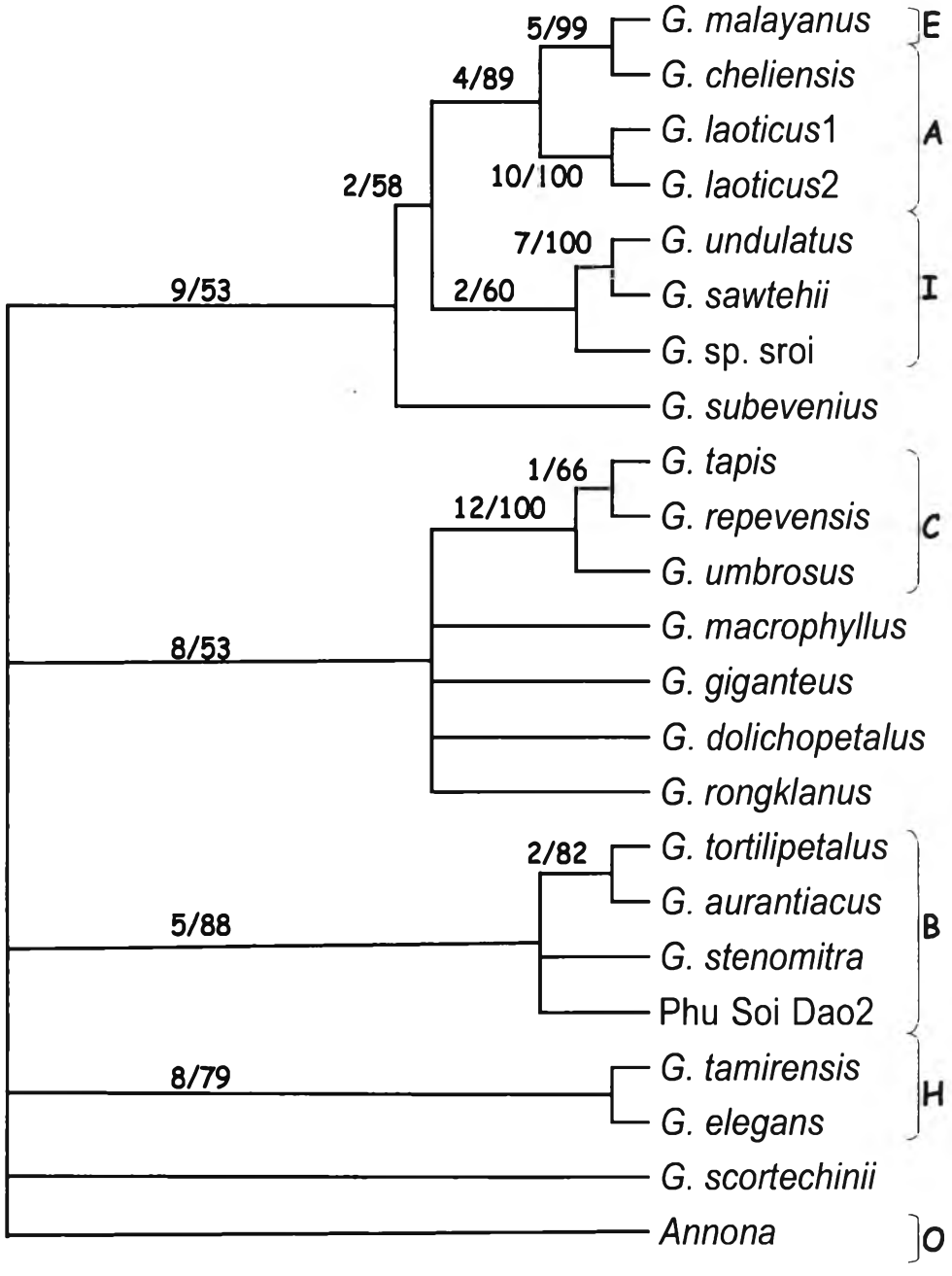



**Figure** 3.2 The strict consensus tree of *Goniothalamus* ITS sequence data with uninformative characters excluded. The clade indicated by "O" was an outgroup. A, B, C, E, H and I were groups for discussion. Number in front of slash was branch length and number after slash was bootstrap value from 100 replicates.

### 3.3.3 Combined data analysis

Due to the constraint on the number of taxa included in the ITS data set, few taxa had to be removed from the *trnL-F* spacer data set in order to be able to carry out the analyses on the combined data set. The aligned sequence of reduced *trnL-F* data set was consisted of 23 taxa and 617 bp. This non-coding region within *Goniothalamus* showed little variation, with only 8 indels ranging in size from 1 to 25 bp being inserted. Of 617 bp, 58 were variable, of these 9 (1.5%) were potentially phylogenetically informative while 49 were singleton. The parsimony analysis resulted in the most tree, with a length of 65 steps, CI=0.9538 and RI=0.8500 (Table 3.2). The tree topology was slightly different from that of full data set (Figure 3.1). The *G. tamirensis-G. elegans* clade still occupied the early branching lineage to the rest of the genus, although without bootstrap support. Therefore, the reduced *trnL-F* spacer data set can be used as representative of *trnL-F* spacer in combined analyses.

Prior to performing analyses on the combined *trnL-F* - ITS data set, partition homogeneity test (i.e. ILD, Mickevich and Ferris, 1981) was conducted to assess incongruence between these of two data sets. The test indicated these two data sets were not statistically incongruent (P value = 0.97). Thus, it was suggested that these data set could be combined.

The parsimony analysis of combined data set yielded only one most parsimonious tree of 373 steps with CI=0.8016 and RI=0.7075 (Table 3.2). This tree was more fully resolved. The monophyly of *Goniothalamus* may not be ascertained as only one outgroup taxon was included (Figure 3.3). Nonetheless, there were resolved clades with low to moderate branch supports. The first and likely to be early branching clade was the *G. tamirensis-G. elegans* clade (H). In the rest of the genus, *G.* cf. *scortechinii* formed the early branching lineage to the rest. The *G. tortilipetalus-G. aurantiacus-G. stenomitra*-Phu Soi Dao2 clade was then separated and two sister subclades, although lacking bootstrap support. However, the bootstrap supports in subterminal branches within these two clades were accasionally high, but often absent.

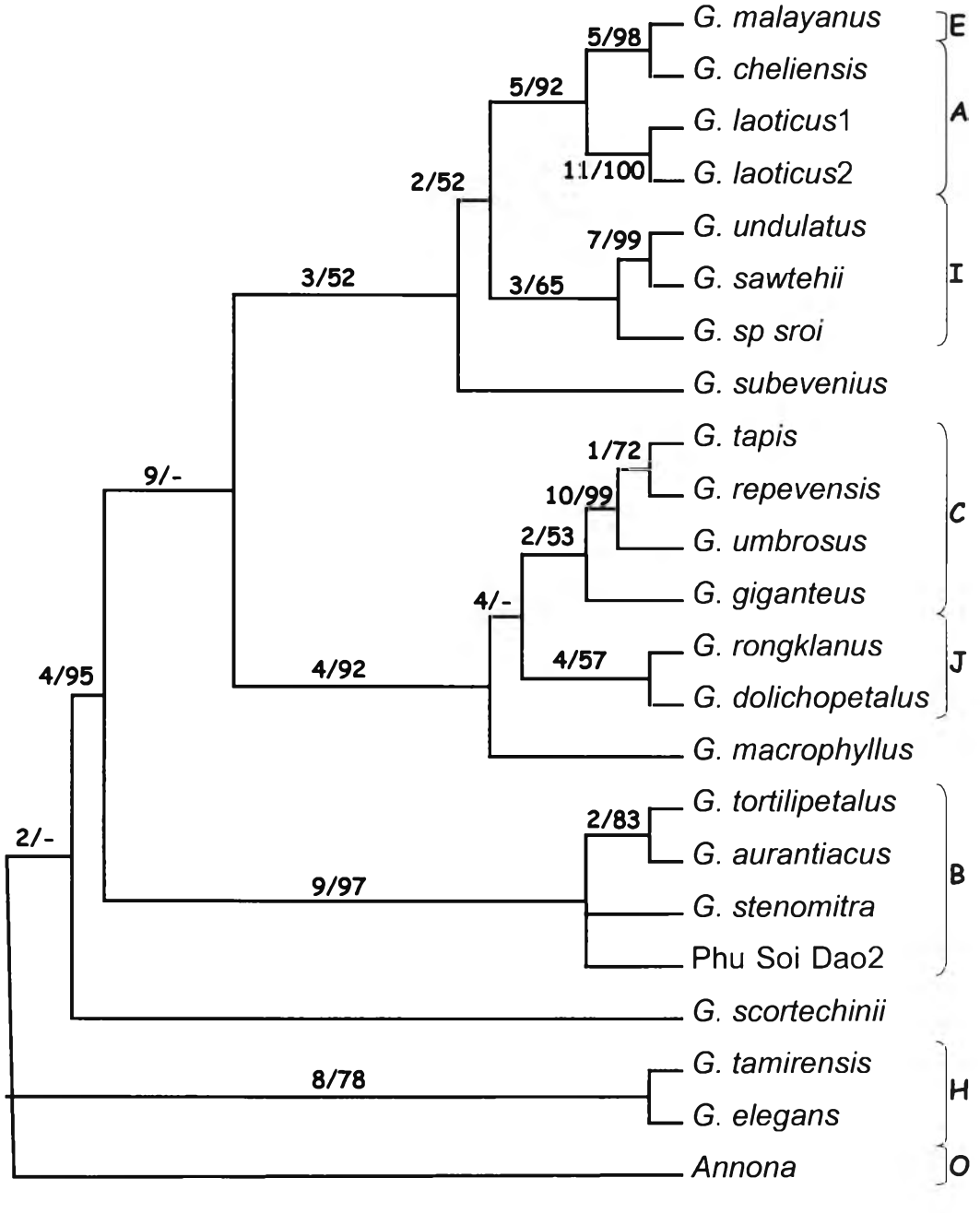



**Figure 3.3** The most parsimonious tree from combined *trnL-F* and ITS data sets. The clade indicated by "O" was an outgroup. A, B, C, E, H, I and J were groups for discussion. Number in front of slash was branch length and number after slash was bootstrap value from 100 replicates. Hyphen indicated bootstrap value below 50%.

## 3.4 DISCUSSION

### 3.4.1 Comparisons among *trnL-F*/ITS/conbined data trees

The phylogenetic trees from *trnL-F* and ITS data sets showed little resolution in the consensus trees due to ambiguous characters supports, and thus low bootstrap values in many clades. The early branching position on the *trnL-F* tree was *G. rotundisepalus*,

whilst that on the ITS tree was the *G. tapis-G. repevensis-G. umbrosus* clade or *G.* cf. *scortechinii* or the *G. tamirensis-G. elegans* clade. For combined tree, the *G. tamirensis-G. elegans* clade was the early branching position, similar to the ITS tree.

Although a few clades were shown (i.e. clades A-I in Figure 3.1 and 3.2), the relationships among taxa and clades may not fully be inferred. In the *trnL-F* tree, clade A was not resolved but it were formed together as sister taxa to clade E in the ITS tree. Similarly, clade B was found on both trees, but the ITS tree indicated that within this clade *G. tortilipetalus* was the sister taxon to *G. aurantiacus*. In contrast, in clade C in the *trnL-F* tree is not present similar to in the ITS tree due to the exclusion of *G. giganteus*. Moreover, clades D, F and G were only present in the *trnL-F* tree but not in the ITS tree. This absence might be due to the limiting taxa in the ITS analysis or lack of bootstrap support in the *trnL-F* tree. *G. tamirensis* was sister to *G. elegans*, as clade H in both trees with moderate bootstrap support. In addition, clade I was only found in the ITS tree. It suggested that *G. undulatus* was closely related to *G. sawtehii*; this relationship was not recovered in the *trnL-F* tree.

From the combined data analyses (Figure 3.3), many subclades in Figure 3.1 and 3.2 existed in the combined tree, including clades A+E (92% bootstrap support), B (97% bootstrap support), C (53% bootstrap support), H (78% bootstrap support) and I (65% bootstrap support), whilst clade J (57% bootstrap support) was shown only in the combined tree. However, a few clades might be reliable. Clade H was early branching positioned as in the ITS tree, although lacking bootstrap support, with *G. scortechinii* early branching to the rest. Clade B was then still appeared and separated as the early

branching lineage to the larger clade. Two sister subclades situated at more latter position on the tree with low bootstrap support. In the first subclade, clades C and J appear as the sister groups (without bootstrap support) and *G. macrophyllus* was the early branching lineage of clade C+J. Clade C, excluding *G. giganteus*, was reliable due to high bootstrap support. This subclade was supported with 53% bootstrap value. In the other subclade, *G. subevenius* was sister to the large clade. Group A was paraphyletic to clade E, with *G. cheliensis* (in group A) sister taxon to *G. malayanus* (in clade E). *Goniothalamus* sp. (sroi) was sister taxon to the *G. undulatus-G. sawtehii* clade, as clade I appeared only in the combined tree. The *G. undulatus-G. sawtehii* clade was reliable due to high bootstrap support, alike the *G. malayanus-G. cheliensis* clade and the *G. laoticus*1-*G. laoticus*2.

In the light of combined analysis, the unclear taxa could be placed into the clades on all trees. *G. laoticus*2 was placed to *G. laoticus*1, suggesting that they are not significant different in molecular data although different location. Therefore, they are counted as one species in next chapter. *Goniothalamus* sp. (sroi) was sister to the *G. undulatus-G. sawtehii* clade. Its relationships might be reliable due to having shared some morphological characters, i.e. truncate staminal connective, glabrous style and stigma, funnel-shaped stigma. In other taxa, *G. rongklanus* was placed with *G. dolichopetalus*. *G. rongklanus* has hairy ovary and glabrous style similar to *G. dolichopetalus*, so their relationships seem likely. *Goniothalamus* from Phu Soi Dao No.2 was placed in the clade B together with *G. stenomitra*, *G. tortilipetalus* and *G. aurantiacus*. This species has large and erect sepals and connate at base similar to these three species, thus its placement might be reliable.

### 3.4.2 Infra-generic relationships within *Goniothalamus*

So far, the only infra-generic classification of *Goniothalamus* was that of Bân (1974b). However, this classification appeared to be somewhat artificial because it was based on 2-3 morphological characters to classify at subgeneric or section level. In addition, my data based on molecular evident was highly incongruent with Bân's classification. Several clades were reliable in our analyses of the *trnL-F*, ITS and

combined data sets (Figure 3.1, 3.2 and 3.3). Clade A comprised five species receiving 61% bootstrap support in *trnL-F* tree. In this clade, *G. wrayi* was likely to be closely related to *G. dewildei*, because of having apiculate staminal connective. This finding agreed with Saunders (2002) suggestion that *G. wrayi* was closely related to *G. dewildei* but differences are in its smaller leaves, fewer secondary vein in the leaves, axillary flowers, distinctively tapered staminal connectives, sparely hairy styles, and ellipsoid monocarps, although these two species was not in ITS and combined trees. A possible morphological synapomorphy for this clade A is erect sepals. Nonetheless, clade A seems as paraphyletic to *G. malayanus* in clade E with 89% and 92% bootstrap supports in ITS and combined trees, respectively.

Clade B was always appeared in all analysis, with high bootstrap supports in ITS and combined trees. Therefore, this clade might be reliable although this clade lacks bootstrap support in *trnL-F* tree. The possible morphological synapomorphies of this clade are connate sepals, sepal persistence in fruit and glabrous seed testa. Moreover, *G. tortilipetalus* was closely related to *G. aurantiacus* with shared some morphological character, i.e. distinctly prominence of secondary vein on adaxial leaf surface, percurrent tertiary vein arrangement, copious mucilage around seed.

Clade C consisted of four species with low bootstrap support in both *trnL-F* and combined trees. Clade C, excluded *G. giganteus*, might be more reliable due to high bootstrap support in ITS and combined trees. The possible morphological synapomorphies of this clade are hairy style and ovary.

Clades D and F are not reliable due to none bootstrap support in *trnL-F* tree and a number of these two clades were not appeared in ITS and combined trees. It is resulted of limiting taxa or none bootstrap support.

Clade E (65% bootstrap support in *trnL-F* tree) showed that *G. malayanus* was sister to *G. borneensis*. Also, somewhat in the same line of Mat-Salleh (2001) and Saunders (2002) suggestion that *G. malayanus* was closely related to *G. borneensis* and *G. giganteus*. My data, however, *G. giganteus* was not included in the clade E due

to having the warty fruits, while the fruits of *G. malayanus* and *G. borneensis* are smooth.

Clade G presented that G. *hookeri* was closely related to *G. gardneri* with 57% bootstrap support in *trnL-F* tree and with the possible synapomorphic character of hairy ovary and shared the same geographic distribution.

Clade H, with 65%, 79% and 78% bootstrap support in *trnL-F*, ITS and combined trees, respectively. This clade explained that *G. tamirensis* was sister to *G. elegans*. The possible synapomorphies are truncate staminal connective, glabrous pistil with large convoluted stigma.

Clade I consisted three taxa with moderate bootstrap support in ITS and combined trees, although this clade was not found in *trnL-F* tree. *G. undulatus* was sister to *G. sawtehii* with 100% bootstrap support. Thus this relationship is reliable and shared morphological characters, namely truncate staminal connective, glabrous style and stigma. Besides clade I might be related clades A and E but low bootstrap support (58% in ITS tree and 52% in combined tree)

Clade J showed the relationship of *G. rongklanus* and *G. dolichopetalus* with 57% bootstrap support in combined tree and shared two morphological synapomorphies, namely reflexed sepal and hairy of seed testa. Nevertheless, this relationship might be not certain, alike the relationship between clade J and clade C with none bootstrap support.

As discussion above, a few clade from these molecular phylogenies was reliable. These clades seem not to agree with Bân (1974b) infra-generic classification of *Goniothalamus*. In many aspects, however, these results seem to agree with Saunders (2002, 2003) view of *Goniothalamus* species relationships. Thus, new classification scheme should be proposed based on new evidence of molecular and morphological analyses.