

# Technical report on the molecular phylogeny of *Camellia* with nrITS: the need for high quality DNA and PCR amplification with *Pfu*-DNA polymerase

Kunjunpillai VIJAYAN and Chih-Hua TSOU\*

*Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan 115, ROC*

(Received May 17, 2007; Accepted December 6, 2008)

**ABSTRACT.** Internal transcribed spacer (ITS) of nrDNA has been widely employed for reconstructing phylogenetic relationships in plants, especially at the species level. Previous attempts to reconstruct the molecular phylogeny of *Camellia* based on nrITS, however, have not succeeded due to technical difficulties. In order to identify the major factors responsible for these difficulties and also to assess the efficacy of nrITS in elucidating the interspecific relationships of *Camellia*, the present investigation was conducted with seven closely or distantly related species. The purity of the DNA was found to be one of the major factors affecting the success of PCR amplification and the errors in the sequences. Therefore, an efficient protocol has been developed for extracting genomic DNA from dried leaf samples of *Camellia*. The purity of the DNA, extracted using this method, was quite good as revealed by the A260/A280 ratio, which ranged from 1.84 to 1.89. Further investigation on the effect of DNA polymerases on PCR induced variations revealed that the PCR error rate was much higher in *Taq*-amplified sequences than *Pfu*-amplified sequences. The effect of the error on phylogenetic analysis was evident from the wide dispersal of *Taq*-amplified sequences across the gene tree while the *Pfu*-amplified sequences from the same sample joined together to form a single clade. Our extensive study of *Camellia* based on *Pfu*-amplified ITS sequences showed well-resolved interspecies relationships. Since the results of the molecular phylogenetic investigation of *Camellia* needs to be reported in a series, due to the technical and scientific complexity of the work, in this first report, we provide technical and scientific insights into the major factors responsible for the failure of the PCR amplification, the occurrence of high sequencing errors, and their effect on the phylogenetic interpretations. The results further stress the potential of nrITS in deducing the phylogenetic relationships in *Camellia*.

**Keywords:** *Camellia*; DNA isolation; ITS; PCR error; *Pfu*; *Taq* polymerase.

## INTRODUCTION

Molecular phylogeny based on DNA sequences has recently been used extensively to resolve intricate problems at various taxonomic levels. The internal transcribed spacer (ITS) of the nuclear ribosomal DNA has been the most widely used molecular marker in resolving phylogeny at the generic and specific levels in the angiosperms (Baldwin et al., 1995; Alvarez and Wendel, 2003). More than two-thirds of the related papers published during 1998 to 2002 and in 2005 included nrITS in the analyses (Alvarez and Wendel, 2003; Feliner and Rossello, 2007). The advantages as well as the disadvantages of nrITS in the phylogenetic application have also been well explored (Buckler et al., 1997; Alvarez and Wendel, 2003; Feliner and Rossello, 2007). Although general concerns have been raised about its sequence complexity and the existence of infra-species or even

infra-individual variations, which impact the accuracy of the phylogeny being deduced, its advantages at a specific level of phylogeny are unsurpassed by any other markers. Recently, it has become customary that whenever nrITS is used for phylogenetic analysis, a segment of plastid or mitochondrial sequences is also incorporated to make a comparison.

The genus *Camellia*, the largest genus in the family Theaceae, is equally important for both the horticultural and beverage industries. It is mainly distributed in East Asia, with Southwest China as the center of distribution (Ming, 2000). Taxonomically, the most popular classification systems proposed by Chang (1981, 1998) and Ming (2000) are very different. For example, Chang recognized 284 species and treated them in 22 sections, but Ming recognized 119 species and 14 sections only. The difficulty in reaching a well-accepted classification system resides in the large number of the species having natural hybridizations (Ming, 2000) and in the great number of

\*Corresponding author: E-mail: chtsou@gate.sinica.edu.tw.

variations in floral characters. For example, the number of bracteoles, sepals, and petals is not fixed in many species, and the degree of filamentous and stilar fusion is variable within species and difficult to quantify. In addition, floral size often varies within a single species. In the past decade, while molecular taxonomy has been overwhelmingly applied to the ranks of all living organisms, some effort has undoubtedly been devoted to the genus *Camellia* as well. The earliest report on the molecular taxonomy of *Camellia*, to our knowledge, was a presentation in a conference by Thakor and Parks (1997). In this study, four cpDNA regions (*rbcL*, *trnT-trnL*, *trnL-trnF*, *atpB-rbcL*) were sequenced from 20 species which showed great morphological diversity, but the sequence variability was as low as 0% to 3%. Later on, in Thakor's unpublished thesis, one more cpDNA region (rpL16 intron) was used, but it showed high intra-specific variation (Xiao, 2001). Thakor's study based on morphology and DNA sequences suggested that *Camellia* was paraphyletic (Xiao, 2001). Xiao (2001) continued the study of molecular phylogeny in *Camellia*, in the same laboratory, with two intron sequences from the RNA polymerase (RPB2) gene. His extensive survey, which included 149 *Camellia* species and around 2000 nucleotides, resulted in a well resolved tree in which four sections of the genus *Camellia* formed distinct clades while the remaining, around three-fifths of the species, formed a few clades that could be defined by the color of the petals and the geographic distribution range (Xiao and Parks, 2003). Nevertheless, this work contributed many new insights and was different from previous treatments done by Sealy (1958), Chang (1998), and Ming (2000). Meanwhile, a few small-scale studies also appeared like that of Tang and Zhong (2002), who constructed a molecular phylogeny of 15 golden camellias with nrITS using DNA from fresh leaf samples. Orel et al. (2003) studied interrelationships of 32 species with cpDNA *trnL-trnF* sequences and ISSR markers and found that the former showed limited variations while the latter appeared to be influenced by phylogenetic as well as non-phylogenetic factors, rendering the tree topology difficult to interpret. Yang et al. (2006) reported their surveys on the utility of four DNA regions from 21 *Camellia* species. They claimed that the cpDNA *trnL-F* and *rpL16* were too conserved to be used while nrDNA-ITS was difficult to sequence technically. A waxy gene, however, appeared to be potentially useful. Although Yang et al. (2006) was the first to mention the difficulties of nrITS sequencing in *Camellia*, our personal contact with many laboratories revealed that, due to difficulties in sequencing, most of the laboratories engaged in the molecular phylogeny of *Camellia* with nrITS have abandoned their efforts. Since our laboratory is involved in the phylogenetic analysis of the family Theaceae, and nrITS appears to have more potential than many other currently available gene sequences in resolving the phylogeny of *Camellia*, we decided to undertake the present investigation to identify and resolve the major bottlenecks in the phylogenetic studies of *Camellia* with nrITS.

Our first major concern was the failure of PCR amplification. Thus, we concentrated on the quality of DNA as many contaminating agents—like polysaccharides, phenolic compounds, tannins, resins, and latex, present in the cell as secondary metabolites, are present in the leaf samples—usually coprecipitate with DNA and interfere with the activity of the DNA polymerase enzyme (Merlo and Kemp, 1976; Shioda and Marakami-Musfushi, 1987; Fang et al., 1992; Panday et al., 1996). In *Camellia sinensis*, i.e. tea, one of the most common beverages, and most likely in many other species of *Camellia* as well, a high content of secondary metabolites is present in the leaves as 40% of the dry leaf weight of *Camellia sinensis* comes from polyphenolic compounds such as flavanols and their glycosides, leucoanthocyanins, and phenolic acids, and more than 12% is polysaccharides (Graham, 1992). Therefore, extraction of high quality genomic DNA from dried leaf samples is challenging but a prerequisite for getting the right sequences through PCR amplification. Initially we tried a number of the existing protocols (Murray and Thompson, 1980; Dellaporta et al., 1983; Rogers and Bendich, 1988; Doyle and Doyle, 1990; Tai and Tanksley, 1990; Matsumoto et al., 1994; Takeuchi et al., 1994; Wachira et al., 1995; Struwe et al., 1998), but none of them was found efficient enough to isolate DNA suitable for consistent PCR amplification. Therefore, we developed a protocol suitable for extracting highly purified genomic DNA, which is detailed in this report.

Our second concern was the possible occurrence of intragenomic variations among the nrITS repeats, as our attempts at direct sequencing of PCR product always failed due to the large amount of noise (unreadable sites) even though the DNA was of good quality. Thus, cloning and sequencing of the PCR products became a necessity to probe into the role of intragenomic heterogeneity in it. Also proofreading DNA polymerase was tried to examine the extent of the sequence variations caused by the routinely used non-proofreading polymerase. Most of the laboratories involved in phylogenetic studies with large number of species and genera generally use the thermostable polymerase from *Thermus aquaticus* (*Taq*) for PCR amplifications due to its remarkable stability at higher temperature, sensitivity, yield, and low economic input. However, the PCR products amplified by *Taq* polymerase are subjected to nucleotide changes during the amplification (Saiki et al., 1988; Tindall and Kunkel, 1988; Kwiatowski et al., 1991; Lundberg et al., 1991; Kobayashi et al., 1999), and it was found that use of a high fidelity DNA polymerase such as *Pfu*, *Vent*, *DeepVent*, or *UITma* could reduce the amplification error in PCR products as compared to *Taq* DNA polymerase (Flaman et al., 1994; Cline et al., 1996). However, the higher cost of these enzymes remains the major constraint on their regular use. Considering this and the need of experimental accuracy, it is prudent to carry out an initial polymerase fidelity assay with the genomic DNA using non-proofreading and proofreading polymerases. Therefore,

we made a comparison of the fidelity of *Taq* polymerase with that of *Pfu* polymerase, and the results showed surprisingly high sequence variations in amplification made by *Taq* polymerases. During the last two years, we have developed a protocol and succeeded in extending the nrITS sequencing to about 100 species of *Camellia*. The results obtained from the initial experiments were insightful and useful in resolving intrageneric relationships of *Camellia*. We intend to present our findings in a series of reports to include both technical and scientific details of the study. In this first report, we provide technical details of the DNA extraction, methods of nrITS sequencing and the major factors responsible for the failure of the PCR amplification and the higher sequencing errors. We also provide a snapshot of the potential of nrITS to resolve the intrageneric relationships in *Camellia*.

## MATERIALS AND METHODS

### Plant material

Leaf samples of seven species belonging to four sections, namely, *C. assamica*, *C. euphlebia*, *C. formosensis*, *C. microphylla*, *C. sasanqua*, *C. sinensis* var. *sinensis*, and *C. vietnamensis* were used for the experiment (Table 1). The materials were kept in a sealable bag containing silica gel immediately after detachment. Voucher specimens were made and kept in the HAST herbarium.

### Solutions and reagents

- Extraction buffer (Carlson et al., 1991). The extraction buffer can be prepared by dissolving 2% (w/v) cetyltrimethyl ammonium bromide (CTAB), 1.4 M sodium chloride, 20 mM EDTA, 100 mM Tris-HCl, 1% PEG 8000. Just before use, add 1.5% (v/v)  $\beta$ -mercaptoethanol to the extraction buffer.

- 2% (w/v) Polyvinyl pyrrolidone (PVP; Sigma, St. Louis, MO; MW 10,000)
- Phenol: Chloroform: isoamyl alcohol (24:24:1)
- Chloroform: isoamyl alcohol (24:1)
- 3 M Sodium acetate solution (pH 4.8)
- Absolute alcohol
- 70% (v/v) ethanol
- RNase-A (10 mg/ml)
- Tris-HCl (pH 8.0)
- EDTA (pH 8.0)
- GENECLEAN III-DNA purification kit.

### DNA extraction protocol

One square cm of the dried leaf lamina was powdered in a tube using a FASTPrep-FPI120 machine (BIO 101 Systems, New York). The extraction buffer along with 2% (w/v) PVP was added into the tube containing the leaf powder, mixed thoroughly, and kept at 4°C for 5 days. The slurry was subsequently incubated at 65°C for 1 h. After incubation, an equal volume of chloroform: isoamyl alcohol (24:1) was added, mixed well, and centrifuged at 14,000 rpm for two min. The supernatant was collected in a fresh tube, and this step was repeated one more time. To the supernatant, a 1/10 volume of 3 M-sodium acetate and 2 volumes of pre-cooled ethyl alcohol were added and mixed slowly by inverting the tube. The tubes were kept at -20°C for 1h to precipitate the DNA before it was spooled out into a fresh tube and air-dried at 37°C. The air-dried DNA was dissolved in TE Buffer (10 mM Tris HCl [pH 8.0] and 1 mM EDTA). RNA contamination was removed by treating the DNA with bovine pancreatic RNase-I at a final concentration of 40  $\mu$ g/ml at 37°C for 30 min. The DNA, after RNase treatment, was re-extracted using a modified protocol of Struwe et al. (1998).

**Table 1.** List of samples, voucher information, and sequence accession numbers.

Species	Section		Collection locality	Voucher (source origin)	Accession number
	Chang (1998)	Ming (2000)			
<i>C. euphlebia</i>	<i>Chrysantha</i>	<i>Archeacamellia</i>	Xishanbana, Yunnan, China	Wong Hong, 2004-3-4-a	<i>Taq</i> : EF544731-EF544734 <i>Pfu</i> : EF544735-EF544737
<i>C. microphylla</i>	<i>Paracamellia</i>	<i>Tuberculata</i>	Internatl. Camellia Gard., Zerjiang, China (cultivated)	Tsou-853T (J. Y. gao)	<i>Taq</i> : EF544763-EF544767 <i>Pfu</i> : EF544768-EF544772
<i>C. assamica</i>	<i>Thea</i>	<i>Thea</i>	Tea Improv. Res. Inst. Nantou, Taiwan (cultivated)	Tsou-455S	<i>Taq</i> : EF544721-EF544725 <i>Pfu</i> : EF544726-EF544730
<i>C. formosensis</i>	<i>Thea</i>	<i>Thea</i>	Nantou, Taiwan	Tsou-318S	<i>Taq</i> : EF544711-EF544715 <i>Pfu</i> : EF544716-EF544720
<i>C. sinensis</i> var. <i>sinensis</i>	<i>Thea</i>	<i>Thea</i>	West Tian-mu Mt., Zerjiang, China	Tsou-894T	<i>Taq</i> : EF544738-EF544741 <i>Pfu</i> : EF544742-EF544745
<i>C. sasanqua</i>	<i>Oleifera</i>	<i>Oleifera</i>	Yang-Ming Mt., Taipei, Taiwan (cultivated)	Tsou-988T	<i>Taq</i> : EF544755-EF544757 <i>Pfu</i> : EF544758-EF544762
<i>C. vietnamensis</i>	<i>Oleifera</i>	<i>Oleifera</i>	Internatl. Camellia Gard., Zerjiang, China (cultivated)	Tsou-856T (J. Y. Gao)	<i>Taq</i> : EF544746-EF544749 <i>Pfu</i> : EF544750-EF544754

An equal volume of Phenol-Chloroform-Isoamyl alcohol (24:24:1) was added to the tube and mixed gently for 10 min. The upper aqueous layer, obtained after centrifuging at 14,000 rpm for 2 min, was transferred into a new tube and an equal volume of Chloroform-Isoamyl alcohol (24:1) was added and mixed gently for 10 min. The tube was subsequently centrifuged at 14,000 rpm for 2 min. From the upper aqueous layer 300  $\mu$ l was transferred into a new tube containing 900  $\mu$ l of sodium iodide solution and 20  $\mu$ l of glass milk prepared from GENECLENE<sup>®</sup> kit (BIO101 Systems, New York), mixed gently for 20 min and centrifuged at 14,000 rpm for 30 s to pellet the glass milk. The supernatant was poured off and 900  $\mu$ l of NEW<sup>™</sup> wash solution was added. The glass milk pellet was resuspended by gently breaking the pellet with a pipette tip and shaking the tube. The tube was centrifuged, and the supernatant was poured off. The washing process was repeated thrice with 900  $\mu$ l NEW<sup>™</sup> wash solution, and a fourth washing was done with 150  $\mu$ l NEW<sup>™</sup> wash solution. After centrifuging, the supernatant was removed completely with a fine tipped pipette without disturbing the pellet. The glass milk pellet was resuspended in TE buffer and placed at 50°C for 10 min to elute the DNA from the glass milk beads. The tube was centrifuged again at 14,000 rpm for 2 min, and the supernatant was collected into a new tube without disturbing the glass milk pellet.

### DNA quantification

The quantity and quality of the DNA was estimated with spectrophotometer (Hitachi U2000; Hitachi High Technologies America, Inc. Life Sciences Division, San Jose, CA, USA) as well as by electrophoresis on 1.0% agarose gel in 1 $\times$ TAE buffer.

### PCR amplification with *Taq* and *Pfu* polymerases

The PCR amplification of the ITS1-5.8S-ITS2 regions of the nrDNA was achieved with the primer pairs ITSleu1 (5'-GTCCACTGAACCTTATCATTAG-3' (Urbatsch et al., 2000) and ITS4 (5'-TCCTCCGCTTATTGATATGC-3') (White et al., 1990). The PCR amplification reaction was first carried out with *Taq* polymerase (1U) in a reaction mixture containing 5 mM Tris-HCl (pH 8.8 at 25°C), 10 mM NaCl, 0.01 mM EDTA, 0.1 mM DTT, 0.5% (v/v) glycerol, 0.1% (v/v) Triton<sup>®</sup> X-100 (Promega corporation, Madison, WI, USA), 2 mM MgCl<sub>2</sub>, 20 ng of Genomic DNA, 100  $\mu$ M dNTPs, and 100  $\mu$ M primers. Then, amplifications were carried out with *Pfu* polymerase (Promega corporation, Madison, WI, USA) with the same primers. The PCR cocktail contained *Pfu* 1U, 20 mM Tris-HCl (pH 8.8 at 25°C), 10 mM KCl, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2.0 mM MgSO<sub>4</sub>, 0.1% (v/v) Triton<sup>®</sup> X-100 and 0.1 mg/ml nuclease-free BSA (Promega Corporation, Madison, WI, USA), 20 ng of Genomic DNA, 100  $\mu$ M dNTPs, and 100  $\mu$ M primers. The PCR thermal cycler (GeneAmp PCR system 2700, Applied Biosystem, Foster City, CA, USA) was programmed as 94°C for 5 min for denaturing the DNA followed by 30 cycles of 1 min at 94°C, 1 min at 50°C, 1 min at 72°C, followed by a final extension of 7 min

at 72°C (Mast, 1998). PCR products were electrophoresed on 1.0% agarose gel in TAE buffer, stained with ethidium bromide (2  $\mu$ g/ml), and visualized under ultraviolet light.

### Cloning and sequencing

PCR products were purified with QIAquick PCR purification kit (QIAGEN Inc., Valencia, CA, USA). PCR products of each sample were cloned to PCR II-TOPO cloning vector and transformed into chemically competent Ecoli- DH5 $\alpha$ <sup>™</sup>-T1<sup>R</sup> cells (provided with the kit) following the instructions of the manufacturer (Invitrogen Life Technologies, Carlsbad, California, USA). Colonies were cultured overnight at 37°C on LB (Luria-Bertani medium) ampicillin/IPTG (isopropyl  $\beta$ -D-1-thiogalactopyranoside)/X-gal (5-bromo-4-chloro-3-ido-yl beta-D-galactoside) selective medium. Plasmids from white colonies were isolated using Mini-M<sup>™</sup> plasmid DNA extraction system (VIOGENE, Sunnyvale, California, USA). The extracted plasmids were digested with *Eco*R1 and tested on 1% agarose gel for inserts. For each species, plasmids from five colonies were sequenced using an ABI PRISM dye terminator cycle sequencer, model 3700 (Applied Biosystems, Foster City, CA, USA).

The full sequences were generated by aligning both forward and reverse sequences of the same clone, and any variation observed in the sequences was checked by examining corresponding peaks in the chromatograms of both forward and reverse sequences. Sequence alignments were constructed with the help of the PILEUP program of GCG, version 8.1 (Genetic Computer Group, 1994). Base composition and length variability were estimated with BioEdit, version 5.0.9 (Hall, 1999). Intra-genomic variation was calculated from the multiple clones as mean number of nucleotide differences per site between pairs of sequences according to Kimura's two-parameter model using MEGA 3.1 (Kumar et al., 2004). Standard errors of the sequence divergence were estimated by applying 1000 bootstrap replications. Total number of transitions and transversions present within a set of sequences was estimated with Tamura-Nei (gamma) algorithm (Tamura and Nei, 1993) as implemented in MEGA 3.1. The PCR error rate was calculated from the observed error frequency on the basis of pair-wise comparison by taking into account the number of doubling cycles using the formula (2  $\times$  observed frequency) / (number of cycles) as described by first Gelfand and White (1990) and later by Kwiatowski et al. (1991). While calculating the PCR error, the variable sites inherently present in the sequence were eliminated based on the existence of patterns of *Pfu*-amplified sequences present in each sample. For instance, in the aligned matrix of the *Pfu*-amplified sequences of each species, if two or more sequence types could be recognized, the variable sites resulting from such differences in the sequence types, as indicated in Figure 1, were identified and eliminated from further analyses of errors.

The gene tree was generated with PAUP\* 4.0b10 (Swofford, 2001) using a maximum likelihood algorithm.

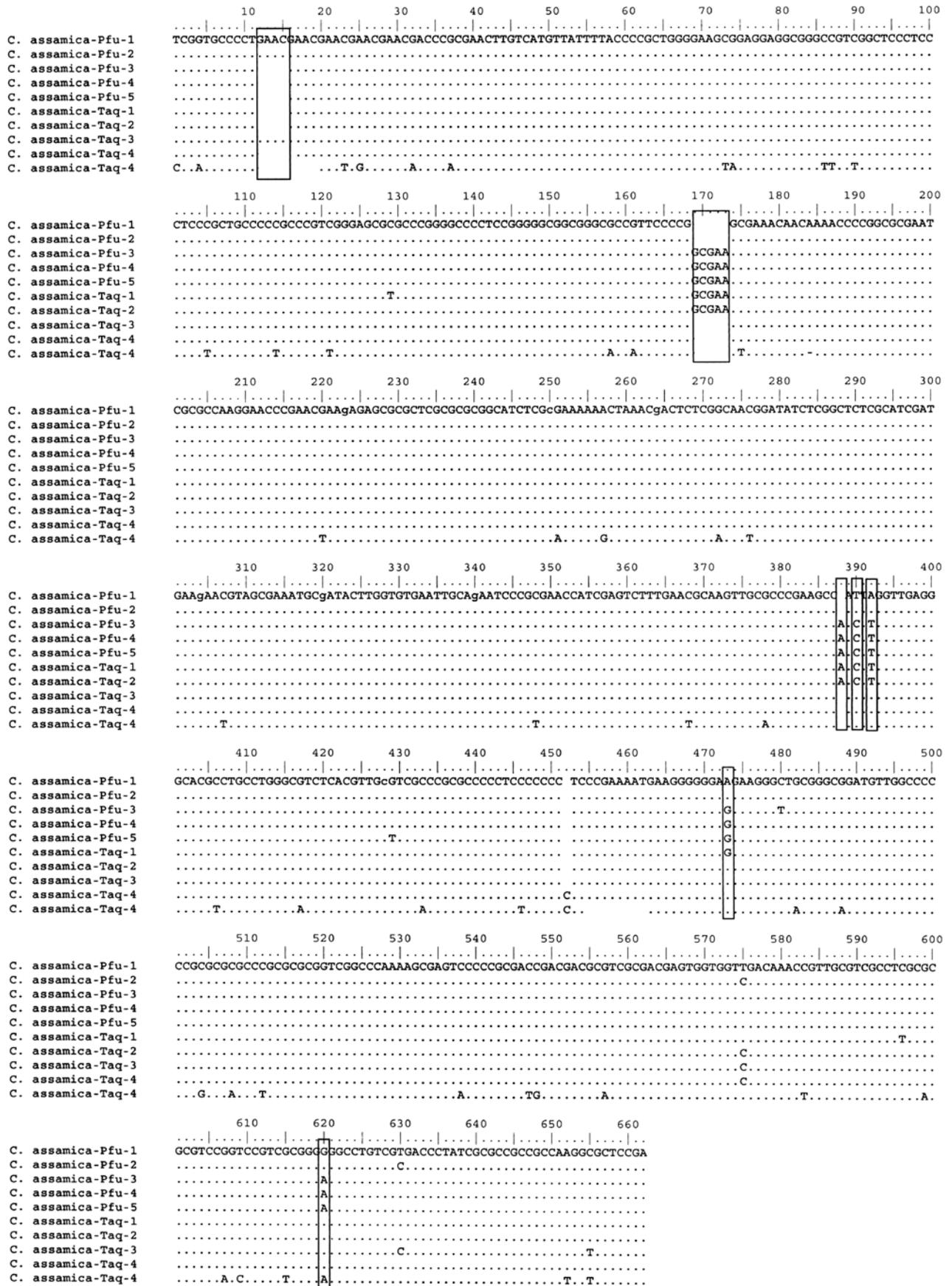


Figure 1. Intragenomic variations inherently present in the sequences of *Camellia assamica*; portions highlighted with rectangles.

To choose the nucleotide substitution model for the maximum likelihood analysis, the computer program Model Test 3.06 (Posada and Crandall, 1998) was used. This program employs two statistical methods, a likelihood ratio test (LRT) and an Akaike information criterion (AIC, Akaike, 1974), to find out the best fitting model to be used for the subsequent analysis with PAUP\* 4.0b10. Based on the results, the likelihood settings from the best-fit model (GTR+G) were selected and implemented in a maximum likelihood analysis with PAUP\* 4.0b10 (Swofford, 2001). A heuristic search, containing 100 random taxon-addition replicates, TBR branch swapping and MulTrees, Collapse and Steepest Descent options were in effect, was conducted with no upper limit imposed on the trees held in memory. To ascertain the relative degree of support for branches in the cladograms, jackknife (Farris et al., 1996) support was estimated with 100 replicates using heuristic searches and a random addition of sequences. Trees were rooted using *Pyrenaria melanogaster* as outgroup since *Pyrenaria* (*Tutcheria*) is so far known as the sister genus of *Camellia* (Prince and Parks, 2001).

## RESULTS AND DISCUSSION

### DNA quality

The DNA isolated from the dry leaf samples of *Camellia* using our protocol was of high quality as is evident from spectrophotometric and gel electrophoresis analyses. The spectrophotometric analysis revealed that the absorption ratio of the DNA at 260/280 nm was in the range of 1.84 (*C. vietnamensis*) and 1.89 (*C. formosensis*). A nearly 2.0 ratio indicates little contamination from proteins, polyphenols, or polysaccharides in the DNA. Electrophoresis of the DNA on agarose gel revealed a conspicuous band nearly 22 kb in size with a little shearing of the DNA. This may have resulted from the purification with the GENE CLEAN<sup>®</sup> III kit as shearing of DNA with higher molecular weight (>10 kb) is a problem often encountered with this kit (Application manual; GENE CLEAN<sup>®</sup> III kit, BIO101 Systems, New York). This purification step was, however, essential as the DNA obtained either from the CTAB extraction method alone or from the method of Struwe et al. (1998) alone was not amplified by PCR. Further, it should be noted that this little shearing of the DNA has not interfered with the subsequent PCR amplification of the DNA. The major components we added into the extraction buffer were PVP, CTAB, and PEG8000 to remove polyphenolic compounds, polysaccharides, and proteins. PVP is reported to form complex hydrogen bonds with phenolic compounds and co-precipitates with cell debris upon cell lyses, and upon centrifugation in the presence of chloroform the PVP complexes accumulate at the interface between the organic and the aqueous phases (Kim et al., 1997; Barnwell et al., 1998). Similarly, CTAB, a cationic detergent which solubilizes membranes, binds to fructans and other polysaccharides to form complexes that are removed during chloroform extraction. PEG-8000 also removes

proteins and polysaccharides (Agudo et al., 1995). Therefore, in our protocol most of the phenolics and polysaccharides must have been removed in the first phase of the extraction. In the second phase, the purification steps with GENE CLEAN<sup>®</sup> III kit removed the remaining proteins, polyphenolics, and polysaccharides along with tannins. This kit allows the DNA to bind onto the silica particles present in the EZ-GLASSMILK under high salt concentrations and to release from them under low salt concentrations. The DNA, while adhering to the silica particles, is washed many times to remove all contaminants from it. Thus at the end, though the DNA was devoid of most contaminants, the process left it slightly sheared. Singh et al. (1999) also reported high degradation of DNA upon its extraction from black tea, attributing it to the extreme processing of the tea leaf and to the binding of some of the phenolic compounds to the DNA upon cell lyses as reported earlier (John, 1992). Nevertheless, in the present study DNA shearing was insufficient to interfere with PCR amplification.

### PCR amplifications with *Taq* and *Pfu* polymerases

The DNA extracted with our method was successfully amplified with *Taq* and *Pfu* DNA polymerases. Among the 70 clones, i.e., five clones from each of the *Taq* and *Pfu* PCR amplifications for each of the seven samples, 63 clones were successfully sequenced (Table 1). However, attempts to sequence the PCR products directly, without cloning, were mostly unsuccessful. Sequences of the five clones from each PCR amplification showed a certain degree of intragenomic variability in the form of indels and substitutions (Table 2). Those variations present among the clones could be the main reason direct sequencing of the PCR products failed. Further, it is obvious that when sequences amplified with the *Taq* polymerase were compared with *Pfu*-amplified ones, the former exhibited much higher intragenomic variability though *Taq* polymerase is much commonly used in routine studies than *Pfu* enzyme. The effect of the two DNA polymerases on the sequence variability will be explicated in the next paragraph.

The length of the ITS1-5.8S-ITS2 regions, of the seven taxa studied, varied from 556 bp in *Camellia sasanqua* to 682 bp in *C. microphylla* when amplified with *Taq* DNA polymerase and varied from 634 bp in *C. sinensis* to 666 bp in *C. microphylla* when amplified with *Pfu* polymerase (Table 1). The average length of sequences with *Taq* was 643.52, and with *Pfu* 644.51. The average G+C content of the sequences with *Taq* was 66.12, and with *Pfu* 67.57. Thus, both the length and G+C content of the sequences amplified by *Pfu* polymerase were slightly higher.

### Sequence variability

The substitution rate (cumulative transitions and transversions) and indels were higher among *Taq*-amplified sequences than among *Pfu*-amplified sequences

(Table 2). The average substitution+indel in sequences amplified with *Taq* polymerase was highest in *C. euphlebica* (15.8 per sequence) and was least in *C. vietnamensis* (3.5 per sequence) while the same in *Pfu* amplified sequences was highest in *C. sinensis* (2.8 per sequence) and least in *C. euphlebica* (0.2 per sequence). When all the *Taq*- and *Pfu*-amplified sequences of the same sample aligned together and a consensus sequence was generated for each species, the most common variations observed were C->T and G->A in both *Taq*- and *Pfu*- amplified sequences. Hofreiter et al. (2001) also found such biased substitutions in the PCR products of DNA from ancient samples. They found that a high frequency of deoxyadenosine residues was incorporated to opposite positions where the template carries deoxycytidine residues. One of the reasons for such a high rate of G->A and C->T substitution is the tendency of the *Taq* polymerase to add deoxyadenosine residues when it reaches the end of templates (Clark, 1988). This has been shown to cause substitutions opposite to deoxycytidine residues when degraded DNA or DNA with high contaminants is used for PCR amplification (Kwok et al., 1990; Pääbo et al., 1990). Another reason for these deamination-like substitutions in sequences amplified by *Taq* polymerase could be the preferential amplification of templates with methylated residues. Methylation has been implicated in nucleolar dominance phenomena, and the resulting non-functional rDNA copies (pseudogenes) therefore typically show patterns of deamination-like substitutions (C->T and G->A) at methylation sensitive sites relative to their active counterparts (Muri et al., 2001; Marquez et al., 2003). Also noteworthy is that a higher occurrence of C->T and G->A substitutions was noticed in this study when G and C were consecutive, either in GC or

CG order, in both *Taq* and *Pfu*-amplified sequences.

In order to prevent the inherently-present sequence variations from influencing the rate of PCR/sequencing errors, we identified and removed the variable sites resulting from the intragenomic variations inherently present in the nrDNA arrays. Out of the seven species studied, three, namely *C. assamica*, *C. microphylla*, *C. sasanqua*, showed definite patterns of variations. The variations at the bases 12 to 15, 169 to 173, 388, 390, 392, 473, and 620 (Figure 1) in *C. assamica*, at 9 to 12 in *C. microphylla*, and at 111 and 654 in *C. sasanqua* were considered variations inherently present in the nrITS arrays, for they were consistently present in more than two sequences amplified by *Pfu*-DNA polymerase. Therefore, these variables sites were excluded from further analyses of PCR/sequencing errors.

The mean pair-wise distance (kimura-2-parameters) revealed that the highest distance was observed among sequences amplified with *Taq* polymerase (Table 2). The number of indels present among the clones of each species varied from 3.0 to 28.0 for *Taq* and 0.0 to 5.0 for *Pfu*. This high frequency of indels in *Taq*-amplified sequences must be due to PCR error. The PCR error rates also showed considerable differences between *Taq*- and *Pfu*-amplified sequences. The error rate in the *Taq*-amplified sequences varied from  $2.3 \times 10^{-4}$ /site/duplication in *C. vietnamensis* to  $1.4 \times 10^{-3}$ /site/duplication in *C. euphlebica* while in the *Pfu*- amplified sequences it varied only from  $0.0 \times 10^{-5}$ /site/duplication in *C. euphlebica* to  $1.5 \times 10^{-4}$ /site/duplication in *C. sinensis* var. *sinensis* (Table 3). This difference in the error rates of *Taq* and *Pfu* polymerases corroborates the earlier reports that the error rate of *Pfu* polymerase

**Table 2.** Variations in length and G-C content of the sequences amplified with *Taq* and *Pfu* DNA polymerases.

Sample	PCR enzyme	No. sequences obtained	Range of length (bp)	Range of G-C content (%)	Indels	Substitutions		
						ti*	tv	Total
<i>C. euphlebica</i>	<i>Taq</i>	4	633-664	62.17-67.32	6	42	15	57
	<i>Pfu</i>	3	647-649	66.72-66.77	1	0	0	0
<i>C. microphylla</i>	<i>Taq</i>	5	650-682	66.86-68.46	28	14	9	23
	<i>Pfu</i>	5	653-666	67.84-68.58	3	2	1	3
<i>C. assamica</i>	<i>Taq</i>	5	640-658	62.97-68.13	6	17	4	21
	<i>Pfu</i>	5	655-657	67.73-68.09	3	3	1	4
<i>C. formosensis</i>	<i>Taq</i>	5	632-648	60.44-68.36	7	19	4	23
	<i>Pfu</i>	5	646	67.96-68.27	0	1	0	1
<i>C. sinensis</i> var. <i>sinensis</i>	<i>Taq</i>	4	616-642	62.05-68.14	14	20	6	26
	<i>Pfu</i>	4	634-639	67.87-68.45	5	4	2	6
<i>C. sasanqua</i>	<i>Taq</i>	4	556-665	68.20-68.50	25	20	6	26
	<i>Pfu</i>	5	657-661	67.37-68.08	3	3	0	3
<i>C. vietnamensis</i>	<i>Taq</i>	4	635-664	66.67-67.62	5	7	2	9
	<i>Pfu</i>	5	635	66.77-67.24	0	3	1	4

\*ti: transition; tv: transversion.

**Table 3.** Sequence divergence calculated using Kimura-2-parameter model the PCR error rates in the sequences amplified with *Taq* and *Pfu* DNA polymerases.

Species	DNA polymerase	No. of clones	Kimura-2-parameter distance	Rate of PCR error*
<i>C. euphlebia</i>	<i>Taq</i>	4	0.097±0.009	1.4×10 <sup>-3</sup>
	<i>Pfu</i>	3	0.000±0.000	0.0×10 <sup>-5</sup>
<i>C. microphylla</i>	<i>Taq</i>	5	0.035±0.005	4.4×10 <sup>-4</sup>
	<i>Pfu</i>	5	0.005±0.002	6.0×10 <sup>-5</sup>
<i>C. assamica</i>	<i>Taq</i>	5	0.033±0.005	4.2×10 <sup>-4</sup>
	<i>Pfu</i>	5	0.006±0.002	8.1×10 <sup>-5</sup>
<i>C. formosensis</i>	<i>Taq</i>	5	0.085±0.010	4.7×10 <sup>-4</sup>
	<i>Pfu</i>	5	0.002±0.001	2.0×10 <sup>-5</sup>
<i>C. sinensis</i> var. <i>sinensis</i>	<i>Taq</i>	4	0.052±0.007	6.8×10 <sup>-4</sup>
	<i>Pfu</i>	4	0.009±0.003	1.5×10 <sup>-4</sup>
<i>C. sasanqua</i>	<i>Taq</i>	4	0.044±0.007	6.8×10 <sup>-4</sup>
	<i>Pfu</i>	5	0.006±0.002	6.07×10 <sup>-5</sup>
<i>C. vietnamensis</i>	<i>Taq</i>	4	0.017±0.004	2.5×10 <sup>-4</sup>
	<i>Pfu</i>	5	0.006±0.002	8.4×10 <sup>-5</sup>

\*Rate of nucleotide substitution/site/per duplication, estimated after eliminating the variable sites that are inherently present in the sequences.

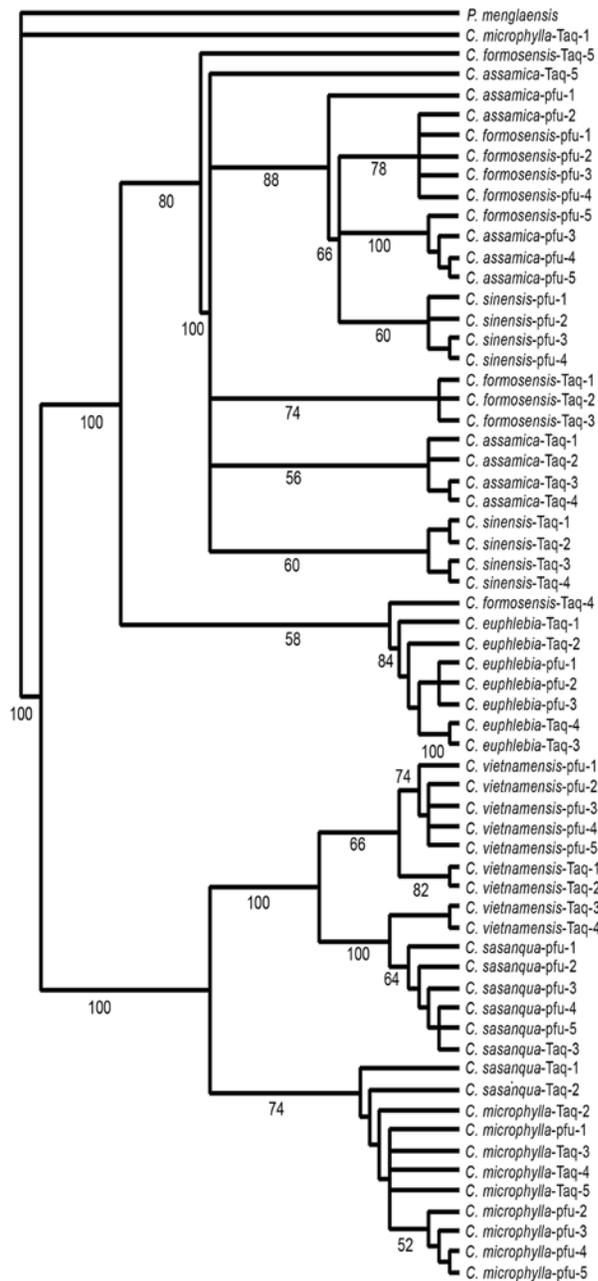
was 5 to 30 fold lower than the error rates of other proof reading and non-proofreading enzymes (Cline et al., 1996; Flaman et al., 1994). It has been noticed as well that the *Taq* polymerase has problems in amplifying regions with high G+C contents (Innis et al., 1988). For instance, while using *Taq* polymerase to amplify a G+C-rich region of the human 18S rRNA gene, Cariello et al. (1991) found formation of deletions in the hairpin regions. This finding supports an earlier report that using *Taq* polymerase in PCR amplification of sequences with secondary structures may give rise to difficulties in the form of deletions and amplification of unknown regions (Triglia et al., 1988; Ochman et al., 1988). Since the ITS regions of *Camellia* have an average of 66% GC content, which is towards the high end of the range recorded for plants (Baldwin et al., 1995), and do form secondary structures during amplifications, the aforementioned problems in the PCR amplification with *Taq* polymerase can be anticipated. We must note that the PCR error rates of the *Pfu* amplified sequences in the seven *Camellia* samples ( $0.0 \times 10^{-5}$ ,  $6.0 \times 10^{-5}$ ,  $8.1 \times 10^{-5}$ ,  $2.0 \times 10^{-4}$ ,  $1.5 \times 10^{-5}$ ,  $6.07 \times 10^{-5}$ ,  $8.4 \times 10^{-5}$ ) were higher than the commercially reported error rate of  $1.3 \times 10^{-6}$  for *Pfu* DNA polymerase (Slater et al., 1998). This could be due to the high G+C content and the tendency to form secondary structures in the nrITS of *Camellia*, as the commercially reported error rate was estimated from the *E. coli* lac I gene, which is devoid of the above problems. Another problem with the *Camellia* nrITS is the presence of mutational hot spots such as mononucleotide stretches and microsatellites, which

are particularly prone to PCR induced errors (Lapegu et al., 1999). An error rate of one in 1000 bp at each cycle results in one error in 400 bp in the final product after 25 cycles (Saiki et al., 1988). Hence, Kobayashi et al. (1999) opined that to obtain sequence data free from PCR error, sequences for several clones are to be made and compared. In the light of these factors, we stress that the use of *Pfu* is essential to reducing PCR errors in GC-rich regions like the nrITS of *Camellia*. Furthermore, the findings of this study may very well explain the failure of previous attempts to sequence nrITS in *Camellia* if they were performed by routinely used *Taq* polymerase on low quality DNA.

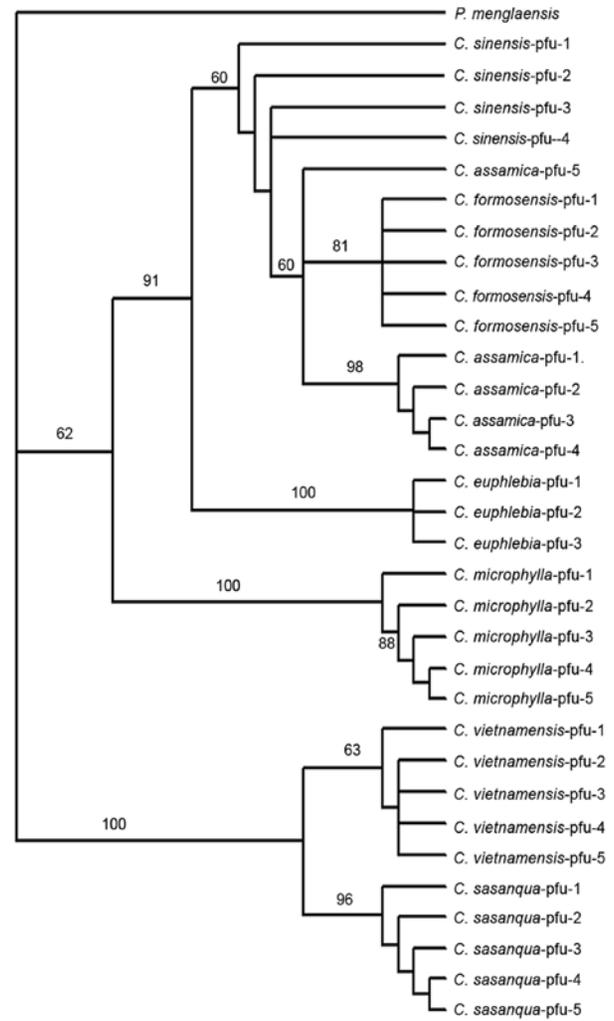
### Phylogenetic applications

Phylogenetic analyses were made to examine the usefulness of nrITS in *Camellia* and the impact of different DNA polymerases on the analyses. The consequence of not eliminating PCR-induced variability from the phylogenetic analysis was clear from the gene trees (Figures 2, 3). From the gene tree obtained from both *Taq* and *Pfu* DNA polymerases (Figure 2), it could be observed that those clones amplified with *Taq* polymerase showed different patterns of grouping. Clones of the same species either joined the sequences of other species or remained as singletons. For instance, the clone *C. formosensis*-4 *Taq* joined with the clade formed by clones of *C. euphlebia*. Similarly, the clones *C. assamica*-3 *Taq*, *C. formosensis*-5 *Taq*, *C. sasanqua*-1 *Taq*, and *C. microphylla*-5 *Taq* remained as isolates showing no definite relationship

with any other groups. In contrast, all clones of the same species amplified with *Pfu* polymerase grouped together into one clade. Thus, it is clear that the *Taq* DNA polymerase induced PCR errors in the sequences and greatly affected the topology of the cladogram. When the *Pfu*-amplified sequences were analyzed separately, all clones of the same taxon grouped together, except in *C. sinensis* (Figure 3). In this gene tree depicting the interspecies relationships of the seven species, four sections were well demonstrated. Species of *C. assamica*, *C. formosensis*, and *C. sinensis* belonging to Section



**Figure 2.** Maximum likelihood 50% majority rule consensus tree derived from the sequences of seven species of *Camellia* amplified with both *Taq* and *Pfu* polymerases.



**Figure 3.** Maximum likelihood 50% majority rule consensus tree derived from the sequences of seven species of *Camellia* amplified with only *Pfu* polymerases.

*Thea* are so closely related that they are often treated as three varieties under *C. sinensis* (Su et al., 2007). The 14 sequences of these three species formed one clade, with the five sequences of *C. formosensis* and four of the five sequences of *C. assamica* forming their subgroups. The clones of *C. sinensis* were not held together, and this was not surprising as *C. sinensis* has a long history of domestication, naturalization, and cultivation. The two species, *C. sasanqua* and *C. vietnamensis*, belonging to the same section viz., sect. *Oleifera*, are grouped together in a clade with 100% support and sequences of each species further joined together as subgroups. As for *C. euphlebica* and *C. microphylla*, they belong to the sections *Chrysantha* and *Paracamellia*, respectively (Chang, 1998), and their sequences form distinct clades. In summary, the four sections (*Chrysantha*, *Oleifera*, *Paracamellia*, and *Thea*) as well as the species within the sections can be well separated in the gene tree generated by the *Pfu*-DNA polymerase amplified sequences.

In short, this investigation has demonstrated that molecular phylogeny with nrITS has great potential to resolve many vexing problems associated with both intersectional and intrasectional relationships of the *Camellia*. However, this is possible only if high quality DNA can be extracted from the dried leaf samples and proofreading enzyme like *Pfu* polymerase can be used for PCR amplification to reduce the PCR/sequencing errors. The DNA extraction protocol developed in this study could be of much use to this end. Furthermore, owing to the possibility of pseudogenes being present in the genome, sequences from multiple clones have to be used to generate a gene tree to distinguish the functional genes from the pseudogenes and to minimize the interference of pseudogenes in the phylogenetic analysis.

**Acknowledgement.** We would like to thank Mong-Huai, Su, Hong Wong and Wen-Ju Zhang for assistance with sample collection and Yaw-wen Yang for technical advice. This research was supported by the National Science Council, Taiwan, Republic of China (Project NSC95-2621-B-001-009-MY2).

## LITERATURE CITED

- Agudo, L.C., P. Gavidia, P. Perez-Bermudez, and J. Segura. 1995. PEG precipitation, a required step for PCR amplification of DNA from wild plants of *Digitalis obscura* L. *BioTechniques* **18**: 766-768.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Contr.* **19**: 716-723.
- Alvarez, I. and J.F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* **29**: 417-434.
- Baldwin, B.G., M.J. Sanderson, J.M. Porter, M.F. Wojciechowski, C.S. Campbell, and M.J. Donoghue. 1995. The ITS regions of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann. Missouri Bot. Gard.* **82**: 247-277.
- Barnwell, P., A. N. Blanchard, J.A. Bryant, N. Smirnov, and A.F. Weir. 1998. Isolation of DNA from highly mucilaginous succulent plant *Sedum telephium*. *Plant. Mol. Biol. Repr.* **16**: 133-138.
- Buckler, E.S., A. Ippolito, and T.P. Holtsford. 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* **145**: 821-832.
- Cariello, N.F., W.G. Thilly, J.A. Swenberg, and R.S. Thamas. 1991. Deletion mutagenesis during polymerase chain reaction: dependence on DNA polymerase. *Gene* **99**: 105-108.
- Carlson, J.E., L.K. Tulsieram, J.C. Glaubitz, V.W. K. Luk, C. Kauffeldt, and R. Rutledge. 1991. Segregation of random amplified DNA markers in F1 progeny of conifers. *Theor. Appl. Genet.* **83**: 194-200.
- Chang, H.T. 1981. A taxonomy of the genus *Camellia*. *Acta Scientiarum Naturalium. Universitatis. Sunyatseni*, Monographic series **1**: 1-180.
- Chang, H.T. 1998. *Flora of Reipublicae Popularis Sinicae, Delectis Florae Republicae Popularis Sinicae, Agendae Academiae Sinicae Edita, Tomus*, Beijing Science Press, China, Vol. 49, pp. 101-113.
- Clark, J.M. 1988. Novel non-templated addition reactions catalyzed by prokaryotic and eukaryotic DNA polymerases. *Nucleic Acids Res.* **16**: 9677-9686.
- Cline, J., J.C. Barman, and H.H. Hogrefe. 1996. PCR fidelity of *Pfu* polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* **24**: 3546-3551.
- Dellaporta, S.L., J. Wood, and J.B. Hicks. 1983. A plant DNA mini preparation a. Version II. *Plant Mol. Biol. Rep.* **1**: 19-23.
- Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-15.
- Fang, G., S. Hammer, and R. Groumet. 1992. A quick and inexpensive method for removing polysaccharides from plant genomic DNA. *Biofeedback* **13**: 52-54.
- Farris, J.S., V.A. Albert, M. Kallersjo, D. Lipscomb, and A.G. Kluge. 1996. Parsimony jackknifing outperforms neighboring. *Cladistics* **12**: 99-124.
- Feliner, G.N. and Rossello. 2007. Better the devil you know? Guidelines for insightful utilization of nrITS in species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* **44**: 911-919.
- Flaman, J.M., T. Frebourg, V. Moreau, F. Charbonnier, C. Martin, C. Ishioka, S.H. Friend, and R. Iggo. 1994. A rapid PCR fidelity assay. *Nucleic Acids Res.* **22**: 3259-3260.
- Gelfand, D.H. and T.J. White. 1990. Thermostable DNA polymerases. *In* M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White (eds.), *PCR Protocols: a Guide to Methods and Applications*. Academic Press, London, pp. 129-141.
- Graham, H.N. 1992. Green tea composition, consumption, and polyphenols chemistry. *Prev. Med.* **21**: 334-350.
- Hall, T.A. 1999. BioEdit: a user friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic. Acid. Sym. Ser.* **41**: 95-98.
- Hofreiter, M., V. Jaenicke, D. Serre, A. Haeseler Av. and S. Pääbo. 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**: 4793-4799.
- Innis, M.A., K. B. Mayambo, D.H. Gelfand, and M.A. Brow. 1988. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of DNA polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci. USA* **85**: 9436-9440.
- John, M.E. 1992. An efficient method for isolation of RNA and DNA from plants containing polyphenolics. *Nucleic Acids. Res.* **20**: 2381.
- Kim, C.S., C.H. Lee, J.S. Shin, Y.S. Chung, and J.E. Hyung. 1997. A simple and rapid method of isolation of high quality genomic DNA from fruit trees and conifers using PVP. *Nucleic Acids Res.* **27**: 66-58.

- Kobayashi, N., K. Tamura, and T. Aotsuka. 1999. PCR error and molecular population genetics. *Biochem. Genetics* **3**: 318-321.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**: 150-163.
- Kwiatowski, J., D. Skarecky, S. Hernandez, D. Pham, F. Quijas, and F.J. Ayala. 1991. High fidelity of polymerase chain reaction. *Mol. Biol. Evol.* **8**: 884-887.
- Kwok, S., D.E. Kellog, N. McKinney, D. Spasic, L. Goda, C. Levenson, and J.J. Sninsky. 1990. Effects of primer-template mismatches on polymerase chain reaction: human immunodeficiency virus type I model studies. *Nucleic Acids Res.* **18**: 999-1005.
- Lapegue, S.D., M.H. Pemonge, L. Gielly, P. Taberlt, and J. Petit. 1999. Amplification of oak DNA from ancient and modern wood. *Mol. Ecol.* **8**: 2137-2140.
- Lundberg, K.S., D.D. Shoemaker, M.W.W. Adams, J. M. Short, J.A. Sogge, and E.J. Marhur. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* **108**: 1-6.
- Marquez, L.M., D.J. Miller, J.B. MacKenzie, and M.J.H. van Oppen. 2003. Pseudogenes contribute to the extreme diversity of nuclear ribosomal DNA in the hard coral *Acropora*. *Mol. Bio. Evol.* **20**: 1077-1086.
- Matsumoto, S., A. Takeuchi, M. Hayatsu, and S. Kondo. 1994. Molecular cloning of phenyle alanine ammonia lyase cDNA and classification of varieties and cultivars of tea lants using the tea PAL cDNA as a probe. *Theor. Appl. Genet.* **89**: 671-675.
- Merlo, D.J. and J.D. Kemp. 1976. Effect of polysaccharides on kinetics of DNA. *Plant Physiol.* **58**: 1522-1526.
- Ming, T. 2000. Monograph of the genus *Camellia*. Kunming Institute of Botany, Chinese Academy of Sciences, Yunnan Science and Technology Press, Kunming, P. R. China.
- Muri, G., C.C. Fleming, and C. Schlotterer. 2001. Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Mol. Biol. Evol.* **18**: 112-119.
- Murray, M.G. and W.F. Thompson. 1980. Rapid isolation of high-molecular-weight plant DNA. *Nucleic Acids Res.* **8**: 4321-4325.
- Ochman, H., A.S. Gerber, and D.L. Hartl. 1988. Genetic application of an inverse polymerase chain reaction. *Genetics* **120**: 621-623.
- Orel, G., A. Marchent, and G. Richards. 2003. Evolutionary relationships of yellow-flowered *Camellia* species from southeast Asia. *Inter. Camellia J.* **35**: 88-96.
- Pääbo, S., D.M. Irwin, and A.C. Wilson. 1990. DNA damage promotes jumping between templates during enzymatic amplification. *J. Biol. Chem.* **265**: 4718-4721.
- Posada, D. and K.A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.
- Prince, L.M. and C.R. Parks. 2001. Phylogenetic relationships of Theaceae inferred from chloroplast DNA sequences data. *Amer. J. Bot.* **88**: 2309-2320.
- Rogers, S.U. and A.J. Bendich. 1988. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant. Mol. Biol.* **5**: 69-76.
- Saiki, R.K., D.H. Geifand, S. Stoffei, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mulfis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.
- Sealy, J.R. 1958. A Revision of genus *Camellia*. The Royal Horticultural Society, pp. 168-191.
- Shioda, M. and K. Marakami-Muofushi. 1987. Selective inhibition of DNA polymerase by a polysaccharide purified from slime of *Physarum polycephalum*. *Biochem. Biophys. Res. Commun.* **146**: 61-66.
- Singh, M., Bandana, and P.S. Ahuja. 1999. Isolation and PCR amplification of Genomic DNA from market samples of dry tea. *Plant. Mol. Biol. Repr.* **17**: 171-178.
- Slater, M., S. Selman, B.L. Mogilevsky, H. Ammons, and J. Hartnett. 1998. Pfu DNA polymerase: A high fidelity enzyme for nucleic acid amplification. *Promega Notes* **68**: 7-12.
- Struwe, L., M. Thiv, J.Kadereit, T. Motley, A.S.R. Pepper, J. Rova, K. Potgieter, P.White, and V.A. Albert. 1998. *Saccifolium*, an endemic of Sierra de La Neblina on the Brazillian-Venezuelan frontier is related to a temperate alpine lineage of Gentianaceae. *Harvard Papers in Bot.* **3**: 199-214.
- Su, M.H. 2007. Taxonomic study of *Camellia formosensis* (Masamune et Suzuki). In M. H. Su, C. F. Hsieh et C. H. Tsou (Theaceae) (eds.), PhD dissertation, Institute of Ecology and Evolutionary Biology College of Life Science, National Taiwan University, Taiwan.
- Swofford, D.L. 2001. PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
- Tai, T.H. and S.D. Tanksley. 1990. A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. *Plant. Mol. Biol. Repr.* **8**: 297-303.
- Takeuchi, A., S. Matsumoto, and M. Hayatsu. 1994. Chalcone synthase from *C. sinensis*: isolation of cDNA and the organ specific and sugar responsive expression of genes. *Plant Cell Physiol.* **35**: 1011-1018.
- Tamura, K. and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**: 512-526.
- Tang, S. and Y. Zhong. 2002. A phylogenetic analysis of nrDNA ITS sequences from Ser. *Chrysanth* (Sect. *Chrysanth*, *Camellia*, Theaceae). *J. Genet. Mol. Biol.* **103**: 105-107.
- Thakor, B.H. and C.R. Parks. 1997. Phylogenetic relationships within the genus *Camellia* (Theaceae). Supplement to *Amer. J. Bot.* **84**: 237 (Abstract only).
- Tindall, K.R. and T.A. Kunkel. 1988. Fidelity of DNA synthesis

- by the *Thermus aquaticus* DNA polymerase. *Biochemistry* **27**: 6008-6013.
- Triglia, T., M.G. Peterson, and D.J. Kemp. 1988. A procedure for *in vitro* amplification of DNA segments that lay outside the boundaries of known sequences. *Nucleic Acids Res.* **16**: 8186.
- Urbatsch, L.E., B.G. Baldwin, and M.J. Donoghue. 2000. Phylogeny of coneflowers and relatives (Heliantheae: Asteraceae) based on nuclear DNA internal transcribed spacer (ITS) sequences and chloroplast DNA restriction site data. *Syst. Bot.* **25**: 1153-565.
- Wachira, F.N., R. Waugh, C.A. Hackett, and W. Powell. 1995. Detection of genetic diversity in tea using RAPD markers. *Genome* **38**: 201-210.
- White, T.J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White (eds.), *PCR Protocols: a Guide to Methods and Applications*. Academic Press, London, pp. 315-322.
- Xiao, T.J. 2001. Phylogenetic relationships of the genus *Camellia* (Theaceae) based on the RNA polymerase II (RPB2) sequences. PhD dissertation, Univ. North Carolina, Chapel Hill.
- Xiao, T.J. and C.R. Parks. 2003. Molecular analysis of the genus *Camellia*. *Inter. Camellia J.* **35**: 57-65.
- Yang, J.B., H.T. Li, S.X. Yang, D.Z. Li, and Y.Y. Yang. 2006. The application of four DNA sequences to studying molecular phylogeny of *Camellia* (Theaceae). *Acta Bot. Yunn.* **28**: 108-114.

## 以 nrITS 研究山茶屬分子親緣關係的技術報告

Kunjupillai Vijayan 鄒稚華

中央研究院 植物暨微生物研究所

核 DNA 的 ITS 片段被廣泛應用在生物親緣關係的重建上，特別是在種的層級。過去曾有許多應用 ITS 於山茶屬的研究，都由於操作過程的困難未能成功。本工作欲克服技術上的障礙，並了解 ITS 解析山茶屬內種間關係的有效性，因此選擇了七個親近或疏離的種進行測試。本研究顯示自山茶屬乾燥葉片中萃取高純度 DNA 的困難度很高卻極為重要，因此發展出一套有效的操作流程，可以萃取純度依 A260/280 讀值高達 1.84 至 1.89 的 DNA。再者，在 PCR 增殖反應中加入 *Taq* 或 *Pfu* DNA 聚合酶進行比較，顯示前者會造成 PCR 產物中相對而言明顯偏高的誤差，因此使用 *Pfu* DNA 聚合酶雖然昂貴卻屬必要。親緣關係的分析則顯示 ITS 片段的 DNA 序列分析能明確顯示這七個種的遠近關係，是分析山茶屬種間關係非常有效的片段。本工作已持續在近 100 種山茶屬植物中順利定序，由於操作技術及分類學理上複雜性均高，需以系列方式報導，本文為第一篇，屬技術性報告。

**關鍵詞：**山茶屬；分子親緣探討；DNA 萃取；ITS；*Pfu*, *Taq* 聚合酶；PCR 增殖誤差。