

Phylogenetic Relationships of the Genus *Taxus* Inferred from Chloroplast Intergenic Spacer and Nuclear Coding DNA

Da Cheng HAO,^{a,c} BeiLi HUANG,^b and Ling YANG^{*,a}

^aLaboratory of Pharmaceutical Resource Discovery, Biotechnology Division, Dalian Institute of Chemical Physics, Chinese Academy of Sciences; Dalian 116023, China; ^bLuShan Botanical Garden, Chinese Academy of Sciences; JiangXi 332900, China; and ^cThe Graduate University, Chinese Academy of Sciences; Beijing 100049, China.

Received September 10, 2007; accepted November 30, 2007; published online November 30, 2007

A cladistic analysis of the medicinal plant *Taxus*, using the sequences of one chloroplast (*trnS-trnQ* spacer) and three nuclear taxadiene synthase (*TS*), 10-deacetylbaaccatin III-10 β -*O*-acetyltransferase (*DBAT*), and 18S *rDNA* molecular markers, was carried out by distance, parsimony, likelihood, and Bayesian methods. Three of the four New World species (*T. brevifolia*, *T. floridana* and *T. globosa*) form a well-supported clade, whereas *T. canadensis* initially branches—appearing distantly related to both Old World taxa and New World species. In Asia, *Taxus chinensis*, *T. mairei*, *T. sumatrana* and *T. wallichiana* cluster together and are sister to a clade containing *T. baccata* and *T. contorta*. *Taxus yunnanensis* is more closely related to *T. wallichiana* than to four other *Taxus* species in our study from China; *T. contorta* is closer to the Euro-Mediterranean *T. baccata* than to the Asian species. This study provides a genetic method for authentication of economically important *Taxus* species and proposes a robust phylogenetic hypothesis for the genus. Using *trnS-trnQ* spacer sequences, we were able to distinguish *T. mairei* from all other species of *Taxus*.

Key words *Taxus*: phylogeny; *trnS-trnQ* spacer; 10-deacetylbaaccatin III-10 β -*O*-acetyltransferase; taxadiene synthase; 18S *rRNA*

Plants of the genus *Taxus* are sources of a number of physiologically and pharmacologically active compounds of different classes, especially the anti-cancer paclitaxel and many other taxane derivatives. There are at least ten species in *Taxus*. The species of *Taxus* are more geographically than morphologically separable (Fig. 1). There is a large variation in taxane content between the different species and cultivars.¹⁾ It is essential to find suitable plants for the various

production protocols of paclitaxel. The correct identification of the *Taxus* species is not only a prerequisite for the relevant plant breeding and selection, and Good Agricultural Practice (GAP), but also a precondition for the chemical and pharmacological investigations of the respective *Taxus* species, and Good Manufacturing Practice (GMP).^{2,3)} *Taxus mairei*, an endemic *Taxus* species of southern China, is the most important source plant for 7-xylosyltaxanes that can be converted to pa-

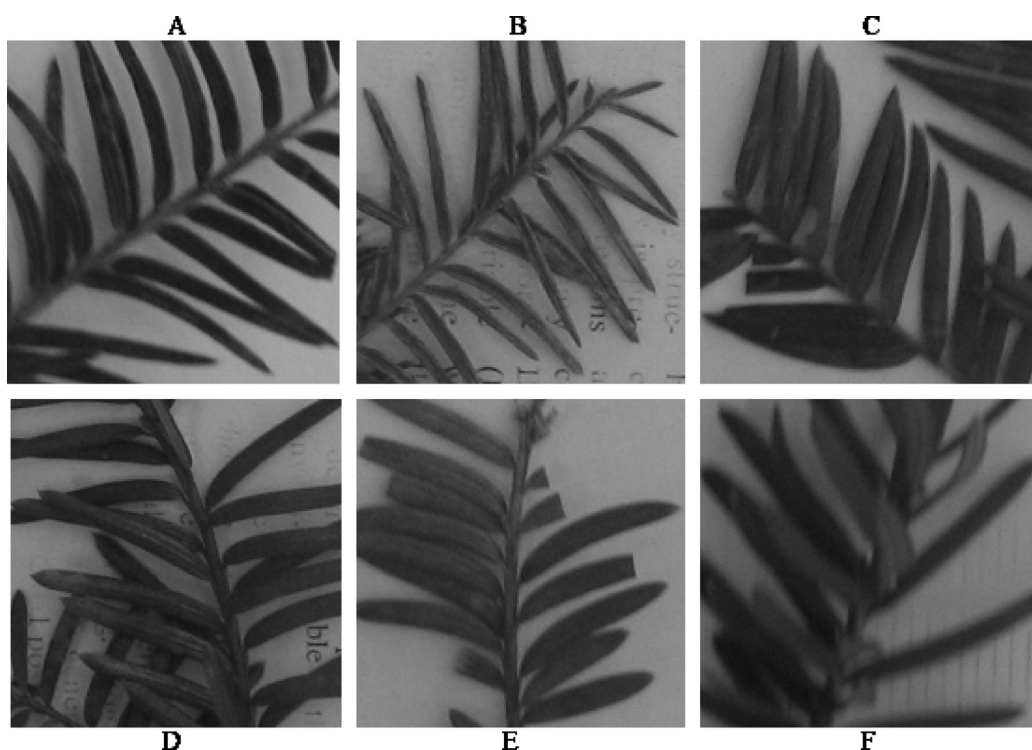


Fig. 1. Leaf Morphology of some Medicinal *Taxus* Species

(A) *T. baccata*; (B) *T. canadensis*; (C) *T. chinensis*; (D) *T. cuspidata*; (E) *T. mairei*; (F) *T. xmedia*.

* To whom correspondence should be addressed. e-mail: yling@dicp.ac.cn

clitaxel and other useful taxanes *via* chemical transformation and biotransformation. However, it is difficult to authenticate and differentiate *T. mairei* from other source plants of taxanes by morphology. Therefore, molecular studies will definitely be required to resolve the problem.

There is currently debate about the number of species, the classification, and the phylogenetic relationship. Ten,⁴⁾ nine,⁵⁾ and ten species,⁶⁾ and different varieties have been proposed by different authors. Farjon treated *T. mairei* as a variety of *T. chinensis*, while Fu *et al.* treated *T. mairei* and *T. chinensis* as varieties of *T. wallichiana*; the relationship between *T. contorta* and others has not been resolved; Spjut recognized *T. yunnanensis* as a variety of *T. wallichiana*, while it is not in the list of Farjon and Fu *et al.* regarded it as a synonym of *T. wallichiana*. *T. sumatrana* from Indonesia and the Philippines, has been recorded by de Laubenfels,⁷⁾ however, *T. sumatrana* is a broad concept which includes the taxa of Fu *et al.*'s treatment such as *T. wallichiana* var. *chinensis* and *T. wallichiana* var. *mairei*. The total number of *Taxus* occurring in Asia is quite controversial and in need of further study.

The objective of our study is to establish genetic profiles for economically valuable medicinal *Taxus* and to provide an assessment of phylogenetic relationships among species in *Taxus* using sequence data from the nuclear (nr) protein-coding genes taxadiene synthase (*TS*) and 10-deacetylbaaccatin III-10 β -O-acetyltransferase (*DBAT*), *18S rRNA* gene, and the chloroplast (cp) non-coding *trnS-trnQ* intergenic spacer. *18S rRNA* sequences have been used extensively for plant molecular phylogenetic analyses,^{8,9)} while the secondary metabolism genes *TS* and *DBAT* and the cp *trnS-trnQ* spacer are for the first time used in evaluating relationships within *Taxus* for comparison to morphology-based classifications and to geographic distribution. In paclitaxel biosynthetic pathway, *TS* catalyzes the cyclization of geranylgeranyl diphosphate to taxadiene and *DBAT* catalyzes the acetylation of 10-deacetylbaaccatin III to baaccatin III.¹⁰⁾ *trnS-trnQ* intergenic spacer is located in the large single copy region of the chloroplast genome, between *trnS* and *trnQ* genes. So far, only one study¹¹⁾ explicitly addressed the question of phylogenetic re-

lationships within *Taxus*, in which taxon sampling was too sparse, especially with respect to Asian species, to make further inferences about phylogenetic relationships. Moreover, the interspecific relationships within *Taxus* have not been addressed by the combined use of multiple molecular markers, including protein-coding gene, RNA-coding gene, and cp non-coding sequence. In the present molecular phylogeny, we extend the set of available *Taxus* molecular sequences to 21 taxa, representing all species. The present study reports the results of the combined analysis of four molecular markers to differentiate all *Taxus* species.

MATERIALS AND METHODS

Taxon Sampling Species, geographic origin of the sequenced material, their voucher numbers, and GenBank accession numbers of the sequences generated in this study, as well as those retrieved from GenBank, are given in Table 1. Twenty-five chloroplast and 41 nuclear sequences were newly generated for this study.

DNA-Extraction, -Amplification, and -Sequencing Fresh or silica-dried leaves were ground into powder. Genomic DNA was extracted by using Universal Genomic DNA Extraction kit (Takara, Dalian, China), following the manufacturer's protocol. A 0.9% agarose gel was run to assess the presence and integrity of the DNA. Quantification was done spectrophotometrically and the concentration of DNA ranged from 50–77 ng per μ l.

A 50 μ l PCR reaction mix consisted of 5 μ l of 10 \times reaction buffer, 4 μ l each of 2.5 mM dNTPs stock, 2.5 μ l of 10 μ M forward and reverse primers (synthesized by Takara, Dalian, China), 0.5 μ l bovine serum albumin (10 mg/ml), and 1.5 units of Ex *Taq* polymerase (Takara, Dalian, China). The primers¹¹⁾ used for amplification of *trnS-trnQ* are, *trnS*^{GGA}: ttaccgggggttcgaatccctc (forward, 5'–3') and 5'*trnL*^{UAA}R: tc-taccgatttcgcatatc (reverse). The primers used for amplification of *18S rDNA* are, F: tcaagattaagccatgcatgctc and R: caacatcgcgcgacggaaacca. The *TS* gene was amplified using atggctcagctctcatttaatgc (forward) and cgcagccgccaattgtcca

Table 1. Voucher Information and GenBank Accession Numbers for Taxa Used in This Study

Species, origin, voucher specimen, and GenBank accession number, <i>trnS-trnQ</i> , <i>18S rDNA</i> , <i>TS</i> , <i>DBAT</i>
Outgroup: <i>Amentotaxus argotaenia</i> , LuShan, JiangXi, China, Huang, LS20061201, EU107153, -, -, -; <i>Amentotaxus formosana</i> , TaiWan, China, Determann, ABG20010355 (ATLAN), -, D38248, -, -; <i>Torreya nucifera</i> , Japan, Quinn, 876809 (NSW), EU107154, D38249, -, -; <i>Torreya jackii</i> , HangZhou, ZheJiang, China, Zhang, HZ20070101 (HHBG), EU107155, -, -, -; <i>Cephalotaxus wilsoniana</i> , TaiWan, China, Determann, ABG20000105 (ATLAN), -, D38241, -, -; <i>Podocarpus macrophyllus</i> , HangZhou, ZheJiang, China, Zhang, HZ20070103, EU107152, -, -, -.
Ingroup: <i>Taxus</i> \times <i>hunnelliana</i> , Waterloo, Canada, WC001 (WAT), EF017313, EF017312, EU107120, EU107132; <i>T.</i> \times <i>hunnelliana</i> , Vancouver, Canada, La Fontaine, UBC200707 (UBC), EU107161, EU107119, EU107121, EU107133; <i>T. chinensis</i> , ShenNongJia, HuBei, China, SNJ001, DQ888590, AY544988*, AY007207*, EU107135; <i>T. chinensis</i> , YunXi, HuBei, China, YX001, EU107157, EU107111, EU107126, EU107142; <i>T.</i> \times <i>media</i> , Dalian, China, DICP001, DQ888587, AY544989*, AY461450*, EF028093; <i>T. cuspidata</i> , Ji'An, JiLin, China, JA001, DQ888591, DQ888593, DQ305407, AF193765*; <i>T. cuspidata</i> var. <i>nana</i> , Japan, DD001, EU107156, EU107115, EU107124, -; <i>T. yunnanensis</i> , ChaYu, Tibet, China, CY001, EU107150, EU107116, EU107129, EU107136; <i>T. baccata</i> , Oxford, U.K., Stevenson, 0000381 (OXF), EF017309, EF017310, AY424738*, AF456342*; <i>T. baccata</i> , Montreal, Canada, Baillieu, 1586–1978 (MTJB), EU107160, EU107110, EU107123, EU107141; <i>T. canadensis</i> , Montreal, Canada, Baillieu, 1960–2000 (MTJB), EF017308, EF017311, AY364470*, EU107134; <i>T. mairei</i> , LiShui, ZheJiang, China, LS001, DQ888589, D16445, AY931015*, AY365031*; <i>T. mairei</i> , NanPing, Fujian, China, NP001, EU107163, -, EU107127, EU107139; <i>T. mairei</i> , JiangXi, China, JX001, EU107159, -, EU107125, EU107140; <i>T. contorta</i> , JiLong, Tibet, China, JL001, EU107146, EU107108, EU107122, EU107138; <i>T. contorta</i> , JiLong, Tibet, China, JL002, EU107158, EU107109, EU107128, EU107137; <i>T. wallichiana</i> , ChaYu, Tibet, China, CYW001, EU107147, EU107112, EU107130, -; <i>T. sumatrana</i> , Unknown, Determann, ABG20051056 (ATLAN), EU107148, EU107113, EU107131, EU107144; <i>T. floridana</i> , Florida, U.S.A., Determann, ABG-TAFLTSP3B (ATLAN), EU107149, EU107114, -, -, -; <i>T. globosa</i> , Mexico, Determann, ABG19971263 (ATLAN), EU107151, EU107117, -, EU107145; <i>T. brevifolia</i> , Vancouver, Canada, La Fontaine, UBC20070701 (UBC), EU107162, EU107118, U48796*, EU107143.

All unmarked vouchers are deposited in Herbarium, LuShan Botanical Garden, Chinese Academy of Sciences, JiangXi, China (LUS). An asterisk indicates a sequence obtained from GenBank, and a dash indicates missing data.

(reverse). The *DBAT* gene was amplified using atggcaggct-caacagaatttg (forward) and tcaagtttagttacatattgtttg (reverse). Approximately 50 ng of genomic DNA was used as a template for the reaction. The reaction mixture was placed in a Takara PCR Thermal Cycler Dice (Takara, Japan). Cycling (38 cycles) condition was described previously.¹²⁾ The annealing temperatures were 53 °C (for *18S rDNA*, *TS*, and *trnS-trnQ*) or 49 °C (for *DBAT*). The PCR products were purified by Agarose Gel DNA Purification Kit (Takara).

All PCR products were subcloned into a TA cloning vector pMD19-T (Takara). The plasmids were purified for sequencing. ABI Prism, BigDye Terminator, and cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA, U.S.A.) were used for sequencing reaction with RV-M and M13-47 primers. The longer clones were sequenced by using oligonucleotides synthesized according to the sequences obtained by RV-M or M13-47 primers. The sequences were detected using an ABI Prism 377 Genetic Analyzer (Applied Biosystems). Two PCR products from one sample were sequenced twice in both directions.

Phylogenetic Analyses Sequence alignment was performed with CLUSTAL W. The aligned chloroplast and nuclear matrix comprised 2538 and 4052 positions, respectively. Each separate DNA region, as well as all combined data, was analyzed with Modeltest 3.8¹³⁾ to find the best model of evolution for the data (Table 2). Employing the Akaike information criterion (AIC), the model with the lowest AIC score was chosen.

Maximum likelihood (ML), Maximum parsimony (MP), and Neighbour joining (NJ) analyses were performed on the separate molecular partitions and on the combined data. ML analysis and bootstrapping were performed using GARLI 0.951.¹⁴⁾ GARLI searches relied on the GTR+G, GTR+I+G, and GTR+G models, which ModelTest selected as the best fitting models for unpartitioned cpDNA, nrDNA, and combined data, respectively. MP analysis was performed using PAUP* 4.0b10.¹⁵⁾ Heuristic searches were performed using tree bisection–reconnection branch-swapping and 10 random sequence addition replicates. All sites were equally weighted and gaps were treated as missing characters. NJ analysis was performed with MEGA4,¹⁶⁾ using Maximum Composite Likelihood method for estimating evolutionary distances between all pair of sequences simultaneously. Strong support for individual node is defined as nodes with Bayesian posterior probabilities (PP) ≥ 0.95 or non-parametric bootstrap (BP) ≥ 60–70. No strongly supported conflicting relationships were recovered from cp and nr datasets, so all data were combined for further analyses. We did not employ the incongruence length difference test as it has been shown to be a poor test of the compatibility of separate data partitions.¹⁷⁾

The data sets were analyzed in combined mixed-model analyses using MrBayes 3.1.2.¹⁸⁾ The analyses of nr, cp, and combined data utilized seven, one, and eight model partitions, respectively (Table 2). Two independent runs with one cold and three heated Markov chains each per analysis were

Table 2. Number of Aligned and Informative Positions, Base Frequencies, Best-Fit Models, $-\ln$ Likelihood with and without Enforcing a Molecular Clock, and Test of Heterogeneity among Lineages for All Loci Analyzed

Partition (no. Taxa)	Alignment positions	Base frequency	P^a	Best-fit model for Bayesian analysis (AIC)	$-\ln$ likelihood		P
	Total Inform. Inform. (%)	A, C, G, T			No molecular clock	$-2 \log LR$ Molecular clock	
<i>trnS^{GCU}-trnQ^{UUG}</i>	2538	0.359, 0.158	0.9982	GTR+G	10404.84	1959.94	<0.001
	929	0.172, 0.312			11384.81		
	36.60						
<i>18S rDNA</i>	1702	0.242, 0.227	1.0	TrN+I (HKY+I) ^{b)}	3089.77	156.36	<0.001
	27	0.275, 0.256			3167.95		
	1.59						
<i>DBAT</i> 1st codon	440	0.272, 0.188	1.0	HKY+I	769.68	21.20	>0.05
	8	0.327, 0.212			780.28		
	1.82						
<i>DBAT</i> 2nd codon	440	0.284, 0.211	1.0	HKY	770.78	19.04	>0.10
	9	0.173, 0.332			780.30		
	2.05						
<i>DBAT</i> 3rd codon	440	0.293, 0.172	1.0	K81uf (GTR) ^{b)}	988.78	34.52	<0.01
	19	0.226, 0.309			1006.04		
	4.32						
<i>TS</i> 1st codon	344	0.307, 0.183	1.0	HKY	752.29	51.06	<0.001
	8	0.321, 0.189			777.82		
	2.33						
<i>TS</i> 2nd codon	343	0.294, 0.249	1.0	HKY+G	663.26	35.14	<0.01
	3	0.179, 0.278			680.83		
	0.87						
<i>TS</i> 3rd codon	343	0.212, 0.228	1.0	TVM+G (GTR+G) ^{b)}	812.54	95.16	<0.001
	14	0.258, 0.301			860.12		
	4.08						
All combined	6590	0.297, 0.194	0	Partitioned mixed model	18440.33	677.0	<0.001
	986	0.227, 0.282			18778.83		
Data (24)	14.96						

a) Significant values were based on χ^2 tests for homogeneity across taxa as implemented in PAUP* 4.0β10. b) Implemented in MrBayes 3.1.2.

performed simultaneously until the average standard deviation of split frequencies between the two runs dropped below 0.01. Analyses were run twice to check for consistency of results. We ran two simultaneous runs for 3×10^6 generations and sampled trees every 100 generations. Topology and branch-length information were summarized in 50% majority rule consensus trees; samples obtained before stationarity of $-\ln$ likelihoods against generations had been reached were discarded as burn-in.

RESULTS AND DISCUSSION

Taxus is notorious for being taxonomically difficult. Except from the split between three North American species and the other species, phylogenetic relationships within the genus have remained enigmatic for the most part, and classification schemes currently in use do not rest upon well-supported hypotheses about the underlying phylogeny. Due to limited taxon sampling, the molecular studies conducted have so far provided only few detailed insights into relationships within *Taxus*.^{11,19,20} In Li *et al.*'s *Taxus* study,¹¹ the controversial *T. sumatrana*, *T. yunnanensis* and *T. wallichiana* were not included and only internal transcribed spacer (ITS) sequences were used to infer the phylogenetic relationship. Collins *et al.*²⁰ noted that *Taxus* species delimitation remained a problem but their study focused on two hybrids and their parental species. In the present study, we have substantially increased taxonomic sampling of nrDNA and cpDNA for *Taxus* and provide a much more comprehensive picture of their phylogeny.

Sixty-six new sequences were generated for 27 taxa. DNA

sequence lengths and general characteristics for each gene and spacer are summarized in Table 2. Final alignments comprised 1030 aligned positions from the *TS* gene, 1320 from the *DBAT* gene, 2538 from the *trnS-trnQ* spacer, and 1702 from 18S rDNA. No significant variation of base frequencies was observed among taxa within each partition (Table 2). *TS* and *DBAT* sequences exhibited no stop codons or frameshift mutations. As expected, 18S rDNA, *DBAT* 1st and *TS* 2nd codon positions evolved slower than other partitions and have fewer parsimony informative sites.

The cp and nr trees showed similar relationships among taxa except a lower interspecific resolution in the nr tree (data not shown). All data were therefore combined to have a "total" evidence hypothesis. In the combined tree (Fig. 2), monophyly of *Taxus* was highly supported, and the clade formed by *T. canadensis* and *T. hunnewelliana* UBC200707 was well supported and basal to the rest of *Taxus*. Two hybrids and their parental species *T. cuspidata* were between *T. canadensis* and a large group formed by other species. *T. contorta* and *T. baccata* formed a basal subclade in this large group, while the others fell into two subclades. Three North American species formed a well-supported subclade (PP=1.00; BP=100) and four Asian species formed the other (PP=0.91). Branches within *Taxus* were much shorter than outside branches (Fig. 2), which indicates that they might have undergone a relatively recent radiation. Alternatively, evolutionary rates might have slowed down in *Taxus* after the lineages split.

de Laubenfels⁷ designated only a single species of Asian yew where Farjon⁴ sees five (*T. chinensis*, *cuspidata*, *contorta*, *sumatrana*, *wallichiana*), all with disjunct distribu-

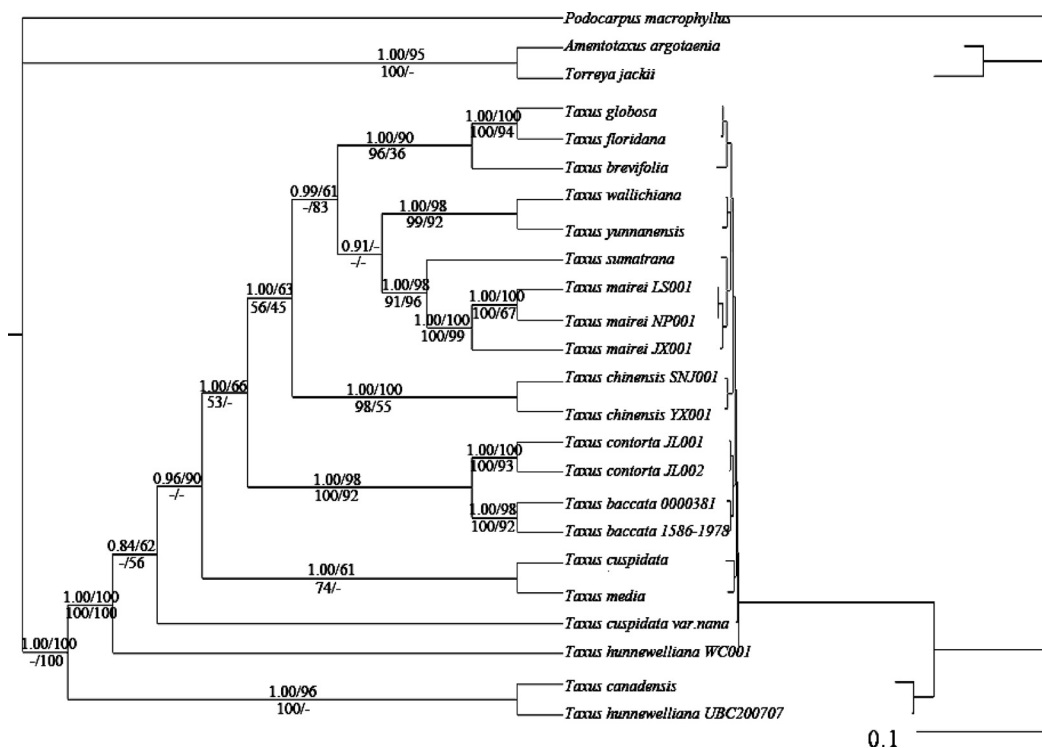


Fig. 2. Bayesian 50% Majority Rule Consensus Tree (10000 Trees Sampled; Burn-in=2500 Trees) Inferred from nr+cp DNA Alignment under the Partitioned Model

Bayesian posterior probabilities are given above branches, before slash (/). ML bootstrap proportions (%) calculated under the GTR+G model are given above branches, after slash (—, clade not included in the tree). MP and NJ bootstrap proportions are shown below branches (MP/NJ). Branch lengths (shown on the right; scale bar, expected number of substitutions per site) are proportional to the mean of the posterior probabilities of the branch lengths of the sampled trees.

Botanical Garden, Canada), and James Stevenson (University of Oxford Botanic Garden, U.K.). This study was supported by the National 973 Project (2007CB707802) of the Department of Science & Technology of China, the Innovation Fund of Drug R & D (2007E11SF0052) of Dalian Bureau of Science & Technology, and the Innovation Fund of Dalian Institute of Chemical Physics, Chinese Academy of Sciences.

REFERENCES

- 1) van Rozendaal E. L., Lelyveld G. P., van Beek T. A., *Phytochemistry*, **53**, 383—389 (2000).
- 2) Khan I. A., *Life Sci.*, **78**, 2033—2038 (2006).
- 3) Chan K., *Chemosphere*, **52**, 1361—1371 (2003).
- 4) Farjon A., “World Checklist and Bibliography of Conifers,” Royal Botanical Gardens at Kew, Richmond, U.K., 1998.
- 5) Fu L., Li N., Mill R. R., “Flora of China,” Vol. 4, ed. by Wu Z., Raven P. H., Science Press, Beijing, Missouri Botanical Garden, St. Louis, 1999.
- 6) Spjut R. W., *J. Bot. Res. Inst. Texas*, **23**, 203—289 (2007).
- 7) de Laubenfels D. J., “Flora Malesiana, Series I,” Vol. 10, Kluwer, Dordrecht, 1988, pp. 337—453.
- 8) Cao H., Sasaki Y., Fushimi H., Komatsu K., *Biol. Pharm. Bull.*, **24**, 1389—1394 (2001).
- 9) Sasaki Y., Fushimi H., Cao H., Cai S. Q., Komatsu K., *Biol. Pharm. Bull.*, **25**, 1593—1599 (2002).
- 10) Walker K., Croteau R., *Phytochemistry*, **58**, 1—7 (2001).
- 11) Li J., Davis C. C., Tredici P. D., Donoghue M. J., *Harvard Papers in Bot.*, **6**, 267—274 (2001).
- 12) Shaw J., Lickey E. B., Beck J. T., Farmer S. B., Liu W., Miller J., *Am. J. Bot.*, **92**, 142—166 (2005).
- 13) Posada D., *Nucleic Acids Res.*, **34** (Web Server issue), W700—W703 (2006).
- 14) Zwickl D. J., “Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion [Ph.D. dissertation],” The University of Texas at Austin, Austin, Texas, U.S.A., 2006.
- 15) Swofford D. L., PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts, 2002.
- 16) Tamura K., Dudley J., Nei M., Kumar S., *Mol. Biol. Evol.*, **24**, 1596—1599 (2007).
- 17) Hipp A. L., Hall J. C., Sytsma K. J., *Syst. Biol.*, **53**, 81—89 (2004).
- 18) Ronquist F., Huelsenbeck J. P., *Bioinformatics*, **19**, 1572—1574 (2003).
- 19) Li X., Zheng B., Su Y., Wang T., *Acta Sci. Nat. Univ. Sunyatseni*, **41**, 70—74 (2002).
- 20) Collins D., Mill R. R., Moller M., *Am. J. Bot.*, **90**, 175—182 (2003).
- 21) Luu N. D. T., Thomas P. I., Kim C. L., “Vietnam/Conifers of Vietnam. Darwin Initiative,” ISBN 1 872291 64 3, 2004.
- 22) van Rozendaal E. L., Kurstjens S. J., van Beek T. A., van den Berg R. G., *Phytochemistry*, **52**, 427—433 (1999).