Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 47 (2008) 1190–1202
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Reticulate evolution in *Thuja* inferred from multiple gene sequences: Implications for the study of biogeographical disjunction between eastern Asia and North America

Dan Peng, Xiao-Quan Wang *

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing 100093, China

Received 28 September 2007; revised 2 February 2008; accepted 5 February 2008
Available online 13 February 2008

Abstract

The eastern Asia–North America disjunction is one of the most interesting biogeographical patterns, but its formation is still in much debate. Here nucleotide sequences of five cpDNA regions, nrDNA ITS and two low-copy nuclear genes (*LEAFY*, *4CL*) were employed to reconstruct the phylogeny and to explore the historical biogeography of *Thuja*, a typical eastern Asia–North America disjunct genus. High topological discordance was observed between chloroplast and nuclear gene trees, even between different nuclear gene trees, suggesting that *Thuja* could have a reticulate evolutionary history due to multiple interspecific hybridization events. The eastern Asian species *Thuja koraiensis* might have obtained its chloroplast genome from the eastern North American species *T. occidentalis* by chloroplast capture, while the western North American species *T. plicata* is very likely to have inherited a recombinant cpDNA. Based on the phylogenetic analysis of multiple genes, DIVA-reconstruction of the distribution history, molecular clock estimation and fossil data, we inferred that *Thuja* could have originated from the high-latitude areas of North America in the Paleocene or earlier with subsequent expansion into eastern Asia through the Bering Land Bridge. The two eastern Asia species *T. standishii* and *T. sutchuenensis* have a sister relationship, and their split could have occurred in the Oligocene or early Miocene. In the present study, the selection of molecular markers in biogeographic studies was also discussed. Since most previous studies on the eastern Asia and North America disjunction are based on uniparentally inherited cpDNA and (or) directly sequenced nrDNA ITS data, the historical reticulate evolution in the studied groups might have been underestimated. Therefore, we suggest that multiple genes from different genomes, especially low-copy nuclear genes, be used in this research area in the future.

Keywords: Molecular phylogeny; Biogeography; cpDNA recombination; Chloroplast capture; Eastern Asia–North America disjunction; Interspecific hybridization; *Thuja*

1. Introduction

The geological changes and climatic cooling in the Cenozoic, especially in the late Tertiary and Quaternary, interrupted the continuity of the boreotropical flora or mixed mesophytic forests in the Northern Hemisphere and brought about the plant biogeographic patterns of intercontinental disjunction. The disjunction between eastern Asia and eastern North America (EA–ENA) is the most attractive one and has magnetized several generations of botanists in the past nearly 150 years (Gray, 1859; Boufford and Spongberg, 1983; Wen, 1999; Manos and Donoghue, 2001). Both eastern Asia and eastern North America sheltered their residents with complex geographical topologies when congeneric species became extinct in Europe and other areas (Tiffney, 1985a,b; Wen, 1999; Tiffney and Manchester, 2001). Tiffney (1985a,b) attributed the EA–ENA disjunct distribution to a layering of many events rather than a single historical event. The two landmark papers

* Corresponding author. Fax: +86 10 62590843.
E-mail address: xiaoq_wang@ibcas.ac.cn (X.-Q. Wang).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.02.001

also emphasized the important roles of the Bering Land Bridge and the North Atlantic Land Bridge, two main land bridges connecting Asia, North America and Europe in the Tertiary, in shaping modern phytogeographic patterns of the Northern Hemisphere, especially the EA–ENA disjunction.

With the help of more and more molecular markers, newly developed analysis methods (e.g., DIVA, Multidivitime, r8s) and fossil records, the evolution of some plant groups with an EA–ENA disjunct distribution has been investigated in much more detail, including interspecific relationships, divergence times of different lineages, and biogeographical histories (e.g., Cunningham and Collins, 1994; Qiu et al., 1995; Wen and Zimmer, 1996; Gould and Donoghue, 1998; Wen, 1999; Xiang et al., 1998, 2000; Wen and Shi, 1999; Xiang and Soltis, 2001; Wang et al., 2003; Nie et al., 2006). However, most of the studies were based on uniparentally inherited chloroplast markers and (or) directly sequenced PCR products of nuclear ribosomal DNA internal transcribed spacers (ITS), which have limitations in discovering reticulate evolution events such as hybridization and introgression. As one mode of speciation in plant, hybrid speciation has occurred much more frequently in nature than previously envisioned. For example, several compelling ancient hybridization events and cryptic hybrid species have been revealed recently with both plastid and nuclear gene markers (e.g., Cronn et al., 2003; Oh and Potter, 2005; Fehrer et al., 2007). A robust phylogeny reconstructed with multiple gene markers from different genomes is very necessary to the historical biogeographical study that infers the development of distribution pattern of a group based on the gene tree topology. On the other hand, previous studies show that the EA–ENA disjunction could have originated at different times and by different pathways (Wen, 1999). In particular, a closer biogeographic relationship between eastern North America and western North America than between eastern Asia and eastern North America was reported from some groups, such as *Aralia* (Wen et al., 1998) and *Cornus* (Xiang et al., 1998). Therefore, more attention should be paid to the western North American species for the study of EA–ENA disjunction.

Thuja L., a small genus of Cupressaceae comprising five extant species, exhibits a disjunct distribution in eastern Asia, eastern and western North America (Farjon, 2005). The three eastern Asian species of the genus have quite restricted distributions, including *Thuja sutchuenensis* in Chengkou, Chongqing, China, *T. koraiensis* in the Changbai Mountain of northeastern China and the Korean Peninsula, and *T. standishii* native to Japan. *T. sutchuenensis* was once listed as a species extinct in the wild by IUCN-SSC and rediscovered in 1999 (Farjon and Page, 1999; Xiang et al., 2002). The two North American species of *Thuja* have much wider distributions. *T. occidentalis* occurs in eastern North America, occupying a geographical range between the subarctic taiga–tundra interface in the north and the belt of deciduous angiosperm forests in the south. *T. plicata* is disjunctly distributed in the Pacific Coastal Mountains and the Rocky Mountains of western North America.

Based on fossil evidence, present distribution and morphological characters, especially seed cone structure, McIver and Basinger (1989) proposed a phylogeny of *Thuja* comprising both extant and fossil species. They inferred that *Thuja* had occupied much of the Northern Hemisphere in the early Tertiary and the subsequent dramatic climatic and geographic changes were responsible for the current disjunction of the genus. This phylogeny is topologically quite different from the molecular phylogeny of *Thuja* reconstructed with the directly sequenced nrDNA ITS region by Li and Xiang (2005), in which an eastern Asia origin and two dispersals to North America were suggested for the genus. The conflict between previous studies could be attributed to the slight morphological differences among species, the discontinuity and incompleteness of the fossil record, and the use of a single molecular marker. In addition, it is better to clone the PCR products when nrDNA ITS is used in the molecular systematic study of gymnosperms, in consideration that this DNA region might have marked intragenomic length and sequence variations due to the incomplete concerted evolution as reported from Pinaceae (Liston et al., 1996; Maggini et al., 1998; Gernandt et al., 2001; Wei et al., 2003; Wei and Wang, 2004a; Campbell et al., 2005; Kan et al., 2007). It is obviously that more DNA markers, including both cytoplasmic and nuclear genes, should be used to reconstruct a robust phylogeny of *Thuja* and then to explore the historical biogeography of the genus.

Recently, several dozens of chloroplast DNA markers suitable for the study of interspecific relationships of plants have been screened out (Shaw et al., 2005, 2007). Meanwhile, the low-copy nuclear gene has the advantages of biparental inheritance and rapid evolutionary rates, and thus is increasingly used in systematic and biogeographic studies (Sang, 2002; Small et al., 2004; Whittall et al., 2006). For example, the *LEAFY* gene involved in the formation of the flower meristem (Frohlich and Parker, 2000) has been used as a single copy gene in the phylogenetic reconstruction of some seed plants (e.g., Grob et al., 2004; Oh and Potter, 2005; Won and Renner, 2005). Also, the *4CL* gene, encoding 4-coumarate: coenzyme A ligase in the general phenylpropanoid metabolism essential to plant development and environmental interactions (Cukovic et al., 2001; Hamberger and Hahlbrock, 2004), has been successfully used in molecular phylogenetic studies of conifers (Wang et al., 2000; Wei and Wang, 2004b; Syring et al., 2005). In the present study, we employed nucleotide sequences of five chloroplast non-coding regions (*rpl*16, *Atp*I–*rpo*C1, *trn*S–*trn*fM, *trn*S–*trn*G, and *trn*T–*trn*F) and two low-copy nuclear genes (*LEAFY* and *4CL*), together with the cloned nrDNA ITS sequences, to reconstruct the phylogeny and to investigate the historical biogeography of *Thuja*. Based on the results, we also discussed the utilities of multiple genes from different genomes

in the biogeographic study, in particular the study of eastern Asian–North American disjunction.

2. Materials and methods

2.1. Plant materials

All the five species of *Thuja* were sampled (Table 1). To detect whether there are intraspecific polymorphisms that occur in the cpDNA markers we used, five individuals of *T. koraiensis* from different populations were analyzed and each of the other congeneric species was represented by more than one individual as possible. *Thujopsis dolabrata* (Thunb. ex L. f.) Sieb. et Zucc. was selected as the outgroup for its sister relationship with *Thuja* (Brunsfeld et al., 1994; Gadek et al., 2000; Kusumi et al., 2000, 2006; Quinn et al., 2002). Voucher specimens were deposited in the herbarium of Institute of Botany, Chinese Academy of Sciences (PE) except Kew 73-10508 of *T. plicata*.

2.2. DNA extraction, gene amplification, cloning, and sequencing

Total DNA was extracted from fresh or silica gel-dried leaves using the modified CTAB method following the protocol of Rogers and Bendich (1988) and used as the template in the polymerase chain reaction. The five chloroplast DNA fragments, *rpl*16 intron, *Atp*I–*rpo*C1 intergenic spacer (IGS), *trn*S–*trn*fM IGS, *trn*S–*trn*G IGS, and *trn*T–*trn*F region (including *trn*T–*trn*L IGS, *trn*L gene, and *trn*L–*trn*F IGS), were amplified with primers used in the previous studies (Demesure et al., 1995; Downie et al., 1996; Small et al., 1998) except the *trn*T–*trn*F region that needs the gymnosperm-specific primer pair (*trn*PTFF: 5′GTCCTCTGCTCTACCAACTGA3′; *trn*PTFR: 5′GATC GAACCGATGACCATCG3′). The nrDNA ITS region was amplified using primers ITS1N from Wei et al. (2003) and ITS4 from White et al. (1990). For the amplification of the nuclear genes *LEAFY* and *4CL*, protein and nucleotide sequence alignments were made across gymnosperms and angiosperms to design degenerate primers from conserved regions. According to results of the initial amplification and sequencing, Cupressaceae-specific primers were designed for the subsequent amplifications, including *4CL*E2Fa (5′GGATAYGGMATGACGGAAGC3′) and *4CL*Rb (5′GAGCTCTTTCACTCTGTCGACGAT3′) for *4CL*, *LFY*E1F3 (5′TGCAGCTTTCTTCAAGTGGGA3′) and *LFY*E3R4 (5′CCAGATTCGAAGCTTTTCATG3′) for *LEAFY*.

PCR amplification was carried out in a volume of 25 μl containing 50–100 ng of DNA template, 6.25 pmol of each primer, 0.2 mM of each dNTP, 2 mM $MgCl_2$, and 0.75 U of *Taq* DNA polymerase. Amplification was conducted in a Tgradient thermal cycler (Biometre, Göttingen, Germany) or an Eppendorf Mastercycler Gradient thermal cycler (Eppendorf Scientific, Westbury, NY). PCR cycles were as follows: one cycle of 4 min at 94 °C, four cycles of 1 min at 94 °C, 30 s at 50–58 °C, and 1–2 min at 72 °C, followed by 36 cycles of 30 s at 94 °C, 30 s at 50–58 °C, and 1–2 min at 72 °C, with a final extension step for 10 min at 72 °C. PCR products were separated on 1.5% agarose gels. The band with the right size was cut out and purified using GFX PCR DNA and Gel Band Purification Kit (Amersham Biosciences). For the cpDNA fragments, purified PCR products were directly sequenced with the PCR primers and several internal primers (*rpo*C-b: 5′TCTGAGTTCTAGCCAAGTCC3′; *rpo*C-e: 5′CTGCTT TAGCAGCTCGATGT3′; *rpo*C-R: 5′GAATGGGTT AGATTGCTTGC3′). For the nuclear gene, the purified PCR products were cloned with pGEM®-T Easy Vector System II (Promega). More than 20 clones with correct insertion (determined by digestion with EcoRI) were picked for each gene of each species and screened with one of the PCR primers. All distinct clones were sequenced in both directions with the DYEnamic ET Terminator Kit (Amersham Pharmacia Biotech), using primers T7 and SP6 and internal primers, including ITSP2N (5′GAGAGCC GAGATATCCGTTG3′), *4CL*E_2S (5′ATCGTCGACAGA GTGAAAGAG3′), *4CL*E3R (5′GCTTCCAGCTCAGCA GGCGCGAC3′), *LFY*IS1 (5′CACATTCTTGGCCAGCT CAG3′), *LFY*E2F1 (5′CCTTTCATTGTTACTGAGCC3′) and *LFY*E2R3 (5′AGCAGAAACTTGCCACACTG3′). After precipitation in 95% EtOH and 3 M NaAc (pH 5.2), the sequencing products were separated on a MegaBACE 1000 automatic sequencer (Amersham Biosciences, Buckinghamshire, UK). The sequences obtained in this study were deposited in GenBank under accession numbers EU183450–EU183455 (*Atp*I–*rpo*C1), EU183456–EU183461 (*trn*S–*trn*fM), EU183444–EU183449 (*rpl*16), EU183462–EU183467 (*trn*S–*trn*G), EU183468–EU183473 (*trn*T–*trn*F), EU183411–EU183416 (*LEAFY*), EU183417–EU183423 (*4CL*), and EU183424–EU183443 (ITS).

Table 1
Sources of materials

Taxa	Sources/individuals	Vouchers
Thuja koraiensis Nakai	Changbai, Jilin, China/5	X.-Q. Wang 2004-T011
T. sutchuenensis Franch.	Chengkou, Sichuan, China/3	Z.-Y. Li 2005-1
T. occidentalis L.	Botanic Garden, Institute of Botany, Beijing/3	X.-Q. Wang 2004-192
T. plicata Donn ex D. Don	Botanic Garden, Institute of Botany, Beijing/1	X.-Q. Wang 2004-205
	Royal Botanic Garden, Kew, UK/1	Kew 73-10508
T. standishii Carriere	Forestry and Forest Products Research Institute, Tsukuba, Japan/1	
Thujopsis dolabrata Siebold & Zucc.	Lushan, Jiangxi, China/1	X.-Q. Wang 2004-204

2.3. Data analysis

Sequence alignments were made with CLUSTAL X (Thompson et al., 1997) and refined manually. The data matrix is available from the corresponding author. To assess congruence between different cpDNA regions, we analyzed the five fragments separately to see if they produced a similar topology and also used the MP-based incongruence length difference test (ILD, Farris et al., 1994) as implemented in PAUP* 4.0b10 (Swofford, 2002) in a pairwise fashion. The ILD test was also used to assess congruence between chloroplast and nuclear genes, and between different nuclear gene datasets. For the nuclear gene with more than one copy, a single clone was randomly chosen to represent a species. The test was performed with 1000 replications of branch-and-bound search. Following Cunningham (1997), we selected a P value of 0.01 as a significance criterion for this test. Considering that the ILD test may be prone to Type I errors by rejecting the null hypothesis of data homogeneity (Yoder et al., 2001), we further used the Shimodaira–Hasegawa (SH) test to evaluate the compatibility between the cpDNA data and the nuclear DNA data (Shimodaira and Hasegawa, 1999). The SH test was also performed with PAUP* 4.0b10, using 50,000 bootstrap replicates with the resampling estimated log-likelihood (RELL) method. The maximum likelihood (ML) tree generated from the cpDNA data was assessed in terms of the combined nuclear DNA data, and the ML tree from the combined nuclear DNA data was evaluated relative to the cpDNA data.

Phylogenetic relationships were inferred using maximum parsimony (MP) and ML optimality criteria as implemented in PAUP* 4.0b10 and MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). MP analyses used branch-and-bound or heuristic searches with 100 random addition sequence replicates, a maximum of 1000 trees saved per round, tree bisection–reconnection (TBR) branch swapping, and MULTREES on. All character states were treated as unordered and equally weighted, with gaps as missing data. To evaluate relative robustness of the clades found in the most-parsimonious trees, the bootstrap analysis (Felsenstein, 1985) was performed with 1000 replicates using the same options as above except that a maximum of 100 trees were saved per round.

The evolutionary models for ML and Bayesian analyses were optimized by hierarchical likelihood ratio tests (LRTs) using ModelTest 3.06 (Posada and Crandall, 1998). The F81 model (Felsenstein, 1981) with gamma-distributed rate variation (F81+G) was chosen for the combined cpDNA data, the Hasegawa–Kishino–Yano (HKY) model (Hasegawa et al., 1985) for *LEAFY* and nrDNA ITS datasets, and the HKY model with gamma-distributed rates (HKY+G) for *4CL* and the combined nuclear gene data. Since model partitioning is important for large datasets in the Bayesian analysis (Brown and Lemmon, 2007), a partitioned Bayesian analysis of the combined nuclear DNA was also implemented by applying the previously determined models to each data partition. For the ML analyses, optimal gene trees were found via branch-and-bound or heuristic searches with 1000 replicates of random sequence addition, TBR branch swapping, and MULTREES on, and clade robustness was estimated by bootstrap analysis of 1000 replicates using heuristic searches with the same options. For the Bayesian inference (BI) analyses, one cold and three incrementally heated Markov chain Monte Carlo (MCMC) chains were run for 1,000,000 cycles. Trees were sampled every 100 generations. For each dataset, MCMC runs were repeated twice to avoid spurious results. The first 300 samples for each run were discarded as burn-in to ensure that the chains have become stationary. Phylogenetic inferences were based on the trees sampled after generation 30,000.

Rate constancy across lineages was examined for each dataset using a likelihood ratio test (LRT) (Felsenstein, 1988), in which likelihood scores of the trees with and without an enforced molecular clock were compared. Significance was evaluated by comparing two times the difference in log likelihoods to a χ^2 distribution with $n-2$ degrees of freedom (n is the number of taxa). Then, the dataset that does not reject a molecular clock hypothesis was used to estimate divergence times of *Thuja* species, using the Langley–Fitch (LF) method implemented in the program r8s (Langley and Fitch, 1974; Sanderson, 2002). The method estimated one substitution rate across the entire tree and a set of calibrated divergence times for all unfixed nodes under the TN algorithm. For comparisons, we also used the nonparametric rate smoothing (NPRS) and the penalized-likelihood (PL) method (Sanderson, 1997, 2002) in the divergence time estimation with r8s. The PL method used a semiparametric approach to optimize rates on branches with a numerical penalty. Optimal values of smoothing were determined by a cross-validation procedure. Standard errors of the divergence times were calculated with a nonparametric bootstrap procedure. We simulated 100 datasets with the SEQBOOT program in PHYLIP Version 3.6a2 (Felsenstein, 1993), then used the matrices to generate new ML trees with PAUP* 4.0b10 and estimated the divergence time on each of the new trees with r8s. The divergence times for node were used to calculate standard errors.

The age of the most recent common ancestor of *Thuja* was fixed at 60 million years before the present (mya) based on the earliest reliable fossil record of the genus from the middle Paleocene of Ellesmere Island in the Canadian Arctic (McIver and Basinger, 1989; Bennike, 1990; Farjon, 2005). Although LePage (2003) described a *Thuja* species collected from the late Cretaceous of Alaska, the identity of this fossil as a species of *Thuja* (or even Cupressaceae) is still in doubt (Farjon, 2005).

We reconstructed biogeographic history of *Thuja* with the program DIVA1.1 (Ronquist, 1996, 1997), in which both dispersal and vicariance are allowed. Ancestral states were inferred by minimizing the number of dispersal events. Thus, vicariance events were emphasized in this

analysis. We did not use an outgroup in the analysis because *Thujopsis*, the sister group of *Thuja*, is monotypic and has a relict distribution in Japan.

3. Results

3.1. Sequence characterization

A summary of the sequences we analyzed is shown in Table 2. For the five cpDNA fragments, no intraspecific sequence variation was observed. The two primers *trn*PTFF and *trn*PTFR failed to amplify the *trn*T–*trn*F region from *T. standishii*. Therefore, specific primers *trn*PLR (5′GAATCGGTAGACGCTACGGA3′) and *trn*PT (5′ACCTCTGAGCTAAGCAGGCTCAT3′) were further designed for this species. The sequence alignments of the *rpl*16 intron, *Atp*I–*rpo*C1 IGS, *trn*S–*trn*fM IGS, *trn*S–*trn*G IGS and *trn*T–*trn*F region are 773, 1720, 781, 756, and 1069 bp in length, respectively. The combined five cpDNA fragments have 5099 bp, of which 38 (0.75%) are variable and 15 (0.29%) are parsimony-informative as shown in Table 3. The *trn*S–*trn*G IGS is the most variable, while the *rpl*16 intron is the least.

The *LEAFY* gene fragment we amplified comprised three exons and two introns, and only one copy was found in the five species of *Thuja*. The aligned *LEAFY* sequence matrix contained 2836 bp with 26 variable sites, and only two of the 10 parsimony-informative sites occurred in the exonic region. *T. standishii* shared a 1-bp deletion in the intron with *T. sutchuenensis*.

The amplified *4CL* gene of *Thuja* consisted of four exons and three introns. All conspecific clones had identical sequence except that two distinct clones were detected from *T. plicata*. To investigate whether other *4CL* members are present in the *Thuja* species but missed due to the PCR bias, two sequence-specific primers *4CL*-Ks (5′CTCTCT ATGTTCTCGAAAGATCATG3′) and *4CL*-Os (5′GATC GAATTCTATGACTCTTCG/CTC3′) were designed for further PCR amplifications and the same results were obtained. The aligned *4CL* matrix contained 1836 characters, of which 60 were variable. For the 30 parsimony-informative sites, two were located at the exon regions, while the remaining 28 were distributed in the introns, especially the third intron. *T. koraiensis* shared a 29-bp indel in the third intron with *T. plicata*, and the clone *T. plicata 205-1* shared a 42-bp indel in the second intron with the outgroup *Thujopsis dolabrata*.

For the nrDNA ITS region, two to five distinct clones were detected from each species of *Thuja*. The aligned ITS matrix, including 19 sequences, consisted of 1203 characters, of which 139 were variable and 56 were parsimony-informative. The informative sites concentrated on ITS1 and ITS2. In addition, seven indels were found in ITS1.

Table 2
A comparison of the eight DNA markers we used

DNA marker	Length[a] (bp)	Alignment[a]	Variable sites[a] (%)	Parsimony-informative sites[a] (%)	% GC[a]	No. of MP tree	Length of MP tree	CI excluding autapomorphies	RI
cpDNA									
*rpl*16	773	773	3 (0.39)	1 (0.13)	30.9	2	27	0.9630	0.6667
*Atp*I–*rpo*C1	1716–1720	1720	9 (0.52)	4 (0.23)	33.0	2	37	0.9730	0.8571
*trn*S–*trn*fM	762–778	781	6 (0.77)	3 (0.38)	32.6	4	34	0.9706	0.8000
*trn*S–*trn*G	737–756	756	7 (0.93)	3 (0.40)	26.1	1	36	1.0000	1.0000
*trn*T–*trn*F	708–1068	1069	13 (1.22)	4 (0.37)	31.6	3	33	0.9697	0.8000
Combined	4715–5074	5099	38 (0.75)	15 (0.29)	31.3	1	169	0.9645	0.7500
nDNA									
LEAFY	2811–2833	2836	26 (0.92)	10 (0.35)	37.39	1	71	0.9859	0.9286
4CL	1666–1804	1836	60 (3.27)	30 (1.63)	37.6	2	95	0.9368	0.8636
ITS	1195–1200	1203	139 (11.55)	56 (4.66)	56.9	2	193	0.9482	0.8636
Combined	5693–5829	5876	105 (1.79)	55 (0.94)	41.8	2	326	0.9448	0.7600

[a] Excluding the outgroup.

Table 3
The parsimony-informative sites from the five chloroplast DNA fragments of *Thuja*

Species	*rpl*16	*Atp*I–*rpo*C1				*trn*S–*trn*FM			*trn*S–*trn*G			*trn*T–*trn*F			
Thuja sutchuenensis	A	G	C	**G**	**A**	A	A	**G**	**T**	G	**A**	**T**	T	**T**	**A**
T. standishii	A	G	C	**G**	**A**	A	A	**G**	**T**	G	**A**	**T**	?	?	?
T. plicata	***G***	***A***	***T***	**G**	**A**	***G***	***C***	**G**	**T**	***T***	**A**	**T**	C	**T**	**A**
T. koraiensis	***G***	***A***	***T***	A	C	***G***	***C***	A	G	***T***	G	C	C	A	C
T. occidentalis	***G***	***A***	***T***	A	C	***G***	***C***	A	G	***T***	G	C	T	A	C
Thujopsis dolabrata	T	G	C	G	C	G	T	A	T	T	G	C	C	G	A

Synapomorphies uniting *T. plicata* with the *T. sutchuenensis*–*T. standishii* group are shown in bold, while those uniting *T. plicata* with the *T. koraiensis*–*T. occidentalis* group are in bolded italics. Question marks indicate missing data.

3.2. Phylogenetic analyses

Separate analyses of individual chloroplast genes yielded one to four most-parsimonious trees (Table 2), and none of the genes had enough resolution for the interspecific relationships of *Thuja*. The ILD test showed no significant incongruence between the five cpDNA fragments ($P > 0.01$), so we combined these sequences into a single dataset for further phylogenetic analysis. Maximum parsimony analysis of the combined cpDNA data generated a single most-parsimonious tree with tree length = 169, consistency index (CI) = 0.9645, and retention index (RI) = 0.7500 (Fig. 1A). The five *Thuja* species were divided into two clades: one comprised *T. plicata* and two eastern Asian species *T. sutchuenensis* and *T. standishii* that formed a well-supported sister group, while the other was composed of the eastern Asian species *T. koraiensis* and the eastern North American species *T. occidentalis*. Bayesian and ML analyses of the cpDNA data generated the same tree topology as the MP tree.

Phylogenetic analysis of the single copy gene *LEAFY* resulted in one MP tree with 71 steps (CI = 0.9859, RI = 0.9286). In the tree, a sister relationship between two eastern Asian species *T. sutchuenensis* and *T. standishii* was robustly supported, whereas the remaining three species formed a strongly supported clade, in which the two North American species *T. occidentalis* and *T. plicata* clustered together (Fig. 1B). Maximum parsimony analysis of the *4CL* gene generated two MP trees with 95 steps (CI = 0.9368, RI = 0.8636). The strict consensus tree is

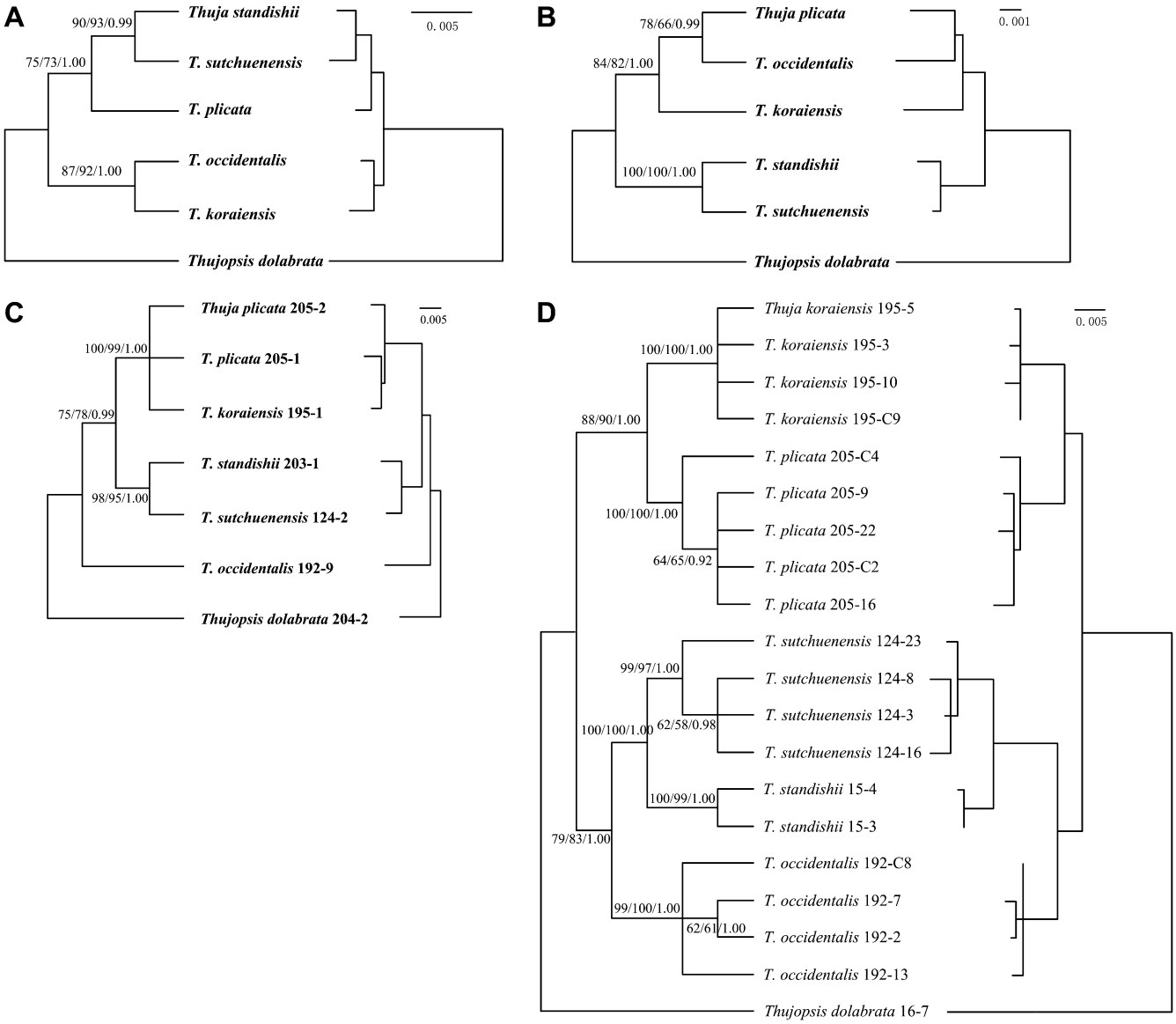

Fig. 1. Strict consensus of most-parsimonious (MP, left) and maximum likelihood (ML, right) trees of *Thuja* based on sequence analyses of the combined cpDNA (A) and different nuclear genes (B, *LEAFY*; C, *4CL*; D, nrDNA ITS). Numbers on the branches denote the bootstrap values of 1000 replicates for MP and ML, and the Bayesian posterior probabilities. In Fig. 1D, the branch length of *Thujopsis dolabrata* is not shown to scale because of the great distance between outgroup and ingroup.

shown in Fig. 1C, in which the eastern North American species *T. occidentalis* branched off first and the rest *Thuja* species were divided into two well-supported sister clades. One clade comprised *T. standishii 203-1* and *T. sutchenensis 124-2* from two eastern Asian species, and the other consisted of *T. koraiensis 195-1* and two members of *T. plicata*, *205-1* and *205-2*. A heuristic search of the ITS dataset found two MP trees with 193 steps (CI = 0.9482, RI = 0.8636). The strict consensus tree is shown in Fig. 1D, in which two sister clades were found and all conspecific clones clustered into a strongly supported monophyletic group. The two eastern Asian species *T. sutchuensis* and *T. standishii* formed a robust subclade sister to the eastern North American species *T. occidentalis*, while the two Pacific species *T. plicata* and *T. koraiensis* clustered together.

Bayesian and ML analyses of the three nuclear genes also generated the same or similar tree topologies as the MP trees (Fig. 1B–D). As shown in Table 4, the ILD test showed no significant incongruence between the three nuclear genes ($P > 0.01$), and thus we combined them for further analysis. This analysis yielded two MP trees, with a tree length of 326 steps (CI = 0.9448, RI = 0.7600). The strict consensus tree is shown in Fig. 2, in which the eastern North American species *T. occidentalis* formed a trichotomy with two strongly supported species pairs, i.e., *T. standishii–T. sutchenensis*, and *T. koraiensis–T. plicata*. In the Bayesian and ML analyses of the combined nuclear DNA, including the partitioned Bayesian analysis, the two species pairs were also robustly supported (Fig. 2). However, the ILD test revealed significant incongruence between cpDNA and the combined nuclear DNA ($P = 0.003$), although cpDNA and the *LEAFY* gene had congruent phylogenetic signals ($P = 0.071$). The SH test indicated that the cpDNA tree is significantly incongruent with the nuclear DNA data ($P = 0.0018$) and the nuclear DNA tree is also significantly incongruent with the cpDNA data ($P = 0.0242$). Therefore, we did not conduct an analysis for the combined chloroplast and nuclear gene data.

The LRT test showed no significant rate heterogeneity among lineages for the *4CL* gene ($\delta = 1.56$, df = 4, $P > 0.05$), but revealed significant rate heterogeneity among lineages for the nrDNA ITS ($\delta = 130.86$, df = 4, $P < 0.001$), *LEAFY* ($\delta = 26.62$, df = 4, $P < 0.001$), and combined cpDNA ($\delta = 155.99$, df = 4, $P < 0.001$) datasets. Therefore, the *4CL* gene in favor of the molecular clock hypothesis was selected to estimate divergence times. To explore the reticulate evolutionary history of *Thuja* (discussed later), the combined cpDNA deviating from clock-like evolution was also used for molecular clock estimation with the PL method.

Based on the *4CL* gene data, the LF analysis suggested the split between the two eastern Asian species *T. standishii* and *T. sutchuenensis* at 23.7 ± 5.04 Mya, the divergence between *T. koraiensis* and *T. plicata* at 14.7 ± 6.06 Mya, and the separation between the *T. standishii–T. sutchuenensis* clade and the *T. koraiensis–T. plicata* clade at 51.1 ± 3.96 Mya (Fig. 3). The results of NPRS and PL analyses are very similar to the LF analysis (Table 5). According to the PL analysis of the combined cpDNA, the divergence between *T. standishii* and *T. sutchuenensis* occurred at 27.5 ± 5.22 Mya, and that between *T. koraiensis* and *T. occidentalis* at 29.8 ± 6.42 Mya.

Compared to other gene trees we constructed, the *4CL* tree has a good resolution for the interspecific relationships of *Thuja*, and thus was used to infer biogeographic history of the genus. The DIVA analysis indicated that the most recent common ancestor (MRCA) of *Thuja* could have a wide distribution in eastern Asia and North America and one dispersal from eastern Asia to North America occurred (Fig. 3).

Table 4
P values from the incongruence length difference (ILD) test

Dataset	*P* value
cpDNA vs. *LEAFY*	0.071
cpDNA vs. *4CL*	0.001
cpDNA vs. ITS	0.001
LEAFY vs. *4CL*	0.015
LEAFY vs. ITS	0.032
4CL vs. ITS	0.135
cpDNA vs. nuclear DNA	0.003

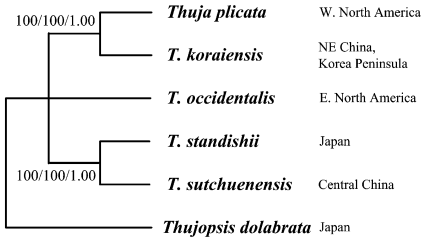

Fig. 2. The strict consensus tree of *Thuja* based on sequence analyses of the combined three nuclear genes with *Thujopsis dolabrata* as the outgroup. Numbers on the branches denote the bootstrap values of 1000 replicates for MP and ML, and the Bayesian posterior probabilities. The distributions of *Thuja* species and the outgroup are shown on the right.

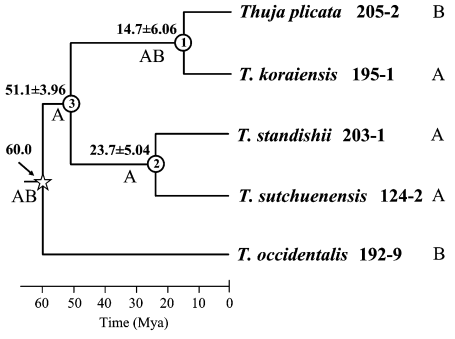

Fig. 3. The distribution history of *Thuja* reconstructed by DIVA and divergence times of the genus estimated by LF based on the *4CL* gene data, with a calibration point of 60 Myr at the root node (indicated by a star) where *Thuja* diverged from the outgroup. The node age estimates with standard errors are marked above branches and the results of DIVA analysis are indicated below branches. A, eastern Asia; B, North America.

Table 5
Molecular clock estimation of divergence times of *Thuja* based on the *4CL* gene (95% credibility intervals in parentheses)

Node	Method		
	LF	NPRS	PL
1	14.7 (±6.06)	12.8 (±6.15)	14.1 (±6.24)
2	23.7 (±5.04)	23.9 (±5.00)	22.7 (±5.00)
3	51.1 (±3.96)	49.8 (±4.03)	50.3 (±5.57)

Node numbers correspond to those marked in Fig. 3.

4. Discussion

4.1. A reticulate evolutionary history of Thuja with chloroplast DNA recombination and chloroplast capture

McIver and Basinger (1989) proposed the first phylogeny of *Thuja* including both extant and fossil species, based on seed cone structure, modern distribution and fossil records of the genus. They inferred that *T. sutchuenensis* might have arisen in the Paleocene and could represent a basal clade together with the fossil species *T. ehrenswaerdii* from Greenland. The fossil species *T. nipponica* found in Japan and Russia and the *T. standishii–T. koraiensis* lineage were hypothesized to form a clade sister to a group comprising the two North American species *T. occidentalis* and *T. plicata* (McIver and Basinger, 1989). The cone characteristics of all *Thuja* species, except *T. sutchenensis*, are remarkably similar and not distinguishable in some poorly preserved fossil specimens (LePage, 2003), and thus should be used carefully in the phylogenetic inference. Using the direct sequence of nrDNA ITS, Li and Xiang (2005) reconstructed a phylogeny of *Thuja*, in which the western North American species *T. plicata* forms a clade with the eastern Asia species *T. koraiensis*, while the eastern North America species *T. occidentalis* is sister to the *T. standishii–T. sutchuenensis* lineage. In the present study, we used multiple genes to reconstruct the phylogeny of *Thuja*, and found high topological discordance between chloroplast and nuclear gene trees, even between different nuclear gene trees (Fig. 1). This finding strongly suggests that the genus *Thuja* could have a reticulate evolutionary history due to multiple interspecific hybridization events.

Two species pairs *T. standishii–T. sutchenensis* and *T. koraiensis–T. plicata* have been revealed by our study (Figs. 1 and 2). The former comprising two eastern Asia species is robustly supported by both chloroplast and nuclear genes, and the latter consisting of two species from both sides of Beringia forms a well-supported clade in the *4CL*, nrDNA ITS and combined nuclear gene trees. It is interesting that *T. koraiensis* shows a sister relationship with *T. occidentalis* in the cpDNA tree rather than with *T. plicata* as indicated in the nuclear gene trees. This incongruence between chloroplast and nuclear gene phylogenies very likely results from interspecific gene flow and subsequent chloroplast capture (Rieseberg and Soltis, 1991; Wei and Wang, 2003; Liston et al., 2007), given that the sister relationship between *T. koraiensis* and *T. plicata* is strongly supported by two independent nuclear genes involving different functions and especially by a bootstrap value of 100% in the combined nuclear gene tree (Figs. 1C, D and 2). That is to say, *T. koraiensis* could have captured the chloroplast genome of *T. occidentalis* when ancient pollen flow and hybridization occurred between the two species. This inference is also supported by the predominantly paternal inheritance of chloroplast in the family Cupressaceae s.l. (Neale et al., 1989; Mogensen, 1996; Shiraishi et al., 2001) and the finding that floral exchange is frequent between eastern Asia and North America in the history (Tiffney, 1985a,b; Milne, 2006).

In the chloroplast gene tree, only the phylogenetic position of *T. plicata* is weakly supported (Fig. 1A). Of the 15 parsimony-informative sites in the cpDNA matrix, all sites are identical between *T. sutchenensis* and *T. standishii* except three missing in *T. standishii*, and 14 are identical between *T. koraiensis* and *T. occidentalis*. Notably, *T. plicata* shares eight sites with the *T. sutchenensis–T. standishii* group and six sites with the *T. koraiensis–T. occidentalis* group (Table 3). This mosaic distribution of nucleotide variations in *T. plicata* could be explained by ancestral cpDNA polymorphism or chloroplast DNA recombination. However, the possibility that *T. plicata* has maintained a different ancestral type of chloroplast DNA could be ruled out, since none of the gene trees we obtained supports it as a primitive species in *Thuja* and its present populations are obviously descendants of a refugial population having experienced a bottleneck during the last ice age (Glaubitz et al., 1999). The chimeric chloroplast sequence in *T. plicata* more likely stems from recombination between the two main cpDNA types found in groups *T. sutchenensis–T. standishii* and *T. koraiensis–T. occidentalis*, respectively, when historical interspecific hybridization occurred. In fact, chloroplast recombination has been reported from some taxa, including conifers such as *Pinus contorta* (Marshall et al., 2001) and *Cycas tatitungensis* (Huang et al., 2001), though non-recombination is regarded as a vital characteristic when researchers use cpDNA in phylogenetic analyses. We infer that *T. koraiensis* might have hybridized with a species in the *T. sutchenensis–T. standishii* group after it captured the cpDNA of *T. occidentalis*, and then the hybrid with a recombinant cpDNA migrated into western North America, giving rise to *T. plicata*. This putative migration event from eastern Asia to North America is also corroborated by the result of the DIVA analysis (Fig. 3), and could have occurred in the middle Miocene according to the molecular clock estimation of the divergence between *T. koraiensis* and *T. plicata* (Table 5).

4.2. Biogeography of Thuja

In the early studies eastern Asia had been regarded as a cradle of many EA–ENA disjunct taxa, though this view was hotly debated (Wen, 1999). Since *T. sutchuenensis*, a species endemic to central China, bears much smaller

mature seed cones (5–8 mm vs. 7–18 mm) and less cone scales (4 pairs vs. 4–6 pairs) than other congeners, it was once regarded as the most ancestral species of *Thuja* (McIver and Basinger, 1989; Farjon, 2005). In addition, the other two eastern Asia species *T. standishii* and *T. koraiensis*, distributed in the two sides of the Korea–Japan Strait, were ever treated as sister species by virtue of morphological characters (McIver and Basinger, 1989). Li and Xiang (2005) proposed an eastern Asia origin and two subsequent dispersals to North America for the genus *Thuja*, based on the directly sequenced nrDNA ITS. However, this point is not supported by the present study. As discussed formerly, the involvement of some species such as *T. koraiensis* and *T. plicata* in the putative ancient hybridization events really makes it difficult to investigate the evolutionary history of *Thuja*, but the gene trees still provide us some biogeographic information and seem to suggest a North American origin of the genus. In the *4CL* gene tree (Fig. 1C), the eastern North American species *T. occidentalis* has a basal position. Also, this species seems to represent an ancient lineage not closely related to any other extant species of *Thuja* in the combined nuclear gene tree (Fig. 2), though a sister relationship between the two North American species is weakly supported by a synapomorphic site mutation in the *LEAFY* gene tree (Fig. 1B).

Actually, the North American origin hypothesis of *Thuja* is also supported by fossil evidence. Although a number of fossils from late Cretaceous and Tertiary sediments were assigned to *Thuja* or *Thuites*, most of the *Thuja*-like vegetative remains should be reexamined and identified due to the similarity of foliage between several cupressaceous genera such as *Thuja*, *Chamaecyparis*, and *Calocedrus* (McIver and Basinger, 1989). LePage (2003) described the fossil species *Thuja smileya* with seed cones that was collected from the late Cretaceous of Alaska, but the identification of this species is still in doubt (Farjon, 2005). Till now, the earliest and most reliable fossils of *Thuja* are *T. ehrenswaerdi* from Greenland and *T. polaris* from the Ellesmere Island of Canada, both occurring in the Paleocene and in the polar regions of the Western Hemisphere (Schweitzer, 1974; McIver and Basinger, 1989). The younger fossils of this genus include *T. nipponica* from Japan and Sikhote Alin of Russia in the late Miocene (Akhmetiev, 1973; Huzioka and Uemura, 1973), and *T. occidentalis* from Peary Land of North Greenland in the Plio-Pleistocene (Bennike, 1990). It is obvious that *Thuja* and *Thuja*-like fossils are restricted to North America and northeastern Asia. *Thuja* grows in a cold and humid habitat and rarely occurs as a dominant species in the community, which might have limited its expansion in the warm climate of the early Tertiary.

Based on the molecular analyses and fossil records, it could be deduced that *Thuja* originated from the high-latitude areas of North America in the Paleocene or earlier. Then the genus was dispersed to eastern Asia through the Bering Land Bridge, and the subsequent vicariance gave rise to the species pair *T. standishii–T. sutchenensis*. Also, the Bering Land Bridge and/or the Aleutian bridge might have mediated further migrations of *Thuja* species and resulted in several interspecific hybridization events. One may argue that the DIVA analysis suggests a wide distribution of the common ancestor of *Thuja* in eastern Asia and North America and does not support the North American origin hypothesis. However, the ancestor distribution inferred from the DIVA analysis is not very reliable when approaching the root node. In particular, no appropriate outgroup can be used for the DIVA analysis in the present study due to the relict distribution of *Thujopsis*, and thus the ancestral area of *Thuja* can not be convincingly deduced by the software, in which vicariance events are emphasized (Fig. 3). It could also be argued that *Thuja* is more likely to have an origin in eastern Asia, where it has more extant species and its sister genus *Thujopsis*. However, it has been widely recognized that modern diversity center is not an indication of geographical place of origin (e.g., Manchester and Tiffney, 2001; Ran et al., 2006). Eastern Asia had served as a refugium for plenty of plant taxa including primitive and advanced types during the late Tertiary and Quaternary climate deterioration. In addition, the Japanese islands, as a typical "stepping stone" for plant migration (Gernandt and Liston, 1999; Wang et al., 2003), has conserved many endemic taxa, including conifers such as *Abies firma*, *Chamaecyparis pisifera*, *Picea polita*, *Pseudotsuga japonica*, and *Thujopsis dolabrata*.

It has been suggested that Eastern Asia could be subdivided into two separate refugial provinces. One includes the Japanese Archipelago, Korea Peninsula and the adjacent areas of northeast China, and the other consists of south and southeast China, with extensions to the Himalayas (Donoghue et al., 2001; Milne and Abbott, 2002). The two provinces had been separated by an arid zone since the Eocene and throughout the Miocene (Tiffney and Manchester, 2001). It could be inferred that the arid zone would have acted as a barrier helpful to the vicariance speciation of *T. standishii* and *T. sutchenensis*. This inference is consistent with the results of molecular clock estimation. Both cpDNA and the *4CL* gene analyses suggest that the split between *T. standishii* and *T. sutchuenensis* could have occurred in the Oligocene or early Miocene (cpDNA, 27.5 ± 5.22 Mya; *4CL*, 23.7 ± 5.04 Mya for LF).

4.3. The selection of molecular markers in the study of biogeographical disjunction between eastern Asia and North America

When it has become a trend that the biogeographical study uses information from many disciplines, including morphological, ecological, paleontological, paleoclimatic, paleogeologic, and molecular data, the most important thing is to reconstruct a robust phylogeny of the organisms under study. To date cpDNA and nrDNA ITS are still the most widely used molecular markers in studying interspecific relationships. The cpDNA has many merits, such as the relatively simple genetics (uniparental inheritance,

single copy, non-recombination) and conserved coding regions for the design of universal primers, but its independent use can not resolve hybridization events. The use of nrDNA ITS in phylogenetic analysis is also limited by paralogy, incomplete concerted evolution and pseudogene (Alvarez and Wendel, 2003), though this marker has advantages of biparental inheritance, low functional constraint underlying large sequence variation, high copy number and the universal primers available for most plants. Recently, the biparentally inherited single or low-copy nuclear genes are increasingly used in the phylogenetic reconstruction. These nuclear gene markers often exhibit high evolutionary rates and can offer multiple unlinked loci for independent phylogenetic inference (e.g., Oh and Potter, 2005; Whittall et al., 2006), which could increase the confidence of the phylogenetic trees. However, low-copy nuclear gene has its limitations such as shortage of universal primers, difficulty of copy number and orthology assessment, and lineage sorting (Sang, 2002; Small et al., 2004).

As discussed above, each gene marker has its own evolutionary trajectory and functional constraints, causing the limitations in molecular systematic studies. An integrative analysis of multiple gene markers with diverse origins can offset the limitations and would be very helpful to the phylogenetic studies of taxa with complex evolutionary history. For example, Cronn et al. (2003) successfully discovered the chimeric genome structure of *Gossypium gossypioidses*, a diploid cotton species native to Mexico, and clearly reconstructed the evolutionary history of the species based on an analysis of four chloroplast and ten nuclear genes. The multiple gene analysis is also very helpful to the study of historical biogeographical patterns such as the eastern Asia and North America disjunction, the formation of which involves many factors.

The hypothesis that floristic relationship is closer between eastern Asia and eastern North America than between eastern and western North America is still in controversy (Tiffney, 1985b; Graham, 1993; Xiang et al., 1998; Wen, 1999; Tiffney and Manchester, 2001). It was believed that the floristic exchange between eastern and western North America would have been interrupted by the mid-continental seaway occupying the central North America in the Late Cretaceous and the subsequent uplift of the Rocky Mountains accompanying with the retreat of the seaway in the Paleocene (Tiffney, 1985b; Graham, 1993). However, the studies of plant groups with a disjunct distribution in eastern Asia and both eastern and western North America indicate that eastern and western North America are biogeographically more closely related in most cases such as *Adiantum pedatum* complex (Paris and Haufler, 1994), *Trillium* (Kato et al., 1995), *Calycanthus*, *Aralia* sect. *Aralia* (Wen et al., 1996, 1998), *Boykinia*, *Cornus*, *Tiarella*, and *Trautvetteria* (Xiang et al., 1998). There are also some exceptional examples in which the eastern and western North America species are derived from different lineages such as *Styrax* (Fritsch, 1999, 2001) and *Chamaecyparis* (Wang et al., 2003). In the present study, we revealed a reticulate evolutionary history of *Thuja*, a genus with a typical disjunct distribution in eastern Asia and both eastern and western North America, based on five cpDNA fragments, nrDNA ITS and two low-copy nuclear genes. However, the reticulate evolution in *Thuja* was not discovered in the previous study using single molecular marker (Li and Xiang, 2005). Considering that most previous studies on plant groups with the eastern Asia and North America disjunction are based on uniparentally inherited cpDNA and (or) directly sequenced nrDNA ITS data, the historical reticulate evolution in these taxa might have been underestimated. This could be partially responsible for the difficulty in elucidating the relationship between eastern Asia, eastern and western Northern America. Therefore, we suggest that multiple genes from different genomes, especially low-copy nuclear genes, be used in the future biogeographical studies.

Acknowledgments

The authors thank the two anonymous reviewers for their insightful comments and suggestions on the manuscript. We also thank Prof. Zhenyu Li and Dr. Yingjun Qiu for their help in the sample collection; Drs. Junko Kusumi and Qiaoping Xiang for providing plant materials of *Thuja standishii* and *T. plicata*, respectively. This study was supported by National Basic Research Program of China (Grant No. 2007CB411600), the Chinese Academy of Sciences (Grant No. KZCX2-YW-415) and the National Natural Science Foundation of China (Grant Nos. 30425028, 30730010).

References

Akhmetiev, M.A., 1973. Paleocene and Eocene floras from southern Far East and their stratigraphic significance and climatic characteristics. Sov. Geol. 7, 14–29.

Alvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29, 417–434.

Bennike, O., 1990. The Kap Kobenhavn formation: Stratigraphy and paleobotany of a Plio-Pleistocene sequence in Peary Land, North Greenland. Meddel. Gronland 23, 1–85.

Boufford, D.E., Spongberg, S.A., 1983. Eastern Asian–eastern North American phytogeographical relationships—a history from the time of Linnaeus to the twentieth century. Ann. Mo. Bot. Gard. 70, 423–439.

Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56, 643–655.

Brunsfeld, S.J., Soltis, P.S., Soltis, D.E., Gadek, P.A., Quinn, C.J., Strenge, D.D., Ranker, T.A., 1994. Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae: evidence from *rbc*L sequences. Syst. Bot. 19, 253–262.

Campbell, C.S., Wright, W.A., Cox, M., Vining, T.F., Major, C.S., Arsenault, M.P., 2005. Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in *Picea* (Pinaceae): sequence divergence and structure. Mol. Phylogenet. Evol. 35, 165–185.

Cronn, R.C., Small, R.L., Haselkorn, T., Wendel, J.F., 2003. Cryptic repeated genome recombination during speciation in *Gossypium gossypioides*. Evolution 57, 2475–2489.

Cukovic, D., Ehlting, J., VanZiffle, J.A., Douglas, C.J., 2001. Structure and evolution of 4-coumarate: coenzyme A ligase (*4CL*) gene families. Biol. Chem. 382, 645–654.

Cunningham, C.W., 1997. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. 14, 733–740.

Cunningham, C.W., Collins, T.M., 1994. Developing model systems for molecular biogeography: vicariance and interchange in marine invertebrates. In: Schierwater, B., Streit, B., Wagner, G.P., DeSalle, R. (Eds.), Molecular Ecology and Evolution: Approaches and Applications. Birkhauser Verlad Base, Switzerland, pp. 405–433.

Demesure, B., Sodzi, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4, 129–131.

Donoghue, M.J., Bell, C.D., Li, J., 2001. Phylogenetic patterns in Northern Hemisphere plant geography. Int. J. Plant Sci. 162, S41–S52.

Downie, S.R., Katz-Downie, D.S., Cho, K.J., 1996. Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpoC*1 intron. Mol. Phylogenet. Evol. 6, 1–18.

Farjon, A., 2005. A Bibliography of Cupressaceae and *Sciadopitys*. Royal Botanic Gardens, Kew.

Farjon, A., Page, C.N., 1999. Conifers: Status Survey and Conservation Action Plan. IUCN-SSC Conifer Specialist Group, Gland, Switzerland and Cambridge, UK.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of incongruence. Cladistics 10, 315–319.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. 22, 521–565.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package) Version 3.6a2. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. Available from: http://evolution.gs.washington.edu/phylip.html.

Fehrer, J., Gemeinholzer, B., Chrtek, J., Brautigam, S., 2007. Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae). Mol. Phylogenet. Evol. 42, 347–361.

Fritsch, P.W., 1999. Phylogeny of *Styrax* based on morphological characters, with implications for biogeography and infrageneric classification. Syst. Bot. 24, 356–378.

Fritsch, P.W., 2001. Phylogeny and biogeography of the flowering plant genus *Styrax* (Styracaceae) based on chloroplast DNA restriction sites and DNA sequences of the internal transcribed spacer region. Mol. Phylogenet. Evol. 19, 387–408.

Frohlich, M.W., Parker, D.S., 2000. The mostly male theory of flower evolutionary origins: from genes to fossils. Syst. Bot. 25, 155–170.

Gadek, P.A., Alpers, D.L., Heslewood, M.M., Quinn, C.J., 2000. Relationships within Cupressaceae *sensu lato*: a combined morphological and molecular approach. Am. J. Bot. 87, 1044–1057.

Gernandt, D.S., Liston, A., 1999. Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). Am. J. Bot. 86, 711–723.

Gernandt, D.S., Liston, A., Piñero, D., 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: implications for molecular systematic studies of pine species complexes. Mol. Phylogenet. Evol. 21, 450–468.

Glaubitz, J.C., El-Kassaby, Y.A., Carlson, J.E., 1999. Nuclear restriction fragment length polymorphism analysis of genetic diversity in western redcedar. Can. J. Forest Res. 30, 379–389.

Gould, K.R., Donoghue, M.J., 1998. Phylogeny and biogeography of *Triosteum* (Dipascales, Lonicereae). Am. J. Bot. 85, 133.

Graham, A., 1993. History of the vegetation: Cretaceous (Maastrichtian)-Tertiary. In: Flora of North America Editorial Committee (Eds.), Flora of North America, vol. 1. Oxford University Press, pp. 57–70.

Gray, A., 1859. Diagnostic characters of new species of phanerogamous plants collected in Japan by Charles Wright, botanist of the U.S. North Pacific Exploring Expedition. Mem. Am. Acad. Arts 6, 377–452.

Grob, G.B.J., Gravendeel, B., Eurlings, M.C.M., 2004. Potential phylogenetic utility of the nuclear *FLORICAULA/LEAFY* second intron: comparison with three chloroplast DNA regions in *Amorphophallus* (Araceae). Mol. Phylogenet. Evol. 30, 13–23.

Hamberger, B., Hahlbrock, K., 2004. The 4-coumarate: CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proc. Natl. Acad. Sci. USA 101, 2209–2214.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.

Huang, S., Chiang, Y.-C., Schaal, B.A., Chou, C.-H., Chiang, T.-Y., 2001. Organelle DNA phylogeography of *Cycas taitungensis*, a relict species in Taiwan. Mol. Ecol. 10, 2669–2681.

Huzioka, K., Uemura, K., 1973. The late Miocene Miyata flora of Akita Prefecture, northeast Honshu, Japan. Bull. Natl. Sci. Mus. Tokyo 18, 661–738.

Kan, X.-Z., Wang, S.-S., Ding, X., Wang, X.-Q., 2007. Structural evolution of nrDNA ITS in Pinaceae and its phylogenetic implications. Mol. Phylogenet. Evol. 44, 765–777.

Kato, H., Kawano, S., Terauchi, R., Ohara, M., Utech, F.H., 1995. Evolutionary biology of *Trillium* and related genera (Trilliaceae): I. Restriction site mapping and variation of chloroplast DNA and its systematic implications. Plant Spec. Biol. 10, 17–29.

Kusumi, J., Sato, A., Tachida, H., 2006. Relaxation of functional constraint on light-independent protochlorophyllide oxidoreductase in *Thuja*. Mol. Biol. Evol. 23, 941–948.

Kusumi, J., Tsumura, Y., Yoshimaru, H., Tachida, H., 2000. Phylogenetic relationships in Taxodiaceae and Cupressaceae *sensu stricto* based on *mat*K gene, *chl*L gene, *trn*L–*trn* F IGS region, and *trn*L intron sequences. Am. J. Bot. 87, 1480–1488.

Langley, C.H., Fitch, W.M., 1974. An examination of the constancy of the rate of molecular evolution. J. Mol. Evol. 3, 161–177.

LePage, B.A., 2003. A new species of *Thuja* (Cupressaceae) from the Late Cretaceous of Alaska: implications of being evergreen in a polar environment. Am. J. Bot. 90, 167–174.

Li, J.-H., Xiang, Q.-P., 2005. Phylogeny and biogeography of *Thuja* L. (Cupressaceae), an Eastern Asian and North American disjunct genus. J. Integr. Plant Biol. 47, 651–659.

Liston, A., Parker-Defeniks, M., Syring, J.V., Willyard, A., Cronn, R., 2007. Interspecific phylogenetic analysis enhances intraspecific phylogeographical inference: a case study in *Pinus lambertiana*. Mol. Ecol. 16, 3926–3937.

Liston, A., Robinson, W.A., Oliphant, J.M., 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. Syst. Bot. 21, 109–120.

Maggini, F., Marrocco, R., Gelati, M.T., De Dominicis, R.I., 1998. Lengths and nucleotide sequences of the internal spacers of nuclear ribosomal DNA in gymnosperms and pteridophytes. Plant Syst. Evol. 213, 199–205.

Manchester, S.R., Tiffney, B.H., 2001. Integration of paleobotanical and neobotanical data in the assessment of phytogeographic history of holarctic angiosperm clades. Int. J. Plant Sci. 162, S19–S27.

Manos, P.S., Donoghue, M.J., 2001. Progress in Northern Hemisphere phytogeography: an introduction. Int. J. Plant Sci. 162, S1–S2.

Marshall, H.D., Newton, C., Ritland, K., 2001. Sequence repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. Mol. Biol. Evol. 18, 2136–2138.

McIver, E.E., Basinger, J.F., 1989. The morphology and relationships of *Thuja polaris* sp. nov (Cupressaceae) from the Early Tertiary, Ellesmere Island, Arctic Canada. Can. J. Bot. 67, 1903–1915.

Milne, R.I., 2006. Northern Hemisphere plant disjunctions: a window on Tertiary Land Bridges and climate change? Ann. Bot. 98, 465–472.

Milne, R.I., Abbott, R.J., 2002. The evolution of Tertiary relict floras. Adv. Bot. Res. 38, 281–314.

Mogensen, H.L., 1996. The hows and ways of cytoplasmic inheritance in seed plants. Am. J. Bot. 83, 383–404.

Neale, D.B., Marshall, K.A., Sederoff, R.R., 1989. Chloroplast and mitochondrial DNA are paternally inherited in *Sequoia sempervirens* D. Don Endl. Proc. Natl. Acad. Sci. USA 86, 9347–9349.

Nie, Z.-L., Sun, H., Li, H., Wen, J., 2006. Intercontinental biogeography of subfamily Orontioideae (*Symplocarpus*, *Lysichiton*, and *Orontium*) of Araceae in eastern Asia and North America. Mol. Phylogenet. Evol. 40, 155–165.

Oh, S.-H., Potter, D., 2005. Molecular phylogenetic systematics and biogeography of tribe Neillieae (Rosaceae) using DNA sequences of cpDNA, rDNA, and *LEAFY*. Am. J. Bot. 92, 179–192.

Paris, C.A., Haufler, C.H., 1994. Geographic isolation and allopatric speciation in ferns with a disjunct distribution in eastern Asia and eastern North America. Am. J. Bot. 77, 1243–1256.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 917–918.

Qiu, Y.-L., Chase, M.W., Parks, C.R., 1995. A chloroplast DNA phylogenetic study of the eastern Asia–eastern North America disjunct section *Rytidospermum* of *Magnolia* (Magnoliaceae). Am. J. Bot. 82, 1582–1588.

Quinn, C.J., Price, R.A., Gadek, P.A., 2002. Familial concepts and relationships in the conifers based on *rbc*L and *mat*K sequence comparisons. Kew Bull. 57, 513–531.

Ran, J.-H., Wei, X.-X., Wang, X.-Q., 2006. Molecular phylogeny and biogeography of *Picea* (Pinaceae): implications for phylogeographical studies using cytoplasmic haplotypes. Mol. Phylogenet. Evol. 41, 405–419.

Rieseberg, L.H., Soltis, D.E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trend Plant 5, 65–84.

Rogers, O.S., Bendich, A.J., 1988. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Mol. Biol. 5, 69–76.

Ronquist, F., 1996. DIVA. Ver. 1.1. Available from: http://www.ebc.uu.se/systzoo/research/diva/diva.html.

Ronquist, F., 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Syst. Biol. 46, 195–203.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sanderson, M.J., 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14, 1218–1231.

Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19, 101–109.

Sang, T., 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit. Rev. Biochem. Mol. Biol. 37, 121–147.

Schweitzer, H.J., 1974. Die "tertiaren" Koniferen Spitzbergens. Palaeontograhica 149B, 1–89.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Windel, C.T., Schilling, E.E., Small, R.L., 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92, 142–166.

Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am. J. Bot. 94, 275–288.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shiraishi, S., Maeda, H., Toda, T., Seido, K., Sasaki, Y., 2001. Incomplete paternal inheritance of chloroplast DNA recognized in *Chamaecyparis obtusa* using an intraspecific polymorphism of the *trn*D–*tm*Y intergenic spacer region. Theor. Appl. Genet. 102, 935–941.

Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear genes for phylogeny reconstruction in plants. Aust. Syst. Bot. 17, 145–170.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. 85, 1301–1315.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods), Version 4. Sinauer, Sunderland, MA.

Syring, J., Willyard, A., Cronn, R., Liston, A., 2005. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. Am. J. Bot. 92, 2086–2100.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Tiffney, B.H., 1985a. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. J. Arnold Arboretum 66, 73–94.

Tiffney, B.H., 1985b. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. J. Arnold Arboretum 66, 243–273.

Tiffney, B.H., Manchester, S.R., 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere tertiary. Int. J. Plant Sci. 162, S3–S17.

Wang, W.-P., Hwang, C.-Y., Lin, T.-P., Hwang, S.-Y., 2003. Historical biogeography and phylogenetic relationships of the genus *Chamaecyparis* (Cupressaceae) inferred from chloroplast DNA polymorphism. Plant Syst. Evol. 241, 13–28.

Wang, X.-Q., Tank, D.C., Sang, T., 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol. Biol. Evol. 17, 773–781.

Wei, X.-X., Wang, X.-Q., 2003. Phylogenetic split of *Larix*: evidence from paternally inherited cpDNA *trn*T–*trn*F region. Plant Syst. Evol. 239, 67–77.

Wei, X.-X., Wang, X.-Q., 2004a. Recolonization and radiation in *Larix* (Pinaceae): evidence from nuclear ribosomal DNA paralogues. Mol. Ecol. 13, 3115–3123.

Wei, X.-X., Wang, X.-Q., 2004b. Evolution of 4-coumarate: coenzyme A ligase (*4CL*) gene and divergence of *Larix* (Pinaceae). Mol. Phylogenet. Evol. 31, 542–553.

Wei, X.-X., Wang, X.-Q., Hong, D.-Y., 2003. Marked intragenomic heterogeneity and geographical differentiation of nrDNA ITS in *Larix potaninii* (Pinaceae). J. Mol. Evol. 57, 623–635.

Wen, J., 1999. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. Annu. Rev. Ecol. Syst. 30, 421–455.

Wen, J., Jansen, R.K., Zimmer, E.A., 1996. Phylogenetic relationships and DNA sequence divergence of eastern Asian and eastern North American disjunct plants. In: Nei, M., Takahata, N. (Eds.), Current Topics on Molecular Evolution. Published Jointly by Pennsylvania State University and the Graduate University for Advanced Studies, Hayama,Japan, pp.37–44.

Wen, J., Shi, S., 1999. A phylogenetic and biogeographic study of *Hamamelis* (Hamamelidaceae), an eastern Asian and eastern North American disjunct genus. Biochem. Syst. Ecol. 27, 55–66.

Wen, J., Shi, S., Jansen, R.K., Zimmer, E.A., 1998. Phylogeny and biogeography of *Aralia* sect. *Aralia* (Araliaceae). Am. J. Bot. 85, 866–875.

Wen, J., Zimmer, E.A., 1996. Phylogeny and biogeography of *Panax* L. (the ginseng genus, Araliaceae): inference from ITS sequences of nuclear ribosomal DNA. Mol. Phylogenet. Evol. 6, 167–177.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Shinsky, J.J., White, T.J. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, pp. 315–322.

Whittall, J.B., Medina-Marino, A., Zimmer, E.A., Hodges, S.A., 2006. Generating single-copy nuclear gene data for a recent adaptive radiation. Mol. Phylogenet. Evol. 39, 124–134.

Won, H., Renner, S.S., 2005. The internal transcribed spacer of nuclear ribosomal DNA in the gymnosperm *Gnetum*. Mol. Phylogenet. Evol. 36, 581–597.

Xiang, Q.-P., Farjon, A., Li, Z.-Y., Fu, L.-K., Liu, Z.-Y., 2002. *Thuja sutchuenensis*: a rediscovered species of the Cupressaceae. Bot. J. Linn. Soc. 140, 93.

Xiang, Q.-Y., Soltis, D.E., 2001. Dispersal-vicariance analyses of intercontinental disjuncts: historical biogeographical implications for angiosperms in the Northern Hemisphere. Int. J. Plant Sci. 162, S29–S39.

Xiang, Q.Y., Soltis, D.E., Soltis, P.S., 1998. The eastern Asian and eastern and western North American floristic disjunction: congruent phylogenetic patterns in seven diverse genera. Mol. Phylogenet. Evol. 10, 178–190.

Xiang, Q.-Y., Soltis, D.E., Soltis, P.S., Manchester, S.R., Crawford, D.J., 2000. Timing of eastern Asian–eastern North American floristic disjunction: molecular clock corroborates paleontological estimates. Mol. Phylogenet. Evol. 15, 462–472.

Yoder, A.D., Irwin, J.A., Payseur, B.A., 2001. Failure of the ILD to determine data combinability for slow Loris phylogeny. Syst. Biol. 50, 408–424.