Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 31 (2004) 542–553
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Evolution of 4-coumarate:coenzyme A ligase (4CL) gene and divergence of *Larix* (Pinaceae)

Xiao-Xin Wei and Xiao-Quan Wang*

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

Received 19 March 2003; revised 20 August 2003

Abstract

The evolutionary dynamics of the 4CL gene encoding 4-coumarate:coenzyme A ligase was investigated in the genus *Larix* (Pinaceae) by comparing copy number, GC content and codon usage, sequence divergence, and phylogenetic analysis. All 4CL clones of *Larix* formed a strongly supported monophyletic group, in which two robust clades (*4cl*A and *4cl*B) derived from an ancient gene duplication event in the common ancestor of *Larix* were identified. Further gene duplication in the *4cl*A clade gave rise to two subclades $4cl\text{A}_1$ and $4cl\text{A}_2$. Frequent duplication/deletion appears to be a common evolutionary phenomenon in the 4CL gene family and paralogous genes differ greatly in their evolution rate. The existence of *L. speciosa* in subclades $4cl\text{A}_1$ and $4cl\text{A}_2$ suggests that this species may represent a primitive form of *Larix* or the closest relative of the common ancestor of the Eurasian Sect. *Multiserialis*. In addition, cpDNA and nrDNA ITS analyses support the hypothesis of an early separation of *Larix* into a North American and a Eurasian clade, which is congruent with the results of previous allozyme and very recent AFLP analyses. The unexpected close relationship between North American larches and the short-bracted species *L. gmelinii* in East Asia, based on the 4CL gene tree, may stem from lineage sorting.

Keywords: 4CL; Molecular evolution; Biogeography; *Larix*; *Pseudotsuga*

1. Introduction

After the surge in the use of cpDNA and nrDNA in studies of plant molecular systematics and evolution, protein-coding nuclear genes have been increasingly used in retrieving the evolutionary history of organisms and elucidating mechanisms of molecular evolution (Clegg et al., 1997; Mason-Gamer et al., 1998; Mathews and Donoghue, 1999; Wen et al., 1997; Zhang et al., 2001). These genes either have high copy numbers such as gene families encoding R (Purugganan and Wessler, 1994) and MADS-box regulatory proteins (Alvarez-Buylla et al., 2000; Becker et al., 2000; Theissen et al., 2000), chalcone synthase (Durbin et al., 1995) and small heat-shock proteins (Waters, 1995a,b), or occur in a single or low number of copies such as *ncpGS* (Emshwiller and Doyle, 1999), *Gpat* (Tank and Sang, 2001), *Adh* (Cronn et al., 1999; Gaut et al., 1996; Morton et al., 1996; Sang et al., 1997; Small and Wendel, 2000), *PHY* (Mathews and Sharrock, 1996; Mathews et al., 1995; Mathews et al., 2000; Schmidt and Schneider-Poetsch, 2002), *PgiC* (Ford and Gottlieb, 1999; Gottlieb and Ford, 1996) and GS-AOP genes (Kliebenstein et al., 2001). As systematists obtain increasing amounts of phylogenetic information from nuclear gene families, more examples of conflicts between gene trees and species trees become apparent. These conflicts can arise from molecular-, population- or organismal-level processes, such as gene duplication, recombination, lineage sorting and horizontal gene transfer (Doyle, 1997; Lyons-Weiler and Milinkovitch, 1997; Page and Charleston, 1997). It is of great interest to investigate the frequency of gene duplication and the evolutionary fate and consequences of duplicate genes (Lynch and Conery, 2000; Zhang et al., 2001). In addition, studies on the evolutionary dynamics of nuclear gene families cannot only be used to resolve the conflicts between gene trees and species trees, but also to shed light on possible

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

1055-7903/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2003.08.015

mechanisms involved in the formation of different genome size and organization.

In comparison to the relatively simple nuclear genomes of angiosperms, gymnosperms are characterized by a large nuclear genome (Murray, 1998) and highly complex gene families (Kvarnheden et al., 1995; Kinlaw and Neale, 1997; Perry and Furnier, 1996; Schubert et al., 1998). For example, the *Adh* gene has at least seven expressed copies in *Pinus banksiana* (Perry and Furnier, 1996) as opposed to usually two or three copies in angiosperms. Also, many processed retropseudogene copies have been found in the *cdc*2 gene family of *Picea* (Kvarnheden et al., 1995). To our knowledge, only a few phylogenetic analyses have been carried out for protein-coding nuclear genes of the gymnosperm families (Kusumi et al., 2002; Schmidt and Schneider-Poetsch, 2002; Wang et al., 2000; Winter et al., 1999). Based on sequence analysis of the low copy nuclear gene 4CL encoding 4-coumarate:coenzyme A ligase in the lignin biosynthetic pathway (Zhang and Chiang, 1997), Wang et al. (2000) constructed a molecular phylogeny of Pinaceae represented by all 11 widely accepted genera, demonstrating a tempo of duplication and deletion in the 4CL gene family that paralogous loci are maintained within rather than between genera. That is to say, frequent duplication of the 4CL gene occurred in respective genera, and thus, resulted in the existence of paralogous loci among congeneric species. Therefore, it is interesting to explore the universality of this tempo by sampling more species from the same genus, and to evaluate the utility of the 4CL gene in phylogenetic studies at the interspecific level. Moreover, an analysis of the structure and organization of gene families is critical for understanding the evolution of the genome in the family Pinaceae (Kinlaw and Neale, 1997), and revealing different evolutionary fates of duplicated gene copies, such as neofunctionalization, subfunctionalization, and nonfunctionalization, which may help explain adaptive evolution (Lynch, 2002).

Larix is a prominent component of the boreal forest and a young genus in the pine family according to fossil records (LePage and Basinger, 1991; Schorn, 1994) and molecular clock estimates based on the *mat*K gene (Wang et al., 2000). It comprises about 15 species, 10 of which are widely accepted (Farjon, 1990, 2001; Fu et al., 1999). Three species, *L. laricina*, *L. occidentalis*, and *L. lyallii*, are endemic to North America while the others are distributed in Eurasia. Patschke (1913) initially divided the genus into two sections based on the morphology of the female cone. Sect. *Larix* (or *Pauciserialis*) comprised species characterized by bracts on the cone that did not extend well beyond the seed scales and Sect. *Multiserialis* included species with bracts extending far beyond the seed scales. The classification of *Larix* in the traditional sense (Farjon, 1990; LePage and Basinger, 1991, 1995; Ostenfeld and Larsen, 1930; Patschke, 1913; Schorn, 1994), using scales of the female cone as the fundamental character, is in conflict with the results of molecular analyses (Gernandt and Liston, 1999; Qian et al., 1995). According to sequence analysis of the nrDNA internal transcribed spacer (ITS), the genus *Larix* was divided into a North American and a Eurasian clade (Gernandt and Liston, 1999). This biogeographical split was consistent with the result of allozyme analysis (Semerikov and Lascoux, 1999). However, the sample size of the Eurasian species of Sect. *Multiserialis* was small or nonexistent in the two studies. Hence, further molecular evidence is needed for the examination of species phylogeny in the genus *Larix*.

In the present study, we cloned and analyzed the low-copy nuclear gene 4CL from nearly all larch species to investigate the evolutionary dynamics of this gene family. In addition to phylogenetic analysis of the 4CL gene, the cpDNA (Wei and Wang, 2003) and nrDNA (Gernandt and Liston, 1999) data, including new nrDNA ITS sequences presented in this study, were reanalyzed in order to deduce the evolutionary history of *Larix* and address the question of gene tree vs. species tree.

2. Materials and methods

2.1. Plant materials

In the 4CL gene study, all recognized species and some varieties of *Larix* (Farjon, 1990, 2001; Fu et al., 1999) were sampled except *L. lyallii*, a long bract species endemic to the subalpine regions of the Cascade and Rocky Mountains of North America. Considering the morphological similarities, geographic proximity, interspecific gene flow (Carlson and Theroux, 1993) and identical ITS sequence (AF041346–AF041347) shared by *L. occidentalis* and *L. lyallii* (Gernandt and Liston, 1999), we employed *L. occidentalis* to represent the North American species of Sect. *Multiserialis*. The 4CL gene sequences of *Pseudotsuga menziesii* were also determined for use as outgroups. In order to compare substitution patterns, divergence rates and evolutionary dynamics of the 4CL gene in *Larix* with several other genera of Pinaceae, we also analyzed some 4CL gene sequences retrieved from GenBank, including those of *Pseudotsuga* (AF144507–AF144510), *Abies* (AF144514–AF144521), and *Pinus* (AF144499–AF144503, U39404–U39405) (Wang et al., 2000; Zhang and Chiang, 1997). Moreover, sequences of the nrDNA ITS region were determined for the same individuals of two varieties of *L. gmelinii*, *L. olgensis* and all Eurasian species of Sect. *Multiserialis* except *L. griffithii* because these species were not sampled in the study of Gernandt and Liston (1999). Additional nrDNA ITS region sequences (AF041343–AF041350; AF041353) (Gernandt and Liston, 1999) and all sequences of the cpDNA *trn*T–*trn*F

Table 1
Sources of materials in this study

Species	Sources/voucher	Clones and GenBank Accession Nos.
Eurasian *Larix*		
Sect. Multiserialis		
L. griffithii Hooker	Linzhi, Xizang, China/LW9802	Clones 1, 2, 3/AF470407–AF470409
L. himalaica Cheng & Fu	Jilong, Xizang, China/LSLJL01	Clones 1, 8, 9/AF470410–AF470412
L. mastersiana Rehder & Wilson	Wolong Natural Reserve, Sichuan, China/LW83101	Clones 8, 9, 14/AF470413–AF470415
L. potaninii var.*australis* Henry & Hand.-Mazz.	Lijiang, Yunnan. China/LKY1115	Clones 12, 14/AF470416–AF470417
L. potaninii var.*chinensis* Fu & Li	Qinling, Shaanxi, China/LW92405	Clones 8, 9, 17/AF470418–AF470420
L. potaninii var. *potaninii* Batalin	Kangding, Sichuan, China/LW9301	Clones 3, 6/AF470421–AF470422
L. speciosa Cheng & Law	Jianchuan, Yunnan, China/LYJC02	Clones 1, 2, 11/AF470423–AF470425
Sect. Larix		
L. decidua Mill.	Campus of Michigan State University, USA/LW9904	Clones 32, 34/AF470426–AF470427
L. gmelinii var. *gmelinii* (Rupr.) Kuzen.	Botanic Garden, Institute of Botany, Beijing/LW966	Clones 1, 2, 14 / AF470428–AF470430
L. gmelinii var. *principis-rupprechtii* (Mayr) Pilger	Wolong Natural Reserve, Sichuan, China/LW83103	Clones 1, 2, 3, 6, 16, 32, 37/AF470431–AF470437
L. kaempferi (Lamb.) Carr.	Cultivated in Linzhi, Xizang, China/LW9806	Clones 3, 4, 27, 29/AF470438–AF470441
L. olgensis Henry	Botanic Garden, Institute of Botany, Beijing/LW2169	Clones 1, 14/AF470442–AF470443
L. sibirica Ledeb.	Altai Mountains, Xinjiang, China/LTXJ11	Clones 8, 12/AF470444–AF470445
North American *Larix*		
Sect. Multiserialis		
L. occidentalis Nutt.	Washington State, USA/LDT2109	Clones 5, 9, 21/AF470446–AF470448
Sect. Larix		
L. laricina (Du Roi) K. Koch	Arboretum of Shenyang, Liaoning, China/LW2006	Clones 20, 31/AF470449–AF470450
Pseudotsuga species		
P. menziesii (Mirbel) Franco	Botanic Garden, Institute of Botany, Beijing/PW965	Clones 2, 3, 12/AF470451–AF470453
	Wang et al. (2000)	Clones 3, 5/AF144507–AF144508
P. sinensis Dode	Wang et al. (2000)	Clones 6, 9 / AF144509–AF144510

region (AF440484–AF440491, AF440493, AF440496–AF440499, AF440501–AF440504) (Wei and Wang, 2003) analyzed in the present study were retrieved from GenBank. The origins of the materials are shown in Table 1. Voucher specimens have been deposited in the Institute of Botany, Beijing (PE) herbarium.

2.2. DNA extraction, gene amplification, cloning and sequencing

Total DNA was extracted from silica gel dried needles using the CTAB method following the protocol of Rogers and Bendich (1988) and used as template in polymerase chain reaction amplification. Primers used for 4CL gene amplification were from Wang et al. (2000) and are as follows: $4CLpF_3$ (5′ CCAATCCTTTYTACA AGCCG) and $4CLpR_3$ (5′ CGGGGAARGGCTYCT TTGC). The positions of the primers are shown in Fig. 1. The nrDNA ITS region was amplified using primers ITS1N (5′ GTCGTAACAAGGTTTCCG-TAGG) on the 18S gene and ITS4 from White et al. (1990). PCR amplification was completed using 25 μl reactions containing 5–50 ng of DNA template,

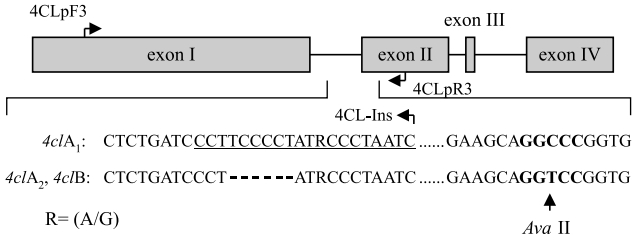

Fig. 1. Schematic representation of the nuclear gene 4CL in Pinaceae. Exons are represented by shaded boxes. Arrows indicate the locations and directions of primers used in the present study. The *4cl*A_1-specific primer (4CL-Ins) was designed according to the sequence underlined. The bases in bold type show the recognition site of *Ava*II.

6.25 pmol of each primer, 0.2 mM of each dNTP, 2 mM $MgCl_2$ and 0.75 Units of *Taq* DNA polymerase. Amplification was conducted in a Peltier Thermal Cycler (PTC-200, MJ RESEARCH). PCR cycles were as follows: one cycle of 4 min at 70 °C, 4 cycles of 40 s at 94 °C, 20 s at 52–55 °C, and 2 min at 72 °C, followed by 36 cycles of 20 s at 94 °C, 20 s at 52–55 °C, and 90 s at 72 °C, with a final extension step for 10 min at 72 °C.

PCR products were separated by 1.5% agarose gel electrophoresis. The only strong band was excised in

order to separate the target product from nonspecific PCR products with very low molecular weight. The gel slices were then purified using GFX PCR DNA and Gel Band Purification Kit (Pharmacia). For the 4CL gene, the purified PCR products were cloned with pGEM-T Easy Vector System II (Promega). For each species, 16–36 clones with correct insertion (determined by digestion with *Eco*RI) were screened by comparing restriction fragments of *Msp*I. All distinct clones were sequenced in both directions. Sequencing reactions were performed with the two PCR primers using ABI Prism Bigdye Terminator Cycle Sequencing Ready Reaction Kit. For the nrDNA ITS region, several internal primers were used. The sequencing reaction products were applied to an ABI 377 automatic sequencer (PE Applied Biosystems) after purification through precipitation with 95% EtOH and 3 M NaAc (pH 5.2).

2.3. Data analysis

DAMBE (Xia and Xie, 2001) was used to investigate the distribution of variable nucleotide sites in different codon positions. MEGA version 2.1 (Kumar et al., 2001) was applied to molecular evolution analyses of the 4CL gene, including GC content and codon usage (bias), as well as distances of synonymous (d_S), nonsynonymous (d_N) and nucleotide (d) substitutions. The d_S and d_N values were calculated according to the Jukes-Cantor model in the modified Nei-Gojobori method (Nei and Gojobori, 1986), while the d value was generated based on the Kimura two-parameter model, in which the transitional and transversional substitution rates were considered (Kimura, 1980).

Sequence alignments were made with CLUSTAL X (Thompson et al., 1997) and refined manually. Some distinctive recombinant sequences of the 4CL gene, which might result from PCR errors (Bradley and Hillis, 1997), were excluded from the dataset and will be investigated in our following studies. Maximum parsimony analyses of the aligned datasets, including 4CL, cpDNA *trn*T–*trn*F and nrDNA ITS regions, were performed using PAUP* (v. 4.0) program (Swofford, 1998). Two *Pseudotsuga* species were used as an outgroup clade because the sister group relationship between *Pseudotsuga* and *Larix* was supported by most previous studies, such as morphological (Hart, 1987), immunological (Price et al., 1987), pollination biology (Owens et al., 1998), and various molecular analyses (Chaw et al., 1997; Tsumura et al., 1995; Wang et al., 1997, 1998a,b, 2000). All character states were specified as unordered and equally weighted with indels as missing data. Heuristic searches were implemented with simple sequence addition, tree-bisection-reconnection (TBR) branch swapping, and MULTREES options. To evaluate relative robustness of the clades found in the most parsimonious trees, the bootstrap analysis (Felsenstein, 1985) was conducted with 500 replicates using the same options as those in the Heuristic search. Considering that the Maximum Likelihood (ML) method is very suitable for analyzing sequence data with unequal rates of substitutions (Kuhner and Felsenstein, 1994), we also constructed the 4CL gene phylogeny with the ML method in PAUP (v. 4.0) based on the Hasegawa–Kishino–Yano (HKY) substitution model (Hasegawa et al., 1985). The heuristic search options for the ML analysis are ASIS addition sequences and TBR branch swapping.

3. Results

3.1. 4CL gene family and phylogenetic analyses

The length of the 4CL region amplified with primers $4CLpF_3$ and $4CLpR_3$ was 825–832 bp in *Larix* compared to 830–833 bp in the two *Pseudotsuga* species. Two to four distinct 4CL gene members were found in all species of *Larix* except *L. gmelinii* var. *principis-rupprechtii*, which included as many as seven 4CL gene members. Likewise, *Pseudotsuga sinensis* and *P. menziesii* had two and five 4CL members, respectively. The aligned sequences of both ingroups and outgroups, excluding some nucleotides on both ends of the 4CL gene fragment, were 756 bp in length, comprising partial sequences of exons I (633 bp) and II (30 bp) as well as the complete intron sequence (93 bp) between them. The alignment contained 160 variable sites of which 108 were phylogenetically informative.

Parsimony analysis using a heuristic search generated 50 most parsimonious trees with a tree length of 200 steps (consistency index = 0.8550 and retention index = 0.9711). One tree that was almost identical with the strict consensus tree in topology is shown in Fig. 2. In this tree all 4CL clones of *Larix* clustered together with strong bootstrap support (96%). Two robust clades (bootstrap value = 100%), each of which comprised all sampled *Larix* species, were found and designated as *4cl*A and *4cl*B, but the *4cl*B clade had very poor resolution of interspecific relationships. The *4cl*A clade was further divided into two subclades, namely $4cl\mathrm{A}_1$ and $4cl\mathrm{A}_2$. The $4cl\mathrm{A}_1$ subclade, comprising the Eurasian species of Sect. *Larix*, *L. speciosa* and the two North American larches sampled, was not very well supported by bootstrap analysis (76%). However, all clones in this subclade were characterized by a 6-bp insertion in the intron and a unique base "C" at nt 24 of exon II from the 5′ end (Fig. 1). In particular, a close relationship between the two North American species and the Eurasian short-bracted species *L. gmelinii* was strongly supported (bootstrap value = 94%). Since this relationship was never found in the previous molecular analyses, we recloned the 4CL gene from *L. gmelinii* var. *gmelinii*

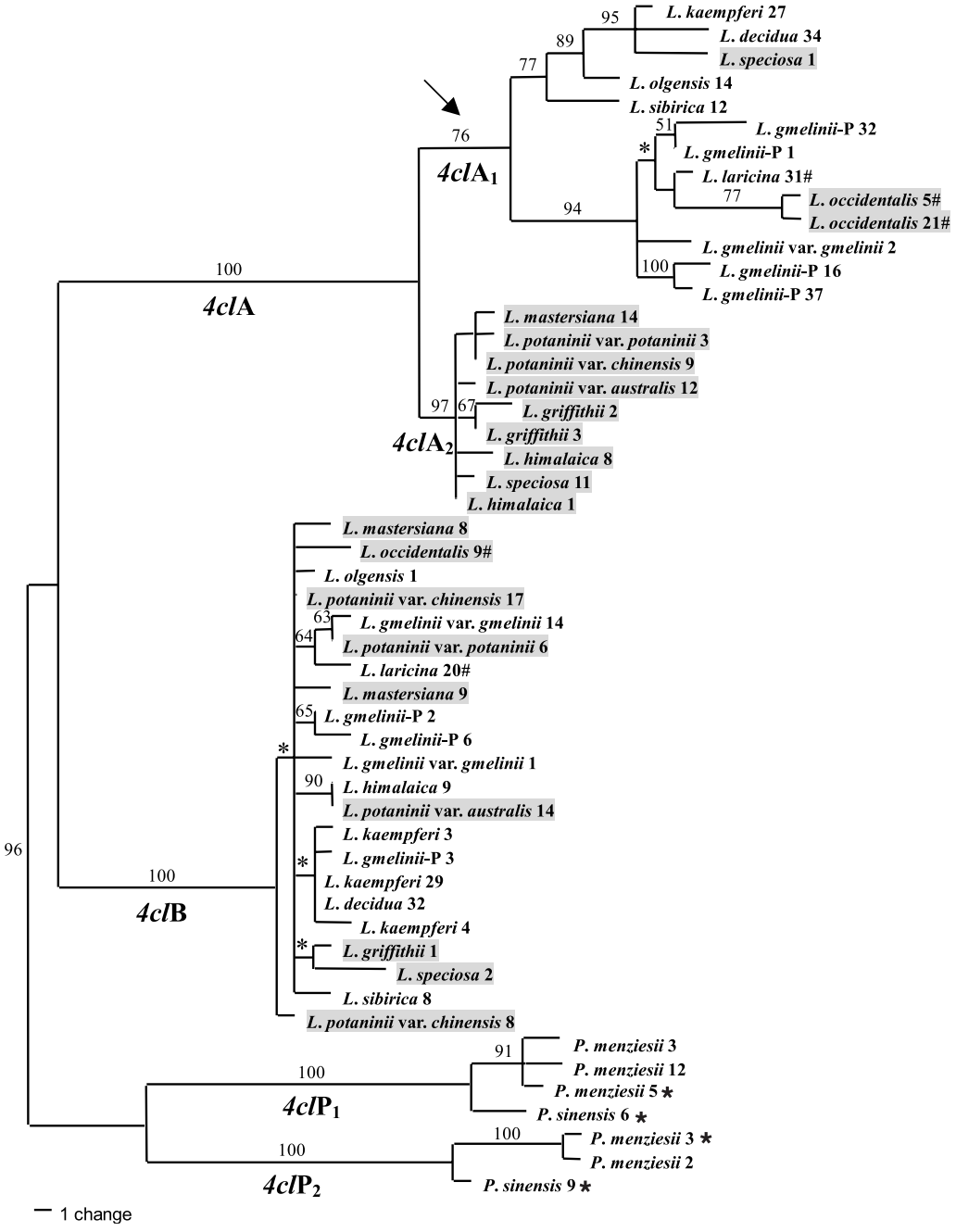

Fig. 2. One of the 50 most parsimonious trees constructed from sequence analysis of the nuclear gene 4CL with two *Pseudotsuga* species as outgroups (length = 200; CI = 0.8550; RI = 0.9711). Numbers associated with branches are bootstrap percentages greater than 50%. Small numbers following a species name represent clone numbers. The arrow indicates that the clones in this clade have a 6-bp insertion. *4cl*A, *4cl*B, *4cl*A_1, *4cl*A_2, *4cl*P_1, and *4cl*P_2 denote different clades. Species with gray shading belong to sect. *Multiserialis*. #, Species from North America; *, branch collapsing on the strict consensus tree; ★, sequences from GenBank. *L. gmelinii*-P: *L. gmelinii* var. *principis-rupprechtii*.

and *L. gmelinii* var. *principis-rupprechtii* and obtained the same result. The *4cl*A_2 subclade was composed of the Eurasian species of Sect. *Multiserialis*. Although this subclade was robustly supported (97%), it had no resolution of interspecific relationships. Surprisingly, *L. speciosa*, a long-bracted species endemic to southwest China, appeared in both subclades *4cl*A_1 and *4cl*A_2. Like *Larix*, all 4CL clones from *Pseudotsuga* were also clustered into two strongly supported clades (100%), namely *4cl*P_1 and *4cl*P_2, and each clade included the two *Pseudotsuga* species. The maximum likelihood analysis of the 4CL data yielded a tree (not shown) with the same topology as that of the MP tree (Fig. 2).

To determine whether the *4cl*A_1-type copy exists in other Eurasian species of Sect. *Multiserialis*, a *4cl*A_1-specific primer (4CL-Ins: 5′-GATTAGGGYATAGGG

GAAGG) was designed based on the obtained sequences (Fig. 1). This primer was used with $4CLpF_3$ in an attempt to amplify this 4CL gene copy from the genomic DNA of each Eurasian species of Sect. *Multiserialis*. Under stringent conditions (low concentration of $MgCl_2$ and high annealing temperature in the amplification), PCR product was only obtained from *L. speciosa* and had the sequence identical to that of *L. speciosa* 1 (Fig. 2). In addition, the *4cl*A_1-type sequences have a unique base ("C" instead of "T") at nt 24 of exon II when compared to *4cl*A_2- and *4cl*B-type sequences (Fig. 1), and therefore, cannot be digested by the restriction enzyme *Ava*II with recognition sequence "G' GWC C". This restriction enzyme was further used to investigate the distribution of the *4cl*A_1-type sequence. Theoretically, the PCR product amplified with primers $4CLpF_3$ and $4CLpR_3$ would be partially digested by *Ava*II for the species with both *4cl*A_1- and *4cl*A_2-type sequences. However, the partial digestion pattern was not found in Eurasian Sect. *Multiserialis* but was found in *L. speciosa*. All of the above evidence confirmed that the *4cl*A_1-type sequence does not exist in the Eurasian species of Sect. *Multiserialis*, except for *L. speciosa*.

The nrDNA ITS region of the two varieties of *L. gmelinii*, *L. olgensis*, and all Eurasian species of Sect. *Multiserialis* except *L. griffithii* ranged from 1761 to 1769 bp in size (GenBank Accession Nos. AF538060–AF538068). The aligned sequences of nrDNA ITS and cpDNA *trn*T–*trn*F regions included 1903 and 1372 characters, respectively. Fifteen most parsimonious trees (length = 383, CI = 0.9452 and RI = 0.9547) were obtained from the analysis of the ITS dataset, one of which is shown in Fig. 3A. In contrast, the analysis of the *trn*T-*trn*F region generated two equally most parsimonious trees with a tree length = 53 (CI = 1.0 and RI = 1.0). The *trn*T–*trn*F phylogeny of *Larix* is shown in Fig. 3B. Both cpDNA and nrDNA analyses strongly support the sister group relationship between North American and Eurasian larch species.

3.2. Sequence composition and divergence of the 4CL gene

The average GC content of the 4CL gene was similar throughout all clades of *Larix* and *Pseudotsuga*, ranging from 54.1 to 54.9% in the exons compared to 44.0–45.4% in the introns. As with most plants, the GC content was lower in the introns than in the exons (Emshwiller and Doyle, 1999). However, this value showed great codon position bias, being 58.0–60.2%, 38.1–39.2%, and 65.5–67.0% at 1st, 2nd, and 3rd positions, respectively. Like most other protein-coding genes (Emshwiller and Doyle, 1999; Wang et al., 1998a), the polymorphic site number (PSN) was significantly higher at the 3rd codon position than at the other two codon

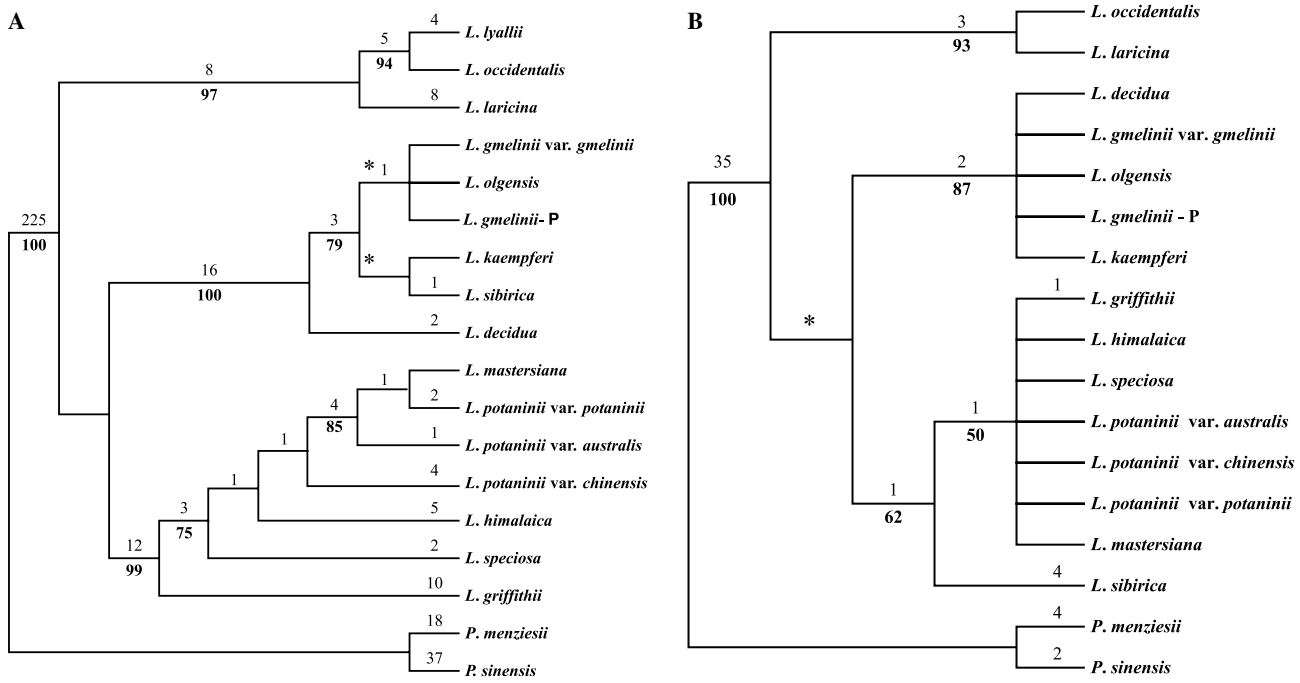

Fig. 3. Phylogenies of nrDNA ITS and cpDNA *trn*T–*trn*F regions of *Larix*. (A) One of the 15 most parsimonious trees constructed from sequence analysis of the nrDNA ITS region (length = 383; CI = 0.9452; RI = 0.9547). (B) One of the two equally most parsimonious trees constructed from sequence analysis of the cpDNA *trn*T–*trn*F region (length = 53; CI = 1.0; RI = 1.0). *, branches collapsing on the strict consensus tree. Numbers above the branches represent branch lengths. Numbers below branches are bootstrap percentages greater than 50%. *L. gmelinii*-P: *L. gmelinii* var. *principis-rupprechtii*.

positions. However, this pattern was not found in *4cl*P_1 and *4cl*P_2, two subclades with only a few sequences analyzed (Table 2).

According to Fig. 1C in Wang et al. (2000), the eight sequences of *Abies* formed two paralogous gene clades (designated here as *4cl*AB_1 and *4cl*AB_2) while the seven pine sequences clustered into two groups corresponding to the two subgenera of *Pinus* (referred here as PsubI and PsubII). The mean d_S, d_N, $R(t_S/t_V)$, and $d(t_S + t_V)$ values within and between different lineages are shown in Table 3. In most cases, the d_S value was substantially higher than the d_N value. For example, the d_S value was more than seven times higher than the d_N value in *Pinus*. However, d_S could be similar to d_N when the d value, the distance of nucleotide substitutions, was very low. The R value ranged in size from 1.173 to 6.055 in different lineages. Among the four genera studied (*Larix*, *Pseudotsuga*, *Abies*, and *Pinus*), *Pinus* had the largest d_S (0.118) and the smallest d_N (0.014) values at the generic level. It is very interesting that each of the d_S, d_N, R, and d values was strikingly higher in *4cl*A than in *4cl*B. Surprisingly, the d value between the two subgenera of *Pinus* was lower than that between *4cl*A and *4cl*B of *Larix*.

Table 2
Average GC content and polymorphic site number (PSN) in each codon position of the 4CL gene within different lineages

	Larix	*4cl*A	*4cl*B	*Pseudo-tsuga*	*4cl*P_1	*4cl*P_2
GC%						
Position-1	58.4	58.0	58.7	59.7	59.3	60.2
Position-2	38.8	38.4	39.2	38.6	38.1	38.1
Position-3	65.8	65.8	65.7	65.8	67.0	65.5
PSN						
Position-1	22	10	4	8	2	3
Position-2	26	12	11	9	4	3
Position-3	58	30	15	25	3	2

Table 3
Mean distance of synonymous (d_S), nonsynonymous (d_N), and nucleotide (d) substitutions, as well as ratio of transitions to transversions (R) calculated from the 4CL exon sequences within and among different lineages lineages (bold values)

Lineage	d_S	d_N	$R(t_S/t_V)$	$d(t_S + t_V)$
Larix	0.092	0.025	2.778	0.041
4clA_1	0.040	0.013	3.977	0.019
4clA_1	0.035	0.011	2.217	0.017
4clA_2	0.007	0.003	1.504	0.004
4clA_1 vs *4cl*A_2	**0.053**	**0.017**	**5.416**	**0.026**
*4cl*B	0.010	0.004	1.456	0.006
*4cl*A vs *4cl*B	**0.155**	**0.040**	**2.621**	**0.068**
Pseudotsuga	0.075	0.018	2.255	0.032
4clP_1	0.012	0.005	3.322	0.007
4clP_2	0.008	0.008	2.518	0.008
4clP_1 vs *4cl*P_2	**0.123**	**0.027**	**1.766**	**0.050**
Abies	0.055	0.022	1.364	0.030
4clAB_1	0.017	0.009	1.558	0.011
4clAB_2	0.004	0.005	1.672	0.005
4clAB_1 vs *4cl*AB_2	**0.090**	**0.033**	**1.173**	**0.047**
Pinus	0.118	0.014	2.303	0.038
PsubI	0.016	0.004	1.505	0.007
PsubII	0.012	0.007	6.055	0.008
PsubI vs Psub II	**0.188**	**0.019**	**1.736**	**0.059**
Larix vs *Pseudotsuga*	**0.154**	**0.035**	**2.638**	**0.063**

4. Discussion

4.1. Evolution of the 4CL gene

Two distinct 4CL genes (*lp4cl-1* and *lp4cl-2*) were previously found in *Pinus taeda* (Zhang and Chiang, 1997), and they shared identical nucleotide sequence in the region we analyzed. In the present study, two to four distinct 4CL gene members were found in most species of *Larix* and as many as seven 4CL members in *L. gmelinii* var. *principis-rupprechtii*. All *Larix* 4CL sequences determined in this study grouped into two robustly supported paralogous clades, *4cl*A and *4cl*B. Each clade comprised clones from all species of *Larix*, and in some cases more than two clones of the same species were found in one clade or subclade (Fig. 2). In contrast with *lp4cl-1* and *lp4cl-2* in *P. taeda*, *4cl*A and *4cl*B, including *4cl*A_1 and *4cl*A_2, of *Larix* differ greatly in both amino acid and nucleotide sequences as evidenced by the high d and d_N values between them (Table 3). Therefore, they are likely to be paralogous genes rather than alleles. This evidence indicates that there are two, three or more 4CL loci in one larch species. A similar pattern also exists in *Pseudotsuga*, the sister group of *Larix*. For example, two and five 4CL members were cloned from *Pseudotsuga sinensis* and *P. menziesii*, respectively (Fig. 2). Our results are in agreement with the findings that two or three major 4CL loci exist in Pinaceae and frequent duplication of the 4CL gene has happened in each genus after the divergence of genera (Wang et al., 2000). The congeneric clustering of 4CL genes in the pine family is similar to the evolution pattern of CAD, another nuclear gene encoding cinnamyl alcohol dehydrogenase in the same lignin biosynthetic pathway (Schubert et al., 1998). In addition, divergent 4CL genes have also been reported from some angiosperms (Cukovic et al., 2001; Uhlmann and Ebel, 1993).

The sister clades *4cl*A and *4cl*B might result from an ancient duplication in the common ancestor of *Larix* given the coexistence of all species in each clade (Fig. 2) and the highest d value (0.068) of *4cl*A vs *4cl*B. After that, they evolved at different rates since all the d_S, d_N, R, and d values are strikingly higher in *4cl*A than in *4cl*B (Table 3). This pattern was also observed in the two subclades of *4cl*A. For example, the divergence within

4clA_1 is nearly four times as great as that within *4cl*A_2. As discussed in the former paragraph, *4cl*A_1 and *4cl*A_2 most likely represent paralogous genes rather than alleles. They might have arisen from another gene duplication event in view of the double placement of *L. speciosa* in the two subclades. However, the duplication giving rise to *4cl*A_1 and *4cl*A_2 from an ancestral *4cl*A-type gene occurred later than that resulting in *4cl*A and *4cl*B considering the topology of the 4CL gene tree and the very low *d* value (0.026) of *4cl*A_1 vs *4cl*A_2 in comparison with that (0.068) of *4cl*A vs *4cl*B. Like the 4CL gene in *Larix*, *Pseudotsuga* and *Abies* (Wang et al., 2000), an ancient gene duplication prior to the species divergence was also reported for the *Adh* gene of *Paeonia* (Sang et al., 1997) and CHS gene family of *Ipomoea* (Durbin et al., 1995). The markedly fast evolutionary rates of *4cl*A and *4cl*A_1, correlating with their high *R* values, may suggest that paralogous 4CL genes have a different fate after the duplication events due to a reduction in selective constraint on at least one duplicate copy together with a commensurate increase in the amino acid replacement rate (Gaut et al., 1999). Walsh (1995) suggested that a lack of selective constraint is consistent with eventual pseudogenization. However, it is possible that there are still a few survivors subsequently experiencing strong purifying selection although a vast majority of gene duplicates are silenced within a few million years (Lynch and Conery, 2000).

Because paralogous 4CL loci are kept within a genus rather than between different genera (Wang et al., 2000), we would expect that sequence divergence between the sister clades *4cl*A and *4cl*B of *Larix* is lower than that between *Larix* and *Pseudotsuga* if this gene has been evolving with a molecular clock. Surprisingly, the *d* value (0.068) of *4cl*A vs *4cl*B is slightly higher than that (0.063) of *Larix* vs *Pseudotsuga* (Table 3), which further supports the point that 4CL genes evolve at very different rates in different lineages. Moreover, all 4CL clones from each subgenus or species of *Pinus* formed a strongly supported monophyletic group. In other words, paralogous 4CL loci do not exist between subgenera or species of *Pinus* (Wang et al., 2000, Fig. 1C). Based on fossil and molecular evidences, divergence of the two subgenera of *Pinus* dates back to at least the late Cretaceous (Florin, 1963; Wang et al., 2000) while that of *Larix* dates to the Eocene (LePage and Basinger, 1991, 1995). Most strikingly, the *d* value (0.059) of PsubI vs PsubII is also lower than that (0.068) of *4cl*A vs *4cl*B (Table 3). The likely explanation of this is that *Pinus* has a quite different evolutionary dynamics of the 4CL gene, which is also supported by the greatest difference between d_S and d_N values in this genus. The nonexistence of paralogous 4CL loci between subgenera or species of *Pinus* further suggests that frequent duplication/deletion is an important evolutionary mechanism of the 4CL gene family in Pinaceae. In fact, the *4cl*A clade of *Larix* has also experienced several independent duplications/deletions of paralogous copies in different species or lineages since copy number of the 4CL gene and species composition are very complicated in the subclades *4cl*A_1 and *4cl*A_2 (Fig. 2). In general, the degree of observable gene family complexity correlates with plant genome size (Kinlaw and Neale, 1997). Due to lack of polyploidy, the evolution of conifers, a group with a large nuclear genome, may have occurred primarily by duplication and dispersal of genes to form complex gene families (Ahuja, 2001). For example, at least seven expressed *Adh* genes in two linked groups have been reported from *Pinus banksiana* (Perry and Furnier, 1996). Our recent study found that the cinnamyl alcohol dehydrogenase (CAD) gene is also frequently duplicated in the two families Pinaceae and Taxodiaceae (unpublished). Frequent gene duplication could be an important mechanism giving rise to the unusually large nuclear genome sizes of conifers.

The average GC content (66.0%) and polymorphic site number (58) are significantly higher at codon position 3 than at the other two positions for all 4CL gene clades, except *4cl*P_1 and *4cl*P_2 (Table 2). This bias may imply that the evolution of paralogous 4CL genes such as *4cl*A and *4cl*B are under functional constraint. At the 3rd codon position, most substitutions are synonymous, while the high GC content may correlate with levels of gene expression (Fennoy and Bailey-Serres, 1993).

4.2. Gene tree, species tree, and biogeography of Larix

Larix is widely distributed across North America and Eurasia. Patschke (1913) initially divided the genus into two sections based on the morphology of the female cone. Sect. *Larix* (or *Pauciserialis*) comprised species characterized by bracts on the cone that did not extend well beyond the seed scales, and Sect. *Multiserialis* included species with bracts extending far beyond the seed scales. Schorn (1994) disagreed with the view that the division of *Larix* into two sections was a natural system of classification, instead considering bract length to be a continuous character, in ontogenetic and historical view. He divided *Larix* into two groups with subdivisions based on the bract scale morphology: group I (Aristatus) has a relatively long exserted, tridentate bract in which the medial fork extends well beyond the lateral forks as a long aristate projection, group II a (Laminatus) is characterized by bract scales that extend only slightly beyond the cone scales, and group II b (Laminatus) possesses bracts that are shorter than the cone scales. LePage and Basinger (1991, 1995) also divided larches into two morphologically distinct groups: those species with bracts that are nonexserted and those with bracts exserted. The species within groups were further divided, in part, on the basis of their geographical distribution, external morphological features of cones, and

the fossil record. The above systematic arrangements and evolutionary hypotheses of *Larix*, using bracts of the female cone as the fundamental character, are in conflict with the results of molecular analyses (Gernandt and Liston, 1999; Qian et al., 1995).

Using RFLP analysis of cpDNA, Qian et al. (1995) contributed the first molecular phylogeny of *Larix*, demonstrating that classification based on exserted or inserted bract scales was inconsistent with the evolutionary history of the genus. However, this analysis used hybridizing fragment occurrence, rather than the presence or absence of restriction sites, as a character. It might introduce bias into the phylogenetic reconstruction since the same length mutation may be scored more than once and the fragments with similar migration rate are not necessarily homologous. Therefore, the cpDNA phylogeny constructed by Qian et al. (1995) may not be very reliable because of the very bad resolution and problems in the data analysis. Based on the sequence analysis of the nrDNA ITS region, Gernandt and Liston (1999) found that the genus *Larix* is divided into a North American and a Eurasian clade. This phylogenetic split according to continent was consistent with the result of allozyme analysis (Semerikov and Lascoux, 1999), and was further supported by the present nrDNA ITS phylogeny (Fig. 3A) with complete sampling and a very recent AFLP analysis conducted by Semerikov et al. (2003). Although the relationships within the two major clades of Eurasian species were not resolved in the cpDNA *trn*T–*trn*F phylogeny, the sister group relationship between North American and Eurasian larches was still strongly supported (Fig. 3B). That is to say, all molecular analyses except 4CL came to the same conclusion that the genus *Larix* has evolved independently on different continents following the breakup of Laurasia. In fact, the paleontological data also support the early divergence of *Larix* into a North American and a Eurasian clade. The oldest fossil records of *Larix* can date to a putatively long-bracted *lyallii*-like form from the Thunder Mountain and Coal Creek, Idaho floras (45 mya) (Axelrod, 1990; Schorn, 1994), and a short-bracted form from the Canadian High Arctic in the Middle to Late Eocene sediments (LePage and Basinger, 1991). Fossil larch cones have also been reported from Oligocene sites in Russia (LePage and Basinger, 1995).

In the present 4CL gene phylogeny of *Larix* (Fig. 2), two paralogous gene clades, *4cl*A and *4cl*B, were robustly supported by bootstrap analysis. Although there is no resolution of interspecific relationships in the *4cl*B clade, the topology of *4cl*A is very interesting. This clade was further divided into two strongly supported subclades $4cl\mathrm{A}_1$ and $4cl\mathrm{A}_2$. The $4cl\mathrm{A}_1$ subclade was composed of the Eurasian species of Sect. *Multiserialis* while the $4cl\mathrm{A}_2$ subclade comprised the other species of *Larix* besides one clone of *L. speciosa*. Particularly, the two North American species were nested within the Eurasian species of Sect. *Larix*, and showed a close relationship with *L. gmelinii* var. *gmelinii* and *L. gmelinii* var. *principis-rupprechtii*. The 4CL gene phylogeny seems to reveal a recent migration origin of the North American larches from East Asia, which is quite different from the congruent result of the cpDNA, nrDNA and allozyme analyses that *Larix* has evolved independently on different continents following the breakup of Laurasia. However, we must be cautious to distinguish between a species tree and a gene tree when reconstructing species phylogeny from a gene family (Doyle, 1997; Sanderson and Doyle, 1992). The processes of gene duplication/deletion, lineage sorting, hybridization and horizontal gene transfer can result in incongruence between the phylogenies of genes and those of species (Doyle, 1997; Gottlieb and Ford, 1996; Page and Charleston, 1997; Pamilo and Nei, 1988; Sang et al., 1997). Though we have distinguished orthologous from paralogous genes, most probably the present 4CL gene phylogeny does not mirror the species phylogeny of *Larix* since it conflicts with the evolutionary history revealed by all other molecular analyses and the fossil evidence.

The unexpected close relationship between North American and East Asian larch species in the 4CL gene phylogeny could result from lineage sorting given the presence of frequent duplication/deletion in the evolution of the 4CL gene family. Since we did not detect the $4cl\mathrm{A}_1$-type copy from Eurasian Sect. *Multiserialis* except *L. speciosa* through cloning, PCR amplification with locus-specific primer and restriction enzyme analysis, it was most probably lost in these species after the gene duplication giving rise to $4cl\mathrm{A}_1$ and $4cl\mathrm{A}_2$. In subclade $4cl\mathrm{A}_1$, it is also possible that *L. gmelinii* orthologs have been lost in *L. kaempferi*, *L. decidua*, *L. olgensis*, *L. sibirica*, and *L. speciosa* although they are still present in the two North American species, *L. laricina* and *L. occidentalis*. Similar lineage sorting phenomena have been reported in many studies such as the *Adh* gene of *Gossypium* (Small and Wendel, 2000).

L. speciosa was recognized as a species by some Chinese researchers (Fu et al., 1999), but this species is very closely related to *L. griffithii* and has a sympatric distribution with *L. griffithii* in northwest Yunnan and southeast Tibet. Farjon (1990, 2001) even reduced *L. speciosa* as a variety of *L. griffithii*. The existence of *L. speciosa* in both subclades $4cl\mathrm{A}_1$ and $4cl\mathrm{A}_2$ suggests that this species inherits more DNA polymorphism from a common ancestor, and thus, may represent a primitive form of *Larix* or the closest relative of the common ancestor of Eurasian Sect. *Multiserialis*. This hypothesis is also supported by other gene trees. In the present nrDNA ITS analysis, *L. speciosa* and *L. griffithii* are basal to the other species of Eurasian Sect. *Multiserialis* (Fig. 3A). According to a study by Semerikov et al. (2003), both the most parsimonious tree based on the cpDNA RFLP data and the neighbor-joining tree based

on the AFLP data show the basal position of *L. griffithii* in the Eurasian Sect. *Multiserialis* clade. However, *L. speciosa* was not sampled.

Acknowledgments

We sincerely thank the two anonymous reviewers and the Associate Editor Prof. Elizabeth Zimmer for valuable comments on the manuscript. We also thank Drs. David Tank (Washington State University), Sang Tao (Michigan State University), and Li Chang-Bao for providing the material of *Larix occidentalis*; Drs. Xu A-Sheng, Peng Pei-Hao, Zhou Shi-Liang, Tan Dun-Yan, Liu Zhan-Lin, Yang Fu-Sheng, and Ding Kai-Yu for their help in the field collection; Miss Sun Ying-Xue for her assistance in the laboratory. This study was supported by grants from State Key Basic Research and Development Plan (Grant No. G2000046804) and Chinese Academy of Sciences (the special fund to Wang Xiao-Quan and Grant kscxz-sw-101A).

References

Ahuja, M.R., 2001. Recent advances in molecular genetics of forest trees. Euphytica 121, 173–195.

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., de Pouplana, L.R., Martinez-Castilla, L., Yanofsky, M.F., 2000. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. USA 97, 5328–5333.

Axelrod, D.I., 1990. Environment of the Middle Eocene (45 Ma) Thunder Mountain flora, central Idaho. Natl. Geogr. Res. 6, 355–361.

Becker, A., Winter, K.-U., Meyer, B., Saedler, H., Theißen, G., 2000. MADS-box gene diversity in seed plants 300 million years ago. Mol. Biol. Evol. 17, 1425–1434.

Bradley, R.D., Hillis, D.M., 1997. Recombinant DNA sequences generated by PCR amplification. Mol. Biol. Evol. 14, 592–593.

Carlson, C.E., Theroux, L.E., 1993. Cone and seed morphology of western larch (*Larix occidentalis*), alpine larch (*Larix lyallii*), and their hybrids. Can. J. For. Res. 23, 1264–1269.

Chaw, S.-M., Zharkikh, A., Sung, H.-M., Lau, T.-C., Li, W.-H., 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear18S rRNA sequences. Mol. Biol. Evol. 14, 56–68.

Clegg, M.T., Cummings, M.P., Durbin, M.L., 1997. The evolution of plant nuclear genes. Proc. Natl. Acad. Sci. USA 94, 7791–7798.

Cronn, R.C., Small, R.L., Wendel, J.F., 1999. Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. USA 96, 14406–14411.

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.

Doyle, J.J., 1997. Trees within trees: genes and species, molecules and morphology. Syst. Biol. 46, 537–553.

Durbin, M.L., Learn, G.H., Huttley, G.A., Clegg, M.T., 1995. Evolution of the chalcone synthase gene family in the genus *Ipomoea*. Proc. Natl. Acad. Sci. USA 92, 3338–3342.

Emshwiller, E., Doyle, J.J., 1999. Chloroplast-expressed glutamine synthetase (*ncpGS*): potential utility for phylogenetic studies with an example from *Oxalis* (Oxalidaceae). Mol. Phylogenet. Evol. 12, 310–319.

Farjon, A., 1990. Pinaceae. Koeltz Scientific Books, Konigstein, Germany.

Farjon, A., 2001. World Checklist and Bibliography of Conifers, second ed. Royal Botanic Gardens, Kew, England.

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

Fennoy, S.L., Bailey-Serres, J., 1993. Synonymous codon usage in *Zea mays* L. nuclear genes is varied by levels of C and G-ending codons. Nucleic Acids Res. 21, 5294–5300.

Florin, R., 1963. The distribution of conifer and taxad genera in time and space. Acta Hort. Berg. 20, 121–312.

Ford, V.S., Gottlieb, L.D., 1999. Molecular characterization of *PgiC* in a tetraploid plant and its diploid relatives. Evolution 53, 1060–1067.

Fu, L.-K., Li, N., Mill, R.R., 1999. Larix. In: Wu, Z.-Y., Raven, P.H. (Eds.), Flora of China (4). Science Press, Beijing and Missouri Botanical Garden Press, St. Louis, pp. 33–37.

Gaut, B.S., Morton, B.R., McCaig, B.C., Clegg, M.T., 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbc*L. Proc. Natl. Acad. Sci. USA 93, 10274–10279.

Gaut, B.S., Peek, A.S., Morton, B.R., Clegg, M.T., 1999. Patterns of genetic diversification within the *Adh* gene family in the grasses (Poaceae). Mol. Biol. Evol. 16, 1086–1097.

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

Gottlieb, L.D., Ford, V.S., 1996. Phylogenetic relationships among the sections of *Clarkia* (Onagraceae) inferred from the nucleotide sequences of *PgiC*. Syst. Bot. 21, 45–62.

Hart, J.A., 1987. A cladistic analysis of conifers: preliminary results. J. Arn. Arb. 68, 269–307.

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.

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Kinlaw, C.S., Neale, D.B., 1997. Complex gene families in pine genomes. Trends Plant Sci. 2, 356–359.

Kliebenstein, D.J., Lambrix, V.M., Reichelt, M., Gershenzon, J., Mitchell-Olds, T., 2001. Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. Plant Cell 13, 681–693.

Kuhner, M.K., Felsenstein, J., 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol. Biol. Evol. 11, 459–468.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: Molecular Evolutionary Genetics Analysis Software. Arizona State University, Tempe, AZ, USA.

Kusumi, J., Tsumura, Y., Yoshimaru, H., Tachida, H., 2002. Molecular evolution of nuclear genes in Cupressaceae, a group of conifer. Mol. Biol. Evol. 19, 736–747.

Kvarnheden, A., Tandre, K., Engström, P., 1995. A *cdc*2 homologue and closely related processed retropseudogenes from Norway spruce. Plant Mol. Biol. 27, 391–403.

LePage, B.A., Basinger, J.F., 1991. A new species of *Larix* (Pinaceae) from the early Tertiary of Axel Heiberg Island, Arctic Canada. Rev. Palaeobot. Palyno. 70, 89–111.

LePage, B.A., Basinger, J.F., 1995. The evolutionary history of the genus *Larix* (Pinaceae). *USDA*. For. Serv. Int. Res. Sta. GTR INT 319, 19–29.

Lynch, M., 2002. Gene duplication and evolution. Science 297, 945–947.

Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155.

Lyons-Weiler, J., Milinkovitch, M.C., 1997. A phylogenetic approach to the problem of differential lineage sorting. Mol. Biol. Evol. 14, 968–975.

Mason-Gamer, R.J., Weil, C.F., Kellogg, E.A., 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. Mol. Biol. Evol. 15, 1658–1673.

Mathews, S., Donoghue, M.J., 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. Science 286, 947–950.

Mathews, S., Sharrock, R.A., 1996. The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of loci found in dicot angiosperms. Mol. Biol. Evol. 13, 1141–1150.

Mathews, S., Lavin, M., Sharrock, R.A., 1995. Evolution of the phytochrome gene family and its utility for phylogentic analyses of angiosperms. Ann. Missouri Bot. Gard. 82, 296–321.

Mathews, S., Tsai, R.C., Kellogg, E.A., 2000. Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. Am. J. Bot. 87, 96–107.

Morton, B.R., Gaut, B.S., Clegg, M.T., 1996. Evolution of alcohol dehydrogenase genes in the palm and grass families. Proc. Natl. Acad. Sci. USA 93, 11735–11739.

Murray, B.G., 1998. Nuclear DNA amounts in gymnosperms. Ann. Bot. 82 (Suppl. A), 3–15.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Ostenfeld, C.H., Larsen, C.S., 1930. The species of the genus *Larix* and their geographic distribution. Biol. Meddel. Kongel. Danske Vidensk. Selsk. 9, 1–107.

Owens, J.N., Takaso, T., Runions, C.J., 1998. Pollination in conifers. Trends Plant Sci. 3, 479–485.

Page, R.D.M., Charleston, M.A., 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. Mol. Phylogenet. Evol. 7, 231–240.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.

Patschke, W., 1913. Über die extratropischen ostasiatischen Coniferen und ihre Bedeutung für die pflanzengeographische Gliederung Ostasiens. Bot. Jahrb. Syst. 48, 626–776.

Perry, D.J., Furnier, G.R., 1996. *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. Proc. Natl. Acad. Sci. USA 93, 13020–13023.

Price, R.A., Olsen-Stojkovich, J., Lowenstein, J.M., 1987. Relationships among the genera of Pinaceae: an immunological comparison. Syst. Bot. 12, 91–97.

Purugganan, M.D., Wessler, S.R., 1994. Molecular evolution of the plant *R* regulatory gene family. Genetics 138, 849–854.

Qian, T., Ennos, R.A., Helgason, T., 1995. Genetic relationships among larch species based on analysis of restriction fragment variation for chloroplast DNA. Can. J. For. Res. 25, 1197–1202.

Rogers, S.O, Bendich, A.J., 1988. Extraction of DNA from plant tissues. Pl. Mol. Biol. Manual A 6, 1–10.

Sanderson, M.J., Doyle, J.J., 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Syst. Biol. 41, 4–17.

Sang, T., Donoghue, M.J., Zhang, D., 1997. Evolution of alcohol dehydrogenase genes in peonies (*Paeonia*): phylogenetic relationships of putative nonhybrid species. Mol. Biol. Evol. 14, 994–1007.

Schmidt, M., Schneider-Poetsch, H.A.W., 2002. The evolution of gymnosperms redrawn by phytochrome genes: the Gnetatae appear at the base of the gymnosperms. J. Mol. Evol. 54, 715–724.

Schorn, H.E., 1994. A preliminary discussion of fossil larches (*Larix*, Pinaceae) from the Arctic. Quat. Int. 22/23, 173–183.

Schubert, R., Sperisen, C., Müller-Starck, G., Scala, S.L., Ernst, D., Sandermann, H., Häger, K.-P., 1998. The cinnamyl alcohol dehydrogenase gene structure in *Picea abies* (L.) Karst.: genomic sequences, southern hybridization, genetic analysis and phylogenetic relationships. Trees 12, 453–463.

Semerikov, V.L., Lascoux, M., 1999. Genetic relationship among Eurasian and American *Larix* species based on allozymes. Heredity 83, 62–70.

Semerikov, V.L., Zhang, H., Sun, M., Lascoux, M., 2003. Conflicting phylogenies of *Larix* (Pinaceae) based on cytoplasmic and nuclear DNA. Mol. Phylogenet. Evol. 27, 173–184.

Small, R.L., Wendel, J.F., 2000. Phylogeny, duplication, and intraspecific variation of *Adh* sequences in New World diploid cottons (*Gossypium* L., Malvaceae). Mol. Phylogenet. Evol. 16, 73–84.

Swofford, D.L., 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), Version 4. Sinauer, Sunderland, MA.

Tank, D.C., Sang, T., 2001. Phylogenetic utility of the glycerol-3-phosphate acyltransferase gene: evolution and implications in *Paeonia* (Paeoniaceae). Mol. Phylogenet. Evol. 19, 421–429.

Theissen, G., Becker, A., Rosa, A.D., Kanno, A., Kim, J.T., Munster, T., Winter, K.-U., Saedler, H., 2000. A short history of MADS-box genes in plants. Plant Mol. Biol. 42, 115–149.

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.

Tsumura, Y., Yoshimura, K., Tomaru, N., Ohba, K., 1995. Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloroplast genes. Theor. Appl. Genet. 91, 1222–1236.

Uhlmann, A., Ebel, J., 1993. Molecular cloning and expression of 4-coumarate:coenzyme A ligase, an enzyme involved in the resistance response of soybean. Plant Physiol. 102, 1147–1156.

Walsh, J.B., 1995. How often do duplicated genes evolve new function. Genetics 139, 439–444.

Wang, X.-Q., Han, Y., Deng, Z.-R., Hong, D.-Y., 1997. Phylogeny of the Pinaceae evidenced by molecular biology. Acta Phytotax. Sin. 35, 97–106.

Wang, X.-Q., Han, Y., Hong, D.-Y., 1998a. A molecular systematic study of *Cathaya*, a relic genus of the Pinaceae in China. Plant Syst. Evol. 213, 165–172.

Wang, X.-Q., Han, Y., Hong, D.-Y., 1998b. PCR-RFLP analysis of the chloroplast gene *trn*K in the Pinaceae, with special reference to the systematic position of Cathaya. Isr. J. Plant Sci. 46, 265–271.

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.

Waters, E.R., 1995a. An evaluation of the usefulness of the small heat shock genes for phylogentic analysis in plants. Ann. Missouri Bot. Gard. 82, 278–295.

Waters, E.R., 1995b. The molecular evolution of the small heat-shock proteins in plants. Genetics 141, 785–795.

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.

Wen, J., Vanek-Krebitz, M., Hoffmann-Sommergruber, K., Scheiner, O., Breiteneder, H., 1997. The potential of *Betv*1 homologues, a nuclear multigene family, as phylogenetic markers in flowering plants. Mol. Phylogenet. Evol. 8, 317–333.

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., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, pp. 315–322.

Winter, K.-U., Becker, A., Munster, T., Kim, J.T., Saedler, H., Theissen, G., 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. Proc. Natl. Acad. Sci. USA 96, 7342–7347.

Xia, X., Xie, Z., 2001. DAMBE: data analysis in molecular biology and evolution. J. Hered. 92, 371–373.

Zhang, X.-H., Chiang, V.L., 1997. Molecular cloning of 4-coumarate:coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol. 113, 65–74.

Zhang, L., Pond, S.K., Gaut, B.S., 2001. A survey of the molecular evolutionary dynamics of twenty-five multigene families from four grass taxa. J. Mol. Evol. 52, 144–156.