Tree Genetics & Genomes (2006) 2: 132–139
DOI 10.1007/s11295-006-0036-2

ORIGINAL PAPER

Ping Lang · Fenny Dane · Thomas L. Kubisiak

Phylogeny of *Castanea* (Fagaceae) based on chloroplast *trn*T-L-F sequence data

Received: 24 June 2005 / Revised: 19 January 2006 / Accepted: 30 January 2006 / Published online: 9 March 2006

Abstract Species in the genus *Castanea* are widely distributed in the deciduous forests of the Northern Hemisphere from Asia to Europe and North America. They show floristic similarity but differences in chestnut blight resistance especially among eastern Asian and eastern North American species. Phylogenetic analyses were conducted in this study using sequences of three chloroplast noncoding *trn*T-L-F regions. The *trn*T-L region was found to be the most variable and informative region. The highest proportion of parsimony informative sites, more and larger indels, and higher pairwise distances between taxa were obtained at *trn*T-L than at the other two regions. The high A+T values (74.5%) in the *Castanea* *trn*T-L region may explain the high proportion of transversions found in this region where as comparatively lower A+T values were found in the *trn*L intron (68.35%) and *trn*L-F spacer (70.07%) with relatively balanced numbers of transitions and transversions. The genus *Castanea* is supported as a monophyletic clade, while the section *Eucastanon* is paraphyletic. *C. crenata* is the most basal clade and sister to the remainder of the genus. The three Chinese species of *Castanea* are supported as a single monophyletic clade, whose sister group contains the North American and European species. There is consistent but weak support for a sister–group relationship between the North American species and European species.

Keywords Fagaceae · *Castanea* · *trn*T-L-F · Phylogenetics

P. Lang · F. Dane (✉)
Department of Horticulture, Auburn University,
Auburn, AL 36849, USA
e-mail: danefen@auburn.edu
Tel.: +334-844-3046
Fax: +334-844-3131

T. L. Kubisiak
USDA-Forest Service, Southern Institute of Forest Genetics,
Saucier, MS 39574, USA

Castanea (Fagaceae), the genus of chestnuts and chinkapins, is widely distributed in the temperate forest of the Northern Hemisphere (Fig. 1). The genus comprises three sections and seven species (Johnson 1988). Section *Eucastanon* comprises five species: *C. mollissima* BL., *C. seguinii* Dode. from China, *C. crenata* Sieb. & Zucc. from Japan, *C. dentata* (Marsh.) Brokh from North America, and *C.sativa* Mill. from Europe, all with the characteristics of three nuts per cupule. Chinese chestnut (*C. mollissima*), Japanese chestnut (*C. crenata*), and European chestnut (*C.sativa*) are economically important as nut producers, and have been cultivated for centuries. The American chestnut (*C. dentata*), formerly one of the dominant trees of the eastern United States, has been almost wiped out by chestnut blight. American chestnut was valued primarily as a timber and tannin source and only secondarily as a nut tree. Section *Balanocastanon*, has a widely disputed taxonomy, with taxa exclusively found in the southeastern United States. It contains just one species, *C. pumila* Mill., with two varieties: var. *pumila* and var. *ozarkensis*, characterized by one nut per cupule. *C. pumila* var *pumila*, the Allegheny chinkapin, has a wide distribution from southern New Jersey and Pennsylvania, west to Indiana and Missouri, and south to Florida and Texas. *C. pumila* var. *ozarkensis*, the Ozark chinkapin, has limited and fragmented distribution in the Ozark Mountains of eastern Oklahoma, southwest Missouri, and north-central Arkansas (Johnson 1988). Chinkapins are less utilized by humans, mainly because it only grows to shrub size and has small nuts. Section *Hypocastanon* consists of just one species, *C. henryi* (Skan) Rehder & Wilson, which is found in a restricted area in southeast China and is sometimes called the Chinese chinkapin because it is characterized by a single nut per cupule (Lang and Huang 1999). This species is also commercially less important because of its small nut size. Henry chestnut grows straight and high and is good for timber.

Castanea is a tertiary genus with a disjunct distribution pattern between eastern Asia and eastern North America. The geographic distribution of *Castanea* in China and the United States has a similar pattern in that geographically

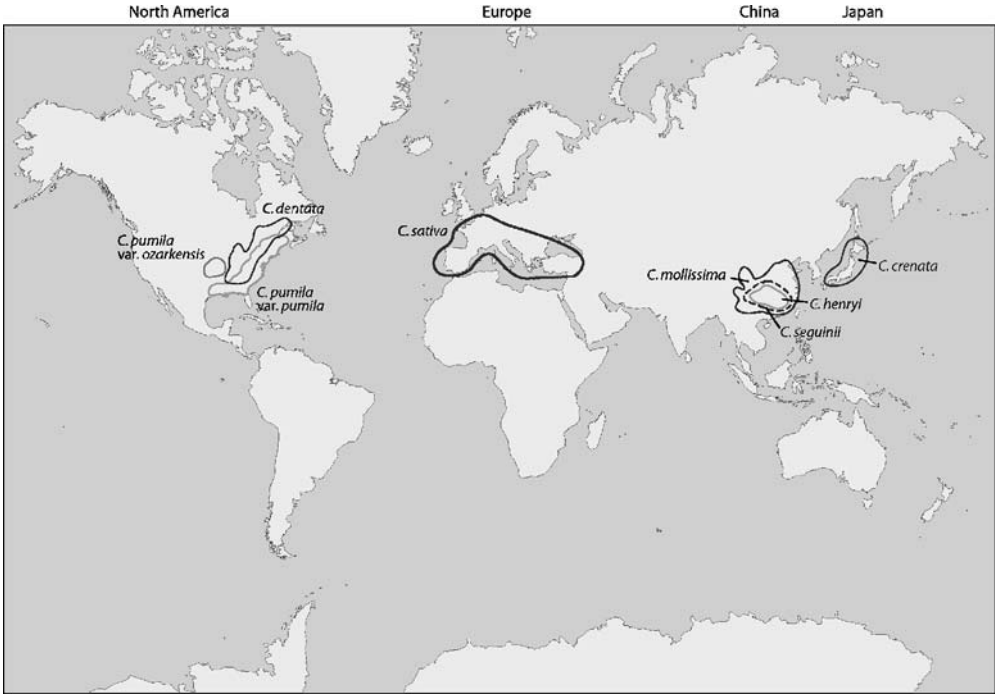

Fig. 1 Distribution map of *Castanea* species

widespread species (*C. mollissima* in China and *C. dentata* in the US) overlap the restricted species (*C. seguinii* and *C. henryi* in China and *C. pumila* in the US). However, significant differences in levels of resistance to *Cryphonectria parasitica* (Murrill) Barr, causal agent of chestnut blight, were found between eastern Asia and eastern North American species. The American species are known to be blight-susceptible, while the Asian species are blight-resistant (Jaynes 1975). The blight fungus is a wound parasite that occludes the phloem of susceptible trees, forming a girdling canker on the main trunk (Jaynes 1975). Because of the introduction of the chestnut blight at the beginning of the 20th century in America, the American chestnut has been reduced from an important timber and nut producing tree to an understory shrub (Graves 1950; Huang et al. 1996).

A number of studies using allozymes or RADP markers have been conducted to detect genetic variation within and among populations of *Castanea* species and propose appropriate conservation strategies (Villani et al. 1991; Huang et al. 1994, 1998; Dane et al. 1999, 2003; Lang and Huang 1999). Studies of chloroplast (cp) DNA variability in European chestnut (*C. sativa*) revealed low levels of differentiation among populations, which has been related to the impact of human colonization (Fineschi et al. 2000). However, systematic studies of the genus *Castanea* on a worldwide basis have been limited or inconclusive. The only comprehensive molecular study on *Castanea* is that of Stanford (1998) who analyzed chloroplast *mat*K and nuclear ITS regions and reported a complicated history of the genus. Because of low levels of informative characters in the *mat*K region and large amounts of missing data, analysis of *Castanea* did not resolve relationships within the genus. Further study to provide more detailed knowledge of the phylogenetic and biogeographic relationships within *Castanea* will be important for their conservation and reintroduction of the American chestnut into the American forest ecosystem.

Chloroplast DNA is well suited for evolutionary and phylogenetic study (Clegg and Zurawaki 1992; Olmstead and Palmer 1994). The *trn*T-L-F region, initially characterized by Taberlet et al. (1991), contains two intergenic spacers *trn*T(UGU)-*trn*L(UAA) and *trn*L (UAA)-*trn*F (GAA), and the *trn*L(UAA) intron, which can easily be amplified using primers homologous to the exons of the *trn*T, *trn*L, and *trn*F gene (Taberlet et al. 1991). The *trn*L-F intergenic spacer, as well as the *trn*L intron, has been used widely for phylogenetic reconstruction from the species level to the family level, however, the *trn*T-L spacer has rarely been used in systematic studies (Taberlet et al. 1991; Small et al. 1998; Shaw et al. 2005). In this study, the complete *trn*T-L-F region will be used to trace phylogenetic relationships for all species within the genus *Castanea*, also, the various phylogenetic utility of three noncoding regions will be addressed.

Materials and methods

Plant material, DNA extraction, PCR, and nucleotide sequencing: Sampling information of 17 taxa representing all seven species of *Castanea* and the outgroup *Quercus falcata* is shown in Table 1. Genomic DNA was extracted from fresh plant material using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). Initial experiments were conducted using representative samples from all species (16 samples from eight populations of *C. dentata* and eight samples from every other *Castanea* species) to detect intraspecific cpDNA polymorphism. Universal primers were used to amplify the noncoding regions of interest.

Table 1 List of investigated taxa and their GenBank accession numbers

Taxon	Sample ID	DNA source/voucher	Chloroplast region (GenBank no.)		
			*trn*T-L	*trn*L	*trn*L-F
C. dentata	D2	CA-44: Ontario, Canada	AY586319	AY586285	AY586302
C. dentata	D12	BR-34: Asheville, North Carolina	AY586320	AY586286	AY586303
C. pumila var. *pumila*	P4	FLE-10: Eglin Air Force Base, Florida	AY586321	AY586287	AY586304
C. pumila var. *pumila*	P6	[a]HH R5T2: Lafayette County, Florida	AY586322	AY586288	AY586305
C. pumila var. *pumila*	Vc1	Snakeden Mountain, Virginia	AY854377	AY854381	AY854385
C. pumila var. *pumila*	Vc9	Snakeden Mountain, Virginia	AY854378	AY854382	AY854386
C. pumila var *ozarkensis*	3–6	Population 5, Sylamore Ranger Distinct, Arkansas[b]	AY854379	AY854383	AY854387
C. pumila var *ozarkensis*	T6D	Population 6, Sylamore Ranger Distinct, Arkansas[b]	AY854380	AY854384	AY854388
C. pumila var *ozarkensis*	P7	CC R-gammaT3: Russelville, Arkansas[a]	AY586323	AY586289	AY586306
C. pumila var *ozarkensis*	P9	CC R-etaT4: Russelville, Arkansas[a]	AY586324	AY586290	AY586307
C. sativa	Sa1	HH R3T2: from the Cavcas Biosphere Reserve[a]	AY586325	AY586291	AY586308
C. sativa	Sa7	24°E 10′ and 46°N 80′, Baia Mare, northern Romania	AY586326	AY586292	AY586309
C. mollissima	M1	CH-MAH: Chinese Mahogany, from North-East China[a]	AY586327	AY586293	AY586310
C. mollissima	M4	WL R37T7: USDA #104061, from Central China[a]	AY586328	AY586294	AY586311
C. seguinii	S3	SL R3T8: USDA #70317, from East-Central China[a]	AY586329	AY586295	AY586312
C. seguinii	S6	HH R4T2: cross of SL R8T4 (female) × SL R2T16 (male)[a]	AY586330	AY586296	AY586313
C. henryi	H1	Hubei Province, P.R. China	AY586331	AY586297	AY586314
C. henryi	H3	Rock R3T9: from Nanjing Botanical Garden, PR China[a]	AY586332	AY586298	AY586315
C. crenata	C1	WL R34T6: USDA#104016, from Japan[a]	AY586333	AY586299	AY586316
C. crenata	C7	92-JH-2a-2: Bare of Mt. Tougo, Kuki-cho, Japan	AY586334	AY586300	AY586317
Quercus falcata		Saucier, Mississippi	AY586335	AY586301	AY586318

[a]The Connecticut Agricultural Experiment Station, New Haven, Connecticut
[b]Population 5 and 6 (Dane et al. 1999)

The *trn*T-L intergenic spacer was amplified with primers "a" and "b"(Taberlet et al. 1991). The *trn*L-F region was amplified using primers "c" and "f" to obtain a product including the *trn*L intron, the 3′ *trn*L exon, and the intergenic spacer between the 3′ *trnL* exon and the *trn*F gene (Taberlet et al. 1991). Double-stranded DNA amplification was performed in a 50-µl volume containing the respective 1× PCR buffer of 20 mmol/l Tris–HCl (pH 8.4) and 50 mmol/l KCl [5 µl of 10× PCR buffer purchased from Invitrogen Life Technologies (Carlsbad, CA, USA)], 2 mmol/l $MgCl_2$, 200 µmol/l of each dNTP, 0.2 µmol/l of each primer, 2 U of *Taq* polymerase (Invitrogen), and 2.5 µl template DNA (25 ng/µl). After an activation step of three cycles of 1 min at 94°C and 30 s at 72°C, the PCR mixture underwent 34 cycles of 1 min at 94°C, 30 s at 56°C, 2 min at 72°C, with a final extension of 7 min at 72°C in Thermo Hybaid thermal cycler (Ashford, Mddlesex, TW, UK). To remove excess primers and dNTPs, PCR products were purified using Concert rapid PCR purification Kit (GibcoBRL, Germany) or Qiaquick PCR purification kit (Qiagen, Valencia, CA, USA), and sequenced by Auburn Genomics and Sequencing Lab with the ABI3100 DNA Analysis Machine (Applied Biosystems Inc. Foster City, CA). To verify complementary stands, primers "a", "b", "c", "d", "e", and "f" (Taberlet et al. 1991) were used to generate the entire sequences.

Sequence alignment and data analyses: Multiple alignment of the sequences was obtained using the AlignX program implemented in the Vector NTI software, and adjusted manually. Gaps were introduced in the alignment to optimize positional homology. Single-base indels were cross-checked to the original chromatograms, to verify that they were not sequencing artifacts missed during base calling. Indels that were potentially parsimonious (e.g., shared by two or more taxa) were scored and added to the end of the data sets as present (1) or absent (0) type characters. Gaps with overlaps were considered nested and treated as a single multistate character according to Simmons and Ochotreana (2000). Areas of ambiguous alignment or with poly-n strings were excluded from all analyses. A maximum parsimony analysis was conducted using PAUP version 4.0d10 software (Swofford 2000). To explore the possibility of multiple islands of most parsimonious trees, searches were conducted over 1,000 random-taxon-addition replicates with tree bisection–reconnection (TBR) branch swapping, with the save multiple trees (MulTrees) option selected and with all characters and states weighted equally and unordered. A 50% majority-rule consensus tree was computed. The robustness of nodes was inferred by bootstrap analysis (Felsenstein 1985) of 1,000 replicates.

Pairwise distance values between sequences were calculated using the Kimura two-parameter model as implemented by PAUP 4.0b10 (Swofford 2000) for all data sets. A series of probable stem/loop-forming regions was identified using mfold 3.0 (Zuker et al. 1999).

Results

Characteristics of trnT-L-F region in Castanea: The chloroplast spacers *trn*T-L and *trn*L-F and the *trn*L intron of *Castanea* are AT rich, with an average A+T content of approximately 70%. Sequences from 17 taxa included in this study provided a total aligned length of 1,955 bp, which included the 3′ *trn*L coding region. The individual *trn*T-L intergenic spacer sequences within *Castanea* ranged in length from 841 bp for the Chinese species to 753 bp for *C. dentata*, with the outgroup sequence being 806 bp. The *trn*L intron varies from 474 bp for one *C. mollissima* haplotype to 501 for *C. pumila* var. *ozarkensis*, with the outgroup sequence being 484 bp. The *trn*L-F region varies from 382 bp for *C. crenata* to 411 bp for *C. henryi*, with the outgroup sequence being 381 bp. As noted by Taberlet et al. (1991) and confirmed by others for different plant groups (Small et al. 1998; McDade and Moody 1999), the *trn*T-L-F noncoding regions experience a high frequency of indel mutations of variable lengths, making sequence alignment problematic in some cases. Eight gaps, ranging from one to 73 bases, were introduced into the sequence alignment (three in the *trn*T-L spacer, two in the *trn*L intron, and three in the *trn*L-F spacer). Most of the indels are autapomorphic, only two of them are parsimony informative because of a 12 bp deletion in *trn*T-L spacer is shared with *C. dentata* and *Q.falcata* and an 18-bp insertion in *trn*L intron can be used to detect the differentiation of *C. pumila*. Three of eight indels were associated with runs of short direct repeats of single base motifs (one A track in *trn*T-L spacer and two T tracks in the *trn*L-F intergenic spacer). Because these variations may be caused by slipped-strand mispairing (SSM) (Levinson and Gutman 1987) or experimental error (Levinson and Gutman 1987; McDade and Moody 1999), these sequences were eliminated from further study. Flanking sequences were examined in the same or closely related taxa to assess if the indels have tandemly repeated sequences, although this cannot always be determined unambiguously, particularly for overlapping gaps (Graham et al. 2000). Four indels are insertions or deletions of exact or nearly exact duplications of immediately adjacent or nearby sequences. These indels involve moderately sized regions from 7 to 25 bp. A comparison of sequence divergence and phylogenetic information of the three chloroplast regions among the *Castanea* is presented in Table 2. The large mutations do provide informative markers for tracing the phylogeny among species of the genus *Castanea*. Because of large indels, *C. dentata* (75 bp deletion in *trn*T-L intergenic region) and *C. henryi* (25 bp insertion in *trn*L-F spacer) can be easily distinguished from the other *Castanea* species using PCR amplification followed by agarose gel separation. These species-specific markers can subsequently be used to detect the mother source of the three species because of the maternal inheritance of the cpDNA. Analysis of additional accessions (16 samples from eight different populations of American chestnut and eight samples each from the other *Castanea* species), which were investigated in the preliminary survey using PCR amplification and agarose gel separation, support this conjecture.

Table 2 Comparison of sequence divergence and phylogenetic information among three chloroplast regions in *Castanea*

Sequence characteristics	*trn*T-L	*trn*L	*trn*L-F
Mean A+T content (%)	74.5	68.4	70.1
Aligned length	884	588	478
Length of sequenced noncoding region (range)	753–841	472–499	373–402
Number of characters included (bp)	798	558	440
Number of variable characters (proportion)	15 (1.88%)	7 (1.25%)	10 (2.27%)
Parsimony informative sites (proportion)	8 (1.00 %)	3 (0.54 %)	3 (0.68 %)
Substitutions among *Castanea* species (Ts:Tv)	13 (4Ts:9Tv)	5 (2Ts:3Tv)	9 (5Ts:4Tv)
Number of indels coded as binary characters (range)	2 (2–73)	2 (7–18)	1 (1–25)
Pairwise distance among *Castanea* species	0.137–0.962%	0–0.644%	0–0.948%
Pairwise distance between *Castanea* and *Quercus*	0.825–1.379%	0.214–0.645%	0.268–1.338%
Tree statistics			
Number of trees	1	6	3
Tree length	20	10	11
Consistency index (CI)	0.950	0.9000	1.0000
Retention index (RI)	0.9737	0.9412	1.0000
Rescaled consistency index (RC)	0.9250	0.8471	1.000
Number of clades in bootstrap consensus with >50% support	9	3	5

The aligned positions 1,246–1,249 were associated with a small inversion of TTTC in most of the *Castanea* species to GAAA in the three Chinese species (*C. mollissima*, *C. seguinii*, and *C. henryi*). This 4-bp region is flanked by two conserved inverted repeats (each 9 bp long), probably forming the stem region of a stem-loop secondary structure, with the inversion corresponding to the loop. The free energy of spontaneous formation equals to -3.21 Kcal/mol^{-1} for the three Chinese species and -1.81 Kcal/mol^{-1} for other species.

Transversions were more prevalent than transitions in the *Castanea trn*T-L intergenic spacer as compared to *trn*L intron and *trn*L-F intergenic spacer. The patterns appear to follow the observations of Morton (1995), in that positions flanked by A or T are more likely to undergo transversions. Some substitutions involved in the recognition sites

of some restriction enzymes can be easily determined using the PCR-RFLP method. For example, at aligned position 457, all Chinese species can be distinguished from the other *Castanea* species by the gain of one *Mse*I restriction site because of the substitution of G→T. Similarly, at position 773, all North American *Castanea* species have lost one *Taq*I restriction site because of a G→A substitution.

Relationships within the genus Castanea: The maximum parsimony (MP) analyses were based upon sequences from three noncoding regions including indels and the combined sequence data of the three regions either excluding or including indels. The analysis of the *trn*T-L intergenic spacer region gave the best resolution (Table 2). When all the sequences were combined and the indels were excluded from the analysis, the total number of variable sites was 34, 16 of which were parsimony informative. Analysis of the 11 samples with equal character weighing generated one most parsimonious tree of 35 steps, with a consistency index (CI) excluding autapomorphies of 0.9412 and a retention index (RI) of 0.9655. When the indels were included, only two of nine indels identified from aligned sequences were informative characters. One most parsimonious tree was 46 steps long, with a consistency index (CI) for informative characters of 0.9130 and a retention index (RI) of 0.9459. The tree based on the combined data sets with indels is consistent with the tree excluding indels as well as the *trn*T-L intergenic spacer. *Castanea* is supported as a monophyletic clade, with the section *Eucastanon* as a paraphyletic group. *C. crenata* is the most basal clade in the genus and sister to the remainder of the genus. The three Chinese species of *Castanea* are supported as a single monophyletic clade, whose sister group contains the North American species clade and the European species (*C. sativa*). There is consistent but weak support for the sister–group relationship between the North American species and European species (Fig. 2) because a phylogenetic informative transversion was shared by the American and the European species (Table 4). However, the European chestnut is characterized by three unique transversions (Table 4).

Discussion

Species in the genus *Castanea* show relatively low levels of morphological variation. Their classification is mainly based on leaf morphology and variation in reproductive character states and cupule to fruit arrangement (Nixon and Crepet 1989). A general observation within Fagaceae by Manos and Stanford (2001) is that sequence divergence among and within genera is very low. Nucleotide variation for noncoding cpDNA intergenic spacer regions was generally low in interspecific comparisons of *Fagus*, *Castanea*, and *Quercus* section *Quercus* as measured by the two-parameter method of Kimura 1980 implemented in PAUP. Pairwise sequence divergence values ranged from roughly 0.1–1.3% (Manos and Stanford 2001). Similarly, this study reveals low levels of pairwise genetic distance varying from 0.137–0.962% for the *trn*T-L spacer, 0–0.644% for *trn*L intron, and 0–0.948% for the *trn*L-F spacer. A possible explanation for the low molecular divergence could be that (1) slow evolution in *Castanea* as compared to other species and (2) divergence in the *Castanea* occurred relatively recently. Although low sequence variations were observed among Chinese as well as North American species based on the sequences of *trn*T-L-F, the relationships within these two clades can be resolved by some autapomorphic indels or substitutions. Substitution G→T is unique to *C. mollissima*, while *C. henryi* can be separated from the other Chinese species by a unique 25-bp insertion in the *trn*L-F intergenic spacer. *C. dentata* is characterized by two main autapomorphic indels (a 12-bp and 73-bp deletion in the *trn*T-L spacer) and two substitutions. Also, two varieties of *C. pumila* can be distinguished from each other by two substitutions.

The *trn*L intron and the *trn*L-F spacer regions in the genus *Castanea* showed a relatively balanced ratio of transitions and transversions (2:2 vs 5:4, respectively), while in the *trn*T-L spacer transversions are more prevalent than transitions (Ts:Tv=3:8). The high A+T values (74.5%) in the *Castanea trn*T-L region may explain the high proportion of transversions found in this region whereas comparatively lower AT values were found in the *trn*L intron (68.35%) and *trn*L-F spacer (70.07%). Also, a lower number of substitutions per site were detected at the intron

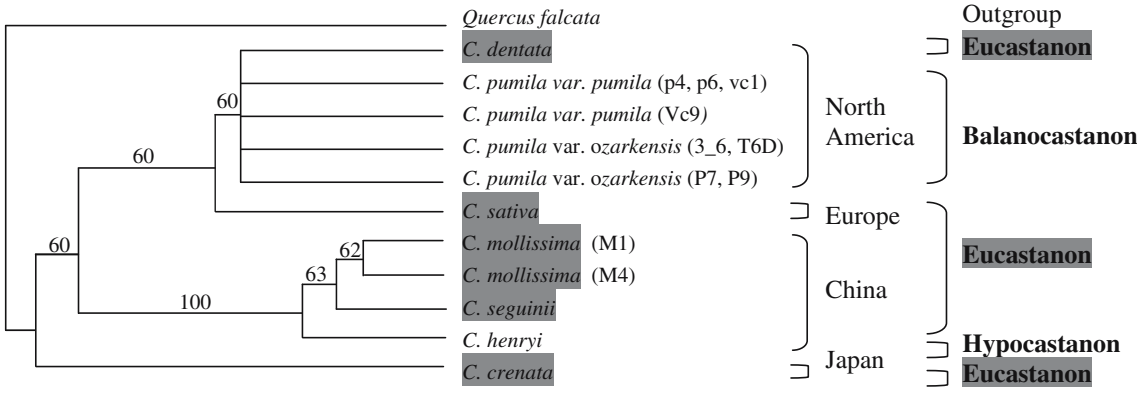

Fig. 2 Fifty percent majority-rule consensus of 1,703 trees inferred from comparative analysis of *Castanea trn*T-L-F sequences (27 steps, CI=0.9630, RI=0.9630). Numbers above branches indicate bootstrap value support for clade

region. A+T rich regions have been found to be highly prone to replication slippage (Levinson and Gutman 1987; Cummings et al. 1994), a mechanism that involves local intrahelical denaturation and displacement of replicating strands, which can lead to further increases in A+T content, possibly enhancing transversion bias (Bakker et al. 2000). Based on the examination of 18 noncoding DNA regions in the chloroplast genome, Morton (1995) described a marked bias towards transversions when point mutations occurred at a base flanked by A or T and especially when flanked on both sides by A or T. This transversion bias coupled with the known AT bias of chloroplast genomes, evidenced here by the >70% AT composition of the *Castanea trn*T-L region, would effectively narrow the pool of potential base mutations at those positions, increasing the probability of parallel mutations or reversals in different lineages (Kelchner and Clark 1997).

It is curious that the *trn*T-L spacer has rarely been used in systematic studies, despite the popularity of the *trn*L intron and *trn*L-F spacer, which have been widely utilized to resolve phylogenetic relationships at both the generic and species level (Small et al. 1998). Consistent with other observations (Small et al. 1998; Fukuda et al. 2001), this study shows that the *trn*T-L region is the most variable of the three. The *trn*T-L has a larger size (both raw and aligned) relative to the *trn*L intron and the *trn*L-F spacer. Also, the *trn*T-L has the highest proportion of parsimony informative sites, and contains more and longer indels and thus is more variable in length. As a result, pairwise distances between taxa are greater than those observed at the other two regions (Table 2). This pattern suggests that the *trn*T-L spacer is evolving faster than the other two regions, and may be under less selection pressure than the other two regions. The *trn*L-F spacer has fewer informative sites and contains fewer length mutations (Table 2). Although frequently used, the *trn*L-F spacer appears to evolve slower, and thus its phylogenetic utility at the intrageneric level is limited. The *trn*T-L spacer appears to be evolving more rapidly than the other two regions, and hence should be useful for phylogenetic studies at lower taxonomic levels (Fukuda et al. 2001).

Most structural mutations in cpDNA are small indels, from 1 to 10 bp (reviewed by Vijverberg and Bachmann 1999), however, a general feature of *trn*T-L-F region is the occurrence of large indels. *Juncus trifidus* is unique among the other *Juncus* species by a 322-bp deletion in the *trn*L intron and an 84-bp insertion in the *trn*L-F spacer. Also, a 334-bp deletion in the *Juncus trn*L intron defines the Southern Hemisphere clade. Similarly, large deletions (350 and 250 bp) in the the *trn*L-F intergenic spacer define clades within Cactaceae subfamily Cactoideae (Applequist and Wallace 2002). Also, phylogenetically significant 70-bp deletions were shared by two species of *Sedum*. In *Castanea*, although indels in *trn*T-L-F are less frequent as compared to the number reported for other species (McDade and Moody 1999), large autapomorphic structural mutations were detected. *C. dentata* is unique among the other species by a 75-bp deletion and an insertion (25 bp) is typical to *C. henryi*. Most of the indels are duplications or deletions of adjacent sequences, but the large deletion detected in *C. dentata* did not show sequence similarity with the flanking region of the indel site. This deletion may be associated with the formation of hairpins or stem-loop structure in DNA secondary structure (Kelchner 2000), and seems to be less prone to homoplasy.

Similar to Kelchner and Wendel's (1996) observation at the *rpl*16 intron in bamboo, the inversions are quite short in length (4 bp) and create no obvious gaps during alignment of sequences. The presence of an inversion could lead to incorrect assessment of homology of the inverted bases and treatment of the single mutation as separate point substitution events. Given the fact that the inversions comprise entire loops of putative stem-loop hairpins, the mechanism may involve excision of four bases at the stem-loop interface and ligation in an inverted orientation, or perhaps recombination within the stemmed region itself, conceivably involving a four-stranded crossover event (Kelchner and Wendel 1996).

In terms of topology, separate analyses of only intron or spacer data, as well as those treating indels in different ways (included or excluded from PM analyses), showed remarkably congruent results. The 50% majority-rule consensus phylogenetic tree based on the combined data shows that *Castanea* species are geographically structured, supporting results from Manos and Stanford (2001). However, relationships within the critical clade containing taxa from Europe, North America, and China could not be resolved in their analysis. In this study, cpDNA data strongly support the monophyly of the Chinese lineage. A 4-bp inversion in the *trn*L intron and four substitutions (two in the *trn*T-L intergenic spacer, one in the *trn*L intron, and the other one in the *trn*L-F spacer) define the monophyletic Chinese clade, while the American and European species are supported in a separate clade (Tables 3 and 4).

The geographically structured cpDNA pattern is not congruent with the current phylogeny based on bur and cupule characteristics. The genus *Castanea* consists of three sections with the section *Eucastanon* being com-

Table 3 Structural mutations in the *trn*T-L-F region within the genus *Castanea*

Position	Size (bp)	Type	Source	Taxa
290–301	12	Deletion	Duplication	*C. dentata*
484–556	73	Deletion		*C. dentata*
1,246–1,249	4	Inversion		*C. mollissima*, *C. seguinii*, *C. henryi*
1,259–1,265	7	Deletion	Duplication	*C. mollissima* (M1)
1,268–1,285	18	Insertion	Duplication	*C. pumila* var. *pumila* (Vc1) and *C. pumila* var. *ozarkensis* (P7,P9)
1,608–1,632	25	Insertion	Duplication	*C. henryi*

Table 4 Substitutions detected within the genus *Castanea* at variable sites of the *trnT*-L-F region

Taxon	*trn*T-L spacer													*trn*L intron					*trn*L-F spacer								
	8	1	1	3	3	3	4	5	5	6	7	8	8	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	7	1	6	5	6	8	5	0	6	4	7	1	4	0	0	1	1	3	4	5	5	5	6	6	7	7	8
		6	6	3	6	8	7	8	4	6	3	9	1	3	5	0	6	3	8	0	8	8	4	8	4	9	1
														6	4	9	5	2	4	4	0	9	9	1	5	6	4
C. dentata	G	G	T	G	C	C	G	–	G	G	A	G	T	G	G	G	G	A	T	C	A	T	G	T	A	C	A
C. pumila var. *pumila* (P4, P6, Vc1)	G	G	T	G	C	C	G	A	G	G	A	G	T	G	G	G	T	A	T	C	A	T	T	T	A	C	G
C. pumila var. *pumila* (Vc9)	G	G	T	G	C	A	G	A	G	G	A	G	T	G	G	G	G	A	T	C	A	T	T	T	A	C	G
C. pumila var. *ozarkensis* (3–6, T6D)	G	G	T	G	C	C	G	A	G	G	A	G	T	G	G	G	G	A	T	C	A	T	T	T	A	C	G
C. pumila var. *ozarkensis* (P7, P9)	C	G	T	G	C	C	G	A	G	G	A	G	T	G	G	G	G	A	T	C	A	T	T	T	A	C	G
C. sativa	G	T	T	A	C	A	G	A	G	G	G	G	T	G	G	G	G	C	T	C	A	G	T	T	A	C	G
C. mollissima (M1)	G	G	T	G	A	C	T	G	G	T	G	T	T	G	G	A	G	A	T	T	A	T	T	T	T	C	G
C. mollissima (M4)	G	G	T	G	A	C	T	G	G	T	G	T	T	G	G	A	G	A	T	T	A	T	T	T	T	C	G
C. seguinii	G	G	T	G	A	C	T	G	G	G	G	T	T	G	G	A	G	A	T	T	A	T	T	T	A	C	G
C. henryi	G	G	T	G	A	C	T	G	G	G	G	T	T	G	G	A	G	A	T	T	A	T	T	T	T	C	G
C. crenata	G	G	G	G	A	C	G	A	A	G	G	G	G	A	T	G	G	A	C	C	G	T	T	C	A	A	G
Quercus falcata	G	G	T	G	A	C	G	A	G	G	G	G	G	A	G	G	G	A	T	C	A	T	T	T	A	C	G

prised of five species from different geographical areas, all characterized by three nuts per cupule. However, the phylogenetic tree based on the cpDNA data shows that the section *Eucastanon* is paraphyletic. The differentiation is best explained by their current geographical ranges. However, species with disjunct distribution on the two continents share a lot of similarity. *C. mollissima* and *C. seguinii* in China and *C. dentata* in North America are all characterized by three nuts per cupule, in contrast, *C. henryi* in China and *C. pumila* in North America are characterized as having only one nut per cupule. However, it is quite possible that a reduction in the number of nuts per cupule may have evolved more than once in *Castanea*. A reduction in flower number from the three-flowered dichasium to a single central flower has occurred in almost all genera of Fagaceae (Nixon and Crepet 1989; Manos et al. 2001). As discussed in a review by Wen (1999), phylogenetic analyses have revealed that many of the hypothesized intercontinental sister species pairs are not the closest relatives. The fact that the putative intercontinental species pairs of Chinese chestnut and American chestnut as well as Chinese chinkapin and American chinkapin do not appear to be sister species pairs may be explained in two ways. First, the species pairs may be relics of formerly widespread ancestral species that have experienced further diversification in eastern Asia and eastern North America after their separation. Thus, the intercontinental species pairs may have been sister species originally but then became further diversified (Wen 1999). Second, the morphologically similar species pairs may be the products of a combination of a common history and convergence in distant but similar environments (eastern Asia and eastern North America have similar climate and latitudes). In this case, they may never have been sister species (Wen 1999). It is critical to distinguish ecological from historical factors that cause the appearance of "disjunction" (Qui et al. 1995). The incongruence between the phylogenetic analysis and the morphological characters may also indicate that the morphological cohesion (or stasis) of three nuts per cupule may be the result of evolutionary constraints and the evolution of one nut per cupule might be independent events on two continents. Chloroplast capture has been considered an important factor that can distort phylogenetic relationships at lower taxonomic levels (Soltis and Kuzoff 1995). Due to the fact that *Castanea* species are insect and wind-pollinated and can hybridize freely, there is a possibility for gene flow via introgression between sympatric species in China as well as in North America. Such events could result in the formation of two separate clades in phylogenetic analyses. In contrast, strong directional selection pressure and human involvement (domestication of economic important chestnut crops) can aggravate morphological divergence despite gene flow in two continents. This hypothesis would be in accordance with geographical structure in *Castanea*, and thus might possibly explain the lack of congruence between chloroplast and morphological characters in phylogenetic analyses. However, it does not explain the position of *C. sativa* in a separate clade with the North American species. The placement of *C. sativa* in the North America clade was further strengthened by recent sequencing analysis of several nuclear DNA regions (Lang, unpublished results). The basal position of *C. crenata* in the genus points to the Far East as the center of diversification, suggesting vicariance from Asia via Europe to North America. More detailed cpDNA analyses are forthcoming to support this conjecture.

Acknowledgements We thank Cliff Parks of the University of North Carolina, Sandra Anagnostakis of the Connecticut Agricultural Research Station, and Alexandru-Lucian Curtu for providing samples.

References

Applequist WL, Wallace RS (2002) Deletions in the plastid *trn*T-*trn*L intergenic spacer define clades within Cactaceae subfamily Cactoideae. Plant Syst Evol 231:153–162

Bakker FT, Culham A, Gomez-Martinez R, Carvalho J, Compton J, Dawtrey R, Gibby M (2000) Patterns of nucleotide substitution in angiosperm cpDNA *trn*L (UAA)-*trn*F (GAA) regions. Mol Biol Evol 17:1146–1155

Clegg MT, Zurawaki G (1992) Chloroplast DNA and the study of plant phylogeny: present status and future prospects. In: Doyle JJ (ed) Molecular systematics of plants. Chapman & Hall, New York, pp 275–294

Cummings M, King L, Kellogg E (1994) Slipped-strand mispairing in a plastid gene: *rpo*C2 in grasses (Poaceae). Mol Biol Evol 11:1–8

Dane F, Hawkins LK, Huang H (1999) Genetic variation and population structure of *Castanea pumila* var. *ozarkensis*. J Am Soc Hortic Sci 124:666–670

Dane F, Lang P, Huang H, Fu Y (2003) Intercontinental genetic divergence of *Castanea* species in eastern Asia and eastern North America. Heredity 91:314–321

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791

Fineschi S, Taurchini D, Villani F, Vendramin GG (2000) Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* mill. (Fagaceae) throughout southern European countries. Mol Ecol 9:1495–1503

Fukuda T, Yokoyama J, Ohashi H (2001) Phylogeny and biogeography of the genus *Lycium* (Solanaceae): inferences from chloroplast DNA sequences. Mol Phylogenet Evol 19:246–258

Graham SW, Patrick A Reeves, Analiese CE Burns, Olmstead RG (2000) Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. Int J Plant Sci 161: S83–S96

Graves AH (1950) Relative blight resistance in species and hybrids of *Castanea*. Phytopathology 49:1125–1131

Huang H, Dane F, Norton JD (1994) Allozyme diversity in Chinese, Seguin and American chestnut (*Castanea* spp.). Theor Appl Genet 88:981–985

Huang H, Carey WA, Dane F, Norton JD (1996) Evaluation of Chinese chestnut cultivars for resistance to *Cryphonectria parasitica*. Plant Dis 80:45–47

Huang H, Dane F, Kubisiak TL (1998) Allozyme and RAPD analysis of the genetic diversity and geographic variation in wild populations of the American chestnut (Fagaceae). Am J Bot 85:1013–1021

Jaynes R (1975) Chestnut. In: Moore J (ed) Advances in fruit breeding. Purdue Univ. Press, West Lafayette, pp 490–503

Johnson GP (1988) Revision of *Castanea* sect. *Balanocastanon* (Fagaceae). J Arnold Arbor 69:25–49

Kelchner SA (2000) The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann Mo Bot Gard 87:482–498

Kelchner SA, Wendel JF (1996) Hairpins create minute inversions in non-coding regions of chloroplast DNA. Curr Genet 30:259–262

Kelchner SA, Clark LG (1997) Molecular evolution and phylogenetic utility of the chloroplast *rpl*16 intron in *Chusquea* and the Bambusoideae (Poaceae). Mol Phylogenet Evol 8:385–397

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

Lang P, Huang H (1999) Genetic diversity and geographic variation in natural populations of the endemic *Castanea* species in China. Acta Bot Sin 41:651–657 (in Chinese)

Levinson G, Gutman G (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Manos PS, Stanford AM (2001) The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere. Int J Plant Sci 162:S77–S93

Manos PS, Zhou ZK, Cannon CH (2001) Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. Int J Plant Sci 162:1361–1379

McDade LA, Moody ML (1999) Phylogenetic relationships among Acanthaceae: evidence from noncoding *trn*L-*trn*F chloroplast DNA sequences. Am J Bot 86:70–80

Morton BR (1995) Neighboring base composition and transversion/transition bias in a comparison of rice and maize chloroplast noncoding regions. Proc Natl Acad Sci USA 92:9717–9721

Nixon KC, Crepet WL (1989) *Trigonobalanus* (Fagaceae): taxonomic status and phylogenetic relationships. Am J Bot 6: 828–841

Olmstead R, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. Am J Bot 81:1205–1224

Qui Y, Chase MW, Parks CR (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

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92:142–166

Simmons MP, Ochotreana H (2000) Gaps as characters in sequence-based phylogenetic analyses. Syst Biol 49:369–381

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (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

Soltis DE, Kuzoff RK (1995) Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). Evolution 49:727–742

Stanford A (1998) The biogeography and phylogeny of *Castanea, Fagus*, and *Juglans* based on *mat*K and ITS sequence data. UNC, Chapel Hill

Swofford DL (2000) PAUP. Phylogenetic Analysis Using Parsimony. Version 4. Sinauer Associates, Sunderland, Massachusetts

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109

Vijverberg K, Bachmann K (1999) Molecular evolution of a tandemly repeated *trn*F(GAA) gene in the chloroplast genomes of *Microseris* (Asteraceae) and the use of structural mutations in phylogenetic analyses. Mol Biol Evol 16:1329–1340

Villani F, Pigliucci M, Benedettelli S, Cherubini M (1991) Genetic differentiation among Turkish chestnut (*Castanea sativa* Mill) populations. Heredity 66:131–136

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

Zuker M, Mathews DH, Turner DH (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide in RNA biochemistry and biotechnology. NATO ASI Series, Kluwer, Boston