Molecular Phylogenetics and Evolution 51 (2009) 157–168

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Evidence for genetic association between East Asian and western North American *Crataegus* L. (Rosaceae) and rapid divergence of the eastern North American lineages based on multiple DNA sequences

Eugenia Y.Y. Lo [a,b,*], Saša Stefanović [b], Knud Ib Christensen [c], Timothy A. Dickinson [a,d]

[a] Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, Ont., Canada M5S 3B2
[b] Department of Biology, University of Toronto at Mississauga, 3359 Mississauga Rd. N Mississauga, Ont., Canada L5L 1C6
[c] University of Copenhagen, Natural History Museum of Denmark, Botanical Garden and Museum, Oe. Farimagsgade 2B, DK-1353 Copenhagen K
[d] Green Plant Herbarium (TRT), Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, Canada M5S 2C6

ARTICLE INFO

Article history:
Received 12 February 2008
Revised 13 November 2008
Accepted 27 January 2009
Available online 6 February 2009

Keywords:
Beringian land bridge
Crataegus
Dispersal-vicariance
East Asia–western North America disjunction
Homoplasy
Hybrid origins
Rapid divergence
Topological incongruence

ABSTRACT

Phylogeographic relationships were constructed for 72 Old and New World *Crataegus* species using combinations of four chloroplast and up to five nuclear regions. Maximum parsimony, maximum likelihood, and Bayesian results yield consistent relationships among major lineages. The close associations of the East Asian and western North American species point toward ancient trans-Beringian migrations. Relationships among eastern North American species are poorly resolved and few groups are identified that are congruent with existing classifications. Scant variation and short internal branches among these species suggest rapid divergence associated with polyploidy and hybridization. Incongruence between the chloroplast and nuclear data, and morphology suggest hybrid origins of three species from an extinct European lineage (the male parent) and three different North American female parents. Europe and eastern North America are suggested as the most recent common areas for *Crataegus*; at least four dispersal events are inferred to explain the present distribution of the genus.

1. Introduction

Disjunct distributions, such as those between the floras of eastern Asia and eastern North America, have been well documented over the last several decades (Graham, 1972; Boufford and Spongberg, 1983; Taylor, 1990; Iwatsuki and Ohba, 1994; Xiang et al., 1998; Guo, 1999; Wen, 1999; Donoghue et al., 2001; Donoghue and Smith, 2004). The North Atlantic land bridge (NALB) and the Beringian land bridge (BLB) have been postulated as routes of floristic interchange between Eurasia and North America in the Tertiary that have contributed to modern global floral and faunal disjunctions (Hopkins, 1967; Tiffney, 1985a,b; Tiffney and Manchester, 2001).

Turning from floristic relationships to the phylogeography of particular groups, Evans and Campbell (2002) concluded that what is now recognized as Rosaceae supertribe Pyrodae (= subfamily Maloideae plus the genus *Gillenia*; Potter et al., 2007) originated in North America. They did so on the basis of phylogenetic analyses of molecular and non-molecular data, and on the basis of the modern distributions of the basal genera in the group. Analyses of the implications of this hypothesis are only just beginning, and here we present the results of our analyses of molecular data from a large, and widely distributed, member of the Pyrodae, the genus *Crataegus* L.

Species of *Crataegus* (hawthorns) are shrubs and small trees. The earliest fossils attributed to the genus date from the mid-Tertiary (reviewed by DeVore and Pigg, 2007). About 140–200 species have been described in the genus, distributed widely in the north temperate regions of both hemispheres (Phipps et al., 1990). Based on cladistic analyses of morphological data, Phipps (1983) suggested that the most basal species in the genus are the two members of *Crataegus* section *Mexicanae* (Loud.) Rehder, one from southern China and the other from Mexico. He further hypothesized that the modern distribution of the genus could have resulted from multiple trans-Beringian migrations (Phipps, 1983). Phipps' hypothesis conflicts with that of an exclusively North American origin for the Pyrodae (Evans and Campbell, 2002), and by implication the North American origin of some or all of the genera included in the supertribe. *Crataegus* is thus a compelling candidate for phylogeographic study because of its wide distribution and rich species diversity in the Northern Hemisphere, and

* Corresponding author. Fax: +1 416 586 7921.
E-mail address: eugenia.lo@utoronto.ca (E.Y.Y. Lo).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.01.018

because the phylogeographic relationships between its species have not yet been evaluated with molecular data.

One of the most recent classifications of *Crataegus* divides the genus into 15 sections and 35 series based on geographical localities and morphologies (Phipps et al., 1990). More than 100 species representing 11 sections are found in the New World, whereas 60 or more species, representing four sections, are known from the Old World (Phipps and Muniyamma, 1980; Phipps et al., 1990; Christensen, 1992; Gu and Spongberg, 2003). Taxonomy of this genus is particularly complicated for many North American species that are characterized by polyploidy and gametophytic apomixis (Muniyamma and Phipps, 1979, 1984; Talent and Dickinson, 2005, 2007). For instance, only two out of the 10 species of series *Cerrones* and *Douglasianae* in western North America (WNA) are diploid (Talent and Dickinson, 2005). The others have been shown to include tetraploids and reproduce by apomixis (Dickinson et al., 2008). In eastern North America (ENA), nearly two-thirds of the species are triploids or tetraploids (Talent and Dickinson, 2005). Together with the diploids, these species are currently divided into eight sections based on a combination of vegetative and reproductive characters (Phipps and Muniyamma, 1980; Phipps et al., 1990). However, this classification has not been tested with molecular data and the cladistic relationships between diploid and polyploid species are unclear. In comparison with the New World species, *Crataegus* of the Old World seems not to be as complicated taxonomically. Over three quarters of the species in Europe (EUR), northern Africa, and East Asia (EA) are diploids (Talent and Dickinson, 2005), and they are morphologically well defined (Phipps et al., 1990; Christensen, 1992; Gu and Spongberg, 2003).

Here we (1) test the biogeographic hypotheses (Phipps, 1983) concerning the Old and New World (EA, ENA, WNA, and EUR) *Crataegus*; (2) identify the origin(s) of three ENA species that are morphologically related to European species (Phipps, 1998); (3) attempt to resolve relationships among ENA species by adding three more nuclear gene regions; and (4) compare relationships based on phylogenetic inference with the existing morphologically-based infrageneric classification. Our ultimate goals are to shed light on the historical biogeography and evolution of diploid and polyploid taxa in *Crataegus*.

2. Materials and methods

2.1. Taxon sampling and DNA regions used

Our sampling aims to maximize the taxonomic and geographical coverage of *Crataegus*. A total of 72 species representing 14 of the 15 sections from both the Old and New World are included regardless of ploidy level (Table 1; Supplementary Appendix 1; the only section not sampled is monospecific sect. *Cuneatae* (Rehder ex Schneider) Rehder). One to six individuals per species and at least five species per section are examined, except for the monospecific sections. Species of *Amelanchier*, *Malus* and *Aronia* are used as outgroups. Samples were either collected in the field or obtained from botanical gardens. Except as noted in Suppl. Appendix 1, voucher specimens are deposited in the Green Plant Herbarium of the Royal Ontario Museum (TRT). Total DNA was extracted from either silica gel dried or frozen tissues using a small-scale modified method of Tsumura et al. (1995). Four intergenic regions of the chloroplast genomes and two nuclear regions including ITS1-5.8S-ITS2 and *LEAFY* second intron, that were used in Lo et al. (2007) for less than half of the taxa studied here, are employed to obtain the fundamental phylogenetic framework of the genus.

For further resolution and statistical support among species of East Asia and North America, sequences of three additional nuclear regions including the *LEAFY* first intron, 3′-portion of the *PISTILLATA* gene, and partial *PEPC* gene are obtained from 53 species representing 10 sections (as indicated with an asterisk in Table 1), together with *Malus angustifolia* as outgroup. The first two nuclear genes are members of the MADS-box gene family that are involved in floral and vegetative development of flowering plants. The second intron of the *LEAFY* gene has been used for phylogenetic reconstructions in genera of Rosaceae (Oh and Potter, 2003, 2005; Lo et al., 2007), whereas the longer first intron has not been investigated. *PISTILLATA* is suggested to be present as a single or low copy gene (Goto and Meyerowitz, 1994). Its longest first intron has been shown to be phylogenetically more informative than the ITS and *trnL* regions in Brassicaceae (Bailey and Doyle, 1999), but its four downstream introns are expected to be even more variable than the first intron because of the absence of transcription factor binding sites (Sieburth and Meyerowitz, 1997). Lastly, the *PEPC* gene has been reported to have low copy number and contains 9 introns and 10 exons in most plants (Matsuoka and Minami, 1989; Lepiniec et al., 1994). Some of these introns have been shown to be phylogenetically informative in different flowering plant families (Gehrig et al., 2001; Olson, 2002; Malcomber, 2002; Helfgott and Mason-Gamer, 2004; Lohmann, 2006).

2.2. PCR amplification and sequencing

Primer sequences and PCR conditions for the chloroplast regions, nuclear ribosomal ITS, and *LEAFY* intron 2 were described in Lo et al. (2007). For the other three nuclear regions, primers and PCR conditions are presented in Table 2. These primers were designed based on the conserved mRNA sequences of other Rosaceae species obtained from the NCBI database as potentially applicable to related genera. Chloroplast sequences were obtained by direct sequencing, whereas PCR products of most nuclear sequences were cloned using pDrive vector (QIAGEN), and at least three clones per individuals were sequenced. Because polyploid individuals may contain multiple allelic sequences within a locus when heterozygous, up to eight clones were sequenced to test for intraspecific polymorphisms. Plasmids were sequenced in both forward and reverse directions on either an ABI 377 or an ABI 3100 (Applied Biosystems) automated DNA sequencer with the DyeNamic or BigDye dye terminator cycle sequencing kits. Representative sequences for each region for each species were deposited in GenBank with the accession numbers presented in Suppl. Appendix 2.

2.3. Computational analyses

2.3.1. Sequence alignment and variation comparisons

Sequences of the nine examined regions were aligned separately with ClustalX (Thompson et al., 1997) and manually adjusted with the Sequence Alignment Editor version 1.d1 (SEAL; Rambaut, 2002). Gaps that are parsimony informative were coded into multistate characters with SeqState version 1.32 (Müller, 2005) and appended to the sequence matrices. Pairwise divergences among taxa for chloroplast and nuclear regions were estimated using DNADIST program of PHYLIP version 3.66c (Felsenstein, 2006). The HKY85 (Hasegawa et al., 1985) model which allows unequal base frequencies and transition and transversion rate was used for divergence comparisons.

2.3.2. Tree reconstructions

Phylogenetic trees were constructed for two major purposes with different datasets. The first was to infer relationships among the species from the four main biogeographic areas (EA, WNA, ENA, and EUR). This was achieved by separate and combined analyses of the four cpDNA regions, ITS, and *LEAFY* intron 2. The second was to further resolve relationships among the East Asian and

Table 1
Summary of *Crataegus* samples included in this study. Locality and voucher data can be found in Supplementary Appendix 1. Number of species that were reported as diploid (2*x*), triploid (3*x*), and tetraploid (4*x*) according to Talent and Dickinson (2005) are indicated for each taxonomic section. Asterisks indicate sections in which species were sequenced with the three additional nuclear regions *LEAFY* intron 1, partial *PEPC* and *PISTILLATA* genes for further phylogenetic analyses. Numbers in bold denote species that contain both diploid and polyploid individuals, as described in greater detail in Suppl. Appendix 1.

Taxonomic section	No. of species	No. of individuals	Biogeographic regions	Ploidy level composition
Mespilus	1	3	Europe	2*x*
*Mexicanae**	1	1	Central America	2*x*
Crataegus	13	37	Central Europe, Eurasia	2*x* (7), 4*x* (6)
*Sanguineae**	11	17	East Asia	2*x* (5), 3*x* (1), 4*x* (2)
Hupehenses	1	2	West Asia	2*x*
*Douglasianae**	8	25	Western North America	2*x* (2), 4*x* (6 **+ 1**)
*Parvifoliae**	1	2	Eastern North America	3*x*
Cordatae	1	2	Eastern North America	3–4*x*
*Virides**	1	2	Eastern North America	2*x*
Microcarpae	1	2	Eastern North America	2*x*
*Lacrimatae**	4	5	Eastern North America	3*x* (1), 4*x* (3)
*Aestivales**	3	7	Eastern North America	2*x* (2), 3*x* (1)
*Crus-galli**	5	10	Eastern North America	2*x* (2), 3*x* (2), 4*x* (**1**)
*Coccineae**	15	30	Eastern North America	2*x* (3), 3*x* (2 **+ 1**), 4*x* (9)
*Brevispinae**	1	4	Eastern North America	2*x*

Table 2
Information of chloroplast and nuclear primers used in the present study. Primers for *LEAFY* intron 1, *PEPC*, and *PISTILLATA* genes were designed based on the conserved regions of the mRNA sequences of Maloideae taxa obtained from the NCBI database and their accession numbers are provided. The two forward primers of the *PISTILLATA* (F1 and F2) are located respectively on the 3′ MADS-domain and 3′ K-domain, and the reverse primer is on the 5′ C-terminus of the gene. Only PIST-F2 and PIST-R are used in this study because the amplified downstream introns are suggested more variable and of readable length for sequencing without internal primers. These nuclear primers are shown applicable to *Crataegus*, *Amelanchier*, *Malus*, and potentially other Pyrinae species.

	Primer sequence (5′–3′)	Tm (°C)	Amplified length (bp)	References/mRNA sequences used
Chloroplast regions				
trnG-trnS	F: GAACGAATCACACTTTTACCAC R: GCCGCTTTAGTCCACTCAGC	58	750	Hamilton (1999)
psbA-trnH	F: GTTATGCATGAACGTAATGCTC R: CGCGCATGGTGGATTCACAAATC	55	500	Sang et al. (1997)
trnH-rpl2	F: CGGATGTAGCCAAGTGGATC R: GATAATTTGATTCTTCGTCGCC	55	500	Vaillancourt and Jackson (2000)
rpl20-rps12	F: TTTGTTCTACGTCTCCGAGC R: GTCGAGGAACATGTACTAGG	58	1000	Hamilton (1999)
Nuclear regions				
ITS1-5.8S-ITS2	F: TCCTCCGCTTATTGATATGC R: GGAAGGAGAAGTCGTAACAAGG	55	700	White et al. (1990)
LEAFY intron 1	F: GGATCCRGATGCCTTCTCTGCGAACTTGTTCAAGTGG R: GTTCTTTTTGCCACGCGCCACCTCCCCCGG-3'	62	1000	*Malus domestica* (AB162034), *Eriobotrya japonica* (AB162039), *Pseudocydonia sinensis* (AB162038), and *Cydonia oblonga* (AB162037)
LEAFY intron 2	F: CACCCACGACCITTYATIGTIACIGARCCIGGIGA R: CCTGCCIACRTARTGICKCATYTTIGGYTT	60	550	Oh and Potter (2003)
PEPC	F: CCGKCTTGCWACACCWGAGCTGGAG R: CCRGGWGCRTACTCGC	58	700	*Eriobotrya japonica* (EF523436) and *Prunus persica* (AJ243415)
PISTILLATA	F1: CARAGAAAATTGGGTAGGGGAAAGGTCGAGAT F2: CYCAGTACTACCARCAAGAAGC R: GTACTGATGATTGGGTTGTAAYGCRTTCACTTG	62	1200	*Malus domestica* (AJ291490) and *Prunus persica* (AY773012)

North American species. This was achieved by the addition of *LEAFY* intron 1, 3′ *PISTILLATA* gene and partial *PEPC* gene sequences. All datasets were analyzed using the maximum parsimony (MP) with equally weighted characters and maximum likelihood (ML) approaches in PAUP* 4.0b (Swofford, 2002), as well as Bayesian inference (BI) in Mr. Bayes 3.0b4 (Huelsenbeck and Ronquist, 2001).

For parsimony analyses, heuristic searches with 1000 random addition, tree bisection and reconnection (TBR) branch swapping, ACCTRAN optimization, MULTREES off, and with no more than 10 trees saved per replicate. The tree output was then used as the starting point for a second round of searches with the same settings except with MULTREES on. To assess clade support, bootstrap analyses (BS; Felsenstein, 1985) were conducted with 500 replicates, 10 random addition per replicate, TBR branch swapping, and MULTREES off options. A reduced dataset with one individual per species was included in the ML analyses to reduce the computational burden. The nucleotide substitution model was determined by the Hierarchical Likelihood Ratio Tests (hLRTs) and the Akaike Information Criterion (AIC) method using Modeltest, version 3.06 (Posada and Crandall, 1998). The best-fitting model and related parameters of datasets were used in both ML and Bayesian analyses. All ML searches were heuristic, with MULPARS and STEEPEST DESCENT options in effect, and TBR swapping. Bayesian analyses were performed with four Markov chains each initiated with a random tree and run for 5,000,000 generations, sampling every 100th generations. Likelihood values were monitored for stationarity and to determine the burn-in cut-off. Trees and other sampling points prior to the burn-in cut-off were discarded and the remaining trees were imported into PAUP* to generate a majority-rule consensus. Posterior probability values (PP; Ronquist and Huelsenbeck, 2003) were used to evaluate support of all nodes in the Bayesian trees.

2.3.3. Test for topological incongruence

Compatibility of tree topologies and bootstrap values were used for initial visual assessments of congruence between datasets. All chloroplast sequences were combined in phylogenetic analyses because these regions are linked as a single unit and no well-supported conflict is detected among individual trees. However, in order to test for the significance of congruence between nuclear datasets, we conducted Incongruence Length Differences tests (ILD; Farris et al., 1994), as implemented in PAUP* (Swofford, 2002; 'partition homogeneity test' option). We ran 1000 homogeneity replicates, each with 10 random sequence additions, under the same settings of MP heuristic searches as outlined above, except that MULTREES option was off. In addition, the statistical significance of incongruence between chloroplast and nuclear trees was assessed by two non-parametric tests in PAUP* (Swofford, 2002). The Templeton test (Wilcoxon signed-rank; Templeton, 1983) was performed under parsimony criterion. The one-tailed Shimodaira–Hasegawa test (SH; Shimodaira and Hasegawa, 1999) was performed with selected substitution models (TVM + I + G for chloroplast data and HKG + I + G for nuclear data), the same ML parameters as outlined above, and FULL optimization with 1000 bootstrap replicates.

2.3.4. Ancestral areas inference

To infer the potential ancestral geographical area(s), we employed both the maximum likelihood (ML−) and maximum parsimony (MP−) based approaches. For both methods, the MP tree generated from the total-evidence analysis (i.e., combined cpDNA and nDNA data, with conflicting taxa removed), was used as the primary input, together with the distribution matrix of taxa. The distribution of *Crataegus* was defined in five geographical areas: eastern North America (ENA), western North America (WNA), East Asia (EA), Europe (EUR), and Central America (CAM). For each terminal taxon, each area was coded either as present or absent.

The ML-based biogeographic inference approach was implemented by using Mesquite version 2.01 (Maddison and Maddison, 2008). The Markov k-state 1 parameter model (Mk1), which assumes a single rate change in characters and does not allow for a bias in gains versus loses was used. The proportional likelihood values of all estimated areas are obtained for each node of the molecular tree. In parallel with this, Dispersal-Vicariance Analysis v1.1 (DIVA; Ronquist, 1997), an MP-based approach, was also used to search for the most parsimonious ancestral geographical areas and relationships among regions where species are distributed. Ronquist's (1994) reversible parsimony assumes that speciation is caused by geographical vicariance. The most probable ancestral area usually has the fewest dispersal and extinction events. The root of each node defined as the most recent common area (MRCA) was obtained by optimal reconstruction with default settings. Maximum areas were constrained from 5 (i.e., the total number of our defined areas) to 2 (the minimal number of areas allowed by the software) in each of the iterations in order to determine the most reliable MRCA and number of dispersal events.

2.3.5. Divergence time estimation

Divergence times among the major lineages of *Crataegus* were estimated by the penalized likelihood (PL) and non-parametric rate-smoothing (NPRS) methods implemented in the program r8s version 1.71 (Sanderson, 2006). The former method is a semiparametric rate-smoothing approach that allows heterogeneous evolutionary rates among branches when estimating node ages in the phylogenetic trees (Sanderson, 2002), whereas the latter uses a least square smoothing approach that compares sum of square differences between branches (Sanderson, 1997). The ML treeblock with branch lengths and the r8s command block with the following settings: 3477 sites, PL/NPRS method, and truncated-Newton (TN)/POWELL optimization for the PL and NPRS methods (see Sanderson, 2006), respectively, were included in the input file. The optimal smoothing parameter is first obtained by cross validation analysis of the data and then implemented in the successive run where the ages of the nodes were calculated. Calibration was constrained with a minimum age of 44 million years ago (the Late Eocene; DeVore and Pigg, 2007), which is the approximate age of the stem group, i.e., the MRCA of the Pyrodae (including *Aronia*, *Malus*, *Amelanchier*, and *Crataegus*) and its sister clade containing members of the tribe Neillieae in the Spiraeoideae (Oh and Potter, 2005; Potter et al., 2007). For the stem group of *Crataegus*, we constrained the minimum age to be 25 mya according to the fossil leaves and fruits record (Oligocene; MacGinitie, 1933; Oliver, 1934; Lamotte, 1952; Hickey, 1984; Wolfe and Wehr, 1988; DeVore and Pigg, 2007), assuming that fossils of the stem lineages leading to the crown group of *Crataegus* should be older than the fossil age of *Crataegus* itself. Confidence intervals of divergence time were further estimated by the non-parametric bootstrap procedure (Baldwin and Sanderson, 1998; Sanderson and Doyle, 2001). One hundred bootstrap sequence matrices were generated from SeqBoot in PHYLIP 3.66c (Felsenstein, 2006). For each matrix, tree of same topology but different branch lengths were generated from the ML heuristic searches and were used for age estimation with the same parameters as outlined above in r8s. The central 95% of the age distribution provides the confidence interval.

3. Results

3.1. Sequence divergence

DNA divergence of chloroplast and nuclear regions were compared among taxa with respect to different geographical groups (Table 3). The variability found in individual chloroplast regions has been reported in Lo et al. (2007). The combined cpDNA sequences in the present study are less variable than those from the nuclear regions (0–2.67%; Table 3). All nuclear sequences reveal less than 1% intraspecific divergence among clones except *PISTILLATA*, which shows several substitutions and indels in the alignments in two out of the 8–10 clones in five of the individuals. These *PISTILLATA* sequences are believed to be paralogous copies of the gene and were excluded from the analyses in order to recover an accurate species phylogeny. Briefly, the average DNA divergence for the total sequences of the EUR species is the highest among all geographical groups and is about twofold higher than that of the ENA species (Table 3).

3.2. Species relationships

Based on the strict consensus parsimony tree the chloroplast data resolved three major clades within *Crataegus* labeled as A–C (Fig. 1A) that correspond to those obtained earlier with the smaller sample of mainly diploid species (Lo et al., 2007). Clade A comprises the EA species of *Crataegus* sect. *Sanguineae* and the WNA species of *Crataegus* sect. *Douglasianae* (series *Cerrones*, 100%BS and 100%PP; series *Douglasianae* plus *C. enderbyensis* in series *Purpureofructi*, Suppl. Appendix 1; 78%BS and 99%PP), together with ENA *C. spathulata* of monospecific sect. *Microcarpae*. Clade B is a sister group of clade A and it contains the ENA species belonging to eight different sections (Fig. 1) including monospecific section *Cordatae* (*C. phaenopyrum*) and monospecific series *Apiifoliae* (*C. marshallii*) from section *Crataegus*. Clade C (sister to clades A and B) contains the EUR and EA species of the remainder of sect. *Crataegus* and *C. hupehensis* (sect. *Hupehenses*) from EA (<50%BS; 84%PP). Although bootstrap support for all of these clades is relatively

Table 3
Estimates of sequence divergence among *Crataegus* species with respect of the four major geographical areas (ENA, WNA, EA, and EUR) for the chloroplast and nuclear regions. Asterisks indicate parsimony results estimated from sequences without *LEAFY* intron 1, *PISTILLATA*, and *PEPC* gene regions. PI, parsimony informative characters; MPTs, most parsimonious trees; CI, consistency index; RI, retention index; RC, rescaled retention index.

	Combined cpDNA	ITS	*LEAFY*2	*LEAFY*1	*PISTILLATA*	*PEPC*	Total
Aligned length (bp)	2487	675	697	1232	1243	683	3859*/7012
No. of taxa	79	79	79	52	52	52	79*/52
No. of sequences	194	234	204	149	118	171	190*/107
Divergence among taxa of EA	0–0.85%	0–5.96%	0.23–3.79%	0.78–1.65%	0.44–2.39%	0.63–1.73%	0.33–3.30%*
Divergence among taxa of ENA	0–2.3%	0.16–3.10%	0–1.74%	0.57–1.92%	0–2.04%	0.78–3.50%	0.58–2.20%*
Divergence among taxa of WNA	0.41–1.72%	0.48–4.25%	0.24–5.97%	0.29–1.97%	0.44–1.32%	0.63–2.86%	0.87–3.13%*
Divergence among taxa of EUR/EA	0.05–0.53%	0–6.86%	0.16–2.16%	—	—	—	0.57–4.28%*
Divergence among all *Crataegus* taxa	0–2.67%	0–7.79%	0–4.92%	0.23–4.20%	0.18–11.34%	0.78–4.65%	0.05–5.49%*
No. of PI	142	227	321	223	336	149	602*/864
No. of MPTs	>50,000	>50,000	>50,000	>50,000	>50,000	>50,000	>50,000*/>50,000
Tree length	476	836	589	521	699	476	1282*/2545
CI	0.721	0.702	0.857	0.716	0.764	0.502	0.738*/0.715
RI	0.849	0.887	0.92	0.899	0.975	0.841	0.894*/0.881
RC	0.612	0.623	0.789	0.644	0.745	0.422	0.571*/0.630

weak, similar topologies (i.e., the three major clades) are found also under the ML and Bayesian criteria (see Suppl. Figs. 1 and 2).

Nuclear data provide stronger bootstrap support and better resolution for clades A–C than do the chloroplast data. The ITS and *LFY*2 sequences were combined for tree reconstructions because no topological conflict was detected in the separate analyses of the two nuclear regions (Suppl. Figs. 3 and 4) and ILD test indicated no significant difference between the two datasets ($P = 0.067$). The strict consensus MPT (Fig. 1B) supports the same three clades (A–C) found on the cpDNA tree. This provides additional support for the association of EA with WNA species in clade A (72%BS; 99%PP), the sister group relationship of species in clade A and B (90%BS; 99%PP), as well as the monophyly of species in clade C (58%BS; 85%PP). Also, better resolution is provided within these clades. For instance, species in sect. *Sanguineae* that are unresolved in clade A of the cpDNA tree (Fig. 1A) form a monophyletic group in the nDNA tree (71%BS; 90%PP; Fig. 1B), adjacent to series *Douglasianae* (and *C. enderbyensis*; 78%BS; 96%PP) and *Cerrones* (79%BS; 99%PP). Likewise, the nuclear data provide some, albeit limited, resolution within clade B (Fig. 1B) that is lacking in the cpDNA tree (Fig. 1A). For example, *C. aestivalis*, *C. opaca*, and *C. rufula* make up sect. *Aestivales*, a group shown here to be monophyletic (76%BS; 96%PP). Species in clade C that formed polytomies with the chloroplast data (Fig. 1A), are resolved into at least two clades within the nuclear data, separating two EA species, *C. pinnatifida* (sect. *Crataegus*) and *C. hupehensis* (sect. *Hupehenses*; 97%BS; 100%PP) from the remainder of EUR species in sect. *Crataegus* (83%BS; 99%PP). Finally, the nuclear data yield a fourth clade (D, Fig. 1B; 86%BS and 99%PP), comprising the three species that represent monospecific sections *Cordatae* (*C. phaenopyrum*), *Microcarpae* (*C. spathulata*), and section *Crataegus* series *Apiifoliae* (*C. marshallii*). This clade shows a sister group relationship to species of clade A and B (Fig. 1B; 84%BS and 99%PP).

3.3. Topological incongruence

The phylogenetic positions of *C. marshalli*, *C. spathulata* and *C. phaenopyrum* (clade D; stars in Fig. 1) represent the only strongly supported incongruence between the cpDNA (Fig. 1A) and nDNA (Fig. 1B) trees. The Templeton tests on the cpDNA data with the nDNA-based topology, and vice versa, indicated significant differences in tree length (52 and 114 extra steps, respectively; $P < 0.001$ for both). Similarly, the SH tests rejected the monophyly of the three species (i.e., nDNA topology) with the cpDNA data ($-lnL = 4483.55$ and 4421.30; $P = 0.005$) as well as the polyphyletic origins of these species (i.e., cpDNA topology) with the nDNA data ($-lnL = 6334.96$ and 6234.17; $P = 0.004$). Based on these tests, we conclude with confidence that these species have significantly different phylogentic positions according to the cpDNA and nDNA results (Fig. 1).

3.4. Total sequence analyses

After exclusion of taxa that showed statistically significant topological incongruence (members of clade D), chloroplast and nuclear regions were combined in a total-evidence phylogeny (3477 bp; 69 taxa). The resulting ML tree (Fig. 2) reveals the same biogeographic and species relationships as found previously in the separate analyses (Fig. 1), except that these relationships received additional resolution and stronger bootstrap support. They include the backbone topology indicating the sister group relationship of clade C to clades A and B (99%BS; 100%PP), the close association of EA and WNA species in clade A (97%BS; 100%PP), and the monophyly of all ENA species in clade B regardless of sectional limits (80%BS; 95%PP). Within clades A and C, respectively, sister group relationships are recovered between species of series *Douglasianae* plus *C. enderbyensis* and series *Cerrones* (76%BS; 95%PP), and between *C. hupehensis* (section *Hupehenses*) and *C. pinnatifida* (section *Crataegus*) (83%BS; 95%PP).

In contrast to clades A and C, the relationships among the ENA species in clade B remain largely unresolved even after the chloroplast and nuclear data were combined. In an attempt to further resolve these relationships, sequences of *LFY*1, *PISTILLATA*, and *PECP* genes were added to the other, previously combined data, for an overall total of 7012 bp. No well-supported topological conflicts were observed among separate trees (Suppl. Figs. 3–7) and no significant differences were detected among datasets with the ILD tests ($P > 0.05$). The MP strict consensus tree (Fig. 3A) offers additional support for the alliance of WNA and EA species (100%BS; 100%PP). The three taxonomic groups, sect. *Sanguineae* (97%BS; 96%PP), ser. *Douglasianae* (99%BS; 100%PP), and ser. *Cerrones* (100%BS; 100%PP), are each shown to be monophyletic. However, the resolution among the ENA species and their bootstrap values still remain low. In clade B, only *Crataegus* section *Aestivales* appears as a monophyletic group, although with weak support (<50%BS and PP). Among species of sect. *Coccineae*, two pairs of sister species, *C. calpodendron* and *C. macracantha* (ser. *Macracanthae*; 76%BS; 100%PP), as well as *C. mollis* and *C. submollis* (ser. *Molles*; 97%BS; 100%PP), are identified. Other species, such as *C. uniflora* (sect. *Parvifoliae*) and *C. viridis* (sect. *Virides*) are mixed with members of sections *Coccineae* and *Crus-galli*, without clear groupings according to the existing taxonomic treatments (Fig. 3A). As can be seen from the phylogram (Fig. 3B), the internal branches in clade B are much shorter than those in clade A (EA and WNA species), suggesting a rapid divergence of the ENA species.

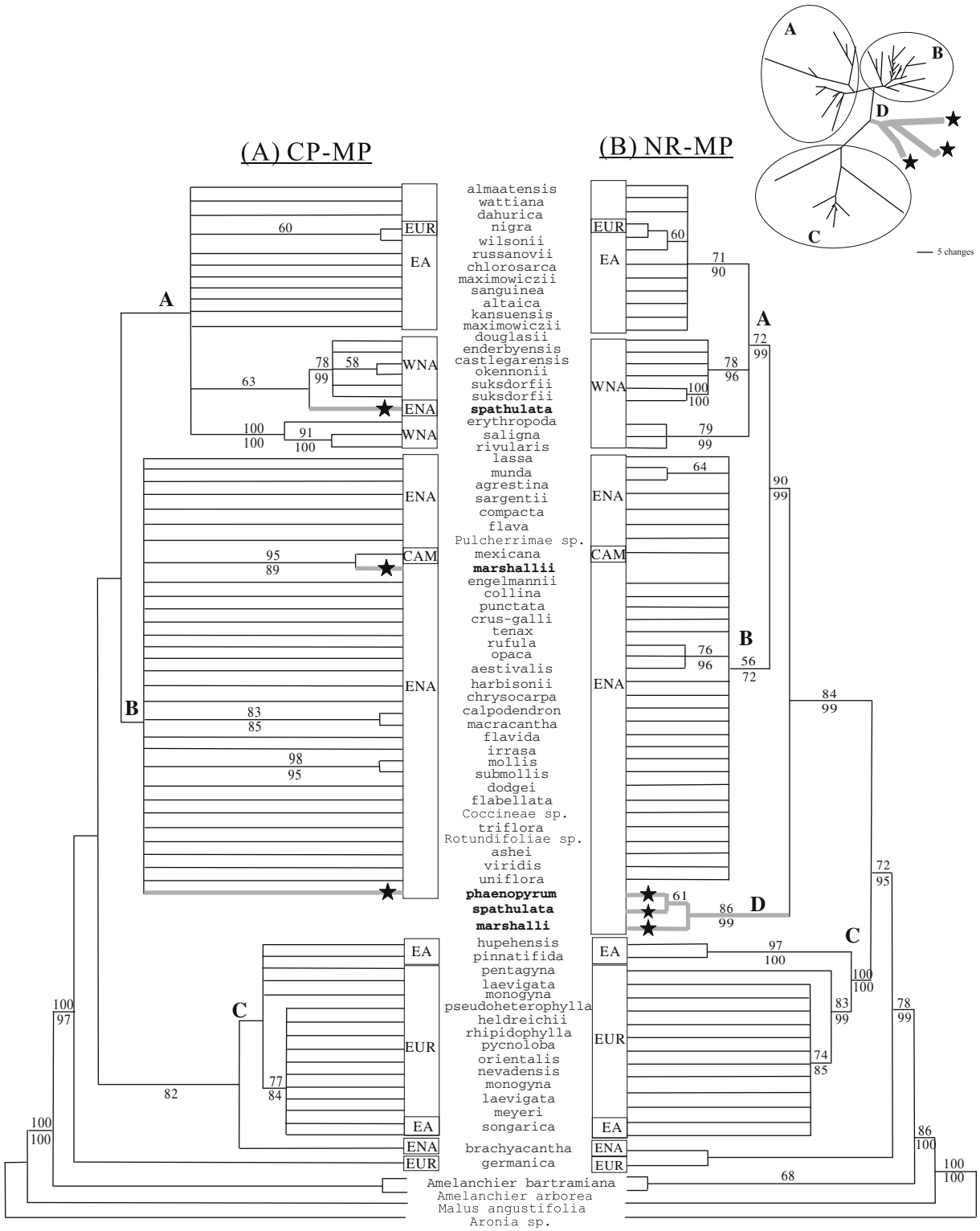

Fig. 1. Strict consensus trees from maximum parsimony (MP) analyses of the (a) combined *trnG-trnS*, *psbA-trnH*, *trnH-rpl2*, and *rps20-rpl12* chloroplast data and (b) combined ITS and *LEAFY* second intron data. The inset (upper right corner) shows an unrooted phylogram representing one of the MP trees from the nuclear dataset chosen to illustrate branch lengths (drawn proportionally to the amount of change). Species of *Amelanchier*, *Malus*, and *Aronia* were used as outgroups. Bars indicate the biogeographic distribution of the examined taxa. EA, East Asia; WNA, western North America; ENA, eastern North America; CAM, Central America; and EUR, Europe. Sections, series, species, and accession number of examined individuals can be found in Appendix 1 (Supplementary material). The three clades are labeled as (A) (taxa of section *Douglasianeae* from western North America, and section *Sanguineae* from East Asia); (B) (taxa of eastern North American sections); and (C) (taxa of sections *Crataegus* and *Hupehensis* from Europe and East Asia). Nodes with bootstrap values (BS; above branches) and posterior probabilities (PP; below branches) >50% are indicated. Stars indicate the three ENA taxa showing significantly conflicting positions between the chloroplast and nuclear trees.

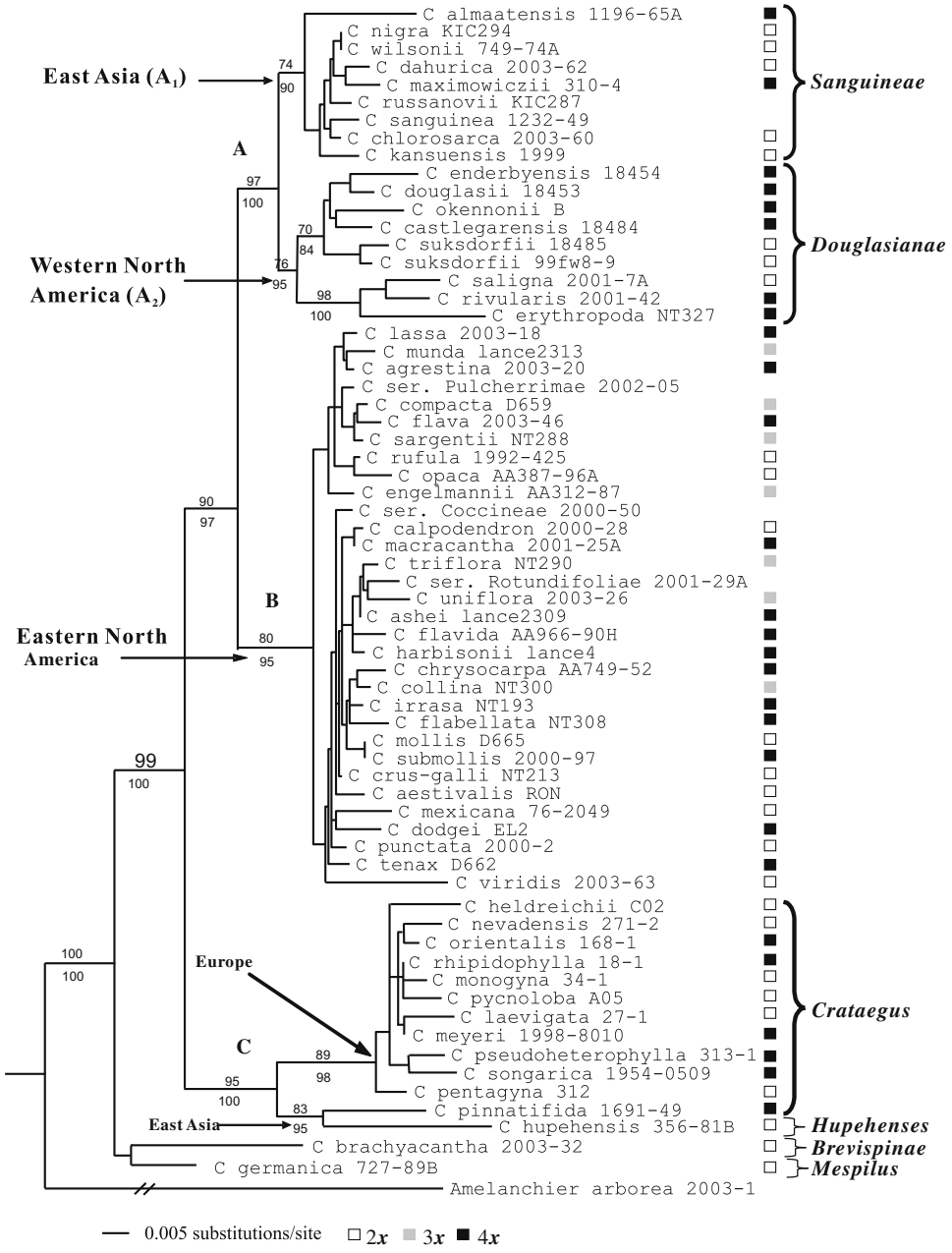

Fig. 2. Maximum likelihood (ML) tree based on combined nuclear and chloroplast data using the TVM model with *ln*L: −6867.76, base frequencies A = 0.3019, C = 0.1798, T = 0.3212, and G = 0.1970, proportion of invariable sites = 0.5747, and gamma shape = 0.6453. *Crataegus marshallii*, *C. phaenopyrum*, *C. spathulata* were omitted from the analyses because of their conflicting positions in the chloroplast and nuclear trees (Fig. 1). Biogeographic regions represented by taxa in each clade are indicated on the left and section and series are indicated on the right. Ploidy level is indicated for each taxon with shaded boxes where known based on Talent and Dickinson (2005). Bootstrap (BS; above branch) and posterior probability (PP; below branch) values >50% are indicated. Branch lengths are drawn proportional according to the TVM model except that of *Amelanchier arborea*. Other outgroup taxa, i.e., species of *Malus* and *Aronia*, are not shown.

3.5. Ancestral areas and estimated divergence times

Two biogeographic areas, eastern North America (proportional likelihoods = 0.23) and Europe (proportional likelihoods = 0.74), are shown to be the most probable ancestral areas for the root of *Crataegus* according to the Mesquite analysis under the ML criterion (WNA = 0.01; EA = 0.01; CAM = 0.01; Fig. 4). In addition, the most recent common areas (MRCA) of *Crataegus* inferred by MP-based approach (DIVA) under different constraints are indicated in Table 4. Briefly, a single solution was obtained for the root of *Crataegus* [ENA (eastern North America), WNA (western North America), EA (East Asia) and EUR (Europe)] when maximum areas were constrained to four or five, indicating that these four areas are equally probable to be the MRCA. When maximum areas were constrained to three, four alternative solutions were provided. When maximum areas were constrained to two, only one solution was obtained (Table 4). The same two areas, eastern North America (ENA) and Europe (EUR), are consistently shown to be MRCA of *Crataegus* in almost all solutions under different constraints. At least four dispersal events are inferred to explain the present distribution of *Crataegus* (Table 4 and Fig. 4). Ancestors of *C. hupehensis*, *C. songarica*, and *C. pinnatifida* are inferred to have dispersed from Europe into Asia. The eastern North American ancestors appear to have diversified into Mexico and Central America, and into western North America. Ancestors from western North America appear to have migrated into East Asia or vice versa (Fig. 4).

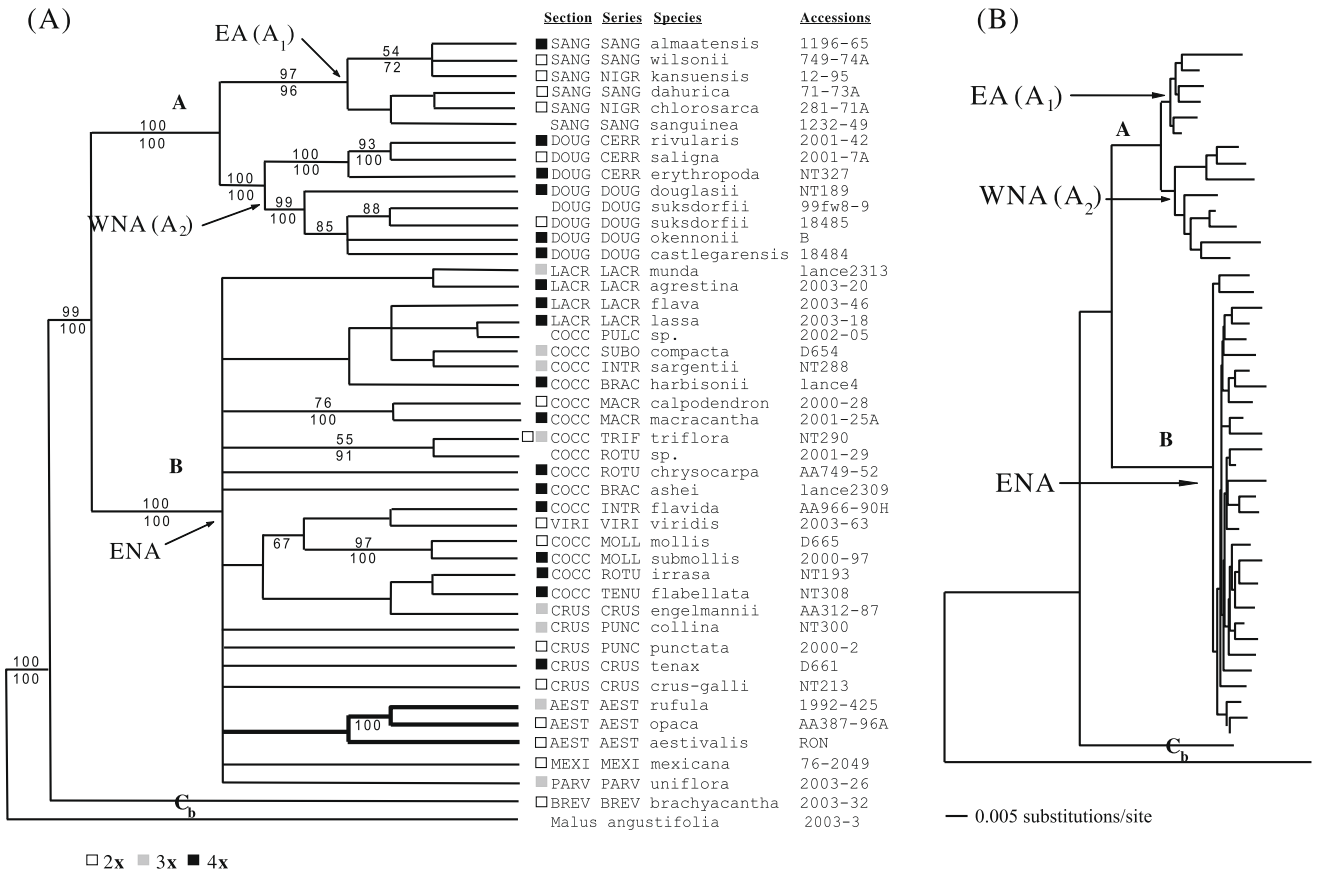

Fig. 3. (A) Strict consensus of 895 equally parsimonious trees from the maximum parsimony (MP) analysis of the eastern North American (ENA), western North American (WNA), and East Asian (EA) taxa using the combination of four chloroplast (*trnG-trnS*, *psbA-trnH*, *trnH-rpl2*, and *rps20-rpl12*) and five nuclear (ITS1-5.8S-ITS2, *LEAFY* intron 1 and 2, partial *PEPC* and *PISTILLATA*) regions. *Malus augustifolia* was used for rooting. Ploidy level, sections, series, species, and accessions of the examined individuals are indicated on the right. Branches in bold denote taxa of section *Aestivales* that are united as a monophyletic group, although with <50% BS values. Nodes with bootstrap values (BS; above branches) and posterior probabilities (PP; below branches) >50% are indicated. (B) Phylogram generated by maximum likelihood analysis showing shorter internal branches within the ENA than the WNA–EA clades.

Using the smoothed value of 1.0 obtained from the cross validation procedure in r8s, the PL method estimated the initial split of *C. brachyacantha* from the rest of the *Crataegus* species to have occurred as recently as 16.5 ± 3.7 mya (Suppl. Table 1). Divergence time between the EUR (clade C) and {EA–NA} lineages (clades A and B) was estimated to be at 14.3 ± 3 mya (approximately in the late Miocene), followed by the split between the ENA and {WNA–EA} lineages at about 9.9 ± 1.7 mya (in the early Pliocene). The WNA taxa appear to have diverged from the EA taxa at 4.6 ± 0.9 mya, which was about the same time when the ENA taxa diversified, i.e., at about 5.2 ± 1.1 mya. All the estimated ages based on the PL method were shown to be 2–5 mya older than those based on the NPRS method (Suppl. Table 1).

4. Discussion

4.1. Ancestral areas of Crataegus

Crataegus has a wide distribution in the Northern Hemisphere. Based on cladistic analyses of morphological characters Phipps (1983) suggested sect. *Mexicanae*, comprising East Asian *C. scabrifolia* (Franchet) Rehder and Central and North American *C. mexicana*, represents the least derived form of *Crataegus*. Although the samples here include only the latter species, our molecular data place it in a derived position within clade B (Figs. 1 and 2) thus falsifying this hypothesis. Biogeographic analyses under both the ML and MP criteria indicate instead that Eastern North America and Europe are the most recent common areas for all species (Table 4 and Fig. 4). These areas are likely to have contained the most primitive stock of *Crataegus*, consistent in part with the hypothesis of a North American origin for all of the supertribe Pyrodae (= Maloideae plus *Gillenia* Moench) on the basis of molecular phylogeny, floral morphology, and fossil evidence (Evans and Campbell, 2002). Moreover, most species of the sister genus of *Crataegus*, *Amelanchier* (and its segregate genera *Malacomeles* and *Peraphyllum*), as well as the other basal genera of the Pyrodae (e.g., *Kageneckia*, *Lindleya*, *Vauquelinia*, and *Gillenia*) are found today only in the New World (Campbell et al., 2007). An extensive rosaceous fossil record in North America and Europe, from the Eocene onwards (DeVore and Pigg, 2007), in addition to the earliest evidences of *Crataegus* fossil from the Okanogan Highlands in North America in at least the late Oligocene (MacGinitie, 1933; Oliver, 1934; Lamotte, 1952; Hickey, 1984; Wolfe and Wehr, 1988; DeVore and Pigg, 2007) support Mesquite and DIVA results in our present study.

In our phylognetic trees (Figs. 1 and 2), *Crataegus germanica* (formerly treated as genus *Mespilus*; Lo et al., 2007; Talent et al., 2008) and *C. brachyacantha* are the two species that are sister to the rest of the genus. *Crataegus germanica* is endemic to eastern Europe and western Asia (Appendix 1; Baird and Thieret, 1989). This area includes glacial refugia for Tertiary biota that have served to preserve the diversity of many extant species (Bennett et al., 1991; Hewitt, 1996, 2000; Taberlet et al., 1998). *Crataegus brachyacantha* occurs naturally in Louisiana and adjacent East Texas and Arkansas (Phipps, 1998), well south of the Pleistocene glacial maximum (Campbell, 1982; Berrggren and Prothero, 1992; Hewitt, 1996). We hypothesize that the ancestors of *C. germanica* and *C.*

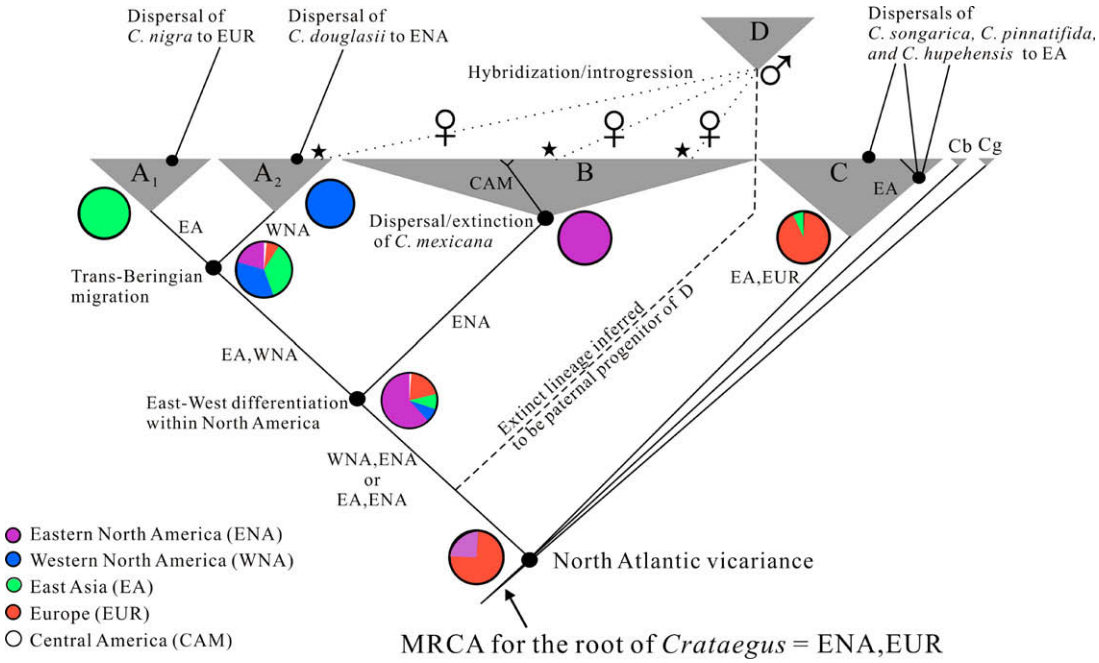

Fig. 4. Biogeographic model for *Crataegus* based on molecular phylogenies and Mesquite and Dispersal-Vicariance (DIVA) results. Biogeographic analyses were performed based on the combined chloroplast and nuclear tree (Fig. 3) and species were defined as five biogeographic areas as indicated below graph. Potential dispersal and vicariance events are indicated as bullets. Pie charts at each node indicate the proportional likelihoods of biogeographic areas to be the ancestral state for each clade of species estimated by the maximum likelihood method (Mesquite). The most recent common areas (MRCA) for each node as well as the root of *Crataegus* based on maximum parsimony method (DIVA) are indicated at the internodes. Species of clades A–D can be found in Figs. 1 and 2. Dashed line denotes the potential paternal lineage of hybrid taxa in clade D which may be extinct. Dotted lines denote potential hybridization occurred between the North American species (A_2 and B) and ancestors of the currently extinct European lineage. Putative maternal lineages of the hybrid taxa are labeled with stars. Cb and Cg denote *C. brachyacantha* and *C. germanica*, respectively.

Table 4
DIVA results indicating the most recent common areas (MRCA) obtained for the root of *Crataegus* and potential number of dispersal events computed by optimal reconstruction under default settings and maximum areas constrained from 5 to 2 in each of the iterations. The five biogeographic areas represented in our samples are stated as below. Only one solution is provided under all constraint tests except when the maximum areas are constrained to three and four alternative solutions are provided.

Maximum areas	MRCA for the root of *Crataegus*				No. of events
5	**ENA**,WNA,EA,**EUR**				4
4	**ENA**,WNA,EA,**EUR**				4
3	**ENA,EUR**	**ENA**,WNA,**EUR**	**ENA**,EA,**EUR**	WNA,EA,**EUR**	5
2	**ENA,EUR**				5

Notes: Eastern North America (ENA); Western North America (WNA); East Asia (EA); Europe (EUR); Central America (CAM).

brachyacantha may represent the earliest lineages of *Crataegus* that have survived, in one case east of the North Atlantic, and in the other to the west of it.

4.2. Intercontinental migrations and hybridization hypotheses

Some migratory pathways are suggested for the present distribution of *Crataegus* by the molecular data (Fig. 4). In the Old World, European species are likely to have migrated into eastern Asia, as evidenced by the phylogenetic association of *C. hupehensis*, *C. songarica*, *C. pinnatifida*, with species of sect. *Crataegus*. *Crataegus nigra* (sect. *Sanguineae*; Fig. 2), on the other hand, appears to have migrated in the opposite direction. The sister group relationships of sect. *Douglasianae* and sect. *Sanguineae* (Figs. 2 and 3), in conjunction with the synapomorphies in morphological features such as nutlet surface (Phipps, 1983), suggest that *Crataegus* could have dispersed between East Asia and western North America via the BLB until its closure in the late Pliocene (Donoghue et al., 2001; Tiffney and Manchester, 2001). Such an EA–WNA disjunction has also been reported in other temperate angiosperms (e.g., Liston et al., 1992; Xiang et al., 1998; Chen and Li, 2004; Li and Xiang, 2005; Oh and Potter, 2005; Nie et al., 2005). The estimated divergence ages between the EA and WNA taxa obtained in some of these studies, e.g., in *Datisca* (in the late Miocene; Liston et al., 1992) and *Kelloggia* (5.4 ± 2.3 mya; Nie et al., 2005), are similar to what we found in *Crataegus* (4.6 ± 0.9 mya; Suppl. Table 1). According to the analyses of divergence times, the separation between the ENA and WNA *Crataegus* appears to occur earlier than that between the EA and WNA taxa (Suppl. Table 1). This chronological order, which has also been documented in other plants (e.g., *Aesculus*; Xiang et al., 1998) and in invertebrates (e.g., lizards and frogs; Macey et al., 2006), supports the hypothesis that mid-continental cooling and aridification of North America preceded the closure of the BLB (Tiffney and Manchester, 2001; Gladenkov et al., 2002; Milne and Abbott, 2002; Milne, 2006).

Although there is no direct evidence for genetic interchange between the European and North American *Crataegus*, the conflicts detected between our chloroplast and nuclear data (Fig. 1) as well as morphological resemblances (Phipps, 1998) suggest that three species from the southeastern United States, *C. marshalli*, *C. spathulata* and *C. phaenopyrum*, are hybrids derived from European and North American ancestors. Conflicts between chloroplast and nuclear phylogenies could be attributed to several factors. One is by the stochastic outcome of the lineage sorting process that happens when variation of the ancestral lineages is not fully represented in the chloroplast and/or nuclear DNA of a taxon (Pamilo and Nei,

1988; Avise, 2004). However, separate analyses of the two independent nuclear loci ITS and *LEAFY*1 with multiple clones from multiple individuals reveal similar relationships (Suppl. Figs. 3 and 4) to those in the combined analyses (Fig. 1B), which eliminate the possibility of contamination and suggest that extant *Crataegus* lineages no longer show incomplete lineage sorting. Another concern is duplication of nuclear genes that could lead to conflicting gene trees when paralogs of some taxa are not detected. In our samples, there is no indication of substantial divergence within taxa in ITS and *LEAFY*1 that appears to be paralogous. Therefore, lineage sorting, contamination, and paralogy as source of conflict in our cpDNA and nDNA trees can be excluded with confidence, leaving hybridization as the most likely biological phenomenon to explain the observed conflict.

The morphology of *C. marshalli*, *C. spathulata*, and *C. phaenopyrum* also support the hybridization hypothesis. All three species have deeply to moderately lobed leaves with secondary veins leading to major sinuses, as in most European species (sect. *Crataegus*; Phipps, 1998). Conversely, reproductive features such as small flowers and small orange-red fruits with 3–5 pyrenes resemble those of other native North American *Crataegus* (Phipps, 1998). According to the chloroplast tree (Fig. 1A), species of North America are likely to be the maternal parents, but no one particular species can be specified because of the limited variation and lack of phylogenetic resolution and support. Paternal progenitors are not clearly indicated in the nuclear tree either (Fig. 1B). However, morphology suggests that these could belong to sect. *Crataegus*. We predict that their paternal lineage(s) could either be extinct or not included in our current sampling. It is noteworthy that such hybridization must have occurred before the NALB was interrupted, in the Oligocene or earlier (Tiffney, 1985b). This ancient paternal lineage apparently has been lost since that time, leaving the three putative hybrid taxa as a monophyletic group showing genetic similarities to both North American and European groups in the nuclear data (Fig. 1B).

4.3. Rapid divergence of eastern North American taxa and taxonomic implications

Some, but not all, infrageneric groups that have been recognized in *Crataegus* on morphological grounds can be recovered in analyses of our molecular data. On the one hand, sections *Brevispinae* (ENA), *Sanguineae* (EA) and *Crataegus* (EUR), and *Douglasianae* (WNA) are clearly supported (Figs. 2 and 3A). On the other hand, one of the striking features of our results is that several Eastern North American sections and series within clade B (Figs. 1 and 2), which have been described on the basis of substantial morphological differences, are *not* resolved even when data from additional gene regions are included (Fig. 3). Such a lack of resolution and extremely short internal branches (Figs. 2 and 3B) has been variously interpreted as a signature of rapid radiation (Rokas et al., 2005) or as the result of extensive hybridization (Campbell et al., 2007).

Taking into account the Quaternary glacial history of Eastern North America (Carrara et al., 1996; Marshall et al., 2002), both these processes could have been implicated. During phases of glacial expansion, vegetation zones occupied by *Crataegus* species were displaced southward. It is likely that species previously isolated during interglacial phases (if they were successful in maintaining refugial populations) came into contact and hybridized. There is abundant evidence of the contemporary potential for hybridization between *Crataegus* species within clade B (Phipps, 2005; Talent and Dickinson, 2007), as well as between North American species in clades A and B on the one hand, and between North American and introduced Eurasian species from clade C on the other (Love and Feigen, 1978; Wells and Phipps, 1989). Following deglaciation, as vegetation zones moved north, radiation into new habitats would have taken place, allowing successful genotypes (i.e., species of polyploid/hybrid origin) to establish themselves. It is perhaps during this phase that the phenological isolation seen between many North American *Crataegus* species (Phipps and Muniyamma, 1980; Campbell et al., 1991; Phipps, 2005) became established.

Our results, notably the monophyly of ENA clade B, could be taken as a support for the recognition of a separate subgenus, *Americanae*, for North American *Crataegus* species (El-Gazzar, 1980). Description of this subgenus was based on a facile interpretation of chromosome number in *Crataegus* at that time, as well as an over-simplification of the contrasts in leaf shape between the two continental *Crataegus* floras (Phipps, 1998). Our results here suggest that the two subgenera *Americanae* and *Crataegus*, as recognized by El-Gazzar (1980), would need to be recircumscribed in order to remove members of clade A (Figs. 1–3) from both subgenera, i.e., section *Douglasianae* from subgenus *Americanae*, and section *Sanguineae* from subgenus *Crataegus*. In addition, *C. brachyacantha* (sect. *Brevispinae*) and *C. germanica* (sect. *Mespilus*) would remain as separate groups in the genus, as discussed in Lo et al. (2007).

4.4. Conclusions

Based on the genetic associations, the present study indicates ancient trans-Beringian movements between East Asian and western North American *Crataegus*. Europe and eastern North America are suggested as the most recent common areas of modern *Crataegus*, and at least four dispersal events are inferred to explain the present distribution of *Crataegus*. Incongruence between the chloroplast and nuclear data, as well as morphology, suggest hybrid origins for *C. marshallii*, *C. phaenopyrum*, and *C. spathulata*. These species were potentially derived from hybridizations between the North American species serving as the maternal parents and the European lineages that might be either extinct or not included in our samples. Relationships among the eastern North American species remain poorly resolved and no clear cladistic groups are identified to support the existing classification. Much as with other groups, poor resolution and short internal branches observed in the ENA species suggest rapid divergence potentially driven by polyploidy and reticulation since these processes can provide an explanation for the substantial morphological differences among these species.

Acknowledgments

The authors thank Simona Margaritescu and Maria Kuzmina for assistance with sequencing; Christopher Campbell and Fannie Gervais for DNA extractions; Steve Brunsfeld, Rodger Evans, Fannie Gervais, Demetra Kandalepas, Ron Lance, Rhoda Love, Christopher S. Reid, Tatiana Romankova, Nadia Talent, Jörg Trau, and Peter Zika for assistance with plant collection and identification; Jenny Bull and Annabel Por for organizing the vouchers; and Tray Lewis (Louisiana), the Arkansas Natural Heritage Commission, The Nature Conservancy, the Toronto and Region Conservation Authority, the Arnold Arboretum of Harvard University, Jardin Botanique de Montréal, the Morton Arboretum, the University of California Botanical Garden, the North Carolina Arboretum, the North Carolina Botanical Garden, and the Royal Botanical Gardens (Burlington, Ontario) for access to trees. Asher Cutter, Jim Eckenwalder, Jean-Marc Moncalvo, Dan Potter, and two anonymous reviewers generously commented on earlier drafts of this work. Financial support from the Natural Sciences and Engineering Research Council of Canada (Grant A3430 to TAD, 326439-06 to SS), the Carlsberg Foundation (Grant 2005-1-462 to KIC), the Botany Department of

the University of Toronto, and the Royal Ontario Museum is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.01.018.

References

Avise, J.C., 2004. Molecular markers, natural history, and evolution, second ed. Sinauer Associates, Sunderland, Massachusetts.

Bailey, C.D., Doyle, J.J., 1999. Potential phylogenetic utility of the low-copy nuclear gene *PISTILLATA* in dicotyledonous plants: comparison to nrDNA ITS and *trnL* intron in *Sphaerocardamum* and other Brassicaceae. Mol. Phylogenet. Evol. 13, 20–30.

Baird, J.R., Thieret, J.W., 1989. The Medlar (*Mespilus germanica*, Rosaceae) from antiquity to obscurity. Econ. Bot. 43, 328–372.

Baldwin, B.G., Sanderson, M.J., 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. USA 95, 9406–9502.

Bennett, K.D., Tzedakis, P.C., Willis, K.J., 1991. Quaternary refugia of North European trees. J. Biogeog. 18, 103–115.

Berrggren, W.A., Prothero, D.R., 1992. Eocene and Oligocene climatic and biotic evolution: an overview. In: Prothero, D.R., Berrggren, W.A. (Eds.), Eocene and Oligocene Climatic and Biotic Evolution. Princeton University Press, Princeton, p. 568.

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

Campbell, J.J.N., 1982. Pears and persimmons: a comparison of temperate forests in Europe and eastern North America. Vegetation 49, 85–101.

Campbell, C.S., Greene, C.W., Dickinson, T.A., 1991. Reproductive biology in subfamily Maloideae (Rosaceae). Syst. Bot. 16, 333–349.

Campbell, C.S., Evans, R.C., Morgan, D.R., Dickinson, T.A., Arsenault, M.P., 2007. Phylogeny of subtribe Pyrinae (formerly the Maloideae, Rosaceae): limited resolution of a complex evolutionary history. Plant Syst. Evol. 266, 119–145.

Carrara, P.E., Kiver, E.P., Stradling, D.F., 1996. The southern limit of cordilleran ice in the colville and pend oreille valleys of northeastern Washington during the late wisconsin glaciation. Can. J. Earth Sci. 33, 769–778.

Chen, Z.D., Li, J.H., 2004. Phylogeny and biogeography of *Alnus* (Betulaceae) inferred from sequences of nuclear ribosomal DNA ITS region. Int. J. Plant Sci. 165, 325–335.

Christensen, K.I., 1992. Revision of *Crataegus* section *Crataegus* and Nothosect. *Crataeguineae* (Rosaceae-Maloideae) in the Old World. Syst. Bot. Mono. 35, 1–199.

DeVore, M.L., Pigg, K.B., 2007. A brief review of the fossil history of the family Rosaceae with a focus on the Eocene Okanogan Highlands of eastern Washington State, USA, and British Columbia, Canada. Plant Syst. Evol. 266, 45–57.

Dickinson, T.A., Lo, E.Y.Y., Talent, N., Love, R.M., 2008. Black-fruited hawthorns of western North America-one or more agamic complexes? Botany 86, 846–865.

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

Donoghue, M.J., Smith, S.A., 2004. Patterns in the assembly of the temperate forests around the Northern Hemisphere. Philos. Trans. R. Soc. Lond. B. 359, 1633–1644.

El-Gazzar, A., 1980. The taxonomic significance of leaf morphology in *Crataegus* (Rosaceae). Bot. Jahrb. Syst. 101, 457–469.

Evans, R.C., Campbell, C.S., 2002. The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. Am. J. Bot. 89, 1478–1484.

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

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

Felsenstein, J., 2006. PHYLIP (phylogeny inference package), version 3.66c. Available from: <http://evolution.genetics.washington.edu/phylip.html>.

Gehrig, H., Heute, V., Kluge, M., 2001. New partial sequences of Phosphoenolpyruvate Carboxylase as molecular phylogenetic markers. Mol. Phylogenet. Evol. 20, 262–274.

Gladenkov, A.Y., Oleinik, A.E., Marincovich, L., Barinov, K.B., 2002. A refined age for the earliest opening of Bering Strait. Palaeogeog. Palaeoclim. Palaeoecol. 183, 321–328.

Goto, K., Meyerowitz, E., 1994. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. Gen. Dev. 8, 1548–1560.

Graham, A., 1972. Outline of the origin and historical recognition of floristic affinities between Asia and eastern North America. In: Graham, A. (Ed.), Floristics and Paleofloristics of Asia and Eastern North America. Elsevier, Amsterdam, pp. 1–18.

Gu, C.Z., Spongberg, S.A., 2003. *Crataegus* L. Flora of China 9, 111–117.

Guo, Q., 1999. Ecological comparisons between Eastern Asia and North America: historical and geographical perspectives. J. Biogeog. 26, 199–206.

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

Helfgott, D.M., Mason-Gamer, R.J., 2004. The evolution of North American *Elymus* (Triticeae, Poaceae) allotetraploids: evidence from Phosphoenolpyruvate Carboxylase gene sequences. Syst. Bot. 29, 850–861.

Hewitt, G.M., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. 58, 247–276.

Hewitt, G.M., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.

Hickey, L.J., 1984. Changes in the angiosperm flora across the Cretaceous-Tertiary boundary. In: Berggren, W.A., Van Couvering, J.A. (Eds.), Catastrophes and Earth History. Princeton University Press, Princeton, NJ, pp. 279–313.

Hopkins, D.M., 1967. The Cenozoic history of Beringia-a synthesis. In: Hopkins, D.M. (Ed.), The Bering Land Bridge. Stanford University Press, Stanford, pp. 451–484.

Huelsenbeck, J.P., Ronquist, F., 2001. Mr. Bayes: a program for the Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Iwatsuki, K., Ohba, H., 1994. The floristic relationship between East Asia and eastern North America. In: Miyawaki, A., Iwatsuki, K., Grandtner, M.M. (Eds.), Vegetation in Eastern North America. Tokyo University Press, Tokyo, pp. 61–74.

Lamotte, R.S., 1952. Catalogue of the Cenozoic plants of North America through 1950, vol. 51. Geol. Soc. Am. Mem. New York, pp. 381.

Lepiniec, L., Vidal, J., Chollet, R., Gadal, P., Cretin, C., 1994. Phosphoenolpyruvate carboxylase: Structure, regulation, and evolution. Plant Sci. 99, 111–124.

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

Liston, A., Rieseberg, L.H., Hansen, M.A., 1992. Geographic partitioning of chloroplast DNA variation in the genus *Datisca* (Datiscaceae). Plant Syst. Evol. 181, 121–132.

Lo, E.Y.Y., Stefanović, S., Dickinson, T.A., 2007. Molecular reappraisal of relationships between Crataegus and Mespilus (Rosaceae, Pyreae)–Two genera or one? Syst. Bot. 32, 596–616.

Lohmann, L.G., 2006. Untangling the phylogeny of neotropical lianas (Bignonieae, Bignoniaceae). Am. J. Bot. 93, 304–318.

Love, R., Feigen, M., 1978. Interspecific hybridization between native and naturalized *Crataegus* (Rosaceae) in western Oregon. Madroño 25, 211–217.

Macey, R.J., Schulte, J.A., Strasburg, J.L., Brisson, J.A., Larson, A., Ananjeva, N.B., Wang, Y., Parham, J.F., Papenfuss, T.J., 2006. Assembly of the eastern North American herpetofauna: new evidence from lizards and frogs. Biol. Lett. 22, 388–392.

Maddison, W.P., Maddison, D.R., 2008. Mesquite: a modular system for evolutionary analysis. Version 2.5. Available from: <http://mesquiteproject.org>.

MacGinitie, H.D., 1933. The trout creek flora of southeastern oregon. In: Fossil Floras of Yellowstone National Park and Southern Oregon, Contributions to paleontology. Carnegie Institution, Washington DC, pp. 21–68.

Malcomber, S.T., 2002. Phylogeny of *Gaertnera* Lam. (Rubiaceae) based on multiple DNA markers: evidence of a rapid radiation in a widespread, morphologically diverse genus. Evolution 56, 42–57.

Marshall, S.J., James, T.S., Clarke, G.K.C., 2002. North American ice sheet reconstructions at the last glacial maximum. Quat. Sci. Rev. 21, 175–192.

Matsuoka, M., Minami, E.I., 1989. Complete structure of the gene for phosphoenolpyruvate carboxylase from maize. Eur. J. Biochem. 181, 593–598.

Milne, R.I., 2006. Northern hemisphere plant disjunctions: a window on Tertiary land bridges and climate change. Ann. Bot. 98, 465–472.

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

Müller, K., 2005. Incorporating information from length-mutational events into phylogenetic analysis. Mol. Phylogenet. Evol. 38, 667–676.

Muniyamma, M., Phipps, J.B., 1979. Cytological proof of apomixis in *Crataegus* (Rosaceae). Am. J. Bot. 66, 149–155.

Muniyamma, M., Phipps, J.B., 1984. Further cytological evidence for the occurrence of apomixis in North American hawthorns. Can. J. Bot. 62, 2316–2324.

Nie, Z.L., Wen, J., Sun, H., Bartholomew, B., 2005. Monophyly of *Kelloggia* Torrey ex Benth. (Rubiaceae) and evolution of its intercontinental disjunction between western North America and eastern Asia. Am. J. Bot. 92, 642–652.

Oh, S.H., Potter, D., 2003. Phylogenetic utility of the second intron of *LEAFY* in *Neillia* and *Stephanandra* (Rosaceae) and implication for the origin of *Stephanandra*. Mol. Phylogenet. Evol. 29, 203–215.

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

Oliver, E., 1934. A Miocene flora from the Blue Mountain, Oregon. In: Middle Cenozoic Floras of Western North America. Contribution to Paleontology. Carnegie Institution, Washington DC, pp. 1–27.

Olson, M.E., 2002. Combining data from DNA sequences and morphology for a phylogeny of Moringaceae (Brassicales). Syst. Bot. 27, 55–73.

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

Phipps, J.B., 1983. Biogeographic, taxonomic, and cladistic relationships between East Asiatic and North American *Crataegus*. Ann. Mo. Bot. Gard. 70, 667–700.

Phipps, J.B., 1998. Synopsis of *Crataegus* series *Apiifoliae*, *Cordatae*, *Microcarpae*, and *Brevispinae* (Rosaceae subfam. Maloideae). Ann. Mo. Bot. Gard. 85, 475–491.

Phipps, J.B., 2005. A review of hybridization in North American hawthorns. Ann. Mo. Bot. Gard. 92, 113–126.

Phipps, J.B., Muniyamma, M., 1980. A taxonomic revision of *Crataegus* (Rosaceae) in Ontario. Can. J. Bot. 58, 1621–1699.

Phipps, J.B., Robertson, K.R., Rohrer, J.R., 1990. A checklist of the subfamily Maloideae (Rosaceae). Can. J. Bot. 68, 2209–2269.

Potter, D., Eriksson, T., Evans, R.C., Oh, S., Smedmark, J.E.E., Morgan, D.R., Kerr, M., Robertson, K.R., Arsenault, M., Dickinson, T.A., Campbell, C.S., 2007. Phylogeny and classification of Rosaceae. Plant Syst. Evol. 266, 5–43.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rambaut, A., 2002. Se–Al Sequence Alignment Editor v2.0a11. University of Oxford, Oxford.

Rokas, A., Kruger, D., Carroll, S.B., 2005. Animal evolution and molecular signature of radiation compressed in time. Science 310, 1933–1938.

Ronquist, F., 1994. Ancestral areas and parsimony. Syst. Biol. 43, 267–274.

Ronquist, F., Huelsenbeck, J.P., 2003. Mr. Bayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

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

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

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

Sanderson, M.J., 2006. r8s, version 1.71, documentation and program. Available from: <http://loco.biosci.arizona.edu/r8s/>.

Sanderson, M.J., Doyle, J.A., 2001. Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. Amer. J. Bot. 88, 1499–1516.

Sieburth, L.E., Meyerowitz, E.M., 1997. Molecular dissection of the *AGAMOUS* control region shows that cis elements for spatial regulation are located intragenically. Plant Cell 9, 355–365.

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

Swofford, D.L., 2002. PAUP*: Phylogenetic analysis using parsimony (and other methods), version 40b10. Sinauer Associates, Sunderland.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7, 453–464.

Talent, N., Dickinson, T.A., 2005. Polyploidy in *Crataegus* and *Mespilus* (Rosaceae, Maloideae): evolutionary inferences from flow cytometry of nuclear DNA amounts. Can. J. Bot. 83, 1268–1304.

Talent, N., Dickinson, T.A., 2007. Ploidy level increase and decrease in seeds from crosses between sexual diploids and asexual triploids and tetraploids in *Crataegus* L. (Rosaceae, Spiraeoideae, Pyreae). Can. J. Bot. 85, 570–584.

Talent, N., Eckenwalder, J.E., Lo, E., Christensen, K.I., Dickinson, T.A., 2008. Proposal to conserve the name *Crataegus* against *Mespilus* (Rosaceae). Taxon 57, 1007–1008.

Taylor, D.W., 1990. Paleobiogeographic relationships of angiosperms form the Cretaceous and early Tertiary of the North American area. Bot. Rev. 56, 279–417.

Templeton, A.R., 1983. Phylogenetic inference from restriction endoclease cleavage site maps with particular reference to the evolution of human and the apes. Evolution 37, 221–244.

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

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

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

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface. flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 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.

Wells, T.C., Phipps, J.B., 1989. Interserial hybridization between *Crataegus monogyna* (series *Oxyacanthae*) and *Crataegus punctata* (series *Punctatae*) in southern Ontario. Can. J. Bot. 67, 2465–2472.

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

Wolfe, J.A., Wehr, W., 1988. Rosaceous *Chamaebatiaria*-like foliage from the paleogene of western North America. Aliso 12, 177–200.

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