doi:10.1111/j.1558-5646.2010.01063.x

RECONSTRUCTING RETICULATION HISTORY IN A PHYLOGENETIC FRAMEWORK AND THE POTENTIAL OF ALLOPATRIC SPECIATION DRIVEN BY POLYPLOIDY IN AN AGAMIC COMPLEX IN *CRATAEGUS* (ROSACEAE)

Eugenia Y. Y. Lo,[1,2,3,4] Saša Stefanović,[2] and Timothy A. Dickinson[3,5]

[1]*Department of Ecology and Evolutionary Biology, Yale University, 21 Sachem St., New Haven, Connecticut 06511*

[2]*Department of Biology, University of Toronto Mississauga, 3359 Mississauga Rd. N Mississauga, Ontario L5L 1C6, Canada*

[3]*Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, Ontario M5S 1B1, Canada*

[4]*E-mail: eugenia.lo@yale.edu*

[5]*Green Plant Herbarium, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada*

Received March 17, 2008
Accepted June 8, 2010

Polyploidy plays a prominent role in the speciation process in plants. Many species are known to be part of agamic complexes comprising sexual diploids and more or less exclusively asexual polyploids. However, polyploid formation has been studied in very few cases, primarily because of the challenges in examining these cases phylogenetically. In this study, we demonstrate the use of a variety of phylogenetic approaches to unravel origins and infer reticulation history in a diploid–polyploid complex of black-fruited *Crataegus*. The tree approaches are shown to be useful in testing alternative hypotheses and in revealing genealogies of nuclear genes, particularly in polyploid organisms that may contain multiple copies. Compared to trees, network approaches provide a better indication of reticulate relationships among recently diverged taxa. Taken together, our data point to both the autopolyploid and allopolyploid origins of triploids in natural populations of *Crataegus suksdorfii*, whereas tetraploids are formed via a triploid bridge, involving the backcross of allotriploid offspring with their diploid *C. suksdorfii* parent, followed by gene introgression from sympatric *C. douglasii*. Our findings provide empirical evidence for different pathways of polyploid formation that are all likely to occur within natural populations and the allopatric establishment of neopolyploids subsequent to their formation.

KEY WORDS: **Allopatric establishment, allopolyploidy, autopolyploidy, *Crataegus*, haplotype network, gene duplications.**

Polyploidy, or the multiplication of entire genomes, has played an important role in the diversification of species lineages and has been found common in plants as well as, increasingly, in fish and amphibians (Ptacek et al. 1994; Becak and Becak 1998; Bowers et al. 2003; Comber and Smith 2004; Cronn and Wendel 2004; Soltis et al. 2004; Meyers and Levin 2006). Two fundamental types of polyploids, autopolyploids and allopolyploids, have been described (Stebbins 1971). Autopolyploidy refers to the multiplication of the whole genome within a species, resulting from the formation and fertilization of unreduced gametes in crosses between conspecific individuals. There is abundant evidence that autopolyploidy is an important mechanism of sympatric speciation, particularly in plants (for a recent review see Soltis et al. 2007). Allopolyploidization also may involve fertilization of unreduced

gametes (i.e., the triploid bridge; Ramsey and Schemske 1998). Much more rarely, somatic chromosome doubling of a homoploid hybrid may also occur (Ramsey and Schemske 1998; Mallet 2007), but in both these cases two or more divergent genomes are involved, as a result of crosses between different species.

In angiosperms, approximately 2–4% of speciation events involve polyploidization (Otto and Whitton 2000). Studies of species-level phylogenies indicated that many speciation events are associated with a change in ploidy level (Soltis et al. 2007; Soltis and Burleigh 2009). Modern genome studies suggested that nearly all angiosperms likely have polyploidy in their evolutionary history (Cui et al. 2006) and the rate of polyploid formation in extant plant species may be comparable to or exceed the genic mutation rate (Ramsey and Schemske 1998). In the Rosaceae, Vamosi and Dickinson (2006) concluded that the increased species richness observed in clades with many polyploids was due to the polyploidization itself, and to the ensuing reproductive isolation from the parental clade(s). Polyploids may establish repeatedly within a species (Ramsey and Schemske 1998; Husband 2004; Husband et al. 2008) and many plant species, including those of *Crataegus* (Talent and Dickinson 2005), have been reported to contain different intraspecific cytotypes (e.g., Burton and Husband 2001; Baack 2004; Stuessy et al. 2004; Nakagawa 2006).

The transition from diploidy to polyploidy may result in correlated changes in breeding system, for example, from self-incompatibility to self-compatibility (Campbell et al. 1991; Mable 2004; Marhold and Lihová 2006; Dickinson et al. 2007; Husband et al. 2008), or from sexuality to apomixis (Mable 2004; Hörandl 2006; Whitton et al. 2008). Moreover, such a transition may significantly change the ecological interactions with animal herbivores and pollinators within a community (Thompson et al. 2004). Other related changes, such as those related to genome structure and gene expression patterns, have been shown to influence phenotype and fitness of an individual, and thus contribute to adaptation (see references in Otto 2007). Although there is an increasing amount of data highlighting the important consequences of polyploidy, studies that address evolutionary relationships among cytotypes in natural populations are still quite limited, mainly due to the challenges in applying phylogenetic methods to their study. Accurate documentations of the origins and reticulate history of diploid and polyploid taxa will facilitate our understanding of how polyploidy influences the speciation process (Otto and Whitton 2000; Soltis et al. 2007) and will help uncover patterns and correlates of evolution associated with polyploidy throughout plant history (Otto 2007; Soltis and Burleigh 2009).

Within last two decades, the advent of sequencing technologies and computational methods has substantially enhanced phylogenetic reconstructions in a large number of organisms, generally resulting in bifurcating hierarchical trees. Nevertheless, there are many phenomena in nature in which genetic material is not transferred in a hierarchical way (Rieseberg et al. 2000; Posada et al. 2002; Richardson and Palmer 2007) and the concept of "tree" may oversimplify our view of evolution. One such event involves the formation of polyploids. In theory, an autopolyploid inherits genomes only from a single progenitor species, and so it will appear as sister to the progenitor of a lower ploidy level on a bifurcating tree. On the other hand, allopolyploidy produces species with a network-like history that cannot be properly represented by bifurcating trees (Funk 1985; McDade 1992; Vriesendorp and Bakker 2005). Forcing reticulation to be displayed in a branching topology might lead to the lack of support for the resolved clades and/or collapse of hierarchical structure (Cassens et al. 2005; Vriesendorp and Bakker 2005; Huber and Moulton 2006). Hence, a network approach using orthologous nuclear genes has become common when examining reticulate relationships (e.g., Oxelman and Bremer 2000; Smedmark et al. 2003; Linder and Rieseberg 2004; Popp et al. 2005; Huber et al. 2006; Brysting et al. 2007).

In this study, we ask two broad questions: (1) what is the extent of reticulate evolution in the black-fruited hawthorn complex and (2) how do reticulate evolution and allopatric establishment of cytotypes contribute to the polyploid speciation process? Here, we employ both tree and network approaches to construct a reticulation model that shows the possible routes to the formation of *Crataegus suksdorfii* cytotypes. In particular, we sought to distinguish auto- versus allopolyploid origins of black-fruited hawthorn triploids and tetraploids from closely related diploids (all assignable to *C. suksdorfii*) and from more distantly related tetraploids (*C. douglasii*). The role of allopatric establishment in the polyploid speciation process is also discussed in the context of the present-day distributions of cytotypes. In addition, we compare the outcomes of a variety of approaches and suggest how they may be employed when constructing genealogical relationships with taxa that involve hybridization and polyploidy.

Materials and Methods

PLANT MATERIALS

Crataegus suksdorfii and *C. douglasii* were sampled at a total of 16 sites (Table 1) in the western United States and eastern Canada. Diploids are found only in *C. suksdorfii*, together with triploids and tetraploids (Talent and Dickinson 2005; Lo et al. 2009a). Diploids occur at lower elevations in northern California and western Oregon, Washington, and British Columbia (Dickinson et al. 2008). Triploids are widespread and locally abundant at higher elevations and further east in Washington and Idaho. Tetraploids, however, are of sporadic occurrence (Dickinson et al. 1996; Lo et al. 2009a). Over parts of their range, polyploid cytotypes of *C. suksdorfii* are sympatric with exclusively tetraploid

Table 1. Summary of *C. douglasii* and *C. suksdorfii* individuals included in amplifications of the two nuclear gene regions *PEPC* and *PISTILLATA*. Two duplicated paralogs, labeled S and L respectively, were identified in each gene. The total number of clones sequenced from each individual and number of clones found for each paralog (in parentheses) are indicated. Bolded letters indicate locality abbreviation used in the text.

Nuclear gene regions—*PEPC* and *PISTILLATA*				
Species	Label	State/Province; County; Locality	No. of *PEPC* clones (S-, L-copy)	No. of *PISTILLATA* clones (S-, L-copy)
Ploidy level	*C. douglasii* Lindl.			
4*x*	**CA3** 2006–13	California; Shasta; Hat Creek	10 (8, 2)	9 (5, 4)
4*x*	**ON20** EL11	Ontario; Grey; Big Bay, Colpoy's range	9 (7, 2)	14 (7, 7)
4*x*	**ON20** EL15	Ontario; Grey; Big Bay, Colpoy's range	8 (5, 3)	8 (3, 5)
4*x*	**ID2** EL138	Idaho; Latah; Little Boulder Creek	13 (8, 5)	10 (5, 5)
4*x*	**ID6** EL166	Idaho; Adams; Last Chance Campground, near Meadows	11 (8, 3)	11 (4, 7)
4*x*	**ID6** EL170	Idaho; Adams; Last Chance Campground, near Meadows	9 (6, 3)	9 (5, 4)
4*x*	**ID15** EL197	Idaho; Lemhi; US 93 S of Gibbonville	11 (5, 6)	8 (5, 3)
4*x*	**ID20** EL121	Idaho; Nez Perce; Little Potlatch Creek	10 (7, 3)	9 (4, 5)
4*x*	**MT2** EL32	Montana; Powell; Kleinschmidt Flat	11 (6, 5)	11 (5, 6)
4*x*	**MT2** EL39	Montana; Powell; Kleinschmidt Flat	15 (12, 3)	9 (4, 5)
4*x*	**MT2** EL41	Montana; Powell; Kleinschmidt Flat	8 (6, 2)	8 (4, 4)
4*x*	**WA5** S0703	Washington; Chelan	11 (7, 4)	11 (3, 8)
4*x*	**WA22** EL155	Washington; Whitman; South of Colfax	11 (9, 2)	8 (3, 5)
4*x*	**WA21** EL89	Washington; Thurston; Mound Prairie	12 (7, 5)	8 (3, 5)
	C. suksdorfii (Sarg.) Kruschke			
2*x*	**CA5** 2006–16	California; Siskiyou; Fay Lane	8 (5, 3)	8 (4, 4)
2*x*	**CA5** 2006–22	California; Siskiyou; Fay Lane	8 (6, 2)	8 (6, 2)
2*x*	**OR1** EL70	Oregon; Linn; Cogswell Foster Reserve	9 (6, 3)	9 (7, 2)
2*x*	**OR1** EL72	Oregon; Linn; Cogswell Foster Reserve	8 (7, 1)	8 (5, 3)
2*x*	**OR1** EL75	Oregon; Linn; Cogswell Foster Reserve	8 (4, 4)	7 (4, 3)
3*x*	**OR6** EL50	Oregon; Lane; Patterson Mountain Prairie	9 (5, 4)	9 (4, 5)
3*x*	**OR6** EL57	Oregon; Lane; Patterson Mountain Prairie	15 (9, 6)	8 (3, 5)
3*x*	**OR6** EL62	Oregon; Lane; Patterson Mountain Prairie	13 (10, 3)	8 (4, 4)
3*x*	**OR6** EL65	Oregon; Lane; Patterson Mountain Prairie	12 (8, 4)	13 (7, 6)
2*x*	**OR11** EL104	Oregon; Columbia; Sauvie Island	9 (6, 3)	9 (5, 4)
2*x*	**OR11** EL115	Oregon; Columbia; Sauvie Island	8 (3, 5)	7 (3, 4)
3*x*	**ID6** EL165	Idaho; Adams; Last Chance Campground, near Meadows	10 (6, 4)	12 (8, 4)
3*x*	**ID6** EL172	Idaho; Adams; Last Chance Campground, near Meadows	11 (8, 3)	9 (3, 6)
3*x*	**ID6** EL173	Idaho; Adams; Last Chance Campground, near Meadows	9 (6, 3)	8 (3, 5)
3*x*	**ID5** EL188	Idaho; Valley; North Beach, Payette Lake	13 (7, 6)	11 (8, 3)
4*x*	**MT2** EL26	Montana; Powell; Kleinschmidt Flat	14 (12, 2)	12 (6, 6)
4*x*	**MT2** EL30	Montana; Powell; Kleinschmidt Flat	13 (10, 3)	14 (7, 7)
4*x*	**MT2** EL36	Montana; Powell; Kleinschmidt Flat	12 (8, 4)	14 (6, 8)
4*x*	**MT2** EL45	Montana; Powell; Kleinschmidt Flat	13 (9, 4)	12 (7, 4)
2*x*	**WA7** Z18485	Washington; Clark	-	7 (4, 3)

C. douglasii. These two species can be distinguished morphologically by variation in the number of stamens per flower (approx. 20 in *C. suksdorfii*, and approx. 10 or fewer in *C. douglasii*). Some of the *C. douglasii* individuals can be ascribed to segregate species, such as *C. castlegarensis* and *C. okennonii* (Phipps and O' Kennon 1998, 2002; Dickinson et al. 2008), but for our purposes here such distinctions are unnecessary given the lack of genetic distinctiveness among these segregates (Lo et al. 2009a). Individuals used here are part of a larger sample of individuals reported elsewhere (Dickinson et al. 2008) and are vouchered by specimens deposited in the Green Plant Herbarium at the Royal Ontario Museum (TRT). Nuclear DNA content of all studied individuals was measured by flow cytometry and is reported as ploidy level (Table 1).

GENE MARKERS AND SEQUENCING STRATEGY

Given that our major aims were to examine the extent of reticulate evolution as well as to infer the genetic relatedness between diploid and polyploid taxa, different sampling strategies were applied to the nuclear and chloroplast sequences. Two unlinked nuclear genes, *PISTILLATA* and *Phosphoenolpyruvate Carboxylase* (*PEPC*), which were suggested to be single copy in other plants (Matsuoka and Minami 1989; Goto and Meyerowitz 1994; Bailey and Doyle 1999), were used in this study. Our sampling strategy sought to fully capture the allelic variation of these nuclear gene loci within individuals, especially those of polyploids in our studied sites. Thus, for these two genes from each individual, 8–15 clones were sequenced, respectively, but only one to four individuals per site were included (Table 1). On the other hand, because chloroplast markers are uniparentally inherited and each individual, regardless of its ploidy level, should contain only one copy, no cloning was necessary to capture variation within individuals. However, to increase the probability of accurately identifying maternal donors, we sought to maximally recover variation among individuals within sites. Therefore, a total of 132 individuals representing *C. douglasii* and *C. suksdorfii* from the 16 sites were sequenced for *psbA-trnH* and *trnH-rpl2* (Table 2). Details of DNA extraction, primers information, PCR conditions, and cloning protocols were described in Lo et al. (2009b).

SEQUENCE ANALYSES

Alignments were first produced automatically with ClustalX (Thompson et al. 1997) and then edited manually with Sequence Alignment Editor version 1.d1 (SE-Al; Rambaut 2002). Indels were coded as multistate characters (Simmons and Ochoterena 2000) with SeqState version 1.32 (Müller 2005) and appended to the sequence matrix for phylogenetic analyses. Because the two chloroplast spacers are linked on a haploid genome, sequences were combined and treated as a single marker for analyses; the two nuclear datasets were treated separately.

Table 2. Summary of haplotypes detected in *C. douglasii* and *C. suksdorfii* individuals. Locality labels can be referred to Table 1. The number of individuals (*N*) included from each site are indicated. Statistical parsimony networks of these haplotypes are shown in Figures 1 and 2B.

Distribution of haplotypes				
Ploidy level	Site	*N*	Chloroplast	*PEPC*-L
C. douglasii				
4*x*	ID2	11	M N P W	F K O
4*x*	ID6	11	M N O P X	F H K M
4*x*	ID15	7	M P T	F H M
4*x*	ON20	9	P Q	F J
4*x*	MT2	12	J L P V	F Q K L
4*x*	WA22	5	P R	F
4*x*	WA21	6	R S	F N
C. suksdorfii				
2*x*	CA5	5	D F	B G
2*x*	OR1	9	A B C E G	A B D E F
3*x*	OR6	20	A B C H	A C
2*x*	OR11	9	A B E G U	A B D E F
3*x*	ID5	2	O P	F M
3*x*	ID6	16	I O P	F G H I M P
4*x*	MT2	13	I J K L O P	F I L Q

Sequence recombination caused by PCR errors is expected to be more common in polyploids when more than one copy is available (Bradley and Hillis 1997; Judo et al. 1998; Cronn et al. 2002). This occurs when the DNA polymerase stops functioning or detaches from the template before elongation is complete. The partially extended product may prime to another allele in a subsequent cycle and result in a recombinant sequence. To identify and exclude potential recombinant sequences from the alignment matrices before phylogenetic analyses, the aligned sequences from all clones of each individual were compared using the pairwise scanning method implemented in the recombination detection program (RDP) version 2b.08 (Martin and Rybicki 2000) in conjunction with TOPALi version 2 (Milne et al. 2004). Sequences detected with recombinant points and/or showing incongruent positions among partitions were excluded from the analyses.

PHYLOGENETIC TREE AND NETWORK CONSTRUCTIONS

To compare the performances of tree and network approaches in revealing reticulation history, the phylogenetic relationships between diploid and polyploid taxa were constructed using different algorithms for all three datasets. Trees were constructed using maximum parsimony (MP) in PAUP* 4.0b (Swofford 2002), maximum likelihood (ML) in RAxML version 7.0.4 (Stamatakis et al. 2005), as well as by means of Bayesian inference (BI) using

Mr. Bayes 3.0b4 (Huelsenbeck and Ronquist 2001). For parsimony analyses, heuristic searches for the most parsimonious (MP) trees with equally weighted data were performed with 1000 random additions, tree bisection-reconnection (TBR) branch swapping, and with no more than 10 trees saved per replicate. The tree output was then used as a starting point for a second round of searches with the same settings except with MULTREES on.

Substitution models for the DNA datasets were selected from the 56 models implemented in ModelTest 3.7 (Posada and Crandall 1998) by employing the Akaike information criterion (AIC). Models selected for the chloroplast, *PEPC*, and *PISTILLATA* matrices were TVM + I, GTR + Γ+ I, and GTR + Γ, respectively. For ML analyses, bootstrap support (BS) was assessed with 1000 replicates with the rapid bootstrap algorithm implemented in the RAxML (Stamatakis et al. 2008). Bayesian analyses were performed with four Markov chains each initiated with a random tree and run for 10,000,000 generations, sampling every 100th generations. Likelihood values were monitored for stationarity with Tracer version 1.4.1 (Rambaut and Drummond 2007). 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 (PP) values (Ronquist and Huelsenbeck 2003) were used to evaluate support of all nodes in the Bayesian trees.

The second approach we used for inferring relationships is network reconstruction using the statistical parsimony (SP) and neighbor-net (NN) algorithms. The network approach is particularly useful when there is contradictory signal or reticulation in the data that cannot be unambiguously represented by a dichotomous phylogeny. The statistical parsimony method of Templeton et al. (1992), as implemented in TCS version 1.13 (Clement et al. 2000), has been shown to be the least erroneous among existing network methods (Woolley et al. 2008). It estimates haplotypes based on the uncorrected *p*-distances above which the parsimony principle is violated with more than 5% probability. All connections are iteratively joined among haplotypes only when the parsimony has a probability of at least 0.95 of being true as determined by coalescence theory, starting with the shortest distance until all haplotypes are joined or the distance exceeds the parsimony limit (Clement et al. 2000). Hence, this method emphasizes what is shared among haplotypes that differ minimally (Posada and Crandall 2001). Here, haplotypes were set to be connected by mutational steps with 95% confidence limits and the "gaps = missing" option. To assess support for these connections, we performed Bayesian analyses using a subset of samples representing all the observed haplotypes to obtain PP values. In addition to the haplotype-based analyses, we also generated an allele-based network to examine the level of reticulation in our nuclear data and to evaluate the potential impact of such to the level of resolution and topological confidence in our resulted trees. This was done with the neighbor-net method as implemented in SplitsTree 4.10 (Huson 1998), which generates a collection of compatible splits through iterations of agglomerative processes similar to the neighbour-joining method (Bryant and Moulton 2004). The result is presented as a splits graph and the shape of the graph is interpreted as a reflection of the extent of reticulation among taxa. To compare reticulations, splits graphs were generated separately for four diploid and polyploid combinations including: (1) 2*x*, and 3*x* from OR; (2) 2*x*, and 3*x* from ID; (3) 2*x*, and 4*x* from MT; (4) all 2*x*, 3*x*, and 4*x* taxa. Internode support was estimated by performing 1000 bootstrap replicates with the genetic distances and model parameters found in ModelTest for respective datasets.

TESTING ALTERNATIVE HYPOTHESES

To compare the statistical support for auto- versus allopolyploid origins of triploid and tetraploid *C. suksdorfii*, we used the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999) as implemented in PAUP* (Swofford 2002) to compare the best ML trees recovered from the respective nuclear data with the constraint trees constructed in MacClade (Maddison and Maddison 1992). Autopolyploidy was considered to be the null hypothesis. According to this hypothesis, all the alleles of a given polyploid cytotype are expected to be inherited only from the diploid taxon. Hence, the unconstrained trees should be statistically indistinguishable from the constrained ones in which the autopolyploid origin was imposed. Three cytotype combinations were constrained to be monophyletic: (1) 2*x*, and 3*x* from OR; (2) 2*x*, and 3*x* from ID; (3) 2*x*, and 4*x* from MT. The constrained trees were loaded separately as a backbone for ML analyses. Heuristic searches were conducted using the parameters from ModelTest to find the optimal trees compatible with the constraint. The likelihood scores of the constrained trees were then compared with the scores of the best unconstrained trees using the one tailed nonparametric SH tests.

Results

CHLOROPLAST GENEALOGIES

The two chloroplast spacers gave a total aligned length of 652 bp in which 23 of the 63 variable positions and six indels were parsimony informative. Although MP analyses (trees not shown) showed very little resolution of our taxa, ML and Bayesian trees (Fig. S1) indicate a strong association of 2*x* and 3*x C. suksdorfii* from Oregon (BS 93%; PP 100%). However, the 3*x* (ID5, ID6) and 4*x* (MT2) *C. suksdorfii*, as well as the 4*x C. douglasii* were not resolved. Compared with the tree results, the network analyses (Fig. 1) reveal clearer patterns of haplotype relationships.

A total of 24 haplotypes labeled as A–X (Table 2) were detected and they were distinguished from each other by—one to

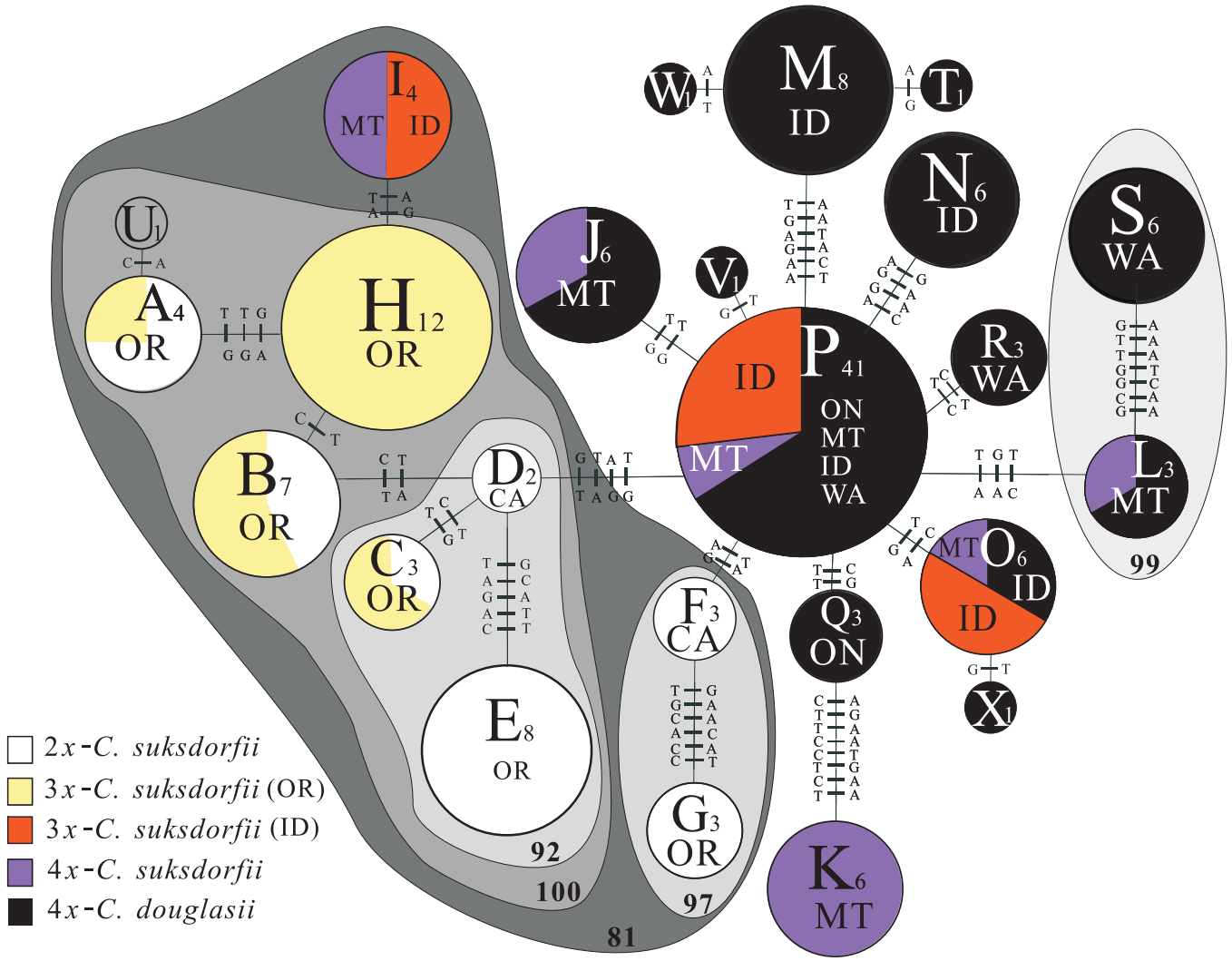

Figure 1. Statistical parsimony network of chloroplast haplotypes A–X (Table 2) obtained from 132 sequences representing diploid and polyploid individuals of *C. suksdorfii* and *C. douglasii* from 16 localities (Table 1). Ploidy level, locality, and number of individuals that share the same haplotype are indicated in each circle. Sizes of circles are approximately proportional to the number of individuals with the given haplotype. Bars on lines between circles represent site changes between haplotypes under the statistical parsimony criterion. Shaded outlines and percentages indicate clades and their level of support in Bayesian analyses.

eight mutation steps in the phylogenetic network (Fig. 1). Diploid *C. suksdorfii* (sites OR1, OR11, and CA5) contains eight haplotypes that were more or less connected except F and G (Table 2; Fig. 1). Three of these haplotypes (A–C) were shared with 3*x C. suksdorfii* from Oregon (OR6). Although haplotype H appeared to be unique among the Oregon triploids, it was differentiated from haplotype B by only one mutation. Haplotype I observed in 3*x C. suksdorfii* from the Idaho (ID5, ID6) and 4*x C. suksdorfii* from Montana (MT2) was found neighbour to haplotype H. The connection of haplotype I to A–H is shown to be supported in the tree analyses (BS 81%; PP 88%; Figs. 1 and S1). Tetraploid *C. douglasii* contains 13 haplotypes (Table 2). Haplotype P was the most common one, found in 41 individuals in total, including not only 4*x C. douglasii* ($N = 27$), but also 3*x* ($N = 11$) and 4*x* ($N = 3$) *C. suksdorfii*. Apart from haplotype P, haplotypes J, L, and O were also shared with either 3*x C. suksdorfii* from the Idaho (ID5, ID6) or 4*x C. suksdorfii* from Montana (MT2), or both (Table 2; Fig. 1). Haplotype K was found uniquely in 4*x C. suksdorfii* (MT2), and to be neighbor to haplotype Q detected in Ontario *C. douglasii* (ON20).

NUCLEAR GENEALOGIES

In this study, we detected two clearly distinguishable copies of *PISTILLATA* and *PEPC* genes in all of our samples, polyploids as well as diploids. This is not surprising in *Crataegus*, or in any other members of Rosaceae tribe Pyreae. This tribe appears to have originated as diploidized, hybrid tetraploids from *Gillenia*-like ancestors (Evans and Campbell 2002; Smedmark

et al. 2003; Freeling and Thomas 2006; Campbell et al. 2007). Other nuclear genes such as *GBSSI* (Evans et al. 2000) and *LEAFY* (Wada et al. 2002; Esumi et al. 2005) have also been reported to contain multiple copies in the Pyreae genera.

PEPC

Of the total of 332 sequences obtained from the 34 individuals, 19 were detected with significant recombinant points and showed incongruent positions among trees of different partitions, and hence were removed from further analyses. Amplicons of the partial *PEPC* genes are of approximately 740 bp in size. Despite intensive cloning efforts, only two copies were detected across all individuals, designated here as L (long) and S (short) copies based on the length (indels) and nucleotide differences. Open reading frames (ORFs) were identified in their exon sequences, suggesting functionality of both copies. The L and S copies of *PEPC* are resolved as two distinct clades in all tree analyses (Figs. 2A and S2).

In the clade containing the L copy (Fig. 2A), very little resolution is obtained from the tree analyses, except the clade containing sequences of 2*x* and 3*x C. suksdorfii* from Oregon (BS 85%; PP 98%). By contrast, the network analyses provide a better indication of taxon relationships. Seventeen haplotypes labeled as A–Q were identified from the 88 *PEPC*-L sequences (Table 2; Fig. 2B). Diploid *C. suksdorfii* contains haplotypes A, B, D, E, F, and G. Triploid *C. suksdorfii* from Oregon (OR6) either shares the same haplotype A or contains related haplotype of A in the diploids (BS 85%; PP 98%), supporting their close association as shown in the chloroplast data (Figs. 1 and S1). On the other hand, individuals of 3*x* (ID5, ID6) and 4*x* (MT2) *C. suksdorfii* were found to share, either separately or in combination, multiple haplotypes with 2*x C. suksdorfii* (F and G) and 4*x C. douglasii* (F, H, I, L, Q, and M; Fig. 2B), reflecting a higher level of heterogeneity within these individuals. In addition, Neighbor-Net (NN) analyses show that 2*x* and 3*x* (OR6) result in a star-like topology, whereas 2*x* and 3*x* (ID5, ID6) as well as 2*x* and 4*x* (MT2) both produce a much more reticulated topology (Fig. 3A–D). Differences in the shape of these NN-splits graphs are consistent with the lack of reticulation or a single origin of 3*x* from Oregon, but multiple origins of 3*x* from Idaho and 4*x* from Montana (whose sequences are found to be scattered in the total network; Fig. 3D).

Compared to the L copy, the S copy is shown to be more variable and resolved taxa into three moderately supported subclades A_1, A_2, and A_3 within the large clade A (Figs. 2A and S2). All sequences of 3*x* individuals from Oregon (OR6) are found to be monophyletic with the 2*x C. suksdorfii* (OR1, OR11) in clade A_2 (BS 80%; PP 99%), which supports a single origin of these triploids. However, sequences of 3*x* (ID) and 4*x* (MT) from other sites are found to be polyphyletic and nested in clades such as A_1 (BS < 50%; PP 68%) and A_3 (BS 76%; PP 86%) together with 2*x C. suksdorfii* (OR) and 4*x C. douglasii* (ID, MT, WA). In the remainder of the S-copy clade (labeled as clade B), only sequences of 4*x C. douglasii* (CA, ID, MT, ON, and WA) and 4*x C. suksdorfii* (MT) are observed (BS 78%; PP 91%; Figs. 2A and S2). Similarly, topologies of the NN-splits graphs (Fig. 3E–H) indicated a greater reticulation when 3*x* (ID) and 4*x* (MT) *C. suksdorfii* sequences were included, consistent with the *PEPC*-L results. The total network (Fig. 3H) suggests a considerable introgression of alleles between polyploid *C. suksdorfii* and *C. douglasii*.

PISTILLATA

Of the total of 321 sequences obtained from the 34 individuals, 13 were detected with significant recombinant points and showed incongruent positions among trees of different partitions, and thus were removed from the subsequent phylogenetic analyses. In the partial *PISTILLATA* gene of 1150 bp, about 140 bp in the first intron were deleted from all analyses because a hypervariable AT-rich region was detected that could not be aligned unambiguously. As with *PEPC*, despite intensive cloning only two copies of *PISTILLATA* were detected. They are designated as S (short) and L (long) and can be clearly identified based on indels and nucleotide differences. The *PISTILLATA*-L copy is shown to be more variable than the S copy (parsimony informative sites: 157 vs. 69) and no clear ORF was observed in their exons, thus suggesting that this is a pseudogene copy. Trees of *PISTILLATA* reveal two distinct and strongly supported clades, corresponding to the S and L copies (Figs. 4 and S3). Sequences of the S copy are almost completely unresolved regardless of the phylogenetic methods used (results not shown). By contrast, phylogenetic resolution was observed with the L copy where sequences are divided into clades A (BS 75%; PP 86%) and B (BS 93%; PP 97%). Within clade A, the subclade A_1 contains only sequences of 4*x C. douglasii* (ID, MT, ON, and WA) and 4*x C. suksdorfii* (MT) (BS 100%; PP 100%; Figs. 4 and S3), equivalent to clade B of *PEPC*-S (Figs. 2A and S2). The remainder of clade A is a mixture of sequences from 2*x* (OR), 3*x* (ID), and 4*x* (MT) *C. suksdorfii* together with those from 4*x C. douglasii* that are also poorly resolved. Sister to clade A is clade B that shows a monophyly of all 3*x C. suksdorfii* from Oregon (OR6; $N = 18$) together with 2*x C. suksdorfii* sequences from Oregon and California (OR1, CA5; $N = 5$; Fig. 4), supporting the single origin of these triploids. A comparable TCS-network constructed from *PISTILLATA*-L sequences was found to contain an excessive number of haplotypes and no clear pattern could be detected. Although the overall level of reticulation shown in the NN-splits graphs (Fig. S4) is not as high as those based on the *PEPC* data, similar trends can still be observed among comparisons.

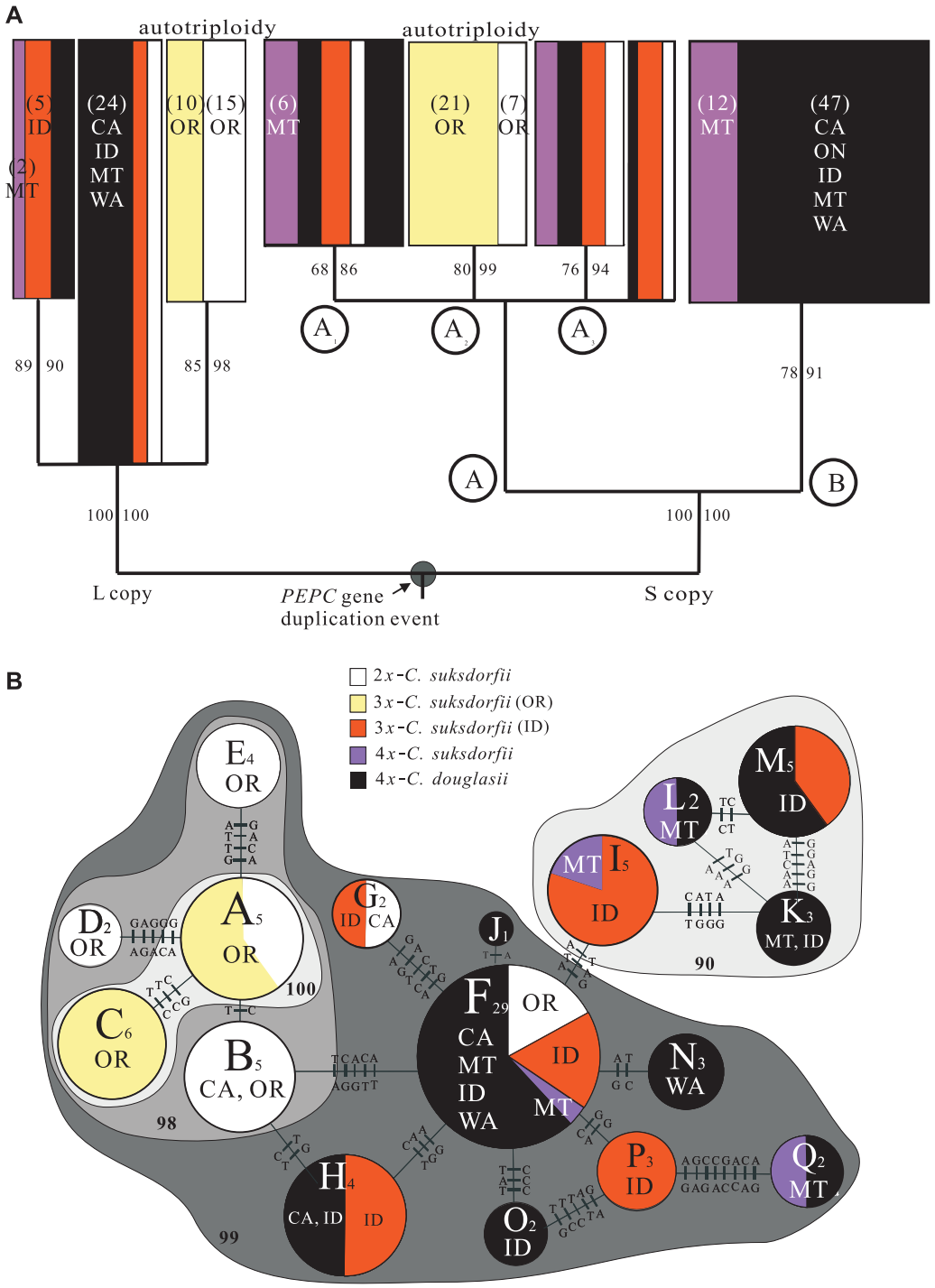

Figure 2. (A) Schematic representation of the strict consensus of MP trees for the L and S copies of *PEPC* (for details, see Fig. S5A,B). Bootstrap (left) and posterior probability (right) values are indicated on branches. Numbers in parentheses represent the number of sequences found in selected clades. Locality labels are indicated in Table 1; (B) Statistical parsimony network of nuclear haplotypes A–Q (Table 2) obtained from 91 *PEPC*-L paralog sequences representing diploid and polyploid individuals of *C. suksdorfii* and *C. douglasii*. Ploidy level, locality, and number of sequences found in the same haplotype are indicated in each circle. Sizes of circles are approximately proportional to the number of sequences with the given haplotype. Bars on lines between circles represent site changes between haplotypes under the statistical parsimony criterion. Shaded outlines and percentages indicate clades and their level of support in Bayesian analyses.

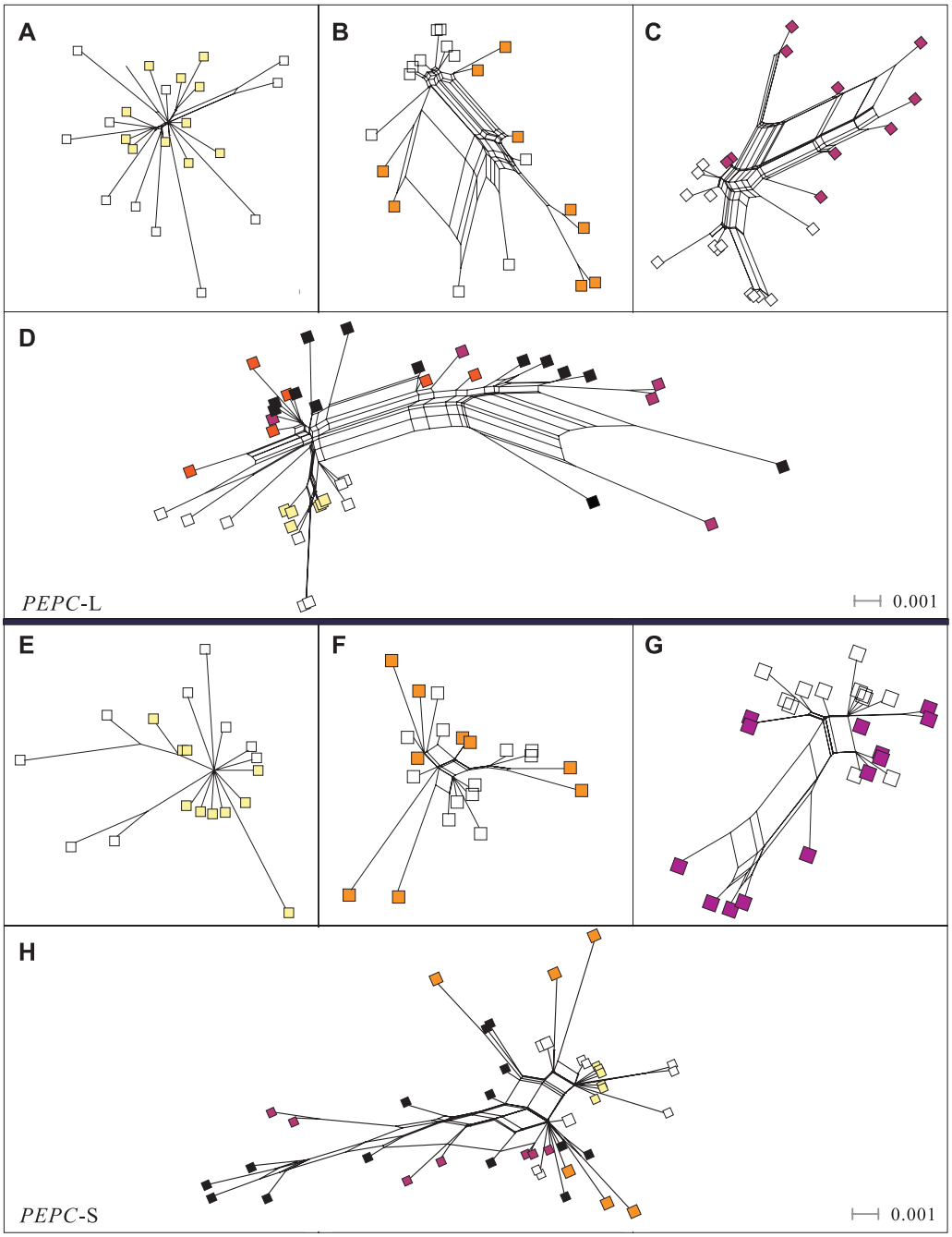

Figure 3. Neighbor-net splits graphs based on the *PEPC*-L (A–D) and *PEPC*-S (E–H) sequences showing limited reticulation (starbursts) in (A) and (E), and extensive reticulation elsewhere (B–D, F–H). In (A) and (E), Oregon 2*x* and 3*x* *C. suksdorfii*; in (B) and (F), Oregon 2*x* and Idaho 3*x* *C. suksdorfii*; in (C) and (G), Oregon 2*x* and Montana 4*x* *C. suksdorfii*. In (D) and (H), inclusion of 4*x* *C. douglasii* indicates the way in which this taxon is also implicated in the reticulation seen here.

TESTING ALTERNATIVE TOPOLOGIES

SH tests based on each of the nuclear datasets failed to reject the null hypothesis that 2*x* and 3*x* individuals from Oregon form one monophyletic group ($P > 0.1$; Table 3). These results are consistent with the tree and network analyses (Figs. 1–4), and support the autopolyploid origin of these 3*x* individuals. On the contrary, the likelihood scores between the unconstrained trees and trees constrained to fit the null hypothesis were significantly different for the other two comparisons involving 2*x* and 3*x* from Idaho and 2*x* and 4*x* from Montana ($P < 0.01$). Instead, as suggested by the tree and network findings (Figs. 1–4), these polyploids contain more heterogeneous gene pools with polyphyletic origins.

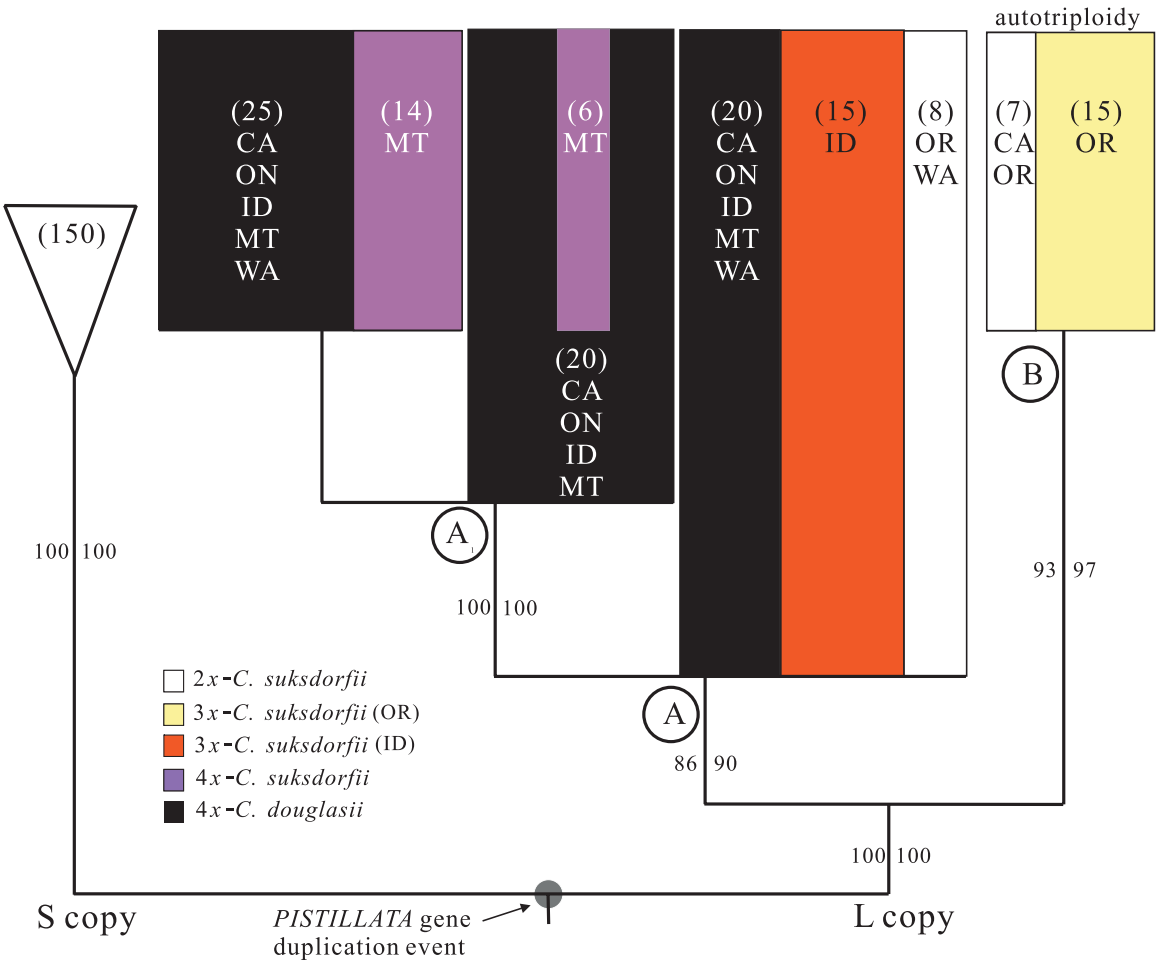

Figure 4. Schematic representation of the strict consensus of MP trees for the S and L copies of *PISTILLATA* (for details, see Fig. S6). Bootstrap (left) and posterior probability (right) values are indicated on branches. Numbers in parentheses represent the number of sequences found within clades. Locality labels are indicated in Table 1. Sequences of the S-copy obtained from all examined individuals (Table 1) are unresolved and are thus represented by a triangle. Two clades (A and B) were identified for the L-copy.

Table 3. Likelihood scores and level of significance of the Shimodaira–Hasegawa tests based on nuclear data. Asterisks denote significant differences in likelihood scores between the unconstrained ML trees and the ML trees constrained to fit the null hypothesis that the groups are monophyletic.

Nuclear data	Groups constrained to be monophyletic	Likelihood score	Differences	*P*-value
PEPC-L	No constraint	−1869.14	-	-
	{2*x* and 3*x* (OR)}	−1855.67	4.46	0.3
	{2*x* and 3*x* (ID)}	−1919.77	59.63	0.002*
	{2*x* and 3*x* (MT)}	−1901.63	41.49	0.006*
PEPC-S	No constraint	−2150.96	-	-
	{2*x* and 3*x* (OR)}	−2156.66	5.70	0.245
	{2*x* and 3*x* (ID)}	−2239.77	88.82	0.004*
	{2*x* and 3*x* (MT)}	−2242.86	91.91	0.003*
PISTILLATA-L	No constraint	−4003.82	-	-
	{2*x* and 3*x* (OR)}	−4017.88	13.98	0.115
	{2*x* and 3*x* (ID)}	−4242.41	230.58	0.0001*
	{2*x* and 3*x* (MT)}	−4409.36	405.54	0.0001*

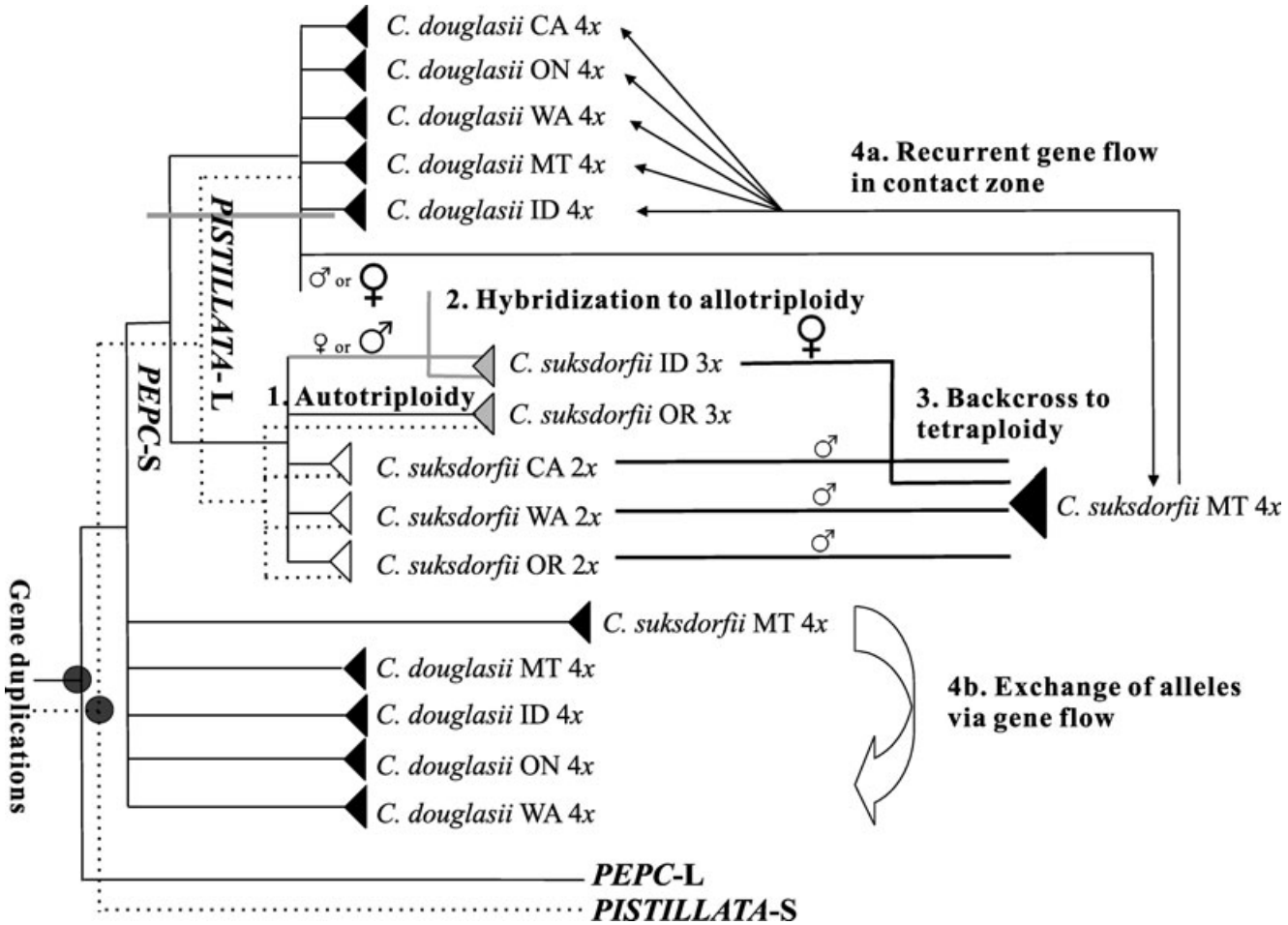

Figure 5. **A summary model synthesized based on the results of the sequence and flow cytometry data, indicating parental lineages and gene flow in the diploid–polyploid complex of *C. suksdorfii* and *C. douglasii*. Solid lines indicate topologies resulting from *PEPC* data whereas dotted lines indicate *PISTILLATA* data. Four routes for polyploid formation are inferred (see Discussion for details). Branches in gray indicate hybridization between 2*x* *C. suksdorfii* and 4*x* *C. douglasii*. Branches in bolded indicate the backcrossing of the allotriploids with their diploid progenitors. Lines with arrows indicate recurrent gene flow between the sympatric 4*x* *C. suksdorfii* and 4*x* *C. douglasii*.**

Discussion

MODELLING POLYPLOID FORMATION

Based on all of our findings, including the evolutionary relationships inferred from the nuclear and chloroplast genes (Figs. 1–4), as well as ploidy-level determinations (Table 1), Figure 5 was synthesized to illustrate the following routes of polyploid formation and reticulate evolution.

Route 1: Autotriploidy

Triploid individuals of *C. suksdorfii* from Oregon were shown to be monophyletic and closely associated only with 2*x* *C. suksdorfii* in the chloroplast and nuclear data (Figs. 1–4). These 3*x* individuals also display a relatively lower haplotype diversity than the others (Table 2), which is to be expected, considering that these triploids are either sterile or less fertile than diploids and tetraploids (Dickinson et al. 1996). Both the nuclear and organelle genomes of the Oregon triploids were inherited only from the 2*x* *C. suksdorfii* with no other taxa involved (Fig. 5). We infer that fertilization of an unreduced egg (2*x* ♀) by a reduced sperm (1*x* ♂), in both cases from diploid parents, is one possible mechanism for this autotriploid formation. In *Crataegus*, the preponderance of evidence supports this scenario, because pollen from diploids typically is reduced (Talent and Dickinson 2007), even though cases of unreduced mate gametes in diploids are seen in other groups (Orjeda et al. 1990; Maceira et al. 1992; Ortiz et al. 1992; Ramsey 2007). Moreover, unreduced female gametes produced by these 3*x* individuals (Dickinson et al. 1996), and gametophytic apomixis generally provide a mechanism for both a degree of reproductive isolation as well as means of reproductive assurance if pollen and/or pollinator services are available.

Route 2: Allotriploidy

Unlike the Oregon triploids, sequences of 3*x* *C. suksdorfii* from Idaho are consistently associated with both 2*x* *C. suksdorfii* and 4*x*

C. douglasii (Figs. 1–4) and reveal a greater amount of haplotype variation (Table 2). The reticulation detected in the data (Figs. 3 and S4) coupled with the rejection of the monophyly (Table 3) points to their allopolyploid origin. Allotriploids could be derived from the fusion of reduced gametes from both 2*x C. suksdorfii* and 4*x C. douglasii* (Fig. 5; Talent and Dickinson 2007). Chloroplast findings suggest that both species have served as the maternal parent, but with the clear preponderance of maternal influence from 4*x C. douglasii* (Figs. 1 and S1). It is unclear whether the bias in the direction of hybridization is caused by differences in the relative abundance of the parental species in sites where these allotriploids were formed, or by environment-dependent selection on the progeny between reciprocal crosses of the two parental species. These factors, either separately or together, were suggested to influence the frequency of a species to mother hybrids within sympatric populations in flowering plants (Tiffin et al. 2001 and references therein).

Route 3: Backcrossing to tetraploidy

Tetraploid *C. suksdorfii* (MT2) and allotriploids (ID) share the same chloroplast haplotypes (I and O; Fig. 1). Sequences of these 4*x* individuals are found in polyphyletic lineages in the nuclear data (Figs. 2–4). One likely pathway for the origin of 4*x C. suksdorfii* is backcrossing between the allotriploid (unreduced 3*x* ♀ gametes) and their 2*x* progenitors (reduced 1*x* ♂ gametes; Fig. 5), given that allotriploids can be nearly as fertile as diploids and tetraploids and are capable of setting seeds (Dickinson et al. 1996). Our results provide additional empirical evidence for the hypothesis of triploid bridge in tetraploid formation (Ramsey and Schemske 1998), as also documented by Husband (2004), Joly and Bruneau (2004), Ayres et al. (2008), and Slotte et al. (2008).

Route 4: Gene flow between sympatric tetraploids

Subsequent to the formation of 4*x C. suksdorfii*, these individuals interbred with 4*x C. douglasii,* as evidenced by the unique associations detected in the chloroplast and nuclear data (e.g., haplotypes J and L in Fig. 1; clade B in Fig. 2A; haplotypes Q and L in Fig. 2B). Individuals of these two taxa occur in sympatry at our Montana site (Table 1) where there is no obvious niche separation between them. In addition, pollen of both taxa is known to be highly stainable at this site (64–82% in *C. suksdorfii* and 86% in *C. douglasii*; Dickinson et al. 1996) so that crossing between these taxa at this ploidy level is possible (Lo et al. 2009a).

SYMPATRIC FORMATION AND ALLOPATRIC ESTABLISHMENT OF POLYPLOIDS

Polyploidy is an important mechanism of sympatric speciation because polyploid descendants are thought to be reproductively isolated from their diploid progenitors (Coyne and Orr 2004; Tate et al. 2005; Ortiz-Barrientos and Rieseberg 2006; Mallet 2007). Our results, as well as those from some other studies (e.g., Ramsey and Schemske 1998; Husband 2004; Slotte et al. 2008) demonstrate that triploid plants (especially allotriploids) may not be completely sterile and can function as seed parents, given gametophytic apomixis, and so can provide a triploid bridge that facilitates gene flow between diploids and at least some tetraploids. Gametophytic apomixis and the production of viable pollen, particularly in *Ranunculus*, some Rosaceae, and some Asteraceae, are thus key components of polyploid speciation (Van Dijk 2003; Joly and Bruneau 2004; Hörandl 2006; Thompson and Whitton 2006; Talent and Dickinson 2007). In addition, physiological differences between polyploids and their diploid progenitors (Stebbins 1971; Ramsey and Schemske 1998), could contribute to allopatric establishment within a species complex. In the case of hawthorns, this will be promoted by their dispersal syndrome and ruderal habit. For instance, *C. suksdorfii* diploids appear to be restricted to lower elevations to the west of the Cascades, whereas polyploids are widespread and often abundant in the Cascades and eastward into the Rocky Mountains (Dickinson et al. 2008). Although our results indicate that polyploids have been derived repeatedly from diploid progenitors, in fact we have no data for the co-occurrence of these diploid and polyploid plants. In the case of *C. suksdorfii* there is also a degree of morphological differentiation (Dickinson et al. 2008) that accompanies the genetic differentiation seen in the microsatellite data (Lo et al. 2009a). To the extent that the establishment of newly formed polyploids may take place more often in allopatry, or are at least found in allopatry with their diploid progenitors, it may be more accurate to see polyploidization as a mechanism leading to allopatric speciation, if geographically isolated cytotypes are recognized as separate species.

Allopatric distribution of cytotypes has been reported in other groups (e.g., Baack 2004; Stuessy et al. 2004; Grundt et al. 2005; Nakagawa 2006) and this could be advantageous for polyploid establishment in two ways. First, the success of neopolyploids in populations of their diploid progenitors is not necessarily automatic. In the absence of uniparental reproduction (e.g., by selfing and apomixis), rare polyploids may suffer from frequency-dependent selection ("minority cytotype exclusion;" Levin 1975; Husband 1998). The ability to colonize areas where parental cytotypes are absent can reduce competition and enhance survivorship. Second, polyploids may be more tolerant than diploids of less-favorable abiotic conditions (Ehrendorfer 1980; Levin 2002). In contrast to our diploid *C. suksdorfii* that occur in mesic lowlands, autotriploid, allotriploid, and tetraploid cytotypes of this taxon are found in colder and/or more xeric environments (Dickinson et al. 2008; E. Y. Y. Lo et al., unpubl. ms.). Autopolyploidy and allopolyploidy could thus both be viewed as processes that facilitate range expansion of diploid progenitors whose polyploid descendants are better adapted to new and less-favorable environments.

THE COMPLEMENTARY ROLES OF TREE AND NETWORK METHODS

It has been long recognized by biologists that there are many instances in natural evolution in which the genetic material is not transferred in a hierarchical way, These include events such as hybrid speciation (Rieseberg et al. 2000; Mallet 2007), horizontal gene transfer (Bergthorsson et al. 2003; Richardson and Palmer 2007), and recombination (Posada et al. 2002; Ruths and Nakhleh 2005). The concept of "bifurcating trees" may oversimplify our view of evolution. For this reason, a variety of methods has been developed based on different strategies to deal with these reticulate phenomena (e.g., Strimmer et al. 2001; Nakhleh et al. 2004; Holland et al. 2005; Huson and Bryant 2006; Morin and Moret 2006). Our study demonstrates that the network and tree approaches play a complementary role for inferring evolutionary history of polyploids at the interface of inter- and intraspecific levels.

With a pool of sequences representing two or more paralogous lineages, tree reconstructions provide a hierarchical set of relationships enabling us to not only identify monophyletic groups but also to position gene duplication events and to infer the number of putative gene copies. This is particularly useful when employing nuclear gene markers in polyploid organisms. Although the tree approach may convey less information than the network when examining closely related taxa, the attempt to search for the optimal tree with a model-based algorithm may reveal a more accurate topology. In addition, alternative topology testing by likelihood methods (e.g., the SH tests; Table 3) provides statistical evaluation of competing evolutionary hypotheses, which cannot be done with the network approach at present.

On the other hand, with a pool of sequences representing orthologous alleles inherited from more than one genome, network constructions provide alternative views (i.e., anastomoses instead of bifurcations) enabling us to infer reticulate relationships, as well as to evaluate the extent of reticulation. For instance, in the absence of reticulation, as in our 2*x* and 3*x* from Oregon (Fig. 3), taxa are united as a well-supported clade in tree analyses (Figs. 1, 2A, and 4). However, when there is a considerable level of reticulation, as detected in the 3*x* from Idaho and 4*x* from Montana (Fig. 3), taxa either form a polytomy or receive weak support, if resolved in the trees (Figs. 1, 2A, and 4). These outcomes corroborate the predictions from simulation studies (e.g., Cassens et al. 2005; Huber and Moulton 2006; Woolley et al. 2008). In these cases, the network approach appears to be more suitable for resolving or visualizing reticulate relationships.

Although networks have been successfully and frequently applied for modelling molecular evolution with gene expression, single nucleotide polymorphism, and protein data (see Morrison 2005 and references therein), this approach is not yet common for inferring reticulation history among organisms in evolutionary studies. One potential reason for this is that the ability of the network approach to accurately estimate the true underlying genealogical relationships has only been examined in a limited number of studies (e.g., Cassens et al. 2005; Woolley et al. 2008). Another reason is that existing network construction software for DNA sequence data mostly employ parsimony (e.g., Clement et al. 2000; Jin et al. 2007; Than et al. 2008) or distance (e.g., Bandelt et al. 1999; Makarenkov 2001; Bryant and Moulton 2004) rather than a model-based method. An extension of a likelihood or Bayesian framework with nucleotide data, which allows model specification or testing of particular hypotheses, would yield accuracy to reflecting genealogical relationships and gain popularity to be used in evolutionary studies.

ACKNOWLEDGMENTS

We thank R. Love, S. Nguyen, and P. Zika for assistance in making plant collections; A. Por, C. Ng, and J. Bull for organizing the vouchers; Nadia Talent for assistance with flow cytometry; M. Kuzmina, K. Choffe, M. Hébert-Lee, and J.-M. Moncalvo for assistance with sequencing; and two anonymous reviewers who generously commented on earlier drafts of this article. Financial support from the Natural Sciences and Engineering Research Council of Canada (grant A3430 to TAD and 326439-06 to SS), the Botany Department of the University of Toronto, and the Department of Natural History of the Royal Ontario Museum is gratefully acknowledged, as is an award from the Royal Ontario Museum Foundation to M. Engstrom and TAD for the purchase of the thermocycler used in part of this work.

LITERATURE CITED

Ayres, D. R., E. Grotkopp, K. Zaremba, C. M. Sloop, M. J. Blum, J. P. Bailey, C. K. Anttila, and D. R. Strong. 2008. Hybridization between invasive *Spartina densiflora* (Poaceae) and native *S. foliosa* in San Francisco Bay, California, USA. Am. J. Bot. 95:713–719.

Baack, E. J. 2004. Cytotype segregation on regional and microgeographic scales in snow buttercups (*Ranunculus adoneus*: Ranunculaceae). Am. J. Bot. 91:1783–1788.

Bailey, C. D., and J. J. Doyle. 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.

Bandelt, J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecies phylogenies. Mol. Biol. Evol. 16:37–48.

Becak, M. L., and W. Becak. 1998. Evolution by polyploidy in Amphibia: new insights. Cytogenet. Cell Genet. 80:28–33.

Bergthorsson, U., K. Adams, B. Thomason, and J. Palmer. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424:197–201.

Bowers, J. E., B. A. Chapman, J. Rong, and A. H. Paterson. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438.

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

Bryant, D., and V. Moulton. 2004. Neighbor-Net: an agglomerative method for the construction of phylogenetic methods. Mol. Biol. Evol. 21:255–265.

Brysting, A. K., B. Oxelman, K. T. Huber, V. Moulton, and C. Brochmann. 2007. Untangling complex histories of genome mergings in high polyploids. Syst. Biol. 56:467–476.

Burton, T. L., and B. C. Husband. 2001. Fecundity and offspring ploidy in matings among diploid, triploid, and tetraploid *Chamerion angustifolium* (Onagraceae): consequences for tetraploid establishment. Heredity 87:573–582.

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

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

Cassens, I., P. Mardulyn, and M. C. Milinkovitch. 2005. Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? Syst. Biol. 54:363–372.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657–1659.

Comber, S. C. L., and C. Smith. 2004. Polyploidy in fishes: patterns and processes. Biol. J. Linn. Soc. 82:431–442.

Cronn, R., M. Cedroni, T. Haselkorn, C. Grover, and J. F. Wendel. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. 104:482–489.

Cronn, R., and J. F. Wendel. 2004. Cryptic trysts, genomic mergers, and plant speciation. New Phytol. 161:133–142.

Cui, L., P. K. Wall, J. H. Leebens-Mack, B. G. Lindsay, D. E. Soltis, J. J. Doyle, J. E. Carlson, K. Arumuganathan, A. Barakat, V. A. Albert, et al. 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res. 16:738–749.

Coyne, J., and H. A. Orr. 2004. Speciation. Sinauer, Sunderland, MA.

Dickinson, T. A., S. Belaoussoff, R. M. Love, and M. Muniyamma. 1996. North American black-fruited hawthorns: I. Variation in floral construction, breeding system correlates, and their possible evolutionary significance in *Crataegus* sect. *Douglasii* Loudon. Fol. Geobot. Phytotax. 31:355–371.

Dickinson, T. A., E. Y. Y. Lo, and N. Talent. 2007. Polyploidy, reproductive biology, and Rosaceae: understanding evolution and making classification. Plant Syst. Evol. 266:59–78.

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

Ehrendorfer, F. 1980. Polyploidy and distribution. Pp. 45–60 *in* W. H. Lewis, ed. Polyploidy—biological relevance, Plenum Press, New York.

Evans, R. C., L. A. Alice, C. S. Campbell, E. A. Kellogg, and T. A. Dickinson. 2000. The granule-bound starch synthase (*GBSSI*) gene in the Rosaceae: multiple loci and phylogenetic utility. Mol. Phylogenet. Evol. 17:388–400.

Evans, R. C., and C. S. Campbell. 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.

Esumi, T., R. Tao, and K. Yonemori. 2005. Isolation of LEAFY and TERMINAL FLOWER 1 homologues from six fruit tree species in the subfamily Maloideae of the Rosaceae. Sexual Plant Reprod. 17:277–287.

Freeling, M., and B. C. Thomas. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. 16:805–814.

Funk, V. A. 1985. Phylogenetic patterns and hybridization. Ann. Missouri Bot. Gard. 72:681–715.

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

Grundt, H. H., R. Obermayer, and L. Borgen. 2005. Ploidal levels in the arctic-alpine polyploid *Draba lactea* (Brassicaceae) and its low-ploid relatives. Bot. J. Linn. Soc. 147:333–347.

Holland, B. R., F. Delsuc, and V. Moulton. 2005. Visualizing conflicting evolutionary hypotheses in large collections of trees: using consensus networks to study the origins of Placentals and Hexapods. Syst. Biol. 54:66–76.

Hörandl, E. 2006. The complex causality of geographical parthenogenesis. New Phytol. 171:525–538.

Husband, B. C. 1998. Contraints on polyploid evolution: a test of the minority cytotype exclusion principle. Proc. R. Soc. Lond. B 267:217–223.

———. 2004. The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy population. Biol. J. Linn. Soc. 82:537–546.

Husband, B. C., B. Ozimec, S. L. Martin, and L. Pollock. 2008. Mating consequences of polyploid evolution in flowering plants: current trends and insights from synthetic polyploids. Int. J Plant Sci. 169:195–206.

Huber, K. T., B. Oxelman, M. Lott, and V. Moulton. 2006. Reconstructing the evolutionary history of polyploids from multilabeled Trees. Mol. Biol. Evol. 23:1784–1791.

Huber, K. T., and V. Moulton. 2006. Phylogenetic networks from multi-labelled trees. J. Math. Biol. 52:613–632.

Huson, D. 1998. SplitsTree: a program for analyzing and visualizing evolutionary data. Bioinformatics 14:68–73.

Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.

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

Jin, G., L. Nakhleh, S. Snir, and T. Tuller. 2007. Inferring phylogenetic networks by the maximum parsimony criterion: a case study. Mol. Biol. Evol. 24:324–337.

Joly, S., and A. Bruneau. 2004. Evolution of triploidy in *Apios americana* (Leguminosae) revealed by genealogical analysis of the histone H3-D gene. Evolution 58:284–295.

Judo, M., A. Wedel, and C. Wilson. 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. 26:1819–1825.

Levin, D. A. 1975. Minority cytotype exclusion in local plant populations. Taxon 24:35–43.

———. 2002. The role of chromosomal change in plant evolution. Oxford Univ. Press, New York.

Lo, E. Y. Y., S. Stefanović, and T. A. Dickinson. 2009a. Population genetic structure of diploid sexual and polyploid apomictic hawthorns (*Crataegus*; Rosaceae) in the Pacific Northwest. Mol. Ecol. 18:1145–1160.

Lo, E. Y. Y., S. Stefanović, K. I. Christensen, and T. A. Dickinson. 2009b. 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. Mol. Phylogenet. Evol. 51:157–168.

Linder, C. R., and L. H. Rieseberg. 2004. Reconstructing patterns of reticulate evolution in plants. Am. J. Bot. 91:1700–1708.

Mable, B. K. 2004. Polyploidy and self-compatibility: is there an association? New Phytol. 162:803–811.

Maceira, N. O., A. A. D. Haan, R. Lumaret, M. Billon, and J. Delay. 1992. Production of 2N gametes in diploid subspecies of *Dactylis glomerata*: occurrence and frequency of 2N pollen. Ann. Bot. 69:335–343.

Maddison, W. P., and D. R. Maddison. 1992. MacClade, version 3.01. Sinauer Associates, Sunderland, MA.

Mallet, J. 2007. Hybrid speciation. Nature 446:279–283.

Makarenkov, V. 2001. T-REX: reconstructing and visualizing phylogenetic trees and reticulation networks. Bioinformatics 17:664–668.

Marhold, K., and J. Lihová. 2006. Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. Plant Syst. Evol. 259:143–174.

Martin, D., and E. Rybicki. 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16:562–563.

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

McDade, L. A. 1992. Hybrids and phylogenetic systematics: the impact of hybrids on cladistic analysis. Evolution 46:1329–1346.

Meyers, L. A., and D. A. Levin. 2006. On the abundance of polyploids in flowering plants. Evolution 60:1198–1206.

Milne, I., F. Wright, G. Rowe, D. F. Marshall, D. Husmeier, and G. McGuire. 2004. TOPALi: software for automatic identification of recombinant sequences within DNA sequence alignments. Bioinformatics 20:1806–1807.

Morin, M. M., and B. M. E. Moret. 2006. NETGEN: generating phylogenetic networks with diploid hybrids. Bioinformatics 22:1921–1923.

Morrison, D. 2005. Networks in phylogenetic analysis: new tools for population biology. Int. J. Parasitol. 35:567–582.

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

Nakagawa, M. 2006. Ploidy, geographical distribution, and morphological differentiation of *Parasenecio auriculata* (Senecioneae; Asteraceae) in Japan. J. Plant Res. 119:51–61.

Nakhleh, T., T. Warnow, and C. R. Linder. 2004. Reconstructing reticulate evolution in species—theory and practice. Pp. 337–346 *in* P. E. Bourne and D. Gusfield, eds, Proceedings of the Eighth Annual International Conference on Research in Computational Molecular Biology, ACM Press, New York.

Orjeda, G., R. Freyre, and M. Iwanaga. 1990. Production of 2*n* pollen in diploid *Ipomoea trifida*, a putative wild ancestor of sweet potato. J. Hered. 81:462–467.

Ortiz-Barrientos, D., and L. H. Rieseberg. 2006. Speciation: splitting when together. Heredity 97:2–3.

Ortiz, R., L. P. Bruederle, T. Laverty, and N. Vorsa. 1992. The origin of polyploids via 2*n* gametes in *Vaccinium* section *Cyanococcus*. Euphytica 61:241–246.

Otto, S. P. 2007. The evolutionary consequences of polyploidy. Cell 131:452–462.

Otto, S. P., and J. Whitton. 2000. Annual Review of Genetics: Polyploid incidence and evolution. Ann. Rev. Genet. 34:401–437.

Oxelman, B., and B. Bremer. 2000. Discovery of paralogous nuclear gene sequences coding for the second-largest subunit of RNA polymerase II (*RPB2*) and their phylogenetic utility in Gentianales of the Asterids. Mol. Biol. Evol. 17:1131–1145.

Phipps, J. B., and R. J. O' Kennon. 1998. Three new species of *Crataegus* (Rosaceae) from western North America: *C. okennonii*, *C. okanagenensis*, and *C. phippsii*. Sida 18:169–191.

———. 2002. New taxa of *Crataegus* (Rosaceae) from the northern Okanagan-southwestern Shuswap diversity center. Sida 20:115–144.

Popp, M., P. Erixon, F. Eggens, and B. Oxelman. 2005. Origin and evolution of a circumpolar polyploid species complex in *Silene* (Caryophyllaceae) inferred from low copy nuclear RNA polymerase introns, rDNA, and chloroplast DNA. Syst. Bot. 30:302–313.

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

———. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA 98:13757–13762.

Posada, D., K. Crandall, and E. Holmes. 2002. Recombination in Evolutionary Genomics. Annu. Rev. Genet. 36:75–97.

Ptacek, M. B., H. C. Gerhardt, and R. D. Sage. 1994. Speciation by polyploidy in treefrogs: multiple origins of the tetraploid, *Hyla versicolor*. Evolution 48:898–908.

Rambaut, A. 2002. Se-Al sequence alignment editor v2.0a11. Univ. of Oxford, Oxford.

Rambaut, A., and A. J. Drummond. 2007. Tracer v1.4. Available at http://beast.bio.ed.ac.uk/Tracer.

Ramsey, J. 2007. Unreduced gametes and neopolyploids in natural populations of *Achillea borealis* (Asteraceae). Heredity 98:143–150.

Ramsey, J., and D. W. Schemske. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Ann. Rev. Ecol. Syst. 29:467–501.

Richardson, A. O., and J. D. Palmer. 2007. Horizontal gene transfer in plants. J. Exp. Bot. 58:1–9.

Rieseberg, L., S. Baird, and K. Gardner. 2000. Hybridization, introgression, and linkage evolution. Pl. Mol. Biol. 42:205–224.

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

Ruths, D., and L. Nakhleh. 2005. Recombination and phylogeny: effects and detection. Int. J. Bioinform. Res. Appl. 1:202–212.

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

Smedmark, J. E. E., T. Eriksson, C. E. Rodger, and C. S. Christopher. 2003. Ancient allopolyploid speciation in Geinae (Rosaceae): Evidence from nuclear Granule-Bound Starch Synthase (GBSSI) gene sequences. Syst. Biol. 52:374–385.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based Soltis, D. E., P. S. Soltis, and J. A. Tate. 2004. Advances in the study of polyploidy since plant speciation. New Phytol. 161:173–191.

Slotte, T., H. Huang, M. Lascoux, and A. Ceplitis. 2008. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). Mol. Biol. Evol. 25:1472–1481.

Soltis, D. E., P. S. Soltis, and J. A. Tate. 2004. Advances in the study of polyploidy since plant speciation. New Phytol. 161:173–191.

Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson, B. C. Husband, and W. S. Judd. 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56:13–30.

Soltis, D. E., and J. G. Burleigh. 2009. Surviving the K-T mass extinction: New perspectives of polyploidization in angiosperms. Proc. Natl. Acad. Sci. USA 106:5455–5456.

Stamatakis, A., T. Ludwig, and H. Meier. 2005. RAxML-III: A fast program for maximum likelihood based inference of larger phylogenetic trees. Bioinformatics 21:456–463.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML Servers. Syst. Biol. 57:758–771.

Stebbins, G. L. 1971. Chromosomal evolution in higher plants. Edward Publishers, Ltd, London.

Strimmer, K., C. Wiuf, and V. Moulton. 2001. Recombination analysis using directed graphical models. Mol. Biol. Evol. 18:97–99.

Stuessy, T. F., H. Weiss-Schneeweiss, and D. J. Keil. 2004. Diploid and polyploid cytotype distribution in *Melampodium cinereum* and *M. leucanthum* (Asteraceae, Heliantheae). Am. J. Bot. 91:889–898.

Swofford, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, MA.

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

———. 2007. The potential for ploidy level increases and decreases in *Crataegus* (Rosaceae, Spiraeoideae, tribe Pyreae). Can. J. Bot. 85:570–584.

Tate, J. A., D. E. Soltis, and P. S. Soltis. 2005. Polyploidy in plants. Pp. 371–426 *in* T. R. Gregory ed., The evolution of the genome, Elsevier Academic Press, San Diego, CA.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132:619–633.

Tiffin, P., M. S. Olson, and L. C. Moyle. 2001. Asymmetrical crossing barrier in angiosperms. Proc. R. Soc. Lond. B268:861–867.

Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9:322–338.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24:4876–4882.

Thompson, J. N., S. L. Nuismer, and K. Merg. 2004. Plant polyploidy and the evolutionary ecology of plant/animal interactions. Biol. J. Linn. Soc. 82:511–519.

Thompson, S. L., and J. Whitton. 2006. Patterns of recurrent evolution and geographic parthenogenesis within apomictic polyploid Easter daises (*Townsendia hookeri*). Mol. Ecol. 15:3389–3400.

Vamosi, J. C., and T. A. Dickinson. 2006. Polyploidy and diversification: A phylogenetic investigation in Rosaceae. Int. J. Plant Sci. 167:349–358.

Van Dijk, P. J. 2003. Ecological and evolutionary opportunities of apomixis: insights from *Taraxacum* and *Chondrilla*. Philos. Trans. R Soc. Lond. B 358:1113–1121.

Vriesendorp, B., and F. T. Bakker. 2005. Reconstructing patterns of reticulate evolution in angiosperms: what can we do? Taxon 54:593–604.

Wada, M., Q. Cao, N. Kotoda, J. Soejima, and T. Masuda. 2002. Apple has two orthologues of FLORICAULA/LEAFY involved in flowering. Plant Mol. Biol. 49:567–577.

Whitton, J., C. J. Sears, E. J. Baack, and S. P. Otto. 2008. The dynamic nature of apomixis in the angiosperms. Int. J. Pl. Sci. 169:169–182.

Woolley, S. M., D. Posada, and K. Crandall. 2008. A comparison of phylogenetic network methods using computer simulation. PLoS ONE 3: 1–12.

Associate Editor: J. Huelsenbeck

Supporting Information

The following supporting information is available for this article:

Figure S1. Bayesian and ML trees of the chloroplast data based on the TVM + I model.

Figure S2. Bayesian and ML trees of *PEPC* data based on the GTR + Γ+ I model.

Figure S3. Bayesian and ML trees of *PISTILLATA* data based on the GTR + Γmodel.

Figure S4. Neighbor-net splits graphs based on the *PISTILLATA* (A–D) sequences.

Figure S5. Strict consensus of equally parsimonious trees for the L and S copies of *PEPC* sequence data that resolved as two distinct clades.

Figure S6. Strict consensus of equally parsimonious trees for the S and L copies of *PISTILLATA* sequence data.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.