Molecular Ecology (2009) **18**, 1145–1160 doi: 10.1111/j.1365-294X.2009.04091.x

Population genetic structure of diploid sexual and polyploid apomictic hawthorns (*Crataegus*; Rosaceae) in the Pacific Northwest

EUGENIA Y. Y. LO,*† SAŠA STEFANOVIĆ† and TIMOTHY A. DICKINSON*‡
*Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON, Canada M5S 3B2, †Department of Biology, University of Toronto at Mississauga, 3359 Mississauga Rd. N Mississauga, ON, Canada L5L 1C6, ‡Green Plant Herbarium (TRT), Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, ON, Canada M5S 2C6

Abstract

Polyploidy and gametophytic apomixis are two important and associated processes in plants. Many hawthorn species are polyploids and can reproduce both sexually and apomictically. However, the population genetic structure of these species is poorly understood. *Crataegus douglasii* is represented exclusively by self-compatible tetraploid pseudogamous apomicts across North America, whereas *Crataegus suksdorfii* found in the Pacific Northwest is known to include self-incompatible diploid sexuals as well as polyploid apomicts. We compare population structure and genetic variability in these two closely related taxa using microsatellite and chloroplast sequence markers. Using 13 microsatellite loci located on four linkage groups, 251 alleles were detected in 239 individuals sampled from 15 localities. Within-population multilocus genotypic variation and molecular diversity are greatest in diploid sexuals and lowest in triploid apomicts. Apart from the isolation of eastern North American populations of *C. douglasii*, there is little evidence of isolation by distance in this taxon. Genetic diversity in western populations of *C. douglasii* suggests that gene flow is frequent, and that colonization and establishment are often successful. In contrast, local populations of *C. suksdorfii* are more markedly differentiated. Gene flow appears to be limited primarily by distance in diploids and by apomixis and self-compatibility in polyploids. We infer that apomixis and reproductive barriers between cytotypes are factors that reduce the frequency of gene flow among populations, and may ultimately lead to allopatric speciation in *C. suksdorfii*. Our findings shed light on evolution in woody plants that show heterogeneous ploidy levels and reproductive systems.

Keywords: *Crataegus*, gametophytic apomixis, gene flow, genotypic diversity, microsatellite, polyploids

Received 5 April 2008; revision received 7 December 2008; accepted 18 December 2008

Introduction

Apomixis, the asexual formation of seeds, is known to occur in over 400 angiosperm genera (Asker & Jerling 1992; Calzada *et al*. 1996; Carman 1997; Whitton *et al*. 2008). There are two major types of apomixis, adventitious embryony and gametophytic apomixis that differ in the way embryos are formed (Nogler 1984; Asker & Jerling 1992; Koltunow & Grossniklaus 2003). The formation of unreduced megagametophytes, and hence of unreduced female gametes, is known as gametophytic apomixis. In gametophytic apomixis, an unreduced egg develops parthenogenetically into an embryo, either autonomously (without pollination) or in a pollination-dependent manner in which fertilization of the central cell nucleus (pseudogamy) is required in order for endosperm to develop successfully (Nogler 1984). Gametophytic apomixis has been reported in *ca*. 140 angiosperm genera (Carman 1997) and this phenomenon is almost always associated with polyploidy and with hybridization (Carman 1997; Whitton *et al*. 2008). Recent molecular investigations of polyploid apomicts

Correspondence: Eugenia Y. Y. Lo, Fax: 416-586-7921;
E-mail: eugenia.lo@yale.edu.

have mainly concentrated on herbaceous plants in the Rosaceae, Poaceae, and Asteraceae (e.g. Menken *et al*. 1995; Noyes & Soltis 1996; Nybom 1998; Esselman *et al*. 1999; Van der Hulst *et al*. 2000; Garnier *et al*. 2002). Where such studies have been carried out on woody Rosaceae, they have been concerned primarily with species relationships (e.g. Bartish *et al*. 2001; Robertson *et al*. 2004; Joly *et al*. 2006; Chester *et al*. 2007). Relatively few attempts have been made to evaluate and compare population genetic structure between sexual and apomictic woody plants, in which growth habit and generation time may influence the pattern and level of genetic variation (Hamrick & Godt 1990; Hamrick *et al*. 1992; Campbell *et al*. 1999; Persson-Hovmalm *et al*. 2004).

One predicted consequence of asexual reproduction is the formation of more genetically uniform local populations than is the case with sexual reproduction. However, these local populations are likely to be differentiated from each other by stochastic colonization events (Dickinson & Campbell 1991; Hamrick *et al*. 1992; Starfinger & Stocklin 1996; McLellan *et al*. 1997). Moreover, polyploidy, mutations, and gene flow via pollen and seed dispersal are suggested as mechanisms that counterbalance the loss of variation through clonal reproduction in agamospermous plants (for reviews see: Ellstrand & Roose 1987; Widén *et al*. 1994; Gornall 1999; Hörandl & Paun 2007; references therein). In addition, gene flow between sympatric cytotypes of *Cerastium* (Hagen *et al*. 2002), *Paspalum* (Daurelio *et al*. 2004), *Sorbus* (Robertson *et al*. 2004; Chester *et al*. 2007), and *Taraxacum* (Meirmans *et al*. 2003) may also contribute to the diversification of agamic complexes (Hörandl & Paun 2007). Even though taken together these processes argue against the 'dead-end' interpretation of apomixis (Stebbins 1950; Darlington 1958) and point to the evolutionary potential of agamic complexes, they are not broadly studied in angiosperms. Data are particularly sparse for the woody groups.

Crataegus (hawthorn) is one of the largest genera in the predominantly woody Rosaceae tribe Pyreae (Campbell *et al*. 2007). In North America, over 60 of the approximately 100 species include polyploid individuals (Talent & Dickinson 2005). Both gametophytic apomixis and hybridization are known or have been inferred (Dickinson *et al*. 2007), so that many groups of hawthorns can thus be thought of as agamic complexes (Stebbins 1950; Dickinson *et al*. 2008). In western North America, series *Douglasianae* occupies the Pacific Northwest with outliers as far east as the upper Great Lakes basin (Dickinson *et al*. 1996; Fig. 2 in Dickinson *et al*. 2008). The two best-known species, *Crataegus douglasii sensu* lato and *Crataegus suksdorfii*, both comprise tetraploids, but *C. suksdorfii* also includes diploids and triploids (Dickinson *et al*. 1996; Talent & Dickinson 2005). Triploids of *C. suksdorfii* appear to have originated both from crosses between conspecific individuals and from crosses between different species, whereas tetraploids are likely to be products of backcrossing between allotriploid offspring and their diploid parents (Talent & Dickinson 2007b; Lo 2008, submitted).

As in other *Crataegus* species, diploid individuals of *C. suksdorfii* have been shown to reproduce sexually, whereas polyploid individuals of *C. suksdorfii* and *C. douglasii* reproduce predominantly by pseudogamous gametophytic apomixis (Dickinson *et al*. 1996; Talent & Dickinson 2007a). However, it remains unclear to what extent such reproductive systems have influenced genetic variation in *C. suksdorfii* and *C. douglasii* at the population level, and whether the gene flow occurring between sympatric cytotypes of the two species has affected local genetic diversity. The overall goal of this paper is, thus, to answer the following questions: (i) how is genetic variation partitioned within and between ploidy levels and between local populations of *C. suksdorfii* and *C. douglasii*? (ii) what does this tell us about the frequency of gene flow within and between these populations? (iii) what do these results imply about the evolutionary potential of polyploid apomicts?

Materials and methods

Plant materials

Our sampling localities (Fig. 1; Table 1) encompassed almost the entire ranges of *Crataegus suksdorfii* and *Crataegus douglasii* in the Pacific Northwest (for the entire North American distribution, see Fig. 2 in Dickinson *et al*. 2008). In large populations, studied trees were chosen by having two collectors using an 'ignorant person' strategy (Ward 1974) to avoid biased sampling. The number of trees collected depended on the population size with a target of at least 15 individuals. For each individual, unexpanded leaves were collected and stored in silica gel and used for DNA extraction. Mature leaves and fruits were collected in the fall and used for ploidy level determinations using a FACSCalibur flow cytometer (Becton-Dickinson), following the protocols of Talent & Dickinson (2005, 2007a). In total, 239 trees representing 16 localities in California, the Pacific Northwest, and Ontario were included, of which 125 tetraploid individuals were *C. douglasii* (including its segregates) and 114 individuals were identified as *C. suksdorfii* (52 diploid, 41 triploid, and 21 tetraploid). Species identification was mainly based on leaf, thorn, and flower or fruit characters such as stamen number and fruit colour (Dickinson *et al*. 2008). Sample vouchers are deposited in the Green Plant Herbarium of the Royal Ontario Museum (TRT). Nuclear DNA content of leaves and seeds (embryo and endosperm values) of all individuals examined are reported in Lo (2008). The ploidy level and reproductive system of individuals studied here are summarized

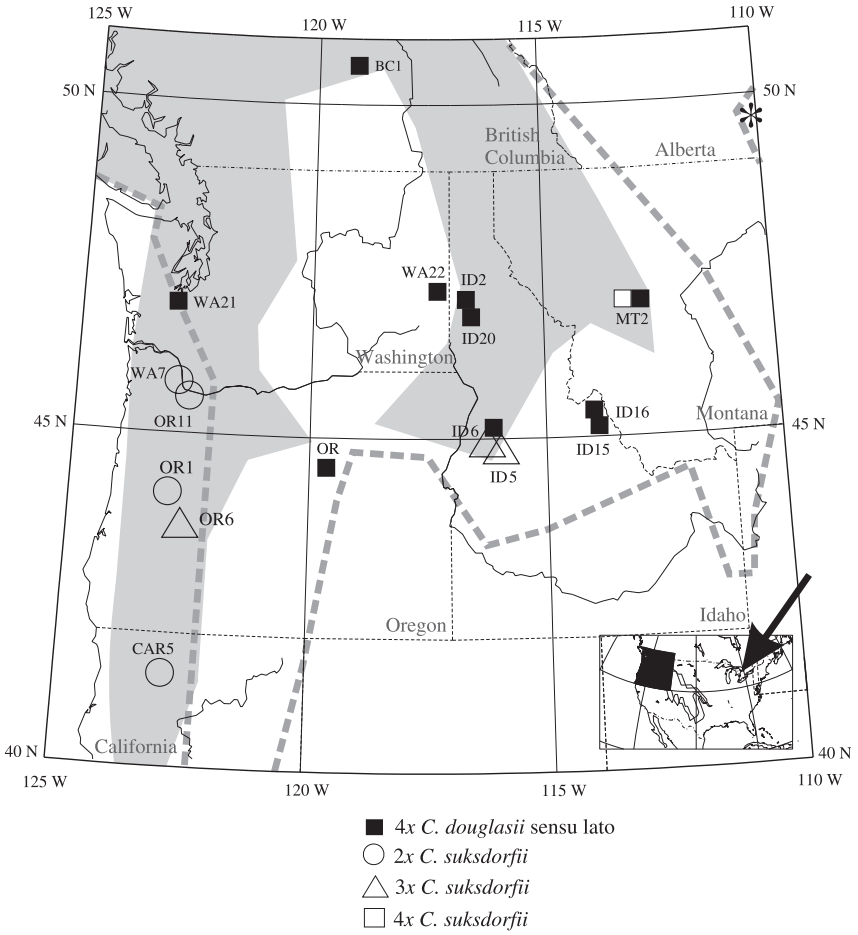

Fig. 1 Approximate distribution of *Crataegus suksdorfii* (shaded) and *Crataegus douglasii sensu* lato (dashed outline) in the Pacific Northwest, showing sampling sites used in the present study (Table 1; disjunct population from Big Bay, Ontario, Canada indicated by arrow in inset): 2*x C. suksdorfii* (○); 3*x C. suksdorfii* (△); 4*x C. suksdorfii* (□); and 4*x C. douglasii sensu* lato (■). Distribution data from Dickinson *et al*. (2008) and Oregon Flora Project (2007). Scale approximately 1:8 000 000; inset shows area covered by detail map on continental scale with solid shading. Disjunct occurrence of *C. douglasii* in the Cypress Hills in southeastern Alberta and adjacent Saskatchewan is indicated by the asterisk at upper right (Fig. 2 in Dickinson *et al*. 2008).

for each sampling locality (Table 1). The distribution map (Fig. 1) was prepared using MicroCAM version 2.05 (Loomer 2003).

DNA extraction and chloroplast sequence analyses

DNA was extracted from leaf tissues according to the modified protocol of Tsumura *et al*. (1995). In order to compare the amount of chloroplast sequence variation within populations with that based on nuclear microsatellite markers, a subset of 132 individuals from 10 of the localities (indicated by asterisks in Table 1), chosen to fully represent the geographical range of *C. douglasii* and *C. suksdorfii*, were sequenced for psbA-trnH and trnH-rpl2, two hypervariable chloroplast intergenic regions. Primers, polymerase chain reaction (PCR) conditions, and sequence protocols followed those described in Lo *et al*. (2007). Because these two regions are linked in the plastid genome, they were combined and treated as a single marker for analyses. The haplotype diversity (H_d) and nucleotide diversity (π) within each locality of *C. suksdorfii* and *C. douglasii* were estimated using DnaSP version 4.10.1 (Rozas *et al*. 2003) based on equations of Nei (1987).

Microsatellite markers

In *Malus domestica*, Liebhard *et al*. (2002) developed and characterized a total of 140 microsatellites. These primers were designed on conserved flanking regions of tandem repeats and the complete procedure of primer development was described in Gianfranceschi *et al*. (1998). Because not all primers developed in *Malus* are transferable to *Crataegus* (Liebhard *et al*. 2002) and because it was too costly to test all the primers, we selected only those loci from linkage groups (LG) that have been identified to contain more than six simple sequence repeat (SSR) regions (Liebhard *et al*. 2002). On average, two loci were selected and tested from a total of 10 LGs (F2, F4, F5, F9, F10, F11, F12, F14, F16, and F17), but up to four loci from LGs F12 and F14 were tested because they were shown to be transferable to other maloid genera (Liebhard *et al*. 2002). Thus, 23 SSR loci representing 10 chromosomal LGs of *Malus domestica* (Liebhard *et al*. 2002) were tested in our preliminary primer screening. Out of those, 13 loci found on at least four LGs (F4, F10, F12, and F14; Table 2) were shown to be transferable to *Crataegus*. These LGs contain perfect dinucleotide tandem repeats in the examined individuals, with the exception of CH05D11,

Table 1 Locality, nuclear DNA content (pg/2C), habitat, and population size (*N*) of *Crataegus douglasii sensu* lato and *Crataegus suksdorfii* included in the present study. Number of unvouchered samples is indicated in parentheses

Species	Mean pg/2C	Ploidy level	*N*	Label	Longitude	Latitude	Elevation (m)	State/Province; County; Locality	Habitat	RS
C. douglasii sensu lato										
	3.02 ± 0.11	4*x*	29 (14)	ON20*	44.80	81.00	225	Ontario; Grey; Colpoys Range	Abandoned farmland	A
	2.96 ± 0.12	4*x*	11 (8)	ID20	46.52	116.73	280	Idaho; Nez Perce; Little Potlatch Creek	Roadside	A
	2.85 ± 0.07	4*x*	12 (3)	ID6*	44.99	116.19	1420	Idaho; Adams; Last Chance Campground,	Along creek	A
	2.93 ± 0.17	4*x*	14 (2)	ID2*	46.77	116.45	811	Idaho; Latah; Little Boulder Creek	Along creek	A
	2.94 ± 0.20	4*x*	10 (4)	ID15*	44.97	113.94	1292	Idaho; Lemhi	Roadside	A
	3.01 ± 0.11	4*x*	15 (1)	MT2*	47.07	112.91	1356	Montana; Powell; Kleinschmidt Flat	Along creek and adjacent slope	A
	2.96 ± 0.14	4*x*	5 (3)	WA22	46.85	117.34	666	Washington; Whitman; South of Colfax[1]	Roadside	A
	3.01 ± 0.13	4*x*	25 (9)	WA21*	46.84	122.98	64	Washington; Thurston; Mound Prairie[2]	Hawthorn colonization	A
	3.01 ± 0.21	4*x*	3	BC1	50.51	119.10	—	British Columbia; Spallumcheen	Roadside	U
	3.20 ± 0.21	4*x*	2	OR	44.58	119.64	655	Oregon; Wheeler; Fossil	Roadside	U
C. suksdorfii sensu lato										
	1.35 ± 0.14	2*x*	8 (1)	CA5*	41.4	122.84	871	California; Siskiyou; Fay Lane	Roadside	S
	1.52 ± 0.17	2*x*	13 (3)	OR1*	44.33	123.12	88	Oregon; Linn; Cogswell Foster Reserve	Hawthorn colonization	S
	1.59 ± 0.07	2*x*	19 (8)	OR11*	45.73	122.77	10	Oregon; Columbia; Sauvie Island	Sandy beach on Columbia River	S
	1.43 ± 0.09	2*x*	12 (7)	WA7	45.83	122.76	15	Washington; Clark.	Roadside	S
	2.24 ± 0.06	3*x*	20	OR6*	43.77	122.62	1295	Oregon; Lane; Patterson Mountain Prairie	Open wet area surrounded by forest	U
	2.17 ± 0.10	3*x*	21 (5)	ID6*	44.99	116.19	1420	Idaho; Adams; Last Chance Campground	Along creek	A
	2.25 ± 0.15	3*x*	2	ID5	45.00	116.06	1524	Idaho; Valley; North Beach, Payette Lake	Sandy delta	A
	2.95 ± 0.07	4*x*	21 (6)	MT2*	47.07	112.91	1356	Montana; Powell; Kleinschmidt Flat	Along creek	A

Reproductive system (RS) is based on flow cytometry determinations of seeds reported in Lo (2008). 'A' denotes apomixis, 'S' denotes sexual and 'U' denotes not determined. Superscripts 1 and 2 denote sites that contain morphological segregates of *C. douglasii* (described as *Crataegus okennonii* and *Crataegus castlegarensis* according to Phipps & O'Kennon 1998). Locality labels marked with asterisks represent samples included in chloroplast sequencing (see Materials and methods; Table 3).

Table 2 Nucleotide sequences of the selected microsatellite primers from four linkage groups (LG) used in the present study. These primers are designed on the conserved SSR-flanking regions of *Malus xdomestica* and are transferable to *Crataegus* and other Maloideae species (Liebhard *et al*. 2002). Information of each locus is described

Locus	Forward primer sequence (5′– 3′)	Reverse primer sequence (5′– 3′)	Total alleles	*D*	*S*	Size (bp)	Maps on LG no.
CH01D03	CCGCTTGGCAATGACTCCTC	ACCCTGAAGCCATGAGGGC	7	4	6	125–149	4
CH01F02	ACCACATTAGAGCAGTTGAGG	CTGGTTTGTTTTCCTCCAGC	25	17	23	149–201	12
CH01F07	CCCTACACAGTTTCTCAACCC	CGTTTTTGGAGCGTAGGAAC	32	17	28	149–263	10
CH03A02	TTGTGGACGTTCTGTGTTGG	CAAGTTCAACAGCTCAAGATGA	17	15	14	122–162	14
CH03C02	TCACTATTTACGGGATCAAGCA	GTGCAGAGTCTTTGACAAGGC	23	17	20	98–148	unclear
CH03D08	CATCAGTCTCTTGCACTGGAAA	TAGGGCTAGGGAGAGATGATGA	16	10	11	120–172	14
CH04F06	GGCTCAGAGTACTTGCAGAGG	ATCCTTAAGCGCTCTCCACA	14	14	7	146–176	14
CH04G04	AGTGGATGATGAGGATGAGG	GCTAGTTGCACCAAGTTCACA	18	12	15	146–186	12
CH05D03	TACCTGAAAGAGGAAGCCCT	TCATTCCTTCTCACATCCACT	15	13	13	141–173	14
CH05D04	ACTTGTGAGCCGTGAGAGGT	TCCGAAGGTATGCTTCGATT	16	10	12	138–174	12
CH05D11	CACAACCTGATATCCGGGAC	GAGAAGGTCGTACATTCCTCAA	13	11	8	161–193	12
CH05G07	CCCAAGCAATATAGTGAATCTCAA	TTCATCTCCTGCTGCAAATAAC	21	20	16	142–192	12 and 14
CH05G11	GCAAACCAACCTCTGGTGAT	AAACTGTTCCAACGACGCTA	34	27	26	173–245	14

D, number of alleles observed in *Crataegus douglasii*; *S*, number of alleles observed in *Crataegus suksdorfii*.

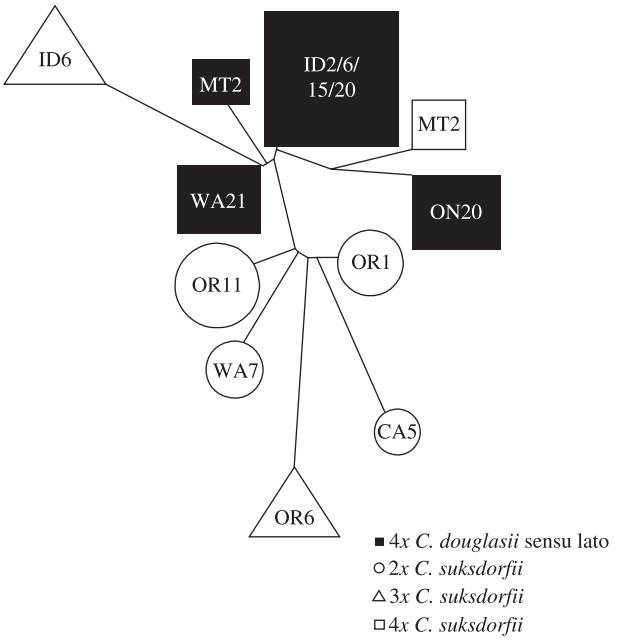

Fig. 2 Unrooted dendrogram based on D_S distances (estimates of *F*-statistics under the IAM; Nei 1978), showing relatedness among sites of *Crataegus douglasii* and *Crataegus suksdorfii* (Table 1). Sites of *C. douglasii* with less than five individuals (e.g. OR, BC1, and WA22) are not included in the analyses. Sizes of the symbols are proportional to the number of individuals sampled from each site.

CH05G07, and CH05G11 that contain imperfect dinucleotide repeats (Liebhard *et al*. 2002). The remaining 10 screened loci (CH01H01, CH01H02, CH02B10, CH02C06, CH02H11, CH03A04, CH04A12, CH04D07, CH04F10, and CH05F03 located on LGs F2, F4, F5, F9, F11, F16, and F17) were proved to be hard to amplify or did not contain tandem repeats in our samples. PCR amplifications were performed in a 15-μL volume containing ~20 ng of genomic DNA, 1.5 μL 10× PCR buffer (Fermentas), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1 U of *Taq* polymerase (Fermentas), and 0.5 μM each of the forward and reverse primers. All forward primers were end-labelled with the fluorescent dyes FAM or HEX. Because the same annealing temperature was used for all SSR primers, two primer pairs were combined in PCR amplifications (multiplexing). We followed the PCR conditions of Gianfranceschi *et al*. (1998). PCR products were analysed on an ABI 3700 automatic sequencer (Applied Biosystems). Peaks were scored using the program GeneMapper version 3.5 (Applied Biosystems). To designate genotypic configurations for triploid and tetraploid individuals, we used the microsatellite DNA allele counting-peak ratios method (MAC-PR; Esselink *et al*. 2004) based on quantitative values for microsatellite allele amplification peak areas to assess the dosage of the amplified products (observed alleles) in the samples.

Linkage disequilibrium analyses

To examine whether the 13 SSR loci represent an independent set of markers in the *Crataegus* genome, genotyping linkage disequilibrium (LD) was tested by Fisher's exact test for each pair of loci (a total of 78 possible pairwise combinations) with GenePop version 3.3 (Raymond & Rousset 1995), using the Markov chain method with 100 batches and 10 000 iterations per batch. Because multiple tests were involved, the sequential Bonferroni correction was applied to test for significance (Rice 1989). Due to the computational burden and technical difficulties in

conducting Fisher's exact test with individuals containing more than two alleles, LD tests were performed only on diploid individuals in our sample. Therefore, because the same sets of SSR loci are used in diploid, triploid, and tetraploid individuals, we assumed that results of LD among loci based on diploids could also be applied to polyploids of *C. suksdorfii* and *C. douglasii*.

Analyses of genotypic variation and structure

For each population, genotypic variation was assessed with GenoDive version 2.0b4 (Meirmans & Van Tienderen 2004). We calculated genetic distances using the method of Bruvo *et al*. (2004) that is based on a two-phase mutation model for microsatellites and scaled by ploidy level, and best described our data. The minimal distance class was set as threshold to identify the following: (i) the number of multilocus genotypes (G); (ii) the proportion of distinguishable multilocus genotypes (PD); (iii) Simpson's diversity index (D), also known as Nei's (1987) genetic diversity corrected for sample size, that ranges from zero, where two randomly chosen individuals in a population share a single genotype, to one where they are genetically different; and (iv) genotype evenness (E) that ranges from zero where one or a few genotypes dominate in a population, to one where all genotypes are of equal frequency in a population.

To estimate molecular diversity of the samples, we calculated ANOVA-based global and pairwise *F-statistics* (based on allele identity under infinite allele model; IAM) and *R*-statistics (based on allele size under stepwise-mutation model; SMM) with significant *P* values of two-sided tests obtained after 1000 random permutations of genes, individuals, and populations with SPAGeDi version 1.2 (Hardy & Vekemans 2002). The standard genetic distances F_{ST} (Wright 1965) and D_S (Nei 1978) based on IAM, as well as R_{ST} (Slatkin 1995) and delta μ^2 (Goldstein *et al*. 1995) based on SMM were estimated. SPAGeDi calculates genetic parameters primarily based on polysomic inheritance (e.g. in autopolyploids; Ronfort *et al*. 1998). Therefore, the F_{ST} values of allopolyploids may be overestimated because of the excess of heterozygotes, owing to disomic inheritance. In turn, this could artificially inflate within-population diversity (Obbard *et al*. 2006), even though the dosage of alleles were estimated in our polyploid samples. Hence, F_{ST} values were also estimated with the phenotype-based approach (i.e. alleles were scored as either presence or absence) using GENALEX version 6.0 (Peakall & Smouse 2006). This approach is suggested to be less affected by ploidy level (Obbard *et al*. 2006) and results were compared with the genotype-based estimates by SPAGeDi. The concordance between F_{ST} and R_{ST} estimates, as well as D_S and $\Delta\mu^2$ distance values were tested by Mantel tests (H_0 = matrices are not correlated) to determine whether our data are influenced by different mutation models. Neighbour-joining (NJ) trees based on D_S and $\Delta\mu^2$ were constructed in PHYLIP version 3.66 (Felsenstein 2006) to infer population relationships. The resulting trees were visualized with the TreeView software (Page 1996).

To examine the distribution of genetic variation, analysis of molecular variance (AMOVA; Excoffier *et al*. 1992) was performed by GENALEX (Peakall & Smouse 2006). The codominant SSR data were first converted to a binary data matrix by treating absence as '0' and presence as '1' of a defined allele. It was followed by computing Jaccard coefficient (JC) based on the binary data, which does not consider the shared absence of a character as similarity (Legendre & Legendre 1983), to obtain an unbiased estimate of pairwise genetic distances between individuals. AMOVA analyses were then performed on the JC matrix with significant tests for 10 000 permutations to determine how genetic diversity is partitioned within and between populations, so as to infer the extent of gene flow.

In addition to AMOVA, a model-based Bayesian analysis was also performed to evaluate genetic structuring of individuals on the basis of their genotypes with Structure version 2.2 (Pritchard *et al*. 2000; Pritchard & Wen 2004). This software attempts to identify *K* groups under varying assumptions about Hardy–Weinberg (HW) and linkage equilibria (Pritchard *et al*. 2000). Because of the potential influence of LD to demographic events, Pritchard & Wen (2004) suggested a rough threshold value of the distance (1.0 cM) between two loci below which the pair of loci should not be used. The relative distances between our markers range from 3.4 to 48 cM (Liebhard *et al*. 2002), well above the suggested threshold value. Also, we employed the linkage models developed by Falush *et al*. (2003), as implemented in this version of Structure to provide more accurate estimates of statistical uncertainty when loci are linked. This was carried out by estimating the origin of chromosomal regions based on the relative distances between our markers within individuals before the estimation of the ancestry of the sampled individuals (Falush *et al*. 2003). The number of clusters *K* is determined by simulating a range of *K* values. The posterior probability of each value was then used to detect the modal value of ΔK, a quantity related to the second order rate of change with respect to *K* of the likelihood function (Evanno *et al*. 2005). Structure estimates posterior probabilities using a Markov chain Monte Carlo (MCMC) method and 1000 000 iterations of each chain following the 1 00 000 iteration burn-in period were performed, as recommended by Pritchard *et al*. (2000). Each MCMC chain for each value of *K* (ranging from 1 to 22) was run for 15 times with the linkage model that takes the linkage disequilibrium between loci into account, together with the 'independent allele frequency' option which assumes allelic frequencies in different populations are reasonably different from each other. These parameters allow individuals with ancestries in

Table 3 Diversity measures from chloroplast sequences of *Crataegus suksdorfii* and *Crataegus douglasii* individuals with respect to localities (Table 1) estimated by DnaSP

Species	Cytotype	Locality id	No. of individuals	No. of poly. sites	*K*	$H_d \pm SD$	π (10^{-3}) ± SD (10^{-3})
C. suksdorfii	Diploid	OR1	9	7	2.87	0.933 ± 0.122	5.31 ± 0.98
		OR11	9	6	2.22	0.889 ± 0.091	4.28 ± 1.11
		CA5	5	5	2.40	0.900 ± 0.161	3.95 ± 0.67
				Average	2.50	0.95 ± 0.034	4.61 ± 0.93
	Triploid	OR6	20	7	0.79	0.284 ± 0.128	1.41 ± 0.68
		ID6	16	4	0.70	0.600 ± 0.127	1.24 ± 0.33
				Average	0.74	0.44 ± 0.010	1.33 ± 0.44
	Tetraploid	MT2	13	11	3.03	0.833 ± 0.081	4.25 ± 1.05
C. douglasii	Tetraploid	ON20	9	9	3.67	0.917 ± 0.073	6.50 ± 1.24
		ID2	11	6	2.22	0.945 ± 0.054	3.95 ± 0.50
		ID6	11	2	0.66	0.327 ± 0.153	1.16 ± 0.54
		ID15	7	4	1.14	0.524 ± 0.209	2.03 ± 1.05
		MT2	12	7	1.17	0.682 ± 0.148	2.07 ± 0.66
		WA21	11	8	2.58	0.873 ± 0.071	4.91 ± 1.03
				Average	2.07	0.73 ± 0.045	3.35 ± 0.60

K, average number of nucleotide differences; H_d, haplotype diversity; π, nucleotide diversity; SD, standard deviation.

more than one group to be assigned into one cluster. Individuals are partitioned into multiple groups according to the membership coefficient (Q) that ranges from 0 (lowest affinity to a group) to 1 (highest affinity to a group) across the *K* groups. Individual assignments can vary across runs if there are weak indications of affinity to particular groups. To address such variation, 100 separate MCMC chains were run for the optimal *K* (where ΔK was a maximum) to test for the consistency of membership coefficient. Analyses were conducted on five sets of individuals to infer genetic structure. These included (i) *C. douglasii* and *C. suksdorfii* together; (ii) *C. douglasii* alone; (iii) *C. suksdorfii* alone; (iv) sympatric site (ID06) of 3*x C. suksdorfii* and 4*x C. douglasii*; (5) sympatric site (MT02) of 4*x C. suksdorfii* and 4*x C. douglasii*. The partitioning of clusters was visualized in the program Distruct (Rosenberg 2004).

Isolation-by-distance analyses

The relationships between genetic distances and geographical distances were analysed for *C. douglasii* and *C. suksdorfii*, respectively, using the Isolation By Distance (IBD) version 1.52 software (Bohonak 2002). The IBD software program assesses the significance and evaluates the strength of the relationships between genetic distances (D_S and $\Delta\mu^2$) and the Euclidean geographical distance (estimated from spatial coordinates using the R Package for multivariate and spatial analysis; Casgrain & Legendre 2004) by Mantel tests (10 000 randomizations) and reduced major axis (RMA) regression, respectively. For *C. suksdorfii*, IBD analyses were also performed separately for all, and for just the diploid populations, in order to examine if neighbouring populations are more genetically similar than distant populations in relation to ploidy level.

Results

Chloroplast sequence variation

The psbA-trnH and trnH-rpl2 regions have a total aligned length of 652 bp in which 63 substitutions and six indels were observed. Haplotype diversity (H_d; Table 3) and nucleotide diversity (π; Table 3) of diploid *Crataegus suksdorfii* (sites OR1, OR11, and CA5) are consistently high and these values are comparable with those of tetraploid *C. suksdorfii* (MT2) and most, but not all of the *Crataegus douglasii* sites (e.g. ON20, ID2, and WA21). Relatively low H_d and π values characterized both the triploid *C. suksdorfii* at sites OR6 and ID6, and the tetraploid *C. douglasii* at site ID6, suggesting the presence of homogeneous cytoplasmic gene pools in each of the cytotypes present at these sites (Table 3).

Genotypic variation and linkage disequilibrium

The 13 SSR loci yielded a total of 251 alleles (7–34 per locus) for our sample of 239 individuals (Table 2). Up to two, three, and four alleles were detected respectively in diploid, triploid, and tetraploid individuals in each locus, consistent with their ploidy levels. Multilocus genotypic variability (G, PG, D, and E) is shown to be the highest in diploid but lowest in triploid and a few tetraploid populations (e.g. ON20, ID15; Table 4), consistent with the findings based on chloroplast sequence markers (Table 3).

Table 4 Descriptive statistics of diploid, triploid, and tetraploid populations of *Crataegus suksdorfii* (indicated with superscript 's') and *Crataegus douglasii* (indicated with superscript 'd') based on the 13 microsatellite loci. The overall estimates for each ploidy level are underlined

	C. suksdorfii								*C. douglasii*									
Species	Diploid					Triploid			Tetraploid									
Ploidy level																		
Population	CA5^s	WA7^s	OR11^s	OR1^s	Average	OR6^s	ID6s	Average	MT2^s	ON20^d	MT2^d	ID15^d	ID6^d	ID20^d	ID2^d	WA22^d	WA21^d	Average^d
N	8	12	19	13	52	20	21	41	21	24	21	10	12	11	14	5	25	122
Mean allele no.	4.33	3.92	9.08	8.15	6.37	3.46	3.69	3.58	5.31	4.85	10.92	5.54	6.69	8.38	6.08	4.31	8	6.68
G	8	12	19	13	13	7	11	9	16	10	11	5	11	9	10	3	13	10
PG	1	1	1	1	1	0.35	0.52	0.44	0.76	0.42	0.52	0.5	0.92	0.82	0.71	0.6	0.52	0.64
D	1	1	1	1	1	0.76	0.84	0.80	0.96	0.71	0.93	0.66	0.98	0.94	0.81	0.7	0.91	0.84
E	1	1	1	1	1	0.52	0.47	0.50	0.74	0.41	0.69	0.5	0.93	0.79	0.63	0.75	0.62	0.66

N, samples size; *G*, number of detected multilocus genotypes; *PG*, proportion of distinguishable genotypes; *D*, genotypic diversity also known as Simpson's diversity index; *E*, genotypic evenness.

Table 5 ANOVA-based *F*- and *R*-statistics for SSR data in *Crataegus douglasii* and *Crataegus suksdorfii*, calculated for all populations as well as separately for *C. douglasii* (tetraploid) and *C. suksdorfii*. Diploids, triploids, and tetraploids of *C. suksdorfii* were analysed both in combination and separately. Because only one site is found to contain tetraploid *C. suksdorfii*, triploids and tetraploids were combined in the analyses

		F-statistics			*R*-statistics		
Populations	*N*	F_{IT}	F_{IS}	F_{ST}	R_{IT}	R_{IS}	R_{ST}
All *C. douglasii* and *C. suksdorfii*	241	0.145	–0.038	0.176	0.195	0.051	0.152
C. douglasii — tetraploids	127	0.054	–0.029	0.079	0.096	0.049*	0.051
C. suksdorfii — diploids, triploids, tetraploids	114	0.245	–0.049	0.280	0.326	0.077**	0.269
C. suksdorfii — diploids	52	0.294	0.189	0.129	0.526	0.375	0.242
C. *suksdorfii* — triploids	41	0.215	–0.339	0.414	0.194	–0.144	0.295
C. *suksdorfii* — triploids and tetraploids	62	0.187	–0.261	0.355	0.202	–0.128	0.293

The two-sided *P* values were all < 0.001 (except *P = 0.02; **P = 0.049). '*N*' denotes samples size.

Diversity indices (D) of some tetraploid populations, e.g. sites MT2, ID6, ID20, and WA21 (Table 4) approach those of diploids.

The LD tests performed on SSR loci in diploid *C. suksdorfii* were shown not to be significant ($P < 0.05$ or smaller, according to the sequential Bonferroni correction) for 70 out of 78 possible pairwise combinations. Tests indicated that 5 of the 13 loci (including CH04G04, CH05D11, and CH05G07 on LG 12 together with CH03D08 and CH05D03 on LG 14) exhibited some degree of linkage. Because the linkage effect cannot be completely avoided in our data and given the constraints of using loci located on a limited number of linkage groups (see Materials and methods), we applied the linkage model in Structure analyses to take the nonrandom association among loci into account when examining genetic structure of our samples.

Partitioning genetic variation among populations

Mantel tests indicated that F_{ST} and R_{ST} indices were significantly correlated in *C. douglasii* ($r = 0.63$, $P = 0.001$), but not in *C. suksdorfii* populations ($r = 0.32$, $P = 0.111$). The genotype-based F_{ST} and R_{ST} estimates (Table 5) and the phenotype-based F_{ST} estimates (Table 6) of *C. suksdorfii* are higher than those of *C. douglasii*, indicative of a higher genetic differentiation among populations in *C. suksdorfii*. According to the AMOVA (Table 6), up to 37% of the variation is found among populations of *C. suksdorfii* ($P = 0.001$), which is higher than that of *C. douglasii* (22%; $P = 0.001$). When *C. suksdorfii* was analysed with respect to ploidy level, up to 35% of the variation was partitioned between diploid and polyploid populations, and 54% among polyploid populations ($P = 0.001$). In contrast, only 15% was detected among diploid populations.

Genetic clustering of individuals

According to the *D*s-based NJ tree (Fig. 2), populations of the west coast *C. suksdorfii* (OR1, OR6, OR11, WA7, and CA5) appear to be distinct from those west of the Cascades, in the Rocky Mountains (e.g. WA21, ID2, ID6, ID15, and MT2), and in the Great Lakes basin (ON20), regardless of

Table 6 Analysis of molecular variance (AMOVA) showing the partitioning of genetic variation among and within populations of *Crataegus douglasii* and *Crataegus suksdorfii*, respectively. The analyses were performed on the Jaccard coefficient matrix of the binary SSR data (absence as '0' and presence as '1' of a defined allele) with 10 000 permutations. Asterisk indicates that all *P* values are ≤ 0.001

Partitioning of variation in the two species	d.f.	Sum of squares	Variance components	Percentage of total variation	F_{ST}*
Among *C. douglasii* populations	7	10.05	0.08	22	0.22
Within *C. douglasii* populations	115	31.53	0.27	78	
Among *C. suksdorfii* populations	6	16.27	0.15	37	0.37
Within *C. suksdorfii* populations	107	27.87	0.26	63	
Partitioning within *C. suksdorfii* with respect to ploidy level					
Among 2*x* and 3*x*	5	12.38	0.15	35	0.35
Within 2*x* and 3*x*	87	23.71	0.27	65	
Among 2*x* and 4*x*	4	7.57	0.11	26	0.26
Within 2*x* and 4*x*	68	21.29	0.31	74	
Among 3*x* and 4*x*	2	9.09	0.21	54	0.54
Within 3*x* and 4*x*	59	10.74	0.18	46	
Among 2*x*	3	3.42	0.06	15	0.15
Within 2*x*	48	16.65	0.35	85	

the ploidy level and taxon. A similar topology was also found in the $\Delta\mu^2$-based dendrogram (tree not shown).

When *C. douglasii* and *C. suksdorfii* individuals were analysed together in Structure (results not shown), the ln[Pr($X|K$)] values kept increasing with higher *K* and a small peak at *K* = 9 was found when the Evanno *et al*. (2005) posterior Δ*K* statistics were applied. In this analysis, the *C. douglasii* membership coefficient values (Q) of respective populations in almost all clusters were less than 0.85, except in the cluster that contains individuals from Ontario (Q = 0.97). This agrees with the results obtained when the *C. douglasii* samples were analysed separately (Fig. 3a) that identified individuals from Ontario as belonging to a single cluster. However, the alleles giving rise to this cluster are shared with individuals from at least one of the western sites (ID6; Fig. 3a). The samples from all but one (ID15) of the western populations each contain individuals whose genotypes resemble more than one cluster in their allelic composition. For *C. suksdorfii*, the optimal partitioning of individuals is obtained for *K* = 5. Diploid individuals from northern California represent predominantly a single cluster (Fig. 3b, pink), whereas the diploids from Washington and Oregon resemble almost equally both the northern California cluster and a cluster common to all diploids, as well as Oregon and Idaho triploids (Fig. 3b, brown). The apparent heterogeneity of gene pools in diploids (Fig. 3b; pink and brown) could be due to an artefact stemming from the use of Structure for both diploids and polyploids. Alternatively, it could have an underlying biological cause. Apart from the alleles shared with the diploids, the two samples of triploid *C. suksdorfii* (OR6, ID6) differ markedly from each other in their allelic composition (Fig. 3b, orange and yellow). Only the Montana tetraploids (MT2) for the most part possess a unique combination of alleles that are not found at any other site (Fig. 3b, blue).

In separate analyses of sites with sympatric cytotypes for example, in Idaho (ID06) where 3*x C. suksdorfii* and 4*x C. douglasii* individuals co-occur, Structure identified two most probable clusters (Fig. S1a, Supporting information) indicating that individuals of these two taxa exhibit almost no admixture of alleles. On the other hand, in the Montana site (MT02) where 4*x C. suksdorfii* and 4*x C. douglasii* co-occur, three clusters were identified in which one of the clusters is shared between the two taxa (Fig. S1b), suggesting that a subset of alleles were mixed by gene flow and shared among these individuals.

Isolation by distance

Relationships between geographical and genetic distances in *C. douglasii* (Fig. 4a) were not significant (D_S: $r = 0.299$, $P = 0.117$; $\Delta\mu^2$: $r = 0.139$, $P = 0.475$). However, Mantel tests revealed a significant association between geographical and genetic distances (D_S) for *C. suksdorfii* ($r = 0.663$, $P = 0.002$; Fig. 4b), regardless of whether within and between ploidy level comparisons were considered, or only the diploid–diploid comparisons (D_S: $r = 0.921$, $P < 0.001$; $\Delta\mu^2$: $r = 0.382$, $P = 0.07$). These results suggest weak barriers to dispersal between populations and/or the similarity of founders shared between neighbouring and distant populations in *C. douglasii*. In contrast, in *C. suksdorfii* it appears that genetic distances measured using the IAM do increase with increasing geographical separation, but the SMM ($\Delta\mu^2$: $r = 0.064$, $P = 0.778$) suggests no distance isolation in *C. suksdorfii* and genetic differentiation among its populations could be attributed to other factors as discussed below.

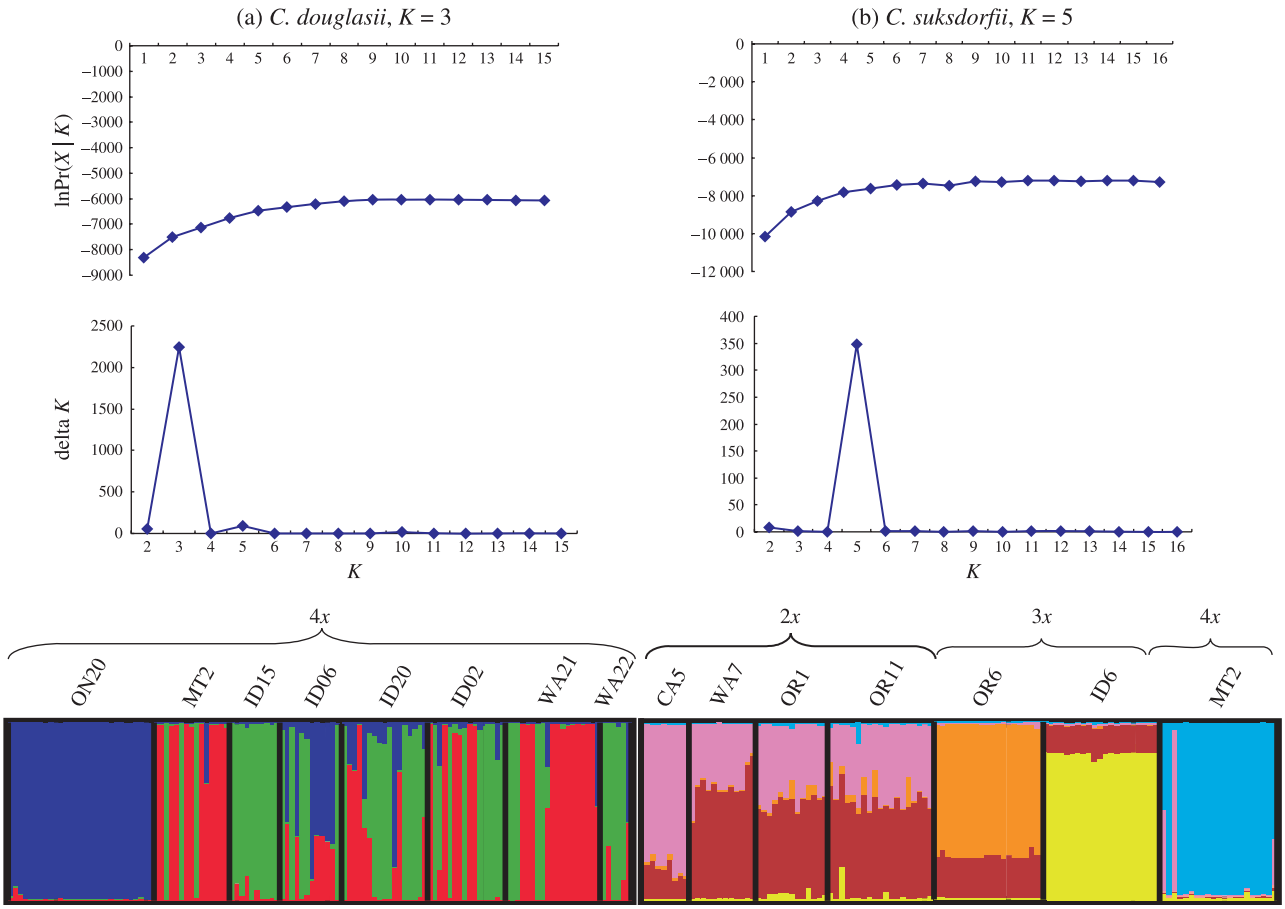

Fig. 3 Bayesian inferences of the *K* clusters estimated by Structure using the linkage model among individuals for (a) *C. douglasii* and (b) *C. suksdorfii*. Plots of ln[Pr(*X* | *K*)] and Δ*K* values against *K* are shown to identify the most probable *K* in each taxon. Clusters in each analysis are represented by colours, and individuals are represented as columns. Within each column (individual), the extent of the component colours indicates the magnitude of the membership coefficient (Q) corresponding to each cluster. In *Crataegus douglasii* (a), three clusters of genotypes are identified, whereas in *Crataegus suksdorfii* (b), five clusters of genotypes are identified.

Discussion

Source of genetic variation in apomictic populations

There are several major factors influencing variation that can result, each by itself or in combination, in the high levels of genetic variation observed in tetraploid *Crataegus suksdorfii* and *Crataegus douglasii* that are known to be predominantly apomictic (Tables 3 and 4; Lo 2008).

Mutation is an important factor that can increase genetic variation in a population and is of special importance in apomictic populations (Hörandl & Paun 2007; Nybom 2007). Theoretical predictions suggest that for populations under stabilizing selection, 0.1% of the genetic variance is attributable to mutation in each generation (Maynard Smith 1998). Such an effect cannot be ignored, particularly with microsatellites, because these regions often exhibit higher mutation rates than other genotyping markers such as amplified fragment length polymorphism, intersimple sequence repeat, and random amplified polymorphic DNA (Rienzo *et al*. 1994; Ellegren 2000; Schlötterer 2000; Estoup *et al*. 2002; Loxdale & Lushai 2003; Nybom 2004). Levels of heterozygosity could be increased by independent mutations in different alleles at different loci, and such mutations may appear faster and accumulate over time, particularly in higher polyploids (e.g. above the tetraploid level; Hörandl & Paun 2007). Even though the ploidy level of our samples is not higher than 4*x*, and the effect of mutation on allelic diversity may not be as high as that in the pentaploids of *Rosa* (Nybom 2004; Nybom *et al*. 2006) and hexaploids of *Ranunculus* (Paun & Hörandl 2006; Paun *et al*. 2006), a factor such as ploidy level may still contribute to allelic variation in our triploid and tetraploid samples. While we cannot avoid the impact of mutations on our microsatellite data, chloroplast markers corroborate results from the nuclear microsatellite markers (Tables 3 and 4).

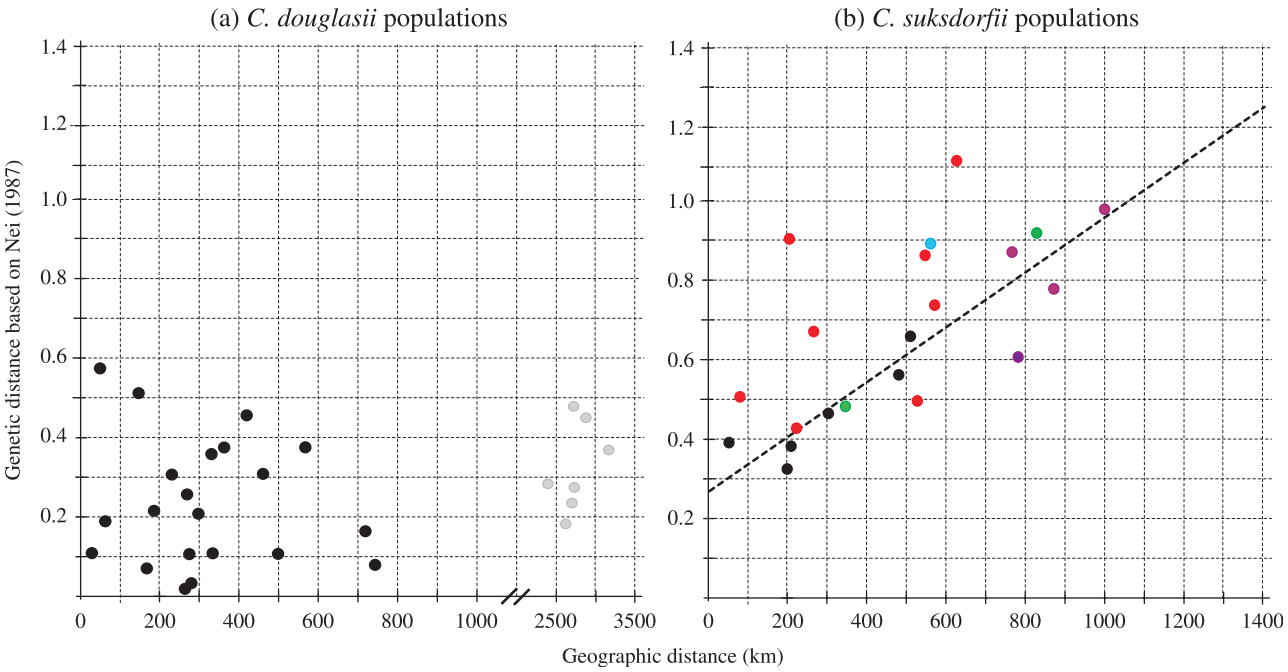

Fig. 4 Scatter plots of pairwise D_S values (IAM-based model; Nei 1978) among populations of *Crataegus douglasii* (a) and *Crataegus suksdorfii* (b) against geographical distances estimated from spatial coordinates (Table 1). In (a), black dots denote comparisons of populations within the Pacific Northwest or western Ontario, while gray dots denote comparisons between the Pacific Northwest and the Ontario (ON20) populations (Table 1). Mantel tests did not indicate a significant relationship between genetic and geographical distances in *C. douglasii* regardless of whether all population comparisons or just those within the Pacific Northwest were considered (see text for details). In (b), comparisons of $2x/2x$ are indicated by black dots, $2x/3x$ by red, $2x/4x$ by purple, $3x/3x$ by blue, and $3x/4x$ by green. Mantel tests comparing genetic and geographical distances rejected H_0: $r \leq 0$ for all *C. suksdorfii* populations, and for the diploid–diploid comparisons (see text for details).

Hence, differences in the levels of genetic variation between our diploid and polyploid taxa are likely to be explained by factors other than mutation within populations.

Polyploid individuals of *C. suksdorfii* and *C. douglasii* have been shown to reproduce through gametophytic apomixis (Dickinson *et al*. 1996; Lo 2008). The lowest chloroplast sequence diversity (H_d and π; Table 3) and SSR-based genotypic diversity (Table 4) in the two triploid populations are to be expected, probably as a consequence of one or more of the following: establishment of the population by seed from a limited number of individuals (Dickinson & Campbell 1991; Dickinson *et al*. 1996), a lack of recruitment of new genotypes due to apomixis, fixation of ancestral genotypes due to failure to set seeds for reasons that are unclear at present (e.g. in site OR06), or frequency-dependent selection for the most fitting genotypes (Van Dijk 2003). This last explanation, however, may not apply to *Crataegus* because it was based on data from clonal herbaceous plants that may differ in generation time and colonizing history (Van Dijk 2003).

On the other hand, the levels of genetic variation observed in some tetraploid populations (e.g. MT2, ID2, and WA21; Tables 3 and 4) are higher than those of the triploids and only slightly lower than the diploids. Despite differences in sensitivity and mutation rate between various marker systems that have to be considered (Loxdale & Lushai 2003; Nybom 2004; Hörandl & Paun 2007), similar levels of genetic variation have also been reported in other polyploid apomicts when compared to their congeneric sexual relatives (e.g. Ellstrand & Roose 1987; Campbell *et al*. 1999; Van der Hulst *et al*. 2003; Paun *et al*. 2006; Palop-Esteban *et al*. 2007). Here we suggest three explanations for the variability observed in our tetraploid apomictic plants. First, tetraploid apomicts of *C. suksdorfii* and *C. douglasii* have been shown to produce highly stainable pollen (Dickinson *et al*. 1996). Apart from endosperm fertilization (i.e. pseudogamy) that is required for seed to be set, sperm nuclei can also fertilize an unreduced egg cell in the embryo sac, given that endosperm balance requirements are often relaxed in polyploid *Crataegus* (Talent & Dickinson 2007a). Occasional sexual recombination to produce offspring with new genotypes is not unusual in other asexual polyploids (D'Souza *et al*. 2004; Hörandl & Paun 2007; Nybom 2007; Whitton *et al*. 2008). Moreover, our tetraploid apomicts are self-compatible (Dickinson *et al*. 1996). Selfing could be already sufficient in generating variability within populations, provided that individuals are highly heterozygous as a result of their hybrid origin

(Nybom 2007). Second, the formation of novel genotypes through allopolyploidization and introgressive hybridization (e.g. in 4*x C. suksdorfii* individuals of site MT02; Fig. S1b) could enrich the initial gene pool (Van Dijk 2003; Hörandl 2006; Hörandl & Paun 2007; Nybom 2007). Third, gene flow via seed dispersal in *C. douglasii* (Fig. 3a) as discussed below could mix gene pools between populations and increase local diversity. All these processes, which are also considered to be important in herbaceous agamosperms (Van Dijk 2003; Hörandl & Paun 2007; Whitton *et al*. 2008; see references therein), may contribute to successful establishment and continued evolution in *Crataegus* apomicts.

Dispersal of apomictic lineages in the Pacific Northwest

Our findings in tetraploid *C. douglasii* corroborate other similar studies in apomictic plants (e.g. Durand *et al*. 2000; Rogstad *et al*. 2002; Robertson *et al*. 2004; Paun *et al*. 2006) that point out the general phenomena of geographical parthenogenesis: genetic variation is not geographically structured but rather is distributed over a broader geographical scale compared to related sexual species (diploid *C. suksdorfii*; Table 6; Fig. 3). In the light of the nonsignificant correlation between genetic and geographical distances of the western populations of *C. douglasii* (Fig. 4a), there are two possible explanations for a lack of geographical structure. First, geographical distance may not be an effective barrier to gene flow in the Pacific Northwest. Gene flow between populations has been documented in several other agamospermous species (see reviews in Hörandl 2006; Whitton *et al*. 2008). It is not surprising for this to occur in *C. douglasii* given that its pollen is viable and common floral visitors (e.g. bees, flies, and beetles) may contribute to some degree of gene mixing between populations. However, because *C. douglasii* individuals reproduce apomictically and are self-compatible (Dickinson *et al*. 1996), this may reduce the frequency of pollen flow. Thus, gene flow between populations is more likely to involve seed instead. Genotypes of *C. douglasii* grow intermingled at sites probably as a result of frequent seed dispersal by birds and small mammals (Courtney & Manzur 1985; Guitián 1998) combined with successful establishment. Such a dispersal would not only allow for rapid colonization of suitable habitats in the Pacific Northwest and elsewhere (e.g. the expansion to the Cypress Hills and upper Great Lakes basin), but would also increase local genotypic diversity of apomictic populations by recruiting new genotypes from other sites (Table 4). An alternative explanation for the lack of geographical structure is that both neighbouring and distant populations were originated from similar founders. The genotypes of these founders may have been multiplied by apomixis independently and thus populations regardless of geographical distances can remain more or less similar in genetic composition even though without much gene exchange.

Potential barriers to gene flow among cytotypes

Reproductive barriers are suggested to be effective in preventing gene flow between cytotypes, as evidenced by the high differentiation and strong genetic structuring across ploidy levels in *C. suksdorfii* (Table 6; Fig. 3b; Fig. S1a). The underlying reason for this is unclear. Although pollination between cytotypes can produce successful seed set in some other *Crataegus* species (Talent & Dickinson 2007b), reproductive isolation is an intuitively obvious consequence of genome multiplication. Reduced fitness in inter-cytotype progeny is expected (e.g. Ramsey & Schemske 1998; Hardy *et al*. 2001; Husband *et al*. 2002), as are epigenetic impacts on the offspring. Such impacts include genomic imprinting leading to phenotypic trait (Liu & Wendel 2003; Scott & Spielman 2006) and endosperm balance number differences between cytotypes (Haig & Westoby 1991; Carputo *et al*. 2003; Kinoshita 2007). Moreover, predominant apomixis and self-compatibility in polyploid *C. suksdorfii* (Dickinson *et al*. 1996) may reduce the extent of pollen-mediated gene flow between populations, even though individuals exhibit similar floral features and phenological regimes (Evans & Dickinson 1996). While we are unable to investigate further the dynamics of tetraploid *C. suksdorfii* due to limited sampling, results from *C. douglasii* clearly demonstrate that frequent dispersal and occasional sexuality are factors involved in the evolution of tetraploid apomicts.

Geographical parthenogenesis has been described in herbaceous angiosperms (Hörandl 2006; references therein) in which apomictic groups almost always exhibit wider niche adaptation and better colonization abilities than their sexual relatives, and lead to a gradual geographical separation of cytodemes without necessarily any corresponding genetic differentiation (Hörandl 2006). In the case of *C. suksdorfii*, the strength of the correlation between genetic and geographical distances detected in diploid but not in polyploid populations (Fig. 4b) suggests that dispersal of diploid sexuals is more constrained by distance compared to polyploid apomicts. Such a contrast in colonization abilities is thus linked both to diploid self-incompatibility, and to the apparently greater ecological amplitude of polyploids.

Evolutionary potential of polyploid apomicts

Our study is the first report of geographical parthenogenesis in woody plant species. In addition, and contrary to the view that apomixis is an evolutionary 'dead-end', our data suggest that apomixis, polyploidy, and other factors discussed above not only maintain genetic variability

within local populations of agamic complexes, but also promote the spread of genotypes across wide geographical distances. It also appears that human disturbance of the landscape may have created further opportunities for originally diploid sexual lineages to diversify and spread as a result of polyploidy and agamospermy (Dickinson *et al*. 2008). In future work we will seek to document the way in which this geographical distribution corresponds to considerable ecological amplitude in the case of tetraploid *C. douglasii* and polyploid *C. suksdorfii*. It is noteworthy that the Pacific Northwest plants that we have referred to here as *C. douglasii sensu* lato include individuals that can be distinguished as the segregate species *Crataegus castlegarensis* and *Crataegus okennonii* (Phipps & O'Kennon 1998, 2002). The role of these factors in the speciation processes thus remains, at this point, still to be discovered because of the difficulty in applying any one species concept to these organisms.

Acknowledgements

The authors thank Rhoda Love and Peter Zika for the extensive collections of spring and fall vouchers, as well as for field directions; Rhoda Love, Sophie Nguyen, and Nadia Talent for assisting us with plant collection and identification; Tara Paton and Simone Russell for the ABI facilities at the Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada; Annabel Por, Cheying Ng, and Jenny Bull for helping organize the vouchers for this study. Asher Cutter, Dan Potter, and three 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 and 326439-06 to SS), the Botany Department of the University of Toronto, and the Royal Ontario Museum is gratefully acknowledged, as is an award from the Royal Ontario Museum Foundation to Mark Engstrom and TAD for the purchase of the thermocycler used in part of this work.

References

Asker S, Jerling L (1992) *Apomixis in Plants*. CRC Press, Boca Raton, Florida.

Bartish IV, Hylmö B, Nybom H (2001) RAPD analysis of interspecific relationships in presumably apomictic *Cotoneaster* species. *Euphytica*, **120**, 273–280.

Bohonak AJ (2002) IBD (Isolation By Distance): a program for analyses of isolation by distance. *Journal of Heredity*, **93**, 153–154.

Bruvo R, Michiels NK, D'Souza TG, Schulenburg H (2004) A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology*, **13**, 2101–2106.

Calzada JPV, Crane CF, Stelly DM (1996) Apomixis: the asexual revolution. *Science*, **274**, 1322–1323.

Campbell CS, Alice LA, Wright WA (1999) Comparisons of within-population genetic variation in sexual and agamospermous *Amelanchier* (Rosaceae) using RAPD markers. *Plant Systematics and Evolution*, **215**, 157–167.

Campbell CS, Evans RC, Morgan DR, Dickinson TA, Arsenault MP (2007) Phylogeny of subtribe Pyrinae (formerly the Maloideae, Rosaceae): limited resolution of a complex evolutionary history. *Plant Systematics and Evolution*, **266**, 119–145.

Carman JG (1997) Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. *Biological Journal of the Linnean Society*, **61**, 51–94.

Carputo D, Frusclante L, Peloquin SJ (2003) The role of 2*n* gametes and endosperm balance number in the origin and evolution of polyploids in the tuber-bearing solanums. *Genetics*, **163**, 287–294.

Casgrain P, Legendre P (2004) *The R Package for Multivariate and Spatial Analysis v. 4.0d10*. Université de Montréal, Montréal, Canada. Available from URL: http://www.bio.umontreal.ca/casgrain/en/labo/R/v4/.

Chester M, Cowan RS, Fay MF, Rich TCG (2007) Parentage of endemic *Sorbus* L. (Rosaceae) species in the British Isles: evidence from plastid DNA. *Botanical Journal of the Linnean Society*, **154**, 291–304.

Courtney SP, Manzur MI (1985) Fruiting and fitness in *Crataegus monogyna*: the effect of frugivores and seed predators. *Oikos*, **44**, 398–406.

D'Souza TG, Storhas M, Schulenburg H, Beukeboom LW, Michiels NK (2004) Occasional sex in an 'asexual' polyploid hermaphrodite. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 1001–1007.

Darlington CD (1958) *Evolution of Genetic Systems*. Oliver and Boyd, Edinburgh, UK.

Daurelio LD, Espinoza F, Quarin CL, Pessino SC (2004) Genetic diversity in sexual diploid and apomictic tetraploid populations of *Paspalum notatum* situated in sympatry or allopatry. *Plant Systematics and Evolution*, **244**, 189–199.

Dickinson TA, Campbell CS (1991) Population structure and reproductive ecology in the Maloideae (Rosaceae). *Systematic Botany*, **16**, 350–362.

Dickinson TA, Belaoussoff S, Love RM (1996) North American black-fruited hawthorns. I. Variation in floral construction, breeding system correlates, and their possible evolutionary significance in *Crataegus* sect. *Douglasii* Loudon. *Folia Geobotanica et Phytotaxonomica*, **31**, 355–371.

Dickinson TA, Lo EYY, Talent N (2007) Polyploidy, reproductive biology, and Rosaceae: understanding evolution and making classification. *Plant Systematics and Evolution*, **266**, 59–78.

Dickinson TA, Lo EYY, Talent N, Love RM (2008) Black-fruited hawthorns of western North America — one or more agamic complexes? *Botany*, **86**, 846–865.

Durand J, Garnier L, Dajoz I, Mousset S, Veuille M (2000) Gene flow in a facultative apomictic Poacea, the savanna grass *Hyparrhenia diplandra*. *Genetics*, **156**, 823–831.

Ellegren H (2000) Microsatellite mutations in the germline. *Trends in Genetics*, **16**, 551–558.

Ellstrand NC, Roose ML (1987) Patterns of genotypic diversity in clonal plant species. *American Journal of Botany*, **74**, 123–131.

Esselink GD, Nybom H, Vosman B (2004) Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. *Theoretical and Applied Genetics*, **109**, 402–408.

Esselman EJ, Jianqiang L, Crawford DJ, Windus JL, Wolfe AD (1999) Clonal diversity in the rare *Calamagrostis porteri* ssp. *insperata* (Poaceae): comparative results for allozymes and random amplified polymorphic DNA (RAPD) and intersimple sequence repeat (ISSR) markers. *Molecular Ecology*, **8**, 443–451.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology Notes*, **14**, 2611–2620.

Evans RC, Dickinson TA (1996) North American black-fruited hawthorns (*Crataegus* section *Douglasii* Loud.): floral development of 10- and 20-stamen morphotypes. *American Journal of Botany*, **83**, 961–978.

Excoffier L, Smouse PE, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Felsenstein J (2006) *PHYLIP (Phylogeny Inference Package), Version 3.66c*. Zoology Department, University of Washington, Seattle, Washington.

Garnier LKM, Durand J, Dajoz I (2002) Limited seed dispersal and microspatial population structure of an agamospermous grass of West African savannahs, *Hyparrhenia diplandra* (Poaceae). *American Journal of Botany*, **89**, 1785–1791.

Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C (1998) Simple sequence repeats for the genetic analysis of apple. *Theoretical and Applied Genetics*, **96**, 1069–1076.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463–471.

Gornall RJ (1999) Population genetic structure in agamospermous plants. In: *Molecular Systematics and Plant Evolution* (eds Hollingsworth PM, Bateman RM, Gornall RJ), pp. 118–138. Taylor & Francis, London.

Guitián P (1998) Latitudinal variation in the fruiting phenology of a bird-dispersed plant (shape *Crataegus monogyna*) in Western Europe. *Plant Ecology*, **137**, 139–142.

Hagen AR, Sæther T, Borgen L, Elven R, Stabbetorp OE, Brochmann C (2002) The arctic-alpine polyploids *Cerastium alpinum* and *C. nigrescens* (Caryophyllaceae) in a sympatric situation: breakdown of species integrity? *Plant Systematics and Evolution*, **230**, 203–219.

Haig D, Westoby M (1991) Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **333**, 1–14.

Hamrick JL, Godt MJ (1990) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **351**, 1291–1298.

Hamrick JL, Godt MJ, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Hardy OJ, Loose MD, Vekemans X, Meerts P (2001) Allozyme segregation and inter-cytotype reproductive barriers in the polyploid complex. *Heredity*, **87**, 136–145.

Hörandl E (2006) The complex causality of geographical parthenogenesis. *New Phytologist*, **171**, 525–538.

Hörandl E, Paun O (2007) Patterns and sources of genetic diversity in apomictic plants: implications for evolutionary potentials. In: *Apomixis: Evolution, Mechanisms, and Perspectives* (eds Hörandl E, Grossniklaus U, Van Dijk PJ, Sharbel TF), pp. 169–194. Koenigstein, Germany.

Husband B, Schemske DW, Burton TL, Goodwilliw C (2002) Pollen competition as a unilaternal reproductive barrier between sympatric *Chamerion angustifolium*. *Proceedings of the Royal Society B: Biological Sciences*, **269**, 2565–2571.

Joly S, Starr JR, Lewis WH, Bruneau A (2006) Polyploid and hybrid evolution in roses east of the Rocky Mountains. *American Journal of Botany*, **93**, 412–425.

Kinoshita T (2007) Reproductive barrier and genomic imprinting in the endosperm of flowering plants. *Genes and Genetic Systems*, **82**, 177–186.

Koltunow AM, Grossniklaus U (2003) Apomixis: a developmental perspective. *Annual Review of Plant Biology*, **54**, 547–574.

Legendre L, Legendre P (1983) *Numerical ecology*. *Developments in Environmental Modelling*. Elsevier Scientific Publishing Company, Amsterdam, The Netherlands.

Liebhard R, Gianfranceschi L, Koller B *et al*. (2002) Development and characterisation of 140 new microsatellites in apple (*Malus* × *domestica* Borkh). *Molecular Breeding*, **10**, 217–241.

Liu B, Wendel JF (2003) Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution*, **29**, 365–379.

Lo EYY (2008) Global and fine scale molecular studies of polyploid evolution in *Crataegus* (Rosaceae). PhD Dissertation. Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada.

Lo EYY, Stefanović S, Dickinson TA (2007) Molecular reappraisal of relationships between *Crataegus* and *Mespilus* (Rosaceae, Pyreae) — two genera or one? *Systematic Botany*, **32**, 596–616.

Loomer SA (2003) *MicroCAM (Microcomputer Aided Mapping) for Windows V2.05*. Available from URL: http://www.csiss.org/map-projections/microcam/.

Loxdale HD, Lushai G (2003) Rapid changes in clonal lines: the death of a 'sacred cow'. *Biological Journal of the Linnean Society*, **79**, 3–16.

Maynard Smith J (1998) *Evolutionary Genetics*. Oxford University Press, Oxford, UK.

McLellan AJ, Prati D, Kaltz O, Schmid B (1997) Structure and analysis of phenotypic and genetic variation in clonal plants. In: *The Ecology and Evolution of Clonal Plants* (eds de Kroon H, Van Groenendael JM), pp. 185–210. Backhuys, Leiden, Netherlands.

Meirmans PG, Van Tienderen PH (2004) GenoType and GenoDive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Meirmans PG, Vlot EC, Den Nijs JCM, Menken SBJ (2003) Spatial ecological and genetic structure of a mixed population of sexual diploid and apomictic triploid dandelions. *Journal of Evolutionary Biology*, **16**, 343–352.

Menken SBJ, Smit E, Hans Den Nijs JCM (1995) Genetical population structure in plants: gene flow between diploid sexual and triploid asexual dandelions (*Taraxacum* section *Ruderalia*). *Evolution*, **49**, 1108–1118.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nogler GA (1984) Gametophytic apomixis. In: *Embryology of*

Angiosperms (eds Johri BM), pp. 475–518. Springer-Verlag, Berlin, Germany.

Noyes RD, Soltis DE (1996) Genotypic variation in agamospermous *Erigeron compositus* (Asteraceae). *American Journal of Botany*, **83**, 1292–1303.

Nybom H (1998) Biometry and DNA fingerprinting detected limited genetic differentiation among populations of the blackberry *Rubus nesssensis* (Rosaceae). *Nordic Journal of Botany*, **18**, 323–333.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, **13**, 1143–1155.

Nybom H (2007) Unique reproduction in dogroses (*Rosa* sect. *Caninae*) maintains successful and highly heterozygous genotypes. In: *Apomixis: Evolution, Mechanisms, and Perspectives* (eds Hörandl E, Grossniklaus U, Van Dijk PJ, Sharbel TF), pp. 281–298. Regnum Vegetabile, Koenigstein, Germany.

Nybom H, Esselink GD, Warlemark G, Leus L, Vosman B (2006) Unique genomic configuration revealed by microsatellite DNA in polyploid dogroses, *Rosa* sect. Caninae. *Journal of Evolutionary Biology*, **19**, 635–648.

Obbard DJ, Harris SA, Pannell JR (2006) Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. *Heredity*, **97**, 296–303.

Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357–358.

Palop-Esteban M, Segarra-Moragues JG, Gonzălez-Candelas F (2007) Historical and biological determinants of genetic diversity in the highly endemic triploid sea lavender *Limonium dufourii* (Plumbaginaceae). *Molecular Ecology*, **16**, 3814–3827.

Paun O, Greilhuber J, Temsch EM, Hörandl E (2006) Patterns, sources and ecological implications of clonal diversity in apomictic *Ranunculus carpaticola* (*Ranunculus auricomus* complex, Ranunculaceae). *Molecular Ecology*, **15**, 897–910.

Paun O, Hörandl E (2006) Evolution of hypervariable microsatellites in apomictic polyploid lineages of *Ranunculus carpaticoloa*: Directional bias at dinucleotide loci. *Genetics*, **174**, 387–398.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Persson-Hovmalm HA, Jepsson N, Bartish IV, Nybom H (2004) RAPD analysis of diploid and tetraploid populations of *Aronia* points to different reproductive strategies within the genus. *Hereditas*, **141**, 301–312.

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

Phipps JB, O'Kennon RJ (2002) New taxa of *Crataegus* (Rosaceae) from the northern Okanagan-southwestern Shuswap diversity center. *Sida*, **20**, 115–144.

Pritchard JK, Wen W (2004) *Documentation for the Structure software Version 2*. Department of Human Genetics, University of Chicago, Chicago, Illinois. Available from URL: http://pritch.bsd.uchicago.edu/software/structure2_2.html.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, **29**, 407–501.

Raymond AE, Rousset F (1995) GenePop (version 3.3): population genetics software for exact test and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rice WR (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

Rienzo AD, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences, USA*, **91**, 3166–3170.

Robertson A, Newton AC, Ennos RA (2004) Multiple hybrid origins, genetic diversity and population genetic structure of two endemic *Sorbus* taxa on the Isle of Arran, Scotland. *Molecular Ecology*, **13**, 123–134.

Rogstad SH, Keane B, Beresh J (2002) Genetic variation across VNTR loci in central North American *Taraxacum* surveyed at different spatial scales. *Plant Ecology*, **161**, 111–121.

Ronfort J, Jenczewski E, Bataillon T, Rousset F (1998) Analysis of population structure in autotetraploid species. *Genetics*, **150**, 921–930.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rozas J, Sánchez-Delbarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**, 365–371.

Scott RJ, Spielman M (2006) Genomic imprinting in plants and mammals: how life history constrains convergence. *Cytogenetic and Genome Research*, **113**, 53–67.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Starfinger U, Stocklin J (1996) Seed, pollen, and clonal dispersal and their role in structuring plant populations. *Progress in Botany*, **57**, 337–355.

Stebbins GL (1950) *Variation and Evolution in Plants*. Columbia University Press, New York.

Talent N, Dickinson TA (2005) Polyploidy in *Crataegus* and *Mespilus* (Rosaceae, Maloideae): evolutionary inferences from flow cytometry of nuclear DNA amounts. *Canadian Journal of Botany*, **83**, 1268–1304.

Talent N, Dickinson TA (2007a) Endosperm formation in aposporous *Crataegus* (Rosaceae, Spiraeoideae, tribe Pyreae): parallels to Ranunculaceae and Poaceae. *New Phytologist*, **173**, 231–249.

Talent N, Dickinson TA (2007b) Ploidy level increase and decrease in seeds from crosses between sexual diploids and asexual triploids and tetraploids in *Crataegus* L. (Rosaceae, Spiraeoideae, Pyreae). *Canadian Journal of Botany*, **85**, 570–584.

Tsumura Y, Yoshimura K, Tomaru N, Ohba K (1995) Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloroplast genes. *Theoretical Applied Genetics*, **91**, 1222–1236.

Van der Hulst RGM, Mes THM, Den Nijs JCM, Bachmann K (2000) Amplified fragment length polymorphism (AFLP) markers reveal that population structure of triploid dandelions (*Taraxacum officinale*) exhibits both clonality and recombination. *Molecular Ecology*, **9**, 1–8.

Van der Hulst RGM, Mes THM, Den Nijs JCM, Bachmann K (2003) Genetic structure of a population sample of apomictic dandelions. *Heredity*, **90**, 326–336.

Van Dijk P (2003) Ecological and evolutionary opportunities of apomixis: insights from *Taraxacum* and *Chondrilla*. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **358**, 1113–1121.

Ward DB (1974) The 'Ignorant Man' technique of sampling plant populations. *Taxon*, **23**, 325–330.

Whitton J, Sears CJ, Bacck EJ, Otto SP (2008) The dynamic nature of apomixis in the angiosperms. *International Journal of Plant Science*, **169**, 169–182.

Widén B, Cronberg N, Widén M (1994) Genotypic diversity, molecular markers, and spatial distribution of genets in clonal plants, a literature survey. *Folia Geobotanica et Phytotaxonomica*, **29**, 245–263.

Wright S (1965) The interpretation of population structure *F*-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.

This work is part of the PhD thesis of Eugenia Y. Y. Lo that uses both molecular markers and flow cytometry to investigate polyploid evolution in the taxonomically complex genus Crataegus (hawthorn). Eugenia Y. Y. Lo is interested in studying the influence of polyploidy on population structure and on pattern and rate of species diversification. Saša Stefanović works on a variety of species-level issues, including evolution of breeding systems, historical biogeography, and molecular systematics and evolution, with a particular focus on nonphotosynthetic plants. Timothy A. Dickinson studies hawthorns, and is interested in the impact of breeding system on population structure and classification.

Supporting Information

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

Fig. S1 Bayesian inferences of the *K* clusters estimated by Structure using the linkage model among sympatric *Crataegus douglasii* and *Crataegus suksdorfii* individuals of sites (a) Idaho (ID06) and (b) Montana (MT02). Clusters in each analysis are represented by different shades of gray, and individuals are represented as columns.

Please note: Wiley-Blackwell are 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.