Received: 22 May 2017 | Revised: 25 July 2017 | Accepted: 8 August 2017

DOI: 10.1111/mec.14331

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Widespread generalist clones are associated with range and niche expansion in allopolyploids of Pacific Northwest Hawthorns (*Crataegus* L.)

J. M. Coughlan[1,2,3,4] | S. Han[1] | S. Stefanović[2] | T. A. Dickinson[1,3]

[1]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada

[2]Department of Biology, University of Toronto, Mississauga, Mississauga, Ontario, Canada

[3]Green Plant Herbarium (TRT), Department of Natural History, Royal Ontario Museum, Toronto, Ontario, Canada

[4]Department of Biology, Duke University, Durham, NC, USA

Correspondence
J. M. Coughlan, Department of Biology, Duke University, Durham, NC, USA.
Email: jennifer.coughlan@duke.edu

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: A3430, 326439, STPGP 381073–09

Abstract

Range and niche expansion are commonly associated with transitions to asexuality, polyploidy and hybridity (allopolyploidy) in plants. The ability of asexual polyploids to colonize novel habitats may be due to widespread generalist clones, multiple ecologically specialized clones, or may be a neutral by-product of multiple, independent origins of asexual polyploids throughout the range. We have quantified niche size and divergence for hawthorns of the Pacific Northwest using data from herbarium vouchers with known cytotypes. We find that all polyploid niches diverge from that of the diploid range, and allopolyploids have the broadest niches. Allotetraploids have the largest niche and the widest geographic distribution. We then assessed the genetic mechanism of range expansion by surveying the ecological and geographic distribution of genotypes within each cytotype from sites in which fine-scale habitat assessments were completed. We find no isolation by either geographic or ecological distance in allopolyploids, suggesting high dispersal and colonization ability. In contrast, autotriploids and diploids show patterns of isolation by geographic distance. We also compared the geographic and ecological distributions of clonal genotypes with those of randomly drawn sites of the most widespread cytotype. We found that most clones are geographically widespread and occur in a variety of habitats. We interpret these findings to suggest that patterns of range and niche expansion in Pacific Northwest Hawthorns may stem from these widespread, ecologically generalist clones of hybrid origin.

KEYWORDS
apomixis, general-purpose genotypes, geographic parthenogenesis, hybridization, polyploidization

1 | INTRODUCTION

Polyploidization is a common evolutionary transition, particularly in plants (Comai, 2005; Otto & Whitton, 2000; Ramsey & Schemske, 1998). In plants, however, polyploidization is also often associated with hybridity (allopolyploidization) and the transition to asexual reproduction (i.e., through gametophytic apomixis; the formation and parthenogenetic development of unreduced female gametes; reviewed in Hörandl, 2006 and De Storme & Mason, 2014). The evolutionary potential of these asexual allopolyploid taxa has long been debated (i.e., Comai, 2005; Mayrose et al., 2011; Otto, 2007; Stebbins, 1940), but one common consequence is the expansion of range and niche in asexual polyploid taxa relative to their sexual, diploid progenitors (Geographic Parthenogenesis; Bierzychudek, 1985; Hijmans et al., 2007; Lowry & Lester, 2006; Vandel, 1928).

Asexuality, polyploidization and hybridization may each contribute to niche expansion, either independently, or in concert. Asexuality allows for single individuals to colonize new geographic areas

without mate limitation (Baker's Law, Baker, 1955; Pannell & Barrett, 1998) or risk of inbreeding depression (Bierzychudek, 1985; Hörandl, 2006). Polyploidization may allow for better adaptive potential, as the probability of a mutation (including beneficial mutations) increases with increasing gene copy number (see Otto & Whitton, 2000 for details). Hybridization may also increase adaptive potential, as hybrid individuals may be able to access one or both parental genomes (Shimizu-Inatsugi et al., 2017), and in some cases, have been shown to have novel patterns of gene expression relative to their progenitors (Paun et al., 2011). These potential benefits are amplified in allopolyploid apomicts, which do not undergo recombination via sexual reproduction and are fixed heterozygotes. An increase in adaptive potential paired with a lack of mate limitation can increase colonization ability and niche breadth, and thus allows for the colonization of geographically and ecologically novel environments. However, we lack a comprehensive understanding of the genetic mechanisms by which range and niche expansions are achieved in asexual polyploids.

Two main hypotheses regarding the distribution of clonal genotypes have been proposed. The "General-Purpose Genotypes" hypothesis posits that a given clonal genotype, having access to multiple alleles at each locus, may have greater plasticity, and hence would be able to live in a greater variety of habitats than their sexual, diploid relatives (Baker, 1965; Lynch, 1984). Evidence for this hypothesis is mixed. Many studies have found that asexual polyploids possess less variance in fitness components across environmental gradients (Bierzychudek, 1989; Hancock & Bringhurst, 1981; Michaels & Bazzaz, 1989; Oplaat & Verhoeven, 2015; although see Bretagnolle & Thompson, 2001) or have more adaptively plastic gene expression across environments (Shimizu-Inatsugi et al., 2017). However, of the few studies that have genotyped range-wide samples of cytotypes, widespread clonal genotypes are often very rare or nonexistent (Cosendai, Wagner, Ladinig, Rosche, & Hörandl, 2013; Ellstrand & Roose, 1987; Van Dijk, 2003). Alternatively, the "Frozen Niche" hypothesis posits that a given asexual polyploid taxon comprises several, genetically diverse, ecologically specialized clonal lineages, and collectively these clones increase niche breadth (Vrijenhoek, 1984; Vrijenhoek & Parker, 2009). While, most clones do tend to be geographically restricted to one or two sites (Ellstrand & Roose, 1987), or even to particular habitats within sites (Paun, Greilhuber, Temsch, & Hörandl, 2006), few studies have distinguished whether these clones are truly ecologically specialized (i.e., locally adapted), or simply just of recent origin and/or with limited dispersal ability. The few studies which have been able to quantify the extent of local adaptation among populations of higher cytotypes (either through quantifying fitness (i.e., Ramsey (2011)), or quantified the spread of alleles across ecological gradients (i.e., Parisod and Christin (2008)) have been carried out in autopolyploids. Thus, we also lack an understanding of how hybridization may contribute to differences in the genetic mechanism by which niche expansion is accomplished in polyploid taxa.

Apparent niche expansion may also simply be a by-product of the demographic history by which asexual polyploid taxa evolved (i.e., Thompson & Whitton, 2006; Van Dijk & Bakx-Schotman, 1997). If an asexual allopolyploid lineage originates multiple times, each time in a separate geographic area or habitat, the combined range and/or niche of these independent allopolyploid lineages could appear larger than the diploid progenitor ranges. For example, Meimberg, Rice, Milan, Njoku, and McKay (2009) have found that range and niche expansions in *Aegilops* are associated with the number of origins of a given asexual cytotype.

Here, we aim to compare the ecological distributions of cytotypes of hawthorns (*Crataegus* L., Rosaceae) of the Pacific Northwest, moderately long-lived woody perennials that exhibit geographic parthenogenesis (Coughlan, Stefanović, & Dickinson, 2014; Lo, Stefanovic, & Dickinson, 2013). Using range-wide data, we first confirm the result that while diploids occupy a relatively limited geographic range, polyploids are considerably more widely distributed. We then test whether the polyploids have undergone niche differentiation and/or niche expansion, and whether this phenomenon differs among allo- and autopolyploids. We then use the geographic and ecological distribution of genotypes to test whether niche expansion in allotetraploids is associated with (i) widespread, ecologically generalist clones, consistent with the general-purpose genotypes hypothesis, (ii) geographically restricted, ecologically specialized clones, consistent with the frozen niche hypothesis or (iii) highly spatially structured populations, consistent with multiple origins.

2 | MATERIALS AND METHODS

2.1 | Study group

Pacific Northwest hawthorns comprise at least four reasonably widespread species, two that are largely confined to this area, and two that have trans-continental ranges that encroach on the Pacific Northwest from the east. The first two species are the most common and abundant representatives of a mainly black-fruited clade recognized as *Crataegus* subgenus *Sanguineae* Ufimov (Eurasia and North America; the taxa dealt with here are exclusively North American in distribution; Zarrei, Stefanović, & Dickinson, 2014). *Crataegus suksdorfii* (Sarg.) Kruschke is recognizable because of its flowers with approximately 20 stamens each. As presently understood, this species comprises at least three cytotypes, diploid, triploid and tetraploid, that are largely allopatric in distribution (Coughlan et al., 2014; Figure 1). *Crataegus douglasii* Lindl. is almost exclusively tetraploid and has flowers with approximately 10 stamens each. Its distribution overlaps with and extends beyond (to a set of disjunct populations in the upper Great Lakes basin) those of the polyploid *C. suksdorfii* cytotypes (Coughlan et al., 2014; Figure 1). The other two species in our comparison belong to a large, at present poorly differentiated clade of red-fruited eastern North American species that make up *C.* subg. *Americanae* El-Gazzar (exclusively North American in distribution). The two members of *C.* subg. *Americanae* in our sample, *C. chrysocarpa* Ashe and *C. macracantha* Lodd. ex Loudon, are both tetraploids, with 10 stamens per flower.

The *Sanguineae* polyploids referred to above have been shown to be allopolyploids, formed through recurrent backcrosses between

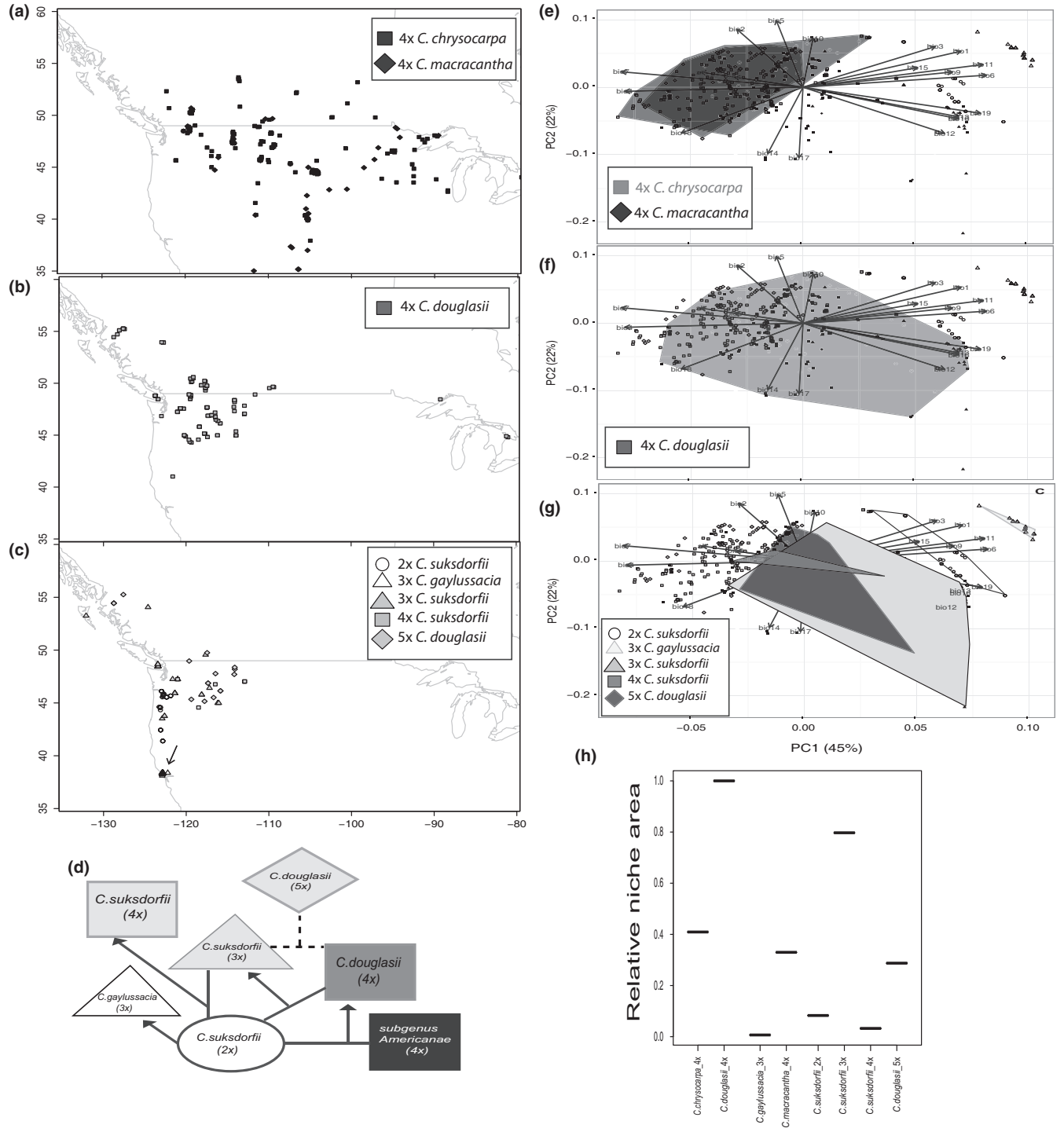


FIGURE 1 Geographic and ecological distributions for Pacific Northwest *Crataegus*, as well as their hypothesized phylogenetic relationships. (a–c) Geographic Distributions based on Royal Ontario Museum herbarium records for (a) *Crataegus* subg. *Americanae*, (b) allotetraploid *C. douglasii* and (c) multiple cytotypes of *C. suksdorfii* as well as autotriploid *C. gaylussacia* and pentaploid *C. douglasii*. Autotriploid *C. gaylussacia* are indicated by the black arrow. (d) Inferred taxonomic relationships among cytotypes of the Pacific Northwest Hawthorns. Solid lines based on Lo et al. (2010) and Zarrei et al. (2014), dotted line is the hypothesized relationship based on geographic distribution and morphological resemblance. (e–g) Ecological distribution of all cytotypes in *Crataegus* subg. *Americanae* (e), allotetraploid *C. douglasii* (f) and multiple cytotypes of *C. suksdorfii* as well as autotriploid *C. gaylussacia*, and pentaploid *C. douglasii* (g) using site data from Royal Ontario Museum databased collections in a PCA of climate data. Polygons represent the convex hulls which includes all sites from a given cytotype. (h) Niche area (as defined by the area of the convex hull in the PC1-PC2 plane that enclosed all individuals) as a proportion of the allotetraploid *C. douglasii* niche

allotetraploid *C. douglasii* and various cytotypes of *C. suksdorfii* (Lo, Stefanović, & Dickinson, 2010; Zarrei et al., 2014; Figure 1). In addition, the sample studied here includes two more black-fruited *Sanguineae* cytotypes: pentaploid *C. douglasii*, and autotriploid *C. gaylussacia* Heller. The pentaploid appears to occur sporadically in the northwestern portion of the range of *C. douglasii* and is hypothesized to have arisen between backcrosses between allotetraploid *C. douglasii* and allotriploid *C. suksdorfii* (Figure 1). Here, we use the name *C. gaylussacia* in its original sense to refer to specimens with 20 stamens per flower from Marin and Sonoma counties in California because of the evidence that it is an autotriploid (Zarrei et al., 2014). This usage is mentioned in *Flora North America* (Phipps, 2015) as an alternative to not distinguishing the different cytotypes present in the complex.

For most North American, polyploid *Crataegus* studied to date the predominant mode of reproduction involves parthenogenetic development of embryos from unreduced female gametes (apospory) that is dependent on the occurrence of central cell fertilization, and hence pollination (pseudogamous gametophytic apomixis; Dickinson, Lo, & Talent, 2007; Lo et al., 2013). Because hawthorns, in common with many other Rosaceae, have gametophytic self-incompatibility, it means that unlike diploids, polyploid *Crataegus* are able to fertilize the central cell with self-pollen (Campbell, Greene, & Dickinson, 1991).

2.2 | Niche assessment of herbarium samples

We created general climate niches for each of the *C.* subg. *Sanguineae* cytotypes using the records for the *Crataegus* specimens in the Royal Ontario Museum Green Plant Herbarium. Although previous researchers have explored the ecological distributions of taxa within this group before, this is the first report which fully explores these ecological distributions from a range-wide sample of these taxa. Because *C. douglasii* is thought to be a hybrid between diploid *C. suksdorfii* and a member of *C.* subg. *Americanae* (Zarrei et al., 2014), we also created climate niches for North American *C.* subg. *Sanguineae* and *C.* subg. *Americanae*, which has not been previously described. The herbarium records represent 590 independent sites, from which at least one individual tree has been collected for either *C.* subg. *Americanae* (n = 214) or *C.* subg. *Sanguineae* (n = 376). For the *Sanguineae*, only sites with individuals of known cytotype (determine by flow cytometry or chromosome squashes) were used to create the distribution map and climate niches (n = 261, compared to Lo et al., 2013; where n = 24). To distinguish autotriploid *C. gaylussacia* from other triploids within the complex, we also use geographical information, as only *C. gaylussaccia* is known to exist in Sonoma and Marin county, California. We cannot exclude the possibility of other autotriploid *C. suksdorfii* that exist further north or inland; however, the only autotriploid *C. suksdorfii* that have been described to date occur adjacent to the range of diploid *C. suksdorfii* (i.e., Lo, Stefanović, & Dickinson, 2009; Lo et al., 2010, 2013). Outside of this area, autotriploid *C. suksdorfii* are thought to be relatively rare (Coughlan et al., 2014). In fact, for populations that were genotyped in the current study and others, no obvious autotriploid *C. suksdorfii* populations were found in the bulk of the triploid *C. suksdorfii* range (Zarrei et al., 2014).

For climate information for each site, we used the 19 Bioclim variables (and elevation) available WorldClim data sets (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). We then completed a PCA in which each variable was scaled to have unit variance (R Core Team, 2012). The first two components of the PCA accounted for 45% and 22% of the variance, respectively, while the third and lower PC axes accounted for 11% or less of the variation within the climate data. We assessed the significance of the variance explained by each PC axis using a Broken Stick Model (Legendre & Legendre, 1998). PCs 1–4 explained a significant proportion of the variance, but as PC1 and PC2 explained approximately 67% of the total variance, we simply use the first two PCs to visualize and test climate niches for each cytotype. For each cytotype, we used the area of the convex hull in the PC1–PC2 plane that enclosed all individuals to estimate the size of the niche, and compared these relative to the largest, that for *C. douglasii*. We completed an ANOVA on each PC to determine if the niches were significantly different from each other, with a post hoc Tukey Honest Significant Difference test (which corrects for multiple comparisons) to determine which cytotypes differed in PC space. All analyses were carried out in R (R Core Team, 2012).

2.3 | Field collections

To understand the genetic basis of niche breadth revealed in our analysis of herbarium records, Pacific Northwest hawthorns were collected, genotyped, and had their cytotype determined by flow cytometry in the spring and summer of 2011. Collections consist of 213 individuals from 52 distinct sites (with an average of four individuals per site, see Table 1 for sample size details. Note that for one "site" for *C. gaylussacia*—Central CA, individuals were pooled from a relatively small geographic region, as they occurred as singleton trees). The sites cover much of the range of *Crataegus* subgenus *Sanguineae* (Coughlan et al., 2014; Dickinson, Lo, Talent, & Love, 2008) from central California to central British Columbia, and as far East as central Montana (Figure 2). Individuals at all sites were collected randomly, using ignorant-person sampling (Ward, 1974). Voucher specimens and flow cytometry were used to confirm species identity upon return.

Two leaf tissue samples were collected from each individual, one of which was immediately desiccated in silica gel sand for further genetic analysis, and the other was kept cool and dry for cytotype determination. Two flowering herbarium vouchers were collected per individual in the spring of 2011, and where possible two fruiting herbarium vouchers were collected per individual later in the summer 2011, 2012 or 2013 and deposited in the Green Plant Herbarium of the Royal Ontario Museum (TRT) for identification (see Table S1 for details). A small number of putative *C. douglasii* individuals (13 of 213) were collected at a developmental stage at which species identification was difficult. We inferred species identities for these individuals based on (i) ploidy level determination, (ii) the morphological features that could be assessed on the specimens and (iii)

TABLE 1 Patterns of genetic diversity and divergence among cytotypes of *Crataegus* subgenus *Sanguineae*

Species	*C. sukdorfii*	*C. gaylussacia*	*C. sukdorfii*	*C. douglasii*
Cytotype	2x	Auto3x	Allo3x	4x
N individuals (*N* populations)	37 (9)	17 (4)	12 (4)	147 (35)
N alleles	7.562	4.75	5	13.562
Effective *N*	2.593	2.323	2.126	2.229
H_O	0.479	0.389	0.477	0.420
H_S	0.662	0.532	0.537	0.55
H_T	0.682	0.563	0.618	0.657
H_T	0.685	0.574	0.645	0.660
F_{IS}	0.276	0.270	0.113	0.237
Nei's D	0.0018	−0.0028	0.006	0.018

N Individuals (*N* Populations), total number of individuals (number of individuals); *N* alleles, the number of alleles observed; Effective *N*, the number of alleles in a population weighted by their frequency; H_O, observed heterozygosity; H_S, within population heterozygosity; H_T, expected heterozygosity; H_T, expected heterozygosity, corrected for small sample size; F_{IS}, inbreeding coefficient; Nei's D, the average genetic distance between populations with a bias correction for small sample sizes.

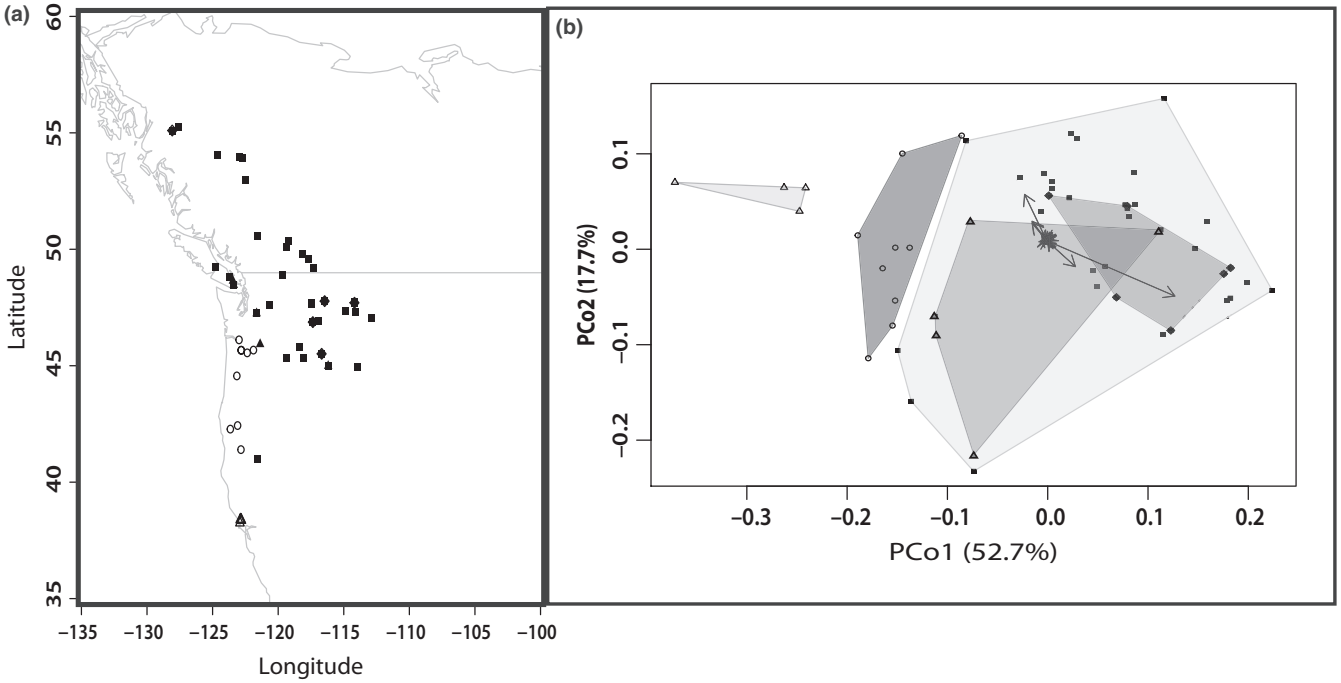


FIGURE 2 (a) Geographic and (b) Ecological distribution (based on PCoA) of field collected individuals. Symbols as in Figure 1: white circles = diploid *C. suksdorfii*, white, triangles = autotriploid *C. gaylussacia*, filled triangles = allotriploid *C. suksdorfii*, filled squares = allotetraploid *C. douglasii*, filled diamonds = pentaploid *C. douglasii*

the species and ploidy level composition of the site. These 13 individuals resided at sites where only *C. douglasii* is found, and the morphological features which could be assessed were consistent with *C. douglasii*. We note that none of these individuals are genetically distinct from *C. douglasii* individuals in which voucher identification was possible.

2.4 | Habitat assessments

In addition to the collecting the specimens described above we surveyed the collection sites to gauge more fine-scale environmental variables and determine if nonclimatic variables contributed substantially to niche divergence and breadth. Following Husband (2000), at each site, we obtained the per cent cover of different vegetation types (i.e., herbaceous, shrub, tree) and substrates (soil, gravel, moss, water) using three randomly placed quadrats.

At each site, a soil core was taken from at least 20 cm before the soil surface to sample the most biologically relevant soil level. Soil samples were housed in air- and watertight containers and were stored in cool, dry conditions until processing. Soil was analysed for soil texture, pH, organic carbon content, colour and water content. Samples were first sieved with a 2-mm mesh to remove debris and

stones. Soil texture was determined by a texture-by-feel test (Arshad, Lowery, & Grossman, 1996). Soil pH was determined on undried soil by making a 1:1 volume:volume slurry of soil and deionized water, pH measurements were taken every 15 min for approximately 60 min, or until readings remained stable through time (Van Lierop & MacKenzie, 1977). Water content was measured by weighing, drying, and reweighing samples, then dividing the dried weight by the wet weight and subtracting from one to attain the per cent water by weight. Organic carbon content was determined using dry samples by burning the samples at 1,000°C for at least 6 hr and measuring the loss on ignition (LOI). Qualitative colour was determined by completing a colour assignment test (Munsell Soil Colour Charts).

As with the herbarium samples, climatic data were collected for 19 World BioClim variables, as well as altitude, and quarterly averages for precipitation, and temperature maximums and minimums (Hijmans et al., 2005). Sites were also measured for their proximity to various disturbance vectors, including major roads and agriculture, as well as nearest aboveground water source. This was performed using Google Earth aerial images and measuring Euclidean distance in ImageJ (National Institute of Health).

We calculated principal coordinates analyses (PCoA) with all ecological variables using a distance matrix calculated using the distance form of Gower's coefficient for mixed data (APE package, R Core Team, 2012). We completed a PCoA for this set of variables, as many of our new ecological variables were multistate categorical, rather than continuous. In this analysis, only the first two components of the PCoA were significant, according to a Broken Stick model, and they accounted for 52.7% and 17.7% of the variance, respectively. PC1 largely accounted for variance in temperature seasonality, and PC2 described variation in annual precipitation.

2.5 | Cytotype determination

For all field collections, cytotype (ploidy level) was determined by flow cytometry of leaf tissue using the method described by Talent and Dickinson (2005). Briefly, samples were run on a FACSCalibur flow cytometer (Becton-Dickinson) soon after collection. We used *Pisum sativum* as an internal standard, added to all samples. Calculations of total DNA content were completed using FlowJo (Tree Star Inc.).

2.6 | DNA extractions and microsatellite amplification

For all field collections, DNA samples were extracted using a standard CTAB extraction protocol (Lo et al., 2009, 2010) then stored at −20°C. Using primers originally developed for *Malus*, a closely related genus (Gianfranceschi, Seglias, Tarchini, Komjanc, & Gessler, 1998; Liebhard et al., 2002), we screened 24 microsatellite loci, including those used by Lo et al. (2009, 2010). Sixteen loci were found to amplify well (see Table S1). Samples were amplified for all 16 microsatellite loci using the following concentrations in a 15 μl reaction: ~20 ng of genomic DNA, 1.5 μl of a high TRIS KCl buffer, 0.2 μM of each dNTP, 0.5 unit of taq polymerase, 4.5 μg of Bovine Serum Albumin, 2.67 μM of forward primer and 4.67 μM of both reverse primer and the fluorescently labelled tag (with NED, VIC or FAM, depending on the loci). The amplification was completed with the following cycles: an initial denaturing period of 94°C for 2 min, then 5 cycles of 94°C for 30 s, 63°C for 45 s and 72°C for 1 min, reducing the annealing temperature by 1°C every cycle to improve specific annealing, followed by 30 cycles of 94°C for 30 s, 58°C for 45 s and 72°C for 1 min, and finally a 10 min final extension period at 72°C. Fluorescently tagged bands were analysed for fragment size using an ABI sequencer (Royal Ontario Museum), and band length peaks were scored by hand using GENEMARKER (ABI software).

2.7 | Population genetic analyses and clone calling

One major hurdle in interpreting genotypic information in polyploids is the problem of allele dosage. For example, if a tetraploid individual possesses the alleles A_1 and A_2, their genotype could be A_1 A_1 A_1 A_2, A_2 A_2 A_2 A_1, or A_1 A_1 A_2 A_2. Two common ways to overcome this problem are either to "fill in" missing genotype information using probabilities that are based on allele frequencies within the population (i.e., Clark & Jasieniuk, 2011; Meirmans, Vlot, Den Nijs, & Menken, 2003), or to convert allelic information to binary form, and treat the microsatellites as dominant markers (Dufresne, Stift, Vergilino, & Mable, 2014). We performed our analyses using both of these techniques, but as both analyses qualitatively agree, we report our findings from the binary data, with the exception of our measurements of diversity (i.e., heterozygosity), which is more intuitive to interpret using the "filled in" ambiguous genotype information.

We used GENODIVE (Meirmans & Van Tienderen, 2004) to calculate measurements of diversity and divergence within each cytotype, using a stepwise-mutation model. We then sought to determine if genetic differentiation occurs along a geographic or ecological axis or neither. We used a mantel test to correlate geographic distance (calculated as the Haversine distance between pairs of latitude and longitude coordinates) and Nei's D genetic distance for each cytotype. Similarly, we also correlated Gower distance based on our ecological variables and Nei's D, but used a partial Mantel test to account for geographic distance. Because microsatellite markers themselves are believed to be relatively neutral, we report these statistics for diploid *C. sukdorfii* for comparison, but acknowledge that these statistics are unlikely to be biologically informative by themselves. On the other hand, we know that polyploids in *Crataegus* subgenus *Sanguineae* are predominantly asexual (Lo et al., 2013); linkages between microsatellite markers and loci under selection are thus unlikely to be disrupted. In this way, the microsatellite markers may be informative in determining correlates of population genetic structure.

We then used GENODIVE to call clonal genotypes within each cytotype. Even clones can differ in multilocus genotypes, simply due to scoring error, or somatic mutations. In long-lived perennials, somatic mutations may contribute substantially to differences between two

individuals of the same clonal genotype. While some cytotypes may display a disjunct distribution, such that there is a small, discrete class of individuals which are more closely related (representing clones), the distribution of pairwise differences may also be more continuous. To address this issue, we called clonal genotypes using a threshold for a maximum number of differences permitted between individuals to still be considered clones. There is little documentation for appropriate threshold calling (Meirmans et al., 2003), but Douhovnikoff and Dodd (2003) proposed a method to objectively set thresholds using estimates of means and standard deviations. We set thresholds by comparing the distributions of the number of clones called at each threshold for diploids (which are obligately outcrossing, and thus should have no clones) and polyploids. These distributions were relatively normal, but tetraploids and triploids (but not diploids), had slightly exaggerated left-hand tails, suggesting an excess of individuals who were very closely related (Figure 3). For allotriploid *C. suksdorfii* and autotriploid *C. gaylussacia*, these distributions showed a discrete class of individuals which were much more genetically similar than the bulk of the sample, and we used this obvious break to call clones. We find 1 clonal genotype for each allotriploid *C. suksdorfii* and autotriploid *C. gaylussacia*.

For cytotypes with a more continuous distribution of pairwise differences (i.e., diploid *C. suksdorfii* and allotetraploid *C. douglasii*), we set the expected 99% confidence intervals for each distribution by calculating three standard deviations from the mean. First, we tested the false positive rate for mis-calling clones in diploid *C. suksdorfii*. In this analysis, three standard deviations from the mean encapsulated almost the entire distribution (99.5% on the lower bound), and when this threshold was used to call "clones" in the diploid data set only two sets of individuals were mis-classified as putative clones. In one of these sets, both individuals resided in the same population, and this population was genetically relatively depauperate. The other set resided approximately 38 km away from each other.

Applying this method to allotetraploid *C. douglasii*, the lower bound of the expected 99% confidence interval excluded a small tail (~2.5%, Fig. S2). When we used this threshold to call clones, we found a total of 94 clonal individuals out of 147 individuals (~64% of individuals). These clonal genotypes consisted of 22 distinct clonal genotypes, with an average of four individuals per clone. All but one population had at least one clone called, and each clonal genotype occurred at three unique geographic locations on average. While *C. douglasii* is self-compatible, these clonal individuals are unlikely to be the result of self-fertilization of the embryo as they were highly heterozygous. In addition, while there likely have been multiple origins in this group (see Section 4), it is unlikely that these putative clonal lineages are the product of identical/very similar parental genotypes recurrently and independently forming genetically identical allotetraploid *C. douglasii* offspring. The probability of recurrently forming the same allotetraploid *C. douglasii* genotypes is simply the probability that multiple identical gametes from diploid *C. suksdorfii* fused with multiple identical gametes of the unknown parent from *C.* subg. *Americanae* (and that this happened for each set of putative clones). While possible, this is highly unlikely, because allelic diversity is relatively high in diploid *C. suksdorfii*, and because this species is obligately outcrossing, we do not expect significant linkage between microsatellite markers from across the genome (and indeed, we see little evidence that this is the case), thus simply drawing the same genotype from a pool of diploid *C. suksdorfii* gametes is unlikely. Indeed, even the two false positive clones that were called for diploid *C. suksdorfii* using our mean and standard deviation approach were more different from each other than the maximum threshold for calling clones in allotetraploid *C. douglasii* individuals (Figure 3), indicating that a low probability of drawing the same diploid *C. suksdorfii* gametes multiple times.

2.8 | Determining the ecological and geographic distributions of clonal genotypes

We then sought to determine the geographic and ecological distributions of clonal genotypes. We focused on clonal genotypes within allotetraploid *C. douglasii*, for two reasons. First, we identified the largest number of clones in the data set for this species and thus had the greatest statistical power to test for differences in the ecological and geographic distributions of these clones. Second, this species is more common and has the widest geographic distribution of all cytotypes within this species complex (Figure 1) and thus may be the most biologically relevant cytotype in which to test how large range size is achieved.

For clonal genotypes that occurred at more than one geographic site we calculated the average pairwise geographic and ecological distances between sites that housed the same clonal genotype (i.e., the same focal clone). We did this using the Haversine distances between the pairs of geographic coordinates (latitude, longitude using the *distHaversine* function in the R package geosphere; Hijmans, 2015), and the Euclidean distances between pairs of scores on PCo1 and PCo2 (using the *dist* function in R). This analysis was performed at the level of site, instead of individual, for two reasons (i) to avoid confounding the potential distribution of clonal genotypes with the fact that a given site is likely to have multiple individuals of the same clonal genotype, and (ii) to determine if clones which did occur at more than one site were occurring in geographically or ecologically proximate locations. To create a distribution of "random" distances, we randomly sampled sets of three allotetraploid sites

FIGURE 3 Threshold histograms for (a) Diploid *C. suksdorfii* (b) autotriploid *C. gaylussacia*, (c) allotriploid *C. suksdorfii* (c) allotetraploid *C. douglasii*. For diploid *C. suksdorfii* and allotetraploid *C. douglasii* thick black bar is the mean, while dashed lines are the expected 2nd *SD*, and dotted lines are the expected 3rd *SD*. For autotriploid *C. gaylussacia* and allotriploid *C. suksdorfii*, dashed line indicates the threshold used to call clones based on the disjunct distributions of pairwise differences. Note that these panels do not occur on the same scale

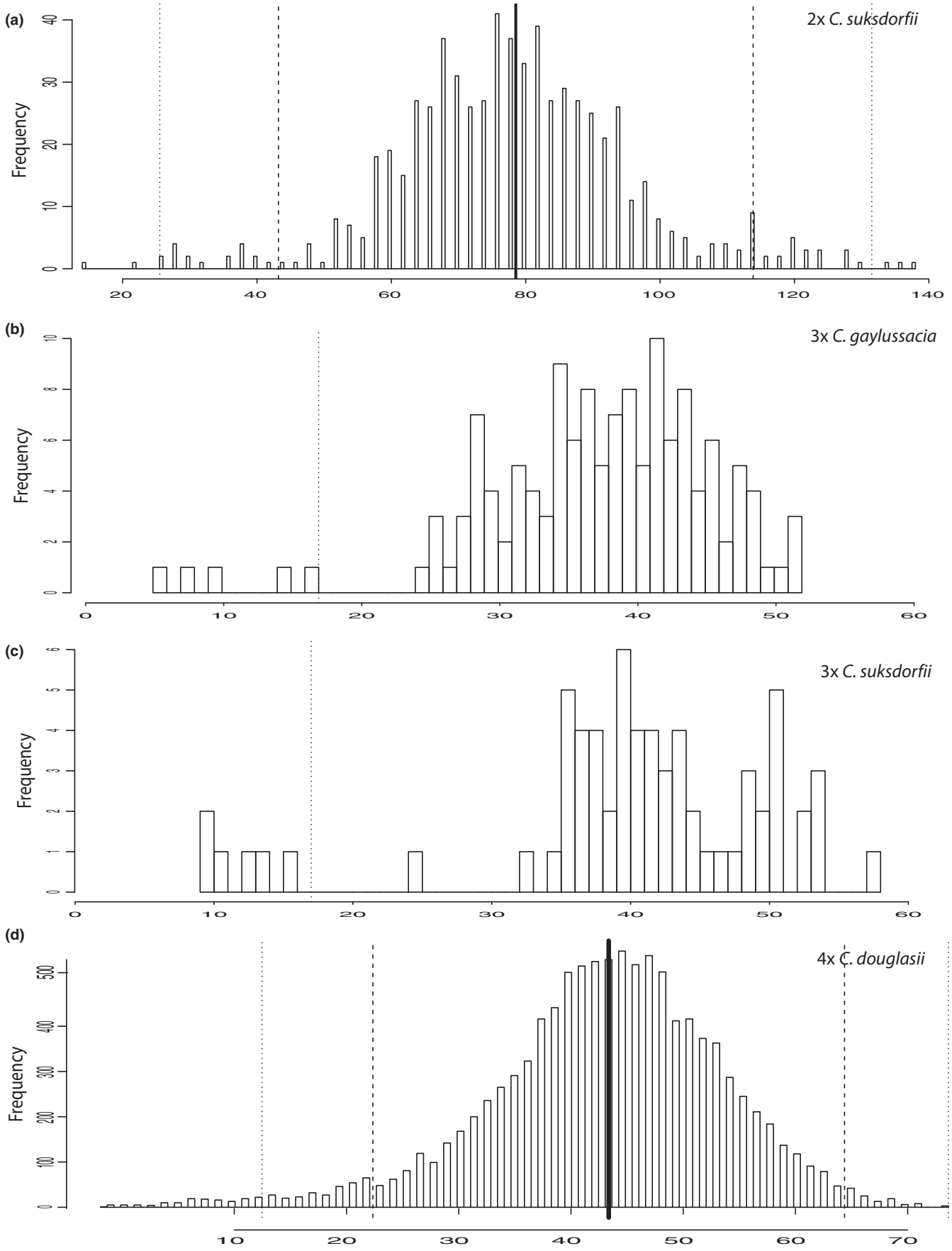

(a)
2x C. suksdorfii
Frequency
0
10
20
30
40
20
40
60
80
100
120
140
(b)
3x C. gaylussacia
Frequency
0
2
4
6
8
10
0
10
20
30
40
50
60
(c)
3x C. suksdorfii
Frequency
0
1
2
3
4
5
6
0
10
20
30
40
50
60
(d)
4x C. douglasii
Frequency
0
100
200
300
400
500
10
20
30
40
50
60
70

FIGURE 4 Isolation by distance and isolation by ecological distance, respectively, for field collected individuals based on microsatellite markers: (a, b) Diploid *C. suksdorfii*; (c, d) Autotriploid *C. gaylussacia*; (e, f) Allotriploid *C. suksdorfii* (g, h) Allotetraploid *C. douglasii*. Lines are the regression between genetic dissimilarity and either geographic or ecological dissimilarity, and only appear on comparisons that were significant or nearly significant based on mantel tests for geography and partial mantel tests for ecology. Note the scales differ in each panel

without replacement and calculated the average geographic or ecological distance among them using the distances described above. We chose sets of three sites to match the average number of sites a given clonal genotype was found at. We completed this random sampling 1,000 times, then used a Kruskal–Wallis test to determine if the average geographic or ecological distance between randomly chosen points and individuals from a given clonal genotype differed.

While many clonal genotypes occurred at multiple sites, 27% of clonal genotypes occurred only at one site. These clones may represent more ecologically specialized clones, and thus, we aimed to assess whether the ecological and geographic distribution of these geographically restricted clones occurred in ecologically extreme habitats, or at higher latitudes or altitudes, relative to the bulk of the allotetraploid *C. douglasii* sample.

Lastly, we correlated measurements of clonal diversity with both latitude and altitude to determine if clonal diversity varies spatially. Specifically, we asked whether populations at the range edge are less genetically diverse than the range centre, which could indicate either a higher propensity for asexual reproduction, or a filtering of unfit genotypes closer to the edge, consistent with the Frozen Niche hypothesis.

3 | RESULTS

3.1 | Geographic distributions

Using range-wide samples from the Royal Ontario Museum *Crataegus* database, we find that *Crataegus* subg. *Sanguineae* display patterns of Geographic Parthenogenesis, wherein allotetraploid *C. douglasii* displays the widest distribution, allopolyploids exhibit an intermediate distribution, and autotriploid *C. gaylussacia* and diploid *C. suksdorfii* exhibit the smallest distribution (Figure 1), which is consistent with previous research (see Coughlan et al., 2014; Lo et al., 2013; McGoey, Chau, & Dickinson, 2014). We also show the novel result that the distribution of *C. douglasii* is also intermediate to its two progenitors; diploid *C. suksdorfii* and a member of *C.* subg. *Americanae*—though it is larger than either (Figure 1).

3.2 | Ecological distributions

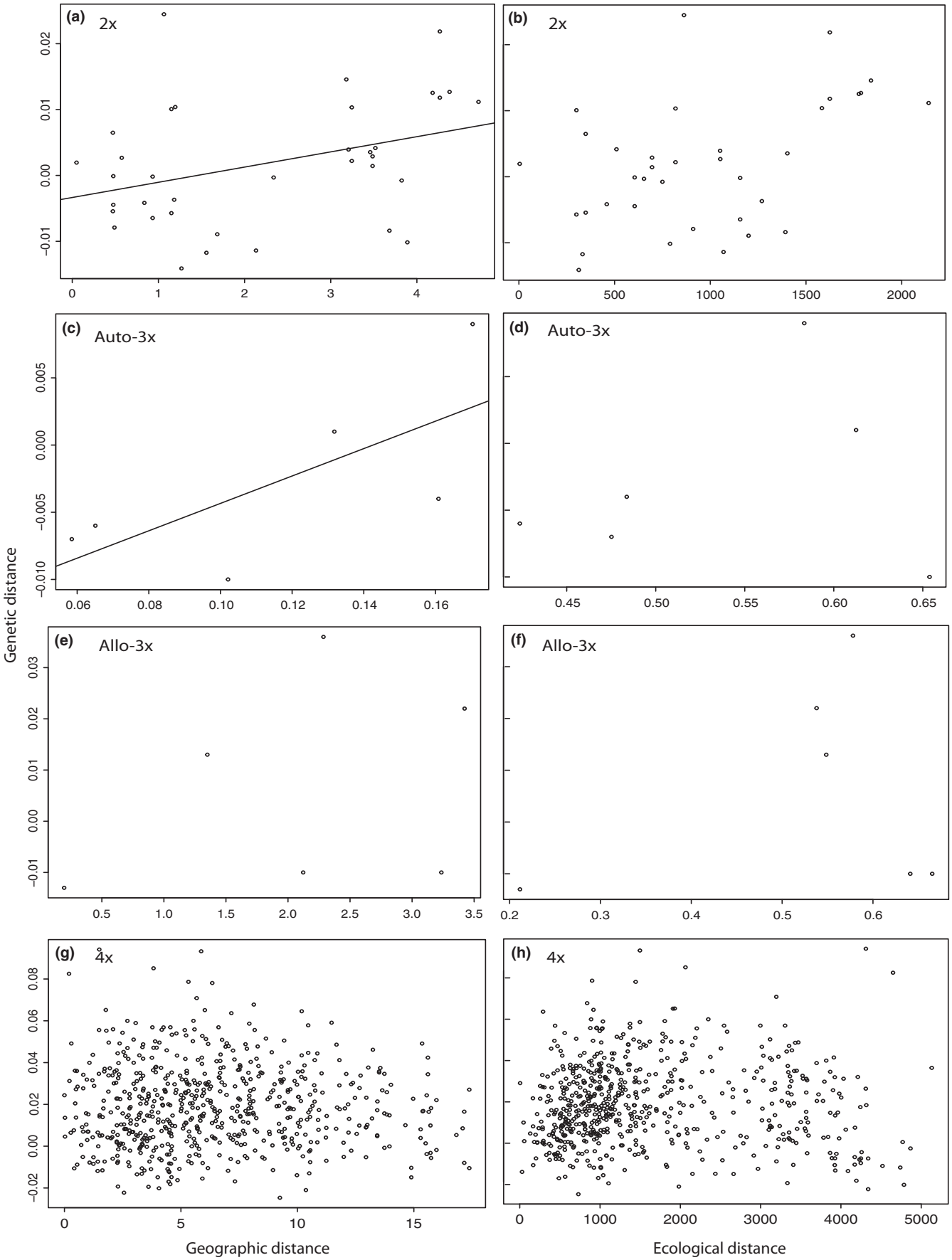
In both the range-wide Herbarium collections as well as our own field collections, cytotypes varied in niche breadth, wherein allopolyploids tended to have larger niches than either autopolyploids or diploids for both the herbarium samples and our field collections (Figures 1 and 2, respectively). Allotetraploid *C. douglasii* has a substantially wider ecological distribution than any of its putative parental species (Figures 1 and 2). Most cytotypes which arose from subsequent backcrossing display niche sizes that are intermediate relative to their parents (as is the case for allotriploid *C. suksdorfii*, the putative backcross between allotetraploid *C. douglasii* and diploid *C. suksdorfii*, and pentaploid *C. douglasii*, the putative backcross between allotetraploid *C. douglasii* and allotriploid *C. suksdorfii*). In contrast, autotriploid *C. gaylussacia* has a substantially smaller niche than any other cytotype in the complex (Figure 1). Thus, polyploidization in itself is not sufficient to explain niche expansion in Pacific Northwest Hawthorns.

Cytotypes also varied in their position within the climate niche space ($F = 196.1$, $df = 10$, $p < .0001$), particularly along PC1 (Figure 1). At one extreme, the Mediterranean climate of the California coastal steppe, inhabited by autotriploid *C. gaylussacia* sites was reflected in their generally warmer conditions, greater diurnal variation in temperature and greater precipitation seasonality. This niche was distinct from that of its sexual, progenitor, diploid *C. suksdorfii*, which inhabits the Marine and Marine Mountain climates of western Oregon and Washington (also summer-dry, with abundant winter precipitation; Bailey, 1997). On the other extreme, *Crataegus*. subg. *Americanae* reside in habitats that are generally cooler, especially in the winter, more variable in temperature, and drier. Though larger than either parental niche, allotetraploid *C. douglasii* inhabits a niche which is intermediate to both of its putative parental species. Along PC1, this niche is distinct from diploid *C. suksdorfii*, as allotetraploid *C. douglasii* is commonly found in drier and cold environments such as the Temperate Steppe and Temperate Desert climate divisions and their montane subdivisions, but overlaps with *Crataegus*. subg. *Americanae*. The climate niches of the allopolyploid *C. suksdorfii* and pentaploid *C. douglasii* are largely overlapping with allotetraploid *C. douglasii*, although all three are intermediate to their respective, putative parental species.

Results from the PCoA based on field collected sites generally recapitulate the PCA results from herbarium records (Figure 2). Autotriploid *C. gaylussacia* and diploid *C. suksdorfii* display distinct niches relative to all allopolyploids in the complex, with autotriploid *C. gaylussacia* occurring in hot, xeric, climatically stable sites, diploid *C. suksdorfii* occurring in much more mesic sites, and all allopolyploids occurring in cooler, drier, higher elevation and more temperately seasonal sites (Figure 2). While the main drivers of niche divergence among cytotypes were climatic, soil profile did differ between the habitat of autotriploid *C. gaylussacia* and all other cytotypes, such that *C. gaylussacia* occurs in more acidic soils (Fig. S3). No other nonclimate ecological variable was important in distinguishing cytotype niches.

3.3 | Patterns of genetic diversity among cytotypes and across space

Genotype data from our field collected individuals show that both allotetraploid *C. douglasii* and diploid *C. suksdorfii* have similar levels


(a) 2x
(b) 2x
(c) Auto-3x
(d) Auto-3x
(e) Allo-3x
(f) Allo-3x
(g) 4x
(h) 4x
Genetic distance
Geographic distance
Ecological distance

of allelic diversity (Table 1), while both allotriploid *C. suksdorfii* and autotriploid *C. gaylussacia* are generally less diverse. Individuals from autotriploid *C. gaylussacia*, allotriploid *C. suksdorfii* and allotetraploid *C. douglasii* exhibit a smaller number of average pairwise differences than individuals of diploid *C. suksdorfii* (Figure 3), likely in part because of the presence of clonal genotypes in these groups. Allotetraploid *C. douglasii* and allotriploid *C. suksdorfii* also exhibited much higher genetic distance among populations than diploid *C. suksdorfii*, while autotriploid *C. gaylussacia* populations showed very little genetic divergence (Table 1).

We find no evidence for isolation by geographic distance (IBD) across the range for allotetraploid *C. douglasii* ($r = -.02143$, $p = .6$; Figure 4) or allotriploid *C. suksdorfii* ($r = -.203$, $p = .9$), but we do for diploid *C. suksdorfii* ($r = .3608$, $p = .05$), and we find nearly significant isolation by distance for autotriploid *C. gaylussacia* ($r = .71$, $p = .1$).

We find evidence of several clonal lineages in allotetraploid *C. douglasii*. Approximately two-third of individuals (94 of 147) were assigned to one of 22 clonal genotypes, with an average of four individuals per clone. While approximately one-third of the individuals were found to be genetically unique, this does not mean that they are the product of sexual reproduction, and may represent clonal lineages which are rarer or older (and thus are harder to find or have accumulated more mutations, respectively). These results are also consistent with Lo et al. (2013) which found that all seeds sampled from allotetraploid *C. douglasii* were the product of asexual reproduction. We also find that these clonal lineages are generally well dispersed. Of the 94 known clonal individuals, 73% of individuals occurred at more than one site, with an average of three sites per clone. Thus, while there was no single clonal genotype which was widespread across the landscape, widespread clones comprise about half of the total data set (69 individuals of 147 belonged to clonal lineages with occurred at more than one site).

In addition, clonal genotypes of allotetraploid *C. douglasii* which occurred at two or more sites are widely dispersed across the landscape. While they are, on average, slightly less geographically

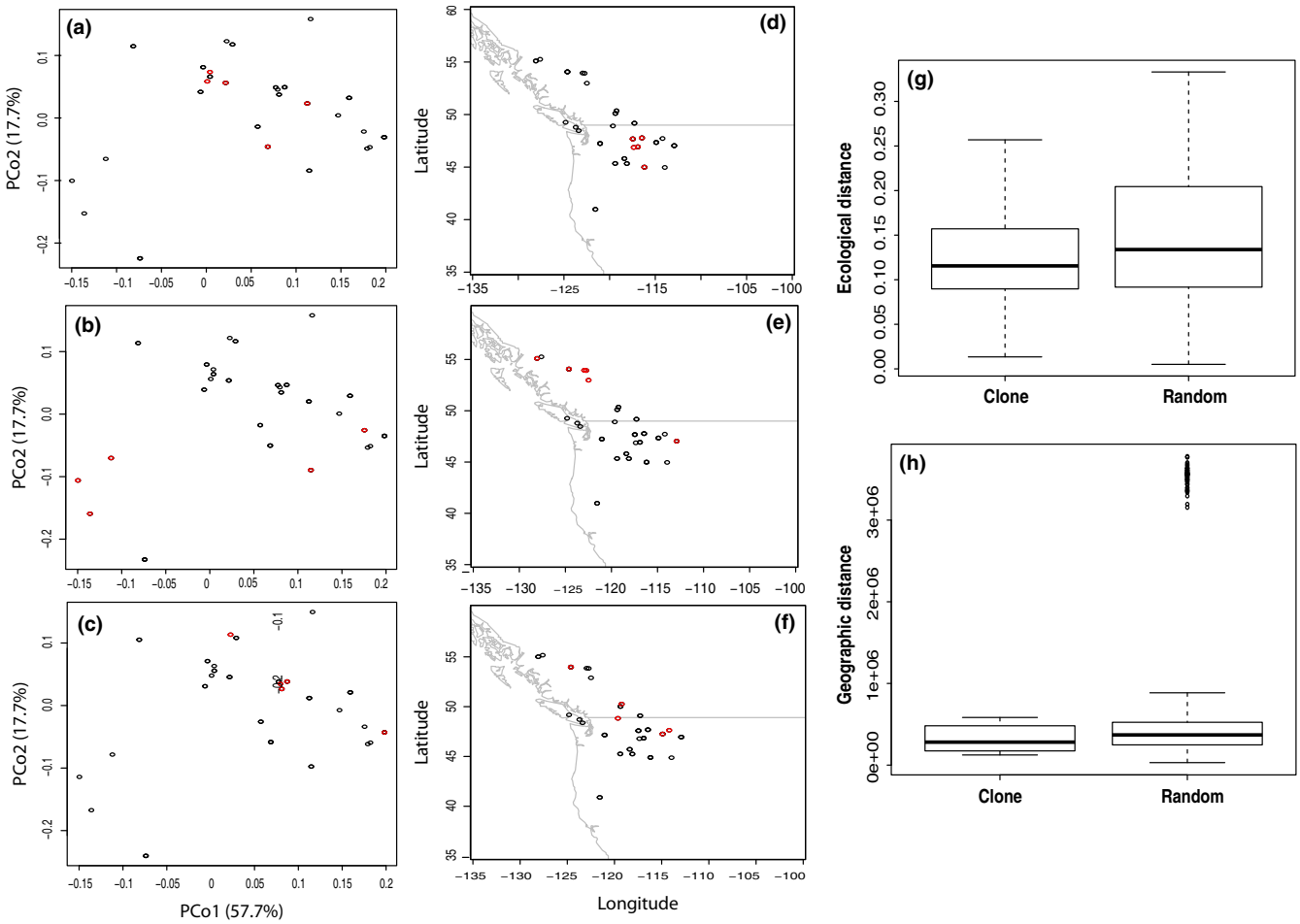


FIGURE 5 (a–f) Three representative examples of the geographic and ecological distribution for clonal genotypes which were found to occur at more than one site for the field collected individuals. Red circles denote sites which housed the same focal clone genotype and black circles denote the remainder of the sample. (g) The average ecological distance among random sites vs sites which housed the same clonal genotype. (h) The average geographic distance among random sites vs sites which housed the same clonal genotype. See Fig. S2 for all clones. All distance measurements, as well as distributions, are based on PCoA analyses completed on the field collected individuals, as in Figure 2 [Colour figure can be viewed at wileyonlinelibrary.com]

dispersed as a random subset of sites, this difference is barely significant ($\chi^2 = 5.1632$, $df = 1$, $p = .0231$, with p .025 being the significance threshold after multiple comparisons; Figure 5). The most common clonal genotype called accounted for 8% of the allotetraploid sample and occurred at 17% of the sampled sites. While on average, a given clonal genotype was well dispersed, 27% clonal genotypes occurred at only one site. While broadly dispersed clonal genotypes are relatively common in allotetraploid *C. douglasii*, some clonal genotypes remain more isolated.

We detect one clonal genotype in each autotriploid *C. gaylussacia* and allotriploid *C. sukdorfii*. In the case of allotriploid *C. suksdorfii*, this clone comprises four individuals and is found at two of the four sites samples, and occur approximately 25 km apart from each other. For autotriploid *C. gaylussacia*, the single clonal genotype comprises three individuals which each occurred at three of four sampled sites (which are all <8 km away from each other). We refrain from further statistical tests of their distributions, given the small sample sizes for these cytotypes.

3.4 | Distribution of genotypes in ecological space

We also find no evidence for isolation by ecological distance, while controlling for geographic distance in any cytotype (diploid *C. suksdorfii*: $r = .3163$, $p = .14$; autotriploid *C. gaylussacia* $r = -.1213$, $p = .5$; allotriploid *C. suksdorfii* $r = -.0175$, $p = .58$; allotetraploid *C. douglasii*: $r = .005337$, $p = .46$; Figure 4). This suggests that successful dispersal and colonization of genotypes are limited by geography more than ecology in both diploids and autotriploids, but by neither geography or ecology in allopolyploids in this group.

In addition, clonal genotypes of allotetraploid *C. douglasii* are not clustered in ecological space. In fact, on average, clonal genotypes are found in ecologically variable habitats (Figures 5, S1) and have an equally variable ecological distribution as a random sample of allotetraploid *C. douglasii* sites ($\chi^2 = 0.6989$, $df = 1$, $p = .4032$; Figure 5). Of the clonal genotypes which do occur at a single site, these sites were no more ecologically extreme than the bulk of the allotetraploid *C. douglasii* range (in terms of their position on PCo1: $\chi^2 = 0.25$, $df = 1$, $p = .61$, their latitude: $\chi^2 = 0.25$, $df = 1$, $p = .61$ or altitude: $\chi^2 = 0.22$, $df = 1$, $p = .64$), suggesting that potentially more specialized clones are not more likely to be found at the edge of the niche or range.

We find no evidence for geographic differences in clonal diversity. The proportion clonal individuals at a given site was not correlated with latitude ($r = -0.07$, $df = 33$, $p = .66$) or altitude ($r = .16$, $df = 32$, $p = .34$). Thus, it appears that clonal reproduction, clonal success, or both are not more likely to occur at higher latitudes or altitudes within the range of allotetraploid *C. douglasii*.

4 | DISCUSSION

We sought to examine whether the increase in range size associated with asexual allopolyploids was also associated with an increase in niche breadth at range-wide scales, and moreover what potential genetic mechanisms may allow for this expansion. Here, we show that while all polyploids exhibit some degree of niche divergence, only allopolyploids exhibit a broader range than that of diploid *C. suksdorfii*, suggesting a potential role of hybridity in niche expansion in this group. We find no evidence of geographic structuring within allotetraploid *C. douglasii*, as would be expected if niche expansion were a by-product of multiple origins. We also do not find evidence that allotetraploid *C. douglasii* genotypes are ecologically clustered in either the overall sample or within a given clonal genotype, as would be expected under the frozen niche hypothesis. However, we find clonal genotypes which are ecologically and geographically widespread, consistent with the generalist genotype hypothesis. We interpret these findings to suggest that high dispersal rates of potentially generalist clones may contribute to niche expansion in allopolyploids of this group.

4.1 | Niche divergence is associated with all polyploids, while niche expansion is associated only with allopolyploids

In North American *Crataegus* subgenus *Sanguineae*, the niches of all polyploids diverge from that of the diploids. Niche differentiation is a relatively common (though not ubiquitous) consequence of the evolution of asexual polyploids (Soltis, Visger, Marchant, & Soltis, 2016), and is thought to be an important barrier to both reproduction and competition among cytotypes, that could facilitate the maintenance of cytotype diversity (Fowler & Levin, 1984; Ramsey & Schemske, 1998). In our results, both auto- and allopolyploids differed substantially from diploids in niche position, and this differentiation was primarily driven by both temperatures, precipitation and altitude (Figures 1, 2, and S2). The relative positions of these cytotypes along PC1/PCo1, however, are likely influenced by the genetic composition of the polyploids involved. All *C.* subg. *Sanguineae* allopolyploids were generally found in cooler and more seasonally variable habitats than diploid *C. suksdorfii*, while autotriploids were found in warmer, drier and less seasonally variable habitats than diploid *C. suksdorfii*. Hybrids tended to fill niche positions which were intermediate to their parents. This is witnessed for allotetraploid *C. douglasii* (putative hybrid of diploid *C. suksdorfii* × *Crataegus* subgenus *Americanae*), as well as allotriploid *C. suksdorfii* (putative backcrosses between diploid *C. suksdorfii* and allotetraploid *C. douglasii*) and pentaploid *C. douglasii* (putative backcrosses between allotriploid *C. suksdorfii* and allotetraploid *C. douglasii*) (Figure 1). In contrast, the niche divergence between diploid *C. suksdorfii* and autotriploid *C. gaylussacia* cannot be explained by hybridization. Polyploidization, apomixis, or their combination may have caused this shift, or patterns of niche divergence between diploid *C. suksdorfii* and autotriploid *C. gaylussacia* may simply be due to historical contingency, local adaptation and limited dispersal among these sites. In *Achillea borealis*, niche divergence among cytotypes is, in part, a direct consequence of polyploidization, and in part due to subsequent local adaptation (Ramsey, 2011). Similarly, in *Boechera*, in which diploid apomicts occur, ploidy, rather than reproductive system, appears to drive niche divergence (Mau et al., 2015).

In *Crataegus* subgenus *Sanguineae*, it appears that polyploidization and mating system shift are sufficient to cause niche differentiation, but the extent and direction of niche divergence may be related to the genetic composition of taxa involved.

In contrast to niche divergence, increased niche breadth relative to diploids is found only in North American allopolyploid *C.* subg. *Sanguineae* taxa, with allotetraploid *C. douglasii* exhibiting the largest niche (Figure 1). This is perhaps not surprising, as allopolyploids in this complex also have markedly larger geographic ranges and have been shown to exhibit the pattern of GP (Coughlan et al., 2014). Stebbins (1940) compared the environmental and geographic distributions of polyploid and diploid *Crepis*, and since then, many other researchers have also shown a wider range of habitats associated with apomictic polyploids compared with their diploid progenitors (i.e., Bierzychudek, 1985; Hancock & Bringhurst, 1981; Lumaret et al., 1987; Meimberg et al., 2009). In the case of *Crataegus* subgenus *Sanguineae*, niche expansion appears to be a consequence of hybridization, either in concert with or independent of apomixis and polyploidization, as allopolyploids have a broader niche than both autopolyploid *C. gaylussacia* and diploid *C. suksdorfii*.

Hybridization has been hypothesized to play an important role in niche expansion, as it can generate substantial genetic diversity as all individuals are fixed heterozygotes at every locus (Kearney, 2005). That said, many cases of niche expansion have been found in autopolyploids (i.e., Bayer, 1991; Yahara, 1990). In *Cardamine*, allopolyploids are able to grow equally well in either parental habitat by nonrandomly expressing the appropriate parental allele (Shimizu-Inatsugi et al., 2017). In *Crataegus* subgenus *Sanguineae*, hybridization has likely played an important role in niche expansion. Allotetraploid *C. douglasii* is the product of hybridization between two relatively disparate groups—that is diploid *C. suksdorfii* (*C.* subg. *Sanguineae*), which is found on in mesic environments in the Pacific NW, and an unknown member of *Crataegus* subg. *Americanae*, extant members of which inhabit sites that are cooler, more variable in temperature, and drier than those occupied by *C. douglasii* (Zarrei et al., 2014; Figures 1, 2 and S2). *Crataegus douglasii* is able to occur in habitats which span the range between its putative parental niches, and thus increasing the breadth of its niche by more than double than either of the putative parental species. Next, we sought to understand the genetic mechanism of this extreme niche expansion in allotetraploid *C. douglasii*.

4.2 | Geographic structuring of populations, and the role of multiple origins in patterns of geographic parthenogenesis

We show that allopolyploids exhibit no isolation by distance and no geographic structuring of clonal genotypes (Figures 4 and 5). This is in contrast to diploid *C. suksdorfii* and autotriploid *C. gaylussacia*, which exhibit some geographic structuring among populations (Figure 4). We interpret these results to suggest that allopolyploids have a higher dispersal and/or colonization ability relative to diploids or autopolyploids in this group. While a lack of structuring has been found in allotetraploid *C. douglasii* across much smaller geographic spaces with fewer sampled populations (i.e., Lo et al., 2009), we recapitulate this pattern with a range-wide sampling effort comprising many more populations and larger geographic spaces. We are the first to report a lack of population structuring both allotriploid *C. suksdorfii* and allotetraploid *C. douglasii* across the range, but evidence for IBD in taxa of nonhybrid origin, as is found in autotriploid *C. gaylussacia* and diploid *C. sukdorfii*. This is congruent with having found that allopolyploids tend to invest proportionally more in fruit mass (dispersal-related), while diploids and autotriploids invest more in seeds (competition-related; Coughlan et al., 2014). This potentially greater ability to disperse and colonize new geographic locations can increase range size (Lester, Ruttenberg, Gaines, & Kinlan, 2007), and so expose a cytotype to a wider range of niches.

Given this lack of geographic structure to our population genetic data for allopolyploids it seems unlikely that niche expansion, at least in allotetraploid *C. douglasii*, is simply a by-product of multiple origins. While we cannot determine how many times allotetraploids have arisen, given the poor phylogenetic resolution in this clade (Zarrei et al., 2014), but there is no evidence of multiple genetic clusters associating with novel portions of the range or niche, unlike in *Plantago media* (Van Dijk & Bakx-Schotman, 1997) or Easter daisies (Thompson & Whitton, 2006). This is not to say that multiple origins of each cytotype have been unimportant in the creation of multiple clonal genotypes which may readily colonize novel habitats, but simply that genotypes in this group appear to be highly dispersive and geographically distinct populations do not appear to be evolutionarily distinct lineages.

4.3 | Ecologically generalist genotypes may contribute to patterns of GP and niche expansion in allotetraploid *C. douglasii*

For cytotypes that show patterns of niche expansion (i.e., allopolyploids), we do not find evidence of ecological clustering of genotypes. There was no cline in genetic similarity based on ecological similarity for any cytotype in this study, but patterns of genetic similarity can be explained by geography for both diploid *C. suksdorfii* and autotriploid *C. gaylussacia*, as discussed above. We note that microsatellites themselves are neutral markers and in the case of sexually reproducing, randomly mating taxa will likely not be correlated with ecological gradients even when populations are locally adapted. However, given that all polyploids in this complex are highly apomictic (Lo et al., 2013), putatively neutral markers are expected to behave non-neutrally, due to close linkage with loci that are under selection. Thus, a lack of both geographic and ecological structuring of population genetic similarity highlights a potentially strong role of dispersal and colonization ability in allopolyploids in this group.

We find that evidence of ecologically and geographically widespread clonal lineages in allotetraploid *C. douglasii*. While no single clone is responsible for this pattern, widespread clones as a class account for approximately half of the individuals sampled. On average, clonal genotypes for allotetraploid *C. douglasii* were not clustered in ecological space, as would be expected if most clonal

genotypes were specialized to particular environmental conditions. We did find that ~27% of clones were geographically restricted, and occurred at only one site. These spatially restricted clones could be examples of ecologically specialized clones, consistent with the Frozen Niche hypothesis (Vrijenhoek, 1984). However, sites which housed unique clones were not more environmentally extreme than the bulk of the allotetraploid *C. douglasii* sites, and these sites often also housed generalist clones. Spatially restricted clones may also be relatively new clonal genotypes which have not had time to disperse across the landscape, or clones which indeed are more broadly distributed, but were un-sampled at other sites. Along these lines, it was also common for the same clonal "generalist" genotype to be found multiple times at the same site. This simply highlights that despite the putatively high dispersing capacity of allotetraploid *C. douglasii*, seeds produced at a given site are more likely to disperse very short distances, while long distance dispersal events should be relatively less common in any given year (Nathan, 2006), but across years these long distance dispersal events could contribute greatly to eroding patterns of isolation by distance.

In contrast to many other studies, we find that ecologically widespread clonal genotypes are relatively common in allotetraploid *C. douglasii*. Sixty-four per cent of allotetraploid individuals were found to belong to one of 22 clonal genotypes (94 of 147 individuals). When clones occurred at more than one geographic site, those sites were not ecologically clustered, relative to the range of allotetraploid *C. douglasii*, suggesting that the same clonal genotype can survive in multiple environments. The frequency and distribution of these clonal genotypes may suggest that allotetraploid *C. douglasii* is composed, in part, by generalist clones. As these plants are long-lived and slow growing, demographic studies and accurate estimates of local fitness are challenging to obtain, and thus, we cannot provide estimates of local fitness across habitats for each clone. However, all individuals sampled were full, adult trees, and thus had survived local selective sieves acting on colonizing seeds, and seedlings. In addition, the trees that were revisited in summer of 2011 all produced fruit and thus are at least able to reproduce.

Support for the general-purpose genotype has been relatively low. Many authors have found that certain clonal genotypes are able to maintain relatively stable fitness measurements across environmental gradients relative to sexual diploids (Bierzychudek, 1989; Hancock & Bringhurst, 1981; Michaels & Bazzaz, 1989), and Oplaat and Verhoeven (2015) have found that generalist clonal genotypes are associated with the leading range edge, while more specialized clonal genotypes tend to be from closer to the range centre in *Taraxacum*. However, if generalist genotypes are contributing to patterns of Geographic Parthenogenesis, these clonal genotypes should be widespread (or at the leading edge, as in Oplaat & Verhoeven, 2015). Very few studies have found widespread clonal genotypes, and when found, these genotypes are often uncommon relative to spatially restricted ones (i.e., Van Dijk, 2003). Here, we show that putatively "generalist" clonal genotypes are widespread and occur in a variety of environments in allotetraploid *C. douglasii*. These clonal genotypes are not found in more extreme environments, or within the range edges (such as Oplaat & Verhoeven, 2015). We suggest that these putatively generalist clonal genotypes may be related to the fact that these allopolyploids are the result of hybridization between two relatively divergent taxa, and thus, these clones may have an increased ability to tolerate habitats ranging from that of *C. suksdorfii* to those of widespread members of *Crataegus* subg. *Americanae* like *C. chrysocarpa* and *C. macracantha*, as well as potentially novel environments. However, more information about the evolutionary history of these clones, as well as measurements of their local fitness are needed to understand the mechanism by which hybridization is contributing to patterns of GP in this group, potentially through these generalist clones.

5 | CONCLUSIONS

We have described patterns of niche divergence and breadth across cytotypes in Pacific Northwest members of *Crataegus* subg. *Sanguineae*. We find that while all polyploids display niche divergence from diploids, only allopolyploids exhibit a broadening of niche. In addition, these allopolyploids show no patterns of isolation by distance, while diploids and autotriploids do, suggesting a difference in dispersal/colonization dynamics between these taxa. Lastly, we emphasize the overall pattern that allotetraploid *C. douglasii* has geographically and ecologically widespread clonal genotypes throughout its range. Taken together, we infer that patterns of niche expansion of geographic parthenogenesis in this group may, in part, be related to highly dispersive, hybrid generalist clonal genotypes.

ACKNOWLEDGEMENTS

We thank C. Shaw, H. Moothoo and M. Zarrei for help with fieldwork. We acknowledge The Cunningham Marsh Preservation Committee, Siskyiyou Field Institute, D. Smith, W. Earle, M. Graham, B. Jensen, G. Cooley, R. Thompson, J. Herrick, R. & M. Johnson, A. Anderson, T. Steen, G. Powell, R. & J. Pojar, C. Delong, B. Rogers, H. Massicotte, L. Tackaberry and J. Lee for information concerning and access to collection sites. We thank N. Talent for flow cytometry expertise, R. Martin for laboratory assistance, and the Green Plant Herbarium for voucher and database organization. A special thanks to C. Rushworth for copious discussions of this work. The Natural Sciences and Engineering Research Council of Canada provided financial support through Discovery Grants (A3430 to T.A.D. and 326439 to S.S.) and Strategic Project Grant 'Hawthorn as a new Canadian agroforestry crop for functional foods and nutraceuticals', STPGP 381073–09 to T.A.D., S.S., P. Shipley (UBC–Okanagan), S. Proctor (University of Alberta) and the Naturally Grown Herb and Spice Producers Co-op. The University of Toronto Department of Ecology and Evolutionary Biology and the Royal Ontario Museum provided additional financial support.

DATA ACCESSIBILITY

Data for this study have been uploaded to Dryad: https://doi.org/10.5061/dryad.75252. Ecological and geographic distributions of *C.*

subg. *Sanguineae* and *C.* subg. *Americanae* based on the database of collections: Crataegus subg Americanae + sanguinae_localities and bioclim data.xlxs. Ecological and geographic distributions data based on the 2010–2014 field collections: Crataegus subg. Sanguineae_collection sites + ecological variables.xlxs. Microsatellite data for the field collections of *C.* subg. *Sanguineae* in binary form, formatted for GENODIVE: Crataegus subg. Sanguineae_microsatellite data_binary form for genodive.xlxs.

AUTHOR CONTRIBUTION

J.M.C. conceived the ideas for this project, completed the field and laboratory work, completed the analyses, and wrote the manuscript. S.H. helped complete laboratory work. S.S. and T.A.D. contributed substantially to the development of ideas and to writing the manuscript.

REFERENCES

Arshad, M. A., Lowery, B., & Grossman, B. (1996). Physical tests for monitoring soil quality. In J. W. Doran, & A. J. Jones (Eds.), *Methods for assessing soil quality*. Soil Sci. Soc. Am. Spec. Publ. 49 (pp. 123–142). Madison, WI: SSSA.

Bailey, R. G. (1997). *Ecoregions of North America*. Washington DC: USDA - Forest Service.

Baker, H. G. (1955). Self-compatibility and establishment after "long-distance" dispersal. *Evolution*, *9*, 347–348.

Baker, H. G. (1965). Characteristics and modes of origin of weeds. In H. G. Baker, & G. L. Stebbins (Eds.), *Genetics of colonizing species* (pp. 137–172). New York: Academic Press.

Bayer, R. J. (1991). Allozymic and morphological variation in *Antennaria* (Asteraceae: Inuleae) from the low arctic of Northwestern North America. *Systematic Botany*, *16*(3), 492–506.

Bierzychudek, P. (1985). Patterns in plant parthenogenesis. *Experientia*, *41*, 1255–1264.

Bierzychudek, P. (1989). Environmental sensitivity of sexual and Apomictic *Antennaria*: Do apomicts have general-purpose genotypes? *Evolution*, *43*(7), 1456–1466.

Bretagnolle, F., & Thompson, J. D. (2001). Phenotypic plasticity in sympatric diploid and autotetraploid *Dactylis glomerata*. *International Journal of Plant Sciences*, *162*(2), 309–316.

Campbell, C. S., Greene, C. W., & Dickinson, T. A. (1991). Reproductive biology in subfamily Maloideae (Rosaceae). *Systematic Botany*, *16*, 333–349.

Clark, L. V., & Jasieniuk, M. (2011). POLYSAT: An R package for polyploid microsatellite analysis. *Molecular Ecology Resources*, *11*(3), 562–566.

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews*, *6*(11), 836–846.

Cosendai, A. C., Wagner, J., Ladinig, U., Rosche, C., & Hörandl, E. (2013). Geographical parthenogenesis and population genetic structure in the alpine species *Ranunculus kuepferi* (Ranunculaceae). *Heredity*, *110*(6), 560–569.

Coughlan, J. M., Stefanović, S., & Dickinson, T. A. (2014). Relative resource allocation to dispersal and competition demonstrates the putative role of hybridity in geographical parthenogenesis. *Journal of Biogeography*, *41*(8), 1603–1613.

De Storme, N., & Mason, A. (2014). Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance. *Current Plant Biology*, *1*, 10–33.

Dickinson, T. A., Lo, E. Y. Y., & Talent, N. (2007). Polyploidy, reproductive biology, and Rosaceae: Understanding evolution and making classifications. *Plant Systematics and Evolution*, *266*(1–2), 59–78.

Dickinson, T. A., Lo, E. Y. Y., Talent, N., & Love, R. M. (2008). Black-fruited hawthorns of western North America—one or more agamic complexes? *Botany-Botanique*, *86*(8), 846–865.

Douhovnikoff, V., & Dodd, R. S. (2003). Intra-clonal variation and a similarity threshold for identification of clones: Application to Salix exigua using AFLP molecular markers. *Theoretical and Applied Genetics*, *106*(7), 1307–1315.

Dufresne, F., Stift, M., Vergilino, R., & Mable, B. K. (2014). Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology*, *23*, 40–69.

Ellstrand, N. C., & Roose, M. L. (1987). Patterns of genotypic diversity in clonal plant species. *American Journal of Botany*, *74*(1), 123–131.

Fowler, N. L., & Levin, D. A. (1984). Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *The American Naturalist*, *124*(5), 703–711.

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

Hancock, J. F., & Bringhurst, R. S. (1981). Evolution in California populations of diploid and octoploid *Fragaria* (Rosaceae): A comparison. *American Journal of Botany*, *68*(1), 1–5.

Hijmans, R. J. (2015). geosphere: Spherical trigonometry. R package version 1.5-1. http://CRAN.R-project.org/package=geosphere

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978.

Hijmans, R. J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A., & Spooner, D. M. (2007). Geographical and environmental range expansion through polyploidy in wild potatoes (Solanum section Petota). *Global Ecology and Biogeography*, *16*(4), 485–495.

Hörandl, E. (2006). The complex causality of geographical parthenogenesis. *The New Phytologist*, *171*(3), 525–538.

Husband, B. C. (2000). Constraints on polyploid evolution: A test of the minority cytotype exclusion principle. *Proceedings of the Royal Society London Biology*, *267*(1440), 217–223.

Kearney, M. (2005). Hybridization, glaciation and geographical parthenogenesis. *Trends in Ecology & Evolution*, *20*(9), 495–502.

Legendre, P., & Legendre, L. (1998). *Numerical ecology: Developments in environmental modelling 20*, 2nd ed. Amsterdam, The Netherlands: Elsevier Scientific Publishing Company.

Lester, S. E., Ruttenberg, B. I., Gaines, S. D., & Kinlan, B. P. (2007). The relationship between dispersal ability and geographic range size. *Ecology Letters*, *10*(8), 745–758.

Liebhard, R., Gianfranceschi, L., Koller, B., Ryder, C. D., Tarchini, R., Van De Weg, E., & Gessler, C. (2002). Development and characterisation of 140 new microsatellites in apple (*Malus* x *domestica* Borkh). *Molecular Breeding*, *10*, 217–241.

Lo, E. Y. Y., Stefanović, S., & Dickinson, T. A. (2013). Geographical parthenogenesis in Pacific Northwest hawthorns. *Botany-Botanique*, *91*, 107–116.

Lo, E. Y. Y., Stefanović, S., & Dickinson, T. A. (2009). Population genetic structure of diploid sexual and polyploid apomictic hawthorns (*Crataegus*; Rosaceae) in the Pacific Northwest. *Molecular Ecology*, *18*(6), 1145–1160.

Lo, E. Y. Y., Stefanović, S., & Dickinson, T. A. (2010). Reconstructing reticulation history in a phylogenetic framework and the potential of allopatric speciation driven by polyploidy in an agamic complex in Crataegus (Rosaceae). *Evolution*, *64*(12), 3593–3608.

Lowry, E., & Lester, S. E. (2006). The biogeography of plant reproduction: Potential determinants of species' range sizes. *Journal of Biogeography*, *33*(11), 1975–1982.

Lumaret, R., Guillerm, J., Delay, J., Loutfi, A. A. L., Izco, J., & Jay, M. (1987). Polyploidy and habitat differentiation in *Dactylis glomerata* L. from Galicia (Spain). *Oecologia*, *73*(3), 436–446.

Lynch, M. (1984). Destabilizing hybridization, general-purpose genotypes and geographic parthenogenesis. *The Quarterly Review of Biology*, *59*(3), 257–290.

Mau, M., Lovell, J. T., Corral, J. M., Kiefer, C., Koch, M. A., Aliyu, O. M., & Sharbel, T. F. (2015). Hybrid apomicts trapped in the ecological niches of their sexual ancestors. *Proceedings of the National Academy of Sciences*, *112*(18), 2357–2365.

Mayrose, I., Zhan, S. H., Rothfels, C. J., Magnuson-Ford, K., Barker, M. S., Rieseberg, L. H., & Otto, S. P. (2011). Recently formed polyploid plants diversify at lower rates. *Science*, *333*(6047), 1257.

McGoey, B. V., Chau, K., & Dickinson, T. A. (2014). Stomata size relation to ploidy level in North American Hawthorns (Crataegus, Rosaceae). *Madrono*, *61*(2), 177–193.

Meimberg, H., Rice, K. J., Milan, N. F., Njoku, C. C., & McKay, J. K. (2009). Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany*, *96*(7), 1262–1273.

Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology*, *4*, 792–794.

Meirmans, P. G., Vlot, E. C., Den Nijs, J. C. M., & Menken, S. B. J. (2003). Spatial ecological and genetic structure of a mixed population of sexual diploid and apomictic triploid dandelions. *Journal of Evolutionary Biology*, *16*(2), 343–352.

Michaels, H. J., & Bazzaz, F. A. (1989). Individual and population responses of sexual and apomictic plants to environmental gradients. *The American Naturalist*, *134*(2), 190–207.

Nathan, R. (2006). Long-distance dispersal of plants. *Science*, *313*(5788), 786–788.

Oplaat, C., & Verhoeven, K. J. F. (2015). Range expansion in asexual dandelions: Selection for general-purpose genotypes? *Journal of Ecology*, *103*(1), 261–268.

Otto, S. P. (2007). The evolutionary consequences of polyploidy. *Cell*, *131*(3), 452–462.

Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics*, *34*, 401–437.

Pannell, J. R., & Barrett, S. C. H. (1998). Baker's Law revisited: Reproductive assurance in a metapopulation. *Evolution*, *52*(3), 657–668.

Parisod, C., & Christin, P.-A. (2008). Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist*, *178*(2), 436–447.

Paun, O., Bateman, R. M., Fay, M. F., Luna, J. A., Moat, J., Hedrén, M., & Chase, M. W. (2011). Altered gene expression and ecological divergence in sibling allopolyploids of *Dactylorhiza* (Orchidaceae). *BMC Evolutionary Biology*, *11*, 113.

Paun, O., Greilhuber, J., Temsch, E. M., & Hörandl, E. (2006). Patterns, sources and ecological implications of clonal diversity in apomictic *Ranunculus carpaticola* (Ranunculus auricomus complex, Ranunculaceae). *Molecular Ecology*, *15*(4), 897–910.

Phipps, J. B. (2015). *Crataegus*. In 1993 + FNAEC (Ed.), *Flora of North America North of Mexico* (pp. 491–643). New York and Oxford: Oxford University Press.

R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ramsey, J. (2011). Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences*, *108*(17), 7096–7101.

Ramsey, J., & Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, *29*, 467–501.

Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J., & Shimizu, K. K. (2017). Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Molecular Ecology*, *26*, 193–207.

Soltis, D. E., Visger, C. J., Marchant, D. B., & Soltis, P. S. (2016). Polyploidy: Pitfalls and paths to a paradigm. *American Journal of Botany*, *103*(10), 1–21.

Stebbins, G. L. (1940). The significance of polyploidy in plant evolution. *The American Naturalist*, *74*(750), 54–66.

Talent, N., & Dickinson, T. A. (2005). Polyploidy in *Crataegus* and *Mespilus* (Rosaceae, Maloideae): Evolutionary inferences from flow cytometry of nuclear DNA amounts. *Canadian Journal of Botany*, *83*(10), 1268–1304.

Thompson, S. L., & Whitton, J. (2006). Patterns of recurrent evolution and geographic parthenogenesis within apomictic polyploid Easter daises (*Townsendia hookeri*). *Molecular Ecology*, *15*(11), 3389–3400.

Van Dijk, P. J. (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*(1434), 1113–1121.

Van Dijk, P. J., & Bakx-Schotman, T. (1997). Chloroplast DNA phylogeography and cytotype geography in autopolyploid *Plantago media*. *Molecular Ecology*, *6*, 345–352.

Van Lierop, W. M., & MacKenzie, A. F. (1977). Soil pH measurement and its application to organic soils. *Canadian Journal of Soil Sciences*, *57*(182), 55–64.

Vandel, A. (1928). La parthénogenese geographique. Contribution a l'étude biologique et cytologique de la parthénogenese naturelle. *Bulletin Biologique de la France et de la Belgique*, *62*, 164–182.

Vrijenhoek, R. C. (1984). Ecological differentiation among clones: The frozen niche variation model. In K. Wöhrmann, & V. Loeschcke (Eds.), *Population biology and evolution* (pp. 217–231). Heidelberg: Springer-Verlag.

Vrijenhoek, R. C., & Parker, E. D. Jr. (2009). Geographical Parthenogenesis: General Purpose Genotypes and Fozen Niche Variation. In I. Schön, K. Martens, & P. Dijk (Eds.), *Lost sex*. The Evolutionary Biology of Parthenogenesis. Dordrecht: Springer Netherlands.

Ward, D. B. (1974). The "Ignorant Man" technique of sampling plant populations. *Taxon*, *233*(2), 325–330.

Yahara, T. (1990). Evolution of agamospermous races in *Boehmeria* and *Eupatorium*. *Plant Species Biology*, *5*, 183–196.

Zarrei, M., Stefanović, S., & Dickinson, T. A. (2014). Reticulate evolution in North American black-fruited hawthorns (*Crataegus* section *Douglasia*; Rosaceae): Evidence from nuclear ITS2 and plastid sequences. *Annals of Botany*, *114*(2), 253–269.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Coughlan JM, Han S, Stefanović S, Dickinson TA. Widespread generalist clones are associated with range and niche expansion in allopolyploids of Pacific Northwest Hawthorns (*Crataegus* L.). *Mol Ecol*. 2017;26:5484–5499. https://doi.org/10.1111/mec.14331