Received: 16 September 2016 | Revised: 12 May 2017 | Accepted: 5 August 2017

DOI: 10.1111/mec.14342

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Reproductive isolation and environmental adaptation shape the phylogeography of mountain pine beetle (*Dendroctonus ponderosae*)

Eddy J. Dowle[1,2] | Ryan R. Bracewell[3] | Michael E. Pfrender[4] | Karen E. Mock[5] | Barbara J. Bentz[5,6] | Gregory J. Ragland[1,2]

[1]Department of Entomology, Kansas State University, Manhattan, KS, USA

[2]Department of Integrative Biology, University of Colorado Denver, Denver, CO, USA

[3]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, USA

[4]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA

[5]Department of Wildland Resources, Utah State University, Logan, UT, USA

[6]USDA Forest Service, Rocky Mountain Research Station, Logan, UT, USA

Correspondence
Eddy Dowle and Gregory Ragland, Department of Integrative Biology, University of Colorado Denver, Denver, CO, USA.
Emails: edwina.dowle@ucdenver.edu; gregory.ragland@ucdenver.edu

Funding information
U.S. Forest Service; Division of Integrative Organismal Systems, Grant/Award Number: 1451274

Abstract

Chromosomal rearrangement can be an important mechanism driving population differentiation and incipient speciation. In the mountain pine beetle (MPB, *Dendroctonus ponderosae*), deletions on the Y chromosome that are polymorphic among populations are associated with reproductive incompatibility. Here, we used RAD sequencing across the entire MPB range in western North America to reveal the extent of the phylogeographic differences between Y haplotypes compared to autosomal and X-linked loci. Clustering and geneflow analyses revealed three distinct Y haplogroups geographically positioned within and on either side of the Great Basin Desert. Despite close geographic proximity between populations on the boundaries of each Y haplogroup, there was extremely low Y haplogroup mixing among populations, and gene flow on the autosomes was reduced across Y haplogroup boundaries. These results are consistent with a previous study suggesting that independent degradation of a recently evolved neo-Y chromosome in previously isolated populations causes male sterility or inviability among Y haplotype lineages. Phylogeographic results supported historic contraction of MPB into three separate Pleistocene glacial refugia followed by postglacial range expansion and secondary contact. Distinct sets of SNPs were statistically associated with environmental data among the most genetically distinct sets of geographic populations. This finding suggests that the process of adaptation to local climatic conditions is influenced by population genetic structure, with evidence for largely independent evolution in the most genetically isolated Y haplogroup.

KEYWORDS

Dendroctonus ponderosae, gene flow, reproductive isolation, sex chromosome

1 | INTRODUCTION

Physical and environmental barriers to dispersal are obvious determinants of phylogeographic patterns. Intrinsic genetic factors such as reproductive incompatibilities and linkage disequilibrium (LD), however, may also play major roles in shaping patterns of historic population differentiation and contemporary gene flow (Coyne & Orr, 2004; Dobzhansky, 1937). For example, secondary contact following extended periods of allopatry can generate allele frequency clines. Such continuous geographic variation may result from (i) selection for different phenotypic optima at opposite ends of a cline (Berry & Kreitman, 1993; Johnsen et al., 2007), (ii) differences in

mutation rate (Charlesworth, Coyne, & Barton, 1987), (iii) differences in effective population sizes (Mank, Vicoso, Berlin, & Charlesworth, 2010) or (iv) Dobzhansky–Muller interactions (Johnson & Lachance, 2012; Turner & Harr, 2014). The steepness of these clines may vary markedly depending on LD and inheritance pattern (Berg et al., 2000; Macholán et al., 2007; Payseur, Krenz, & Nachman, 2004). For example, mitochondrial DNA and chromosomes harbouring male-determining genes (i.e., Y or Z chromosomes) may exhibit much steeper clines than autosomal loci (Carling & Brumfield, 2008; Geraldes, Basset, et al., 2008). In extreme cases, genetic differences on sex chromosomes may be nearly fixed, while gene flow continues in autosomal regions (Geraldes, Basset, et al., 2008; Geraldes, Carneiro, et al., 2008), implying the existence of sex-linked loci associated with genetic reproductive incompatibilities.

Similar to a classic ring species (Mayr, 1942), mountain pine beetles (MPB; *Dendroctonus ponderosae* Hopkins, Coleoptera: Curculionidae, Scolytinae) exhibit genetic structure that follows a broad genetic isolation-by-distance pattern around the Great Basin and Mojave Deserts of the western United States (Batista, Janes, Boone, Murray, & Sperling, 2016; Mock et al., 2007). However, there is clear evidence of postzygotic isolation among MPB populations, despite minimal genetic differentiation among populations detected by AFLP markers (Bracewell, Pfrender, Mock, & Bentz, 2011). Consistent with Haldane's rule (Haldane, 1922), when populations from either side of the Great Basin are mated, hybrid male sterility and hybrid male inviability, in the form of delayed development, occur (Bentz, Bracewell, Mock, & Pfrender, 2011; Bracewell, Bentz, Sullivan, & Good, 2017; Bracewell et al., 2011). Reproductive isolation in MPB is hypothesized to have resulted from periods of allopatry following Pleistocene climate fluctuations that shifted host tree species distributions, as documented for multiple Scolytine bark beetle species (Cognato, Harlin, & Fisher, 2003; Cullingham et al., 2011; Mock et al., 2007; Stock, Pitman, & Guenther, 1979).

In addition to phylogeographic genetic structure, there are also clear life history differences among geographic populations that appear to reflect adaptation to local seasonality (Bentz et al., 2011). In particular, because MPB lack a well-defined dormancy, seasonal synchronization of reproduction with favourable conditions is largely determined by larval development rates and thresholds (Bentz, Logan, & Amman, 1991). Development rates vary predictably across geography, with northern populations developing faster than southern populations to conform to relatively shorter growing seasons (Bentz, Logan, & Vandygriff, 2001; Bentz et al., 2014; Bracewell, Pfrender, Mock, & Bentz, 2013). Therefore, we expect that geographically variable selection on development time will produce locally adapted genotypes, modulated by the amount of standing genetic variation within populations and the movement of alleles among populations.

Here, we present the results of a range-wide genetic sampling to infer the evolutionary history of MPB autosomes and sex chromosomes in the context of postglacial range expansion. MPB have 11 autosomes, a large neo-X chromosome resulting from a recent fusion of the ancestral X with the largest autosome, and a neo-Y chromosome formed concurrently with the neo-X (Lanier & Wood, 1968). We conducted phylogeographic and environmental association analyses of genomewide polymorphism in MPB, comparing evidence of admixture on autosomes vs. sex chromosomes, and exploring how restricted gene flow might influence the process of local seasonal adaptation. We present evidence for strong spatial restriction of Y haplotype lineages despite moderate gene flow at autosomal loci. Further, environmental association analyses suggest that the constriction of gene flow across Y haplotype lineage (hereafter, "haplogroup") boundaries influences adaptation to local climatic conditions. This pattern is consistent with (i) previously reported reproductive incompatibilities of different Y haplogroups and (ii) secondary contact following expansion from three historically disjunct glacial refugia in the western (Sierra Nevada), eastern (Rocky Mountains) and central (Great Basin sky islands and Northern Arizona) portions of the MPB geographic range.

2 | METHODS

2.1 | Study system

Mountain pine beetle is one of the most economically important native pests of western North American forest ecosystems. MPB attacks and infests species within the *Pinus* genus, and their geographic range tracks both favourable climate and suitable host species distribution. Northward range expansion is ongoing (Cullingham et al., 2011; Mock et al., 2007), and MPB currently occur from Baja California Norte, Mexico, to northern British Columbia and western Alberta, Canada (Wood, 1982).

Mountain pine beetle have a long evolutionary history with *Pinus* (Bracewell et al., 2011; Brunelle, Rehfeldt, Bentz, & Munson, 2008; Sequeira, Normark, & Farrell, 2000), resulting in evolved adaptations in both the insect and host trees. MPB larval feeding in the phloem girdles the host tree, which in most cases results in tree death. *Pinus*, however, are not passive prey and have evolved resinous defences that are induced when attacked by native predators including MPB (Franceschi, Krokene, Christiansen, & Krekling, 2005). Because overwhelming resinous defences requires a large number of MPB attacking host trees in a relatively short period of time, MPB have in turn evolved strategies to increase the probability of successful tree colonization including (i) metabolism of tree defence compounds to produce pheromones that aggregate conspecifics (Blomquist et al., 2010; Raffa & Berryman, 1983), (ii) mutualistic associations with phytopathogenic fungi (Six, 2012) and (iii) temperature-driven physiological processes that aid in promoting seasonality and synchronization of adult emergence (Bentz et al., 1991; Logan & Bentz, 1999). Although landscapes of suitable host trees must be present for population outbreaks (Fettig, Gibson, Munson, & Negrón, 2014), temperatures that promote survival and synchronized adult emergence are also critical (Powell & Bentz, 2009). Temperature-driven developmental timing results in synchronous adult emergence that is critical for coordinated attacks (Bentz et al., 1991; Logan & Bentz, 1999; Powell & Bentz, 2009, 2014). Due to the importance of seasonality,

divergent natural selection caused by differences in growing season length across the extensive geographic range of MPB has driven marked local adaptation via traits related to developmental timing (Bentz et al., 2001, 2011; Bracewell et al., 2013).

2.2 | Sample collection

Mountain pine beetle were collected from across the species' range in one of three ways: (i) manual extraction from beneath the bark of multiple infested trees at a site, (ii) collection of adults emerging from naturally infested trees that were cut and transported to the laboratory, or (iii) use of traps baited with a MPB-specific aggregation pheromone (*trans*-verbenol, myrcene, *exo*-brevicomin, Phero-Tech, Inc., Delta, BC; Appendices S1, S2). When collections were made from beneath the bark of infested trees, collections came from multiple galleries within a tree and also from multiple trees at a geographic site and included adults and larvae. Adults emerging in the laboratory from bolts taken from naturally infested trees were collected from across the adult emergence distribution. Adults from pheromone-baited traps were taken from across weeks during flight times. All adults and larvae were placed in 95% ethanol while alive and subsequently stored at −80°C.

2.3 | DNA extraction and sequencing

In total, 707 individuals from across the MPB range were sequenced via a double-digest restriction-associated DNA (RAD) approach (Parchman et al., 2012) modified for paired-end sequencing. The full protocol is provided in Appendix S3. Briefly, DNA was extracted from the thorax of individual beetles (to limit fungal/bacterial contamination) using Gentra Puregene reagents (Qiagen, Valencia, CA) and following manufacturer's recommendations, and 90–110 ng extracted DNA from each individual was transferred to a digestion reaction using the restriction enzymes EcoRI and MseI (New England Biolabs, NEB). Following digestion, adapters were ligated (the EcoRI end contains a unique barcode and forward Illumina sequencing primer; the universal MseI end contains no barcode and reverse Illumina sequencing primer), and then PCR (30 cycles)-amplified using primers complementary to the adapters. PCR products for each individual were visualized on an agarose gel to ensure amplification, and then individuals were pooled into groups of ~96. Each pool was size-selected for 300–500 bp fragments using a BluePippin System (Sage Science, Beverly, MA) and then cleaned with Ampure beads (Ampure beads Beckman Coulter, Indianapolis, IN). Cleaned libraries were assessed using a Bioanalyser 2100 system (Agilent) then submitted to the BGI Americas Inc., Davis, CA, facility for seven lanes of 100-bp paired-end sequencing on an Illumina HiSeq 2000 (v3 chemistry).

2.4 | Data filtering and genotyping

Although we have achieved high-quality reverse reads using the same protocol in other studies (M. E. Pfrender, unpublished), quality of the reverse reads for the MPB RAD libraries were of poor quality relative to the high-quality forward reads. Thus, only forward reads were analysed. Reads were demultiplexed and cleaned using custom scripts (Appendices S1, S4) before mapping using BWA-MEM v 0.7.12 (Li & Durbin, 2009) to the MPB draft genome (Keeling et al., 2013). Simultaneously calling and scoring SNPs for all individuals was computationally intractable. Thus, SNP calling was first performed on seven equal-sized, random groups of individuals using the Unified Genotyper in GATK v 3.3.0 (McKenna et al., 2010) retaining only SNP locations and qualities. Called SNPs were merged across groups to produce a comprehensive list before applying filters to remove loci with (i) a total depth per locus (across all individuals) of <50, (ii) evidence for overassembly (i.e., when multiple RAD fragments from a single individual erroneously map to the same location; $p > .05$ from chi-squared test of null hypothesis of 50% representation of each allele in heterozygotes), (iii) an average Phred base quality score of <20 as in (Egan et al., 2015) and (iv) a minor allele frequency (MAF) of <0.05 (across all individuals). Next, we assigned missing values for base calls (per read, per SNP) with (i) a base quality Phred score of <20, (ii) a read mapping quality Phred score of <20. Additional filtering of the loci and individual coverage filters varied depending on analyses (see sections below). Genotype likelihoods were then generated using the GATK model in ANGSD v 0.901 (Korneliussen, Albrechtsen, & Nielsen, 2014; McKenna et al., 2010).

Sex chromosome scaffolds (X and Y) were identified based on mapping to the draft genome (Keeling et al., 2013) and chromosomal assignments based on read coverage in Bracewell et al. (2017). Individuals with <30,000 mapped reads were removed from subsequent analyses. Most beetles were not visually sexed, but we were able to assign sex based on the percentage of total reads mapping to the X and Y scaffolds for each individual (see Appendix S1). In total, 122 individuals with poor coverage were removed from all subsequent analyses, 322 individuals were scored as female, 249 were scored as male, and 14 could not be sexed but had sufficient coverage to be included in the autosomal analyses. Subsequent analyses were performed separately on three groups of loci: (i) autosomal male and female, (ii) female X and (iii) male Y with an average of 16, 9 and 6 individual beetles per geographic site, respectively. Because of the recent sex chromosome rearrangement (neo-XY) within MPB (Lanier & Wood, 1968), some scaffolds in the genome are poorly assembled and are chimeric, comprised of both Neo-X and Neo-Y. We therefore performed all X analyses only on females to avoid this artefact (Appendix S1).

2.5 | Population genetic analyses

Within-population statistics were estimated using ANGSD and NGSTOOLS (Fumagalli, Vieira, Linderoth, & Nielsen, 2014; Fumagalli et al., 2013; Korneliussen et al., 2014). Neutrality statistics, Watterson's estimator, Tajima's theta and Tajima's D were calculated in ANGSD via a sliding window analysis, while Tajima's Pi was calculated in POPOOLATION (Kofler et al. 2011; Tajima 1989; Watterson 1975) (Appendix S1).

We estimated population structure separately for the autosomes, X and Y chromosomes. Pairwise F_{ST} estimates were generated using

allele frequencies estimated for each geographic site generated in ANGSD. A minimum of five individuals per geographic site were required at each locus for autosomes, while four (or all individuals per site if $n < 4$) individuals were required for X and Y loci. F_{ST} values were calculated using the Hudson estimator (Hudson, Slatkin, & Maddison, 1992), which is not biased by population size (Bhatia, Patterson, Sankararaman, & Price, 2013). We also estimated F_{ST} using several alternative methods, and results were qualitatively similar (Appendices S5d,e and S6). We applied a permutation approach to test whether average pairwise F_{ST} between geographic sites (weighted by the inverse square root of geographic distance) was greater within compared to among Y haplogroups, by estimating the empirical averages (within and across) and comparing against the distribution of estimates from 10,000 permutations where the Y haplogroups were randomly shuffled (two-tailed test). To ensure comparable geographic scales, we discarded all pairwise comparisons across Y haplogroups involving two geographic sites separated by greater than the maximum distance within either of the Y haplogroups (Appendix S7).

NgsADMIX (Skotte, Korneliussen, & Albrechtsen, 2013) was used to examine population structure and admixture among geographic sites. NgsADMIX estimates individual ancestry in a manner similar to that of the more commonly used software ADMIXTURE (Alexander, Novembre, & Lange, 2009), but ngsADMIX takes into account uncertainty in the genotype call, using genotype likelihoods as opposed to called genotypes (Skotte et al., 2013). For each analysis, loci were limited to those that were present in a minimum of 200, 100 and 100 individuals for autosomes, X and Y, respectively, in addition to the filters described above (Appendices S1, S9, S10, S11, S12).

Population structure was also visualized using an unguided principle component analysis (PCA). A covariance matrix was generated using the genotype likelihood scores for the three data sets via ngsCovar in the NGSTOOLS package, accounting for genotype uncertainty by weighting based on the likelihood for each possible genotype (Fumagalli et al., 2013, 2014; Patterson, Price, & Reich, 2006). To avoid high levels of missing data, we included only loci with nonmissing genotype probabilities, for a minimum of 400, 200 and 100 individuals for autosomes, X and Y analyses, respectively. Principal components were calculated in R v 3.2.1 using "eigen," and PCA plots were drawn using the GGPLOT2 and ELLIPSE packages (R Core Team 2013; Wickham, 2009).

Unrooted phylogenetic trees were used to determine hierarchical relationships among populations. A maximum-likelihood tree for each of the three data sets was generated using SNPHYLO (Lee, Guo, Wang, Kim, & Paterson, 2014). Because this analysis relies on called genotypes, we assigned missing values for all genotypes with quality scores <30, a polymorphic loci p-value of $<1 \times 10^{-6}$ and SNP posterior cut-off of <0.95 in ANGSD. We then further filtered individuals with missing genotypes at a high proportion of loci (Appendix S1). A total of 1,000 bootstrap replicates were run for each analysis. Trees were visualized and manipulated in FIGTREE v1.4.2 (Rambaut, 2009).

Using called autosomal genotypes filtered as above, we estimated relationships among geographic sites using SNAPP (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012; Drummond, Suchard, Xie, & Rambaut, 2012), implemented in BEAST 2.0. SNAPP uses biallelic data to infer rooted population trees using a coalescent model. Because it is computationally intractable to infer relationships among >15 groups (here, geographic sites) at a time, we ran two analyses using different sets of sites. For each of the geographic sites, the three individuals with the fewest missing data across loci were selected. The first analysis included ten geographic sites with at least one representative from all major clusters identified by SNPHYLO (see above). The second analysis included thirteen geographic sites that were close to Y haplogroup boundaries. SNAPP was run for 1,500,000 generations sampling every 500 generations. Convergence was assessed using TRACER v1.6.0 (Rambaut & Drummond, 2007) with a burnin of 375,000. Trees were drawn in DENSITREE using the same burnin (Bouckaert, 2010).

2.6 | Tests for gene flow

We tested for gene flow among geographically separated sites using the ABBA-BABA test as implemented in POPSTATS (Skoglund et al., 2015). The ABBA-BABA method uses allele frequency estimates to test for an excess of shared derived alleles between populations (Green et al., 2010; Reich, Thangaraj, Patterson, Price, & Singh, 2009). The Jeffery pine beetle (*Dendroctonus jeffreyi*, sequence Bracewell et al. (2017)) was used as the outgroup in all analyses. POPSTATS uses called genotypes, and loci were generated in ANGSD using a minimum mapping and locus quality of 30, a polymorphic loci p-value of $<1 \times 10^{-6}$ and posterior cut-off of >0.95. Population structure analyses identified three distinct male Y haplotype lineages, each with discrete geographic ranges and little evidence of mixed Y haplogroups within populations. However, there was evidence for autosomal admixture across the Y haplogroup boundaries (see Section 3). ABBA-BABA tests using autosomal loci therefore focused on comparisons of geographically proximal sites separated by a Y haplogroup boundary. The ABBA-BABA test requires three samples (here geographic sites) and an outgroup (here *D. jeffreyi*). Sample one was chosen from the same autosomal genetic cluster and Y haplogroup (determined from ngsADMIX) as sample two; however, the geographic site chosen for sample one was spatially isolated (i.e., unlikely to be influenced by recent gene flow) from any Y haplogroup boundary, whereas the site chosen for sample two bordered a Y haplogroup boundary. The third sample was taken from a geographic site near the same Y haplogroup boundary as sample two but which contained an alternative Y haplogroup. As per previous studies, a z-score (D-statistic/SE) with an absolute value of three or more was considered evidence of significant gene flow (Green et al., 2010; Reich et al., 2009).

2.7 | SNP-to-environment associations

We calculated 30-year averages (1971–2000) for 14 climatic variables that are known to influence MPB seasonality for each geographic location using BIOSIM (BioSIM 2014; Régnière, 1996). To avoid running multiple tests on correlated variables, we used PC1

and PC2 values from a PCA of all fourteen variables, which together explained 68% of the variance (Appendices S2, S15a). Associations between SNP variation and PC1/2 were tested across all locations and independently within each Y haplogroup. Development/generation time of MPB is tightly coupled to seasonality and varies predictably with latitude, with northern populations experiencing shorter growing seasons and demonstrating relatively faster development rates (Bracewell et al., 2011, 2013).

Four separate analyses were performed to test for statistical associations between autosomal and X-linked loci and PC1/2: a range-wide analysis and three analyses within each of the three Y haplogroups identified in the preceding analyses (see Results; western, eastern, central Figure 1; poorly represented geographic sites were removed, see Appendix S1). We used two statistical models implemented in (i) Bayenv2 and (ii) latent factor mixed model (LFMM; Frichot, Schoville, Bouchard, & François, 2013; Günther & Coop, 2013). Bayenv2 models allele frequencies as a function of population structure (a variance–covariance matrix of allele frequencies among populations) and an environmental variable (Günther & Coop, 2013). LFMM uses a latent factor mixed model with a matrix of allele frequencies as the response, environmental variables as fixed factors and population structure as a latent factor (Frichot et al., 2013). The LFMM models for the central Y haplogroup converged poorly; thus, we report only the Bayenv2 results for the analysis of this Y haplogroup. Bayenv2 does not tolerate missing data. Thus, to enable comparisons between groups we only included loci represented in all three groups (range-wide, eastern and western) for which LFMM and Bayenv2 were run. These stringent criteria yielded 3,101 loci that were present in at least five individuals per geographic site (or all individuals if $n < 5$) for range-wide, western and eastern Y haplogroup analyses using both LFMM and Bayenv2. Due to higher levels of missing data, we analysed only 1,173 loci for the central Y haplogroup. Subsequently, all comparisons to the central Y haplogroup were also limited to the same 1,173 loci.

Evidence for environmental associations evolving in parallel between Y haplogroups was evaluated in three ways. First, we tested whether pairs of Y haplogroups shared more significantly associated autosomal loci than expected by chance using Fisher's exact tests. For comparisons with both LFMM and Bayenv2, "significant" loci were required to have a false discovery rate <0.05% in LFMM and Bayenv2 *z*-score >0.40. For the central Y haplogroup, we were only able to run the Bayenv2 model, so for this group we required a more stringent significance criteria, $z > 0.45$. Second, we asked whether the directionality of the allele–environment relationships was the same or different between Y haplogroups (same or different sign on the regression coefficient; Appendix S1). Finally, we applied functional enrichment analyses for sets of SNPs significantly associated with environmental variables, using DAVID (Dennis et al., 2003) to test for over-representation of Gene Ontology (GO) and additional (e.g., Kyoto Encyclopedia of Genes and Genomes) functional categories. These analyses included only SNPs in or within 100 bp of gene models with blastx hits ($e < 0.01$) to the flybase amino acid database or Uniprot.

3 | RESULTS

In total, 19,904 SNP loci were identified as passing our minimal filters. Of these, 2,522 mapped to the X chromosome and 434 mapped to the Y chromosome. Nucleotide divergence/diversity statistics mostly suggest neutral evolution on autosomal and sex-linked scaffolds (see Appendices S1 and S8 for estimates of Watterson estimator, Tajima's theta and Tajima's D).

3.1 | Population admixture and phylogeography

Admixture analysis and F_{ST} values suggest substantial admixture across geography for autosomal and X-linked loci, whereas Y-linked

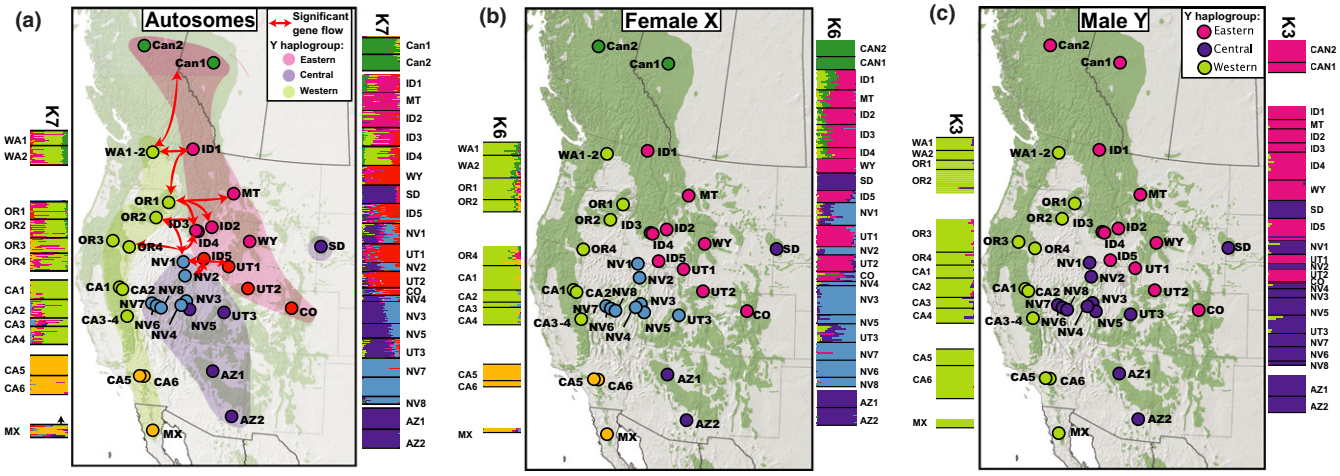

FIGURE 1 (a) NgsADMIX results for autosomal (male and female combined) (a), female X (b) and male Y (c) loci. Geographic site names include the abbreviation for the state from which beetles were collected. Sites on the maps are colour-coded by the majority genetic cluster found within each site. Significant autosomal gene flow (tested using ABBA-BABA) crossing Y haplogroup boundaries is indicated with red arrows in panel a [Colour figure can be viewed at wileyonlinelibrary.com]

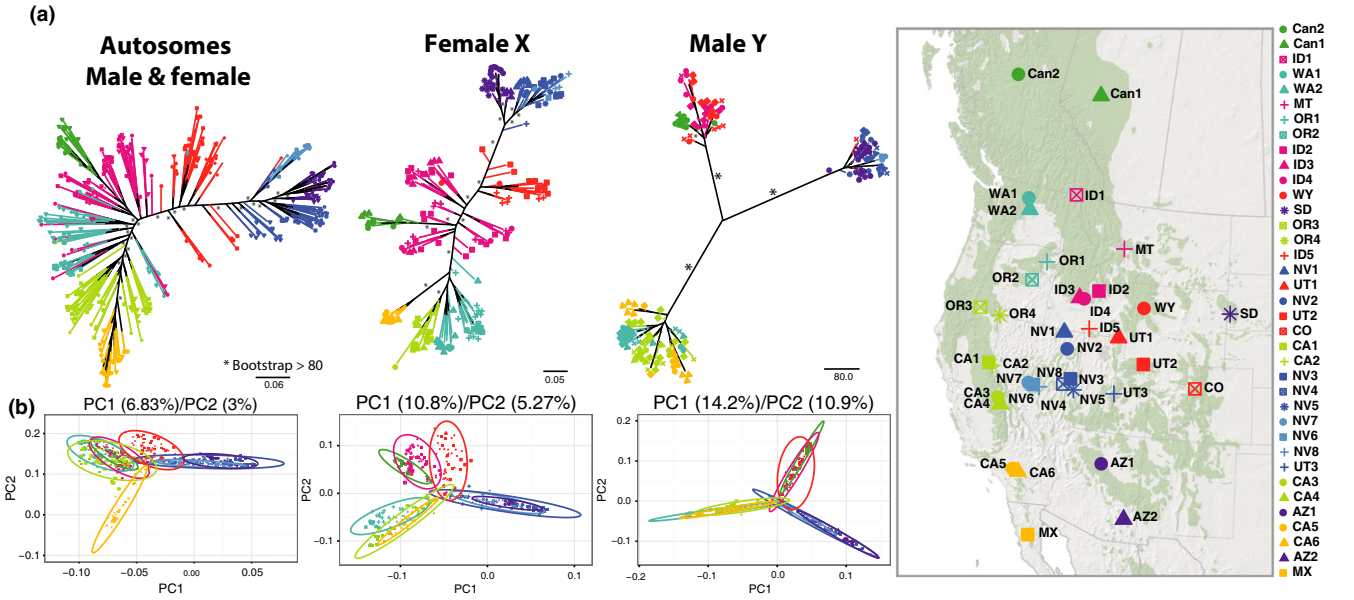

FIGURE 2 Maximum-likelihood phylogenies inferred from SNPhylo for autosomal (male and female combined), female X and male Y loci. Bootstrap values >80 are shown. (b) PCA clustering of autosomal, female X and male Y loci [Colour figure can be viewed at wileyonlinelibrary.com]

loci segregated as a single haplotype with multiple variants that were almost entirely monomorphic within geographic sites (Figures 1 and 2; Appendices S5, S9–S12). We identified three discreet Y haplogroups, referred to as western, eastern and central. Admixture analysis suggested that only a single Y haplogroup was present at each geographic site except for ID5, where we identified nine individuals with the eastern Y haplogroup and a single individual with the central Y haplogroup. However, admixture analysis of autosomal loci clustered this individual most closely with geographic sites in the eastern Y haplogroup. The three Y haplogroups occurred separately in well-defined geographic regions, except for a spatially disjunct geographic site in South Dakota (SD) that shared the central Y haplogroup with sites in the Great Basin (e.g., NV1-8), southern Utah (UT3) and Arizona (AZ1-2) rather than the eastern Y haplogroup of the most spatially proximal site in Colorado (CO). This relationship was also supported by autosomal and X-linked loci (Figures 1 and 2; autosomal F_{ST} for CO-SD = 0.16; F_{ST} for SD-AZ1 = 0.09). In general, population structure inferred from autosomal and X-linked loci was consistent with admixture and generally lower F_{ST} values among spatially proximal sites, with population clusters largely conforming to geography. Although the population structure was very similar on the autosomes and X chromosome, pairwise F_{ST} scores were higher for X-linked loci (average F_{ST} autosomes = 0.24, X = 0.37). Similarly, admixture was reduced on the X compared to the autosomes, particularly between geographic sites with different Y haplogroups (Figure 1). The highest F_{ST} estimates (>0.4) were found between sites in CA and AZ, the southern ends of the MPB distribution on either side of the Great Basin Desert (Appendix S5).

Admixture across Y haplogroup boundaries was inferred for both autosomal and X-linked loci, particularly between sites in Washington and Oregon (western Y haplogroup) and sites in Idaho and Montana (eastern Y haplogroup). These sites lie along the northernmost boundary between the western and eastern Y haplogroups. Permutation tests also supported substantial autosomal admixture between the eastern and western Y haplogroups, although admixture between the central Y haplogroup and both the western and eastern Y haplogroups was restricted. Average, geographic distance-standardized F_{ST} was not significantly greater between geographic sites spanning the eastern–western Y haplogroup boundary compared to F_{ST} between sites within the eastern Y haplogroup (p = .68; Appendix S7). Average F_{ST} was actually greater within the western Y haplogroup compared to F_{ST} across the eastern/western Y haplogroup boundary (p = .036). This result seems to be driven primarily by genetic isolation of the geographic sites in southern California (CA) and northern Baja California, Mexico (MX). When these geographic sites were removed from the analyses, F_{ST} did not significantly differ. In contrast, autosomal F_{ST} was significantly higher between eastern and central Y haplogroups compared to F_{ST} within either the eastern or central Y haplogroups (p < .0001), with similar results comparing western and central Y haplogroups (p < .0001).

Both the PCA results and unrooted phylogenetic trees generated from Y-linked loci suggested three highly distinct groups of geographic sites corresponding to the same three Y haplogroups (Bootstrap value = 100) identified using ngsADMIX (Figure 2a). The unrooted phylogenies for autosomal and X-linked loci were similar, revealing more population subdivision within the Y haplogroups. However, the tree generated using X-linked loci suggested that beetles collected from the Oregon (OR1-2) and Idaho (ID1-4) sites each formed monophyletic groups, while autosomal-generated trees suggested that these groups were polyphyletic. Autosomal and X-linked PCA clustering and unrooted trees suggested that British Columbia and Alberta, Canada (Can), sites cluster most closely with ID sites. These patterns agree with the pattern revealed by Y-linked loci showing that the Canadian sites share the Y haplogroup found in ID; that is, that Canadian sites are part of the eastern Y haplogroup.

Population trees generated in SNAPP using autosomal loci (Figure 3; Appendix S13) were also consistent with autosomal introgression across geography and across Y haplogroups. The first tree including ten geographic sites centrally located within the geographic range of each Y haplogroup (Appendix S13) had good convergence across all parameters (effective sample size (ESS), i.e., an estimate of the number of independent samples in an MCMC run >200), while the second tree including sites near Y haplogroup geographic boundaries (Figure 3) had reasonable convergence across tree parameters (ESS > 100). However, ESS values suggested poor convergence on effective population size estimates; these estimates were thus not reported. Both trees reveal two main clades; the tree produced with centrally located geographic sites suggests monophyly of Y haplogroups within clades (Appendix S13), but the tree produced with sites located near Y haplogroup boundaries suggests more uncertainty in phylogeographic relationships (Figure 3). In the latter tree, geographic sites associated with the eastern and western Y haplogroups did not form monophyletic groups and sites from the eastern Y haplogroup occurred in both of the major clades. Subsequent geneflow analysis (see below) identified significant autosomal gene flow between geographic sites in the eastern Y haplogroup and sites within both other Y haplogroups.

3.2 | Gene flow

Geneflow analysis (ABBA-BABA test) identified significant exchanges of alleles between geographic sites spanning Y haplogroup boundaries at autosomal loci, yet with the limited exception noted above, Y haplogroup was monomorphic within sites (Figure 1; Appendix S14). We detected significant autosomal gene flow among geographic sites with different Y haplogroups in the north-central portion of the MPB range (Can, WA, OR, ID, MT, NV1-2, UT1), corroborating admixture results. No significant autosomal gene flow was detected between central Nevada (NV3-NV8; central Y haplogroup) and California (CA1-CA4; western Y haplogroup) despite their geographic proximity (~100 km). Additionally, no autosomal gene flow was detected between South Dakota (SD), which genetically clustered with AZ (central Y haplogroup), and the geographically proximal sites in Wyoming (WY), Utah (UT) and Colorado (CO).

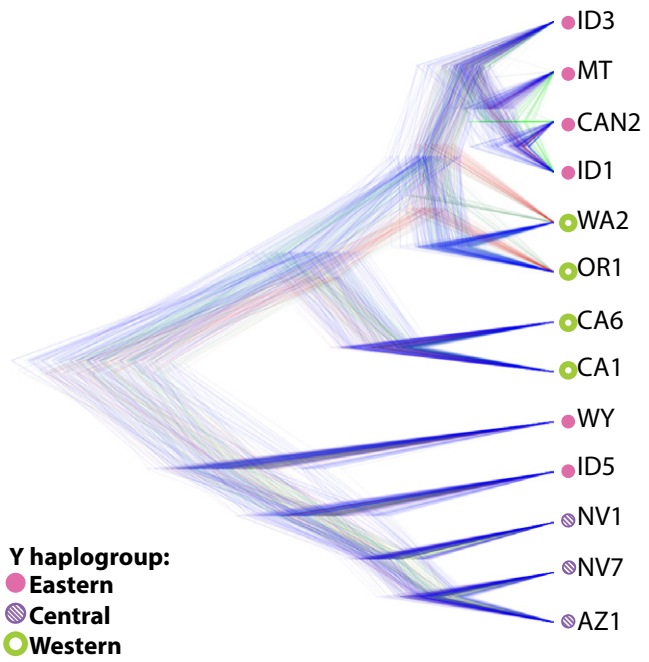

FIGURE 3 SNAPP-generated tree visualized using DENSITREE. Relationships among twelve samples, chosen because of their proximity to Y haplogroup boundaries, from across the MPB range are shown. Samples of Y haplogroup are indicated by dot colour and pattern [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | Environmental adaptation

Environmental association analyses identified sets of loci correlated with PC1/2 of the 14 environmental variables (Table 1). Most of the environmental variables, for example mean temperature and degree days, loaded on PC1, while number of days with precipitation, latitude, longitude and elevation loaded on PC2 (Appendix S15a). Analyses were performed separately for geographic sites within the western and eastern Y haplogroups, in addition to an analysis including all sites (range-wide analysis). The range-wide analyses identified 161/219 loci significantly associated with PC1/2, while the separate analyses within the western and eastern Y haplogroups identified 154/284 and 184/245 significantly associated loci, respectively. The range-wide analysis captured some of the loci identified in the separate analyses of each Y haplogroup (>40% and >30% of loci identified in the western and eastern analysis, respectively). Significantly more loci than expected by chance alone were associated with PC2 between the western and eastern Y haplogroup analyses (37 common loci; $p = <4 \times 10^{-4}$, Fisher's exact test) but not for PC1 (14 common loci; $p = .1117$). Evidence for congruent sets of loci associated with PC2 was strengthened by the observation that for 25 of these 37 common loci (Table 1), the signs of the regression coefficients (the direction of the relationship with PC2) were the same in both Y haplogroups. Comparisons with the central Y haplogroup were limited to Bayenv2 results because LFMM models would not converge in the central Y haplogroup analysis (Appendix S1). Based on the Bayenv2 results, there was no statistical evidence for over-representation of common, significant loci between the central Y haplogroup and any other Y haplogroup or the range-wide analysis of both PC1/2 (nonsignificant Fisher's exact tests; Appendix S16). The analysis of PC2 in the central Y haplogroup only identified nine loci reflecting the limited amount of variation found in PC2 for this Y haplogroup (Appendix S15b). Finally, very few loci demonstrating a significant environmental association (157 of 932 total) mapped to gene models (Appendix S16b:i), and no sets were enriched for any GO functional categories.

4 | DISCUSSION

The mountain pine beetle is a *Pinus* host generalist that can undergo wind-borne transport above the forest canopy at rates of 30–110 km/

TABLE 1 Numbers of loci significantly associated (FDR < 0.05, LFMM; $Z > 0.40$, Bayenv2) with PC1/2 in both LFMM and Bayenv2 models in three analyses: (i) including only geographic sites from the western Y haplogroup, (ii) including only sites from the eastern Y haplogroup and (iii) including sites from across the MPB range. Figure 1 indicates Y haplogroup affiliation of each geographic population. Significant loci with the same directionality that overlap between comparisons are left of the/while significant loci that differ in direction are to the right. Directionality was determined from LFMM *z*-scores. *p* values indicate whether the number of overlapping loci was significantly different from chance (Fisher's exact test).

Y Haplogroup and variable	Significant loci that overlap Western Y PC1	Significant loci that overlap Eastern PC1	Significant loci that overlap Range-wide PC1	Significant loci that overlap Western Y PC2	Significant loci that overlap Eastern PC2	Significant loci that overlap Range-wide PC2
Significant loci that overlap Western Y	**154**			**264**		
Significant loci that overlap Eastern Y	12/2 ($p = .112$)	**184**		25/12 ($p < .001$)	**245**	
Significant loci that overlap Range-wide	77/0 ($p < .001$)	50/0 ($p < .001$)	**161**	100/0 ($p < .001$)	89/0 ($p < .001$)	**219**

Bold/diagonals, significant loci within region; off-diagonals, N overlapping SNPs.

day (Jackson, Straussfogel, Lindgren, Mitchell, & Murphy, 2008), which in the absence of reproductive barriers should facilitate gene flow among many of the geographically proximal sites that we sampled. Indeed, two previous studies have inferred substantial admixture among geographic localities, one using AFLP markers and mainly geographic sites across the United States (Mock et al., 2007), and one using SNP markers and mainly sites across Canada (Batista et al., 2016). Despite admixture, both of these studies inferred population structure across the landscape, and crossing experiments identified male reproductive incompatibilities between both geographically proximal and geographically isolated US populations (Bentz et al., 2011; Bracewell et al., 2011). Most recently, a genomic analysis of selected geographic sites sampled across the MPB range identified three distinct male Y haplogroups (Bracewell et al., 2017). Consistent with these previous studies, we found evidence for genetic structure with gene flow across the MPB range in autosomal and X-linked loci. In contrast, Y-linked loci segregated as haplogroup blocks that were nearly completely monomorphic within populations, with three distinct Y haplogroups dispersed in mostly geographically contiguous western, central and eastern portions of the MPB range. These Y haplogroups, or collections of populations sharing the same Y haplotype lineage, are separated by very short geographic distances in portions of the MPB range, suggesting barriers to gene flow despite the possibility of physical dispersal among populations.

4.1 | Gene flow and reproductive incompatibilities

Y-linked loci appear to restrict, but not prevent autosomal and X-linked loci from introgressing across the MPB range, an inference supported by all admixture, clustering and geneflow analyses. In particular, there is clear evidence for gene flow among many geographically proximal populations at autosomal loci, but tight geographic delineation of distinct Y chromosome haplogroups. We identified three distinct Y chromosome haplogroups across the MPB range, and with the exception of one individual from a single geographic site (ID5), each site contained only a single Y haplotype lineage (Figures 1 and 2). This is the equivalent of a stepwise allelic cline (frequency changes directly from 0 to 1 across geography), which has also been observed (though not as clearly delineated) in Z- or X-linked loci in birds (Carling & Brumfield, 2008) and mammals (Geraldes, Basset, et al., 2008; Geraldes, Carneiro, et al., 2008). Bracewell et al. (2017) identified the same three Y haplogroups from a more limited set of geographic sites (Montana, Idaho, California, Utah, Colorado and Arizona). The more comprehensive sampling in the current study demonstrates a lack of Y haplogroup mixing across the entire range, even when populations carrying alternate Y haplogroups are geographically proximal and not separated by obvious barriers such as a lack of host trees (e.g., WA1-2 and ID1). In contrast to the Y chromosome, autosomal admixing between populations with both the same and alternative Y haplogroups was extensive (Figure 1). However, permutation tests of the central Y haplogroup suggested a restriction in autosomal gene flow between vs. within Y haplogroups, supporting an association between Y-linked loci and barriers to gene flow.

The clear delineation of Y haplogroups in the face of gene flow, combined with previous reports of low male fitness in population crosses (Bracewell et al., 2011, 2017), is consistent with Y-linked reproductive isolation. Crossing experiments have illustrated reproductive incompatibilities in male offspring when populations from the western (e.g., CA5-6) and eastern (e.g., ID4, UT1) sides of the MPB range were crossed (Bentz et al., 2011; Bracewell et al., 2011). Moreover, these reproductive incompatibilities have been explicitly linked to structural variation on the Y chromosome (Bracewell et al., 2017). MPB have a neo-XY system in which the ancestral X chromosome fused with the largest autosome (Lanier & Wood, 1968). The X chromosome is thus composed of both ancestral and neo-X sequence while the entire Y is of recent autosomal origin. Bracewell et al. (2017) found large deletion polymorphisms between MBP Y haplogroups consistent with degeneration of the neo-Y chromosome that has occurred independently in each Y lineage. Those results and our current findings are consistent with Haldane's rule, which predicts that crosses between Y haplogroups will cause reproductive incompatibilities in F1 male, but not female hybrids (Bracewell et al., 2011;

Haldane, 1922). The genetic structure among Y haplogroups identified here is consistent with the geographic boundaries previously identified by male F1 hybrid sterility in crossing experiments (Bracewell et al., 2011). Crossing experiments found strong symmetric sterility and delayed development in F1 hybrids from southern California (CA6) and Arizona (AZ2), while weak asymmetric hybrid male sterility was found when crossing beetles from geographically proximal sites in Oregon (OR2) and Idaho (ID3).

Given the near-complete lack of Y haplogroup mixing among populations, natural selection must effectively cull all MPB male hybrids and backcrosses in one to two generations. This would be a necessary condition to maintain reciprocal fixation given any appreciable migration in simple one-island, two-island and more complex migration–selection balance models (Felsenstein, 1976). Past and current migration among populations is strongly suggested by evidence for admixture on the autosomes, and our estimates of migration rates between geographically proximal populations with alternate Y haplogroups (e.g., NV1 and ID5) were very high (~30 migrants per generation; Appendix S1). Thus, we infer that male hybrid (cross-Y haplogroup) fitness is effectively zero in nature, while female hybrids maintain reasonably high fitness. Examples of complete, reciprocal hybrid male sterility (e.g., CA6 to AZ2) strongly support this conclusion. However, we also observe a complete lack of Y admixture between populations that yield asymmetric, or no F1 male sterility in the laboratory (e.g., UT1 and AZ1; B. J. Bentz, unpublished). In such cases, severe backcross viability/sterility may prevent Y-linked gene flow, but it is also possible that ecological selection acts to cull hybrids with deleterious, intermediate phenotypes. For example, male progeny of CA2 and AZ2 crosses are not only sterile but also developmentally delayed (Bracewell et al., 2017). Natural selection for seasonal synchronization is particularly strong in eruptive species such as MPB that must overwhelm host defences by mass attack (Bentz et al., 2001). Thus, both intrinsic genetic and postzygotic reproductive incompatibilities may contribute to the complete restriction of gene flow among Y haplogroups in MPB.

Restriction of sex-linked gene flow may be a hallmark of the early stages of speciation. Previous work firmly establishes the prominent role sex chromosomes play in species formation (Carling & Brumfield, 2008; Coyne & Orr, 2004; Geraldes, Carneiro, et al., 2008; Payseur et al., 2004; Presgraves, 2008). Evidence from Bracewell et al. (2017) supports a Dobzhansky–Muller model for the evolution of reproductive incompatibility in MPB wherein Y degradation that occurred independently in separate lineages (i.e., the three Y haplogroups) leads to epistatic genetic incompatibilities with other genes in the genome. It remains unclear which specific genes may interact with the Y, and whether they are autosomal, X-linked, or both, although Bracewell et al. (2017) provide gene expression and reproductive phenotypic evidence supporting several candidates.

4.2 | MPB phylogeography

The presence of three distinct and mostly geographically contiguous MPB Y haplogroups suggests a history of allopatry resulting from Pleistocene glaciation that is consistent with genetic discontinuities observed in multiple plant and animal species (Jaramillo-Correa, Beaulieu, Khasa, & Bousquet, 2009; Shafer, Cullingham, Cote, & Coltman, 2010; Swenson & Howard, 2005). These results support three separate MPB glacial refugia during the last glacial maximum (LGM), one on the western side of the Great Basin (western Y haplogroup), one on the eastern side (eastern Y haplogroup) and one generally centred in the Great Basin and Arizona (central Y haplogroup). Previous work modelling glacial refugia among western North American tree species showed that refugia for some *Pinus* sp. may have reached as far north as Montana and Oregon, during the LGM (Godbout, Fazekas, Newton, Yeh, & Bousquet, 2008; Richardson, Brunsfeld, & Klopfenstein, 2002; Roberts & Hamann, 2015). Moreover, the MPB contact zone in Oregon (OR) and Idaho (ID) first identified in Bracewell et al. (2011) and further resolved here is similar to contact zones described in MPB host species *P. albicaulis* (Richardson et al., 2002), *P. contorta* (Godbout et al., 2008) and *P. ponderosae* (Potter, Hipkins, Mahalovich, & Means, 2015). In fact, this same geographic region is a well-described contact zone for many tree species that reflects northward expansion from the Sierra Nevada/Cascades and Rocky Mountain Ranges (Jaramillo-Correa et al., 2009; Shafer et al., 2010; Swenson & Howard, 2005).

Sampling MPB from sites in the Great Basin (NV) revealed an additional contact zone (with autosomal gene flow) between the eastern Y haplogroup in the Rocky Mountains and the central Y haplogroup that includes the Arizona Y haplogroup identified in Bracewell et al. (2017). A phylogenetic break between the Sierra Nevada and Rocky Mountains, which are physically separated by the Great Basin (NV), is commonly observed in forest-dwelling vertebrates (Malaney et al., 2013; Manthey, Klicka, & Spellman, 2011; Spellman & Klicka, 2007) and plants (Gugger, Sugita, & Cavender-Bares, 2010; Potter et al., 2015). MPB from sites in the Great Basin form a lineage unique from MPB in both the Sierra Nevada and Rocky Mountain ranges, similar to what has been found in woodpeckers, which are also a forest-associated species (Klicka, Spellman, Winker, Chua, & Smith, 2011). Furthermore, Great Basin bristlecone pine (*P. longaevae*), which has a distribution encompassing NV1-8 and UT2 all within the central Y haplogroup, was recently found to have defences that confer resistant to MPB potentially due to a long-term evolutionary relationship (Bentz, Hood, Hansen, Vandygriff, & Mock, 2017). In all analyses of autosomal, X-linked and Y-linked loci, sites from NV cluster with those from AZ. This pattern is congruent with that of two host tree species: *P. monophylla* is thought to have expanded north recently (i.e., <12,000 years BP) from Southern Nevada, while *P. edulis* is thought to have expanded recently from central/southern Arizona to Utah and Northeast Arizona (Cole, Fisher, Arundel, Cannella, & Swift, 2008; Cole, Fisher, Ironside, Mead, & Koehler, 2013).

Although there is evidence for autosomal gene flow across the contact zones described above, there are also several contiguous geographic populations that appear to experience no migration. For example, the SD geographic site clusters within the geographically noncontiguous central Y haplogroup comprised of AZ and NV sites

(Figure 1) and exhibits no evidence of autosomal gene flow with surrounding sites from the eastern Y haplogroup. At first glance, this is a phylogeographic oddity, but very similar patterns have been observed in shrews (Hope, Speer, Demboski, Talbot, & Cook, 2012), woodpeckers (Klicka et al., 2011) and *Pinus ponderosae* (Potter, Hipkins, Mahalovich, & Means, 2013). This pattern, previously identified using autosomal SNPs, has been suggested to be the result of unintentional translocations (Batista et al., 2016), but a more parsimonious explanation is that the historical connection between AZ and SD has been disrupted by expansion southwards of the eastern Y haplogroup. Without further sampling, it is not possible to determine exactly where the two Y haplogroups now meet. Significant autosomal gene flow occurs between Nevada and sites to the north, south and east (AZ, ID, UT, OR), but a distinct geneflow boundary occurs between western NV and geographically proximal sites directly west in CA (CA1-4). Discontinuous habitat currently separates the Nevada sites from those on the Sierra Nevada (CA) range, and historical glaciers and lakes, for example Lake Lahontan to the east and Lake Bonneville to the west, may have previously created barriers between Nevada and both the Sierra and Rocky mountain ranges (Waltari & Guralnick, 2009). Although MPB from sites in southern CA (CA5/6) produce sterile male offspring when crossed with AZ2 populations (Bracewell et al., 2017), it has yet to be determined whether crosses with MPB from central CA and eastern NV also produce sterile males.

Beyond the phylogeographic signal suggesting three lineages with distinct Y haplogroups, extensive autosomal and X-linked gene flow likely obscures more detailed phylogeographic relationships. For example, ABBA-BABA tests suggested that geographic sites from the eastern Y haplogroup experience extensive autosomal gene flow with sites from both the western and central Y haplogroups, and there is clear evidence for admixture within each of the Y haplogroups. SNAPP analysis of autosomal loci yielded a relatively stable, well-supported tree when sites from the central geographic portions of each Y haplogroup were included (Appendix S13). However, when sites close to Y haplogroup boundaries were included, autosomal SNAPP analysis yielded unstable and poorly supported trees (Figure 3). Thus, autosomal gene flow among populations likely causes the unstable tree topology that limits our ability to make inferences about postglacial range expansion. Geographic sites within the central Y haplogroup (AZ and NV) do appear monophyletic relative to the other Y haplogroups. However, the placement of geographic sites with the eastern Y haplogroup in the autosomal tree is unclear, and they may occur sister or within both alternative clades of the tree depending on the sites included in the analysis (Figure 3).

4.3 | Evidence for adaptation within and across Y haplogroups

Adaptation to climate, rather than generating broad environmental genetic clines across the species range, has led to more locally limited clines in MPB. In particular, environmental association analysis suggested largely distinct sets of loci among Y haplogroups associated with variation in the first two principal components of 14 environmental variables (Table 1; Appendix S15). There was statistically significant overlap (i.e., more common, significant loci than expected by chance) in PC2-associated loci between the western and eastern Y haplogroups, most exhibiting the same directionality of the relationship (i.e., same sign on regression coefficients). It is possible that linkage disequilibrium inflates the estimate of common, independent loci, but 30 of the 37 shared loci mapped to different scaffolds with a mean length of 1,230 kb, suggesting relatively free recombination in the absence of major structural polymorphisms. Days with precipitation, latitude, longitude and elevation loaded most strongly on PC2 suggesting greater overlap of adaptive loci for these variables compared to those that load strongly on PC1 (e.g., mean temperature and degree days). However, environmental association analysis applied to frost-free days alone (loads strongly on PC1) identified significant common loci in the eastern and western Y haplogroups similar to the pattern observed for PC2 (Appendix S16). Overall, the results suggest a relatively weak signal for common adaptive loci; most significant locus-to-environment associations were specific to each haplogroup. Moreover, there was no statistical evidence for overlap of environmentally associated loci between either the western or eastern Y haplogroups and the central Y haplogroup.

We suggest that this evidence for adaptation at distinct sets of loci in different portions of the geographic range is consistent with restricted gene flow and independent, adaptive evolution among Y haplogroups. Seasonality exerts strong selection on MPB life history traits, generating clear latitudinal phenotypic clines (Bentz et al., 2001). These clines are apparent on both sides of the Great Basin, and northern populations require less time to develop in common garden experiments, compared to southern populations, matching shorter northern growing seasons (Bentz et al., 2014). Cosmopolitan species such as *Drosophila melanogaster* often demonstrate strong and continuous genetic clines along comparable or greater latitudinal ranges that track phenotypic differentiation (Berry & Kreitman, 1993; Hoffmann & Weeks, 2006). Environmental association analysis in this study, however, identified distinct sets of clinal loci (environmentally associated, corrected for population structure) that may underlie seasonal adaptation in the three different geographic regions corresponding to different Y haplogroups (Table 1). The extent of shared, adaptive polymorphism is consistent with our estimates of gene flow; the western and eastern Y haplogroups that exhibit higher cross-Y haplogroup gene flow share more environmentally associated loci than do the western and central Y haplogroups that have marked cross-Y haplogroup geneflow restriction (the central Y haplogroup was the only monophyletic Y haplogroup, and F_{ST} was significantly reduced between vs. within Y haplogroups). The Great Basin itself probably restricts east–west gene flow as well, but given the proximity of some NV populations with Sierra Nevada populations, this restriction is clearly bolstered by reproductive incompatibilities. Thus, sex chromosome evolution in MPB appears to constrain the continuous exchange of alleles, therefore disrupting continuous allele frequency clines corresponding to clines in environmental factors.

The underlying genetic architecture of development time, the trait most strongly associated with seasonality in MPB, may also influence the formation of geographic clines in allele frequency. Development time is a complex, polygenic trait, and there are likely many allelic combinations at many loci that will produce similar developmental phenotypes. Clines in such polygenic traits do not necessarily predict underlying genetic clines. This is nicely illustrated by a study in pitcher plant mosquitoes demonstrating that QTL analyses of diapause timing, another key trait for seasonal adaptation, identify different QTL in different genetic backgrounds using crosses of different pairs of geographic populations (Bradshaw, Emerson, Catchen, Cresko, & Holzapfel, 2012).

5 | CONCLUSION

Genetic structure across the MPB range is consistent with periods of allopatry during the LGM, during which rapid evolution on the Y chromosome in each refugium led to varying degrees of reproductive isolation after postglacial expansion and secondary contact. This process, driven by differential degradation of the Y in different refugia (Bracewell et al, 2017), has resulted in restricted autosomal gene flow across secondary contact zones, despite currently permissive habitats for dispersal. Although Great Basin sky island *Pinus* forests could provide a possible route for gene flow, our data suggest distinct lineages (Y haplogroups) occur on the western and eastern sides of the Great Basin. This may reflect real geographic barriers (past or current), but it is also consistent with Y-linked reproductive incompatibilities limiting gene flow. In combination with barriers caused by Y-linked reproductive isolation, the process of adaptation along environmental gradients appears to have unfolded largely independently within each of the three distinct Y haplogroups. These patterns suggest that relatively recent histories of range contraction and expansion can produce genetic differentiation on the sex chromosomes leading to genetic divergence and reproductive barriers to gene flow that may reflect the early stages of species formation.

ACKNOWLEDGEMENTS

We thank Kathy Bleiker, Tom Coleman, Al Dymerski, Ken Gibson, Matt Hansen, Camille Jensen, Sandy Kegley, Staffan Lindgren, Ann Lynch, Joel McMillin, Connie Mehmel, Jose Negron, Dana Perkins, Greta Schen, Don Scott, Dave Schultz, Ken Sterling, Jim Vandygriff and Craig Willcox for help with field collections, Jacqueline Lopez and the Notre Dame Genomics Core for assistance with library preparation and QC, Joseph Sarro for assistance with data processing and funding from the UDSA Forest Service, Rocky Mountain Research Station and NSF IOS 1451274 to GJR.

DATA ACCESSIBILITY

Data are accessible through dryad: https://doi.org/10.5061/dryad.vf8sm and NCBI: Bioproject PRJNA398557. This includes the following: raw reads; sample barcode information (for demultiplexing); lists of autosomal, X and Y scaffolds, and VCF file with genotype probabilities for all individuals.

AUTHOR CONTRIBUTIONS

All authors contributed to the experimental design, B.J.B. and R.R.B. collected beetles and larvae, G.J.R. and M.E.P. prepared the RAD libraries, E.J.D. and G.J.R. analysed the data and wrote the initial draft, and all authors contributed to draft revisions.

ORCID

Eddy J. Dowle http://orcid.org/0000-0003-2265-6281
Gregory J. Ragland http://orcid.org/0000-0002-6501-3641

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664.

Batista, P. D., Janes, J. K., Boone, C. K., Murray, B. W., & Sperling, F. A. H. (2016). Adaptive and neutral markers both show continent-wide population structure of mountain pine beetle (*Dendroctonus ponderosae*). *Ecology and Evolution*, *6*, 6292–6300.

Bentz, B., Bracewell, R., Mock, K., & Pfrender, M. (2011). Genetic architecture and phenotypic plasticity of thermally-regulated traits in an eruptive species, *Dendroctonus ponderosae*. *Evolutionary Ecology*, *25*, 1269–1288.

Bentz, B., Hood, S. A., Hansen, E. M., Vandygriff, J. C., & Mock, K. E. (2017). Defense traits in the long-lived Great Basin bristlecone pine and resistance to the native herbivore mountain pine beetle. *New Phytologist*, *213*, 611–624.

Bentz, B. J., Logan, J. A., & Amman, G. D. (1991). Temperature-dependent development of the mountain pine beetle (Coleoptera: Scolytidae) and simulation of its phenology. *Canadian Entomologist*, *123*, 1083.

Bentz, B., Logan, J., & Vandygriff, J. (2001). Latitudinal variation in *Dendroctonus ponderosae* (Coleoptera: Scolytidae) development time and adult size. *The Canadian Entomologist*, *133*, 375–387.

Bentz, B., Vandygriff, J., Jensen, C., Coleman, T., Maloney, P., Smith, S., … Schen-Langenheim, G. (2014). Mountain pine beetle voltinism and life history characteristics across latitudinal and elevational gradients in the western United States. *Forest Science*, *60*, 434–449.

Berg, R. J. W., Rebel, H., van der Horst, G. T. J., van Kranen, H. J., Mullenders, L. H. F., van Vloten, W. A., & de Gruijl, F. R. (2000). Impact of global genome repair versus transcription-coupled repair on ultraviolet carcinogenesis in hairless mice. *Cancer Research*, *60*, 2858–2863.

Berry, A., & Kreitman, M. (1993). Molecular analysis of an allozyme cline: Alcohol dehydrogenase in Drosophila melanogaster on the east coast of North America. *Genetics*, *134*, 869–893.

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting F_{ST}: The impact of rare variants. *Genome Research*, *23*, 1514–1521.

BioSIM. (2014). Retrieved from https://cfs.nrcan.gc.ca/projects/133

Blomquist, G. J., Figueroa-Teran, R., Aw, M., Song, M., Gorzalski, A., Abbott, N. L., … Tittiger, C. (2010). Pheromone production in bark beetles. *Insect Biochemistry and Molecular Biology*, *40*, 699–712.

Bouckaert, R. R. (2010). DENSITREE: Making sense of sets of phylogenetic trees. *Bioinformatics*, *26*, 1372–1373.

Bracewell, R. R., Bentz, B. J., Sullivan, B. T., & Good, J. M. (2017). Rapid neo-sex chromosome evolution and incipient speciation in a major forest pest. Nature Communications, https://doi.org/10.1038/s41467-017-01761-4

Bracewell, R. R., Pfrender, M. E., Mock, K. E., & Bentz, B. J. (2011). Cryptic postzygotic isolation in an eruptive species of bark beetle (*Dendroctonus ponderosae*). *Evolution*, *65*, 961–975.

Bracewell, R. R., Pfrender, M. E., Mock, K. E., & Bentz, B. J. (2013). Contrasting geographic patterns of genetic differentiation in body size and development time with reproductive isolation in *Dendroctonus ponderosae* (Coleoptera: Curculionidae, Scolytinae). *Annals of the Entomological Society of America*, *106*, 385–391.

Bradshaw, W. E., Emerson, K. J., Catchen, J. M., Cresko, W. A., & Holzapfel, C. M. (2012). Footprints in time: Comparative quantitative trait loci mapping of the pitcher-plant mosquito, *Wyeomyia smithii*. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*, 4551–4558.

Brunelle, A., Rehfeldt, G. E., Bentz, B., & Munson, A. S. (2008). Holocene records of Dendroctonus bark beetles in high elevation pine forests of Idaho and Montana, USA. *Forest Ecology and Management*, *255*, 836–846.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*, 1917–1932.

Carling, M. D., & Brumfield, R. T. (2008). Haldane's rule in an avian system: Using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the Passerina bunting hybrid zone. *Evolution*, *62*, 2600–2615.

Charlesworth, B., Coyne, J. A., & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist*, *130*, 113–146.

Cognato, A. I., Harlin, A. D., & Fisher, M. L. (2003). Genetic structure among pinyon pine beetle populations (Scolytinae: *Ips confusus*). *Environmental Entomology*, *32*, 1262–1270.

Cole, K. L., Fisher, J., Arundel, S. T., Cannella, J., & Swift, S. (2008). Geographical and climatic limits of needle types of one- and two-needled pinyon pines. *Journal of Biogeography*, *35*, 257–269.

Cole, K. L., Fisher, J. F., Ironside, K., Mead, J. I., & Koehler, P. (2013). The biogeographic histories of *Pinus edulis* and *Pinus monophylla* over the last 50,000 years. *Quaternary International*, *310*, 96–110.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.

Cullingham, C. I., Cooke, J. E. K., Dang, S., Davis, C. S., Cooke, B. J., & Coltman, D. W. (2011). Mountain pine beetle host-range expansion threatens the boreal forest. *Molecular Ecology*, *20*, 2157–2171.

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Clifford Lane, H., & Lempicki, R. A. (2003). DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biology*, *4*, 1–11.

Dobzhansky, T. (1937). *Genetics and the origin of species*. New York: Columbia University Press.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUTI and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973.

Egan, S. P., Ragland, G. J., Assour, L., Powell, T. H. Q., Hood, G. R., Emrich, S., … Feder, J. L. (2015). Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecology Letters*, *18*, 817–825.

Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. *Annual Review of Genetics*, *10*, 253–280.

Fettig, C. J., Gibson, K. E., Munson, A. S., & Negrón, J. F. (2014). Cultural practices for prevention and mitigation of mountain pine beetle infestations. *Forest Science*, *60*, 450–463.

Franceschi, V. R., Krokene, P., Christiansen, E., & Krekling, T. (2005). Anatomical and chemical defenses of conifer bark against bark beetles and other pests. *New Phytologist*, *167*, 353–376.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*, 1687–1699.

Fumagalli, M., Vieira, F. G., Korneliussen, T. S., Linderoth, T., Huerta-Sánchez, E., Albrechtsen, A., & Nielsen, R. (2013). Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*, *195*, 979–992.

Fumagalli, M., Vieira, F. G., Linderoth, T., & Nielsen, R. (2014). NGSTOOLS: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*, *30*, 1486–1487.

Geraldes, A., Basset, P., Gibson, B., Smith, K. L., Harr, B., Yu, H. T., … Nachman, M. W. (2008). Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*, *17*, 5349–5363.

Geraldes, A., Carneiro, M., Delibes-Mateos, M., Villafuerte, R., Nachman, M. W., & Ferrand, N. (2008). Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Molecular Ecology*, *17*, 4489–4499.

Godbout, J., Fazekas, A., Newton, C. H., Yeh, F. C., & Bousquet, J. (2008). Glacial vicariance in the Pacific Northwest: Evidence from a lodgepole pine mitochondrial DNA minisatellite for multiple genetically distinct and widely separated refugia. *Molecular Ecology*, *17*, 2463–2475.

Green, R. E., Krause, J., Briggs, A. W., Maricic, R., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the neandertal genome. *Science*, *328*, 710–722.

Gugger, P. F., Sugita, S., & Cavender-Bares, J. (2010). Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: Testing hypotheses from the fossil record. *Molecular Ecology*, *19*, 1877–1897.

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*, 205–220.

Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, *12*, 101–109.

Hoffmann, A. A., & Weeks, R. A. (2006). Climatic selection on genes and traits after a 100 year-old invasion: A critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica*, *129*, 133–147.

Hope, A. G., Speer, K. A., Demboski, J. R., Talbot, S. L., & Cook, J. A. (2012). A climate for speciation: Rapid spatial diversification within the *Sorex cinereus* complex of shrews. *Molecular Phylogenetics and Evolution*, *64*, 671–684.

Hudson, R. R., Slatkin, M., & Maddison, W. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*, 583–589.

Jackson, P. L., Straussfogel, D., Lindgren, B. S., Mitchell, S., & Murphy, B. (2008). Radar observation and aerial capture of mountain pine beetle, Dendroctonus ponderosae Hopk. (Coleoptera: Scolytidae) in flight above the forest canopy. *Canadian Journal of Forest Research*, *38*, 2313–2327.

Jaramillo-Correa, J. P., Beaulieu, J., Khasa, D. P., & Bousquet, J. (2009). Inferring the past from the present phylogeographic structure of North American forest trees: Seeing the forest for the genes. *Canadian Journal of Forest Research*, *39*, 286–307.

Johnsen, A., Fidler, A. E., Kuhn, S., Carter, K. L., Hoffmann, A., Barr, I. R., … Kempenaers, B. (2007). Avian Clock gene polymorphism: Evidence for a latitudinal cline in allele frequencies. *Molecular Ecology*, *16*, 4867–4880.

Johnson, N. A., & Lachance, J. (2012). The genetics of sex chromosomes: Evolution and implications for hybrid incompatibility. *Annals of the New York Academy of Sciences*, *1256*, E1–22.

Keeling, C., Yuen, M., Liao, N., Docking, T. R., Chan, S. K., Taylor, G. A., ... Bohlmann, J. (2013). Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. *Genome Biology*, *14*, R27.

Klicka, J., Spellman, G. M., Winker, K., Chua, V., & Smith, B. T. (2011). A phylogeographic and population genetic analysis of a widespread, sedentary North American Bird: The hairy woodpecker (*Picoides villosus*). *The Auk*, *128*, 346–362.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., ... Schlötterer, C. (2011). POPOOLATION: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE*, *6*, e15925.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*, 356.

Lanier, G., & Wood, D. (1968). Controlled mating, karyology, morphology, and sex-ratio in the *Dendroctonus ponderosae* complex. *Annals of the Entomological Society of America*, *61*, 517–526.

Lee, T.-H., Guo, H., Wang, X., Kim, C., & Paterson, A. H. (2014). SNPHYLO: A pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, *15*, 162.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760.

Logan, J. A., & Bentz, B. J. (1999). Model analysis of mountain pine beetle (Coleoptera: Scolytidae) seasonality. *Environmental Entomology*, *28*, 924–934.

Macholán, M., Munclinger, P., Šugerková, M., Dufková, P., Bímová, B., Božíková, E., ... Piálek, J. (2007). Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution*, *61*, 746–771.

Malaney, J. L., Conroy, C. J., Moffitt, L. A., Spoonhunter, H. D., Patton, J. L., & Cook, J. A. (2013). Phylogeography of the western jumping mouse (*Zapus princeps*) detects deep and persistent allopatry with expansion. *Journal of Mammalogy*, *94*, 1016–1029.

Mank, J. E., Vicoso, B., Berlin, S., & Charlesworth, B. (2010). Effective population size and the faster-X effect: Empirical results and their interpretation. *Evolution*, *64*, 663–674.

Manthey, J. D., Klicka, J., & Spellman, G. M. (2011). Cryptic diversity in a widespread North American songbird: Phylogeography of the Brown Creeper (*Certhia americana*). *Molecular Phylogenetics and Evolution*, *58*, 502–512.

Mayr, E. (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. London, UK: Harvard University Press.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303.

Mock, K. E., Bentz, B. J., O'Neill, E. M., Chong, J. P., Orwin, J., & Pfrender, M. E. (2007). Landscape-scale genetic variation in a forest outbreak species, the mountain pine beetle (*Dendroctonus ponderosae*). *Molecular Ecology*, *16*, 553–568.

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21*, 2991–3005.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*, e190.

Payseur, B. A., Krenz, J. G., & Nachman, M. W. (2004). Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, *58*, 2064–2078.

Potter K. M., Hipkins V. D., Mahalovich, M. F., & Means, R. E. (2013). Mitochondrial DNA haplotype distribution patterns in *Pinus ponderosa* (Pinaceae): Range-wide evolutionary history and implications for conservation. *American Journal of Botany*, *100*, 1562–1579.

Potter, K. M., Hipkins, V. D., Mahalovich, M. F., & Means, R. E. (2015). Nuclear genetic variation across the range of ponderosa pine (*Pinus ponderosa*): Phylogeographic, taxonomic and conservation implications. *Tree Genetics & Genomes*, *11*, 1–23.

Powell, J. A., & Bentz, B. J. (2009). Connecting phenological predictions with population growth rates for mountain pine beetle, an outbreak insect. *Landscape Ecology*, *24*, 657–672.

Powell, J. A., & Bentz, B. J. (2014). Phenology and density-dependent dispersal predict patterns of mountain pine beetle (*Dendroctonus ponderosae*) impact. *Ecological Modelling*, *273*, 173–185.

Presgraves, D. C. (2008). Sex chromosomes and speciation in Drosophila. *Trends in Genetics*, *24*, 336–343.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, pp. Retrieved from http://www.r-project.org/

Raffa, K. F., & Berryman, A. A. (1983). Physiological aspects of lodgepole pine wound responses to a fungal symbiont of the mountain pine beetle, *Dendroctonus ponderosae* (Coleoptera: Scolytidae). *The Canadian Entomologist*, *115*, 723–734.

Rambaut, A. (2009). FIGTREE v1.3.1. Retrieved from http://tree.bio.ed.ac.uk/software/figtree

Rambaut, A., & Drummond, A. (2007). *TRACER v1. 4*. Retrieved from http://tree.bio.ed.ac.uk/software/tracer/

Régnière, J. (1996). Generalized approach to landscape-wide seasonal forecasting with temperature-driven simulation models. *Environmental Entomology*, *25*, 869–881.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*, 489–494.

Richardson, B. A., Brunsfeld, S. J., & Klopfenstein, N. B. (2002). DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (*Pinus albicaulis*). *Molecular Ecology*, *11*, 215–227.

Roberts, D. R., & Hamann, A. (2015). Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society of London B: Biological Sciences*, *282*, 20142903.

Sequeira, A. S., Normark, B. B., & Farrell, B. D. (2000). Evolutionary assembly of the conifer fauna: Distinguishing ancient from recent associations in bark beetles. *Proceedings of the Royal Society of London B: Biological Sciences*, *267*, 2359–2366.

Shafer, A. B. A., Cullingham, C. I., Cote, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, *19*, 4589–4621.

Six, D. L. (2012). Ecological and evolutionary determinants of bark beetle—fungus symbioses. *Insects*, *3*, 339–366.

Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M. L., ... Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature*, *525*, 104–108.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*, 693–702.

Spellman, G. M., & Klicka, J. (2007). Phylogeography of the white-breasted nuthatch (*Sitta carolinensis*): Diversification in North American pine and oak woodlands. *Molecular Ecology*, *16*, 1729–1740.

Stock, M. W., Pitman, G. B., & Guenther, J. D. (1979). Genetic differences between Douglas-fir beetles (*Dendroctonus pseudotsugae*) from Idaho and Coastal Oregon. *Annals of the Entomological Society of America*, *72*, 394–397.

Swenson, N., & Howard, D. (2005). Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist*, *166*, 581–591.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595.

Turner, L. M., & Harr, B. (2014). Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. *eLife*, *3*, e02504.

Waltari, E., & Guralnick, R. P. (2009). Ecological niche modelling of montane mammals in the Great Basin, North America: Examining past and present connectivity of species across basins and ranges. *Journal of Biogeography*, *36*, 148–161.

Watterson, G. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*, 256–276.

Wickham, H. (2009). GGPLOT2: *Elegant graphics for data analysis*. Houstan , TX: Springer Science & Business Media.

Wood, S. L. (1982). The bark and ambrosia beetles of North and Central America (Coleoptera: Scolytidae), a taxonomic monograph. *Great Basin Naturalist Memoirs*, *6*, 1–1356.

How to cite this article: Dowle EJ, Bracewell RR, Pfrender ME, Mock KE, Bentz BJ, Ragland GJ. Reproductive isolation and environmental adaptation shape the phylogeography of mountain pine beetle (*Dendroctonus ponderosae*). *Mol Ecol*. 2017;26:6071–6084. https://doi.org/10.1111/mec.14342

SUPPORTING INFORMATION

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