Received: 6 April 2020 | Accepted: 28 June 2021

DOI: 10.1111/mec.16056

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Multigenerational backcrossing and introgression between two woodrat species at an abrupt ecological transition

Joshua P. Jahner[1,2] | Thomas L. Parchman[1,3] | Marjorie D. Matocq[3,4]

[1]Department of Biology, University of Nevada, Reno, Nevada, USA

[2]Department of Botany, University of Wyoming, Laramie, Wyoming, USA

[3]Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, Nevada, USA

[4]Department of Natural Resources and Environmental Science, University of Nevada, Reno, Nevada, USA

Correspondence
Joshua P. Jahner, Department of Botany, University of Wyoming, Laramie, WY, USA.
Email: jpjahner@gmail.com

Funding information
Office of Integrative Activities, Grant/Award Number: 1826801; Division of Integrative Organismal Systems, Grant/Award Number: 1457209; National Science Foundation

Abstract

When organisms experience secondary contact after allopatric divergence, genomic regions can introgress differentially depending on their relationships with adaptation, reproductive isolation, recombination, and drift. Analyses of genome-wide patterns of divergence and introgression could provide insight into the outcomes of hybridization and the potential relationship between allopatric divergence and reproductive isolation. Here, we generate population genetic data (26,262 SNPs; 353 individuals) using a reduced-representation sequencing approach to quantify patterns of ancestry, differentiation, and introgression between a pair of ecologically distinct mammals—the desert woodrat (*N. lepida*) and Bryant's woodrat (*N. bryanti*)—that hybridize at a sharp ecotone in southern California. Individual ancestry estimates confirmed that hybrids were rare in this bimodal hybrid zone, and entirely consisted of a few F_1 individuals and a broad range of multigenerational backcrosses. Genomic cline analyses indicated more than half of loci had elevated introgression from one genomic background into the other. However, introgression was not associated with relative or absolute measures of divergence, and loci with extreme values for both were not typically found near detoxification enzymes previously implicated in dietary specialization for woodrats. The decoupling of differentiation and introgression suggests that processes other than adaptation, such as drift, may underlie the extreme clines at this contact zone.

KEYWORDS
bimodal hybrid zone, ecotone, genomic clines, genotyping-by-sequencing, hybridization, *Neotoma*

1 | | INTRODUCTION

The evolutionary outcome of secondary contact and hybridization between diverging lineages is regulated by the strength of reproductive barriers and the genomic architecture of differentiation (Barton & Hewitt, 1985; Burke & Arnold, 2001; Gompert et al., 2017). For strongly differentiated lineages, the factors contributing to reproductive isolation are often diverse, and can include both prezygotic and postzygotic mechanisms (Coyne & Orr, 2004). In extreme instances, hybrid breakdown can occur, where advanced-generation hybrids such as F_2 hybrids or backcrosses are rarely if ever viable (Burton et al., 2013; Clausen, 1951; Dobzhansky, 1970). However, if reproductive isolation is incomplete (i.e., if at least some hybrids are viable and fertile), introgression and novel recombinants between parental genomes can arise. In this scenario, genomic regions that diverged in allopatry can introgress at different rates, potentially depending upon their relationship with adaptation and reproductive isolation (Christe et al., 2016; Gompert et al., 2012; Lindtke & Buerkle, 2015). As such, variation in ancestry generated across hybrid zones provides an opportunity to evaluate the ecological and evolutionary factors associated with reproductive isolation (Gompert et al., 2017; Harrison & Larson, 2016; Ravinet et al., 2017).

Genome-wide heterogeneity in the magnitude of differentiation is often quantified to identify regions of elevated divergence

potentially involved in adaptation (Hoban et al., 2016; Seehausen et al., 2014; Wolf & Ellegren, 2017). While divergence can arise from adaptation and be associated with reproductive isolation, it can also be influenced by genetic drift, recombination rate heterogeneity, structural rearrangements, and indirect selection (Burri, 2017; Burri et al., 2015; Cruickshank & Hahn, 2014; Haasl & Payseur, 2016; Ravinet et al., 2017; Semenov et al., 2019; Stankowski et al., 2019; Wolf & Ellegren, 2017). Despite such limitations, genome scans have yielded insight into the genetic basis of adaptation in many systems (e.g., Christe et al., 2017; Hohenlohe et al., 2010; Lamichhaney et al., 2015; Pfeifer et al., 2018; Tavares et al., 2018), even for reduced-representation sequencing data sets for which marker density is often limited relative to the extent of linkage disequilibrium (Catchen et al., 2017; McKinney et al., 2017; but see Lowry et al., 2017).

While genome scans have often been used for describing regions shaped by adaptation or influenced by variation in gene flow, the characterization of hybrid ancestry and introgression can offer additional perspective on the genetic basis of reproductive isolation (e.g., Lindtke et al., 2014; Mandeville et al., 2017; Schumer et al., 2017). Multiple frameworks are available for quantifying introgression (Payseur & Rieseberg, 2016), including phylogenetic (e.g., *D*-statistics; Green et al., 2010) and cline-based approaches for population genetic data. Geographic clines use physical transects across hybrid zones to quantify allele frequency changes across space, with cline width serving as an estimate of the strength of dispersal and selection in a tension zone at equilibrium (Baldassarre et al., 2014; Barton & Hewitt, 1985; Endler, 1977; Souissi et al., 2018). In contrast, genomic clines are not reliant on spatial sampling, and instead quantify how locus-specific ancestry changes across the range of genome-wide ancestry found in hybrids (i.e., hybrid index; Fitzpatrick, 2013; Gompert & Buerkle, 2011; Macholán et al., 2011; Szymura & Barton, 1986).

The identification of loci with extreme genomic cline parameters could facilitate an understanding of the genetic basis of reproductive isolation, as elevated cline centre (α) estimates are indicative of strong introgression from one genomic background into the other, and elevated cline rate (β) estimates are consistent with either heightened or suppressed introgression at a locus (Gompert & Buerkle, 2011; Gompert, Parchman, et al., 2012). Loci with both strong differentiation and exceptional rates of introgression are particularly intriguing because these combinations could be expected for adaptive introgression (fast introgression of divergent alleles) or reproductive barriers (slow introgression of divergent alleles) (Gompert & Buerkle, 2011; McFarlane et al., 2021). Multiple previous studies have reported an association between locus-specific estimates of introgression and differentiation (e.g., Baiz et al., 2019; Gompert, Lucas, et al., 2012; Nosil et al., 2012; Oswald et al., 2019; Parchman et al., 2013; Schield et al., 2017; Taylor et al., 2014), suggesting that heavily differentiated genomic regions might also influence hybrid fitness. It is worth noting, however, that extreme genomic cline estimates can arise from stochastic processes (Gompert & Buerkle, 2011; Gompert, Parchman, et al., 2012; McFarlane et al., 2021), and thus do not necessarily stem from selection or reproductive isolation. Because genomic clines do not depend on sampling geographic transects, they allow for the analysis of introgression for mosaic or spatially restricted hybrid zones (e.g., McFarlane et al., 2021; Simon et al., 2021). Here, we employ genomic clines to quantify genome-wide heterogeneity in introgression for a pair of hybridizing woodrats at a highly localized contact zone, and then ask how introgression relates to differentiation across the genome.

The North American genus *Neotoma* (Cricetidae) consists of ~20 species of woodrats that consume a wide diversity of plants and are known to occasionally hybridize (Coyner et al., 2015; Hunter et al., 2017; Mauldin et al., 2021; Patton et al., 2007). For example, the sister species *N. lepida* and *N. bryanti* are estimated to have diverged in the mid Pleistocene (~1.6 million years ago; Patton et al., 2007), yet they meet and hybridize in southern California across a sharp ecological transition (Figure 1a). The outcome is a bimodal hybrid zone (Jiggins & Mallet, 2000) in which individuals with pure parental ancestry are common and hybrids are relatively rare (~13% of the population has admixed ancestry; Shurtliff et al., 2014). Here, *N. bryanti* and backcross (BC) *bryanti* are largely restricted to the boulder-strewn, relatively mesic habitat of the southeastern slopes of the Sierra Nevada ("the hill"), whereas *N. lepida* and BC-*lepida* are found in the adjacent Mojave Desert scrub habitat ("the flats") (Figure 1a). Each species is tightly associated with its respective habitat and maintains a habitat specific diet (Matocq et al., 2020; Nielsen & Matocq, 2021; Shurtliff et al., 2014). The diet of *N. bryanti* is dominated by California coffeeberry (*Frangula californica*: Rhamnaceae), whereas the diet of *N. lepida* is predominantly composed of desert almond (*Prunus fasciculata*: Rosaceae) (Nielsen & Matocq, 2021). Both plants contain secondary compounds (*Frangula*, anthraquinones; *Prunus*, cyanogenic glycosides) that are potentially toxic to these mammals (Dong et al., 2016; Newton et al.,1981) and may require specialized metabolic processing. Woodrats have been shown to use a family of cytochrome p450 enzymes to detoxify plant secondary compounds (Kitanovic et al., 2018; Magnanou et al., 2009; Malenke et al., 2012), and the specific metabolic processing needed to consume the plants available at this site may be facilitated by differentiation or introgression of alleles at these loci.

Using a suite of microsatellite loci and a 4-year mark-recapture analysis, Shurtliff et al. (2014) showed that F_1 individuals are rare (2% of total population) and proposed that the strong habitat-based spatial segregation at the site played a role in limiting interspecific mating opportunities. Further, F_1 individuals are most likely to backcross with the habitat-specific pool of pure parentals in which they live. Laboratory trials demonstrated that even when pure parentals come into contact, *N. lepida* females strongly prefer to mate with conspecific males, whereas *N. bryanti* females will mate with males of both species, suggesting that most F_1 hybrids are generated by crosses of *N. bryanti* females and *N. lepida* males (Shurtliff et al., 2013). Shurtliff et al. (2014) also found evidence of postzygotic isolation in this system, where juvenile hybrids survived to their first year of adulthood at less than half the rate of their purebred counterparts. In sum, habitat-based mating opportunities and asymmetric mate-choice, coupled with putative extrinsic (i.e., ecological) and

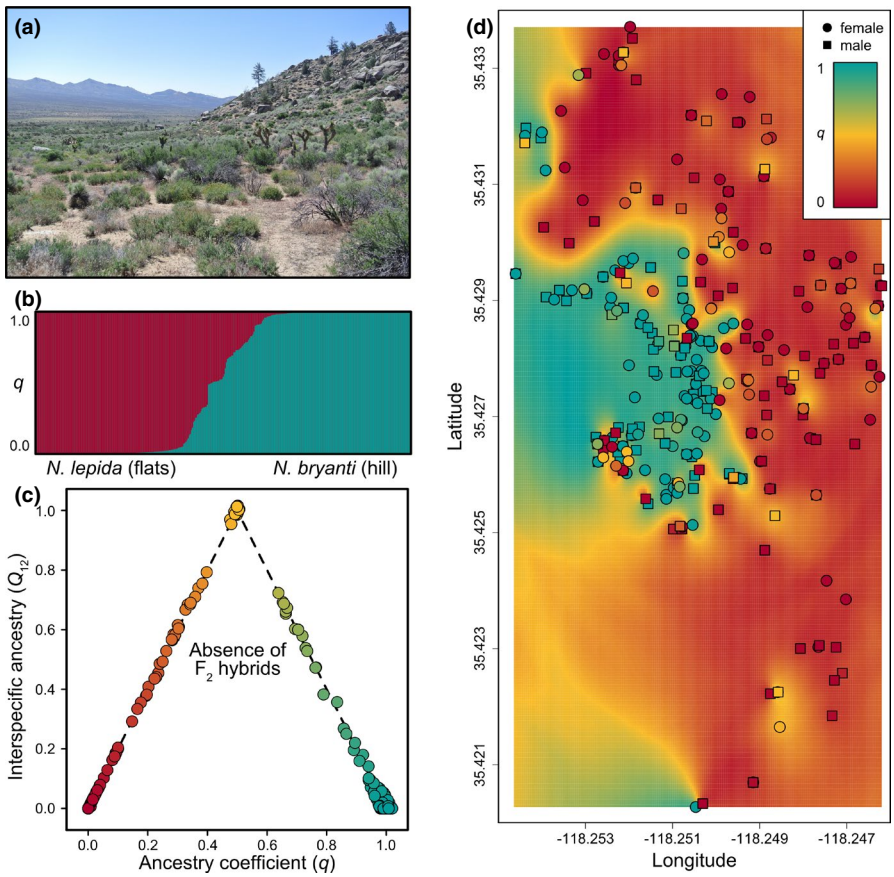

FIGURE 1 (a) The contact zone at Whitney Well, California spans a sharp ecotone across the xeric "flats" habitat typically occupied by *N. lepida* and the mesic "hill" habitat typically occupied by *N. bryanti* (photo by Danny Nielsen, taken from the north looking southeast). (b) The distribution of ancestry coefficients (q) is consistent with a bimodal hybrid zone. Individuals are ordered along the x-axis by increasing values of q. See Figure S2 for estimates of uncertainty. (c) Individual ancestry classes were assigned based on estimates of ancestry coefficients (q) and interspecific ancestry (Q_{12}) generated with `entropy` (Gompert et al., 2014; Shastry et al., 2021). Pure parental *N. lepida* and *N. bryanti* individuals have ancestry coefficients of zero and one, respectively. All individuals residing on the dashed lines, which depict the theoretical maximum value of Q_{12} for a given q, have at least one parent with pure parental ancestry (e.g., pure parentals, F_1 hybrids, backcrosses), whereas individuals in the centre of the triangle have two parents with admixed ancestry (e.g., F_2 and F_3 hybrids) (Gompert et al., 2014; Shastry et al., 2021). Points were jittered to avoid overplotting. (d) Individual ancestry coefficients (q) are plotted at each individual's trap location. The interpolated q surface was generated based on jittered locality data using the `kriging` package (Olmedo, 2014) in `R`. Although *N. bryanti* are largely restricted to the "hill" to the west and most *N. lepida* are found in the "flats" to the east (a few were trapped on the hill), hybrid individuals can be found throughout the study area [Colour figure can be viewed at wileyonlinelibrary.com]

intrinsic (i.e., genomic) sources of postzygotic selection, potentially play roles in determining patterns of introgression in this system.

Here, we characterized genome-wide patterns of differentiation and introgression between *N. bryanti* and *N. lepida* to further examine the genomic consequences of ecological segregation and hybridization in this system. Specifically, we sought to: (i) characterize the distribution of ancestry in hybrids and determine the relative frequencies of potential hybrid classes; (ii) quantify genome-wide variation in introgression and differentiation, and ask how these estimates relate to one another; and (iii) examine the potential for loci with exceptional levels of differentiation or introgression to be proximal to cytochrome P450 genes, since these loci may be involved in metabolic processing of the diets available across the sharp environmental gradient of this contact zone. We found evidence for multigenerational backcrossing between *N. lepida* and *N. bryanti* along with the complete absence of F_2 hybrids, and detect a large number of loci with accentuated α estimates indicating elevated but asymmetric introgression. However, there was no correspondence between divergence and introgression, illustrating the potential for processes other than selection to influence extreme genomic clines in this system.

2 | | MATERIALS AND METHODS

2.1 | | Sample collection

Field collections were conducted as described in Shurtliff et al. (2014) at the Whitney Well contact zone in southern California (Figure 1a). Both areas were extensively surveyed using Sherman and Tomahawk live-traps over a 4-year period; each individual was uniquely marked with metal ear tags and a small ear biopsy was taken as a source

of DNA (Shurtliff et al., 2014). Juveniles emerge from the nest and can be live-trapped at approximately 1 month of age, so all genetic data are from individuals that survived at least to this age. DNA was extracted from ear tissue samples using Qiagen DNeasy Blood and Tissue kits (Qiagen Inc.). Original DNA extracts from the previous study ($N = 600$) were quality-checked and quantified using fluorometry. To focus our sequencing effort in the present study on the full spectrum of hybrid ancestries recovered by Shurtliff et al. (2014), we included nearly all hybrid individuals from their study (19 BC-*bryanti*, 27 BC-*lepida*, and 21 individuals designated as either F_1/F_2 or "other") and a subset of the individuals they identified as parentals (156 *N. bryanti*; 160 *N. lepida*) for a total of 383 individuals. See Supporting Information Methods for an analysis of how sample size influences the probability of sampling offspring generated from hybrid × hybrid crosses.

2.2 | | Genotyping-by-sequencing, assembly, and variant calling

We constructed reduced-representation libraries for Illumina sequencing using a genotyping-by-sequencing approach (GBS; Parchman et al., 2012) that is analogous to ddRADseq (Peterson et al., 2012). We used two restriction enzymes, *Eco*RI and *Mse*I, to digest genomic DNA, and ligated DNA barcoded adaptors to digested fragments from each individual. We ligated Illumina adaptors embedded with eight to 10 base-pair (bp) DNA barcode sequences to the *Eco*RI ends of fragments, and ligated standard Illumina adaptors to the *Mse*I cut sites. Barcoded restriction-ligation products from all individuals were subsequently pooled and PCR amplified with Illumina primers and a proofreading polymerase (Iproof; BioRad). Libraries were size-selected for a region between 350 and 425 bp using a PippinPrep quantitative electrophoresis unit (Sage Science). We generated single end, 100 bp reads on three lanes of the Illumina HiSeq 2500 at the University of Texas Genomic Sequencing and Analysis Facility.

We first discarded potential contaminant reads (PhiX, *Escherichia coli*) or those matching portions of any Illumina oligos using `bowtie 2 _ db` (Langmead & Salzberg, 2012) and a series of scripts designed for contaminant cleaning (https://github.com/ncgr/tapioca). We used a Perl script to match barcode sequences to individuals and to correct one or two base barcode errors. Individual fastq files are available at Dryad (Jahner et al., 2021). Due to poor sequencing performance, 13 individuals were not considered further. Reads from the remaining 370 individuals were aligned to the draft *N. lepida* assembly (Campbell et al., 2016) using the *aln* and *samse* algorithms in `bwa 0.7.8` (Li & Durbin, 2009), specifying an edit distance of four and otherwise default settings. This early draft assembly is highly fragmented (scaffold N50 = 138 kbp, N90 = 7 kbp; Campbell et al., 2016), but is nonetheless useful as a reference for aligning reduced-representation reads for variant calling. Sequence alignment map (SAM) files generated by `bwa` were converted to binary alignment (BAM) format with `samtools 1.3` (Li et al., 2009).

Single nucleotide polymorphisms (SNPs) were identified and genotype likelihoods calculated with `samtools 1.3` and `bcftools 1.3` (Li et al., 2009), requiring minimum base, site, and mapping qualities of 20, and a minimum genotype quality of 10. The potential pool of loci was further filtered using `vcftools 0.1.14` (Danecek et al., 2011). We retained a single biallelic SNP per contig, where at least 60% of individuals were represented by at least one read and the minor allele frequency (maf) was >0.05. Individuals were additionally excluded if they were missing data for more than half of the remaining loci. Mapping reads to the draft *N. lepida* reference genome should minimize the over-assembly of paralogous genomic regions and associated genotype error in these variants. Nonetheless, to further safeguard against this issue, and the potential of over-assembly across repetitive regions, we removed loci with a mean coverage >10× per individual (3.2% of loci).

2.3 | | The distribution of ancestry across a sharp ecotone

Based on the genotype likelihoods calculated above, we coincidentally estimated genotype probabilities and global ancestry coefficients (q) using a hierarchical Bayesian model (the q model of `entropy`, described in Gompert et al., 2014; Shastry et al., 2021) that is based on the correlated allele frequency model of `structure` (Falush et al., 2003; Pritchard et al., 2000). Importantly, this model estimates genotype probabilities and ancestry coefficients while incorporating uncertainty arising from variability in sequencing coverage as well as sequencing and mapping error across individuals and loci into parameter estimation (Buerkle & Gompert, 2013; Gompert et al., 2014). The incorporation of genotype uncertainty can be critical for low to medium coverage reduced-representation sequencing data sets, as simulations have demonstrated that parameter estimation is more precise and less biased when researchers maximize individual sample sizes as opposed to sequencing depth (Buerkle & Gompert, 2013; Fumagalli et al., 2013). `entropy` does not make a priori assumptions about the population origin of individuals, and assumes that individual genomes consist of loci with ancestry from one of k ancestral populations. To aid MCMC convergence, we initialized ancestry coefficients in chains using cluster membership probabilities generated from a k-means clustering of principal components estimated from genotype likelihoods in a linear discriminant analysis using the `MASS` package in `R v3.4.3` (R Core Team, 2020; Venables & Ripley, 2002). This approach was used to provide starting values of q to initialize MCMC, and does not limit posterior sampling. We executed `entropy` at $k = 2$ for 70,000 steps following 30,000 steps that were discarded as burnin, and sampled every 10th step. The matrix of genotype probabilities generated by `entropy` is available at Dryad (Jahner et al., 2021).

As a model-based view of the distribution of ancestry in hybrids, we also used the ancestry complement model of `entropy` to estimate interspecific ancestry coefficients (Q_{12}), or the proportion of loci with interclass ancestry from both parental species (Gompert

et al., 2014). This parameter allows for more refined categorization of ancestry classes for hybrids and is particularly valuable for delineating between F_1 and F_2 individuals (Gompert & Buerkle, 2016; Lindtke et al., 2014; Shastry et al., 2021), which are both expected to have $q = 0.5$, as well as identifying the offspring from other advanced generation hybrid crosses. Pure individuals have $Q_{12} = 0$ because their parents contributed the same ancestry to their offspring's genome, whereas F_1 hybrids have $Q_{12} = 1$ because their parents each contributed different ancestries. Further, F_2 hybrids have on average $Q_{12} = 0.5$ because recombination results in a genome consisting of roughly equal parts same-source and intersource ancestry, and previously published simulations have demonstrated that backcrossed individuals reside on lines drawn between F_1 hybrids ($Q_{12} = 1.0$; $q = 0.5$) and parental individuals ($Q_{12} = 0$; $q = 0$ or 1) (Lindtke et al., 2014; Shastry et al., 2021). The distinction between F_1 and F_2 hybrids is particularly relevant in the context of postzygotic incompatibilities in F_2 hybrids because they have recombinant genomes (i.e., chromosomes contain ancestry from both parental species), which could impose lower fitness relative to F_1 or parental individuals (Barton, 2001; Dobzhansky, 1970; Maheshwari & Barbash, 2011). It is worth noting that because meiosis generally produces an equal number of recombinant and nonrecombinant chromatids, roughly half of an F_2 hybrid's chromatids will be recombinant, though this varies probabilistically due to random assortment.

We ran five replicate MCMC chains of `entropy` for 40,000 steps, discarded the first 10,000 steps, and retained every 10th step. We utilized the comparison of q and Q_{12} to characterize the distribution of ancestry at the contact zone and assign individuals to hybrid classes (Gompert & Buerkle, 2016; Lindtke et al., 2014; Mandeville et al., 2017). Because the genomic clines analyses described below were only performed using a subset of ancestry informative markers (AIMs), defined as those loci with allele frequency differences >0.30 between pure parental individuals (*N. lepida*: $q < 0.05$; *N. bryanti*: $q > 0.95$; $N = 2565$), we also ran `entropy` on the AIM subset specifying the same parameters as above. Allele frequencies were calculated as the mean genotype probability divided by two (genotype probabilities range from zero to two [first and second homozygous state, respectively], with heterozygotes having probabilities of one).

2.4 | Genome-wide patterns of differentiation, divergence, and introgression

To characterize patterns of differentiation across the genome, we calculated locus-specific Hudson's F_{ST} (Hudson et al., 1992) based on allele frequencies from genotype probabilities (as described above) for parental individuals with $q < 0.05$ (*N. lepida*) or $q > 0.95$ (*N. bryanti*). Following Tavares et al. (2018), we also decomposed F_{ST} into its two components of within-population diversity (π_w) and between-population diversity (π_b) (Charlesworth, 1998).

$$F_{ST} = 1 - \overline{\pi_w}/\pi_b \quad (1)$$

$$\pi_w = 2p_i\left(1 - p_i\right) \quad (2)$$

$$\pi_b = d_{XY} = p_i\left(1 - p_j\right) + p_j\left(1 - p_i\right) \quad (3)$$

We follow the terminology of Ravinet et al. (2017) by using "genetic differentiation" when referring to F_{ST} (a relative measure) and "genetic divergence" when referring to π_b (an absolute measure that is equivalent to d_{XY}; Hahn, 2018). Relative measures can be extreme due to either high divergence between or reduced diversity within populations (Baird, 2017; Charlesworth, 1998; Charlesworth et al., 1997; Cruickshank & Hahn, 2014), thereby limiting the ability to delineate between gene flow and potential selective sweeps as underlying causes (but see Tavares et al., 2018). In contrast, absolute measures of divergence are not influenced by levels of within-population diversity and can be more reflective of between-population differentiation (Charlesworth, 1998; Cruickshank & Hahn, 2014; Hahn, 2018; Nei, 1973, 1987). It is worth noting that π_b is typically calculated for longer sequences, as locus-specific estimates from short read data (such as reduced-representation approaches) can have little precision (Cruickshank & Hahn, 2014). Despite this, we calculated locus-specific π_b because (i) the relatively low marker density of our GBS data set resulted in relatively few loci per scaffold, and (ii) we were interested exploring the potential relationship between locus-specific estimates of divergence and introgression (see below). Manhattan plots were generated using the `qqman` package (Turner, 2014) in `R` to visualize the distribution of pairwise F_{ST}, π_b, and π_w across the 25 largest scaffolds of the reference assembly ($2n = 52$ chromosomes; Baker & Mascarello, 1969). To determine which loci resided in or near coding regions, we aligned the 200 bp sequence flanking each SNP in the reference genome to the *N. lepida* gene reference (Campbell et al., 2016) using `bwa` as described above.

To quantify the strength of linkage disequilibrium across the genome, the squared correlation coefficient (r^2; Hill & Robertson, 1968) among all loci located on the same genomic scaffolds was calculated using the --geno-r2 option in `vcftools` (Danecek et al., 2011). The expected decay in r^2 among adjacent sites was estimated using the equations provided by Remington et al. (2001), which were derived from earlier work by Sved (1971) and Hill and Weir (1988):

$$E\left(r^2\right) = \left[\frac{(10 + C)}{(2 + C)(11 + C)}\right]\left[1 + \frac{(3 + C)\left(12 + 12C + C^2\right)}{n(2 + C)(11 + C)}\right] \quad (4)$$

$$C = 4Nc, \quad (5)$$

where n is twice the number of individuals in the sample, N is the effective population size, and c is the recombination fraction between sites. C was estimated from the data via nonlinear regression (Bates & Watts, 1988) using the *nls* function in `R`, specifying 0.1 as the start value. The physical distance at which r^2 decayed below 0.20 was used as a proxy for the strength of linkage disequilibrium (Hahn, 2018), following other recent authors (e.g., Acosta et al., 2019; Leforestier et al., 2015). This analysis was performed on (i) all individuals ($N = 353$); (ii) a subset of

parental individuals with $q < 0.05$ or $q > 0.95$ ($N = 277$); and (iii) a subset of hybrid individuals with $0.18 < q < 0.82$ ($N = 55$) (see explanation of hybrid ancestry cutoffs below).

Bayesian genomic clines (Gompert & Buerkle, 2011) were employed to quantify variation in locus-specific introgression. At each locus, two parameters (cline centre [α] and cline rate [β]) were estimated based on the relationship between the proportion of parental ancestry (ϕ) and the hybrid index (h), both of which range between 0 (pure *N. bryanti*) and 1 (pure *N. lepida*) and are predicted to be equivalent under a null expectation (Gompert & Buerkle, 2011; Gompert, Parchman, et al., 2012). The α parameter measures deviations in ϕ for a locus relative to h (if $\phi = h$, $\alpha = 0$). Positive values of α reflect introgression of *N. lepida* ancestry into the *N. bryanti* genomic background ($\phi > h$), whereas negative values of α are consistent with introgression of *N. bryanti* ancestry into the *N. lepida* genomic background ($\phi < h$). In contrast, the β parameter measures the rate of transition between *N. bryanti* and *N. lepida* ancestry across the genomic cline. Positive values of β can result from strong selection against hybrids and could be associated with potential barrier loci (e.g., BDM incompatibilities; Gompert & Buerkle, 2012; Gompert, Parchman, et al., 2012; but see Lindtke & Buerkle, 2015), though extreme β estimates can also arise during some scenarios of neutral introgression (Gompert & Buerkle, 2011; McFarlane et al., 2021). Conversely, negative values of β are consistent with broad clines and could be indicators of alleles that are potentially favoured in both genomic backgrounds (Sung et al., 2018). In addition to characterizing variation in the magnitude, direction, and rate of introgression, we were interested in asking to what extent historical differentiation and divergence between the species predicts introgression, because any relationship could suggest that divergent selection during allopatry may also be associated with reproductive isolation in the contact zone (Baiz et al., 2019; Gompert, Lucas, et al., 2012; Schield et al., 2017).

Given the computational effort required for chains to successfully mix and converge for this analysis, we only considered AIMs (described above; $N = 2565$ loci) and included the 55 hybrid individuals with the strongest evidence of mixed ancestry ($0.18 < q < 0.82$). These ancestry cutoffs for hybrids were selected based on the identification of relatively natural breaks in the distribution of q across all individuals. Additionally, 40 parental individuals were randomly selected from both species (*N. lepida*: $q < 0.05$; *N. bryanti*: $q > 0.95$) for the pure reference populations, as this was sufficient for precise estimates of parental allele frequencies. Five replicate chains of the genotype uncertainty model of `bgc v1.04b` (Gompert & Buerkle, 2012) were implemented for 1,000,000 MCMC iterations, specifying a burnin of 200,000 and thinning every 100 steps. In order to explore the potential for the use of AIMs to bias the results from `bgc` (McFarlane et al., 2021), we ran additional analyses based on 10 random subsets of the same number of loci (2565) used in the AIM analysis that were sampled from the entire distribution of allele frequency differences (see Supporting Information Methods). Chain convergence and mixing were assessed via correlations of parameters among chains and the calculation of effective sample sizes in the `coda` package (Plummer et al., 2006) in `R`. Across loci, the relationships among π_b, F_{ST}, and absolute genomic cline parameters (i.e., $|\alpha|$ and $|\beta|$) were assessed using linear models in `R`.

3 | RESULTS

3.1 | Assembly and variant calling

After removing individuals with insufficient sequencing data and filtering potential contaminants from the raw data, we retained 634,404,152 reads for analysis. A total of 406,609,782 reads mapped onto the *N. lepida* draft genome assembly. After thorough bioinformatic and quality filtering, we retained a data set of 353 individuals and 26,262 SNPs with a mean coverage per individual per locus of 3.31x. The full data set of 26,262 loci were used for all analyses except those based on Bayesian genomic clines and a confirmatory `entropy` analysis, which instead used the subset of 2565 AIMs described above.

3.2 | The distribution of ancestry across a sharp ecotone

Genetic differentiation and divergence were pronounced between pure individuals of the parental species (mean $\pi_b = 0.214$; mean $F_{ST} = 0.178$; based on the full data set of 26,262 SNPs), consistent with substantive divergence in allopatry (Patton et al., 2007; Shurtliff et al., 2014). Ancestry coefficients (q) completely assigned pure *N. lepida* and *N. bryanti* to separate ancestries, and putative hybrids had a range of intermediate q estimates (Figure 1b). Estimates of individual ancestry coefficients (q) from `entropy` were broadly consistent with those from a previously published microsatellite data set ($r = .995$; Figure S1; Shurtliff et al., 2014). However, 95% credible intervals were dramatically smaller for the GBS data set, especially for hybrids (Figure S2).

The distribution of ancestry coefficients (q) and interspecific ancestry estimates (Q_{12}) from `entropy` were highly concordant among independent MCMC chains (Figure S3) and were consistent with a broad distribution of multigenerational backcrosses ($0.05 < q < 0.95$; Figure 1c). All but 16 individuals were assigned to the same ancestry category (i.e., *N. lepida*; BC-*lepida*; F_1 hybrid; BC-*bryanti*; *N. bryanti*) based on the ancestry coefficients from both data sets (SSRs and GBS; Figure S1). All hybrids were designated as F_1 ($N = 16$; $q = 0.4–0.6$), BC-*lepida* ($N = 28$; $q = 0.1–0.4$), or BC-*bryanti* ($N = 20$; $q = 0.6–0.9$). No individuals had ancestry estimates that are expected for F_2 and later generation F_N hybrids ($q \approx 0.5$; $Q_{12} \approx 0.5$), or for other hybrid ancestries (e.g., BC × BC), even though simulations found sufficient power to detect hybrid × hybrid crosses with sample sizes >200 (Supporting Information Results; Figure S4). The results obtained from the full `entropy` model including all 26,262 loci were qualitatively identical to those generated from an additional confirmatory model based on the subset of 2565 ancestry informative

markers (mean AIMs π_b = 0.571; mean AIMs F_{ST} = 0.545; Figure S5)—individual parameter estimates were highly correlated for both q (r = .9998) and Q_{12} (r = .9971) (Figure S6). Further, estimates were consistent even when the initialized ancestry coefficients based on LDA were randomly assigned across individuals (Figure S7).

3.3 | | Genome-wide patterns of differentiation, divergence, and introgression

The genome-wide distributions of differentiation and divergence between *N. lepida* and *N. bryanti* were highly correlated (r = .888) and both contained a unimodal peak (mean π_b = 0.214; median π_b = 0.166; mean F_{ST} = 0.178; median F_{ST} = 0.147) and a long tail of loci with more extreme estimates (Figure 2). Further, both distributions were similar for the subset of 4636 loci (17.7%) that mapped to coding regions (mean π_b = 0.220; median π_b = 0.166; mean F_{ST} = 0.184; median F_{ST} = 0.148; Figure S8). Of the 31 loci residing in regions coding for cytochrome P450 detoxification enzymes (mean π_b = 0.255; mean F_{ST} = 0.177; Table S1), four were above the 95th quantile of the π_b distribution (π_b > 0.499) (Cyp4b1, scaffold 2527, position 76723, π_b = 0.975, F_{ST} = 0.975; Cyp3a11, scaffold 29245, position 1971, π_b = 0.679, F_{ST} = 0.639; Cyp2b1, scaffold 1904, position 35291, π_b = 0.501, F_{ST} = 0.006; Cyp2C25, scaffold 8345, position 44827, π_b = 0.500, F_{ST} = 0.498) and three were below the fifth quantile (π_b < 0.119) (Cyp3a13, scaffold 7071, position 35662, π_b = 0.100, F_{ST} = 0.098; Cyp3a11, scaffold 7303, position 40349, π_b = 0.106, F_{ST} = 0.040; Cyp2C25, scaffold 3799, position 72216, π_b = 0.110, F_{ST} = 0.107) (Tables S1 and S2). Based on permutations, the number of cytochrome p450 loci in either tail did not differ from null expectations (upper 95th quantile: p = .100; lower fifth quantile: p = .215; Figure S9). Within-population genetic diversity was markedly higher for *N. lepida* (mean π_w = 0.198; median π_w = 0.230) than for *N. bryanti* (mean π_w = 0.114; median π_w = 0.024), a pattern driven by a relatively large number of loci with π_w ranging from 0.20 to 0.35 for *N. lepida* (Figure 2). For most scaffolds, estimates of differentiation, divergence, and diversity were not visibly clustered, but instead exhibited patterns suggesting a lack of genomic autocorrelation (Figure S10). This result was consistent with a moderately rapid erosion of LD across scaffolds, with expected r^2 decaying below 0.20 at 1466 bp (estimate based on all individuals; Figure S11). As expected, r^2 decayed below 0.20 more slowly for the subset of parental individuals (2157 bp) and more quickly for the subset of hybrids (1373 bp). Importantly, these estimates of LD decay occur at a scale that is two orders of magnitude smaller than that covered by our marker density (1 SNP per ~90–110 kbp).

Estimates of genomic cline centre (α) and rate (β) parameters were concordant across replicate MCMC chains (Figure S12), and all effective sample sizes were >200. The distribution of α was highly heterogeneous (Figure 3a), with 60.7% of the ancestry informative

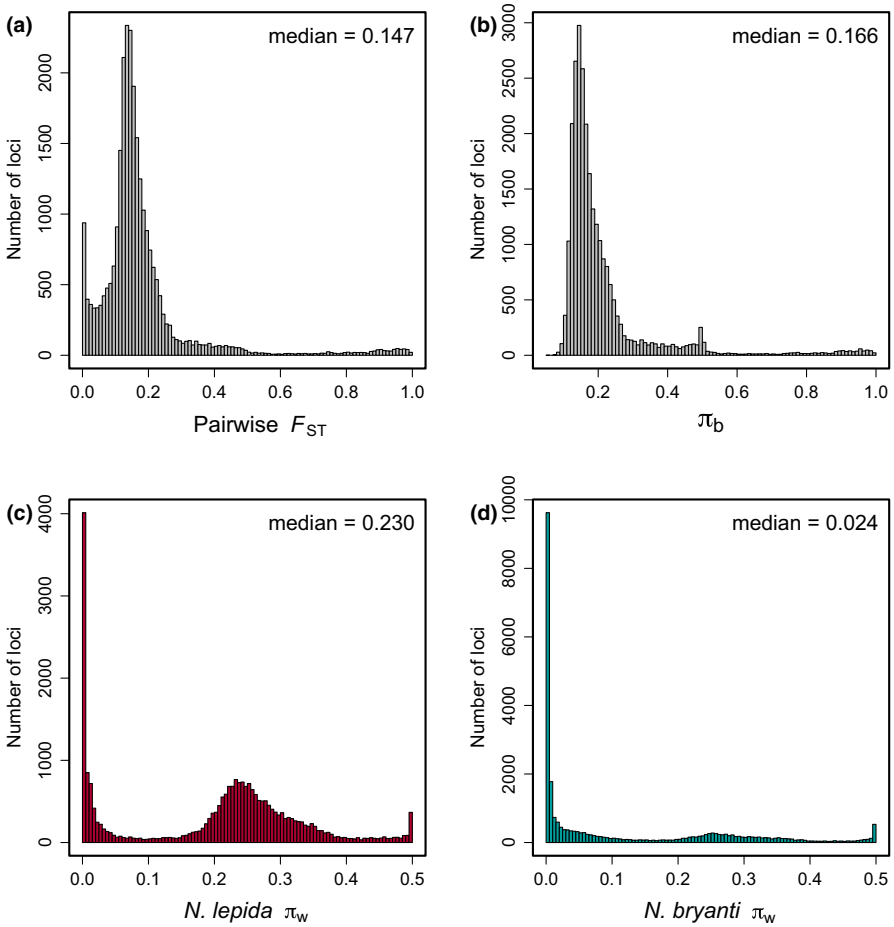

FIGURE 2 Locus-specific estimates of differentiation (F_{ST}; panel a) and divergence (π_b; panel b) between *N. lepida and N. bryanti* were highly heterogeneous across the genome. Additionally, within-population diversity (π_w) estimates were higher for *N. lepida* (panel c) than for *N. bryanti* (panel d). All estimates were calculated based on allele frequencies from parental *N. bryanti* (q > 0.95) and *N. lepida* (q < 0.05) individuals [Colour figure can be viewed at wileyonlinelibrary.com]

markers (mean AIMs $\pi_b = 0.571$; $F_{ST} = 0.545$) having 95% credible intervals that did not overlap zero. Estimates of differentiation and divergence did not differ for the subset of α outliers ($N = 1558$; $\pi_b = 0.570$; $F_{ST} = 0.543$) relative to the full set of AIMs ($N = 2565$; $\pi_b = 0.571$; $F_{ST} = 0.545$). There was no difference in the number of loci with either positive or negative α estimates ($N = 803$ vs. 755; $\chi^2 = 1.479$; df = 1; $p = .224$), but the magnitude of estimates were larger for negative α outliers (mean = −1.817; SD = 0.575) than for positive α outliers (mean = 1.109; 0.131) (Figure 3b). Additionally, there was no relationship between π_b and $|\alpha|$ ($R^2 < .001$; $p = .205$; Figure 3C; Table S3). No loci had 95% credible intervals for β that did not include zero, and $|\beta|$ was not predicted by pairwise π_b ($R^2 < .001$; $p = .418$; Table S3). Finally, of the four loci residing in regions associated with cytochrome p450 detoxification enzymes that were included in the genomic cline analysis, a single locus had an α estimate with credible intervals that did not overlap zero (Cyp3a11, scaffold 5775, position 58030, $\pi_b = 0.386$, $F_{ST} = 0.324$, $\alpha = -2.304$; 95% CI = −3.878, −1.083; Table S1). Our `bgc` results were consistent across different subsets of individuals and loci (Supporting Information Results; Figures S13–S23; Table S3).

4 | DISCUSSION

4.1 | Distribution of ancestry across the contact zone

Estimates of Q_{12} provided new insight into patterns of mating among parental and hybrid ancestry classes that survive to emergent juvenile or adult stages in this system. All individuals were restricted to the outer diagonals of the triangle in Figure 1c, indicating that all had at least one parent with pure *N. lepida* or *N. bryanti* ancestry (Gompert et al., 2014; Lindtke et al., 2014; Shastry et al., 2021). That is, hybrids at this site appear to only be the product of matings that are: $P_1 \times P_2$ (F_1 offspring), $F_1 \times P_1$ or P_2 (BC_1 offspring), and $BC_N \times P_1$ or P_2 (BC_{N+1} offspring). The apparent absence of F_2 hybrids may not be surprising given the rarity of F_1 hybrids in this bimodal hybrid zone (2% of individuals; Shurtliff et al., 2014), as only 1 in 2500 offspring would be expected to be F_2 under random mating. However, there was also no evidence for offspring resulting from any type of hybrid × hybrid cross, despite the null expectation that 1.69% of mating events should fall into this category under random mating based on the estimate that 13% of individuals have admixed ancestry (latter estimate from Shurtliff et al. [2014], who sampled individuals from this contact zone at random).

Many bimodal hybrid zones are regulated by prezygotic barriers to gene flow that promote strong assortative mating (Billerman et al., 2019; Jiggins & Mallet, 2000; Schumer et al., 2017). In this system, the strong spatial segregation of the parental species and their respective backcrosses, albeit spanning distances well within the dispersal capacity of woodrats, probably contributes to the rarity of F_1 hybrids and could be involved in limiting the occurrence of subsequent hybrid × hybrid crosses. Indeed, <5% of large inter-annual movements (>20 m) involve transitions across the ecotone that divides the sampling area (Figure 1; Shurtliff et al., 2013), so individuals are less likely to encounter potential heterospecific mates. Further, even when pure parentals meet, laboratory mate choice trials show that *N. lepida* females display strong aversion to mating with *N. bryanti* males, whereas *N. bryanti* females will mate with males from either species (Shurtliff et al., 2013). This suggests that F_1 hybrids predominantly result from *N. bryanti* female × *N. lepida* male crosses, further limiting the potential for heterospecific matings.

While it is possible that the prezygotic factors listed above combined with the overall rarity of hybrids could fully explain the absence of hybrid × hybrid matings, postzygotic barriers could also contribute to this pattern. Specifically, the apparent absence of individuals that would have been produced via hybrid × hybrid crosses could be consistent with some form of intrinsic genomic incompatibility (i.e., hybrid breakdown; Barton, 2001; Dobzhansky, 1970; Stelkens et al., 2015). If parental ancestry blocks are vital for survival, the genetic variation found in hybrids with recombinant genomes could have negative fitness consequences. Such intrinsic incompatibilities could act in addition to the extrinsic postzygotic mechanisms proposed for the low survival of juvenile hybrids (Shurtliff et al., 2014). Together these results suggest that hybrids may require at least one pure parent to have the genomic stability to reach the age of nest emergence, but that these individuals may still face a survival disadvantage as they begin interacting with the environment during their first year of life.

4.2 | Heterogeneous patterns of differentiation, divergence, and introgression

The distributions of differentiation (F_{ST}, a relative measure) and divergence (π_b, an absolute measure) were strongly correlated ($r = .888$; Figure 2; Table S3), indicating they similarly illustrate genome-wide variation in this case. Both contained long tails of extreme values (Figure 2), including some loci found in coding regions (Table S2), which could represent genomic regions involved in adaptation (but see Baird, 2017). Due to the fragmented state of the *N. lepida* draft genome we utilized (Campbell et al., 2016), as well as the potential pitfalls associated with detecting loci associated with adaptation using genome scans of reduced-representation data (Cruickshank & Hahn, 2014; Hoban et al., 2016; Lowry et al., 2017; Ravinet et al., 2017), we view these distributions only as a cursory characterization of genome-wide heterogeneity. That said, we were interested in exploring candidates associated with cytochrome P450s, as these detoxification enzymes have been implicated in driving dietary specialization in a wide range of organisms, including woodrats (Kitanovic et al., 2018; Magnanou et al., 2009; Malenke et al., 2012), greater sage-grouse (Oh et al., 2019), koalas (Johnson et al., 2018), and swallowtail butterflies (Berenbaum & Feeny, 2008). Our analyses provide little evidence for divergent selection influencing genetic variation at cytochrome P450 loci. The mean estimates of π_b and F_{ST} for cytochrome P450 associated loci were similar to

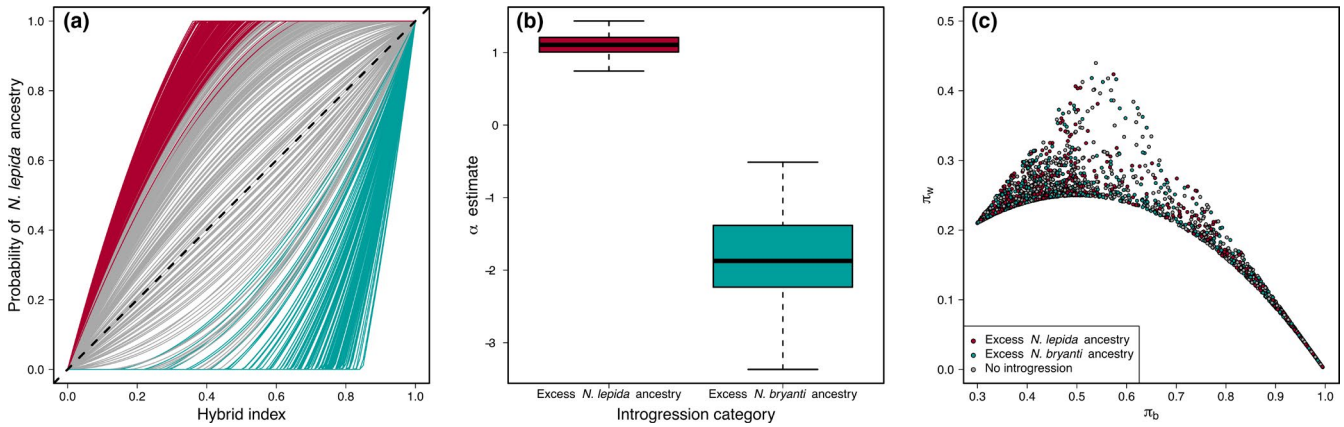

FIGURE 3 (a) Bayesian genomic clines for 1000 representative ancestry informative markers (AIMs) are drawn based on cline centres (α) and rates (β) estimated with `bgc` (Gompert & Buerkle, 2012). The dashed diagonal black line gives the predicted shape of clines when the probability of *N. lepida* ancestry (ϕ) is equal to hybrid index (h; *N. bryanti* = 0; *N. lepida* = 1). Maroon genomic clines represent loci with excess *N. lepida* ancestry (lower 95% credible interval of $\alpha > 0$), teal genomic clines represent loci with excess *N. bryanti* ancestry (upper 95% credible interval of $\alpha < 0$), and grey genomic clines represent loci where α is not different from zero. Zero loci had β estimates that were different from zero. (b) Boxplots display the distribution of α parameter estimates for loci with excess *N. lepida* ancestry ($N = 803$) and excess *N. bryanti* ancestry ($N = 755$). (c) Across all ancestry informative markers (AIMs), the magnitude of introgression (α) is not associated with between-population or within-population divergence (π_b and π_w) [Colour figure can be viewed at wileyonlinelibrary.com]

the genome-wide averages, and only four of the 31 loci occurring in or near p450 genes were in the upper 5% of the genome-wide π_b distribution (Table S2; Figure S9). Despite the paucity of cytochrome P450 candidates, the rapid decay of linkage disequilibrium (Figure S11) coupled with the relatively low marker density of our data places substantial limits on our ability to describe genomic differentiation and its relationship to traits of ecological significance. Ultimately, more genomically thorough sequencing will be needed for analyses of the genetic basis of dietary adaptation in this system.

Given the absence of hybrid × hybrid crosses, we expected to find loci with steep genomic clines (i.e., positive β), which might be associated with reproductive isolation (Gompert & Buerkle, 2012; Gompert, Parchman, et al., 2012). However, we did not detect any loci with β estimates not overlapping zero. The absence of β outliers has been previously interpreted as an absence of selection against hybrids (e.g., Akopyan et al., 2020; Menon et al., 2018; Oswald et al., 2019). In contrast, ~60% of loci had genomic cline centre (α) credible intervals that did not overlap zero, confirming that the history of multigenerational backcrossing (Figure 1c) has enabled substantial introgression. The fraction of loci with extreme α estimates we report is larger than in most other studies using genomic clines (e.g., Kingston et al., 2017; Nosil et al., 2012; Parchman et al., 2013; Sung et al., 2018; Taylor et al., 2014; but see Janoušek et al., 2015; Souissi et al., 2018), although this result is likely influenced by restricting our analyses to AIMs. In another recent study, McFarlane et al. (2021) recovered a much higher proportion of loci with excess α estimates when considering only the AIM subset of SNPs (70% excess α for AIMs vs. 11% excess α for all loci). When we performed genomic cline analysis using random subsets of loci sampled from the full range of allele frequency differences (instead of considering only AIMs; see Supporting Information), only 46% of loci had extreme α estimates on average. Additionally, while we nearly exhaustively sampled hybrids from this bimodal hybrid zone, the general rarity of hybrids resulted in the inclusion of relatively few admixed individuals in this analysis ($N = 55$), which may have contributed to the variability of genomic cline parameters (Gompert & Buerkle, 2011). Further, the ability to distinguish between backcrosses and pure parental individuals rapidly erodes after even a few generations (Anderson & Thompson, 2002; Boecklen & Howard, 1997), so we cannot rule out that this hybrid zone was formed relatively recently (i.e., within many tens of generations) despite the apparent presence of multigenerational backcrosses (Figure 1c). As such, the large number of loci with extreme α estimates could be influenced by the stochasticity of drift and recombination (Baird et al., 2003; Gompert & Buerkle, 2011; McFarlane et al., 2021).

Extreme α outliers should have higher detectability when alleles are moving into less diverse genomic backgrounds (McFarlane et al., 2021), which could bias our results towards detecting stronger introgression of *N. lepida* alleles (*N. lepida*: mean $\pi_w = 0.198$; median $\pi_w = 0.230$; *N. bryanti*: mean $\pi_w = 0.114$; median $\pi_w = 0.024$; Figure 2). Counter to this expectation, the magnitude of negative α outliers was more pronounced (Figure 3b), consistent with stronger introgression of *N. bryanti* ancestry into the *N. lepida* genetic background than vice versa. The one locus tagging a cytochrome p450 (Cyp3a11; Table S1) that appears to have introgressed from *N. bryanti* to *N. lepida* is intriguing, but requires further verification. Cyp3a has particularly broad substrate specificity and is abundantly expressed in the liver and intestines of mammals, where it acts as a barrier against xenobiotics (e.g., plant secondary compounds) from entering the circulation system, while also regulating the expression of other liver detoxification pathways (van Waterschoot et al., 2009). The functional significance and potential selective advantage that might be conferred from the introgression of Cyp3A alleles requires further investigation.

While there was plenty of variability in measures of divergence, differentiation, and genomic clines (and thus power for analysis of their comparison), there was no evidence of an association between divergence (or differentiation) and introgression in this study. Genetic regions with extreme genomic cline parameters did not strongly coincide with those having extreme differentiation or divergence (Figure 3; Table S3), nor did they have estimates of differentiation and divergence that differed from the full set of ancestry informative markers (α outliers: $\pi_b = 0.570$; $F_{ST} = 0.543$; all AIMs: $\pi_b = 0.571$; $F_{ST} = 0.545$). Multiple previous studies using similar data and analyses have reported some level of coincidence between introgression and differentiation (e.g., Baiz et al., 2019; Ebersbach et al., 2020; Gompert, Lucas, et al., 2012; Kingston et al., 2017; Menon et al., 2018; Nosil et al., 2012; Oswald et al., 2019; Parchman et al., 2013; Schield et al., 2017; Taylor et al., 2014; but see Zieliński et al., 2019) and interpreted this as evidence that divergence in allopatry can predict reproductive isolation. This is despite the fact that most of these studies used the relative divergence metric F_{ST} (but see Schield et al., 2017), which could obscure such a relationship. Nonetheless, there are multiple explanations for why divergence and introgression might be unrelated in this *Neotoma* hybrid zone. First, our consideration of only AIMs for Bayesian genomic clines might limit the ability to detect relationships between introgression and divergence metrics, as previous studies recovered reduced correlations between F_{ST} and α when employing higher minimum F_{ST} thresholds (Gompert, Lucas, et al., 2012; Nosil et al., 2012). Additional genomic clines analyses based on random sampling of loci across the entire range of allele frequency differences did recover evidence that π_b and F_{ST} were associated with α in one of ten subsets each (Supporting Information Results; Table S3). However, these predictors explained a small proportion of the variance in genomic cline parameters even when associations could be detected ($R^2 < .003$; Table S3), suggesting that any putative relationships between differentiation, divergence, and introgression in this study are subtle. Second, simulations of hybridization have demonstrated that genetic drift may underlie variation in extreme genomic cline estimates under some demographic scenarios. For example, using simulations of neutral introgression Gompert and Buerkle (2011) found higher variability in cline estimates—particularly for α—when sample sizes were small (i.e., $N = 100$). Using similar simulations of neutral introgression, McFarlane et al. (2021) recovered high variability in the proportion of β outliers when hybridization was rare or relatively recent. As similar conditions could characterize the *Neotoma* hybrid zone, the variability in genomic clines we report could be strongly influenced by stochastic processes, hence precluding the type of association between differentiation or divergence and introgression reported in previous studies.

A final caveat worth considering is that our results are likely influenced by the fact that our sampling of parental ancestry is restricted to the area of active hybridization (Figure 1d) wherein "parentals" may themselves be partly introgressed. Indeed, a recent study of *Mytilus* hybridization reported higher proportions of loci with extreme genomic cline estimates when using spatially distant parental populations in comparison to parental populations near sites of active hybridization (Simon et al., 2021). Thus, future studies should include additional parental populations that are less likely to have a recent history of hybridization to allow insight into patterns of differentiation and introgression beyond the immediate area of secondary contact. Likewise, as with most studies, our sampling of admixed genomes in the area of contact is limited to a narrow window of time that may not characterize longer-term patterns of admixture between these species, even at this site. The demography and, thus, opportunity for interspecific mating in woodrats has been shown to be weather-driven and likely fluctuates through time (e.g., Hunter et al., 2017). Ultimately, clarity into the causes and consequences of hybridization in this system will require augmented temporal, spatial, and genomic sampling, including comparative analyses of replicate contact zones.

ACKNOWLEDGEMENTS

We thank F. Bonhomme, A. Buerkle, T. Faske, J. Fordyce, P.-A. Gagnaire, Z. Gompert, C. Nice, D. Nielsen, and P. Nosil for helpful discussions. We thank Associate Editor J. Pemberton, S. Baird, and three anonymous reviewers for thoughtful comments that improved the manuscript. We thank Q. Shurtliff for the initial fieldwork and sample preparation associated with this project. This research was funded by National Science Foundation grants to MDM (IOS-1457209 and OIA-1826801).

AUTHOR CONTRIBUTIONS

Joshua P. Jahner led the analysis of the data and wrote the initial draft of the manuscript. Thomas L. Parchman constructed the genotyping-by-sequencing libraries and contributed to the analysis of the data. Marjorie D. Matocq developed the study system, secured funding for this research, and collected samples. All authors contributed to manuscript preparation.

DATA AVAILABILITY STATEMENT

Individual fastq files and a matrix of genotype probabilities have been made available at the Dryad Digital Repository (https://doi.org/10.5061/dryad.bnzs7h4bd).

ORCID

Joshua P. Jahner https://orcid.org/0000-0001-8121-6783
Thomas L. Parchman https://orcid.org/0000-0003-1771-1514
Marjorie D. Matocq https://orcid.org/0000-0003-4829-3749

REFERENCES

Acosta, J. J., Fahrenkrog, A. M., Neves, L. G., Resende, M. F. R., Dervinis, C., Davis, J. M., & Kirst, M. (2019). Exome resequencing reveals evolutionary history, genomic diversity, and targets of selection in the conifers *Pinus taeda* and *Pinus elliottii*. *Genome Biology and Evolution*, *11*(2), 508–520.

Akopyan, M., Gompert, Z., Klonoski, K., Vega, A., Kaiser, K., Mackelprang, R., Rosenblum, E. B., & Robertson, J. M. (2020). Genetic and phenotypic evidence of a contact zone between divergent colour morphs

of the iconic red-eyed treefrog. *Molecular Ecology*, *29*(22), 4442–4456. https://doi.org/10.1111/mec.15639

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*(3), 1217–1229. https://doi.org/10.1093/genetics/160.3.1217

Baird, S. J. E. (2017). The impact of high-throughput sequencing technology on speciation research: Maintaining perspective. *Journal of Evolutionary Biology*, *30*(8), 1482–1487. https://doi.org/10.1111/jeb.13099

Baird, S. J. E., Barton, N. H., & Etheridge, A. M. (2003). The distribution of surviving blocks of an ancestral genome. *Theoretical Population Biology*, *64*(4), 451–471. https://doi.org/10.1016/S0040-5809(03)00098-4

Baiz, M. D., Tucker, P. K., & Cortés-Ortiz, L. C. (2019). Multiple forms of selection shape reproductive isolation in a primate hybrid zone. *Molecular Ecology*, *28*(5), 1056–1069. https://doi.org/10.1111/mec.14966

Baker, R. J., & Mascarello, J. T. (1969). Karyotypic analyses of the genus *Neotoma* (Cricetidae, Rodentia). *Cytogenetics*, *8*(3), 187–198.

Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution*, *68*(9), 2644–2657. https://doi.org/10.1111/evo.12457

Barton, N. H. (2001). The role of hybridization in evolution. *Molecular Ecology*, *10*(3), 551–568. https://doi.org/10.1046/j.1365-294x.2001.01216.x

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. Wiley and Sons.

Berenbaum, M. R., & Feeny, P. P. (2008). Chemical mediation of host-plant specialization: The papilionid paradigm. In K. J. Tilmon (Ed.), *Specialization, speciation, and radiation: The evolutionary biology of herbivorous insects* (pp. 3–19). University of California Press.

Billerman, S. M., Cicero, C., Bowie, R. C. K., & Carling, M. D. (2019). Phenotypic and genetic introgression across a moving hybrid zone. *Molecular Ecology*, *28*(7), 1692–1708.

Boecklen, W. J., & Howard, D. J. (1997). Genetic analysis of hybrid zones: Numbers of markers and power of resolution. *Ecology*, *78*(8), 2611–2616. https://doi.org/10.1890/0012-9658(1997)078[2611:GAOHZN]2.0.CO;2

Buerkle, C. A., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology*, *22*(11), 3028–3035. https://doi.org/10.1111/mec.12105

Burke, J. M., & Arnold, M. L. (2001). Genetics and the fitness of hybrids. *Annual Review of Genetics*, *35*, 31–52. https://doi.org/10.1146/annurev.genet.35.102401.085719

Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, *1*(3), 118–131. https://doi.org/10.1002/evl3.14

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, *25*(11), 1656–1665.

Burton, R. S., Pereira, R. J., & Barreto, F. S. (2013). Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 281–302. https://doi.org/10.1146/annurev-ecolsys-110512-135758

Campbell, M., Oakeson, K. F., Yandell, M., Halpert, J. M., & Dearing, D. (2016). The draft genome sequence and annotation of the desert woodrat *Neotoma lepida*. *Genomics Data*, *9*, 58–59. https://doi.org/10.1016/j.gdata.2016.06.008

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, *17*(3), 362–365. https://doi.org/10.1111/1755-0998.12669

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*(5), 538–543. https://doi.org/10.1093/oxfordjournals.molbev.a025953

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research*, *70*(2), 155–174.

Christe, C., Stölting, K. N., Bresadola, L., Fussi, B., Heinze, B., Wegmann, D., & Lexer, C. (2016). Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. *Molecular Ecology*, *25*(11), 2482–2498.

Christe, C., Stölting, K. N., Paris, M., Fraïsse, C., Bierne, N., & Lexer, C. (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology*, *26*(1), 59–76.

Clausen, J. (1951). *Stages in the evolution of plant species*. Cornell University Press.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer.

Coyner, B. S., Murphy, P. J., & Matocq, M. D. (2015). Hybridization and asymmetric introgression across a narrow zone of contact between *Neotoma fuscipes* and *N. macrotis* (Rodentia: Cricetidae). *Biological Journal of the Linnean Society*, *115*(1), 162–172.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R.; 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Dobzhansky, T. (1970). *Genetics of the evolutionary process*. Columbia University Press.

Dong, X., Fu, J., Yin, X., Cao, S., Li, X., Lin, L., & Ni, J. (2016). Emodin: A review of its pharmacology, toxicity, and pharmacokinetics. *Phytotherapy Research*, *30*(8), 1207–1218. https://doi.org/10.1002/ptr.5631

Ebersbach, J., Posso-Terranova, A., Bogdanowicz, S., Gómez-Díaz, M., García-González, M. X., Bolívar-García, W., & Andrés, J. (2020). Complex patterns of differentiation and gene flow underly the divergence of aposematic phenotypes in *Oophaga* poison frogs. *Molecular Ecology*, *29*(11), 1944–1956.

Endler, J. A. (1977). *Geographic variation, speciation, and clines*. Princeton University Press.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587. https://doi.org/10.1093/genetics/164.4.1567

Fitzpatrick, B. M. (2013). Alternative forms for genomic clines. *Ecology and Evolution*, *3*(7), 1951–1966. https://doi.org/10.1002/ece3.609

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*(3), 979–992. https://doi.org/10.1534/genetics.113.154740

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*(10), 2111–2127. https://doi.org/10.1111/j.1365-294X.2011.05074.x

Gompert, Z., & Buerkle, C. A. (2012). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, *12*(6), 1168–1176. https://doi.org/10.1111/1755-0998.12009.x

Gompert, Z., & Buerkle, C. A. (2016). What, if anything, are hybrids: Enduring truths and challenges associated with population structure and gene flow. *Evolutionary Applications*, *9*(7), 909–923. https://doi.org/10.1111/eva.12380

Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, *23*(18), 4555–4573. https://doi.org/10.1111/mec.12811

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, *66*(7), 2167–2181. https://doi.org/10.1111/j.1558-5646.2012.01587.x

Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, *48*, 207–229. https://doi.org/10.1146/annurev-ecolsys-110316-022652

Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012). Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 439–450. https://doi.org/10.1098/rstb.2011.0196

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., & Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722.

Haasl, R. J., & Payseur, B. A. (2016). Fifteen years of genomewide scans for selection: Trends, lessons and unaddressed genetic sources of complication. *Molecular Ecology*, *25*(1), 5–23. https://doi.org/10.1111/mec.13339

Hahn, M. W. (2018). *Molecular population genetics*. Oxford University Press.

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*(11), 2454–2466.

Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, *38*(6), 226–231. https://doi.org/10.1007/BF01245622

Hill, W. G., & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, *33*(1), 54–78. https://doi.org/10.1016/0040-5809(88)90004-4

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*(4), 379–397. https://doi.org/10.1086/688018

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*(2), 583–589. https://doi.org/10.1093/genetics/132.2.583

Hunter, E. A., Matocq, M. D., Murphy, P. J., & Shoemaker, K. T. (2017). Differential effects of climate on survival rates drive hybrid zone movement. *Current Biology*, *27*(24), 3898–3903. https://doi.org/10.1016/j.cub.2017.11.029

Jahner, J. P., Parchman, T. L., & Matocq, M. D. (2021). Data from: Multigenerational backcrossing and introgression between two woodrat species at an abrupt ecological transition. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.bnzs7h4bd

Janoušek, V., Munclinger, P., Wang, L., Teeter, K. C., & Tucker, P. K. (2015). Functional organization of the genome may shape the species boundary in the house mouse. *Molecular Biology and Evolution*, *32*(5), 1208–1220. https://doi.org/10.1093/molbev/msv011

Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*, *15*(6), 250–255. https://doi.org/10.1016/S0169-5347(00)01873-5

Johnson, R. N., O'Meally, D., Chen, Z., Etherington, G. J., Ho, S. Y. W., Nash, W. J., Grueber, C. E., Cheng, Y., Whittington, C. M., Dennison, S., Peel, E., Haerty, W., O'Neill, R. J., Colgan, D., Russell, T. L., Alquezar-Planas, D. E., Attenbrow, V., Bragg, J. G., Brandies, P. A., … Belov, K. (2018). Adaptation and conservation insights from the koala genome. *Nature Genetics*, *50*(8), 1102–1111. https://doi.org/10.1038/s41588-018-0153-5

Kingston, S. E., Parchman, T. L., Gompert, Z., Buerkle, C. A., & Braun, M. J. (2017). Heterogeneity and concordance in locus-specific differentiation and introgression between species of towhees. *Journal of Evolutionary Biology*, *30*(3), 474–485. https://doi.org/10.1111/jeb.13033

Kitanovic, S., Orr, T. J., Spalink, D., Cocke, G. B., Schramm, K., Wilderman, P. R., Halpert, J. R., & Dearing, M. D. (2018). Role of cytochrome P450 2B sequence variation and gene copy number in facilitating dietary specialization in mammalian herbivores. *Molecular Ecology*, *27*(3), 723–736. https://doi.org/10.1111/mec.14480

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), 371–375.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Leforestier, D., Ravon, E., Muranty, H., Cornille, A., Lemaire, C., Giraud, T., Durel, C.-E., & Branca, A. (2015). Genomic basis of the differences between cider and dessert apple varieties. *Evolutionary Applications*, *8*(7), 650–661. https://doi.org/10.1111/eva.12270

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lindtke, D., & Buerkle, C. A. (2015). The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution*, *69*(8), 1987–2004. https://doi.org/10.1111/evo.12725

Lindtke, D., Gompert, Z., Lexer, C., & Buerkle, C. A. (2014). Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Molecular Ecology*, *23*(17), 4316–4330.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152. https://doi.org/10.1111/1755-0998.12635

Macholán, M., Baird, S. J. E., Dufková, P., Munclinger, P., Bímová, B. V., & Piálek, J. (2011). Assessing multilocus introgression patterns: A case study on the mouse X chromosome in central Europe. *Evolution*, *65*(5), 1428–1446. https://doi.org/10.1111/j.1558-5646.2011.01228.x

Magnanou, E., Malenke, J. R., & Dearing, M. D. (2009). Expression of biotransformation genes in woodrat (*Neotoma*) herbivores on novel and ancestral diets: Identification of candidate genes responsible for dietary shifts. *Molecular Ecology*, *18*(11), 2401–2414.

Maheshwari, S., & Barbash, D. A. (2011). The genetics of hybrid incompatibilities. *Annual Review of Genetics*, *45*, 331–355. https://doi.org/10.1146/annurev-genet-110410-132514

Malenke, J. R., Magnanou, E., Thomas, K., & Dearing, M. D. (2012). Cytochrome p450 2B diversity and dietary novelty in the herbivorous, desert woodrat (*Neotoma lepida*). *PLoS One*, *7*(8), e41510. https://doi.org/10.1371/journal.pone.0041510

Mandeville, E. G., Parchman, T. L., Thompson, K. G., Compton, R. I., Gelwicks, K. R., Song, S. J., & Buerkle, C. A. (2017). Inconsistent

reproductive isolation revealed by interactions between *Catostomus* fish species. *Evolution Letters*, *1*(5), 255–268.

Matocq, M. D., Ochsenrider, K. M., Jeffrey, C. S., Nielsen, D. P., & Richards, L. A. (2020). Fine-scale differentiation in diet and metabolomics of small mammals across a sharp ecological transition. *Frontiers in Ecology and Evolution*, *8*, 282. https://doi.org/10.3389/fevo.2020.00282

Mauldin, M. R., Haynie, M. L., Vrla, S. C., & Bradley, R. D. (2021). Temporal evaluation of a woodrat (genus *Neotoma*) hybrid zone based on genotypic and georeferenced data. *Journal of Mammalogy*, *102*(2), 541–557. https://doi.org/10.1093/jmammal/gyaa164

McFarlane, S. E., Senn, H. V., Smith, S. L., & Pemberton, J. M. (2021). Locus-specific introgression in young hybrid swarms: Drift may dominate selection. *Molecular Ecology*, *30*(9), 2104–2115. https://doi.org/10.1111/mec.15862

McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: Comment on Breaking RAD by Lowry et al (2016). *Molecular Ecology Resources*, *17*(3), 356–361.

Menon, M., Bagley, J. C., Friedline, C. J., Whipple, A. V., Schoettle, A. W., Leal-Sàenz, A., & Eckert, A. J. (2018). The role of hybridization during ecological divergence of southwestern white pine (*Pinus strobiformis*) and limber pine (*P. flexilis*). *Molecular Ecology*, *27*(5), 1245–1260.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(12), 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Newton, G. W., Schmidt, E. S., Lewis, J. P., Conn, E., & Lawrence, R. (1981). Amygdalin toxicity studies in rats predict chronic cyanide poisoning in humans. *The Western Journal of Medicine*, *134*(2), 97–103.

Nielsen, D. P., & Matocq, M. D. (2021). Differences in dietary composition and preference maintained despite gene flow across a woodrat hybrid zone. *Ecology and Evolution*, *11*(9), 4909–4919. https://doi.org/10.1002/ece3.7399

Nosil, P., Parchman, T. L., Feder, J. L., & Gompert, Z. (2012). Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequencing data in *Timema* stick insects. *BMC Evolutionary Biology*, *12*, 164.

Oh, K. P., Aldridge, C. L., Forbey, J. S., Dadabay, C. Y., & Oyler-McCance, S. J. (2019). Conservation genomics in the sagebrush sea: Population divergence, demographic history, and local adaptation in sage-grouse (*Centrocercus* spp.). *Genome Biology and Evolution*, *11*(7), 2023–2034. https://doi.org/10.1093/gbe/evz112

Olmedo, O. E. (2014). kriging: Ordinary kriging. R package version 1.1. https://CRAN.R-project.org/package=kriging

Oswald, J. A., Harvey, M. G., Remsen, R. C., Foxworth, D. U., Dittmann, D. L., Cardiff, S. W., & Brumfield, R. T. (2019). Evolutionary dynamics of hybridization and introgression following the recent colonization of glossy ibis (Aves: *Plegadis falcinellus*) in the New World. *Molecular Ecology*, *28*(7), 1675–1691.

Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A. C., Zhang, G., Jarvis, E. D., Schlinger, B. A., & Buerkle, C. A. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, *22*(12), 3304–3317. https://doi.org/10.1111/mec.12201

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*(12), 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x

Patton, J. L., Huckaby, D. G., & Álvarez-Castañeda, S. T. (2007). *The evolutionary history and a systematic revision of woodrats of the* Neotoma lepida *group*. University of California Press.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360. https://doi.org/10.1111/mec.13557

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pfeifer, S. P., Laurent, S., Sousa, V. C., Linnen, C. R., Foll, M., Excoffier, L., Hoekstra, H. E., & Jensen, J. D. (2018). The evolutionary history of Nebraska deer mice: Local adaptation in the face of strong gene flow. *Molecular Biology and Evolution*, *35*(4), 792–806. https://doi.org/10.1093/molbev/msy004

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11479–11484. https://doi.org/10.1073/pnas.201394398

Schield, D. R., Adams, R. H., Card, D. C., Perry, B. W., Pasquesi, G. M., Jezkova, T., Portik, D. M., Andrew, A. L., Spencer, C. L., Sanchez, E. E., Fujita, M. K., Mackessy, S. P., & Castoe, T. A. (2017). Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. *Ecology and Evolution*, *7*(11), 3951–3966. https://doi.org/10.1002/ece3.2996

Schumer, M., Powell, D. L., Delclós, P. J., Squire, M., Cui, R., Andolfatto, P., & Rosenthal, G. G. (2017). Assortative mating and persistent reproductive isolation in hybrids. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(41), 10936–10941. https://doi.org/10.1073/pnas.1711238114

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Semenov, G. A., Safran, R. J., Smith, C. C. R., Turbek, S. P., Mullen, S. P., & Flaxman, S. M. (2019). Unifying theoretical and empirical perspectives on genomic differentiation. *Trends in Ecology & Evolution*, *34*(11), 987–995. https://doi.org/10.1016/j.tree.2019.07.008

Shastry, V., Adams, P. E., Lindtke, D., Mandeville, E. G., Parchman, T. L., Gompert, Z., & Buerkle, C. A. (2021). Model-based genotype and ancestry estimation for potential hybrids with mixed-ploidy. *Molecular Ecology Resources*, *21*, 1434–1451. https://doi.org/10.1111/1755-0998.13330.

Shurtliff, Q. R., Murphy, P. J., & Matocq, M. D. (2014). Ecological segregation in a small mammal hybrid zone: Habitat-specific mating opportunities and selection against hybrids restrict gene flow on a fine spatial scale. *Evolution*, *68*(3), 729–742. https://doi.org/10.1111/evo.12299

Shurtliff, Q. R., Murphy, P. J., Yeiter, J. D., & Matocq, M. D. (2013). Experimental evidence for asymmetric mate preference and

aggression: Behavioral interaction in a woodrat (*Neotoma*) hybrid zone. *BMC Evolutionary Biology*, *13*, 220.

Simon, A., Fraïsse, C., El Ayari, T., Liautard-Haag, C., Strelkov, P., Welch, J. J., & Bierne, N. (2021). How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology*, *34*(1), 208–223. https://doi.org/10.1111/jeb.13709

Souissi, A., Bonhomme, F., Manchado, M., Bahri-Sfar, L., & Gagnaire, P.-A. (2018). Genomic and geographic footprints of differential introgression between two divergent fish species (*Solea* sp.). *Heredity*, *121*(6), 579–593.

Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLoS Biology*, *17*(7), e3000391. https://doi.org/10.1371/journal.pbio.3000391

Stelkens, R. B., Schmid, C., & Seehausen, O. (2015). Hybrid breakdown in cichlid fish. *PLoS One*, *10*(5), e0127207.

Sung, C.-J., Bell, K. L., Nice, C. C., & Martin, N. H. (2018). Integrating Bayesian genomic cline analyses and association mapping of morphological and ecological traits to dissect reproductive isolation and introgression in a Louisiana iris hybrid zone. *Molecular Ecology*, *27*(4), 959–978. https://doi.org/10.1111/mec.14481

Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, *2*(2), 125–141. https://doi.org/10.1016/0040-5809(71)90011-6

Szymura, J. M., & Barton, N. H. (1986). Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, *40*(6), 1141–1159.

Tavares, H., Whibley, A., Field, D. L., Bradley, D., Couchman, M., Copsey, L., Elleouet, J., Burrus, M., Andalo, C., Li, M., Li, Q., Xue, Y., Rebocho, A. B., Barton, N. H., & Coen, E. (2018). Selection and gene flow shape genomic islands that control floral guides. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(43), 11006–11011. https://doi.org/10.1073/pnas.1801832115

Taylor, S. A., Curry, R. L., White, T. A., Ferretti, V., & Lovette, I. (2014). Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution*, *68*(11), 3066–3081. https://doi.org/10.1111/evo.12510

Turner, S. D. (2014). qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. biorXiv. https://doi.org/10.1101/005165

van Waterschoot, R. A. B., Rooswinkel, R. W., Wagenaar, E., van der Kruijssen, C. M. M., van Herwaarden, A. E., & Schinkel, A. H. (2009). Intestinal cytochrome P450 3A plays an important role in the regulation of detoxifying systems in the liver. *The FASEB Journal*, *23*(1), 224–231. https://doi.org/10.1096/fj.08-114876

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), 87–100. https://doi.org/10.1038/nrg.2016.133

Zieliński, P., Dudek, K., Arntzen, J. W., Palomar, G., Niedzicka, M., Fijarczyk, A., Liana, M., Cogălniceanu, D., & Babik, W. (2019). Differential introgression across newt hybrid zones: Evidence from replicated transects. *Molecular Ecology*, *28*(21), 4811–4824. https://doi.org/10.1111/mec.15251

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Jahner, J. P., Parchman, T. L., & Matocq, M. D. (2021). Multigenerational backcrossing and introgression between two woodrat species at an abrupt ecological transition. *Molecular Ecology*, 30, 4245–4258. https://doi.org/10.1111/mec.16056