

ORIGINAL ARTICLE

Patterns of divergence across the geographic and genomic landscape of a butterfly hybrid zone associated with a climatic gradient

Sean F. Ryan^{1,2}  | Michael C. Fontaine³  | J. Mark Scriber^{4,5} | Michael E. Pfrender^{2,6} | Shawn T. O'Neil^{2,7} | Jessica J. Hellmann^{2,8}¹USDA ARS, Gainesville, FL, USA²Department of Biological Sciences, University of Notre Dame, South Bend, IN, USA³Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, AG Groningen, The Netherlands⁴Department of Entomology, Michigan State University, East Lansing, MI, USA⁵McGuire Center for Lepidoptera and Diversity, University of Florida, Gainesville, FL, USA⁶Environmental Change Initiative, University of Notre Dame, South Bend, IN, USA⁷Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA⁸Institute on the Environment and Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN, USA**Correspondence**

Sean F. Ryan, USDA ARS, Gainesville, FL, USA.

Email: citscisean@gmail.com

Funding information

Division of Environmental Biology, Grant/Award Number: DEB-0918879, DEB-0919147

Abstract

Hybrid zones are a valuable tool for studying the process of speciation and for identifying the genomic regions undergoing divergence and the ecological (extrinsic) and nonecological (intrinsic) factors involved. Here, we explored the genomic and geographic landscape of divergence in a hybrid zone between *Papilio glaucus* and *Papilio canadensis*. Using a genome scan of 28,417 ddRAD SNPs, we identified genomic regions under possible selection and examined their distribution in the context of previously identified candidate genes for ecological adaptations. We showed that differentiation was genomewide, including multiple candidate genes for ecological adaptations, particularly those involved in seasonal adaptation and host plant detoxification. The Z chromosome and four autosomes showed a disproportionate amount of differentiation, suggesting genes on these chromosomes play a potential role in reproductive isolation. Cline analyses of significantly differentiated genomic SNPs, and of species-diagnostic genetic markers, showed a high degree of geographic coincidence (81%) and concordance (80%) and were associated with the geographic distribution of a climate-mediated developmental threshold (length of the growing season). A relatively large proportion (1.3%) of the outliers for divergent selection were not associated with candidate genes for ecological adaptations and may reflect the presence of previously unrecognized intrinsic barriers between these species. These results suggest that exogenous (climate-mediated) and endogenous (unknown) clines may have become coupled and act together to reinforce reproductive isolation. This approach of assessing divergence across both the genomic and geographic landscape can provide insight about the interplay between the genetic architecture of reproductive isolation and endogenous and exogenous selection.

KEYWORDS

climate, cline, genomic, hybridization, intrinsic incompatibilities, RAD

1 | INTRODUCTION

Hybrid zones are ideal systems for exploring the interplay between exogenous (ecological factors, such as climate) and endogenous (genetic incompatibilities) selective pressures that drive genomic

divergence. Although climate has been implicated in the maintenance of a number of hybrid zones across a broad range of taxa (Taylor, Larson, & Harrison, 2015), our understanding of how climate drives patterns of divergence and gene flow between the genomes of hybridizing species remains limited to a handful of systems

(Hamilton, la Torre, & Aitken, 2014; Taylor et al., 2014; de Villemerueuil, Frichot, Bazin, François, & Gaggiotti, 2014).

It has long been recognized that an advantage to studying hybrid zones is their utility in understanding the genetic basis of species divergence, that is, identifying those regions of the genome under selection and potentially involved in reproductive isolation (Barton & Hewitt, 1985). A classical and still widely used approach is to look at how genetic differences between species vary across geography and species boundaries, through a geographic cline analysis. Complementary to cline analysis, there are a number of population genomic approaches to identify genomic regions under selection (i.e., via “genome scans”) that do not require spatial sampling along a cline. Genome scans can be used to identify outlier loci from a null expectation based on genomewide ancestry that are potential indicators of divergent selection (Foll & Gaggiotti, 2008). Given that geographic cline analyses and genome scans estimate selection dynamics operating on different spatiotemporal scales, from recent to more long-term timescale, respectively, combining these approaches increases the power to detect regions of the genome involved in local adaptation and reproductive isolation (de Lafontaine, Prunier, Gérardi, & Bousquet, 2015).

Placing patterns of genetic differentiation within the genomic landscape can also provide insight as to how genetic architecture itself influences introgression across the genome—the “genomic architecture of divergence” (Gompert et al., 2012; Teeter et al., 2010). A number of studies suggest that gene flow may be heterogeneous across the genome of hybridizing species (Mallet, 2005; Teeter et al., 2008), with large amounts of the genome diffusing between species while a few regions maintain exceptionally high levels of divergence (Fontaine et al., 2015). Specifically, regions of low recombination such as inversions, centromeres and sex chromosomes are often found to have elevated levels of divergence and often harbour a disproportionate amount of the factors responsible for reproductive isolation (Coyne & Orr, 1989; Presgraves, 2008).

Both exogenous and endogenous factors can act to inhibit introgression and maintain species boundaries. Endogenous factors have often been found to play a major role in driving clinal patterns of genetic variation across many hybrid zones (Bierne, Welch, Loire, Bonhomme, & David, 2011). For example, most tension zones, hybrid zones maintained by a balance between selection and migration, are believed to be the result of intrinsic incompatibles. The geographic selection gradient model (GSGM), a mathematically similar model to the tension zone model, can also produce similar clinal patterns, with the main difference being selection is primarily exogenous (e.g., related to abiotic factors; Endler, 1977). However, endogenous and exogenous barriers are not mutually exclusive (Barton & Hewitt, 1985; Bierne et al., 2011; Hewitt, 1988). In fact, these barriers may become “coupled” and act synergistically to increase reproductive isolation and maintain the location of the hybrid zone, often resulting in tension zones becoming associated (co-occurring) within an ecological gradient (Barton & Hewitt, 1985; Bierne et al., 2011).

The hybrid zone between the eastern tiger swallowtail (*Papilio glaucus*) and the Canadian tiger swallowtail (*P. canadensis*) butterfly is an excellent system for exploring how climate influences genomic patterns of divergence. The ecology of this system has been studied extensively for over thirty years and climate is implicated as a prominent factor in maintaining reproductive isolation between the two species (Scriber, 2011). However, an understanding of how endogenous and exogenous selective forces maintain divergence across the genome remains unresolved. The two species are estimated to have diverged 0.5–0.6 million years ago (Cong, Borek, Otwinowski, & Grishin, 2015; Kunte et al., 2011) and previous genetic and morphological analysis of individuals suggests they form a narrow (<50 km) east–west hybrid zone in the northeastern United States, extending from Minnesota to New England (Figure 1a; Scriber, 1990). A number of phenotypic traits, including those that are physiological (Kukal, Ayres, & Scriber, 1991; Mercader & Scriber, 2008; Scriber, Lindroth, & Nitao, 1989), behavioural (Deering & Scriber, 2002; Mercader, Aardema, & Scriber, 2009; Scriber, Giebink, & Snider, 1991) and morphological (Lehnert, Scriber, Gerard, & Emmel, 2012; Luebke, Scriber, & Yandell, 1988), appear to coincide (exhibit similar clines) with latitudinal and altitudinal variation in climate (Hagen & Scriber, 1989; Scriber, 2002; Scriber, Stump, & Deering, 2003). A better understanding of how genes putatively involved in ecological adaptations in this system covary with environmental factors (the geographic landscape) and how they are distributed across the genome (the genomic landscape) will help to elucidate how ecological selection pressures drive reproductive isolation within the genomes of these hybridizing species.

The hybrid zone between *P. glaucus* and *P. canadensis* is believed to be maintained primarily by a climate-mediated developmental threshold, produced by geographic variation in the length of the growing season (Scriber & Lederhouse, 1992). One line of evidence implicating the growing season is that the hybrid zone coincides with a change in a number of ecologically adaptive traits related to seasonality including the transition of facultative diapause induction of *Papilio glaucus* to the obligate diapause induction of *P. canadensis*. It is presumed that *P. glaucus* alleles involved in facultative diapause would be disadvantageous north of the hybrid zone where on average the growing season is too short to complete two generations (Ritland & Scriber, 1985). Conversely, alleles underlying obligate diapause phenotype may be disadvantageous south of the hybrid zone, as they would likely lead to phenotypes with lower fitness than those that emerged for a second generation.

In the *P. glaucus*–*P. canadensis* system, diapause is largely Z-linked (Hagen & Scriber, 1989); like birds, Lepidoptera has a ZW system ($Z = X$, $W = Y$) of sex determination and females are the heterogametic sex. Divergent selection acting on genes involved in diapause has been invoked as a possible explanation for the observed lack of introgression of mtDNA and Z-linked markers (Scriber, 2011; Scriber, Elliot, Maher, McGuire, & Niblack, 2014; Scriber, Ordng, & Mercader, 2008). Previous studies have documented a dramatic change in allele frequencies of Z-linked allozymes (LDH and PGD), an mtDNA gene (cytochrome oxidase I) and physiological

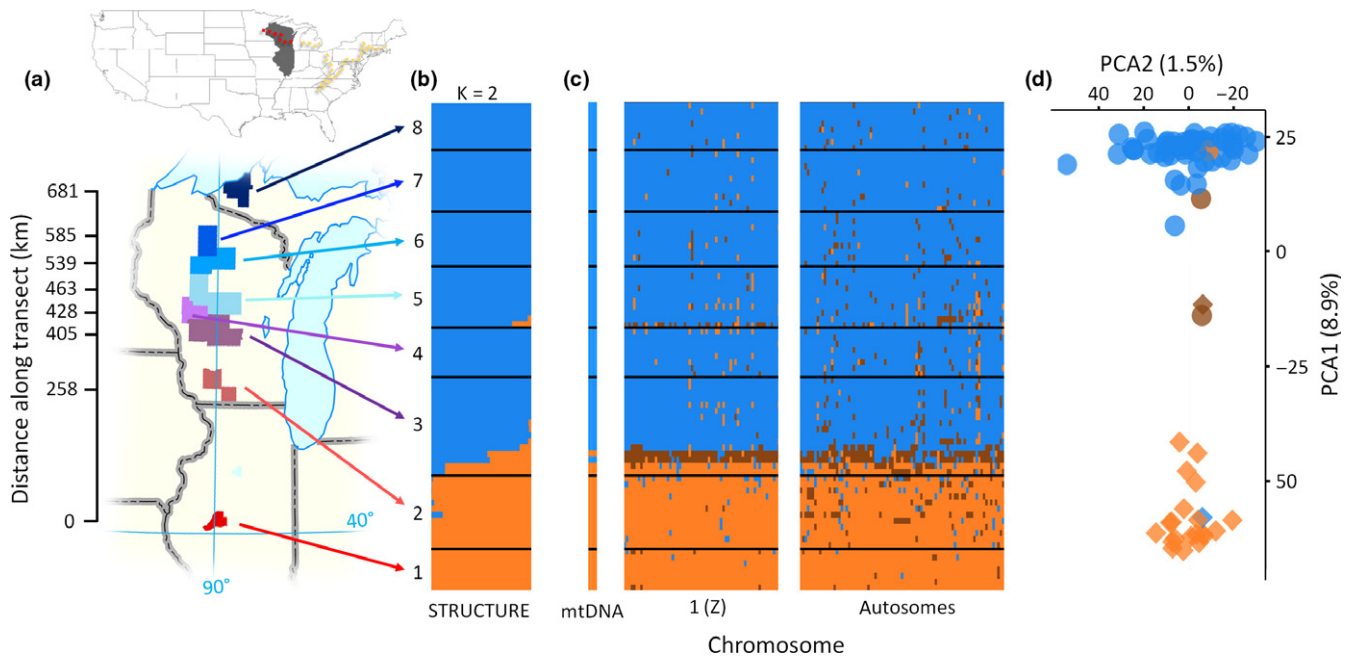


FIGURE 1 The genomic landscape of divergence: evidence of ecological selection across the genome. (a) Map showing sampling locations for each population; counties with the same colour were pooled for analysis. (b) Bayesian clustering of ancestry proportion to $K = 2$. (c) Maximum-likelihood estimates of hybrid index of one mtDNA marker and 164 SNPs. Each row is an individual and column a genetic marker. Colour of blocks indicates genotype or mtDNA haplotype (for the mtDNA)—orange and blue blocks represent homozygous genotype for *P. glaucus* and *P. canadensis*, respectively, and brown blocks indicate heterozygote genotypes. Black horizontal lines separate individuals by location. (d) Principal components analysis showing the individual scores for the first two PCs. The proportion of variance explained by each axis is indicated between parentheses. Colour indicates whether individuals had the *P. glaucus* (orange), *P. canadensis* (blue) or were heterozygous (brown) for a species-diagnostic SNP in *Ldh* and shape indicates an individual's mtDNA haplotype (*P. glaucus*: diamonds, *P. canadensis*: circles) [Colour figure can be viewed at wileyonlinelibrary.com]

traits (geographic distribution of host-specific detoxification abilities and host plant preferences) that coincide with the environmental gradient that is the length of the growing season (reviewed in Scriber, 2011). Given that diapause appears to be an important ecological trait that differs between these species and may be involved in maintaining the location of the hybrid zone, Scriber (2011) suggested that the Z chromosome may harbour a disproportionate amount of divergence.

In addition to climatic variation, other extrinsic factors may be involved in the maintenance of this hybrid zone, including the distribution of host plants (Lindroth, Scriber, & Hsia, 1988; Mercader et al., 2009). While both species are highly phytophagous, they differ in their preferences for, and ability to metabolize, linear and angular furanocoumarins, resulting in differential detoxification abilities for some host species (e.g., *Populus tremuloides* can be detoxified by *P. canadensis*, but not *P. glaucus*, with the opposite true for *Liriodendron tulipifera*; Hung, Berenbaum, & Schuler, 1997). Candidate genes involved in, or linked to, traits associated with various ecological adaptations, such as host plant detoxification, larval chemical defences and diapause, have been identified (Cong et al., 2015; Hagen & Scriber, 1989; Li, Petersen, Schuler, & Berenbaum, 2002; Li, Schuler, & Berenbaum, 2003). Yet, how these candidate genes are distributed within the genome and coincide with patterns of genomic divergence, or how they vary with geography and ecological

gradients is not well understood. Further, given that extrinsic factors vary in their geographic distribution, they may produce different underlying selective landscapes that could lead to different patterns of introgression across the hybrid zone.

In this study, we characterized genetic differentiation across the genomic and geographic landscape of the ecologically well-studied *P. glaucus* and *P. canadensis* hybrid zone. Specifically, we screened a reduced representation of genetic diversity across the genome using a ddRADseq protocol and characterized genomewide patterns of differentiation between *P. glaucus* and *P. canadensis* along a 650-km latitudinal transect in the Wisconsin portion of the hybrid zone. First, we set out to characterize genetic structure between the two species and identify regions of the genome potentially impacted by natural selection. We then placed these patterns of differentiation within the genomic landscape to explore the genomic architecture of differentiation and assess whether differentiation is genomewide or variable among chromosomes (particularly the Z chromosome). Further, we evaluated the prediction that the distribution of genes previously identified as being putatively involved in ecological adaptations would be associated with regions of the genome under selection and exhibit steep clinal variation across the hybrid zone (i.e., provide further evidence for their being involved in ecological adaptations). Finally, we used a geographic cline approach to explore how significantly differentiated genomic regions (F_{ST} outliers in

allopatry) vary across geography (and exhibit steep clinal variation in sympatry—hybrid zone) and evaluate the relative roles endogenous and exogenous factors may be contributing to patterns of differentiation observed within the genome.

2 | METHODS AND MATERIALS

2.1 | Sample collection and preservation

Three hundred and fourteen specimens were collected between 2007 and 2013 from across a latitudinal transect that spans the *P. canadensis* and *P. glaucus* hybrid zone, from the Upper Peninsula of Michigan through Wisconsin to central Illinois (Table 1; Figure 1a). Of these specimens, all 314 were used to genotype individuals for species-diagnostic mitochondrial haplotypes in the cytochrome oxidase subunit 1 (*COI*) gene (Sperling, 1993), 192 for a species-diagnostic single nucleotide polymorphism (SNP) in the lactate dehydrogenase (*Ldh*) gene and 101 (80 after filtering) for the generation of reduced representation of the genomewide diversity using the ddRADseq protocol. Only males were used in all analyses because females cannot be heterozygous for loci on the Z chromosome (females are hemizygous).

2.2 | DNA extraction and genotyping/phenotyping of diagnostic markers

Briefly, gDNA was isolated from leg tissue using a phenol-chloroform protocol (Sambrook, Fritsch, & Maniatis, 1989). The two species-diagnostic mtDNA haplotypes (*COI* gene) were scored using a restriction fragment length polymorphism (RFLP) protocol adapted from Putnam, Scriber, and Andolfatto (2007). A TaqMan SNP genotyping procedure was used to genotype individuals for the Z-linked SNP located in the coding region of *Ldh* (Table S1). This mutation leads to a nonsynonymous substitution that is fixed between the two species and is believed to lead to different LDH isozymes (Putnam et al., 2007) that may result in thermophysiological differences between the species (see Fig. S1). More details on DNA extraction genotyping of diagnostic markers are provided in supplementary material; all data can be found in Data S1.

2.3 | ddRADseq library preparation, sequencing and bioinformatic postprocessing

Two reduced-complexity libraries were generated for 101 individuals using a variation of the restriction-site-associated DNA fragment (RAD) procedure (adapted from Nosil et al., 2012) involving a double digest (ddRADseq; Peterson et al., 2012). Briefly, ~600 ng of genomic DNA was digested using the restriction enzymes *EcoRI* and *MseI*. For each individual, the *EcoRI* adapter contained a unique 8- to 10-base pair (bp) index for barcoding. PCR product was then pooled (into a five and 96 individual library), purified with QIAquick Purification Kit (Qiagen) and size-selected (400–600 bp range) on a BluePipin System (Sage Science). One library (containing five individuals)

TABLE 1 Sampling locations along the transect

Location	County	Latitude	Distance			
			(km)	n_{COI}	n_{Ldh}	n_{ddRAD}
1	Mason	40.38352	0	24	10	11
	Putnam	41.22890	94	1	–	–
	Rock Island	41.528373	127	–	–	–
2	Green	42.64672	252	10	10	2
	Iowa	43.05780	297	39	15	12
	Dane	43.076369	299	–	–	–
	Richland	43.37849	333	4	–	–
3	Sauk	43.42960	339	11	8	2
	Columbia	43.60436	358	7	7	2
	Green Lake	43.838402	384	–	–	–
	Marquette	43.89078	390	5	5	3
	Monroe	43.94569	396	1	–	–
	Adams	44.03663	406	14	13	9
4	Juneau	44.16089	420	19	16	5
	Jackson	44.23612	428	16	15	9
	Wood	44.44880	452	16	15	6
	Portage	44.47409	455	1	1	1
5	Clark	44.69498	479	21	15	8
	Marathon	44.89816	502	7	–	–
	Taylor	45.16705	532	30	10	6
6	Lincoln	45.36137	554	19	11	3
	Oneida	45.57978	578	9	8	1
	Price	45.64626	585	14	14	10
7	Vilas	45.95206	619	32	10	1
	Iron	46.26355	654	5	–	–
	Ontonagon	46.51145	681	9	9	10
	Total			314	192	101

The transect is measured as the distance from Mason Co. using the "Haversine equation" and 90°N as longitude and 6,371 km for the radius of the earth.

n_{COI} : number of individuals genotyped for species-diagnostic SNP in *COI* gene.

n_{Ldh} : number of individuals genotyped for species-diagnostic SNP in *Ldh* gene.

n_{RADseq} : number of individuals sequenced with ddRADseq.

was run on an Illumina MiSeq (Notre Dame Genomics Core) and one (containing 96 individuals) was run on an Illumina HiSeq 2000 (BGI, UC Davis) using paired-end sequencing (150-bp reads on MiSeq and 100-bp reads on the HiSeq).

Raw reads were demultiplexed using a custom Perl script and trimmed and cleaned with the program TRIMMOMATIC (v.0.32, Bolger, Lohse, & Usadel, 2014) using default settings. The first 20 bp from all reads was clipped to remove barcodes and cutsite and only reads with a minimum read length of 30 bp were kept. The 3' end of forward unpaired reads were of lower quality and were trimmed to 80 and 60 bp for MiSeq and HiSeq reads, respectively. Reads were then aligned to the *Papilio glaucus* genome v1 (Cong et al., 2015) using BWA-*aln* (v.0.7.8, Li & Durbin, 2009). Read groups were added using

custom Perl scripts, soft-clipped using PICARD (v.1.33; <http://pica.sourceforge.net>) and realigned using GATK v.3.3 INDELREALIGNER (DePristo et al., 2011). Variant calling was performed using GATK HAPLOTYPICALLER (DePristo et al., 2011; McKenna et al., 2010) with the default settings. Filters were applied to all SNPs in the following order. Variants were filtered for biallelic SNPs and indels removed using VCFTOOLS v0.1.12a (Danecek et al., 2011). Individuals with excessive missing data (>90%) were removed ($n = 3$). Rare alleles with a frequency >95% and <5% (pooling all individuals) were removed using VCFTOOLS. GATK's default "hard filtering" was applied to keep only high-quality variants ($QD > 2.0$, $MQ > 30$, $ReadPosRankSum > -8.0$, $HaplotypeScore > 13$, $MappingQualityRankSum > -12.5$). Only genotypes with a ~99% likelihood of being correct were kept by removing SNPs with a genotype quality (GQ) <20. Following these filters, a minimum of 6× coverage for at least 75% of individuals in the "pure" species populations (*P. glaucus*: locations 1 and 2; *P. canadensis*: locations 7 and 8) was then applied (referred to as RAD1 data set; 28,417 SNPs). For clinal analyses, we used a subset of the RAD1 data set that had the additional requirement that a SNP had a minimum of 6× coverage in at least 75% of individuals from all eight locations. Following this filter, individuals with less than an average of 14× coverage (for all SNPs) were removed (referred to as RAD2 data set; 18,316 SNPs) because the estimates of heterozygosity for individuals below this cut-off could be biased by low coverage (Fig. S2).

2.4 | Putative chromosome assignment of genetic markers

To determine the chromosome and orientation of each *P. glaucus* scaffold (while maintaining the orientation of SNPs within the *P. glaucus* scaffold), we used blastx (ncbi-blast-2.2.30+; Altschul, Gish, Miller, Myers, & Lipman, 1990; Shiryev, Papadopoulos, Schäffer, & Agarwala, 2007) to blast all peptide sequences within each scaffold of the *P. glaucus* assembly against the *Bombyx mori* genome using SILKDB v.2.0 (Wang et al., 2005). The *B. mori* scaffold with the most significant blast hits (based on Bit scores) was retained and used to determine the chromosome of each *P. glaucus* scaffold and their orientation within a chromosome. Given the high level of synteny between *P. glaucus* and *B. mori* (Cong et al., 2015), we expect this approach to reflect a moderately accurate representation of the genomic distribution of our ddRADseq loci.

2.5 | Genetic diversity and structure

Genetic structure was assessed using the Bayesian clustering approach of STRUCTURE v2.3.4 (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000) using an admixture model with correlated allele frequencies, a burn-in of 100,000 Monte Carlo Markov chains (MCMCs) and a run length of 500,000 iterations, testing clusters (K) from 1 to 5; 15 replicates were run for each value of K . We used the RAD2 data set pruned for linkage disequilibrium (LD) by keeping only 1 SNP per 1 kb ($n = 6,769$ SNPs). The

results were postprocessed using CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015), and the best K was chosen following Evanno's method (Evanno, Regnaut, & Goudet, 2005) as implemented in STRUCTURE HARVESTER v.0.6.94 (Dent & vonHoldt, 2012).

A principal component analysis (PCA) was also performed in R v 3.2.2 (R Core Team 2013) using the LD-pruned RAD1 data set and the package ADEGENET v.1.4-2 (Jombart & Ahmed, 2011) to provide a complementary visualization of the genetic structure without relying on any model-based assumptions.

We estimated the levels of genetic diversity within each locality from the RAD2 data set using the per-site nucleotide diversity (π ; Nei & Li, 1979), as well as the inbreeding index (F_{IS}) using VCFTOOLS v0.1.12b. Linkage disequilibrium (LD) was calculated as mean r^2 for all pairwise SNPs within each location using a modified version of plink-1.07 that allows for our large number of "linkage groups" (scaffolds; Malenfant, Coltman, & Davis, 2015). Confidence intervals were generated by randomly subsampling 100 times seven individuals (smallest n of all locations) from each location. Using these estimates of r^2 , we then calculated the distance (in kb) at which r^2 is predicted to fall below 0.2 for each chromosome (1–28) by fitting the relationship between r^2 and distance between SNPs within each scaffold to a non-linear model (Hill & Weir, 1988) describing the rate of LD decay similar to Marroni et al. (2011). A subset of the RAD2 data set, using SNPs with $F_{ST} > 0.9$ and filtered to 1 SNP per 1 kb (i.e., if more than one SNP were within <1 kb within a scaffold, only the first SNP was retained; 164 SNPs) were used to calculate the hybrid index and heterospecific heterozygosity of each individual using the hi.index and int.het functions in the INTROGRESS package (Gompert & Buerkle, 2010) in the R computing environment (R Core Team 2013).

2.6 | Detecting (outlier) loci potentially under selection

We investigated the pattern of interspecific differentiation using two genome scan approaches conducted on the allopatric populations at the extremities of the transect, locations 1 and 2 for *P. glaucus* ($n = 18$) and locations 7 and 8 for *P. canadensis* ($n = 18$). After removal of monomorphic loci, we first quantified the level of differentiation in allelic frequencies for each SNP using the data set filtered using just the allopatric populations (RAD1; 28,417 SNPs) with the F_{ST} estimator of Weir and Cockerham (1984) using VCFTOOLS. We identified SNPs potentially evolving under natural selection using the F_{ST} outliers approach originally described in Beaumont and Nichols (1996) and implemented in ARLEQUIN v.3.5 (Excoffier & Lischer, 2010). We estimated the 99.9% neutrality envelope of F_{ST} as a function of the heterozygosity using 500,000 coalescent simulations under a neutral model of population evolution and a Hierarchical Island Model consisting of 10 groups and 100 demes. SNPs above the upper 99.9% quantile were considered under potential divergent selection while those under the lower 0.001% quantile to be under potential balancing selection.

In addition, we used a Bayesian method implemented in BAYESCAN v.2.1 to estimate the probability that each locus is subject to divergent selection (Foll & Gaggiotti, 2008). To account for multiple comparisons and reduce the amount of false positives, we set the prior odds for the neutral model (pr_odds) to 10,000 and used a false discovery rate of 0.01. Twenty pilot runs of 5,000 iterations were used with a thinning interval of 10 and a burn-in of 50,000 iterations followed by 5,000 iterations for output.

Loci were considered as outliers for potential divergent selection if they were identified as significant with both methods (BAYESCAN and ARLEQUIN). To test whether the outlier loci were nonrandomly distributed across chromosomes, a Fisher's exact test was used to compare the distribution of outlier loci for divergent selection and balancing selection to a null distribution, using all linkage groups as well as a separate analysis with just autosomes; a Bonferroni correction was applied to control for multiple testing ($\alpha = 0.01/\text{number of linkage groups}$; $\alpha = 0.00034$ and 0.00038 when using all linkage groups and just autosomes, respectively). This analysis was also performed using the "divergence hot spots" (genes that always show larger interspecies variation on both the protein and DNA sequence levels) identified in Cong et al. (2015) to determine whether these highly divergent regions were nonrandomly distributed across linkage groups and compare their distribution to outlier regions from our study.

2.7 | Geographic cline analysis

We evaluated coincidence (amount of congruence among estimates for cline centre for each SNP) and concordance (amount of congruence among estimates for cline width for each SNP) using the RAD2 data set for ddRADseq loci identified as outliers that were in HWE in each location and filtered to 1 SNP per 1 kb (228 SNPs), as well as mtDNA (*COI*) and Z-linked (*Ldh*) loci. We fitted a combination of equations that describe the sigmoidal shape at the centre of each cline and two exponential decay curves (Szymura & Barton, 1986, 1991). A total of 15 different models that varied in shape (five possibilities: having no tails, symmetrical tails, asymmetrical tails, left tail only or right tail only) and scaling of tails (three possibilities: using no scale, empirical estimates or best fit) were compared using a Metropolis–Hasting algorithm implemented in the HZAR package in R (Derryberry, Derryberry, Maley, & Brumfield, 2014). Posterior distribution of each model parameter was estimated using three independent chains run using 500,000 MCMC steps after a burn-in of 100,000 steps. This process was repeated three times for each model, and each MCMC trace was visually inspected to ensure convergence of the results. Model selection was based on the lowest Akaike information criterion (AIC) value and parameter estimates of these selected models were used (Akaike, 1974). The cline centre and width of any two loci were considered noncoincident or discordant if the 2-log-likelihood unit (comparable to 95% confidence intervals) for the upper and lower bound estimates of cline centres or width did not overlap, respectively.

To determine whether the clinal patterns of loci coincide with a climatic gradient, the estimated centre of clinal loci was compared to geographic variation in length of the growing season. Growing degree-days (GDDs) for each county along a transect spanning the hybrid zone were calculated with PRISM time series data sets modelled using climatologically aided interpolation at a 4-km resolution (PRISM Climate Group 2014). For each year from 1980 to 2012, GDDs were calculated as the sum of the mean daily temperature $((T_{\min} - T_{\min})/2 - T_{\text{base}})$, between March 1st–October 31. A T_{base} of 10° C was chosen based on the lower activity threshold of *P. glaucus* estimated by Ritland and Scriber (1985) at 4-km pixels in ArcGIS® (ESRI v10.2) and then averaged to the county level. The 32-year mean and standard deviation among years for each county were used to examine the relationship between the total GDDs and distance across this transect. It has been estimated that the GDD requirement for *P. glaucus* to complete two generations is at least 1,300 GDDs (a conservative estimate; Ritland & Scriber, 1985). A linear model was fit to the GDD data by location (distance along the transect). The fit from this model was then used to predict the location where 1,200, 1,300, and 1,400 GDDs occur; 1,200 GDD and 1,400 GDD were used to get an idea of the geographic variance of this climate-mediated threshold. The distribution of other ecological factors—that is, hosts associated with differential detoxification, *Liriodendron tulipifera*, *Populus tremuloides* and *Ptelea trifoliata* (Little, 1971, 1976) and the Batesian model *Battus philenor* for which *P. glaucus* is a mimic—was also visually compared to the genetic clines (estimated centre of the hybrid zone).

3 | RESULTS

3.1 | Summary of ddRADseq data

One lane of Illumina HiSeq 2000 and one MiSeq run produced 346,321,096 and 12,818,490 paired reads, respectively, derived from 101 individuals. After mapping the reads to the *P. glaucus* reference genome, we called 1,835,135 SNPs, of which 28,417 loci showed a minimum of $6\times$ coverage for 75% of individuals from "pure" parental populations and were used for F_{ST} outlier analysis. Out of these, 18,316 SNPs showed a minimum of $6\times$ for at least 75% individuals for all populations, with an average coverage of $31.1X \pm 8 SD$ and average of $0.07\% \pm 0.6 SD$ missing data across individuals. This second set was used for all other analysis. A detailed summary of sequencing and SNP filtering statistics is provided in supplementary results (Table S2).

3.2 | Putative chromosome assignment of genetic markers

We used the close synteny of *Bombyx mori* and *P. glaucus* to infer the linkage group of our ddRADseq SNPs. Synteny is relatively high within Lepidoptera ($>80\%$), and *P. glaucus* and *B. mori* share $>85\%$ of their genes in micro-syntenic blocks (Cong et al., 2015). For the 28,417 ddRADseq loci used for outlier analysis, 97% mapped to a

B. mori scaffold, with 92% mapping to a linkage group (Fig. S3, Data S2), spanning all 28 *B. mori* chromosomes. As expected if the chromosome assignments were correct, we found a significantly positive relationship ($r^2 = .69$; p value $<.0001$) between *B. mori* chromosome size and number of mapped loci (Fig. S4). Thus, our ddRADseq SNPs appear to be well distributed across the genomes of *P. glaucus* and *P. canadensis*.

3.3 | Population structure and diversity

The Bayesian clustering analysis of *STRUCTURE* showed the highest likelihood values and highest rate of likelihood change as the number of clusters (K) increased from 1 to 5 was obtained for $K = 2$ (Figure 1b; Fig. S5). Most individuals (18 *P. glaucus*, 55 *P. canadensis*) clustered into two distinct groups corresponding to the two species with more than 95% assignment probability to a species cluster (Figure 1b; Data S3). The principal component analysis (PCA) revealed a similar clustering (Figure 1d).

Seven individuals showed an admixed ancestry with $<95\%$ of their genome coming from either species based on *STRUCTURE* analysis (Figure 1b; Fig. S5). Of the seven putative hybrids, two appeared to have mixed ancestry resembling early-generation hybrids ($<60\%$ assignment to either cluster and nearly all diagnostic loci between the two species are heterozygous for both individuals; see below and Figure 1c) and five appear to be later-generation hybrids (Figure 1b,d; Fig. S6). Five of these putative hybrids were found at sampling location 3, previously identified as the location of the hybrid zone (Scriber et al., 2003) where hybrids made up 31% of the individuals from this location. However, using the species-diagnostic marker, *Ldh*, for which we have a larger sample size and assuming individuals heterozygous for this marker are early-generation hybrids (a reasonable assumption), we get a lower estimate of 15% (6/41) of individuals being putative hybrids in location 3, and only 3% (6/192) across the entire transect.

3.4 | Genetic differentiation and detecting outlier loci

Genetic differentiation, measured by mean (weighted; Weir & Cockerham, 1984) F_{ST} between the two species (RAD1 data set), was moderate (0.17). However, the distribution of F_{ST} was slightly bimodal, made up largely of loci with F_{ST} value <0.25 (89%) and a small subset (3%) with estimates >0.75 (Figure 2c).

For the 28,417 SNPs we identified, of those that were at least 1 kb apart, 368 (1.3%) were significantly more differentiated in both the F_{ST} outlier (Fig. S7A) and *BAYESCAN* (Fig. S7B) analysis and could potentially be under divergent selection (Figure 2b). As for those SNPs that were less differentiated than expected under a neutral expectation, we identified 148 (0.5%) falling outside the lower 99.9% envelope (Figure 2a; Fig. S7A). Genetic differentiation between the two species was highly heterogeneous across the genome (Figure 2a; Fig. S8). While most SNPs showed low levels of differentiation (F_{ST}), a small subset displayed very high F_{ST} values, with

239 fixed differences (of 28,417 loci), 301 SNPs with F_{ST} between 0.90 and 0.99 and 234 SNPs with F_{ST} between 0.80 and 0.89 (Figure 2c). These highly differentiated ($F_{ST} > 0.90$) SNPs that were also identified as outliers were distributed across all but five chromosomes (2, 3, 18, 20 and 26) of the *Bombyx* reference genome (Figure 2b; Fig. S8). Some of these chromosomes contained highly divergent regions spanning multiple scaffolds within the chromosome, most notably on the Z chromosome where regions with high F_{ST} span nearly the entire chromosome. The Z chromosome had an average F_{ST} 4.4 times greater than the genomic background of differentiation compared to autosomes—Z-linked SNPs had an average F_{ST} of 0.36 ± 0.37 (median = 0.19) and autosomes an average F_{ST} of 0.08 ± 0.18 (median = 0.0006).

Outlier SNPs were not randomly distributed across linkage groups (Fig. S3). A significantly greater number of SNPs putatively under divergent selection as well as previously identified divergence hot spots (Cong et al., 2015) mapped to the Z chromosome (p value $<.0001$; Figure 2a,b; Fig. S3A). Among the autosomes, six chromosomes (4, 9, 10, 11, 12 and 19) were significantly enriched for outlier loci under divergent selection (p value $<.001$; Fig. S3B). Of these, chromosomes 4, 10, 12 and 19 were also significantly enriched for divergence hot spots. Contrastingly, there were twelve chromosomes (5, 7, 8, 16, 17, 18, 20, 21, 25, 26, 27 and 28) that had significantly fewer outliers than expected (p value $<.001$; Fig. S3A).

3.5 | Geographic cline analysis

The prediction that there will be a high degree of coincidence (similar cline centre) among genetic and phenotypic clines was highly supported by our data. A majority of the outlier SNPs that were in HWE in each location and filtered to 1 SNP per 1 kb were coincident (81%; 187/230) and concordant (80%; 184/230) with the estimated median cline centre of 387 and 57 km median width (Figure 3b). This estimate places the centre of the hybrid zone at $\sim 43.8^\circ\text{N}$ (Juneau Co., Wisconsin; Figure 3b). The estimated centre of the hybrid zone (location 3) also coincided with a steep cline in hybrid index and a peak in LD (r^2) and π (Figure 3c). Estimates of LD blocks (distance in bp between SNPs with $r^2 > .2$) were also largest in the hybrid zone (locations 3–5 averaging 75 kb), ranging from 36 to 78 kb across the transect and averaging 61 kb ($SE \pm 6.5$ kb) in size.

While there was an overall high level of coincidence and concordance across SNPs, 77% of chromosomes (17 of the 22 chromosomes containing SNPs for this analysis) had at least one locus that was discordant or noncoincident (based on nonoverlapping confidence intervals with the average estimate; Data S4). Of the 44 SNPs with noncoincident clines, the majority (40/44; 90%) were shifted in the direction of *P. glaucus*, while only 10% (4/44) were in the direction of *P. canadensis* (Figure 4). SNPs with clines shifting towards *P. glaucus* were found on thirteen different chromosomes. Of the four markers shifting towards *P. canadensis*, two were found on chromosomes 23 while one was on chromosome 1 (Z) and all of these

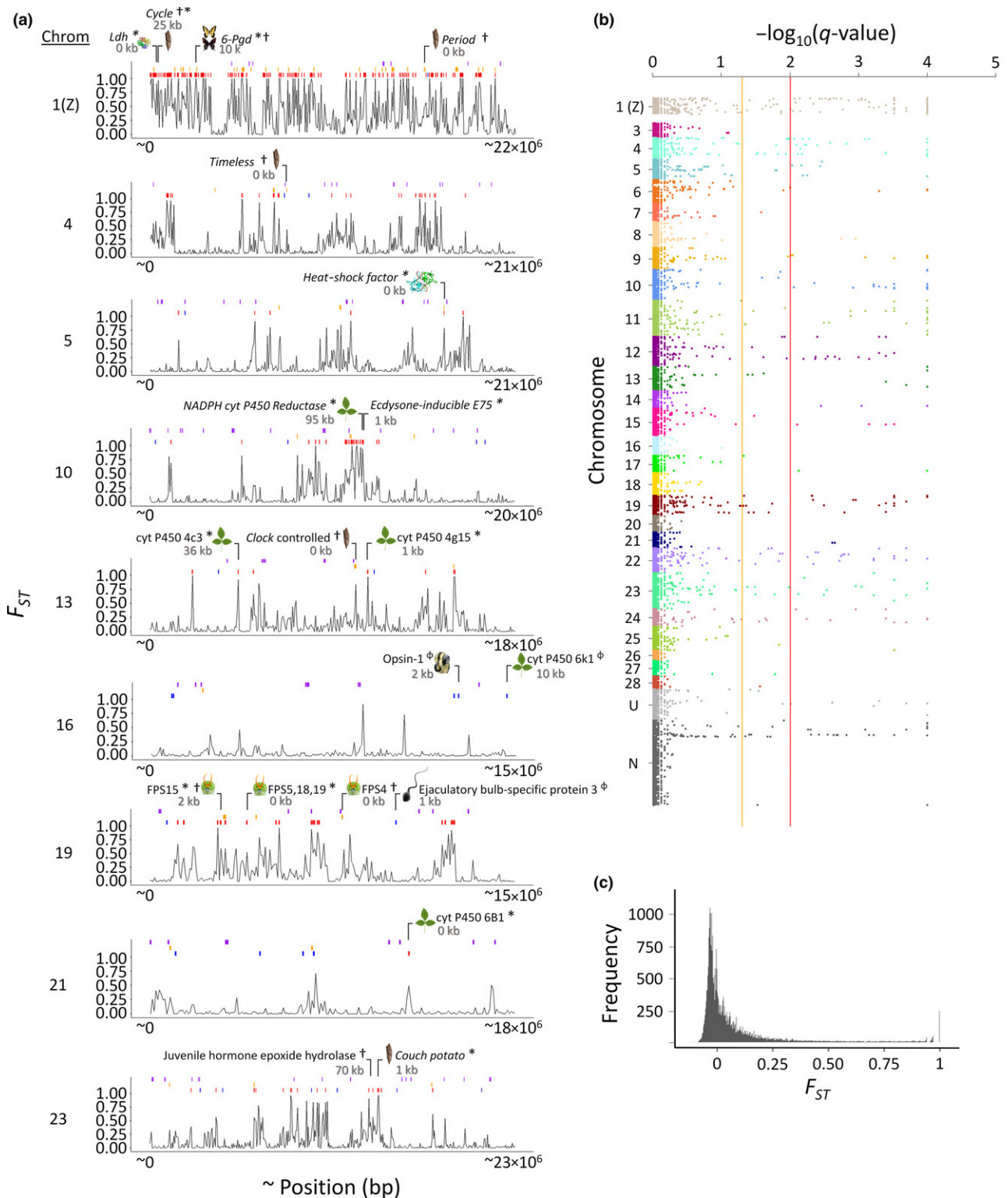


FIGURE 2 Genomic landscape of divergence between *P. glaucus* and *P. canadensis*. (a) Differentiation (F_{ST} averaged in nonoverlapping 1-kb windows) within linkage groups containing candidate genes for ecologically divergent traits (position of candidate genes shown at the top of each linkage group along with distance from closest ddRADseq locus); note that these distances are <1 cM and the average LD decay of ~80 kb for $r^2 > .2$. *, ϕ indicate genes associated with loci under divergent and balancing selection, respectively. † indicates genes that were identified as divergence hot spots by Cong et al. (2015). Red and blue bars indicate loci under divergent and balancing selection, respectively (F_{ST} outlier); orange and purple bars represent loci within 1 kb of genes identified as being divergence hot spots or under positive selection by Cong et al. (2015), respectively. (b) Manhattan plot of $\log_{10}(Q\text{-values})$ from BAYESCAN analysis by linkage group. Coloured lines indicate threshold for false discovery rate (orange = 0.05; red = 0.01). Chromosome “U” and “N” represent loci that did not have linkage group information or did not map to a *B. mori* scaffold, respectively. (c) Histogram of F_{ST} values for all 28,417 ddRADseq loci [Colour figure can be viewed at wileyonlinelibrary.com]

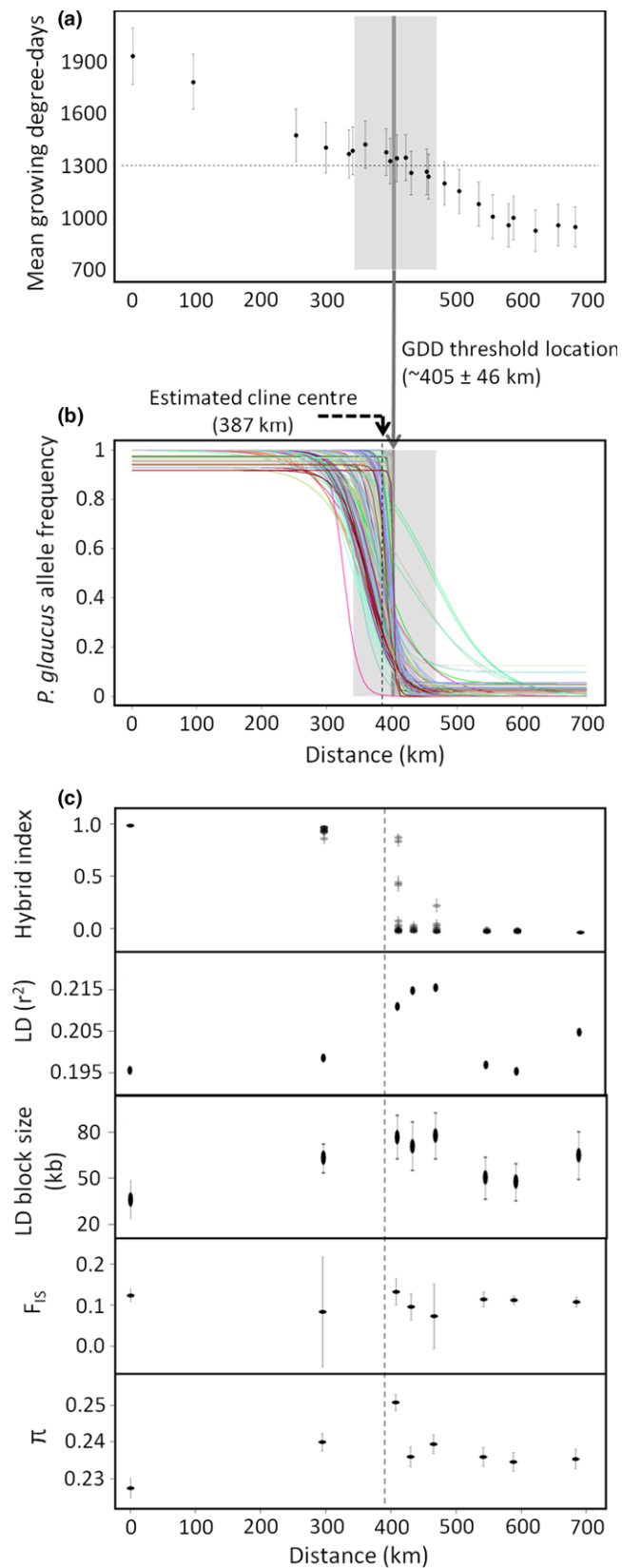


FIGURE 3 The geographic landscape of divergence. (a) 32-year average length of the growing season (mean growing degree-days) across the *P. glaucus*-*P. canadensis* hybrid zone. Error bars are standard deviation. The dark grey shaded region illustrates the (1,300) minimum growing degree-day requirement (light grey regions ± 100 GDD) for *P. glaucus* to complete two generations (Ritland & Scriber, 1985). (b) Maximum-likelihood cline for 230 outlier SNPs (in HWE in each location and filtered to 1 SNP per 1 kb). Allele frequency clines that share the same colour map to the same (*B. mori*) chromosome (See Figure 2b). The shaded region illustrates the threshold required for two generations (panel A). (c) Hybrid index, linkage disequilibrium (LD), LD block size (average distance in bp between SNPs with $r^2 > .2$), fixation index (F_{IS}) and nucleotide diversity (π) across the *P. glaucus*-*P. canadensis* hybrid zone. Hybrid index ranges from 0 (*P. glaucus*-like) to 1 (*P. canadensis*-like). Error bars represent 95% confidence intervals. Estimates for LD were created by randomly subsampling seven individuals (smallest n of all locations) from each location 100 times. Vertical dashed lines represent the estimated centre of the hybrid zone based on genetic clines [Colour figure can be viewed at wileyonlinelibrary.com]

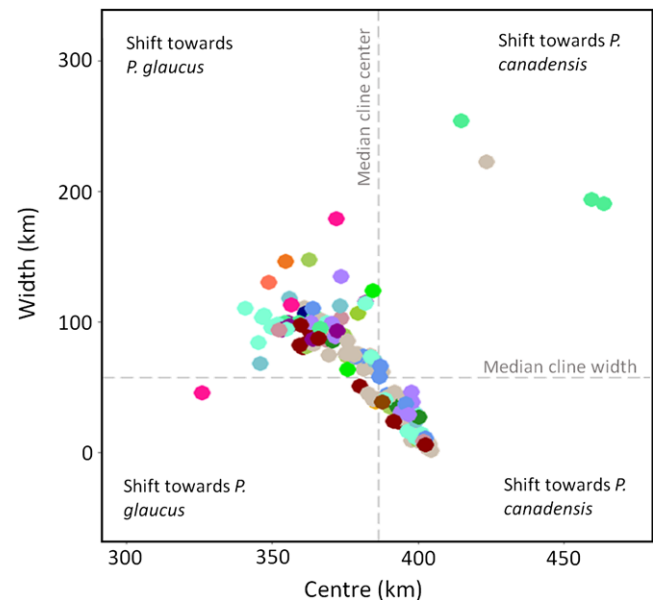


FIGURE 4 Maximum-likelihood estimates of cline centre and width for 230 (outlier) SNPs (circles). Dashed lines indicate estimated median cline centre (387 km) and width (57 km) based on all genetic markers. Colour indicates putative chromosome assignment (see Figure 2b) [Colour figure can be viewed at wileyonlinelibrary.com]

markers were also significantly wider (along with another locus on chromosome 23, they ranked among the top four widest of all clines; Figure 4).

Of the only 20% (47/230) of loci that were discordant, 98% (46/47) of these were significantly wider and 2% (1/47) narrower than the median estimate (Data S4). The majority of the significantly wider clines that were also noncoincident (24/28) were shifted to the south of the transect. The only locus found to be significantly narrower than the median cline was a RADseq locus located in a scaffold that did not map to a *B. mori* linkage group. However, it should be noted that given hybrids are rare and make up a small portion of our samples within the hybrid zone, we are only able to detect modest to large differences in cline width among clinal loci.

3.6 | Evaluating ecological factors maintaining divergence

The estimated centre of clinal loci was compared to geographic variation in length of the growing season. As expected, we found that the boundary of the growing degree-day (GDD) requirement to complete two generations coincided with the estimated centre of the hybrid zone. The cline centre of the genetic loci was 387 km compared with the cline centre for the GDD threshold which was at 405 km, putting this climate-mediated developmental threshold ~18 km north of the genetic cline centre and similar in cline width (to the genetic loci)—GDD “width” (variance in estimate based on ± 100 GDD) was 64 km and genetic cline width was ~57 km (Figures 3b and 4a). Contrastingly, the range boundaries of three host plants (*Liriodendron tulipifera*, *Populus tremuloides*, *Ptelea trifoliata*), for which differences in detoxification abilities exist between *P. glaucus* and *P. Canadensis*, did not coincide with the hybrid zone. The east-

west boundary where *Ptelea trifoliata* reaches its northern limit and *Populus tremuloides* reaches its southern limit is ~100+ km from the hybrid zone in Wisconsin (Figure 5c–e). Similarly, the northern range boundary of the Batesian mimicry model *Battus philenor* (for the female black morph of *P. glaucus*) does not coincide with the hybrid zone, as its most northern extent ends ~100 km south of the estimated centre of the hybrid zone (Figure 5b).

4 | DISCUSSION

4.1 | The genomic landscape of divergence

Placing patterns of divergence within a genomic and geographic context can provide a better understanding of how natural selection and neutral processes drive and maintain species boundaries. Our results, based on a reduced representation of genomic variation, showed that genetic divergence and differentiation between *P. glaucus* and *P.*

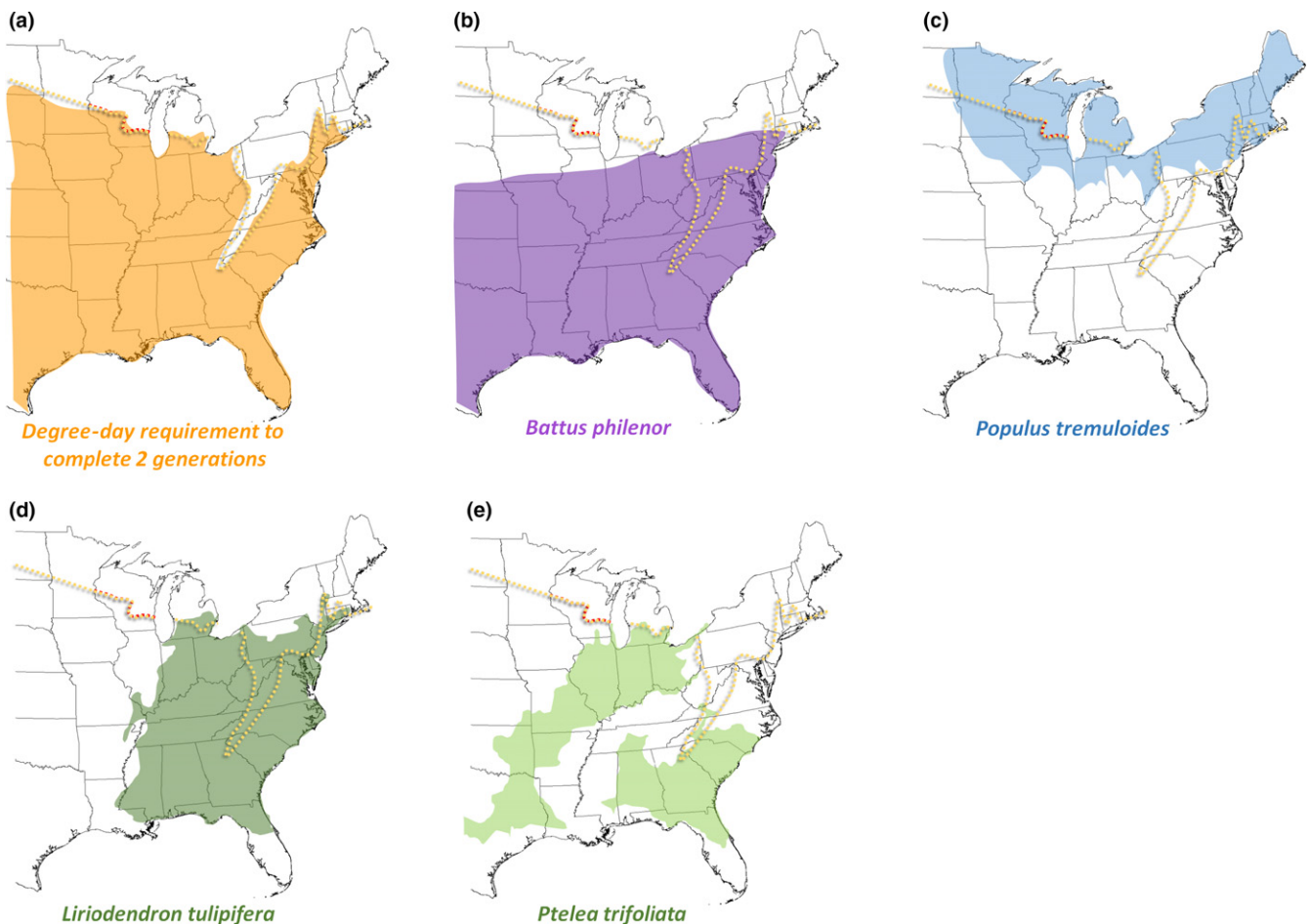


FIGURE 5 Approximate distributions of ecological factors associated with traits known to be ecological divergent between *P. glaucus* and *P. canadensis* in relation to the estimated location of the hybrid zone. The yellow dotted line indicates the estimated location of hybrid zone and the red portion highlights the region of the hybrid zone sampled in this study. (a) The estimated spatial distribution of growing degree-days (GDD) required to complete two generations associated with facultative diapause; adapted from Scriber, 2011 a threshold of $>1,300$ GDD. (b) The distribution of the Batesian model (*Battus Philenor*) associated with the female black morph in *P. glaucus*; adapted from previously published distributions of *Battus Philenor* (Fordyce and Nice 2003). (c,d) Estimated distribution of two hosts plants (*Populus tremuloides* and *Liriodendron tulipifera*) associated with differential detoxification abilities between *P. glaucus* and *P. canadensis* (Li, Baudry, Berenbaum, & Schuler, 2004; Lindroth et al., 1988); adapted from range maps published in Little (1971, 1976) [Colour figure can be viewed at wileyonlinelibrary.com]

canadensis is widespread across the genome and is highly heterogeneous in distribution and size within and across chromosomes. A subset of chromosomes (Z chromosome and the autosomes 4, 10, 12 and 19) showed a disproportionate amount of these differentiated and divergent regions. This pattern of loci putatively under divergent selection distributed across multiple linkage groups, with a sex chromosome harbouring what appears to be a disproportionate amount of these divergent regions, is similar to what has been reported in a number of other systems including other butterflies such as *Heliconius* (Martin et al., 2013), mosquitoes (Fontaine et al., 2015), flies (Tao, Chen, Hartl, & Laurie, 2003), mice (Teeter et al., 2008) and birds (Ellegren et al., 2012). This provides further evidence that regions of low recombination, and especially sex chromosomes, may include features that could play a disproportionate role in maintaining or facilitating divergence in the speciation process, as previously suggested (Fontaine et al., 2015; Mitchell et al., 2015).

Multiple factors can produce variable patterns of differentiation across the genome, and interpreting the causes of these patterns depends in part on whether this hybrid zone is a primary contact zone (divergence with gene flow) or result of secondary contact (divergence in allopatry), which is not known in this system. However, there are a number of reasons why this hybrid zone is more likely the result of secondary contact. As with many hybrid zones in North America, it is unlikely that this hybrid zone survived climatic fluctuations over the last 0.6 million years of divergence between these two species (Cong et al., 2015). More likely, this hybrid zone formed as a result of secondary contact following the last glaciation (Barton & Hewitt, 1985). While not conclusive, the skewed ("F-shape") F_{ST} distribution we observe is similar to the pattern observed by Nosil et al. (2012) for allopatric *Timema cristinae* stick insect populations, in contrast to the "L distribution" observed in parapatric *T. cristinae* populations experiencing gene flow. Some models predict that allopatric isolation should result in a more genomewide distribution of genome divergence (Noor, Grams, Bertucci, & Reiland, 2001; Nosil, Funk, & Ortiz-Barrientos, 2009; Rieseberg, 2001; Yeaman & Whitlock, 2011), in contrast to divergence with gene flow which is predicted to result in a more clustered distribution of divergent loci. Our finding that multiple outlier loci are distributed across all but five chromosomes is thus more consistent with a process of allopatric divergence. However, discordance in the estimated divergence times of Z-linked markers suggests that gene flow has been ongoing during the speciation process (Putnam et al., 2007).

Regardless of the geographic origin of divergence, both selection and neutral processes can produce the heterogeneity in differentiation we observe across these genomes. In the case of the Z chromosome, the nearly chromosome-wide high level of differentiation is likely, in part, the result of the increased genetic drift resulting from the reduced recombination rate and lower effective population size of this chromosome (Oyler-McCance, Cornman, Jones, & Fike, 2015). However, there are a few regions that appear to be largely undifferentiated on the Z chromosome. Whether these regions reflect further variation in recombination within this chromosome, variable gene flow, or are regions that are no longer in synteny (i.e.,

are actually located on a different chromosome) is not known. In the case of autosomes, it seems unlikely that variation in recombination alone would explain a pattern that consists of a large number of differentiated regions, distributed widely both within and across nearly all chromosomes. However, there are a number of relatively large differentiated regions that may be the result of suppressed recombination, illustrated by the introgression of what appear to be large genomic blocks (Figure 1c; Fig. S8). The importance and potential of inversions in facilitating divergence during speciation has been well documented in a number of systems (Hoffmann & Rieseberg, 2008; Kirkpatrick & Barton, 2006; Love et al., 2016; Noor et al., 2001; Trickett & Butlin, 1994); however, their role in facilitating divergence in this system remains unknown. A recombination and linkage map for *P. glaucus* would help in identifying possible inversions and more generally the role recombination is playing in producing variable patterns of differentiation in this system.

Alternatively, these patterns could reflect variable gene flow. There is now some controversy over whether patterns of heterogeneous differentiation across the genome, assessed by relative measures of differentiation such as F_{ST} , are caused by variable gene flow or processes that reduce variation in one population (e.g., selective sweeps or variation in recombination rates; Cruickshank & Hahn, 2014; Noor & Bennett, 2009). Thus, caution should be taken when using these measures alone. However, Cong et al. (2015) previously identified "divergence hot spots" (genes with sequence divergence at both the DNA and protein level that is greater between than within species; regions with elevated absolute divergence), as well as regions under positive selection and found little overlap (11%) between these regions. Similarly, only 10% (36/368) of our ddRAD-seq SNPs (1 SNP per 1 kb) under divergent selection were closest to a gene identified as being under positive selection by Cong et al. (2015). Although these results are based on a reduced representation of the genetic variation across the genome, this lack of overlap between loci under divergent and positive selection suggests that positive selection is likely only playing a minor role in the variation in differentiation we observe. For variable introgression to be driving this pattern, there would have to be ongoing gene flow. That all (F_{ST}) outliers in allopatric populations had relatively narrow clines in the hybrid zone and were largely coincident and concordant suggests the presence of strong reproductive barriers and very little (later-generation) hybridization and gene flow. Thus, the observed variation in differentiation across the genome may reflect heterogeneous selection more than heterogeneous gene flow, although both may be playing a factor.

4.2 | Evidence of ecological divergence throughout the genome

A major question remaining in speciation biology is how genomic architecture influences divergence (Nosil & Feder, 2012). Resolving this question requires an understanding of how genes involved in ecological adaptation or reproductive isolation are arrayed within the genome. We find that candidate genes for divergent ecological traits

(seasonal adaptations, host plant detoxification and predator defence gland) in *P. glaucus*–*P. canadensis* system are distributed throughout the genome (across at least seven chromosomes), somewhat clumped and almost all associated with regions of elevated differentiation. Loci found to be significantly differentiated in allopatric populations (F_{ST} outliers) that also exhibit steep clines across the hybrid zone are strong candidates for reproductive isolation (Harrison & Larson, 2016). Interestingly, we find that nearly all ddRADseq outlier loci, spanning multiple chromosomes, exhibit steep clines across the hybrid zone. That the majority of these loci were also highly coincident and concordant suggests that reproductive isolating factors are spread throughout the genome and/or selection against hybrids is very strong.

While loci involved in reproductive isolation may be distributed across the genome, five chromosomes seem to play a disproportionate role. Chromosomes 1 (Z), 4, 10, 12 and 19 harbour a significantly greater number of loci potentially evolving under divergent selection as well as containing divergence hot spots, and all exhibited steep clinal variation and (with the exception of chromosome 12) contained candidate genes for ecologically divergent traits. The Z chromosome contains near chromosome-wide elevated differentiation (nearly four times the background of autosomes) and is enriched with genes putatively involved in all traits that are known to be ecologically divergent between these species, such as diapause (*cycle* and *period*), thermal tolerance (*Ldh*) and a suppressor gene involved in female mimetic dimorphism (linked to *6-Pgd*), all of which are in, or are near (<30 kb from) highly differentiated and divergent regions. Reduced recombination of this chromosome is likely facilitating some of this divergence, as a region as large as ~700 kb has been estimated to be in linkage on the Z chromosome (Winter & Porter, 2010). Regions of the genome experiencing low recombination such as near centromeres (Ellegren et al., 2012), sites of inversions (Twyford & Friedman, 2015) and often sex chromosomes (Sæther et al., 2007; Teeter et al., 2008) have been shown to accumulate genes involved in adaptive divergence. Both *period* and *cycle* of the Z chromosome are central to the circadian clock system that regulates the timing of eclosion (Blanchardon et al., 2001; Zhu et al., 2008) for which Cong et al. (2015) identified clusters of mutations on the surface of these proteins that likely modulate differences in diapause between these species (obligate vs. facultative diapause). Interestingly, while *period* was identified as a divergence hot spot by Cong et al. (2015), we did not identify it as being under divergent selection (based on a number of RADseq loci within 1 kb of this gene). While only 11% of genes under positive selection were also divergence hot spots (as determined in Cong et al., 2015), of the few that were, four were found on the Z chromosome, and include the gene *period*, suggesting that positive selection, not divergent selection, may be driving divergence in *period*. Thus, positive selection on traits involved in seasonal adaptation(s), including *period*, may be facilitating divergence on this chromosome (Dean, Harrison, Wright, Zimmer, & Mank, 2015).

Four autosomes (4, 10, 12 and 19) also stand out with their enrichment of loci under potential divergent selection (with steep clinal variation) and divergence hot spot genes. Chromosome 10 contains one candidate gene for host plant detoxification (NADPH cytochrome P450 reductase; 95 kb from a RADseq locus under divergent selection) and one associated with development (ecdysone-inducible protein) that clustered closely in the centre of this chromosome within what appears to be a large region of differentiation. Surprisingly, while chromosome four contains the gene *timeless*, a gene identified as a divergence hot spot (Cong et al., 2015) involved in seasonal adaptations (diapause induction; Tauber et al., 2007), it was not identified as an F_{ST} outlier in our study. In the fruit fly (*Drosophila melanogaster*) and the monarch butterfly (*Danaus plexippus*), this gene is involved in post-translational modifications of the core clock proteins and regulates differences in the incidence of diapause in response to changes in light and temperature (Reppert, 2007; Tauber et al., 2007). Given that *P. canadensis* has an obligatory diapause and no longer uses photoperiodic cues, the lower-than-expected (moderate) differentiation we observed for a SNP in this gene may reflect relaxed selection on this trait in *P. canadensis* (Ryan, Valella, Thivierge, Aardema, & Scriber, 2016).

Interestingly, a number of candidate genes for farnesyl pyrophosphate synthase (FPPS) homologues hypothesized to synthesize terpenes for a predator defence gland (osmeterium) unique to Papilionidae (Cong et al., 2015) were found on chromosome 19. While the chemistry of this gland is known to be affected by genetic background and varies substantially through larval development, it is not known to what extent the chemical composition of this gland differs between *P. glaucus* and *P. canadensis* (Frankfater, Tellez, & Slattery, 2009). We identified at least two FPPS genes in close proximity (<2 kb) to loci under divergent selection as well as another FPPS gene identified as a divergence hot spot by Cong et al. (2015). Given the large geographic distribution of these species, it is possible that variation in predator communities could be driving divergence in this trait, but this remains to be tested.

Of the few RADseq loci identified as being under balancing selection, three were in close proximity to genes involved in traits that have been implicated as being under divergent or balancing selection in other systems. Two of these SNPs were located on chromosome 16. This includes a SNP within 2 kb of an opsin gene (*opsin-1*). A recent study of a North American brush-footed butterfly (*Limenitis arthemis*) found clinal variation in a majority of opsin genes, with many also under balancing selection (Frentiu et al., 2015). Interestingly, the variation in the long-wavelength opsin genes of *Limenitis arthemis* did not affect spectral sensitivity and instead is speculated to play a role in unknown adaptive functions such as mediating responses to temperature or photoperiod. The other SNP on chromosome 16 identified as a candidate for balancing selection was closets (~10 kb) to a cytochrome P450 gene (*CYP6k1*). Hybrids of *P. glaucus* and *P. canadensis* exhibit detoxification abilities of both parents (Mercader et al., 2009). Whether overdominance is driving an increase in diversity of this cytochrome P450 gene (or others) by increasing fitness, for example by expanding an individual's host

breadth, is not known and warrants further investigation. Chromosome 19 contained a SNP identified as a candidate for balancing selection within 1 kb of a gene that is closest ortholog is the gene *EbpIII* that is specifically expressed in the ejaculatory bulb and seminal fluid of *D. melanogaster*. Balancing selection has been proposed as an explanation for high allelic diversity in a number of seminal fluid accessory proteins (Acps), proteins that influence reproductive traits such as sperm transfer, sperm storage, female receptivity, ovulation and oogenesis (Chapman, 2001). While these are an interesting set of candidate genes for balancing selection in this system, future studies involving more fine-scale genomic resolution and functional assays will help to evaluate these candidates more thoroughly.

It should be noted that most of the distances (in bp) between these candidate genes and our RADseq loci are generally <1 cM based on our estimate of a physical-to-map recombination distance of ~322 kb/cM (with the genome being 376 Mb and the total map length 1167 cM). This value is similar to the 180 kb/cM estimated for *H. melpomene* (Jiggins et al., 2005). More importantly, these distances were also usually far less than our estimate of the average size of regions in linkage disequilibrium (i.e., those with $r^2 > .2$), particularly those in the hybrid zone that were estimated to be ~70–80 kb. Moreover, given that regions under selection would likely be at the high end of such estimates, these regions may even be greater than 100 kb in size.

4.3 | Climate a salient factor in hybrid zone maintenance

While there are a number of ecologically divergent traits between the two species, climate seems to be a salient factor in maintaining genomic divergence. The first and most prominent reason is that nearly all divergent regions of the genome show a cline pattern that covaries with a climatic gradient—the estimated centre of the hybrid zone fell just 20 km south of the estimated location of the growing degree-day (1,300 GDD) threshold to complete two generations; all of the narrowest clines (clines < 5 km in width; those putatively under the strongest selection) are even closer to (~1 km south of) this GDD threshold. This estimate is in agreement with previous analyses based on a few species-diagnostic allozymes (Hagen, 1990; Scriber et al., 2003). Furthermore, while the genetic basis of diapause induction and termination remains elusive (Denlinger, 2002), some of the genes most consistently implicated as playing a role in regulating diapause or eclosion phenology—*cycle*, *period*, *timeless*, *clock*—all mapped to chromosomes identified as harbouring a disproportionate amount of divergence. Third, candidate genes for other ecologically divergent traits (i.e., host plant detoxification and Batesian mimicry; Mercader et al., 2009; Scriber, Hagen, & Lederhouse, 1996; Zandt Brower, 1957) do not covary with the ecological gradients that would be associated with these traits, but instead largely covary with a climatic gradient. In the case of three of the host plants associated with the greatest differences in detoxification abilities

between *P. glaucus* and *P. canadensis*—*Populus tremuloides*, *Ptelea trifoliata* and *Liriodendron tulipifera*—the range boundary of these plants falls hundreds of kilometres south of the hybrid zone. Similarly, the northern range boundary of the model *Battus philenor* is located hundreds of kilometres south of the hybrid zone (at least in Wisconsin). Thus, it seems unlikely that the distribution of host plants or a model for Batesian mimicry is playing a major role in determining or maintaining the location of this hybrid zone (at least in Wisconsin). Last, there is evidence of steep clinal variation in genetic markers and a number of physiological, phenotypic and behavioural traits largely coincide with this same climate-mediated developmental threshold in other portions of this hybrid zone (e.g., in Michigan and Vermont; Scriber, 2011).

Examples of climatic gradients driving clines in genetic and phenotypic variation have been well documented across a breadth of taxa, particularly in plants and insects (Franks & Hoffmann, 2012; Knibb, Oakeshott, & Gibson, 1981). Specifically, geographic gradients in the length of the growing season that result in changes in voltinism (number of generations in a growing season), the so-called “voltinism transition zones,” often coincide with genetic and phenotypic changes (Bradford & Roff, 1995; Ikten, Skoda, Hunt, Molina-Ochoa, & Foster, 2011; Mallet, Wynne, & Thomas, 2011). Other examples include adaptation to variation in the length of the growing season that has resulted in clinal variation in the body size of *Pteronemobius* spp. of cricket (Roff, 1980), allele frequencies of a circadian gene in the European corn borer (*Ostrinia nubilalis*; Levy, Kozak, Wadsworth, Coates, & Dopman, 2015) and the critical photoperiod of the pitcher plant mosquito (*Wyeomyia smithii*; Bradshaw, 1976). These examples demonstrate the powerful selective force that variation in the length of the growing season can impose, even in the face of gene flow, given that many of these clines exist within these species' ranges. The *P. glaucus* and *P. canadensis* hybrid zone also coincides with a voltinism transition zone, from bivoltine (facultative diapause of *P. glaucus*) to univoltine (obligate diapause of *P. canadensis*) and it has been hypothesized that strong climate-mediated divergent selection is acting to restrict gene flow, as genes associated with obligate and facultative diapause would be selected against south and north of the hybrid zone, respectively (Scriber, 2011). Our results lend further support for this hypothesis. Whether the obligate diapause phenotype (of *P. canadensis*) is an adaptation that predates divergence of these species, or came about post divergence is still not known.

While most clines were largely coincident and concordant, the majority of discordant clines exhibited a pattern of widening (“leaning”) in the direction of *P. glaucus*. This pattern can be an indication of hybrid zone movement (Moran, 1981; i.e., reflect a recent northward shift) or reflect biased gene flow towards *P. glaucus* that could be due to asymmetrical selective pressures or assortative mating (Buggs, 2007). Biased introgression into *P. glaucus* would agree with previous estimates of greater gene flow into *P. glaucus* (Cong et al., 2015 and Zhang, Kunte, & Kronforst, 2013). Although introgression appears biased in the direction of *P. glaucus* (southward), the most

discordant of all clines were those located on chromosome 23 that were shifted north of the hybrid zone. One explanation of such a pattern could be adaptive introgression (Harrison & Larson, 2016). That is, *P. glaucus* alleles are introgressing into *P. canadensis* because they confer some fitness advantage in both genetic backgrounds. Interestingly, two of the SNPs identified as being under divergent selection on this chromosome were within 1 kb of the gene *couch potato*; an amino acid polymorphism in the *couch potato* gene has been implicated in climatic adaptation in *Drosophila melanogaster* by altering reproductive diapause (Schmidt et al., 2008). Another SNP was closest to a juvenile hormone epoxide hydrolase that is implicated in insect development by regulating juvenile hormone (Touhara & Prestwich, 1993). Whether these genes are acting to restrict the introgression of favourable alleles on this chromosome is not known.

4.4 | Potential for “coupling”

Although climate appears to be a salient factor in the maintenance of this hybrid zone, there are multiple lines of evidence that suggest intrinsic barriers might also be present in this system. One possibility is that the *P. glaucus*–*P. canadensis* hybrid zone is a tension zone that has become associated with a climatic gradient. In support of this idea, we observe a relatively high level of coincidence (81%) and concordance (80%) among both clinal ddRADseq loci that also coincide with a peak in *LD*. These patterns along with what is estimated as a relatively moderate portion of the genome (1.3%) being under divergent selection have been proposed as an indication of the existence of intrinsic barriers (Bierne et al., 2011). Thus, it is possible that endogenous barriers have become coupled with an exogenous (climate-mediated) cline to impede hybrid zone movement, due to either lowered population density across this ecotone (Barton & Hewitt, 1985) or strong extrinsic selection. Unfortunately, as is the case with many hybrid zones, we do not have robust estimates of population density across the hybrid zone due to the difficulty in obtaining such data. Thus, we cannot completely rule out this possibility. Such patterns do not require invoking endogenous barriers, however, as a geographic selection gradient model can also produce clinal patterns indistinguishable from a tension zone model. Further, the large amount of divergence across the genome may be explained by “genomic hitchhiking” as (extrinsic) selection on many loci, such as that on the candidate genes for ecological adaptations we observe, can result in an increase in genomewide divergence, even in regions unlinked to known adaptive loci (Feder, Gejji, Yeaman, & Nosil, 2012).

Isolation during allopatry can lead to divergence in allelic variation, that when admixed, such as during secondary contact, results in genotypes that are less fit due to the creation of combinations of alleles that were previously hidden from selection (e.g., Dobzhansky–Muller incompatibilities; Coyne & Orr, 2004). If the *P. glaucus* and *P. canadensis* hybrid zone is a secondary contact zone as has been proposed (Scriber et al., 1991) and is partly supported by results from this study, it is plausible that neutral process and selection have

resulted in the build-up of at least some intrinsic incompatibilities within these genomes. Intriguingly, 40 years of research in this system have produced limited evidence of intrinsic incompatibilities, with reductions in fitness often negligible. Tens of thousands of laboratory pairings from numerous experiments have only observed hybrid offspring exhibiting slightly reduced egg viability (~9%), low-to-moderate pupal mortality (~28%) and a reduction or even increase in fecundity, depending on direction of cross (Scriber, Lederhouse, & Dowell, 1995); there is even some evidence of heterosis, translating into faster developmental rates and larger pupal size (Scriber et al., 2003). It is possible that the fitness consequences of hybridization do not manifest until after multiple bouts of hybridization—that is, in later-generation hybrids, such as has been observed in a number of hybridizing species (Bierne, Bonhomme, Boudry, Szulkin, & David, 2006; Pritchard & Edmands, 2013). However, given that we found few early-generation hybrids, this suggests that most hybrids are not surviving long enough for multiple bouts of hybridization to occur. Alternatively, it may be that we have been focusing on the wrong phenotype. It has been observed that hybrids of these species exhibit a dramatically altered phenology (emerge weeks to months prior to, or after, parentals) as well as aberrant diapause phenotypes (i.e., failure to diapause even under short-day conditions; Ryan et al., 2016). Presumably, these aberrant phenotypes could be less fit regardless of environmental conditions and explain strong selection against admixed genomic backgrounds. Future studies that evaluate the fitness of hybrids and parentals in the field, reared in sympatric (hybrid zone) and allopatric (parental) populations, would help to identify possible intrinsic incompatibilities in this system.

Other nonenvironmental barriers, such as prezygotic barriers (i.e., mate preferences) could also be preventing hybridization and aiding in the maintenance of this hybrid zone. There is some evidence for mate preference, with what appears to be a preference by males of both species for *P. glaucus* females (Deering & Scriber, 2002). However, these preference assays have assumed that males are making the choice, which has never been confirmed and there is actually evidence that females in this system are making the choice when mating (Krebs, 1988). Further, hybrids based on our mtDNA and Z-linked species-diagnostic markers have a relatively even mix of both *P. glaucus* and *P. canadensis* mothers, suggesting that if there is a preference, it does not appear to be substantial.

The finding that candidate genes putatively associated with host plant detoxification overlap more with a climate-mediated developmental threshold than the geographic distribution of differentially detoxified host plants is intriguing. The simplest explanation for this is that the candidate genes associated with these outlier loci do not actually result in phenotypic divergence in these traits (e.g., differences in detoxification ability). Another possibility is that they are merely linked with other genes under exogenous (climate-mediated) or endogenous selection. This may very well be the case, as the specific CYP450 genes we identified as being under divergent selection are not those implicated in producing the greatest difference in furanocoumarin metabolism. In fact, the RADseq loci closest to CYP6B4 genes (those shown to have the greatest level of

furancoumarin metabolism) exhibited very low levels of differentiation. It is also possible that the CYP6 and CYP4 genes that were implicated as being under divergent selection are associated with detoxification abilities for host plants not explored in this study (*P. glaucus* can feed on plants from >34 families), or may only play a minor role. An another explanation is that clines for ecological traits not involved in adaptations to climate (e.g., CYP6B and CYP64 genes) have become associated with a climatic gradient as a result of historical cline movement. There is some evidence that tension zones can become “attracted” and move together after contact (Mallet & Barton, 1989; Mallet & Turner, 1998). Recently, Rosser, Das-mahapatra, and Mallet (2014) have shown that moving clines can also become trapped when they encounter stationary clines maintained by exogenous selection along an ecotone due to an increase in per locus selection resulting from disequilibrium among clines. Future studies that explore how clinal patterns for a set of genome-wide SNPs vary across the multiple independent portions of this hybrid zone where the distribution of host plants (as well as other ecological gradients) varies should shed light on this question as well as the relative role of endogenous and exogenous barriers in this system (Harrison & Larson, 2016).

This study demonstrates the power of assessing how genomic differentiation varies along a gradient of allopatry (pure parental populations) to sympatry (hybrid zone) to reveal insights into the genetic architecture of reproductive isolation. With this approach, we show that clinal variation in a climate-mediated developmental threshold can act to maintaining divergence between the genomes of hybridizing species. We also find evidence that previously unrecognized endogenous barriers may be contributing to reproductive isolation in this system, providing some support for the coupling hypothesis that hybrid zones with steep clinal variation associated with an ecotone are the result of a coupling of endogenous and exogenous barriers. Last, similar to what is now a growing number of systems, we find evidence that the Z chromosome may be playing a disproportionate role in the process of speciation.

ACKNOWLEDGEMENTS

Butterfly collections were undertaken by Sean F. Ryan, Mark Scriber, Jason Dzurisin and Sarah Richman. Many thanks to Benjamin Clifford, Meredith Doellman and Jacqueline Lopez and Melissa Stephens in the Notre Dame Genomics & Bioinformatics Core Facility for their advice in various aspects of experimental design, NGS data collection and analysis. This work was sponsored by NSF grants DEB-0918879 to JMS and DEB-0919147 to JJH and a grant to SFR from the Environmental Change Initiative at the University of Notre Dame.

DATA ACCESSIBILITY

Demultiplexed ddRAD sequence data were deposited in NCBI's SRA archives under BioProject ID PRJNA388284. Scripts and associated files for processing and analysing all data in this manuscript were deposited in DRYAD (<https://doi.org/10.5061/dryad.t9221>).

AUTHOR CONTRIBUTIONS

SFR, JMS, MCF, MEP and JJH designed this research. SFR did field collections and all laboratory work. SFR performed all analysis, with help from MCF and SO. SO and MCF developed software and scripts for a portion of the genomic analyses. SFR, MCF, JJH, JMS and MEP were involved in writing and editing this manuscript.

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410.
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113–148.
- Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, *263*, 1619–1626.
- Bierne, N., Bonhomme, F., Boudry, P., Szulkin, M., & David, P. (2006). Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*, 1253–1260.
- Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, *20*, 2044–2072.
- Blanchardon, E., Grima, B., Klarsfeld, A., Chélot, E., Hardin, P. E., Prémat, T., & Rouyer, F. (2001). Defining the role of *Drosophila* lateral neurons in the control of circadian rhythms in motor activity and eclosion by targeted genetic ablation and PERIOD protein overexpression. *European Journal of Neuroscience*, *13*, 871–888.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). TRIMMOMATIC: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120.
- Bradford, M. J., & Roff, D. A. (1995). Genetic and phenotypic sources of life history variation along a cline in voltinism in the cricket *Allonemobius socius*. *Oecologia*, *103*, 319–326.
- Bradshaw, W. E. (1976). Geography of photoperiodic response in diapausing mosquito. *Nature*, *262*, 384–386.
- Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity*, *99*, 301–312.
- Chapman, T. (2001). Seminal fluid-mediated fitness traits in *Drosophila*. *Heredity*, *87*, 511–521.
- Cong, Q., Borek, D., Otwinowski, Z., & Grishin, N. V. (2015). Tiger swallowtail genome reveals mechanisms for speciation and caterpillar chemical defense. *Cell Reports*, *10*, 910–919.
- Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution*, *43*, 362–381.
- Coyne, J. A., & Orr, A. H. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.
- 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*, 3133–3157.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Group, 1000 Genomes Project Analysis (2011). The variant call format and VCFTOOLS. *Bioinformatics*, *27*, 2156–2158.
- Dean, R., Harrison, P. W., Wright, A. E., Zimmer, F., & Mank, J. E. (2015). Positive selection underlies faster-Z evolution of gene expression in birds. *Molecular Biology and Evolution*, *32*, 2646–2656.
- Deering, M. D., & Scriber, M. J. (2002). Field bioassays show heterospecific mating preference asymmetry between hybridizing North American *Papilio* butterfly species (Lepidoptera: Papilionidae). *Journal of Ethology*, *20*, 25–33.

- Denlinger, D. L. (2002). Regulation of diapause. *Annual Review of Entomology*, 47, 93–122.
- Dent, E. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361.
- de Lafontaine, G., Prunier, J., Gérardi, S., & Bousquet, J. (2015). Tracking the progression of speciation: Variable patterns of introgression across the genome provide insights on the species delimitation between progenitor–derivative spruces (*Picea mariana* × *P. rubens*). *Molecular Ecology*, 24, 5229–5247.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43, 491–498.
- Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). hzar: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, 14, 652–663.
- de Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, 23, 2006–2019.
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ... Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491, 756–760.
- Endler, J. A. (1977). *Geographic variation, speciation, and clines*. Princeton, NJ: Princeton Univ. Press.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, 14, 2611–2620.
- Excoffier, L., & Lischer, H. E. L. (2010). ARLEQUIN suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567.
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 461–474.
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993.
- Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., ... Besansky, N. J. (2015). Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science (New York, NY)*, 347, 1258524.
- Frankfater, C., Tellez, M. R., & Slattery, M. (2009). The scent of alarm: Ontogenetic and genetic variation in the osmeterial gland chemistry of *Papilio glaucus* (Papilionidae) caterpillars. *Chemoecology*, 19, 81–96.
- Franks, S. J., & Hoffmann, A. A. (2012). Genetics of climate change adaptation. *Annual Review of Genetics*, 46, 185–208.
- Frentiu, F. D., Yuan, F., Savage, W. K., Bernard, G. D., Mullen, S. P., & Briscoe, A. D. (2015). Opsin clines in butterflies suggest novel roles for insect photopigments. *Molecular Biology and Evolution*, 32, 368–379.
- Gompert, Z., & Buerkle, C. A. (2010). INTROGRESS: A software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, 10, 378–384.
- 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, 2167–2181.
- Hagen, R. H. (1990). Population structure and host use in hybridizing subspecies of *Papilio glaucus* (Lepidoptera: Papilionidae). *Evolution*, 44, 1914–1930.
- Hagen, R. H., & Scriber, J. M. (1989). Sex-linked diapause, color, and allozyme loci in *Papilio glaucus*: linkage analysis and significance in a hybrid zone. *Journal of Heredity*, 80, 179–185.
- Hamilton, J. A., la Torre, A. R. D., & Aitken, S. N. (2014). Fine-scale environmental variation contributes to introgression in a three-species spruce hybrid complex. *Tree Genetics & Genomes*, 11, 817.
- Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, 25, 2454–2466.
- Hewitt, G. M. (1988). Hybrid zones—natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, 3, 158–167.
- Hill, W. G., & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, 33, 54–78.
- Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, 39, 21–42.
- Hung, C. F., Berenbaum, M. R., & Schuler, M. A. (1997). Isolation and characterization of CYP6B4, a furanocoumarin-inducible cytochrome P450 from a polyphagous caterpillar (Lepidoptera: papilionidae). *Insect Biochemistry and Molecular Biology*, 27, 377–385.
- Ikten, C., Skoda, S. R., Hunt, T. E., Molina-Ochoa, J., & Foster, J. E. (2011). Genetic variation and inheritance of diapause induction in two distinct voltine ecotypes of *Ostrinia nubilalis* (Lepidoptera: Crambidae). *Annals of the Entomological Society of America*, 104, 567–575.
- Jiggins, C. D., Mavarez, J., Beltrán, M., McMillan, W. O., Johnston, J. S., & Bermingham, E. (2005). A genetic linkage map of the mimetic butterfly *Heliconius melpomene*. *Genetics*, 171, 557–570.
- Jombart, T., & Ahmed, I. (2011). ADEGENET 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics (Oxford, England)*, 27, 3070–3071.
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, 173, 419–434.
- Knibb, W. R., Oakeshott, J. G., & Gibson, J. B. (1981). Chromosome inversion polymorphisms in *Drosophila melanogaster*. I. latitudinal clines and associations between inversions in Australasian populations. *Genetics*, 98, 833–847.
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15, 1179–1191.
- Krebs, R. A. (1988). The mating behavior of *Papilio glaucus* (Papilionidae). *Journal of Research on the Lepidoptera*, 26, 27–31.
- Kukal, O., Ayres, M. P., & Scriber, J. M. (1991). Cold tolerance of the pupae in relation to the distribution of swallowtail butterflies. *Canadian Journal of Zoology*, 69, 3028–3037.
- Kunte, K., Shea, C., Aardema, M. L., Scriber, J. M., Juenger, T. E., Gilbert, L. E., & Kronforst, M. R. (2011). Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. *PLoS Genetics*, 7, e1002274.
- Lehnert, M. S., Scriber, J. M., Gerard, P. D., & Emmel, T. C. (2012). The “Converse to Bergmann’s Rule” in tiger swallowtail butterflies: Boundaries of species and subspecies wing traits are independent of thermal and host-plant induction. *American Entomologist*, 58, 156.
- Levy, R. C., Kozak, G. M., Wadsworth, C. B., Coates, B. S., & Dopman, E. B. (2015). Explaining the sawtooth: Latitudinal periodicity in a circadian gene correlates with shifts in generation number. *Journal of Evolutionary Biology*, 28, 40–53.
- Li, X., Baudry, J., Berenbaum, M. R., & Schuler, M. A. (2004). Structural and functional divergence of insect CYP6B proteins: From specialist to generalist cytochrome P450. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 2939–2944.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25, 1754–1760.
- Li, W., Petersen, R. A., Schuler, M. A., & Berenbaum, M. R. (2002). CYP6B cytochrome p450 monooxygenases from *Papilio canadensis* and

- Papilio glaucus*: Potential contributions of sequence divergence to host plant associations. *Insect Molecular Biology*, 11, 543–551.
- Li, W., Schuler, M. A., & Berenbaum, M. R. (2003). Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: Specificity and substrate encounter rate. *Proceedings of the National Academy of Sciences of the United States of America*, 100 (Suppl 2), 14593–14598.
- Lindroth, R. L., Scriber, J. M., & Hsia, M. T. S. (1988). Chemical ecology of the tiger swallowtail: Mediation of host use by phenolic glycosides. *Ecology*, 69, 814–822.
- Little, E. L. J. (1971). *Atlas of United States trees, volume 1, conifers and important hardwoods*. U.S. Department of Agriculture Miscellaneous Publication, 9 p., 200 maps.
- Little, E. L. J. (1976). *Atlas of United States trees, volume 3, minor Western hardwoods*. U.S. Department of Agriculture Miscellaneous Publication, 13 p., 290 maps.
- Love, R. R., Steele, A. M., Coulibaly, M. B., Traore, S. F., Emrich, S. J., Fontaine, M. C., & Besansky, N. J. (2016). Chromosomal inversions and ecotypic differentiation in *Anopheles gambiae*: The perspective from whole-genome sequencing. *Molecular Ecology*, 25, 5889–5906.
- Luebke, H. J., Scriber, J. M., & Yandell, B. S. (1988). Use of multivariate discriminant analysis of male wing morphometrics to delineate a hybrid zone for *Papilio glaucus glaucus* and *P. g. canadensis* in Wisconsin. *American Midland Naturalist*, 119, 366–379.
- Malenfant, R. M., Coltman, D. W., & Davis, C. S. (2015). Design of a 9K illumina BeadChip for polar bears (*Ursus maritimus*) from RAD and transcriptome sequencing. *Molecular Ecology Resources*, 15, 587–600.
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20, 229–237.
- Mallet, J., & Barton, N. (1989). Inference from clines stabilized by frequency-dependent selection. *Genetics*, 122, 967–976.
- Mallet, J., & Turner, J. R. G. (1998). Biotic drift or the shifting balance: Did forest islands drive the diversity of warningly coloured butterflies? In Grant, P.R. (Ed.), *Evolution on Islands*. (pp. 262–280). Oxford, UK: Oxford University Press.
- Mallet, J., Wynne, I. R., & Thomas, C. D. (2011). Hybridisation and climate change: Brown argus butterflies in Britain (*Polyommatus subgenus Aricia*). *Insect Conservation and Diversity*, 4, 192–199.
- Marroni, F., Pinosio, S., Zaina, G., Fogolari, F., Felice, N., Cattonaro, F., & Morgante, M. (2011). Nucleotide diversity and linkage disequilibrium in *Populus nigra cinnamyl* alcohol dehydrogenase (CAD4) gene. *Tree Genetics & Genomes*, 7, 1011–1023.
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, 23, 1817–1828.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., ... DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297–1303.
- Mercader, R. J., Aardema, M. L., & Scriber, J. M. (2009). Hybridization leads to host-use divergence in a polyphagous butterfly sibling species pair. *Oecologia*, 158, 651–662.
- Mercader, R. J., & Scriber, J. M. (2008). Asymmetrical thermal constraints on the parapatric species boundaries of two widespread generalist butterflies. *Ecological Entomology*, 33, 537–545.
- Mitchell, S. N., Kakani, E. G., South, A., Howell, P. I., Waterhouse, R. M., & Catteruccia, F. (2015). Evolution of sexual traits influencing vectorial capacity in anopheline mosquitoes. *Science*, 347, 985–988.
- Moran, C. (1981). Genetic demarcation of geographical distribution by hybrid zones. *Ecological Society of Australia*, 11, 67–73.
- Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, 76, 5269–5273.
- Noor, M. A. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103, 439–444.
- Noor, M. A., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 12084–12088.
- Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society B*, 367, 332–342.
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18, 375–402.
- Nosil, P., Gompert, Z., Farkas, T. E., Comeault, A. A., Feder, J. L., Buerkle, C. A., & Parchman, T. L. (2012). Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B*, 279, 5058–5065.
- Oyler-McCance, S. J., Cornman, R. S., Jones, K. L., & Fike, J. A. (2015). Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. *Heredity*, 115, 452–459.
- 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>
- Presgraves, D. C. (2008). Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics*, 24, 336–343.
- Prism Climate Group. (2014). Oregon State University, <http://prism.oregonstate.edu>, created 4 Feb 2014.
- Pritchard, V. L., & Edmands, S. (2013). The genomic trajectory of hybrid swarms: Outcomes of repeated crosses between populations of *Tigriopus californicus*. *Evolution*, 67, 774–791.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Putnam, A. S., Scriber, J. M., & Andolfatto, P. (2007). Discordant divergence times among Z-chromosome regions between two ecologically distinct swallowtail butterfly species. *Evolution*, 61, 912–927.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reppert, S. M. (2007). The ancestral circadian clock of monarch butterflies: Role in time-compensated sun compass orientation. *Cold Spring Harbor Symposia on Quantitative Biology*, 72, 113–118.
- Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, 16, 351–358.
- Ritland, D. B., & Scriber, J. M. (1985). Larval developmental rates of three putative subspecies of tiger swallowtail butterflies, *Papilio glaucus*, and their hybrids in relation to temperature. *Oecologia*, 65, 185–193.
- Roff, D. (1980). Optimizing development time in a seasonal environment: The “ups and downs” of clinal variation. *Oecologia*, 45, 202–208.
- Rosser, N., Dasmahapatra, K. K., & Mallet, J. (2014). Stable *Heliconius* butterfly hybrid zones are correlated with a local rainfall peak at the edge of the Amazon basin. *Evolution; International Journal of Organic Evolution*, 68, 3470–3484.
- Ryan, S. F., Valella, P., Thivierge, G., Aardema, M. L., & Scriber, J. M. (2016). The role of latitudinal, genetic and temperature variation in the induction of diapause of *Papilio glaucus* (Lepidoptera: Papilionidae). *Insect Science*, <https://doi.org/10.1111/1744-7917.12423>
- Sæther, S. A., Sætre, G.-P., Borge, T., Wiley, C., Svedin, N., Andersson, G., ... Qvarnström, A. (2007). Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science*, 318, 95–97.
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual*. Cold Springs Harbor, New York: Cold Spring Harbor Laboratory Press.
- Schmidt, P. S., Zhu, C.-T., Das, J., et al. (2008). An amino acid polymorphism in the couch potato gene forms the basis for climatic

- adaptation in *Drosophila melanogaster*. *Proceedings of the National Academies of Sciences*, 105, 16207–16211.
- Scriber, J. M. (1990). Interaction of introgression from *Papilio glaucus canadensis* and diapause in producing "spring form" Eastern tiger swallowtail butterflies, *P. glaucus*. *Great Lakes Entomologist*, 23, 127–138.
- Scriber, J. M. (2002). Evolution of insect-plant relationships: Chemical constraints, coadaptation, and concordance of insect/plant traits. *Entomologia Experimentalis et Applicata*, 104, 217–235.
- Scriber, J. M. (2011). Impacts of climate warming on hybrid zone movement: Geographically diffuse and biologically porous "species borders". *Insect Science*, 18, 121–159.
- Scriber, J. M., Elliot, B., Maher, E., McGuire, M., & Niblack, M. (2014). Adaptations to "Thermal Time" constraints in *Papilio*: Latitudinal and local size clines differ in response to regional climate change. *Insects*, 5, 199–226.
- Scriber, J. M., Giebink, B. L., & Snider, D. (1991). Reciprocal latitudinal clines in oviposition behavior of *Papilio glaucus* and *P. canadensis* across the great lakes hybrid zone: Possible sex-linkage of oviposition preferences. *Oecologia*, 87, 360–368.
- Scriber, J. M., Hagen, R. H., & Lederhouse, R. C. (1996). Genetics of mimicry in the tiger swallowtail butterflies, *Papilio glaucus* and *P. canadensis* (Lepidoptera: Papilionidae). *Evolution*, 50, 222–236.
- Scriber, J. M., & Lederhouse, R. C. (1992). The thermal environment as a resource dictating geographic patterns of feeding specialization of insect herbivores. *Resource distribution and animal-plant interactions* (pp. 429–466). New York: Academic Press.
- Scriber, J. M., Lederhouse, R. C., & Dowell, R. V. (1995). Hybridization studies with North American swallowtails. *Swallowtail butterflies: Their ecology and evolutionary biology* (pp. 367–391). Chicago, Illinois, USA: University of Chicago Press.
- Scriber, J. M., Lindroth, R. L., & Nitao, J. (1989). Differential toxicity of a phenolic glycoside from quaking aspen to *Papilio glaucus* butterfly subspecies, hybrids and backcrosses. *Oecologia*, 81, 186–191.
- Scriber, J. M., Ording, G. J., & Mercader, R. J. (2008). Hybrid introgression and parapatric speciation in a hybrid zone. *Specialization, speciation, and radiation: The evolutionary biology of herbivorous insects* (pp. 68–87). Berkeley: Univ. California Press.
- Scriber, J. M., Stump, A., & Deering, M. (2003). *Ecology and evolution taking flight: Butterflies as model study systems*. Chicago, Illinois, USA: University of Chicago Press.
- Shiryev, S. A., Papadopoulos, J. S., Schäffer, A. A., & Agarwala, R. (2007). Improved BLAST searches using longer words for protein seeding. *Bioinformatics (Oxford, England)*, 23, 2949–2951.
- Sperling, F. A. (1993). Heredity - Abstract of article: Mitochondrial DNA variation and Haldane's rule in the *Papilio glaucus* and *P. troilus* species groups. *Heredity*, 71, 227–233.
- Szymura, J. M., & Barton, N. H. (1986). Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina orientalis* and *B. variegata*, Near Cracow in Southern Poland. *Evolution*, 40, 1141–1159.
- Szymura, J. M., & Barton, N. H. (1991). The genetic structure of the hybrid zone between the fire-bellied toads *Bombina orientalis* and *B. variegata*: Comparisons between transects and between loci. *Evolution*, 45, 237–261.
- Tao, Y., Chen, S., Hartl, D. L., & Laurie, C. C. (2003). Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. *Genetics*, 164, 1383–1397.
- Tauber, E., Zordan, M., Sandrelli, F., Pegoraro, M., Osterwalder, N., Breda, C., ... Costa, R. (2007). Natural selection favors a newly derived timeless allele in *Drosophila melanogaster*. *Science*, 316, 1895–1898.
- Taylor, S. A., Larson, E. L., & Harrison, R. G. (2015). Hybrid zones: Windows on climate change. *Trends in Ecology & Evolution*, 30, 398–406.
- Taylor, S. A., White, T. A., Hochachka, W. M., Ferretti, V., Curry, R. L., & Lovette, I. (2014). Climate-mediated movement of an avian hybrid zone. *Current Biology*, 24(6), 671–676. <https://doi.org/10.1016/j.cub.2014.01.069>
- Teeter, K. C., Payseur, B. A., Harris, L. W., Bakewell, M. A., Thibodeau, L. M., O'Brien, J. E., ... Tucker, P. K. (2008). Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, 18, 67–76.
- Teeter, K. C., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., Nachman, M. W., & Tucker, P. K. (2010). The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution; International Journal of Organic Evolution*, 64, 472–485.
- Touhara, K., & Prestwich, G. D. (1993). Juvenile hormone epoxide hydrolase. Photoaffinity labeling, purification, and characterization from tobacco hornworm eggs. *Journal of Biological Chemistry*, 268, 19604–19609.
- Trickett, A. J., & Butlin, R. K. (1994). Recombination suppressors and the evolution of new species. *Heredity*, 73(Pt 4), 339–345.
- Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution*, 69, 1476–1486.
- Wang, J., Xia, Q., He, X., Dai, M., Ruan, J., Chen, J., ... Yu, J. (2005). SILKDB: A knowledgebase for silkworm biology and genomics. *Nucleic Acids Research*, 33, D399–D402.
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.
- Winter, C. B., & Porter, A. H. (2010). AFLP linkage map of hybridizing swallowtail butterflies, *Papilio glaucus* and *Papilio canadensis*. *The Journal of Heredity*, 101, 83–90.
- Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution; International Journal of Organic Evolution*, 65, 1897–1911.
- Zandt Brower, J. V. (1957). Experimental studies of mimicry in some North American butterflies. *Nature*, 180, 444.
- Zhang, W., Kunte, K., & Kronforst, M. R. (2013). Genome-wide characterization of adaptation and speciation in tiger swallowtail butterflies using de novo transcriptome assemblies. *Genome Biology and Evolution*, 5, 1233–1245.
- Zhu, H., Sauman, I., Yuan, Q., Casselman, A., Emery-Le, M., Emery, P., & Reppert, S. M. (2008). Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. *PLoS Biology*, 6, e4.

SUPPORTING INFORMATION

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

How to cite this article: Ryan SF, Fontaine MC, Scriber JM, Pfrender ME, O'Neil ST, Hellmann JJ. Patterns of divergence across the geographic and genomic landscape of a butterfly hybrid zone associated with a climatic gradient. *Mol Ecol*. 2017;26:4725–4742. <https://doi.org/10.1111/mec.14236>