

GENOMIC REGIONS WITH A HISTORY OF DIVERGENT SELECTION AFFECT FITNESS OF HYBRIDS BETWEEN TWO BUTTERFLY SPECIES

Zachariah Gompert,^{1,2} Lauren K. Lucas,¹ Chris C. Nice,³ James A. Fordyce,⁴ Matthew L. Forister⁵ and C. Alex Buerkle¹

¹Department of Botany and Program in Ecology, University of Wyoming, Laramie, Wyoming 82071

²E-mail: zgomper@uwyo.edu

³Department of Biology, Texas State University, San Marcos, Texas 78666

⁴Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996

⁵Department of Biology, University of Nevada, Reno, Nevada 89557

Received August 11, 2011

Accepted January 13, 2012

Data Archived: Dryad: doi:10.5061/dryad.f0b2f083

Speciation is the process by which reproductively isolated lineages arise, and is one of the fundamental means by which the diversity of life increases. Whereas numerous studies have documented an association between ecological divergence and reproductive isolation, relatively little is known about the role of natural selection in genome divergence during the process of speciation. Here, we use genome-wide DNA sequences and Bayesian models to test the hypothesis that loci under divergent selection between two butterfly species (*Lycaeides idas* and *L. melissa*) also affect fitness in an admixed population. Locus-specific measures of genetic differentiation between *L. idas* and *L. melissa* and genomic introgression in hybrids varied across the genome. The most differentiated genetic regions were characterized by elevated *L. idas* ancestry in the admixed population, which occurs in *L. idas*-like habitat, consistent with the hypothesis that local adaptation contributes to speciation. Moreover, locus-specific measures of genetic differentiation (a metric of divergent selection) were positively associated with extreme genomic introgression (a metric of hybrid fitness). Interestingly, concordance of differentiation and introgression was only partial. We discuss multiple, complementary explanations for this partial concordance.

KEY WORDS: Divergent selection, hybridization, introgression, population genomics, speciation.

Speciation is a fundamental evolutionary process that occurs when genetic differentiation leads to reproductive isolation (e.g., a reduced propensity to mate or reduced fitness of immigrants or hybrid offspring) between divergent lineages. As evidenced by empirical data, species boundaries are porous rather than absolute, particularly during the early stages of divergence, and reproductive isolation is a property of genetic loci and not a property of populations or lineages (Harrison and Rand 1989; Wu 2001). Studies of population divergence and speciation with gene flow have consistently documented heterogeneity across the genome

in the extent of genetic differentiation between lineages and introgression in hybrid zones (Dowling and Hoeh 1991; Rieseberg et al. 1999; Turner et al. 2005; Hohenlohe et al. 2010; Lawniczak et al. 2010). This genome-wide heterogeneity in genetic differentiation and introgression can arise from variation among loci in their contribution to fitness in parental and hybrid populations (Payseur et al. 2004; Harr 2006; Nosil et al. 2009a). Individual genes responsible for local adaptation and reproductive isolation have been identified (e.g., Mihola et al. 2009; Tang and Presgraves 2009; Barr and Fishman 2010; Nosil and Schluter

2011). Nonetheless, and despite considerable efforts and progress, we are only beginning to understand the process of genomic divergence during speciation and the importance of divergent selection and local adaptation for the evolution of reproductive isolation.

Theory and data demonstrate that reproductive isolation is entwined with divergent selection and local adaptation (Mayr 1963; Endler 1977; Dodd 1989; Nosil et al. 2002; Rundle and Nosil 2005), as originally envisioned by (Darwin 1859). Considering speciation with gene flow, divergent selection is a direct cause of reproductive isolation if immigrants or their hybrid offspring have low fitness. Similarly, divergent selection can cause adaptive divergence in habitat use, phenology, or mating signals and reduce the frequency of interspecific matings (Jiggins et al. 2001; Nosil et al. 2002; Nosil 2004). When species diverge in geographic isolation and in the absence of gene flow, divergent selection and reproductive isolation upon secondary contact have the potential to be partially decoupled. For example, reproductive isolation upon secondary contact in geographically abutting or admixed populations could be due to mechanisms other than historical divergent selection, if the environment or genetic composition of populations differed between periods of allopatric divergence and secondary contact. Selection pressures are certainly known to vary in time (Grant and Grant 2002; Siepielski et al. 2009), and studies contrasting introgression among multiple hybrid zones provide evidence that the effect of a locus on hybrid fitness can be environment-dependent (Nolte et al. 2009; Teeter et al. 2010).

Similarly, genetic differences that cause reproductive isolation upon secondary contact can evolve through genetic hitchhiking or as byproduct of nonadaptive forces, such as genetic drift or gene conversion (Gavrilets et al. 1998; Fierst and Hansen 2010). For example, although Dobzhansky–Muller (DM) incompatibilities evolve most rapidly if they are a pleiotropic byproduct of loci experiencing divergent selection, DM incompatibilities can also readily evolve through tight linkage with selected loci or in the absence of selection (Turelli et al. 2001; Gavrilets 2003). For example, the proliferation of transposable elements can cause hybrid inviability and can readily occur within geographically isolated nonadmixed populations unfacilitated by selection (Hurst and Werren 2001; Lynch 2007). Likewise, reciprocal loss or deactivation of duplicated genes can occur without selection and cause inviability in hybrid offspring (Lynch and Force 2000; Mizuta et al. 2010).

Even if the same suite of loci are affected by selection in geographically isolated nonadmixed populations and admixed populations, the genomic consequences of selection could differ. Specifically, because admixture generates linkage disequilibrium, the size of the genetic region affected by selection in an admixed population could be much larger than the size of the genetic region affected by divergent selection between geographically isolated

nonadmixed populations. In other words, differences in linkage disequilibrium could alter the genomic response to natural selection of the same loci in admixed and nonadmixed populations. This effect would be intensified if divergent selection causes soft selective sweeps rather than hard selective sweeps, as the genetic footprint of selection would be further reduced (Hermisson and Pennings 2005).

Consequently, whether loci under divergent selection between geographically isolated nonadmixed populations also experience selection or contribute to reproductive isolation in hybrid zones or admixed populations is an open question that must be addressed empirically. This is an interesting question because selection in hybrids is a direct measure of a major component of reproductive isolation and comparing selection in geographically isolated nonadmixed and admixed populations will increase our understanding of (1) divergent selection's importance for the evolution of reproductive isolation, (2) geographic variation in the genetic basis of reproductive isolation, and (3) the effect of admixture linkage disequilibrium on the genomic consequences of selection. Indirect, population genetic metrics associated with selection in geographically isolated nonadmixed and admixed populations can be used to address these issues.

Genetic differentiation and introgression are affected by selection and various stochastic processes (Buerkle et al. 2011; Gompert et al. 2012). Thus, there is not a simple correspondence between selection and measures of genetic differentiation (i.e., measures of allele frequency differences) or introgression. Nonetheless, because divergent selection reduces genetic diversity and increases genetic differentiation for the selected locus and linked loci (i.e., genetic hitchhiking; Maynard-Smith and Haigh 1974; Gillespie 2000), divergently selected loci are more likely to reside in highly differentiated regions of the genome (Beaumont and Nichols 1996; Beaumont and Balding 2004). We define introgression as the movement of alleles from one gene pool into another through hybridization and repeated backcrossing. Introgression can be measured by quantifying the movement of alleles across the geographic range of hybridizing lineages (hereafter, geographic introgression; e.g., Barton and Hewitt 1985; Szymura and Barton 1986; Porter et al. 1997) or between different genomic backgrounds (hereafter, genomic introgression; e.g., Anderson 1949; Rieseberg et al. 1999; Lexer et al. 2007; Gompert and Buerkle 2009). These two forms of introgression can, but need not, be coincident (Teeter et al. 2010). First, considering geographic introgression, relative to genetic regions that do not affect fitness, the geographic range of introgression should be reduced for genetic regions that harbor variants that decrease fitness in hybrids (i.e., steeper geographic clines with a rapid, exponential decay of foreign allele frequencies away from the hybrid zone center; Barton 1983; Barton and Hewitt 1985; Szymura and Barton 1986, 1991). Similarly, considering genomic introgression within

an admixed population, genetic regions associated with reduced hybrid fitness should have a low frequency of the negatively selected alleles from one parental species or have alleles from each parental species confined primarily to alternative genomic backgrounds (Rieseberg et al. 1999; Gompert et al. 2012). Thus, loci that contribute to isolation through fitness variation in hybrids are more likely to reside in regions of the genome with specific, extreme patterns of geographic or genomic introgression.

In this article, we analyze genome-wide DNA sequence data with novel Bayesian models to quantify and compare genetic differentiation between *Lycaeides idas* and *L. melissa* (Lepidoptera: Lycaenidae) and genomic introgression in a hybrid zone. *Lycaeides idas* and *L. melissa* are among five nominal *Lycaeides* species that occur in North America and likely diverged from a common Eurasian ancestor within the last 2.4 million years (Gompert et al. 2006a, 2008a; Vila et al. 2011; Forister et al. 2011). *Lycaeides idas* and *L. melissa* differ in male genitalic morphology, wing pattern, host plant use, phenology, and behavior (Scott 1986; Fordyce et al. 2002; Lucas et al. 2008; Gompert et al. 2010b). Diversification of North American *Lycaeides* likely occurred during periods of Pleistocene glacial advance. However, many of these lineages have hybridized in regions of secondary contact (Gompert et al. 2006a,b, 2010a,b). In the current study, we focus on a hybrid zone between *L. idas* and *L. melissa* in northwestern Wyoming (USA), specifically in Jackson Hole and the Gros Ventre mountains (Nabokov 1949; Gompert et al. 2010b). Secondary contact in this geographic region occurred recently, as much of this area was glaciated until 14,000 years before present (Harris et al. 1997). Henceforth, we refer to the admixed population in Wyoming as Jackson Hole *Lycaeides*. The habitat occupied by Jackson Hole *Lycaeides* is similar to the habitat occupied by *L. idas* (relatively mesic forest and montane environments), and differs from the habitat occupied by *L. melissa* (xeric sites, often associated with agricultural fields; Gompert et al. 2010b). *Lycaeides* populations have a patchy, discontinuous distribution that is tied to the distribution of their host plants. Jackson Hole *Lycaeides* feed on *Astragalus miser*, which is the same host plant used by many *L. idas* populations in the central Rocky Mountains (Scott 1986; Gompert et al. 2010b). Conversely, nearby *L. melissa* populations feed either on feral *Medicago sativa* or *A. bisulcatus*. The Jackson Hole *Lycaeides* populations are not directly connected with parental populations and there is only very limited gene flow between Jackson Hole and parental populations (Gompert et al. 2010b).

Herein, we use DNA sequence data from thousands of loci and Bayesian models to document considerable variation across the genome in the magnitude of genetic differentiation between *L. idas* and *L. melissa* and genomic introgression in Jackson Hole *Lycaeides*. We then test the hypothesis that divergent selection between geographically disjunct *L. idas* and *L. melissa* popu-

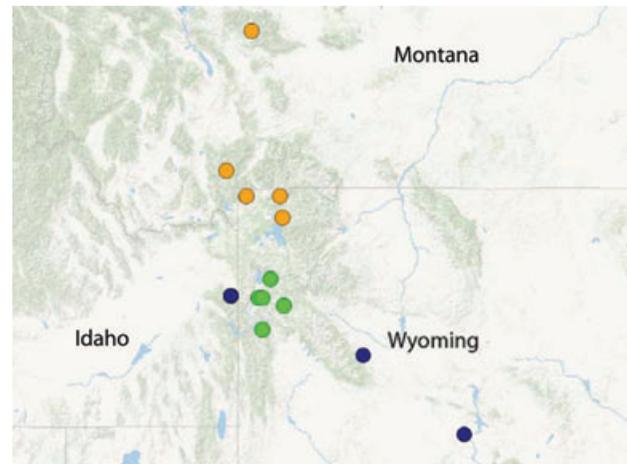


Figure 1. Population sample locations for *L. idas* (orange points), *L. melissa* (blue points), and Jackson Hole *Lycaeides* (green points; Table 1).

lations predicts selection and the consequences of selection in admixed Jackson Hole *Lycaeides*. We evaluate two specific predictions that would support this hypothesis: (1) because *L. idas* and Jackson Hole *Lycaeides* occupy similar habitats and use the same host plant, *L. idas* alleles at highly differentiated loci that could be associated with adaptation to habitat or host plant will be found at disproportionately high frequency in Jackson Hole *Lycaeides*, and (2) locus-specific measures of genetic differentiation (a metric of divergent selection) will be positively correlated with extreme genomic introgression (a metric of selection in hybrids). The first prediction is clearly supported by our results, whereas support for the second is mixed and less conclusive. For example, we document concordant patterns of differentiation and genomic introgression, particularly for highly divergent regions of the genome. However, we also document weakly differentiated loci with exceptional patterns of introgression and highly differentiated loci with patterns of introgression not different from the genome average. We discuss several possible causes for this discordance.

Methods

SEQUENCE DATA COLLECTION AND ASSEMBLY

We generated DNA sequence data from 116 *L. idas* (five localities), 76 *L. melissa* (three localities), and 186 Jackson Hole *Lycaeides* (five localities; Fig. 1, Table 1). We collected butterflies within U.S. National Parks (NP) in accordance with NP Service permits: Yellowstone NP (YELL-2008-SCI-5682) and Grand Teton NP (GRTE-2008-SCI-0024). We isolated and purified DNA from each of the sampled butterflies from approximately 10 mg of thoracic tissue using Qiagen's DNeasy 96 Blood and Tissue Kit (Cat. No. 69581; Qiagen Inc., Valencia, CA). We generated reduced genomic complexity libraries for each individual using a restriction

Table 1. Population sample information (JH = Jackson Hole *Lycaeides*; *N* = sample size).

Locality	Taxon	ID	<i>N</i>	Latitude (°N)	Longitude (°W)	Elevation (m)
King's Hill, MT	<i>L. idas</i>	KHL	18	46.8407	110.6990	2239
Garnet Peak, MT	<i>L. idas</i>	GNP	20	45.4323	111.2245	1910
Bunsen Peak, WY	<i>L. idas</i>	BNP	20	44.9337	110.7212	2260
Trout Lake, WY	<i>L. idas</i>	TRL	20	44.9019	110.1291	2124
Hayden Valley, WY	<i>L. idas</i>	HNV	38	44.6823	110.4945	2344
Mt. Randolph, WY	JH	MRF	37	43.8547	110.3918	2221
Upper Slide Lake, WY	JH	USL	36	43.5829	110.3328	2246
Teton Science School, WY	JH	TSS	39	43.6974	110.6102	2180
Blacktail Butte, WY	JH	BTB	38	43.6382	110.6820	2220
Bull Creek, WY	JH	BCR	36	43.3007	110.5530	2195
Victor, ID	<i>L. melissa</i>	VIC	25	43.6590	111.1114	1850
Lander, WY	<i>L. melissa</i>	LAN	26	42.6533	108.3551	1787
Sinclair, WY	<i>L. melissa</i>	SIN	25	41.8517	107.0917	1961

fragment-based procedure (van Orsouw et al. 2007; Gompert et al. 2010a; Andolfatto et al. 2011). We labeled fragments from each butterfly with 10 base pair (bp) individual identification sequences (i.e., barcodes). DNA sequencing of the reduced genomic complexity libraries was performed by the National Center for Genome Research (Santa Fe, NM) using the Illumina GAII platform. We used SeqMan NGen 3.0.4 (DNASTAR) to perform a de novo assembly for a subset of the sequences (12 million) and generate a reference sequence. We then assembled the full sequence dataset (110 million sequences) to the reference using SeqMan xng 1.0.3.3 (DNASTAR). We used custom Perl scripts in conjunction with samtools and bcftools (Li et al. 2009) to identify variant sites in the assembled sequence data and to determine the number of reads supporting each alternative nucleotide state for each individual and locus. We identified 119,677 variable sites using stringent criteria. The mean number of sequences per variable site per individual was 2.21. Because sequence coverage was low, we did not attempt to call genotypes, but rather incorporated genotype uncertainty in all analyses (see below). A detailed description of sequence data collection, assembly, and variant calling is included in the *Supporting information*.

GENETIC VARIATION, LINKAGE DISEQUILIBRIUM, AND POPULATION STRUCTURE

We used a Bayesian model to estimate allele frequencies for each of 119,677 variable nucleotides based on the observed sequence data (see Population allele frequency model in the *Supporting information*). The model incorporates uncertainty in genotypic state arising from low-coverage next-generation sequence data, but is otherwise equivalent to standard methods used to estimate allele frequencies (e.g., Gillespie 2004). In other words, we treated the genotype at each locus and the population allele frequency as unknown model parameters, which we estimated from the DNA

sequence data. We estimated posterior probability distributions for genotypic state and allele frequencies separately for each of the 13 sampled localities (populations), *L. idas*, *L. melissa*, Jackson Hole *Lycaeides*, and all sampled *Lycaeides*. We obtained posterior parameter estimates for the model parameters (allele frequency and genotypic state) using Markov chain Monte Carlo (MCMC). Each analysis consisted of a single chain iterated for 20,000 steps (we recorded samples every fourth step).

We estimated Burrow's composite measure of Hardy–Weinberg and linkage disequilibrium (Δ) between pairs of variable sites (Weir 1979). This metric does not require phased data or assume Hardy–Weinberg equilibrium, but instead provides a composite measure of intralocus and interlocus disequilibria estimated directly from the genotype frequencies (Weir 1979). We used a Monte Carlo procedure to estimate Δ while accounting for incomplete knowledge of genotypes. Specifically, for each population and taxon, we iteratively sampled genotypic states for each individual from the posterior probability distribution of genotypic states (as estimated above) and calculated $\Delta_{ii'}$ for each locus pair (i.e., locus *i* and locus *i'*). We iterated this procedure 100 times for each locus pair and used the mean value of $\Delta_{ii'}$ as an estimate of $\Delta_{ii'}$. We calculated Δ for the subset of 17,693 variable sites with minor allele frequencies greater than 0.1. We discarded low-frequency variants for this analysis as they are less informative regarding deviations from linkage equilibrium. Similarly, we did not use these loci for the analyses of genetic differentiation and genomic introgression described below, as low-frequency variants provide little information about ancestry and genomic introgression. We summarized the distribution of the absolute value of $\Delta_{ii'}$ (i.e., $|\Delta_{ii'}|$) separately for all pairs of variable sites ($\sim 1.57 \times 10^8$ pairs) and pairs of variable sites associated with the same DNA fragment (2986 pairs or $\sim 0.002\%$ of all locus pairs). Finally, for each admixed population, we determined the proportion of locus

pairs with $\Delta_{ii'} > 0$ (we defined coupling and repulsion alleles so admixture generates positive values of Δ). An excess of locus pairs with $\Delta_{ii'} > 0$ would constitute evidence of admixture linkage disequilibrium in Jackson Hole *Lycaeides*. We used C to implement this Monte Carlo estimation procedure.

We used principal component analysis (PCA) to summarize population genetic structure. Specifically, we used the estimated genotypic state probabilities for two of three genotypes (the heterozygous genotype and one homozygous genotype) at each locus as variables for PCA ($119,677 \times 2 = 239,354$ variables total). We performed the PCA in R using the `prcomp` function after centering, but not scaling, the genotype probabilities. In other words, we used the covariance matrix rather than the correlation matrix for PCA. We also described population genetic structure based on estimates of genome-level F_{ST} between pairs of populations (see the *Supporting information*, F-model).

GENETIC DIFFERENTIATION ACROSS THE GENOME

We used two methods to quantify genome-wide genetic differentiation between *L. idas* and *L. melissa*: (1) we estimated F_{ST} using a hierarchical Bayesian implementation of the F-model, (2) we calculated G_{ST} directly from allele frequency estimates (these models are described in the *Supporting information*). The hierarchical Bayesian F-model treats F_{ST} as an evolutionary parameter and allows information sharing among loci, whereas G_{ST} is modeled as a simple summary of allele frequencies for each locus (Holsinger and Weir 2009). The F-model provides a metric of genetic differentiation that is analogous to F_{ST} under several evolutionary models (Balding and Nichols 1995; Nicholson et al. 2002; Falush et al. 2003). Here, we equate this metric of differentiation with F_{ST} . We provide a detailed description of our specification and implementation of the hierarchical Bayesian F-model in the *Supporting information*. Related models for F_{ST} outlier analysis without genotype uncertainty have been proposed and are in wide use (Beaumont and Balding 2004; Foll and Gaggiotti 2008; Guo et al. 2009).

We obtained Bayesian estimates of F_{ST} using 17,693 loci with a global minor allele frequency ≥ 0.1 . We excluded less variable loci from this and all subsequent analyses as they provide little information about genetic divergence or ancestry. We estimated posterior probabilities of F_{ST} for each locus and a metric of F_{ST} for the genome using MCMC. Additionally, we designated outlier loci with higher than expected levels of genetic differentiation between *L. idas* and *L. melissa* given the estimated genome-wide F_{ST} distribution. In other words, we concurrently estimated F_{ST} for individual loci and model parameters describing the expected distribution of F_{ST} across the genome from the DNA sequence data. We then used the expected distribution of F_{ST} across the genome to designate outlier loci. This general procedure for designating outlier loci is described by Gompert and

Buerkle (2011a; 2011b) and in the *Supporting information* (see F-model). F_{ST} outlier loci are more likely to have been influenced by divergent selection than nonoutlier loci (Beaumont and Nichols 1996; Nosil et al. 2009b; Gompert and Buerkle 2011b). We ran two independent chains for 25,000 steps each. Samples were recorded every 10th step following a 1000 step burn-in. We combined the output of the two chains after inspecting the MCMC output to assess convergence to the stationary distribution.

We measured genetic differentiation between pairs of populations to determine whether F_{ST} outlier loci between *L. idas* and *L. melissa* were also highly differentiated between individual pairs of populations. We quantified genome-wide, population-level genetic differentiation using both F_{ST} and G_{ST} , as described for the species-level analyses. We designated outlier loci for each population comparison and quantified the number of outlier loci shared among individual pairs of *L. idas*, *L. melissa*, and Jackson Hole *Lycaeides* populations, and individual pairs of heterospecific populations. We also determined the number of F_{ST} outliers between *L. idas* and *L. melissa* that were also outliers between individual *L. idas* and *L. melissa* populations.

GENOMIC INTROGRESSION

We measured genomic introgression of *L. idas* and *L. melissa* genetic regions in admixed Jackson Hole *Lycaeides*. We were not concerned with the geography of introgression (i.e., geographic introgression), but rather with the movement of genetic material from one genomic background to another within a geographic region (i.e., genomic introgression; Rieseberg et al. 1999; Lexer et al. 2007; Gompert and Buerkle 2009). We quantified locus-specific genomic introgression using the Bayesian genomic cline model on the basis of two locus-specific genomic cline parameters (Gompert and Buerkle 2011a). These cline parameters specify the probability that an individual with hybrid index $H = h$ inherited a gene copy at locus $I = i$ from *L. idas* (denoted ϕ ; the probability of *L. melissa* ancestry is $1 - \phi$). The base probability of *L. idas* ancestry for a locus is equal to an individual's hybrid index. The genomic cline center parameter, α , specifies an increase (positive values of α) or decrease (negative values of α) in the probability of *L. idas* ancestry for a locus relative to the base expectation. The genomic cline rate parameter, β , specifies an increase (positive values) or decrease (negative values) in the rate of transition from low to high probability of *L. idas* ancestry as a function of hybrid index (Fig. S1; Gompert and Buerkle 2011a). The parameter β is a measure of the average ancestry-based pairwise linkage disequilibrium between a marker locus and all other marker loci. More formally,

$$\Phi_{ih} = h + 2(h - h^2)(\alpha_i + \beta_i(2h - 1)), \quad (1)$$

where ϕ_{ih} is given by a simple transformation of Φ_{ih} to ensure $0 \leq \phi \leq 1$ and that ϕ is a monotonically increasing function

of hybrid index (Gompert and Buerkle 2011a). Simulations have demonstrated that selection against specific hybrid genotypes (i.e., locus-specific reproductive isolation), whether arising from single locus (underdominance) or multilocus (DM) incompatibilities, affects α and β , but the effect of selection on α is often more pronounced, particularly if dispersal from parental populations is limited (Gompert and Buerkle 2011a; Gompert et al. 2012). Specifically, underdominance and epistatic incompatibilities cause extreme positive or negative genomic cline center parameters (α) and high, positive genomic cline rate parameters (β). The same simulation studies demonstrate that selection favoring a locally favored homozygous genotype will affect α , but has little to no effect on β (Gompert and Buerkle 2011a; Gompert et al. 2012).

We used a modified implementation of the Bayesian genomic cline model to quantify genome-wide variation in introgression in Jackson Hole *Lycaeides* (see Genomic cline model with genotype uncertainty in the *Supporting information*). The modified model incorporates uncertainty in genotypic state inherent in next-generation sequence data, but is otherwise identical to the model described by Gompert and Buerkle (2011a). We estimated marginal posterior probability distributions for hybrid indexes and cline parameters (α and β) using MCMC. We ran five independent chains for 50,000 steps each and recorded samples from the posterior distribution every 20th step following a 30,000 step burn-in. We combined the output of the five chains after inspecting the MCMC output to assess convergence to the stationary distribution.

COMPARATIVE ANALYSES

The following comparative analyses, which test the hypothesis that locus-specific genetic differentiation between *L. idas* and *L. melissa* predicts locus-specific introgression in Jackson Hole *Lycaeides*, consider only the genomic cline parameter α , as we did not detect meaningful variation for β (see Results). Jackson Hole *Lycaeides* inherited the majority of their genome from *L. idas* (see Results) and use the same host plant and occupy similar habitats to nearby *L. idas* populations (Scott 1986; Gompert et al. 2010b). Thus, we predict that, for loci that are highly differentiated between *L. idas* and *L. melissa*, selection in Jackson Hole *Lycaeides* should favor *L. idas* alleles more frequently than *L. melissa* alleles. To test this hypothesis, we asked whether *L. idas* \times *L. melissa* F_{ST} outlier loci were more likely to have elevated *L. idas* ancestry than expected by chance. By elevated *L. idas* ancestry, we simply mean that the point estimate of α (i.e., $\hat{\alpha}$) was greater than zero. Because genomic cline parameter α represents a deviation from the ancestry probability predicted solely from hybrid index and because the α parameters are constrained to sum to zero, we used the expectation that 50% of outlier loci should have $\hat{\alpha} > 0$ as a null hypothesis. We tested for a significant deviation from

this expectation based on a binomial probability distribution with $p_{L.idas} = 0.5$. Additionally, we obtained Bayesian posterior estimates of the probability that F_{ST} outlier loci had estimates of $\hat{\alpha} > 0$ by specifying a binomial likelihood for the number of F_{ST} outlier loci with $\hat{\alpha} > 0$ and an uninformative beta prior on $p_{L.idas}$ (i.e., $p_{L.idas} \sim \text{beta}[1, 1]$). We repeated these analyses with all 17,693 loci with estimates of F_{ST} and α for comparison.

We were also interested in whether there was a general correlation between the genetic regions affected by selection in the geographically disjunct parental populations and Jackson Hole *Lycaeides*. A correlation between locus-specific F_{ST} and the absolute value of locus-specific genomic cline center parameter, α , could provide support for this hypothesis. However, estimates of α could be affected by the ancestry information contained in the sequence data for each locus (i.e., the correspondence between what population an allele was inherited from and the allelic state). Ancestry information is dependent on the allele frequency differential between parental species and thus should be positively correlated with locus-specific F_{ST} . Therefore, in addition to calculating the correlation between locus-specific F_{ST} and the absolute value of the genomic cline center parameter ($|\alpha|$) for the *Lycaeides* sequence data, we estimated the same correlation for simulated datasets. We first analyzed 100×4 replicate datasets simulated using four different sets of demographic conditions to ask whether a correlation between F_{ST} and $|\alpha|$ might arise in the absence of selection. Each dataset included 204 biallelic loci. Ideally, we would examine simulated datasets with the same demographic history as the Jackson Hole *Lycaeides*. However, given that this history is not known and would be quite difficult to infer with precision, we instead chose to explore a range of demographic conditions. We simulated 10 additional datasets with a large number of loci (15,010) to determine how the correlation between F_{ST} and $|\alpha|$ for the simulated and observed data was affected by the minimum ancestry information (i.e., minimum value of F_{ST}) required for a locus to be included in the analysis. The conditions we used to simulate these 10 datasets gave a range of estimated values for F_{ST} and α that were very similar to the range of estimated values from the *Lycaeides* data. Full details regarding the simulation procedures and analysis of simulated datasets are included as *Supporting information* (see Admixture simulations).

Results

We sequenced 110.2 million, 108 bp DNA fragments from 116 *L. idas* (five localities), 76 *L. melissa* (three localities), and 186 Jackson Hole *Lycaeides* (five localities; Fig. 1, Table 1). These data contained 119,677 variable nucleotides distributed among 52,239 short (~ 92 bp long) contigs. Estimates of $\Delta_{i'}$, which is a joint metric of deviations from Hardy–Weinberg and linkage equilibrium, were quite low (Table 2). Specifically, the median value

Table 2. Quantiles of the empirical distribution of estimated pairwise composite linkage disequilibria ($|\Delta_{i'j'}|$) for *Lycaeides* populations and taxa (combined populations). “Linked loci” refers to pairs of variable sites from the same DNA fragment. Population IDs are from Table 1.

Taxon	Population	All loci				Linked loci			
		$q_{0.25}$	$q_{0.50}$	$q_{0.75}$	$q_{0.99}$	$q_{0.25}$	$q_{0.50}$	$q_{0.75}$	$q_{0.99}$
<i>L. idas</i>	BNP	0.0026	0.0076	0.0164	0.0502	0.0040	0.0151	0.0386	0.1238
	GNP	0.0021	0.0063	0.0148	0.0469	0.0031	0.0141	0.0355	0.1236
	HNV	0.0027	0.0065	0.0122	0.0366	0.0062	0.0157	0.0369	0.1226
	KHL	0.0017	0.0046	0.0116	0.0408	0.0020	0.0070	0.0253	0.0971
	TRL	0.0027	0.0079	0.0166	0.0500	0.0045	0.0155	0.0371	0.1226
	Combined	0.0018	0.0038	0.0069	0.0197	0.0059	0.0148	0.0344	0.1072
<i>L. melissa</i>	LAN	0.0020	0.0064	0.0142	0.0449	0.0034	0.0153	0.0358	0.1280
	SIN	0.0017	0.0059	0.0141	0.0463	0.0023	0.0138	0.0339	0.1346
	VIC	0.0017	0.0055	0.0132	0.0433	0.0023	0.0131	0.0326	0.1226
	Combined	0.0016	0.0042	0.0084	0.0262	0.0048	0.0149	0.0339	0.1213
JH	BCR	0.0032	0.0073	0.0133	0.0375	0.0070	0.0175	0.0388	0.1176
	BTB	0.0029	0.0068	0.0124	0.0356	0.0073	0.0172	0.0388	0.1135
	MRF	0.0033	0.0075	0.0137	0.0392	0.0076	0.0187	0.0417	0.1198
	TSS	0.0032	0.0073	0.0133	0.0379	0.0079	0.0180	0.0430	0.1217
	USL	0.0032	0.0075	0.0138	0.0394	0.0074	0.0188	0.0405	0.1249
	Combined	0.0016	0.0035	0.0062	0.0168	0.0081	0.0186	0.0401	0.1161

of $|\Delta_{i'j'}|$ for each population and each taxon was less than 0.01. The distribution of pairwise disequilibria was similar in *L. idas*, *L. melissa*, and Jackson Hole *Lycaeides* populations. Estimated values of $|\Delta_{i'j'}|$ for physically linked variable sites that occurred on the same DNA fragment (i.e., fewer than 100 bp apart) were two to three times greater than for other loci (mean [across all populations] median value of $|\Delta_{i'j'}|$: linked loci = 0.0155, all loci = 0.0062; Table 2). The proportion of locus pairs with $\Delta_{i'j'} > 0$ in each admixed population was 0.50. PCA of estimated multilocus genotypes and estimates of genome-level F_{ST} demonstrated genetic differentiation between *L. idas* and *L. melissa* populations (Fig. 2; $F_{ST} = 0.074$, 95% credible interval [CI]: 0.072 – 0.075). These analyses indicated that Jackson Hole *Lycaeides* possess an intermediate gene pool relative to *L. idas* and *L. melissa*, but are more genetically similar to *L. idas* (*L. idas* × Jackson Hole $F_{ST} = 0.017$, 95% CI: 0.016 – 0.018; *L. melissa* × Jackson Hole $F_{ST} = 0.047$, 95% CI: 0.046 – 0.048). Ordination of pairwise population genetic differences (measured by F_{ST}) and morphometric analyses gave similar results (Figs. S2–S4, Table S1). Bayesian estimates of hybrid index for individual Jackson Hole *Lycaeides* ranged from 0.63 to 0.78 (Fig. S5; hybrid index measures the proportion of an individual’s genome that was inherited from *L. idas*).

GENETIC DIFFERENTIATION AND GENOMIC INTROGRESSION

We detected considerable interlocus variation in genetic differentiation between *L. idas* and *L. melissa* (Fig. 3 A). Locus-specific

estimates of F_{ST} between *L. idas* and *L. melissa* ranged from 0.059 to 0.298 (mean = 0.074). F_{ST} estimates for 876 loci (out of 17,693) exceeded 0.1. Eighty loci exceeded the 95th quantile of the expected genome-wide distribution of F_{ST} and were classified as statistical outlier loci (21 loci exceeded the 99th quantile of the expected genome-wide distribution and represent extreme outlier loci; Fig. 3A). This means that the probability of the F_{ST} value for these loci conditional on the estimated variation in F_{ST} across the genome was less than 0.05 (outlier loci) or 0.01 (extreme outlier loci). F_{ST} outlier loci are likely enriched for genetic regions tightly linked to loci experiencing divergent selection between *L. idas* and *L. melissa* (Beaumont and Nichols 1996; Beaumont and Balding 2004; Nosil et al. 2009a). All 80 F_{ST} outlier loci between *L. idas* and *L. melissa* were also classified as outlier loci between at least two pairs of *L. idas* and *L. melissa* populations (Table S2). Moreover, all loci classified as outliers between at least 10 of the 15 individual pairs of *L. idas* by *L. melissa* populations were also classified as outliers between *L. idas* and *L. melissa*. Estimates of G_{ST} were highly correlated with estimates of F_{ST} ($r = 0.979$, $P < 2.2 \times 10^{-16}$), but were considerably more variable (mean = 0.054, SD = 0.079, minimum < 0.001, 95th quantile = 0.198, maximum = 0.938; Fig. S6).

Genomic introgression in Jackson Hole *Lycaeides* varied across the genome (Figs. 3B, 4, S7). Genomic cline parameter α was particularly variable, with a minimum of $\alpha = -1.79$ and a maximum of $\alpha = 1.03$. Considering an individual with $h = 0.7$, the probability of *L. idas* ancestry for a locus with $\alpha = -1.79$ is ~ 0 , whereas the probability of *L. idas* ancestry for a locus

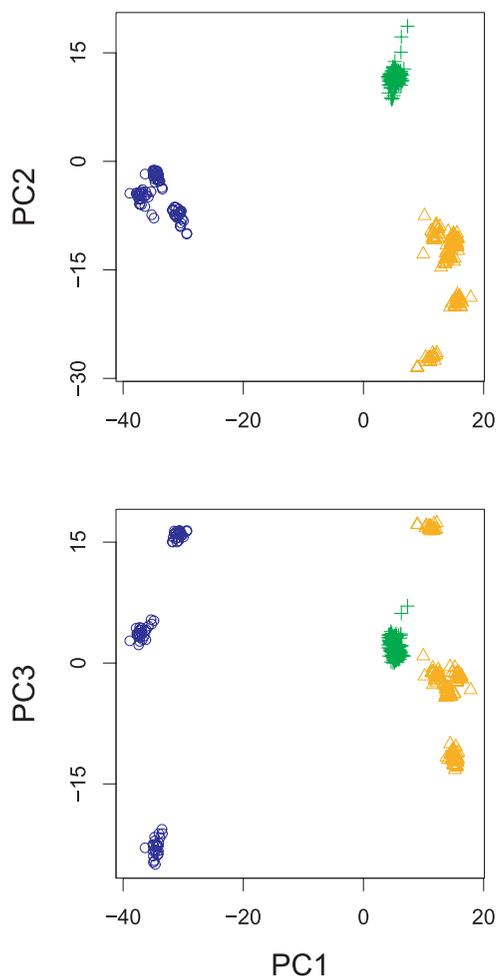


Figure 2. Statistical summary of population genetic structure in *Lycaeides* based on principal component axis one (PC1), axis two (PC2), and axis three (PC3). These axes explain 4.2% (PC1), 2.1% (PC2), and 1.1% (PC3) of the variation in genotype estimates for 119,677 variable nucleotides. Colored symbols represent individuals: *L. idas* (orange triangles; five localities), *L. melissa* (blue circles; three localities), and Jackson Hole *Lycaeides* (green crosses; five localities).

with $\alpha = 1.03$ is ~ 1 . We detected excess *L. idas* ancestry (i.e., the lower bound of the 95% CI for α was greater than zero) for 1791 loci (10.1% of the loci) and excess *L. melissa* ancestry (i.e., the upper bound of the 95% CI for α was less than zero) for 1583 loci (8.9% of the loci; Fig. 4). By excess ancestry, we mean that the 95% CI for α did not include zero. Thus, genomic introgression for approximately 19% of the sampled loci differed from the genome average predicted by hybrid index. In general, extreme values of α are expected for loci that reside in genetic regions affected by selection in Jackson Hole *Lycaeides*. Genomic cline parameter β was less variable (min = -0.298 , max = 0.230). The 95% CI for β encompassed zero for all loci.

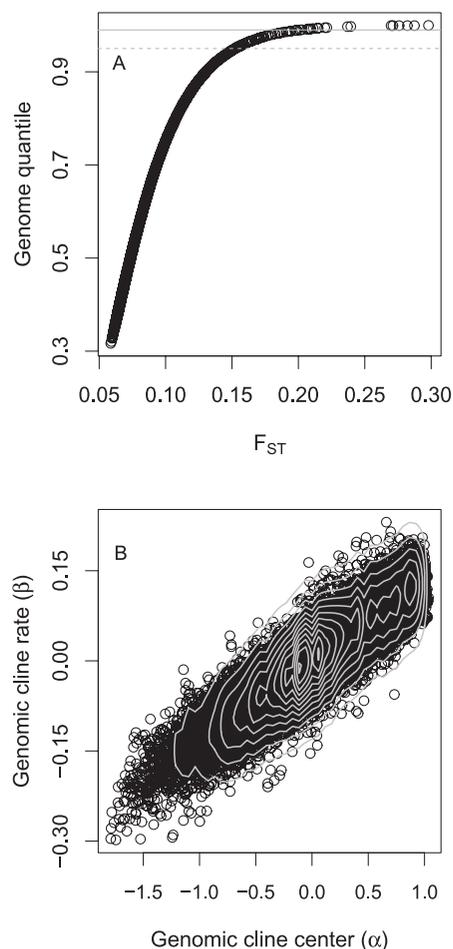


Figure 3. Scatterplots depicting (A) genetic differentiation (measured by F_{ST}) between *L. idas* and *L. melissa*, and (B) variable genomic introgression in Jackson Hole *Lycaeides* populations based on genomic cline center (α) and genomic cline rate (β). Each point represents one of the 17,693 loci. The 99th (solid horizontal line) and 95th (dashed horizontal line) quantiles of the genome-wide distribution of F_{ST} are shown in panel A. Gray contour lines in panel B depict the joint density of estimated cline parameters. Note, posterior estimates of β are not significantly different from zero.

COMPARATIVE ANALYSES

Estimates of genomic cline center ($\hat{\alpha}$; defined as the median of the posterior distribution for α) for 72 of the 80 F_{ST} outlier loci were greater than zero. This represents significantly more outlier loci with an elevated probability of *L. idas* ancestry (i.e., $\hat{\alpha} > 0$) than expected by chance (binomial prob. test, $P = 0.0002$; Bayesian estimate of the binomial probability, $p_{L.idas} = 0.89$, 95% CI: $0.81 - 0.95$). We obtained similar results for the 21 extreme outlier loci. Point estimates of genomic cline center ($\hat{\alpha}$) for 19 of these loci were greater than zero, which is indicative of an elevated probability of *L. idas* ancestry for extreme outlier loci (binomial prob. test, $P = 5.4 \times 10^{-14}$; Bayesian estimate of the binomial

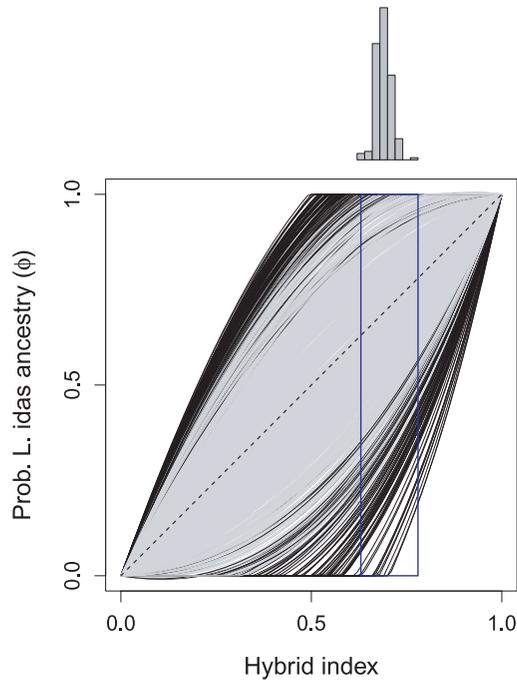


Figure 4. Estimated genomic clines for 1000 representative loci. Each black (locus's 95% CI for α does not include zero) or gray (locus's 95% CI for α includes zero) line represents the genomic cline for a single locus. In other words, each line gives the probability of *L. idas* ancestry at an individual locus as a function of hybrid index. The dashed black line gives $\phi = h$. The marginal histogram shows the distribution hybrid indexes for the sampled individuals. Estimates of the probability of *L. idas* ancestry are only reliable over the range of observed hybrid indexes (0.63–0.78). This range is delineated with a rectangle.

probability, $p_{L.idas} = 0.87$, 95% CI: 0.71 – 0.97). Conversely, $\hat{\alpha}$ was greater than zero for 8546 of the 17,693 loci (i.e., all loci, not just F_{ST} outlier loci). This result is also inconsistent with binomial expectation (binomial prob. test, $P = 6.4 \times 10^{-6}$). However, the probability of a locus with $\hat{\alpha} > 0$ was much closer to 0.5 (Bayesian estimate of the binomial probability, $p_{L.idas} = 0.483$, 95% CI: 0.476 – 0.490) and indicative of a slight excess of loci with $\hat{\alpha} < 0$ (i.e., elevated *L. melissa* ancestry rather than elevated *L. idas* ancestry).

Locus-specific measures of genetic differentiation between nonadmixed populations were correlated with locus-specific estimates of $\hat{\alpha}$ in Jackson Hole *Lycaeides* (Fig. 5). Specifically, we detected a positive correlation between F_{ST} and the absolute value of $\hat{\alpha}$ ($r = 0.152$, $P < 2.2 \times 10^{-16}$). Our analysis of datasets simulated under a variety of demographic conditions indicates that the observed correlation could easily be explained in the absence of selection and could simply be due to variation in the ancestry-information content of loci (see the *Supporting information: Admixture simulations and Fig. S8*). However, for the simulated datasets, the correlation between F_{ST} and $|\hat{\alpha}|$ approached zero

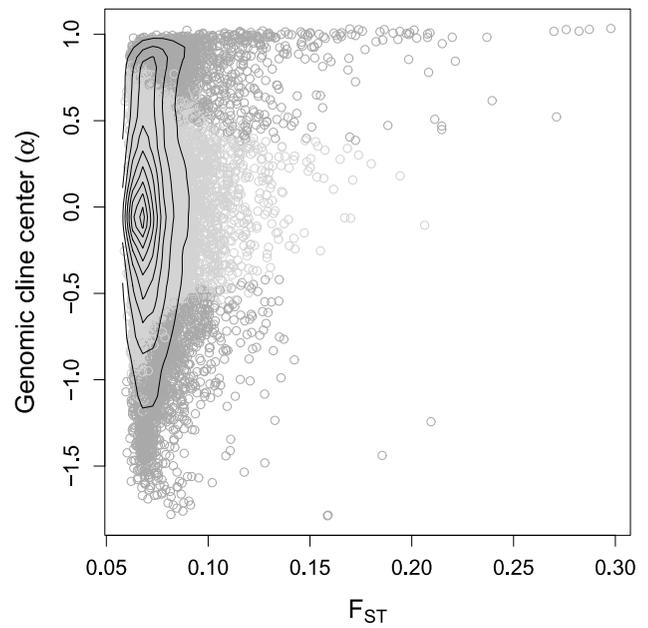


Figure 5. Scatterplot depicting the relationship between F_{ST} between *L. idas* and *L. melissa* and genomic cline parameter α for 17,693 loci. Each circle represents a single locus and contour lines denote the bivariate density. Dark gray circles denote loci with 95% CI for α that do not include 0. Positive values of α correspond to elevated *L. idas* ancestry and negative α correspond to elevated *L. melissa* ancestry.

when only loci with F_{ST} greater than 0.08 or 0.1 were analyzed (Fig. 6). Conversely, for the *Lycaeides* data, the correlation between F_{ST} and $|\hat{\alpha}|$ increased when loci with low estimates of F_{ST} were removed from the analysis. When only highly differentiated loci were used to calculate the correlation between F_{ST} and $|\hat{\alpha}|$, results for individual simulated datasets were erratic because few loci met the criterion for inclusion; however, the mean correlation coefficient across 10 simulated datasets remained near zero (Fig. 6). The correlation coefficient for the *Lycaeides* data was positive regardless of the loci included.

Discussion

Patterns of genetic differentiation and genomic introgression in *Lycaeides* are consistent with the hypothesis that genetic regions experiencing divergent selection between geographically disjunct *L. idas* and *L. melissa* populations also affect hybrid fitness in Jackson Hole *Lycaeides*. As predicted, Jackson Hole *Lycaeides* had an elevated probability of *L. idas* ancestry at a far greater proportion of F_{ST} outlier loci than expected by chance. This result is not an artifact of the overall genomic composition of Jackson Hole *Lycaeides*, as we measured the probability of locus-specific ancestry relative to the genomic admixture proportion or hybrid index, but rather is best explained by selection. There are two

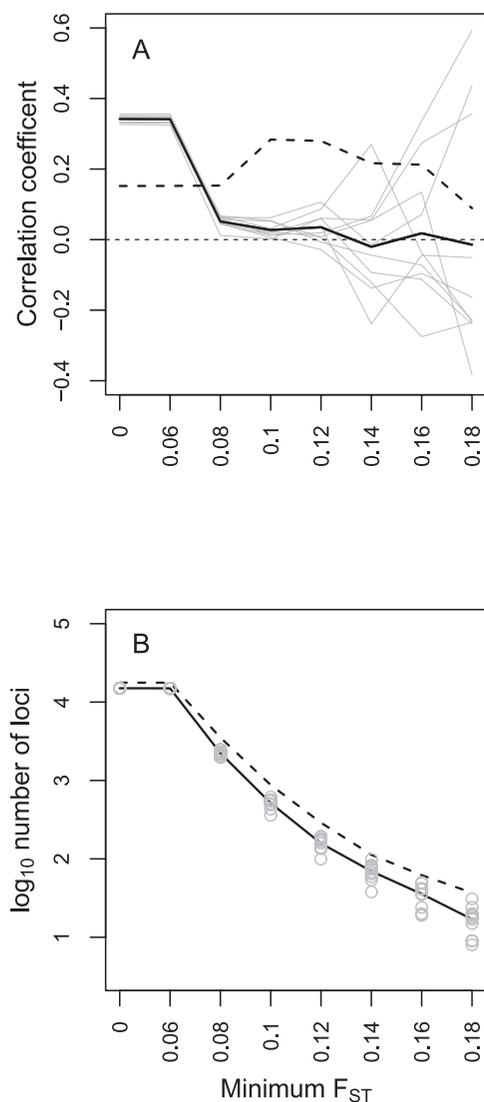


Figure 6. Correlation between F_{ST} and the absolute value of α for loci with $F_{ST} \geq \{0.0, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18\}$ (A) and \log_{10} number of loci meeting this criterion (B). Solid black lines show the mean correlation coefficient or number of loci across 10 simulated datasets, dashed black lines denote the mean for the observed *Lycaeides* sequence data. Gray lines (A) or open circles (B) show results for individual replicates.

primary reasons that selection might favor elevated *L. idas* ancestry at highly differentiated loci. First, a subset of highly differentiated loci between *L. idas* and *L. melissa* could have evolved by divergent selection in response to differences in habitat and host plant, and these same loci could affect habitat or host plant associated fitness in Jackson Hole *Lycaeides*. Alleles inherited from *L. idas* would likely be favored at these loci, because *L. idas* and Jackson Hole *Lycaeides* populations occupy similar habitat and feed on the same larval host plant. This explanation is consistent with an important role for habitat or host plant associated adaptation during the speciation process, as has been suggested

for numerous phytophagous insects (Rand and Harrison 1989; Via et al. 2000; Nosil et al. 2002; McBride and Singer 2010) and other species (Szymura and Barton 1986; Rundle et al. 2000; Schluter and Conte 2009). Elevated *L. idas* ancestry at highly differentiated loci could also be explained by gene interactions, such as DM incompatibilities, in which alleles are disproportionately favored in particular genomic backgrounds. Because Jackson Hole *Lycaeides* have inherited a greater proportion of their genome from *L. idas* than *L. melissa* populations, selection would likely favor *L. idas* ancestry at such loci. We would need several admixed populations occupying different habitats or with different overall genomic compositions to distinguish between these two explanations.

As predicted, we detected a positive association between locus-specific measures of genetic differentiation and genomic introgression. Specifically, locus-specific estimates of F_{ST} were correlated with the absolute value of locus-specific estimates of genomic cline center ($|\hat{\alpha}|$) in Jackson Hole *Lycaeides*. This positive association could reflect consistent selection between parental populations and in the admixed population and suggests that further research to investigate the relationship between selection in parental lineages and hybrids is warranted. However, the association could also simply reflect variation in the ancestry-information content of individual loci. Analysis of datasets simulated under a model of neutral admixture and a variety of demographic conditions certainly supports the possibility that such a correlation could arise as an artifact of variation in ancestry information. However, the correlation between F_{ST} and $|\hat{\alpha}|$ in *Lycaeides* is greater when only loci with $F_{ST} > 0.1$ are examined, whereas the correlation disappears in a series of 10 simulated datasets under comparable conditions. This contrast between neutral simulations and the empirical *Lycaeides* data suggests that some fraction of the loci under divergent selection between *L. idas* and *L. melissa* also affect hybrid fitness in Jackson Hole *Lycaeides*. This conclusion should be robust to different demographic histories as long as they generate roughly the same genome-wide distribution of F_{ST} and α , but further simulation studies might lead to additional insights.

Reproductive isolation requires differences in allele frequencies at the causal loci, but allele frequency differences at linked marker loci are much more likely if isolation evolved by selection rather than drift. Thus, as hybrid fitness is a component of reproductive isolation, the association of F_{ST} , which measures allele frequency differences, and $|\hat{\alpha}|$ is consistent with the hypothesis that genetic differentiation caused by divergent selection in allopatry contributes to reproductive isolation and affects hybrid fitness in Jackson Hole *Lycaeides*. Additional empirical evidence from a variety of taxonomic groups links reproductive isolation and divergent selection; this includes comparative studies of isolation and ecological divergence (Bolnick et al. 2006; Funk et al. 2006),

experimental evolution studies where replicate populations are exposed to different environmental conditions (Dodd 1989), and molecular population genetic characterization of known speciation genes (Tang and Presgraves 2009; Nosil and Schluter 2011). The current results suggest that multiple loci that affect hybrid fitness might have differentiated between *L. idas* and *L. melissa* as a consequence of divergent selection. Future research combining population genomics and experimental crosses to study the association between historical divergent selection and hybrid fitness at many loci across the genome will likely refine this conclusion.

Although our results suggest that a subset of the loci under divergent selection between *L. idas* and *L. melissa* also affect hybrid fitness in Jackson Hole *Lycaeides*, genetic differentiation and genomic introgression were discordant for a substantial portion of the genome. The locus-specific probability of *L. idas* ancestry in Jackson Hole *Lycaeides* for many loci that were markedly differentiated between *L. idas* and *L. melissa* was not different from the genome-wide average probability of *L. idas* ancestry (i.e., hybrid index). Conversely, many loci were only weakly differentiated between *L. idas* and *L. melissa*, but introgressed disproportionately relative to the genome-wide average. As stated previously, there are numerous reasons why locus-specific measures of genetic differentiation might not predict locus-specific measures of genomic introgression. For example, the subset of weakly differentiated loci with high positive or negative estimates of genomic cline center could correspond to loci involved in DM incompatibilities that differentiated either by stochastic processes in allopatry (e.g., Gavrillets et al. 1998), or by selection on standing genetic variation. Although we find little evidence of elevated linkage disequilibrium presently in Jackson Hole *Lycaeides*, admixture linkage disequilibrium is likely to have existed in the past (Fig. S10). If this were the case, these marker loci might have been in linkage disequilibrium with DM incompatibility loci in Jackson Hole *Lycaeides*, but not in *L. idas* and *L. melissa* populations. Clearly, these possible explanations are not mutually exclusive, and our results are consistent with theoretical and empirical results suggesting DM incompatibilities are important for reproductive isolation (e.g., Turelli and Orr 2000; Orr and Turelli 2001; Gavrillets 2003; Brideau et al. 2006; Moyle and Nakazato 2009). Finally, although the habitat, host plant, and genomic composition of Jackson Hole *Lycaeides* are more similar to that of *L. idas* than *L. melissa*, differences certainly exist. Such ecological and genomic differences might alter selection and contribute to the documented discordance in genetic differentiation and genomic introgression.

Importantly, our results demonstrate that both genetic differentiation and genomic introgression vary considerably across the genome in *Lycaeides*. These results are consistent with patterns of heterogeneous genetic differentiation during speciation reported for other taxa (Harr 2006; Egan et al. 2008; Nosil et al. 2008;

Hohenlohe et al. 2010), but are based on considerably more loci than are generally available for taxa that are not genetic model organisms. Genome-wide heterogeneity in genetic differentiation or introgression is often interpreted as evidence that selection shapes variation across the genome (Nolte et al. 2009; Nosil et al. 2009a). This inference is strengthened in *Lycaeides* because many species-level outlier loci were consistently differentiated between individual pairs of *L. idas* and *L. melissa* populations. This consistency is expected if genetic differentiation is the result of selection operating similarly in all populations within each species, but would not be expected if genetic differentiation at these loci was caused by idiosyncratic genetic drift (or selection) operating differently in different conspecific populations (Nosil et al. 2009a). However, recent common ancestry among conspecific populations could also account for this pattern. Despite current genetic differentiation among conspecific populations, this possibility cannot be excluded for *L. idas* or *L. melissa*. Moreover, the potential for stochastic processes to generate variation in genetic differentiation or introgression across the genome has not been fully characterized (Buerkle et al. 2011). Finally, the genomic locations of the loci analyzed in this study are currently unknown. Knowledge of the linkage relationships and ancestry block sizes would lead to additional insights on the distribution of differentiation and introgression across the genome (e.g., Ungerer et al. 1998; Buerkle and Rieseberg 2008; Lawniczak et al. 2010). We are currently constructing genetic and physical maps for this purpose.

The utility of the Bayesian genomic cline model is affected by linkage disequilibrium and the distribution of hybrid indexes in the admixed population(s). As we discuss in the following section, our results indicate little linkage disequilibrium and a rather narrow distribution of hybrid indexes in Jackson Hole *Lycaeides*. These are not ideal conditions for the application of this analytical framework and are likely responsible for the uninformative estimates of genomic cline rate parameter β . It is unclear how our results might differ in a different admixed population with increased linkage disequilibrium and a broader range of hybrid indexes. Nonetheless, we were able to generate informative estimates of α that provide intuitive measures of genomic introgression.

HYBRIDIZATION, INTROGRESSION, AND JACKSON HOLE LYCAEIDES

Linkage disequilibrium in Jackson Hole *Lycaeides* was remarkably low for an admixed population, except between a small subset of loci within the short (< 100 bp) DNA fragments, and was similar to linkage disequilibrium in *L. idas* and *L. melissa* populations. This is consistent with previous estimates of linkage disequilibrium in Jackson Hole *Lycaeides* based on six microsatellite and single nucleotide polymorphism (SNP) loci

(Gompert et al. 2010b). Moreover, we found no evidence of admixture linkage disequilibrium as positive and negative estimates of Δ were observed with nearly equal frequency. In fact, because most loci had similar allele frequencies in *L. idas* and *L. melissa*, admixture is not expected to generate very high levels of linkage disequilibrium (Fig. S10). Likewise, the range of hybrid indexes for Jackson Hole *Lycaeides* was rather narrow (0.63–0.78). Previous estimates of admixture proportions based on AFLP loci analyzed with the admixture model implemented in structure suggested the genomic composition of Jackson Hole *Lycaeides* was more variable than our current results indicate; both sets of results show that hybrids are generally more *L. idas*-like in genomic composition. However, this increased variability from AFLPs was coupled with increased uncertainty in admixture proportion estimates (Gompert et al. 2010b). Moreover, the AFLP dataset included populations at the southern and northern ends of the hybrid zone, whereas our current study focused on populations from the geographic center of the hybrid zone (Gompert et al. 2010b). Together these differences likely account for the discrepancy in hybrid indexes. Regardless, the observed low levels of linkage disequilibrium and narrow range of hybrid indexes in Jackson Hole *Lycaeides* suggest that current gene flow from *L. idas* and *L. melissa* populations to the center of the hybrid zone is minimal to nonexistent. Although the formation of the Jackson Hole *Lycaeides* populations clearly required a period of gene flow and admixture in the past, they presently appear to be evolving with little ongoing influence of nearby parental populations. Genome-wide variation in estimates of genomic cline parameter α and thus the locus-specific probability of *L. idas* ancestry suggest that a subset of loci might have fixed for chromosomal blocks inherited from *L. idas* or *L. melissa*, but that much of the genome still contains segregating variation from both species. The extent of genome stabilization and size of parental chromosomal blocks can be informative about demographic and evolutionary processes affecting admixed populations (Ungerer et al. 1998; Buerkle and Rieseberg 2008), but presently we are limited in our ability to assess these quantities in the absence of mapped genetic loci.

The admixed Jackson Hole *Lycaeides* populations offer an interesting contrast to the homoploid hybrid *Lycaeides* species in the alpine region of the Sierra Nevada of western North America (Gompert et al. 2006a). The alpine species was formed following hybridization between *L. anna* (formerly *L. idas anna*) and *L. melissa*. Unlike the Jackson Hole *Lycaeides*, the alpine hybrid species occupies a different and extreme habitat relative to *L. anna* and *L. melissa* and uses a different, alpine-endemic host plant (Gompert et al. 2006a). Moreover, the alpine hybrid species has novel traits that are adaptive in the alpine habitat (Fordyce and Nice 2003; Gompert et al. 2006a). The alpine hybrid species has a mosaic genome with alleles inherited from both *L. anna* and *L. melissa*, but has evolved novel derived alleles not shared

with either parental species. Conversely, Jackson Hole *Lycaeides* occur in similar habitat to nearby *L. idas* populations and use the same host plant. Moreover, Jackson Hole *Lycaeides* are largely segregating for parental alleles. These differences could simply reflect a more recent origin for Jackson Hole *Lycaeides* relative to the alpine species in the Sierra Nevada. However, colonization of a novel habitat and host plant in the Sierra Nevada, rather than simple geographic isolation, has likely contributed to these different outcomes of hybridization. Additional geographically isolated hybrid lineages exist in the Warner mountains and White mountains of western North America (Gompert et al. 2008b, 2010a), and perhaps in the San Juan range in southern Colorado and the Wassatch range in Utah (Nabokov 1943, 1949). Contrasting the genomic outcomes of secondary contact and admixture in each of these ranges has the potential to provide important insights into how selection affects speciation.

CONCLUSIONS

Our results provide convincing evidence that multiple loci affected by divergent selection between geographically disjunct *L. idas* and *L. melissa* populations also affect fitness in admixed Jackson Hole *Lycaeides*. The idea that divergent selection drives the evolution of reproductive isolation is pervasive and a signature of divergent selection associated with individual speciation genes has been detected previously (Darwin 1859; Endler 1977; Turelli et al. 2001; Barbash et al. 2003; Tang and Presgraves 2009). In some instances, divergent selection can be equated with reproductive isolation, however divergent selection and reproductive isolation can also be decoupled. This study shows that even when selection in parental and hybrid lineages is potentially decoupled, we see an association between divergent selection in allopatry and hybrid fitness (a component of reproductive isolation) at many loci across the genome. Interestingly, some highly differentiated genetic regions do not appear to be associated with hybrid fitness, whereas genomic introgression of some weakly differentiated genetic regions is affected by selection in hybrids. This indicates that selection associated with species barriers might vary by environment or genomic context and that the genetic basis and evolution of reproductive isolation could be quite complex.

ACKNOWLEDGMENTS

This manuscript was improved by comments from C. Lexer, P. Nosil, T. Parchman, two anonymous reviewers, and Editor N. Barton. We thank the following for donating *Lycaeides* specimens or help in the field: M. Diaz, R. Lund, M. Moore, P. Opler, M. Pfeifer, M. Palmer, C. Schmidt, and M. Spurrier. This research was facilitated by the UW-NPS field station in Grand Teton National Park and the research staff at Yellowstone and Grand Teton National Parks. This research was funded by the National Science Foundation (DDIG-1011173 to ZG, NSF EPSCoR WySTEP summer fellowship to LKL, IOS-1021873 and DEB-1050355 to CCN, DEB-0614223 and DEB-1050947 to JAF, DEB-1020509 and DEB-1050726 to MLF, and DBI-0701757 and DEB-1050149 to CAB).

LITERATURE CITED

- Anderson, E. 1949. Introgressive hybridization. John Wiley, New York.
- Andolfatto, P., D. Davison, D. Erezylmaz, T. T. Hu, J. Mast, T. Sunayama-Morita, and D. L. Stern. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* 21: 610–617.
- Balding, D. J., and R. A. Nichols. 1995. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* 96:3–12.
- Barbash, D., D. Siino, A. Tarone, and J. Roote. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc. Natl. Acad. Sci.* 100:5302–5307.
- Barr, C. M., and L. Fishman. 2010. The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. *Genetics* 184: 455–465.
- Barton, N. H.. 1983. Multilocus clines. *Evolution* 37: 454–471.
- Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16: 113–148.
- Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* 13: 969–980.
- Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B* 263:1619–1626.
- Bolnick, D. I., T. J. Near, and P. C. Wainwright. 2006. Body size divergence promotes post-zygotic reproductive isolation in centrarchids. *Evol. Ecol. Res.* 8: 903–913.
- Brideau, N. J., H. A. Flores, J. Wang, S. Maheshwari, X. Wang, and D. A. Barbash. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* 314: 1292–1295.
- Buerkle, C. A., Z. Gompert, and T. L. Parchman. 2011. The $n = 1$ constraint in population genomics. *Mol. Ecol.* 20: 1575–1581.
- Buerkle, C. A., and L. H. Rieseberg. 2008. The rate of genome stabilization in homoploid hybrid species. *Evolution* 62: 266–275.
- Darwin, C. R. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London, England.
- Dodd, D. M. B. 1989. Reproductive isolation as a consequence of adaptive divergence in *Drosophila pseudoobscura*. *Evolution* 43:1308–1311.
- Dowling, T. E., and W. R. Hoeh. 1991. The extent of introgression outside the contact zone between *Notropis cornutus* and *Notropis chrysocephalus* (Teleostei: Cyprinidae). *Evolution* 45: 944–956.
- Egan, S. P., P. Nosil, and D. J. Funk. 2008. Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution* 62:1162–1181.
- Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton Univ. Press, Princeton, NJ.
- Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
- Fierst, J. L., and T. F. Hansen. 2010. Genetic architecture and postzygotic reproductive isolation: evolution of Bateson-Dobzhansky-Muller incompatibilities in a polygenic model. *Evolution* 64: 675–693.
- Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977–993.
- Fordyce, J. A., and C. C. Nice. 2003. Variation in butterfly egg adhesion: adaptation to local host plant senescence characteristics? *Ecol. Lett.* 6:23–27.
- Fordyce, J. A., C. C. Nice, M. L. Forister, and A. M. Shapiro. 2002. The significance of wing pattern diversity in the Lycaenidae: mate discrimination by two recently diverged species. *J. Evol. Biol.* 15:871–879.
- Forister, M. L., Z. Gompert, J. A. Fordyce, and C. C. Nice. 2011. After 60 years, an answer to the question: what is the Karner blue butterfly. *Biol. Lett.* 7: 399–402.
- Funk, D. J., P. Nosil, and W. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc. Natl. Acad. Sci. USA* 103: 3209–3213.
- Gavrilets, S. 2003. Perspective: models of speciation: what have we learned in 40 years? *Evolution* 57:2197–2215.
- Gavrilets, S., H. Li, and M. D. Vose. 1998. Rapid parapatric speciation on holey adaptive landscapes. *Proc. R. Soc. Lond. B* 265:1483–1489.
- Gillespie, J. 2004. Populations genetics : a concise guide. 2nd ed. Johns Hopkins Uni. Press, Baltimore, MD.
- Gillespie, J. H. 2000. Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* 155:909–919.
- Gompert, Z., and C. A. Buerkle. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Mol. Ecol.* 18: 1207–1224.
- . 2011a. Bayesian estimation of genomic clines. *Mol. Ecol.* 20:2111–2127.
- . 2011b. A hierarchical Bayesian model for next-generation population genomics. *Genetics* 187:903–917.
- Gompert, Z., J. A. Fordyce, M. L. Forister, A. M. Shapiro, and C. C. Nice. 2006a. Homoploid hybrid speciation in an extreme habitat. *Science* 314:1923–1925.
- Gompert, Z., C. C. Nice, J. A. Fordyce, M. L. Forister, and A. M. Shapiro. 2006b. Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Mol. Ecol.* 15:1759–1768.
- Gompert, Z., J. A. Fordyce, M. L. Forister, and C. C. Nice. 2008a. Recent colonization and radiation of North American *Lycaeides (Plebejus)* inferred from mtDNA. *Mol. Phylogenet. Evol.* 48:481–490.
- Gompert, Z., M. L. Forister, J. A. Fordyce, and C. C. Nice. 2008b. Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Mol. Ecol.* 17:5231–5244.
- Gompert, Z., M. L. Forister, J. A. Fordyce, C. C. Nice, R. Williamson, and C. A. Buerkle. 2010a. Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of *Lycaeides* butterflies. *Mol. Ecol.* 19:2455–2473.
- Gompert, Z., L. K. Lucas, J. A. Fordyce, M. L. Forister, and C. C. Nice. 2010b. Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive introgression and a patchy hybrid zone. *Mol. Ecol.* 19:3171–3192.
- Gompert, Z., T. L. Parchman, and C. A. Buerkle. 2012. Genomics of isolation in hybrids. *Phil. Trans. R. Soc. Lond. B* 1587:439–450.
- Grant, P. R., and B. R. Grant. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* 296:707–711.
- Guo, F., D. K. Dey, and K. E. Holsinger. 2009. A Bayesian hierarchical model for analysis of single-nucleotide polymorphisms diversity in multilocus, multipopulation samples. *J. Am. Stat. Assoc.* 104:142–154.
- Harr, B.. 2006. Genomic islands of differentiation between house mouse subspecies. *Genome Res.* 16: 730–737.
- Harris, A. G., E. Tuttle, and S. D. Tuffe. 1997. Geology of national parks. 5th ed. Kendall/Hunt Publishing, Dubuque, IA.
- Harrison, R. G., and D. M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp. 110–133 in D. Otte and J. Endler, eds. Speciation and its consequences. Sinauer Associates, Sunderland, MA.
- Hermisson, J., and P. S. Pennings. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169:2335–2352.

- Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6: e1000862.
- Holsinger, K. E., and B. S. Weir. 2009. Fundamental concepts in genetics: genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . *Nat. Rev. Genet.* 10:639–650.
- Hurst, G. D., and J. H. Werren. 2001. The role of selfish genetic elements in eukaryotic evolution. *Nat. Rev. Genet.* 2:597–606.
- Jiggins, C., R. Naisbit, R. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. *Nature* 411:302–305.
- Lawniczak, M. K. N., S. J. Emrich, A. K. Holloway, A. P. Regier, M. Olson, B. White, S. Redmond, L. Fulton, E. Appelbaum, J. Godfrey, et al. 2010. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* 330:512–514.
- Lexer, C., C. A. Buerkle, J. A. Joseph, B. Heinze, and M. F. Fay. 2007. Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity* 98: 74–84.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Proc. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Lucas, L. K., J. A. Fordyce, and C. C. Nice. 2008. Patterns of genitalic morphology around suture zones in North American *Lycaeides* (Lepidoptera: Lycaenidae): implications for taxonomy and historical biogeography. *Ann. Entomol. Soc. Am.* 101:172–180.
- Lynch, M. 2007. *The origins of genome architecture*. Sinauer Associates, Sunderland, MA.
- Lynch, M., and A. Force. 2000. The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* 156:590–605.
- Maynard-Smith, J., and J. Haigh. 1974. Hitch-hiking effect of a favorable gene. *Genet. Res.* 23:23–35.
- Mayr, E. 1963. *Animal species and evolution*. Harvard Univ. Press, Cambridge, MA.
- McBride, C. S., and M. C. Singer. 2010. Field studies reveal strong postmating isolation between ecologically divergent butterfly populations. *PLoS Biol.* 8:e1000529.
- Mihola, O., Z. Trachtulec, C. Vlcek, J. C. Schimenti, and J. Forejt. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323:373–375.
- Mizuta, Y., Y. Harushima, and N. Kurata. 2010. Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proc. Natl. Acad. Sci. USA* 107:20417–20422.
- Moyle, L. C., and T. Nakazato. 2009. Complex epistasis for Dobzhansky-Muller hybrid incompatibility in *Solanum*. *Genetics* 181:347–351.
- Nabokov, V. 1943. The nearctic forms of *Lycaeides* Hub. (Lycaenidae, Lepidoptera). *Psyche* 50: 87–99.
- . 1949. The nearctic members of *Lycaeides* Hübner (Lycaenidae, Lepidoptera). *Bull. Mus. Comp. Zool.* 101: 479–541.
- Nicholson, G., A. V. Smith, F. Jonsson, O. Gustafsson, K. Stefansson, and P. Donnelly. 2002. Assessing population differentiation and isolation from single-nucleotide polymorphism data. *J. R. Stat. Soc. B* 64:695–715.
- Nolte, A. W., Z. Gompert, and C. A. Buerkle. 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Mol. Ecol.* 18: 2615–2627.
- Nosil, P. 2004. Reproductive isolation caused by visual predation on migrants between divergent environments. *Proc. R. Soc. Lond. B* 271:1521–1528.
- Nosil, P., B. J. Crespi, and C. P. Sandoval. 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* 417:440–443.
- Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: “isolation by adaptation” and multiple roles for divergent selection. *Evolution* 62: 316–336.
- Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009a. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18:375–402.
- Nosil, P., L. J. Harmon, and O. Seehausen. 2009b. Ecological explanations for (incomplete) speciation. *Trends Ecol. Evol.* 24:145–156.
- Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. *Trends Ecol. Evol.* 26:160–167.
- Orr, H. A., and M. Turelli. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55:1085–1094.
- van Orsouw, N. J., R. C. J. Hogers, A. Janssen, F. Yalcin, S. Snoeijsers, E. Verstege, H. Schneiders, H. van der Poel, J. van Oeveren, H. Verstegen, et al. 2007. Complexity reduction of polymorphic sequences (CRoPS™): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE* 2: e1172.
- Payseur, B. A., J. G. Krenz, and M. W. Nachman. 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution* 58: 2064–2078.
- Porter, A. H., R. Wenger, H. Geiger, A. Scholl, and A. M. Shapiro. 1997. The *Pontia daplidice-edusa* hybrid zone in northwestern Italy. *Evolution* 51: 1561–1573.
- Rand, D. M., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution* 43: 432–449.
- Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152: 713–727.
- Rundle, H., and P. Nosil. 2005. Ecological speciation. *Ecol. Lett.* 8:336–352.
- Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* 287: 306–308.
- Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. *Proc. Natl. Acad. Sci.* 106:9955–9962.
- Scott, J. 1986. *The butterflies of North America: a natural history and field guide*. Stanford Univ. Press, Stanford, CA.
- Siepielski, A. M., J. D. DiBattista, and S. M. Carlson. 2009. It’s about time: the temporal dynamics of phenotypic selection in the wild. *Ecol. Lett.* 12: 1261–1276.
- Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution* 40: 1141–1159.
- . 1991. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution* 45: 237–261.
- Tang, S., and D. C. Presgraves. 2009. Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* 323:779–782.
- Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman, and P. K. Tucker. 2010. The variable genomic architecture of isolation between hybridizing species of house mouse. *Evolution* 64: 472–485.
- Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. *Trends Ecol. Evol.* 16: 330–343.
- Turelli, M., and H. A. Orr. 2000. Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* 154:1663–1679.
- Turner, T., M. Hahn, and S. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3:1572–1578.
- Ungerer, M. C., S. J. E. Baird, J. Pan, and L. H. Rieseberg. 1998. Rapid hybrid speciation in wild sunflowers. *Proc. Natl. Acad. Sci.* 95: 11757–11762.

- Via, S., A. Bouck, and S. Skillman. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* 54: 1626–1637.
- Vila, R., C. D. Bell, R. Macniven, B. Goldman-Huertas, R. H. Ree, C. R. Marshall, Z. Bálint, K. Johnson, D. Benyamini, and N. E. Pierce. 2011. Phylogeny and palaeoecology of *Polyommatus* blue butterflies show Beringia was a climate-regulated gateway to the New World. *Proc. R. Soc. Lond. B* 278: 2737–2744.
- Weir, B. S.. 1979. Inferences about linkage disequilibrium. *Biometrics* 35: 235–254.
- Wu, C. I. 2001. The genic view of the process of speciation. *J. Evol. Biol.* 14:851–865.

Associate Editor: N. Barton

Supporting Information

The following supporting information is available for this article:

Table S1. Genome level $F_{ST}(\mu)$ for pairs of populations.

Table S2. Number of loci that were classified as statistical F_{ST} outliers ($q \geq 0.95$) in at least n (column names) out of N pairwise population analyses.

Figure S1. Plots depict hypothetical genomic clines.

Figure S2. Goodness of fit for NMDS.

Figure S3. Ordination-based summary of genetic differentiation in *Lycaeides* based on nonmetric multidimensional scaling (NMDS).

Figure S4. Statistical summary of morphological variation in *Lycaeides* based on male genitalic morphology PC1 and wing pattern PC1.

Figure S5. Estimated distribution of hybrid index in Jackson Hole *Lycaeides* ($0 = L. melissa$, $1 = L. idas$).

Figure S6. Scatterplot depicting genetic divergence (measured by Neis G_{ST}) between *L. idas*, *L. melissa*, and Jackson Hole *Lycaeides*.

Figure S7. Estimated distribution of genomic cline parameters α and β in Jackson Hole *Lycaeides*.

Figure S8. Correlation between F_{ST} and the absolute value of α for simulated datasets.

Figure S9. Scatterplot depicting the relationship between G_{ST} between *L. idas* and *L. melissa* and genomic cline parameter α .

Figure S10. Histogram of expected admixture linkage disequilibrium (LD).

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

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.