doi:10.1111/evo.12021

GENOME DIVERGENCE AND THE GENETIC ARCHITECTURE OF BARRIERS TO GENE FLOW BETWEEN *LYCAEIDES IDAS* AND *L. MELISSA*

Zachariah Gompert,[1,2,3] Lauren K. Lucas,[2] Chris C. Nice,[2] and C. Alex Buerkle[1]

[1]*Department of Botany, University of Wyoming, Laramie, Wyoming*

[2]*Department of Biology, Texas State University, San Marcos, Texas*

[3]*E-mail: zg1002@txstate.edu*

Received July 24, 2012

Accepted October 30, 2012

Data Archived: Dryad doi:10.5061/dryad.n136c

Genome divergence during speciation is a dynamic process that is affected by various factors, including the genetic architecture of barriers to gene flow. Herein we quantitatively describe aspects of the genetic architecture of two sets of traits, male genitalic morphology and oviposition preference, that putatively function as barriers to gene flow between the butterfly species *Lycaeides idas* and *L. melissa*. Our analyses are based on unmapped DNA sequence data and a recently developed Bayesian regression approach that includes variable selection and explicit parameters for the genetic architecture of traits. A modest number of nucleotide polymorphisms explained a small to large proportion of the variation in each trait, and average genetic variant effects were nonnegligible. Several genetic regions were associated with variation in multiple traits or with trait variation within- and among-populations. In some instances, genetic regions associated with trait variation also exhibited exceptional genetic differentiation between species or exceptional introgression in hybrids. These results are consistent with the hypothesis that divergent selection on male genitalia has contributed to heterogeneous genetic differentiation, and that both sets of traits affect fitness in hybrids. Although these results are encouraging, we highlight several difficulties related to understanding the genetics of speciation.

KEY WORDS: **Bayesian variable selection regression, genome-wide association mapping, morphology, plant-insect interactions, population genetics, speciation.**

Genome divergence during population divergence or speciation is a complex evolutionary process that is affected by geographic and ecological context, genome structure, selection, and the genetic architecture of inherent barriers to gene flow (Bustamante et al. 2005; Neafsey et al. 2010; Gompert et al. 2012b; Jones et al. 2012; Nosil et al. 2012). Despite decades of theoretical and empirical research, many conflicting views persist about the genetic basis of adaptation and barriers to gene flow (e.g., Orr 2005; Rockman 2012). Building on previous theoretical analyses of Fisher's geometric model (Fisher 1930; Kimura 1983), Orr (1998) showed that the mutations substituted during an entire bout of adaptation to a stable optimum approximate an exponential distribution that includes a modest number of large effect alleles and a greater number of small effect alleles. The same conclusion holds for an adaptive walk in sequence space and is largely independent of the distribution of mutant fitnesses (Gillespie 1984; Orr 2002). Yeaman and Whitlock (2011) showed that adaptation with gene flow results in a more concentrated architecture with functional variants that have larger effects and are more tightly linked. Moreover, studies of the genetic basis of variation in traits that affect fitness or function as barriers to gene flow have identified large effect alleles (e.g., Bradshaw and Schemske 2003; Stinchcombe et al. 2004; Linnen et al. 2009; Joron et al. 2011).

But theory does not require large effect alleles, and known large effect alleles might be anomalies that are unrepresentative

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

of functional variants that generally cause barriers to gene flow (Rockman 2012). When a population adapts to a gradually changing environment, the expected distribution of adaptive substitutions is not exponential, and large effect alleles contribute less to adaptation (Kopp and Hermisson 2009). Likewise, Hermisson and Pennings (2005) showed that alleles with smaller effects are much more likely to contribute to adaptation when selection acts on standing variation. Lab experiments indicate that adaptation from standing genetic variation often involves many loci and occurs by modest changes in allele frequencies rather than allele substitutions (Teotonio et al. 2009; Burke et al. 2010; Pritchard et al. 2010). Finally, most large effect variants affect relatively discrete adaptive traits (e.g., Bradshaw and Schemske 2003; Colosimo et al. 2005; Steiner et al. 2007), whereas putatively adaptive quantitative trait variation is often better explained by many functional variants with small or even infinitesimal phenotypic effects (e.g., Weber et al. 1999; Weiss 2008; Flint and Mackay 2009; Yang et al. 2010).

Various experimental and statistical procedures exist to characterize the genetic basis of reproductive isolation. Artificial crosses have identified genetic regions associated with adaptive phenotypic differences and reproductive isolation (e.g., Coyne et al. 1998; Mackay 2001; Good et al. 2008). But crosses often lack sufficient recombination for fine-scale mapping, are not practical for many organisms, and might not identify functional variants that are relevant in natural populations (Buerkle and Lexer 2008; Weiss 2008). Alternatively, genome-wide association mapping uses linkage disequilibrium in natural populations to identify genetic markers associated with phenotypic variation (e.g., Aranzana et al. 2005; Cho et al. 2009). This approach ensures that the variants identified are relevant in nature and enables fine-scale mapping, but requires a large sample size and many genetic markers because less linkage disequilibrium generally exists in natural populations than artificial crosses (Hirschhorn and Daly 2005). Also, association mapping can only identify variants that segregate within populations, which means functional variants that are fixed between species cannot be mapped. Most published genome-wide association mapping studies (GWAS) have conducted independent tests of association for each SNP. These methods rely on stringent significance threshold to detect associations, and often fail to identify small effect variants or estimate the genetic architecture of traits (Manolio et al. 2009; Yang et al. 2010; Rockman 2012). Recently developed multilocus models that use Bayesian variable selection regression for genome-wide association mapping ameliorate these limitations (see the Methods section for details; Guan and Stephens 2011; Carbonetto and Stephens 2012; Peltola et al. 2012).

Herein we use Bayesian variable selection regression to quantitatively describe aspects of the genetic architecture of two putative barriers to gene flow between the butterfly species *Lycaeides idas* and *L. melissa*, male genitalic morphology and oviposition preference. Both putative barriers are composed of a set of traits. We address three specific questions regarding the genetic basis of variation in male genitalic morphology and oviposition preference: (i) do functional variants with large effects contribute to trait variation, (ii) do the same functional variants affect trait variation within and among populations, and (iii) does variation in related traits (i.e., components of male morphology or oviposition preference) have a common genetic basis? Next, we consider the genetic basis of these putative barriers to gene flow in the context of genome divergence and ask whether genetic regions associated with trait variation exhibit exceptional genetic differentiation between *L. idas* and *L. melissa* or exceptional introgression in admixed *Lycaeides* populations. Such an association would be consistent with the hypothesis that variation in male morphology or oviposition preference affects fitness and constitutes a barrier to gene flow, and that selection on these traits has contributed to genome divergence. Conversely, the lack of an association between the genetics of male morphology or oviposition preference and genetic differentiation or introgression could be explained by multiple alternative hypotheses, which we describe in the Discussion section.

Methods

STUDY SYSTEM

Lycaeides idas and *L. melissa* (Lepidoptera: Lycaenidae) diverged from one or more Eurasian ancestors that colonized North America about 2.4 million years ago (Gompert et al. 2008a; Vila et al. 2011). Reproductive isolation between these nominal species is incomplete, and hybridization occurs in areas of secondary contact (Gompert et al. 2008b, 2010). For example, admixed *Lycaeides* populations exist in the central Rocky mountains area, specifically in Jackson Hole valley and the Gros Ventre mountains (Gompert et al. 2010). Hereafter, we refer to these admixed populations collectively as Jackson Hole *Lycaeides*. These populations are not directly adjacent to nonadmixed *L. idas* or *L. melissa* populations, and genetic data indicate that Jackson Hole *Lycaeides* experience little or no ongoing gene flow with nonadmixed populations (Gompert et al. 2012b).

Nominal *Lycaeides* species and populations differ in male genitalic morphology, wing pattern, habitat use, host plant use, oviposition preference, male mate preference, egg adhesion, diapause, and voltinism, and many of these differences might act as barriers to gene flow (Nabokov 1949; Fordyce et al. 2002; Fordyce and Nice 2003; Lucas et al. 2008; Gompert et al. 2010, 2012a). In the current manuscript, we consider male genitalic morphology and female oviposition preference in *Lycaeides* populations that occupy the central Rocky mountains (Fig. 1). The posterior scleritized portion of the male genitalia interacts with

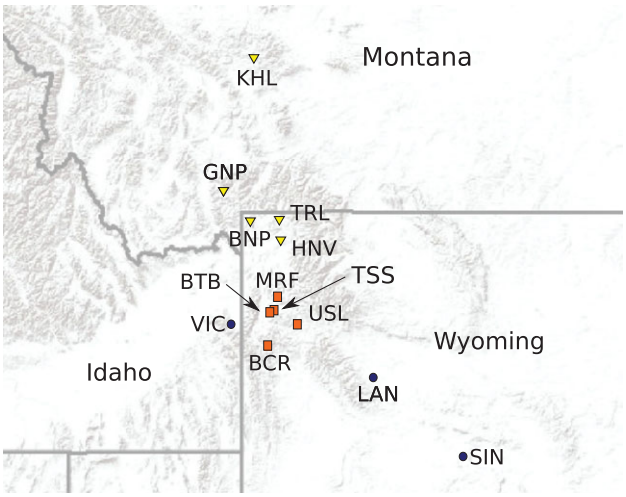

Figure 1. Population sample locations for *L. idas* (triangles), *L. melissa* (circles), and Jackson Hole *Lycaeides* (squares). See Table S1 for population abbreviations.

female reproductive morphology during mating, and is thought to be important for copulation (Nabokov 1949; Nice and Shapiro 1999). This structure is short and wide in *L. idas*, but long and thin in *L. melissa* (Fig. 2; Lucas et al. 2008). *Lycaeides* mate on or near their host plant, and differences in female oviposition preference might indicate local adaptation and could limit interspecific gene flow (Nice et al. 2002). Populations vary in host plant use, but many *L. idas* populations in the central Rocky mountains feed on *Astragalus miser*, whereas nearby *L. melissa* populations feed on *Medicago sativa* or *A. bisulcatus* (Gompert et al. 2012a). Female *Lycaeides* vary in oviposition preference, but often lay more eggs on their natal host plant in choice tests (Gompert et al. 2012a).

Gompert et al. (2012b) documented considerable variation across the genome in the magnitude of genetic differentiation between *L. idas* and *L. melissa* and introgression in Jackson Hole *Lycaeides* based on 17,693 sequenced nucleotide polymorphisms. Measures of genetic differentiation and introgression were consistent with the hypothesis that fitness in hybrids depends on

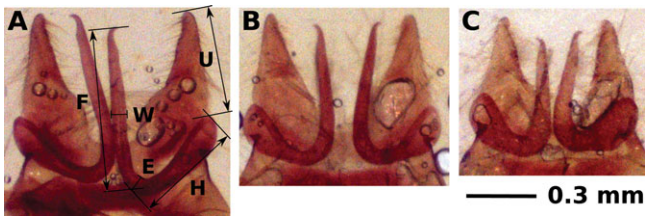

Figure 2. Photographs depict the posterior scleritized portion of dissected male genitalia for *L. melissa* (A), Jackson Hole *Lycaeides* (B), and *L. idas* (C). We measured five linear distances forearm length (*F*), humerelus length (*H*), uncus length (*U*), forearm width (*W*), and elbow width (*E*). We also calculated four measurement ratios: *F* by *W*, *F* by *H*, *F* plus *H* by *E*, and *H* by *U*.

host plant or habitat and that genetic variants under divergent selection between geographically disjunct *L. idas* and *L. melissa* populations also frequently affect hybrid fitness in Jackson Hole *Lycaeides*. Specifically, a disproportionate number of the SNPs that were most differentiated between *L. idas* and *L. melissa* exhibited excess *L. idas* ancestry in hybrids, and locus-specific estimates of genetic differentiation (F_{ST}) between *L. idas* and *L. melissa* were positively correlated with locus-specific introgression in hybrids (the absolute value of the genomic cline parameter α; Gompert et al. 2012b).

Here, we build on the findings of Gompert et al. (2012b) and investigate the relationship between genetic differentiation, introgression, and aspects of the genetic architecture of trait differences that putatively function as barriers to gene flow. The current study is based on previously published DNA sequence data (Gompert et al. 2012b) and phenotypic data from 116 *L. idas* (five populations), 76 *L. melissa* (three populations), and 186 Jackson Hole *Lycaeides* (five populations; Fig. 1, Table S1). We included Jackson Hole *Lycaeides* in the current study because historical admixture and recombination should break up parental genotype combinations and facilitate association mapping (Pfaff et al. 2001; Rieseberg and Buerkle 2002; Buerkle and Lexer 2008).

TRAIT VARIATION

We quantified genitalic morphology of 184 male *Lycaeides* (Table S1). We removed the posterior-most abdominal segments from each male butterfly and digested the soft tissues in hot (100°C) 5 M potassium hydroxide. We then dissected and removed the scleritized portion of the male genitalia. We photographed the genitalia using a dissecting scope with an embedded camera. We then measured forearm length (*F*), humerelus length (*H*), uncus length (*U*), forearm width (*W*), and elbow width (*E*) using the `imagej` software (Fig. 2). We also calculated four shape measurements from the ratios of the length and width measurements (Fig. 2).

Female oviposition preference data were previously described and analyzed by Gompert et al. (2012a) to quantify phenotypic variation in *Lycaeides*. We included oviposition preference data for 167 of those female butterflies in the current study (Table S1). As described by Gompert et al. (2012a), we placed individual, wild-caught adult female butterflies in oviposition chambers with a few sprigs of plant material from each of two host plants, *A. miser* and *M. sativa*. We counted the number of eggs laid by each female on each host plant species after two days. We considered three measures of oviposition preference: the number of eggs laid on *A. miser* (Num. *Ast.*), the number of eggs laid on *M. sativa* (Num. *Med.*), and the proportion of eggs laid on *A. miser* (Prop. *Ast.*).

Character correlations and the distribution of variation within and among populations affect the interpretation of association

mapping results. Thus, we quantified correlations among all male genitalia and oviposition preference characters. We also estimated the proportion of total variance in each trait partitioned among populations. We estimated variance components in a linear model with population as a random effect using restricted maximum likelihood and the R function lmer (Bates et al. 2011; R Development Core Team 2011).

GENETIC VARIATION

The DNA sequence data analyzed in the current manuscript were previously described and analyzed by Gompert et al. (2012b). As described by Gompert et al. (2012b), we constructed reduced genomic complexity DNA libraries for each butterfly included in the morphological or oviposition preference analyses using a restriction fragment-based procedure. We obtained approximately 110 million 108-base pair (bp) individual-indexed short-read DNA sequences using the Illumina GAII platform. We first performed a de novo assembly with a subset of the sequences to generate a reference sequence. We then assembled the full sequence dataset to the reference using SeqMan xng 1.0.3.3 (DNASTAR). We used samtools 0.1.18 (Li et al. 2009), bcftools 0.1.18, and custom Perl scripts to identify variable sites in the assembly and count the number of sequences containing each of two nucleotides for each variable site and individual. We designated 119,677 bi-allelic variable sites (single nucleotide polymorphisms or SNPs) distributed among 51,428 restriction fragments (genetic regions) with a mean of 2.2 sequences per individual per SNP. See Gompert et al. (2012b) for a detailed description of the sequence data, assembly, and variant calling. The physical and genetic map locations of the SNPs are unknown.

We estimated genotypes and population allele frequencies using a hierarchical Bayesian model. Gompert et al. (2012b) used a similar model with these data, but unlike the model described by Gompert et al. (2012b), the analyses in this manuscript incorporated sequence errors and included a conditional prior that describes the genome-wide distribution of allele frequencies (denoted by θ). We describe the new model in the Supporting Information. We allowed the sequence error rate to differ among SNPs. We calculated the mean error probability for each variable site from the base alignment qualities, which we obtained using the mpileup command in samtools 0.1.18 (Li et al. 2009), and used this value as the site-specific error probability. We obtained samples from the joint posterior probability distribution of genotypes ($\mathbf{g}$), allele frequencies ($\mathbf{p}$), and θ using Markov chain Monte Carlo (MCMC). We ran two 26,000 iteration chains for each population with a 1000 iteration burn-in. We recorded parameter values every 10th iteration, and we verified likely convergence to the stationary distribution using qualitative and quantitative analysis of sample histories and parameter estimates.

We estimated two measures of Hardy–Weinberg and linkage disequilibrium for each population and pair of variable sites to quantify statistical associations among SNPs. The first measure, Burrow's Δ, is a composite measure of intralocus and interlocus disequilibria and is estimated directly from genotype frequencies (Weir 1979). The second measure, which is a standardized composite measure of intralocus and interlocus disequilibria, is given by $\Delta' = \frac{\Delta}{\Delta_{MAX}}$ as described by Zaykin (2004). We used a Monte Carlo method to estimate $\hat{\Delta_{ik}} = \sum_{\mathbf{g}} \mathrm{E}[\Delta_{ik}|\mathbf{g}]P(\mathbf{g})$ and $\hat{\Delta'_{ik}} = \sum_{\mathbf{g}} \mathrm{E}[\Delta'_{ik}|\mathbf{g}]P(\mathbf{g})$ for each pair of variables sites (each SNP i by SNP k).

GENETIC ARCHITECTURE OF TRAIT VARIATION

We used the Bayesian variable selection regression model proposed by Guan and Stephens (2011) and implemented in the computer software pimass to quantify aspects of the genetic architecture of male genitalic morphology and female oviposition preference in *Lycaeides* butterflies. With this method, we were able to simultaneously evaluate alternative models with different subsets of SNPs and generate model-averaged parameter estimates. We assumed a linear model for phenotype ($\mathbf{y}$), such that $P(\mathbf{y}|\boldsymbol{\gamma}, \tau, \mu, \boldsymbol{\beta}, \hat{\mathbf{g}}) \sim \mathrm{N}(\mu + \boldsymbol{\beta}_{\boldsymbol{\gamma}}\hat{\mathbf{g}}_{\boldsymbol{\gamma}}, \tau)$. Here, μ is the mean phenotype, $\frac{1}{\tau}$ is the residual variance, $\boldsymbol{\beta}$ is the vector of regression coefficients, $\boldsymbol{\gamma}$ is a vector of binary indicator variables that indicate which SNPs are in the model, and $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is the vector of regression coefficients for SNPs that are in the model (Table 1; Guan and Stephens 2011). $\hat{\mathbf{g}}$ is a matrix that contains the posterior expected value of the genotype for each SNP i and individual j, $\hat{g_{ij}} = \sum_{y=\{0,1,2\}} yP(g_{ij} = y|x_{ij}, p_i, \theta)$. Genetic regions that harbor SNPs statistically associated with phenotypic variation are identified by the posterior distribution of $\boldsymbol{\gamma}$, and the $\boldsymbol{\beta}$ are estimates of the phenotypic effect associated with each SNP. The model contains additional parameters that are estimated from the data and describe higher level aspects of the genetic architecture of the trait (Table 1). These include the proportion of variance explained by the SNPs (*PVE*), the conditional prior probability of a SNP being included in the model (p_{SNP}), the number of SNPs in the regression model (N_{SNP}), and the average phenotypic effect associated with a SNP that is in the model (σ_{SNP}). Importantly, estimates of the regression coefficients ($\boldsymbol{\beta}$) and genetic architecture parameters (*PVE*, N_{SNP}, and σ_{SNP}) incorporate uncertainty regarding which SNPs are associated with trait variation, e.g., $\hat{\beta}_i = \mathrm{E}[\beta_i|\gamma_i = 1]P(\gamma_i = 1)$. Similarly, $\hat{\sigma_{SNP}}$ is the expected value of β_γ, rather than the expected value of $\hat{\beta}$ for the subset of genetic markers with posterior inclusion probabilities greater than an arbitrary threshold.

A second key advantage of the Bayesian variable selection regression method is that the effects of SNPs in the model are controlled for when considering whether additional SNPs should be added to the model (Guan and Stephens 2011). This aspect of

Table 1. Descriptions of key Bayesian variable selection regression and population genomic parameters, analyses, and sets of loci.

Symbol	Description
β	Regression coefficients that describe the additive phenotypic effect associated with each SNP
β_γ	Regression coefficients for the subset of SNPs in the model
PVE	A parameter that describes the proportion of phenotypic variation explained by all of the SNPs in the model
p_{SNP}	The prior probability that a SNP is included in the model
N_{SNP}	The number of SNPs included in the model; this parameter is an estimate of the number of functional variants
σ_{SNP}	A parameter that describes the average additive effect of a SNP included in the model
F_{ST}	A locus-specific evolutionary parameter that describes the expected variance in allele frequencies among populations or species
α	A parameter in the Bayesian genomic cline model describes the expected locus-specific ancestry in hybrids
LN	(*Lycaeides* naive) regression analysis that includes all butterflies with no adjustment for population structure; incorporates interspecific and interpopulation genetic variation
AN	(Admixed naive) regression analysis that includes only butterflies from the admixed populations; incorporates variation segregating in hybrids
LR	(*Lycaeides* population residuals) regression analysis that removes the effects of population structure; only measures intrapopulation genetic variation
PIP	Posterior inclusion probability; we refer to the 0.1% of genetic regions with the highest posterior inclusion probabilities as "highest PIP regions"
$PIP_{0.01}$	The set of genetic regions with a PIP greater than 0.01

the model should reduce (but not negate) the tendency of population structure to result in many false positives, because in model selection, those SNPs already in the model can control for population structure and make the addition of a spuriously associated SNP less likely.

We conducted three analyses to quantitatively describe the genetic basis of each trait: the *Lycaeides* naive (LN) analysis, the admixed naive (AN) analysis, and the *Lycaeides* population residuals (AR) analysis (Table 1). We designed these analyses to test for SNP-by-trait associations and estimate aspects of the genetic architecture of male morphology and oviposition preference in ways that differentially emphasize intraspecific and interspecific variation. In the LN analysis, we included butterflies from all populations without adjusting variables for population structure. This analysis was partially confounded by population structure (genome-average F_{ST} for *L. idas* × *L. melissa* was 0.074; Gompert et al. 2012b), but was more likely to identify SNPs associated with functional variants that differed in frequency among populations or species. And, as stated in the previous paragraph, the effect of population structure on Bayesian variable regression selection is reduced relative to traditional methods that evaluate marker-by-trait associations one at a time. In the AN analysis, we included only butterflies from the five admixed populations (i.e., Jackson Hole *Lycaeides*). This analysis might allow us to identify SNPs associated with functional variants that were segregating in Jackson Hole *Lycaeides*, and that explain species-level phenotypic differences between *L. idas* and *L. melissa*. In general, recombination and independent assortment should erode linkage disequilibrium in hybrids relative to the combined parental populations. This benefit should be especially pronounced in Jackson Hole *Lycaeides*, because these admixed populations exhibit little or no population structure and little variation in hybrid index (genome average F_{ST} for pairs of admixed populations was between 0.002 and 0.004; Gompert et al. 2012b). In the LR analysis, we calculated the difference between the population mean genotype ($\frac{1}{n_j}\sum_j^{n_j} \hat{g_{ij}}$, where the sum is over the individuals in a population) or phenotype and the grand mean genotype ($\bar{\hat{g}_{i.}}$) or phenotype for each SNP, trait, and population, and subtracted this difference from each butterfly's posterior expected genotype ($\hat{g_{ij}}$) or phenotype. We then analyzed the population-adjusted genotypes and phenotypes for butterflies from all populations. This analysis fully removed the confounding effects of population structure, but could only identify SNPs associated with functional variants that explain within population trait variation. We normal quantile transformed the phenotypic data for each trait prior to the analyses as suggested by Guan and Stephens (2011).

We used the computer software `pimass` (Guan and Stephens 2011) to obtain MCMC samples from the joint posterior probability distribution of the model parameters. We placed a uniform prior on log p_{SNP} with lower bound $\log(\frac{1}{N_g})$ and upper bound $\log(\frac{100}{N_g})$, where N_g is the total number of SNPs. For each analysis, we used three 4×10^6 iteration chains. We discarded the first 10^5 iterations as a burn-in, and recorded the parameter values every 400th iteration. We calculated the posterior inclusion probability for each genetic region by estimating the probability that one or more SNPs in the genetic region was associated with

phenotypic variation (we defined a genetic region as the continuous 92 bp DNA sequence that was sequenced from one end of a restriction fragment).

We used genetic region posterior inclusion probabilities to determine whether the same genetic regions were associated with phenotypic variation in different analyses or for different traits. We considered the 0.1% of genetic regions (≈50 genetic regions) with the highest posterior inclusion probabilities for each analysis and trait (hereafter, highest PIP regions). First, for each trait and each pair of analyses, we calculated the number of highest PIP regions identified in both analyses and the number of shared highest PIP regions expected by chance (i.e., assuming independent genetic region-by-trait associations in the different analyses). Likewise, for each analysis and each pair of morphological or behavioral traits, we calculated the number of highest PIP regions found for both traits and the number of shared highest PIP regions expected by chance.

GENOME DIVERGENCE AND GENETIC REGION-BY-TRAIT ASSOCIATIONS

Next, we asked whether genetic regions with the highest posterior inclusion probabilities resided in regions of the genome characterized by specific patterns of genetic differentiation or introgression. Divergent selection causes elevated genetic differentiation in the vicinity of the selected variants, and nonrandom mating or selection in hybrids causes exceptional introgression of chromosomal segments that contain the causal alleles (Barton and Hewitt 1989; Beaumont and Nichols 1996; Nosil et al. 2009; Gompert et al. 2012c). Consequently, an association between the genetic basis of trait variation and genetic differentiation or introgression is expected if trait variation affects fitness in *L. idas*, *L. melissa*, or admixed populations. We previously quantified locus-specific genetic differentiation between *L. idas* and *L. melissa* and introgression in Jackson Hole *Lycaeides* using a subset (17,693 SNPs with a minor allele frequency greater than 0.1) of the sequence data used for association mapping (Gompert et al. 2012b). We quantified genetic differentiation between *L. idas* and *L. melissa* with the evolutionary parameter F_{ST}, in the context of a hierarchical Bayesian F-model (Gompert et al. 2012b). Specifically, we allowed F_{ST} to vary among SNPs, and we modeled the locus F_{ST} parameters dependent on a genome-average F_{ST} and the genome-wide variance in F_{ST}. We quantified introgression with a parameter, α, that specifies an increased (positive values of α) or decreased (negative values of α) probability of locus-specific *L. idas* ancestry in a hybrid relative to null expectations from hybrid index (Gompert et al. 2012b). We estimated α using the Bayesian genomic cline model proposed by Gompert and Buerkle (2011), with modifications described by Gompert et al. (2012b). We took the largest estimates of F_{ST} and $|\alpha|$ for any SNP in a genetic region (i.e., residing in the same restriction fragment) as the estimate of F_{ST} and α for that genetic region.

For each association mapping analysis and trait, we calculated the mean genetic region F_{ST} and $|\alpha|$ for (i) highest PIP regions and (ii) any regions with a posterior inclusion probability greater than or equal to 0.01 (hereafter, $PIP_{0.01}$ regions). We then repeatedly (10,000 times) permuted F_{ST} or $|\alpha|$ values among genetic regions to calculate the probability of obtaining a mean F_{ST} or $|\alpha|$ for these sets of genetic regions that is as high or higher than the observed mean under the null hypothesis that posterior inclusion probabilities and F_{ST} or $|\alpha|$ were independent.

Results

TRAIT VARIATION

We detected large positive correlations among male genitalia length measurements (i.e., F, H, and U; Fig. S1), consistent with previous studies (Nice and Shapiro 1999; Lucas et al. 2008; Gompert et al. 2010). In general, correlations with width and ratio measurements were weaker, but, as expected, ratio measurements were often correlated with one or more of their component measurements. We detected a weak positive correlation between the number of eggs laid on *A. miser* and the number of eggs laid on *M. sativa* (Fig. S2). We found a positive correlation between the proportion of eggs laid on *A. miser* and the number of eggs laid on *A. miser*, and a negative correlation between the proportion of eggs laid on *A. miser* and the number of eggs laid on *M. sativa*. Much of the variation in male genitalic morphology, particularly for length and ratio traits, was partitioned among populations (29.3–89.7% of the variation; Table S2). Considerably less of the variation in male genitalia width traits and oviposition preference was partitioned among populations (less than 14.1%).

GENETIC VARIATION

Genome-wide genetic diversity (θ) was greater, on average, in admixed populations than nonadmixed populations, but this difference was minor (Figs. 3A). The allele frequency distribution in each population was distinctly U-shaped, meaning that most SNPs had one rather common and one rather rare allele (Figs. 3B and S3). Deviations from Hardy–Weinberg and linkage equilibrium (Δ) were generally low, but were higher for SNPs in the same restriction fragment (i.e., genetic region) than SNPs in different genetic regions (Fig. 3C). Evidence of nonzero deviations from Hardy–Weinberg or linkage equilibrium was apparent for a greater number of loci based on the scaled measure of composite disequilibrium (Δ'; Fig. 3D). But, the possible values of Δ are quite restricted by Δ_{MAX} and Δ_{MIN} causing an excess of loci with high Δ', particularly when high- and low-frequency alleles are common (Zaykin 2004). Unlike estimates of Δ, estimates of Δ' were higher, on average, in admixed populations than nonadmixed populations (Fig. 3).

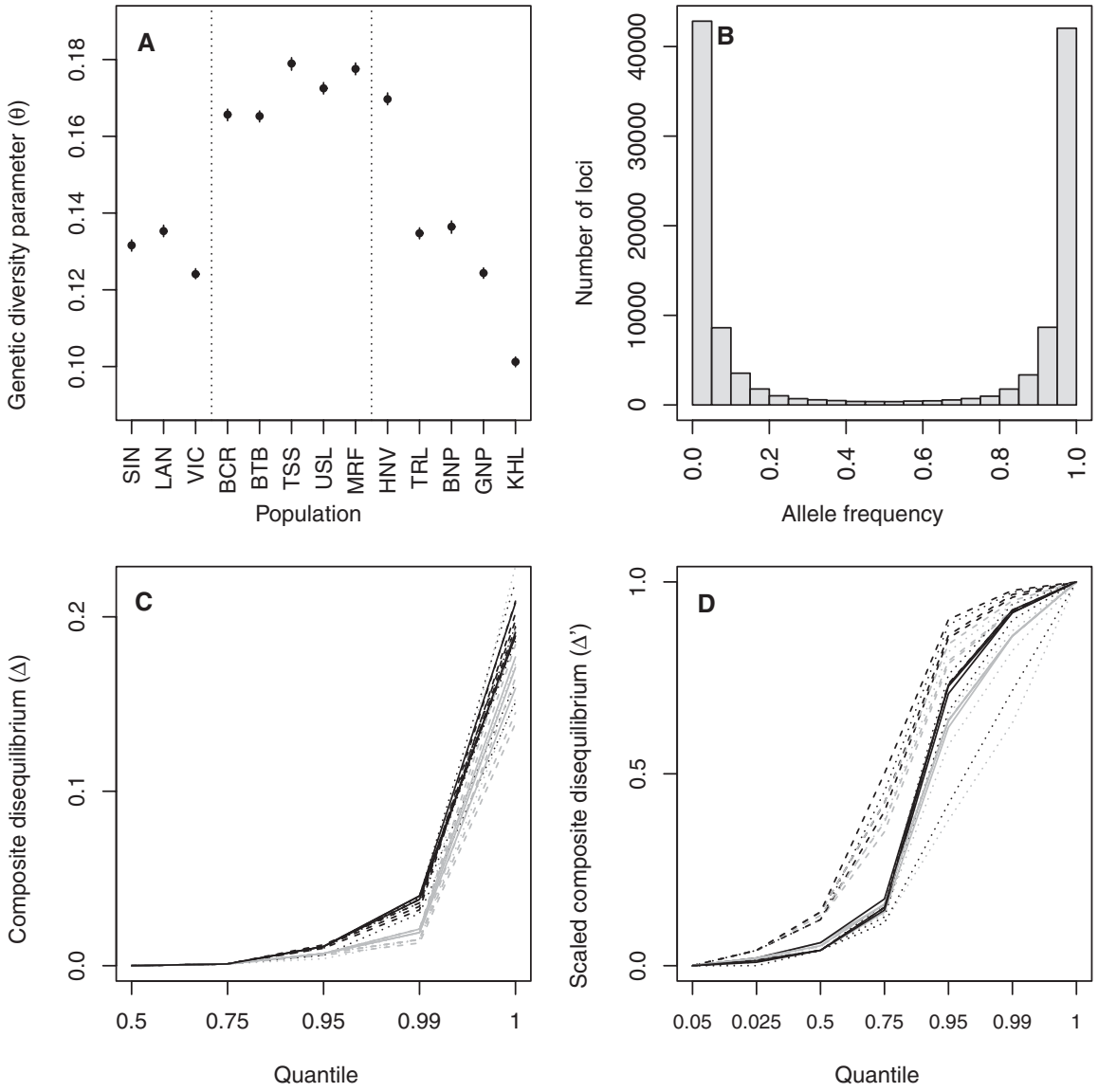

Figure 3. **Plots summarize genetic variation (A) and (B) and disequilibria in *Lycaeides* (C) and (D). Pane A displays estimates of genetic diversity (θ) for each population. We show the 95% ETPI with solid vertical lines. We define population abbreviations in Table S1 . Dotted vertical lines in pane A separate *L. melissa* (left), Jackson Hole *Lycaeides* (center), and *L. idas* (right) populations. Pane B is a histogram of the estimated reference allele frequency for all loci in the BCR population. The allele frequency distributions for the other populations are similar (Fig. S3). Panes C and D depict empirical quantiles of the distribution of pairwise composite disequilibria (Δ; C) and scaled pairwise composite disequilibrium ($\Delta' = \frac{\Delta}{\Delta_{MAX}}$; D): *L. idas* = solid lines; Jackson Hole *Lycaeides* = dashed lines; *L. melissa* = dotted lines. We denote disequilibria quantiles for pairs of variable sites sequenced on a single DNA fragment (i.e., variables sites within 100 bp of each other) with black lines and quantiles for other pairs of loci with gray lines.**

GENETIC ARCHITECTURE OF TRAIT VARIATION

The SNP data explained a considerable proportion of variation in male genitalia length and ratio measurements in the LN analysis ($\widehat{PVE} = 0.329 - 0.818$; Table 2). These traits also differed most among populations (Table S2). The SNP data explained a more modest, but nonnegligible, proportion of variation in male genitalia width measurements and oviposition preference ($\widehat{PVE} = 0.066 - 0.251$). Similarly, the SNP data explained a modest proportion of phenotypic variation in the AN analysis ($\widehat{PVE} = 0.049$–0.241; Table 3), and somewhat less of the phenotypic variation in the LR analysis ($\widehat{PVE} = 0.048$–0.095; Table 4). Point estimates of genetic architecture parameters for most traits implicated a modest number of SNPs (LN, $\hat{N}_{SNP} =$ 13–63; AN, $\hat{N}_{SNP} =$ 9–20; LR, $\hat{N}_{SNP} =$ 11–17; Tables 2–4), with measurable average effects (LN, $\hat{\sigma}_{SNP} = 0.367$–$0.658$; AN, $\hat{\sigma}_{SNP} = 0.443$–$0.638$; LR, $\hat{\sigma}_{SNP} = 0.428$–$0.521$; effect sizes are measured in standard deviations). Importantly, for most traits and analyses, 95% equal-tail probability intervals (ETPIs) for genetic

Table 2. Genetic architecture parameter estimates and 95% ETPI for the naive analysis with all populations (LN analysis). The parameters are the proportion of variance explained (*PVE*), the conditional prior probability of a SNP being in the model (p_{SNP}), the number of SNPs in the model (N_{SNP}), and the average phenotypic effect associated with a SNP in the regression model (σ_{SNP}).

Trait	*PVE*	p_{SNP}	N_{SNP}	σ_{SNP}
F	0.818 (0.764−0.857)	0.00014 (3e-05−0.00052)	16 (4−60)	0.658 (0.376−0.942)
H	0.507 (0.381−0.606)	0.00052 (1e-04−0.00088)	63 (16−99)	0.367 (0.255−0.647)
U	0.621 (0.521−0.701)	0.00023 (3e-05−0.00076)	27 (4−90)	0.488 (0.288−0.939)
W	0.157 (0.029−0.315)	0.00021 (1e-05−0.00082)	26 (2−95)	0.370 (0.131−1.670)
E	0.066 (0.002−0.232)	0.00012 (1e-05−0.00077)	14 (1−91)	0.393 (0.084−2.982)
F/W	0.677 (0.587−0.745)	0.00022 (4e-05−0.00073)	27 (6−87)	0.500 (0.307−0.856)
F/H	0.693 (0.603−0.762)	0.00021 (3e-05−0.00074)	25 (4−88)	0.526 (0.310−0.940)
[F+H]/E	0.508 (0.381−0.614)	0.00022 (2e-05−0.00077)	26 (4−91)	0.484 (0.274−0.968)
H/U	0.329 (0.195−0.475)	0.00015 (1e-05−0.00076)	18 (2−90)	0.456 (0.226−1.148)
Num. *Ast*.	0.251 (0.079−0.407)	0.00029 (2e-05−0.00083)	34 (3−96)	0.422 (0.193−1.303)
Num. *Med*.	0.081 (0.003−0.290)	0.00011 (1e-05−0.00076)	13 (1−90)	0.462 (0.090−3.012)
Prop. *Ast*.	0.180 (0.023−0.358)	0.00026 (2e-05−0.00082)	32 (2−96)	0.395 (0.139−1.558)

Table 3. Genetic architecture parameter estimates and 95% ETPI for the naive analysis with admixed populations (AN analysis). The parameters are described in Table 2.

Trait	*PVE*	p_{SNP}	N_{SNP}	σ_{SNP}
F	0.132 (0.003−0.373)	0.00011 (1e-05−0.00077)	13 (1−92)	0.543 (0.113−3.302)
H	0.241 (0.015−0.489)	0.00013 (1e-05−0.00076)	15 (1−90)	0.638 (0.177−2.22)
U	0.059 (0.001−0.277)	9e-05 (1e-05−0.00076)	11 (1−90)	0.478 (0.091−4.571)
W	0.093 (0.002−0.323)	0.00012 (1e-05−0.00077)	14 (1−91)	0.482 (0.103−3.472)
E	0.074 (0.001−0.296)	1e-04 (1e-05−0.00076)	12 (1−90)	0.487 (0.094−4.497)
F/W	0.102 (0.003−0.333)	0.00014 (1e-05−0.00079)	16 (1−93)	0.466 (0.101−3.209)
F/H	0.065 (0.001−0.293)	1e-04 (1e-05−0.00077)	12 (1−92)	0.474 (0.091−4.105)
[F+H]/E	0.11 (0.003−0.341)	0.00016 (1e-05−0.00079)	20 (1−93)	0.443 (0.103−3.075)
H/U	0.049 (0.001−0.247)	7e-05 (1e-05−7e-04)	9 (1−83)	0.508 (0.093−5.277)
Num. *Ast*.	0.172 (0.007−0.391)	0.00015 (1e-05−0.00079)	17 (1−93)	0.536 (0.144−2.865)
Num. *Med*.	0.11 (0.003−0.343)	0.00017 (1e-05−8e-04)	20 (1−94)	0.443 (0.102−3.052)
Prop. *Ast*.	0.155 (0.005−0.388)	0.00014 (1e-05−0.00077)	17 (1−91)	0.526 (0.139−2.964)

Table 4. Genetic architecture parameter estimates and 95% ETPI for the population-mean-adjusted (LR) analysis. The parameters are described in Table 2.

Trait	*PVE*	p_{SNP}	N_{SNP}	σ_{SNP}
F	0.070 (0.002−0.258)	0.00012 (1e-05−0.00077)	14 (1−92)	0.464 (0.094−2.930)
H	0.093 (0.004−0.280)	0.00014 (1e-05−0.00078)	17 (1−92)	0.482 (0.108−2.739)
U	0.051 (0.001−0.224)	9e-05 (1e-05−0.00075)	11 (1−88)	0.466 (0.091−3.255)
W	0.051 (0.002−0.214)	0.00011 (1e-05−0.00077)	13 (1−91)	0.430 (0.084−3.324)
E	0.048 (0.001−0.209)	9e-05 (1e-05−0.00076)	11 (1−91)	0.450 (0.088−3.367)
F/W	0.065 (0.002−0.239)	9e-05 (1e-05−0.00072)	11 (1−86)	0.521 (0.104−3.454)
F/H	0.095 (0.003−0.312)	0.00014 (1e-05−0.00078)	17 (1−92)	0.458 (0.108−2.632)
[F+H]/E	0.068 (0.002−0.252)	0.00012 (1e-05−0.00077)	15 (1−90)	0.460 (0.095−2.827)
H/U	0.069 (0.002−0.265)	0.00013 (1e-05−0.00076)	16 (1−91)	0.428 (0.093−2.772)
Num. *Ast*.	0.059 (0.002−0.243)	0.00011 (1e-05−0.00078)	14 (1−92)	0.446 (0.087−3.193)
Num. *Med*.	0.090 (0.003−0.284)	0.00014 (1e-05−0.00079)	17 (1−93)	0.468 (0.102−3.374)
Prop. *Ast*.	0.067 (0.002−0.271)	9e-05 (1e-05−0.00076)	11 (1−90)	0.520 (0.096−4.548)

Table 5. The number of genetic regions with posterior inclusion probabilities greater than or equal to the 99.9th empirical quantile for two sets of analyses. The expected number of shared genetic regions if posterior inclusion probabilities for pairs of traits are independent is less than one (approximately $\frac{1}{20}$; LN = all *Lycaeides*, naive analysis; AN = admixed populations, naive analysis; LR = all *Lycaeides*, population-mean-adjusted analysis).

Trait	LN×AN	LN×LR	AN×LR
F	2	2	1
H	4	1	2
U	1	1	6
W	7	6	3
E	6	11	4
F/W	3	5	4
F/H	6	8	5
[F+H]/E	3	5	6
H/U	6	9	5
Num. *Ast*.	9	4	1
Num. *Med*.	12	10	3
Prop. *Ast*.	5	1	2

architecture parameters were quite large. For example, in the AN analysis of F, the 95% ETPIs included 1–92 SNPs (N_{SNP}) with an average effect of 0.113–3.302 (σ_{SNP}).

Model-averaged estimates of the phenotypic effects associated with the SNPs ($\hat{\beta}$) varied among traits and analyses (Fig. S4). For example, considering only SNPs with posterior inclusion probabilities greater than 0.01 (e.g., Table 4), we detected a single SNP with $\hat{\beta}$ greater than 0.5 in the LN analysis of forearm length (F) and many SNPs with quite small effects, but estimates of $\hat{\beta}$ were less than 0.2 for all SNPs in the AN and LR analyses of F (Figs. S4A–C). We identified fewer SNPs with large $\hat{\beta}$ for the proportion of eggs laid on *A. miser* (Figs. S4G–I). The estimated effect of each SNP is an average over models including and excluding the SNP and will generally be less than the average effect of an associated SNP, σ_{SNP} (when a SNP is not in the model $\beta = 0$; Guan and Stephens 2011).

The distribution of posterior inclusion probabilities for genetic regions varied considerably among traits and analyses (Fig. S5), and we based maker-trait association comparisons on the 0.1% of genetic regions (approximately 50) with the highest posterior inclusion probabilities (highest PIP regions). For each trait, highest PIP regions identified in one analysis were identified as highest PIP regions in other analyses more often than expected by chance. Specifically, there were 1–11 shared highest PIP regions for each trait and shared highest PIP regions occurred 17.4–219.8 more times than expected if marker-trait associations were independent among different analyses (Table 5). More highest PIP regions were also shared among traits than expected by chance. This pattern was particularly evident in the LN analysis for male genitalia length measurements, male genitalia ratio measurements and their component length measurements, and oviposition preference measurements (Table S3). This pattern persisted, but was less apparent in the LR analysis (Table 6). Finally, we detected more highest PIP regions that were shared between male genitalic morphology and oviposition preference traits in the AN analysis than in other analyses (Table S4).

GENOME DIVERGENCE AND GENETIC REGION-BY-TRAIT ASSOCIATIONS

Mean F_{ST} for (i) highest PIP regions or (ii) genetic regions with posterior inclusion probabilities greater than or equal to 0.01 ($PIP_{0.01}$ regions) was greater than expected under the null hypothesis of no association between F_{ST} and whether a SNP was associated with male genitalic morphology traits in the LN analysis (Fig. 4, Table 7). Likewise, mean $|\alpha|$ for these genetic regions was greater than expected under this null hypothesis for most male genitalic morphology traits in the LN analysis. Interestingly, several genetic regions with very high F_{ST} and $|\alpha|$ were associated with variation in multiple male genitalic morphology traits in the LN analysis (Fig. 4). We found evidence for an association between F_{ST} or $|\alpha|$ and highest PIP regions for a few specific male genitalia traits in the AN or LR analysis (e.g., W in the LR analysis), but we found little to no evidence of such an association for most male genitalia traits (Table 7). Mean $|\alpha|$ for (i) highest PIP regions, or (ii) $PIP_{0.01}$ regions was greater than expected under the null hypothesis of no association between $|\alpha|$ and whether a SNP was associated with oviposition preference parameters in several instances, but we did not detect a similar pattern for mean F_{ST} (Fig. 4, Table 7).

Discussion

We found that a modest number of SNPs explained a small to large proportion of the variation in male genitalic morphology and oviposition preference and had nonnegligible average effects (Tables 2–4). Although phenotypic variation was best explained by this moderately complex genetic architecture, parameter estimates exhibited considerable uncertainty and in many cases, we could not exclude considerably simpler or more complex genetic architectures. The SNP data explained more variation in male genitalia length and ratio traits in the LN analysis ($\widehat{PVE} \geq 0.329$) than variation in other traits and analyses ($\widehat{PVE} \leq 0.251$). Genitalia length and ratio traits differed more among populations than the genitalia width and oviposition traits (Table S2), and population structure elevates linkage disequilibrium (when treating all populations as a single unit), even among unlinked variants (e.g., Price et al. 2006; Rosenberg and Nordborg 2006). This means that statistical associations in the LN analysis for genitalia length and ratio traits are more likely to occur without physical linkage

Table 6. **The number of genetic regions with posterior inclusion probabilities greater than or equal to the 99.9th empirical quantile for each pair of traits. The expected number of shared genetic regions if posterior inclusion probabilities for pairs of traits are independent is less than one (approximately $\frac{1}{20}$). Genitalic measurements F, H, U, W, E, F/W, F/H, [F+H]/E, and H/U are depicted in Figure 2, and oviposition traits are the number or proportion of eggs laid on *Medicago* or *Astragalus*. These results are for the population-mean-adjusted analysis that includes all *Lycaeides* populations (LR analysis; see Figs. S3 and S4 for the LN and AN analyses).**

	F	H	U	W	E	F/W	F/H	[F+H]/E	H/U	Num.*Ast.*	Num.*Med.*	Prop.*Ast.*
F	52	2	1	0	0	0	0	0	0	1	0	0
H	2	52	1	1	0	0	1	0	0	0	0	0
U	1	1	53	0	0	0	1	0	2	0	0	0
W	0	1	0	53	0	17	0	0	1	0	0	3
E	0	0	0	0	53	1	0	22	0	0	0	0
F/W	0	0	0	17	1	54	0	1	0	0	0	0
F/H	0	1	1	0	0	0	52	0	3	0	0	1
[F+H]/E	0	0	0	0	22	1	0	54	1	0	0	0
H/U	0	0	2	1	0	0	3	1	52	0	0	1
Num. *Ast.*	1	0	0	0	0	0	0	0	0	50	0	2
Num. *Med.*	0	0	0	0	0	0	0	0	0	0	52	2
Prop. *Ast.*	0	0	0	3	0	0	1	0	1	2	2	56

between a SNP and a functional variant than statistical associations in the AN and LR analyses, and this difference could explain differences in the proportion of variance explained for different traits and analyses. Additional factors that might contribute to differences in the proportion of the variation explained for each trait and analysis include differences in heritability or in the proportion of functional variants with infinitesimal effects.

Average effect size estimates support the hypothesis that functional variants with moderate to large effects on male genitalic morphology and oviposition preference exist (e.g., $\hat{\sigma}_{SNP}$ was between 0.367 and 0.638 standard deviations for the AN and LR analyses). Different effect size distributions are expected under different ecological conditions (Kopp and Hermisson 2009). Colonization of a novel host plant species should cause an abrupt change in the plant morphology and chemistry encountered by ovipositing female butterflies, and might select for novel oviposition behaviors. Thus, the oviposition preference genetic architecture parameter estimates are consistent with theory that predicts moderate and large effect alleles to contribute to adaptation following a sudden and discrete change in the optimal phenotype (Orr 1998, 2005). This same theory might explain known large effect alleles associated with adaptation to freshwater lakes in sticklebacks (Colosimo et al. 2005) or discrete mimicry groups in *Heliconius* butterflies (Joron et al. 2006; Reed et al. 2011).

One or more highest PIP genetic regions were shared among the three analyses for each trait (Table 5). These results are consistent with the hypothesis that some of the same functional variants affect phenotypic variation within and among *Lycaeides* populations and species (a similar result was found in *Helianthus*; Lexer et al. 2005). Similarly, and perhaps not surprisingly, some of the same genetic regions were statistically associated with multiple traits, especially correlated traits (Tables 6, S3, and S4). Thus, phenotypic correlations were likely caused, at least in part, by a shared genetic architecture. But, in some cases, we detected major differences in genetic architecture despite phenotypic correlations. For example, despite the strong positive correlation between the male genitalia length traits F and H ($r = 0.732$) and several shared highest PIP genetic regions, genetic architecture parameter estimates from the LN analysis suggest that variation in F was determined by fewer loci with larger effects ($\hat{N}_{SNP} = 16$, 95% ETPI = 4–60; $\hat{\sigma}_{SNP} = 0.66$, 95% ETPI = 0.38–0.94), than variation in H ($\hat{N}_{SNP} = 63$, 95% ETPI = 16–99; $\hat{\sigma}_{SNP} = 0.37$, 95% ETPI = 0.26–0.65). Interestingly, there was more evidence for the same genetic regions being associated with both male genitalic morphology and oviposition preference in the AN analysis than the other analyses. This result might be expected if novel statistical associations among functional variants arose by selection or genetic drift in Jackson Hole *Lycaeides* following admixture.

GENETIC DIFFERENTIATION, INTROGRESSION, AND THE GENETICS OF BARRIERS TO GENE FLOW

Many genetic regions associated with male genitalic variation in the LN analysis were quite differentiated between *L. idas* and *L. melissa* (i.e., high F_{ST}) and contained elevated *L. idas* or *L. melissa* ancestry in Jackson Hole *Lycaeides* (i.e., high $|\alpha|$; Table 7, Fig. 4). This result is consistent with the hypothesis that morphological variation in male genitalia affects fitness and has limited gene flow between *L. idas* and *L. melissa*. Quantitative trait loci for adaptive phenotype differences between species or populations have also been documented in highly differentiated genetic regions in lake whitefish (Rogers and Bernatchez 2005, 2007), *Heliconius* butterflies (Nadeau et al. 2012), and three-spine

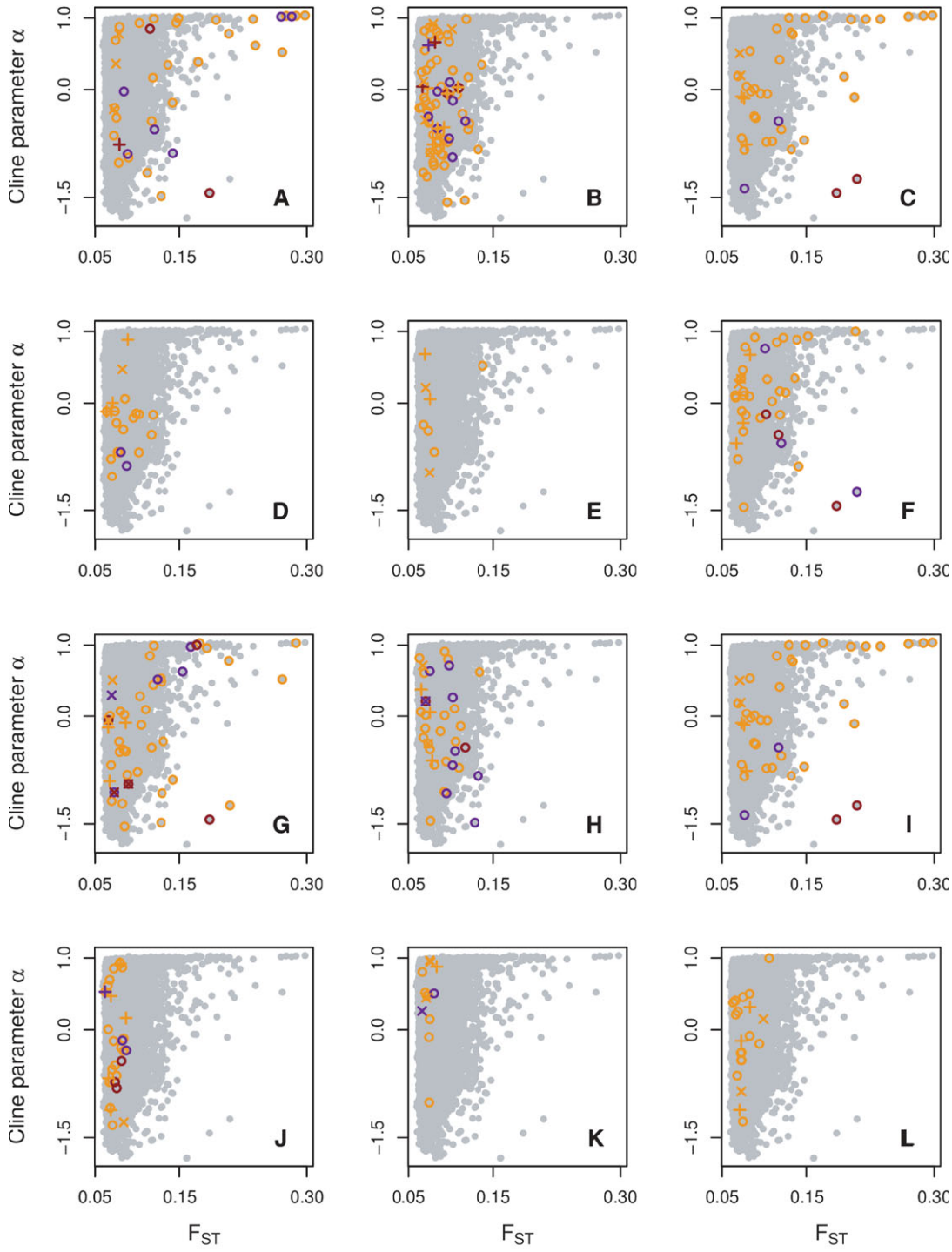

Figure 4. Scatter plots depict the relationship between F_{ST}, cline parameter α, and posterior inclusion probabilities for each genetic region and character: forearm length (A), humerelus length (B), uncus length (C), forearm width (D), elbow width (E), *F* by *W* (F), *F* by *H* (G), *F* plus *H* by *E* (H), *H* by *U* (I), number of eggs on *Astragalus* (J), number of eggs on *Medicago* (K), and proportion of eggs on *Astragalus*. We denote F_{ST} and α for each genetic region with gray closed circles. We use the larger estimate of F_{ST} and α when multiple variable sites occur within a genetic region. We denote genetic regions with modest to high posterior inclusion probabilities with colored symbols, where the symbol gives the analysis and the color gives the strength of evidence: all *Lycaeides*, naive analysis (LN; open circle); admixed populations, naive analysis (AN; +); all *Lycaeides*, population-mean-adjusted analysis (LR; ×); posterior inclusion probability <0.01 (orange), posterior inclusion probability <0.05 (blue), posterior inclusion probability <0.1(dark red).

Table 7. Probability of mean F_{ST} or $|\alpha|$ for genetic regions with posterior inclusion probabilities (i) greater than or equal to the 99.9th empirical quantile [highest PIP regions; $P(F|q)$ or $P(\alpha|q)$], or (ii) greater than or equal to 0.01 [$PIP_{0.01}$ regions; $P(F|p)$ or $P(\alpha|p)$] under the null hypothesis that posterior inclusion probability and F_{ST} or $|\alpha|$ are independent (LN = all *Lycaeides*, naive analysis; AN = admixed populations, naive analysis; LR = all *Lycaeides*, population-mean-adjusted analysis). Probabilities less than 0.05 are in bold font.

	LN				AN				LR			
Trait	$P(F\|q)$	$P(\alpha\|q)$	$P(F\|p)$	$P(\alpha\|p)$	$P(F\|q)$	$P(\alpha\|q)$	$P(F\|p)$	$P(\alpha\|p)$	$P(F\|q)$	$P(\alpha\|q)$	$P(F\|p)$	$P(\alpha\|p)$
F	**0.0000**	**0.0001**	**0.0000**	**0.0000**	0.3946	0.3125	0.2319	0.1819	0.4993	0.8529	0.4026	0.4371
H	**0.0006**	0.4691	**0.0000**	**0.0004**	0.1897	0.4453	0.0614	0.7754	0.1042	0.0629	0.2257	0.1023
U	**0.0000**	**0.0000**	**0.0000**	**0.0000**	0.2249	0.5525	0.3529	0.7220	0.2615	0.3155	0.5441	0.6242
W	**0.0089**	0.5158	**0.0073**	0.5463	0.4908	0.6953	0.5328	0.8516	**0.0329**	0.2441	0.1812	0.4111
E	**0.0033**	0.3766	**0.0196**	0.4779	0.3539	0.1567	0.5057	0.5931	0.3941	0.1939	0.4941	0.2552
F/W	**0.0000**	**0.0323**	**0.0000**	0.0841	**0.0364**	0.7242	0.3857	0.3670	0.2463	0.4281	0.5789	0.6984
F/H	**0.0000**	**0.0000**	**0.0000**	**0.0000**	0.2944	0.3776	0.5138	0.6144	0.4887	**0.0169**	0.5187	0.1912
[F+H]/E	**0.0005**	**0.0072**	**0.0000**	**0.0306**	0.1746	**0.0187**	0.5194	0.6579	0.1678	0.2274	0.7031	0.4830
H/U	**0.0000**	**0.0000**	**0.0000**	**0.0000**	0.2215	0.5533	0.3473	0.7131	0.2547	0.3151	0.5369	0.6262
Num. *Ast*.	0.5771	**0.0225**	0.5286	**0.0257**	0.4541	0.2342	0.6721	0.0792	0.7608	0.4723	0.1637	**0.0154**
Num. *Med*.	0.5407	0.1285	0.6291	0.2973	0.5826	**0.0323**	0.1794	0.1135	0.4945	0.2113	0.6147	0.1105
Prop. *Ast*.	0.3963	0.5252	0.3809	0.4277	0.3536	0.0741	0.3665	0.3121	0.3974	**0.0160**	0.1086	0.3771

sticklebacks (Hohenlohe et al. 2010). This pattern might be expected because male genitalic morphology varies considerably among populations. Thus, allele frequencies must vary among populations at genetic variants that affect male genitalic morphology, and linkage disequilibrium between these functional variants and a subset of sequenced SNPs with similar interpopulation allele frequency differences is expected because of population structure even without physical linkage. But the association with introgression is more remarkable, and is consistent with the hypothesis that a combination of alleles that conferred more *L. idas*-like or *L. melissa*-like male genitalia were favored by selection following admixture in Jackson Hole *Lycaeides*. Similarly, for some oviposition preference traits and analyses, genetic regions associated with trait variation were characterized by elevated *L. idas* or *L. melissa* ancestry in Jackson Hole *Lycaeides* (Table 7, Fig. 4). This result is consistent with the hypothesis that oviposition preference affects migrant or hybrid fitness, and that selection on oviposition preference occurred in Jackson Hole *Lycaeides*. These admixed populations feed on the *L. idas* host plant, *A. miser*, and the mean preference in these populations exceeds that in *L. idas* populations (Gompert et al. 2012a), which could be explained by selection following transgressive segregation in hybrids (e.g., Rieseberg et al. 2003a,b). Conversely, we found little correspondence between genetic regions associated with morphological variation in the AN or LR analyses and patterns of genetic differentiation or introgression (Table 7, Fig. 4). This discordance could mean that, unlike interspecific genitalic variation, intrapopulation variation in male genitalic morphology has little effect on fitness.

Caution is required when interpreting the relationship between trait genetics and genetic differentiation or introgression because (i) we studied a subset of the traits that might affect fitness, and (ii) genetic differentiation and introgression provide only limited information about underlying population genetic processes. If an unmeasured trait was under selection and was correlated with one of the measured traits, the genetic regions associated with variation in the measured trait might exhibit exceptional genetic differentiation even if the measured trait variation was neutral. The possibility for a spurious association between genetic regions that explain trait variation and exceptionally differentiated genetic regions could be exacerbated by population structure, and this phenomenon could explain the concordance between the set of male genitalic morphology highest PIP regions and the set of genetic regions with high F_{ST}. But population structure coupled with selection on unmeasured traits would be much less likely to affect introgression in hybrids, because recombination and independent assortment in an admixed population rapidly erodes statistical associations between physically unlinked functional variants and phenotypes (Buerkle and Lexer 2008). A second consideration is that genetic differentiation and introgression are affected by genetic drift and selection, and drift could contribute substantially to heterogeneous introgression, particularly if in small populations or if selection is weak (Morjan and Rieseberg 2004; Charlesworth 2009; Gompert et al. 2012c). Moreover, the effect of selection on genetic differentiation is determined by linkage disequilibrium and the local recombination rate (Maynard-Smith and Haigh 1974; Gillespie 2000; Hermisson and Pennings 2005), and heterogeneous genetic differentiation might better reflect variation in recombination rates than variation in selection (Noor and Bennett 2009; Neafsey et al. 2010; Turner and Hahn 2010).

LIMITATIONS OF THE CURRENT STUDY

Potential limitations of this study need to be acknowledged. First, the potential confounding effect of population structure in association mapping studies is well known (e.g., Aranzana et al. 2005; Price et al. 2006; Rosenberg and Nordborg 2006). Specifically, if the mean phenotype differs among populations, SNPs with allele frequency differences among populations might be statistically associated with trait variation even if they are not linked to a functional variant. Methods exist to control for population structure (e.g., Price et al. 2006), but these methods often also reduce one's ability to identify SNPs that differ in frequency among populations and are associated with functional variants. Thus, although the LN analysis might allow us to identify genetic regions that control interspecific differences that could not be identified in the LR analysis, these results must be interpreted cautiously because of population structure. Importantly, in the presence of population structure, analyses using Bayesian variable selection regression models should be less likely to grossly overestimate the number of associated SNPs or percent variance explained by the SNPs than single-locus frequentist methods. This is true because each SNP is not tested individually, and the subset of SNPs that covary most with the functional variants will have high model inclusion probabilities and will control for population structure in a manner analogous to including fixed population structure covariates. Nonetheless, the effect of population structure on Bayesian multilocus association mapping requires further theoretical and statistical analysis. So-called spurious associations (i.e., statistical associations not due to tight physical linkage) caused by population structure would still be a serious problem if we were interested in pursuing the identity of functional variants. But this was not the aim of our study and would not be very practical based on the genomic resources currently available for *Lycaeides* (we currently lack a recombination-based or physical genetic map for *Lycaeides*). Instead, we were interested in quantifying the number and effects of functional variants, which simply required strong statistical associations between SNPs and functional variants; the cause(s) of these statistical associations were not really relevant.

A second limitation of the current study is that the posterior estimate of many genetic architecture model parameters included substantial uncertainty. This uncertainty reflects the inherent complexity of genome-wide association mapping and the modest number of individuals (fewer than 200 for each trait) and SNPs (about 118,000) we were able to analyze. Additional SNPs and individuals could be included in future studies, but considerable uncertainty might still persist if trait variation is affected by both (i) functional variants with measurable, even large, effects (as suggested by this study), and (ii) many additional functional variants with infinitesimal effects (e.g., Rockman 2012). Lastly, we quantified aspects of genetic basis of two sets of traits that probably constitute barriers to gene flow between *L. idas* and *L. melissa* in the wild, but other traits, such as diapause timing and wing pattern, might also contribute to reproductive isolation (Gompert et al. 2010, 2012a). Indeed, the contribution of each inherent barrier to gene flow likely varies among populations and generations because of genetic differences and spatial or temporal heterogeneity in the environment. Thus, even if an allele explains much of the phenotypic variation that causes a specific barrier to gene flow, it might not explain much of the variation in reproductive isolation.

Conclusion

We found evidence that functional variants with moderate to large effects probably contribute to variation in male genitalic morphology and oviposition preference in *Lycaeides* butterflies. The frequency of these functional variants in natural populations is likely determined by multiple population genetic processes including genetic drift, natural selection, and recombination. For example, the population genetic patterns we described are consistent with the hypothesis that some functional variants affecting male genitalic morphology have experienced divergent selection in nonadmixed populations, but also indicate that genetic drift has likely contributed to genome-wide variation in genetic differentiation. A full understanding of the extent that selection on specific traits translates into reproductive isolation and affects genome divergence (as a whole and in specific genetic regions) requires detailed knowledge of the genetic basis of the many traits that determine fitness. This is a potentially daunting task, which we have barely begun. But these initial results suggest that individual functional variants could be important, and that selection on individual variants might cause localized genetic differentiation during speciation. Nonetheless, because of uncertainty in parameter estimates and the limitations described in the previous section, an alternative possibility cannot yet be entirely excluded: fitness in *Lycaeides* could be determined by a multitude of functional variants, most with infinitesimal effects, and selection would have had little effect on local genetic differentiation.

Functional variants that contribute to local adaptation or reproductive isolation are known and many of these have major effects on a phenotype or fitness (e.g., Mihola et al. 2009; Tang and Presgraves 2009; Barr and Fishman 2010; Nosil and Schluter 2011). But many of these so-called speciation genes affect traits that are effectively discrete, and it is not clear whether these results can be generalized to more continuous trait variation (Rockman 2012). Our results suggest that barriers to gene flow caused by continuous, quantitative phenotypic variation might also be affected by moderate or large effect variants. But we do not know whether the unexplained variation in each trait we studied was caused by plasticity or many variants with infinitesimal effects. We believe that progress in speciation genetics requires that we

know the relative contribution to reproductive isolation of functional variants with measurable and infinitesimal effects, and that more research is needed to address this question. If moderate and large effect alleles are generally important in the evolution of reproductive isolation, then we can understand the genetics of speciation by identifying these variants and studying their histories and functions. But if a multitude of infinitesimals is more important in speciation, we learn much less by studying individual variants. Instead, it might be more productive to ask questions regarding why the genetic architecture of reproductive isolation is frequently so diffuse, and to focus on statistical descriptions of these genetic architectures, such as the distribution of effect sizes or the frequency distribution of epistatic interactions among functional variants.

ACKNOWLEDGMENTS

This research was facilitated by J. Fordyce, M. Forister, Y. Guan, H. Harlow, C. Hendrix, A. Krist, 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, and DBI-0701757 and DEB-1050149 to CAB). The authors have no conflict of interest to declare.

LITERATURE CITED

Aranzana, M. J., S. Kim, K. Y. Zhao, E. Bakker, M. Horton, K. Jakob, C. Lister, J. Molitor, C. Shindo, C. L. Tang, et al. 2005. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genet. 1:531–539.

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., and G. M. Hewitt. 1989. Adaptation, speciation and hybrid zones. Nature 341:497–503.

Bates, D., M. Maechler, and B. Bolker. 2011. lme4: linear mixed-effects models using S4 classes. Available via http://CRAN.R-project.org/package=lme4. R package version 0.999375-42.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. Roy. Soc. B—Biol. Sci. 263:1619–1626.

Bradshaw, H., and D. Schemske. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature 426:176–178.

Buerkle, C. A., and C. Lexer. 2008. Admixture as the basis for genetic mapping. Trends Ecol. Evol. 23:686–694.

Buerkle, C. A., and L. H. Rieseberg. 2001. Low intraspecific variation for genomic isolation between hybridizing sunflower species. Evolution 55:684–691.

Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, and A. D. Long. 2010. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 467:587–590.

Bustamante, C. D., A. Fledel-Alon, S. Williamson, R. Nielsen, M. T. Hubisz, S. Glanowski, D. Tanenbaum, T. White, J. Sninsky, R. Hernandez, et al. 2005. Natural selection on protein-coding genes in the human genome. Nature 437:1153–1157.

Carbonetto, P., and M. Stephens. 2012. Scalable variational inference for Bayesian variable selection in regression, and its accuracy in genetic association studies. Bayesian Anal. 7:73–107.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10:95–205.

Cho, Y. S., M. J. Go, Y. J. Kim, J. Y. Heo, J. H. Oh, H.-J. Ban, D. Yoon, M. H. Lee, D.-J. Kim, M. Park, et al. 2009. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat. Genet. 41:527–534.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307:1928–1933.

Coyne, J. A., S. Simeonidis, and P. Rooney. 1998. Relative paucity of genes causing inviability in hybrids between *Drosophila melanogaster* and *D*. simulans. Genetics150:1091–1103.

Dufková, P., M. Macholan, and J. Pialek. 2011. Inference of selection and stochastic effects in the house mouse hybrid zone. Evolution 65:993–1010.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford Univ. Press, Oxford, UK.

Flint, J., and T. F. C. Mackay. 2009. Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res. 19:723–733.

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.

Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi. 2009. GNU Scientific Library: Reference Manual. Network Theory Ltd.

Gillespie, J. 1984. Molecular evolution over the mutational landscape. Evolution 38:1116–1129.

———. 1994. The causes of molecular evolution. Oxford Univ. Press, Oxford, UK.

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.

———. 2011. Bayesian estimation of genomic clines. Mol. Ecol. 20:2111–2127.

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., L. K. Lucas, J. A. Fordyce, M. L. Forister, and C. C. Nice. 2010. 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., L. K. Lucas, C. C. Nice, J. A. Fordyce, C. A. Buerkle, and M. L. Forister. 2012a. Geographically variable, multifarious phenotypic divergence during the speciation process. Ecol. Evol. *In press*.

Gompert, Z., L. K. Lucas, C. C. Nice, J. A. Fordyce, M. L. Forister, and C. A. Buerkle. 2012b. Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. Evolution 66:2167–2181.

Gompert, Z., T. L. Parchman, and C. A. Buerkle. 2012c. Genomics of isolation in hybrids. Philos. Trans. R. Soc. B Biol. Sci. 367:439–450.

Good, J. M., M. A. Handel, and M. W. Nachman. 2008. Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. Evolution 62:50–65.

Grant, P. R., and B. R. Grant. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. Science 296:707–711.

Guan, Y., and M. Stephens. 2011. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. Ann. Appl. Stat.5:1780–1815.

Hermisson, J., and P. S. Pennings. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335–2352.

Hirschhorn, J. N., and M. J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6:95–108.

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.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, et al. 2012. The genomic basis of adaptive evolution in three spine sticklebacks. Nature 484: 55–61.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477:203–206.

Joron, M., R. Papa, M. Beltran, N. Chamberlain, J. Mavarez, S. Baxter, M. Abanto, E. Bermingham, S. J. Humphray, J. Rogers, et al. 2006. A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. PLoS Biol. 4:1831–1840.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge Univ. Press, Cambridge, UK.

Kopp, M., and J. Hermisson. 2009. The genetic basis of phenotypic adaptation II: the distribution of adaptive substitutions in the moving optimum model. Genetics 183:453–1476.

Lexer, C., D. M. Rosenthal, O. Raymond, L. A. Donovan, and L. H. Rieseberg. 2005. Genetics of species differences in the wild annual sunflowers, *Helianthus annuus* and *H*petiolaris. Genetics 169:2225–2239.

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.

Linnen, C. R., E. P. Kingsley, J. D. Jensen, and H. E. Hoekstra. 2009. On the origin and spread of an adaptive allele in deer mice. Science 325:1095–1098.

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.

Mackay, T. F. C. 2001. The genetic architecture of quantitative traits. Annu. Rev. Genet. 35:303–339.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, et al. 2009. Finding the missing heritability of complex diseases. Nature 461:747–753.

Maynard-Smith, J., and J. Haigh. 1974. Hitch-hiking effect of a favorable gene. Genet. Res. 23:23–35.

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.

Morjan, C. L., and L. H. Rieseberg. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol. Ecol. 13:1341–1356.

Nabokov, V. 1949. The nearctic members of *Lycaeides* Hübner (Lycaenidae, Lepidoptera). Bull. Museum Comp. Zool. 101:479–541.

Nadeau, N. J., A. Whibley, R. T. Jones, J. Davey, C. K. Dasmahapatra, S. W. Baxter, M. Joron, R. H. ffrench-Constant, M. Blaxter, J. Mallet, et al. 2012. Evidence for genomic islands of divergence among hybridizing *Heliconius* butterflies obtained by large-scale targeted sequencing. Philos. Trans. R. Soc. B Biol. Sci. 367:343–353.

Neafsey, D. E., M. K. N. Lawniczak, D. J. Park, S. N. Redmond, M. B. Coulibaly, S. F. Traore, N. Sagnon, C. Costantini, C. Johnson, R. C. Wiegand, et al. 2010. SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. Science 330:514–517.

Nice, C. C., J. A. Fordyce, A. M. Shapiro, and R. Ffrench-Constant. 2002. Lack of evidence for reproductive isolation among ecologically specialised lycaenid butterflies. Ecol. Entomol. 27:702–712.

Nice, C. C., and A. M. Shapiro. 1999. Molecular and morphological divergence in the butterfly genus *Lycaeides* (Lepidoptera: Lycaenidae) in North America: evidence of recent speciation. J. Evol. Biol. 12:936–950.

Nielsen, R., C. Bustamante, A. G. Clark, S. Glanowski, T. Sackton, M. Hubisz, A. Fledel-Alon, D. Tanenbaum, D. Civello, T. White, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3:976–985.

Noor, M. A. F., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103:439–444.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375–402.

Nosil, P., Z. Gompert, T. A. Farkas, A. Comeault, J. L. Feder, C. A. Buerkle, and T. L. Parchman. 2012. Genomic consequences of multiple speciation processes. Proc. R. Soc. B Biol. Sci. doi: 10.1098/rspb.2012.0813. [Epub ahead of print].

Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. Trends Ecol. Evol. 26:160–167.

Orr, H. A. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution 52:935–949.

———. 2002. The population genetics of adaptation: the adaptation of DNA sequences. Evolution 56:1317–1330.

———. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Parchman, T. L., Z. Gompert, F. Schilkey, C. W. Benkman, and C. A. Buerkle. 2012. Genome-wide association genetics of cone serotiny in lodgepole pine. Mol. Ecol. 21:2991–3005.

Peltola, T., P. Marttinen, A. Jula, V. Salomaa, M. Perola, and A. Vehtari. 2012. Bayesian variable selection in searching for additive and dominant effects in genome-wide data. PLoS ONE 7:e29115.

Pfaff, C., E. Parra, C. Bonilla, K. Hiester, P. McKeigue, M. Kamboh, R. Hutchinson, R. Ferrell, E. Boerwinkle, and M. Shriver. 2001. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am. J. Hum. Genet. 68:198–207.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909.

Pritchard, J. K., J. K. Pickrell, and G. Coop. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. 20:R208–R215.

Purcell, S. M., N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan, P. Sklar, D. M. Ruderfer, J. L. Moran, M. J. Daly, et al. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460:748–752.

R Development Core Team. 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reed, R. D., R. Papa, A. Martin, H. M. Hines, B. A. Counterman, C. Pardo Diaz, C. D. Jiggins, N. L. Chamberlain, M. R. Kronforst, R. Chen, et al. 2011. *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333:1137–1141.

Rieseberg, L. H., and C. A. Buerkle, 2002. Genetic mapping in hybrid zones. Am. Nat. 159:S36–S50.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. 2003a. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211–1216.

Rieseberg, L. H., A. Widmer, A. M. Arntz, and J. M. Burke. 2003b. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. Philos. Trans. R. Soc. Lond. B 358:1141–1147.

Rockman, M. V. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66:1–17.

Roesti, M., A. P. Hendry, W. Salzburger, and D. Berner. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. Mol. Ecol. 21:2852–2862.

Roff, D. A., and T. A. Mousseau. 1987. Quantitative genetics and fitness - lessons from *Drosophila*. Heredity 58:103–118.

Rogers, S. M., and L. Bernatchez. 2005. Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). Mol. Ecol. 14:351–361.

Rogers, S. M., and L. Bernatchez. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonas* sp Salmonidae) species pairs. Mol. Biol. Evol. 24:1423–1438.

Rosenberg, N. A., and M. Nordborg. 2006. A general population-genetic model for the production by population structure of spurious genotype-phenotype associations in discrete, admixed or spatially distributed populations. Genetics 173:1665–1678.

Singer, M. C., D. Ng, and C. D. Thomas. 1988. Heritability of oviposition preference and its relationship to offspring performance within a single insect population. Evolution 42:977–985.

Steiner, C. C., J. N. Weber, and H. E. Hoekstra. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol. 5:e219.

Stinchcombe, J. R., C. Weinig, M. Ungerer, K. M. Olsen, C. Mays, S. S. Halldorsdottir, M. D. Purugganan, and J. Schmitt. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. Proc. Natl. Acad. Sci. USA 101:4712–4717.

Szymura, J. M., and N. H. Barton. 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.

Teotonio, H., I. M. Chelo, M. Bradic, M. R. Rose, and A. D. Long. 2009. Experimental evolution reveals natural selection on standing genetic variation. Nat. Rev. Genet. 41:251–257.

Turner, T. L., and M. W. Hahn. 2010. Genomic islands of speciation or genomic islands and speciation? Mol. Ecol. 19:848–850.

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. B. Biol. Sci. 278:2737–2744.

Wade, M. J., N. A. Johnson, R. Jones, V. Siguel, and M. McNaughton. 1997. Genetic variation segregating in natural populations of *Tribolium castaneum* affecting traits observed in hybrids with *T* freemani. Genetics147:1235–1247.

Weber, K., R. Eisman, L. Morey, A. Patty, J. Sparks, M. Tausek, and Z. Zeng. 1999. An analysis of polygenes affecting wing shape on chromosome 3 in *Drosophila melanogaster*. Genetics 153:773–786.

Weir, B. S. 1979. Inferences about linkage disequilibrium. Biometrics 35:235–254.

Weiss, K. M. 2008. Tilting at quixotic trait loci (QTL): an evolutionary perspective on genetic causation. Genetics 179:1741–1756.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:0097–0159.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565–569.

Yeaman, S., and M. C. Whitlock. 2011. The genetic architecture of adaptation under migration-selection balance. Evolution 65:1897–1911.

Zaykin, D. V. 2004. Bounds and normalization of the composite linkage disequilibrium coefficient. Genet. Epidemiol. 27:252–257.

Associate Editor: P. Nosil

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1.

Table S1. Population sample information (JH = Jackson Hole *Lycaeides*; N_{male} = male sample size; N_{female} = female sample size).

Table S2. Proportion of phenotypic variance explained by population.

Table S3. The number of genetic regions with posterior inclusion probabilities greater than or equal to the 99.9th empirical quantile for each pair of traits.

Table S4. The number of genetic regions with posterior inclusion probabilities greater than or equal to the 99.9th empirical quantile for each pair of traits.

Figure S1. Histograms summarize the variation for each morphological trait (diagonal) and scatter plots depict the covariance between pairs of characters (off-diagonal; light gray = *L. idas*, gray = Jackson Hole *Lycaeides*, black = *L. melissa*).

Figure S2. Histograms summarize the variation for each oviposition preference trait (diagonal) and scatter plots depict the covariance between pairs of characters (off-diagonal; light gray = *L. idas*, gray = Jackson Hole *Lycaeides*, black = *L. melissa*).

Figure S3. Histograms depict the reference allele frequency distribution for all loci and each population. We define population abbreviations in Table S1.

Figure S4. Histograms depict estimated effect sizes for SNPs with posterior inclusion probabilities greater than 0.01.

Figure S5. Plots depict genetic region posterior inclusion probabilities for forearm length (A) and (B), humerelus length (C), and the proportion of eggs laid on *Astragalus* (D).