ORIGINAL ARTICLE

HYBRID SPECIATION AND INDEPENDENT EVOLUTION IN LINEAGES OF ALPINE BUTTERFLIES

Chris C. Nice,[1,2] Zachariah Gompert,[3] James A. Fordyce,[4] Matthew L. Forister,[5] Lauren K. Lucas,[3] and C. Alex Buerkle[3]

[1]Department of Biology, Population and Conservation Biology Program, Texas State University, San Marcos, Texas 78666

  [2]E-mail: ccnice@txstate.edu

[3]Department of Botany, University of Wyoming, Laramie, Wyoming 82071

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

[5]Department of Biology/MS 314, University of Nevada, Reno, Nevada 89557

The power of hybridization between species to generate variation and fuel adaptation is poorly understood despite long-standing interest. There is, however, increasing evidence that hybridization often generates biodiversity, including via hybrid speciation. We tested the hypothesis of hybrid speciation in butterflies occupying extreme, high-altitude habitats in four mountain ranges in western North America with an explicit, probabilistic model, and genome-wide DNA sequence data. Using this approach, in concert with ecological experiments and observations and morphological data, we document three lineages of hybrid origin. These lineages have different genome admixture proportions and distinctive trait combinations that suggest unique and independent evolutionary histories.

Hybridization is often detrimental and maladaptive, such as when hybrid individuals suffer reduced fitness. Alternatively, hybridization can facilitate gene and trait exchange, and hybrid individuals sometimes experience greater fitness than parental forms, especially in novel, and often extreme, environments (Grant and Grant 1992; Rieseberg et al. 2003a; Arnold 2006; Mallet 2007; Fitzpatrick and Shaffer 2007). One of the most dramatic outcomes of hybridization is homoploid hybrid speciation (HHS), the formation of an isolated lineage of hybrid origin without a change in base chromosome number (Buerkle et al. 2000; Mallet 2007; Abbott et al. 2010). Cases of HHS constitute windows on the mechanisms of genomic evolution during speciation because the extant parental species provide a basis for comparison. Cases of HHS also offer excellent opportunities to examine divergence in the face of gene flow and to investigate the contribution of ecological processes to speciation.

The propensity for hybridization to generate genetic and evolutionary novelty, including hybrid species, and the fate of that novelty, is often determined largely by the ecological context in which hybridization occurs (Buerkle et al. 2000; Rieseberg et al. 2003b; Martin et al. 2006; Taylor et al. 2009). Although there has long been debate about the relative fitness of hybrid genotypes in parental and novel habitats, it is clear that the ecological context is of fundamental importance (Arnold and Martin 2010). For example, substantial ecological differences between the parent species' habitat and the habitat of the hybrids are predicted to facilitate the formation of species of hybrid origin (Buerkle et al. 2000, 2003). Empirical data support this prediction: hybrid species are commonly ecologically differentiated and associated with novel or even extreme environments relative to the parent species (Rieseberg et al. 2003b; Gompert et al. 2006; Abbott et al. 2010; Kunte et al. 2011).

When considering the contribution of ecological processes to hybrid speciation, there is the additional possibility of repeated origins of hybrid species leading to multiple isolated lineages of hybrid origin (multiple origins) (Schwarzbach and Rieseberg 2002; Gross et al. 2007; Abbott et al. 2010). The evolution of multiple hybrid lineages from the same parental species, driven by ecological factors, has been investigated in several plant systems (Wang et al. 2001; Schwarzbach and Rieseberg 2002; Gross et al. 2003; Rieseberg et al. 2003b; Abbott et al. 2010). For example, in *Helianthus deserticola*, there is evidence of as many as three origins, presumably involving three admixture events (Gross et al. 2007). It is also possible that multiple disjunct populations of hybrid origin could become isolated following a single hybridization event. This might occur if segments of an admixed lineage became isolated in separate populations, perhaps associated with islands, prior to genome stabilization (Buerkle and Rieseberg 2008). These scenarios (single or multiple origins) present opportunities to study adaptation and differentiation among admixed populations. Hybrid populations can evolve either independently in differing selective environments, in concert with selection fixing adaptive alleles in each population, or in parallel in similar selective environments (Gross et al. 2007). Investigations of multiple hybrid populations provide a framework for understanding how populations or species maintain cohesion or undergo differentiation. Here, we investigate four lineages of butterflies, each putatively arising from hybridization between the same parental species, and each occupying high-altitude habitats in different mountain ranges in western North America.

Hybridization between the butterfly species *Lycaeides melissa* and *L. anna* (formerly *L. idas anna*) resulted in an isolated hybrid lineage in the Sierra Nevada of California and Nevada that was documented with multiple molecular marker systems and investigations of morphological and ecological variation (Gompert et al. 2006). *Lycaeides melissa* populations generally occur in Great Basin habitats on the east side of the Sierra Nevada and the species range extends to southern Canada and the central USA, whereas *L. anna* populations occupy wet meadows at mid-elevation on the west slope of these mountains and in isolated populations in the central Coast Range of California (Scott 1986; Nice and Shapiro 1999; Guppy and Shepard 2001; Nice et al. 2002, 2005; Gompert et al. 2006). The hybrid lineage in the Sierra Nevada occupies a novel, alpine (above treeline, oreal) habitat, utilizes an alpine-endemic host plant, and exhibits both intermediate and transgressive traits (Nice et al. 2002; Fordyce and Nice 2003; Gompert et al. 2006). Some of these traits, such as a unique absence of egg adhesion to the larval host plant and strong female oviposition preference for their natal host plant, appear to be adaptive in the hybrid lineage's environment and might contribute to reproductive isolation (Nice et al. 2002; Fordyce
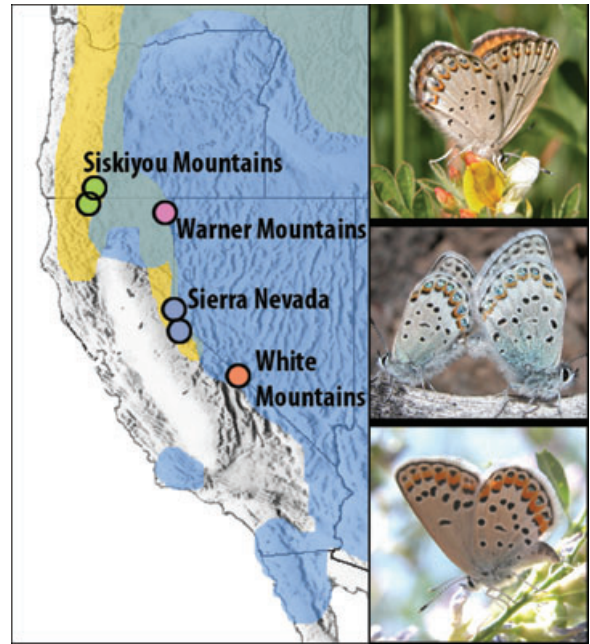
Figure 1. Left: Map depicting the location of putative hybrid taxa (colored circles) in western North America. Each circle represents a sampled population. The approximate geographic ranges of *Lycaeides anna* (yellow) and *L. melissa* (blue) are shown, the area of range overlap is shown as intermediate. Right: *L. anna* (top), the hybrid lineage from the Sierra Nevada (middle), and *L. melissa* (bottom).

and Nice 2003; Gompert et al. 2006). These observations suggest an important role for ecological factors in the evolution of the Sierran hybrid lineage, as predicted by theory (Buerkle et al. 2000). Additional investigations have detected possible evidence of genetic admixture and unique morphological and ecological variation in high-altitude populations of *Lycaeides* in three other nearby mountain ranges (Nice et al. 2005; Gompert et al. 2008a,b, 2010b) (Fig. 1, Table 1).

In the present study, we first used genome-wide DNA sequence variation to test the hypothesis of hybrid speciation in each of these lineages. In doing this, we use an explicit, model-based test of the hypothesis of HHS, which contrasts with previous studies (e.g., Wang et al. 2001; Gompert et al. 2006; Mavárez et al. 2006) that have used marker or phenotypic additivity to detect patterns that are consistent with HHS. We used this model-based approach to estimate historical, demographic parameters for each lineage, including initial genomic admixture proportions and time of the admixture (speciation) event. We also measured contemporary genetic differentiation among lineages and estimated contemporary admixture proportions in each lineage. We then quantified and summarized ecological and morphological variation (including reanalysis of data from previous studies) to address whether these lineages exhibit similar traits that are potentially adaptive in alpine habitats or, alternatively, exhibit measurable

Table 1. Summary of ecological and morphological differentiation among four putative hybrid lineages and the parental species, *L. anna* and *L. melissa*.

Lineage/character	*L. anna*	Sierra Nevada	Siskiyou Mts.	Warner Mts.	White Mts.	*L. melissa*
(A) Larval host plant	*Lupinus, Lotus*	*Astragalus*	*Lupinus*	*Astragalus*	*Astragalus*	*Medicago* and low-elevation *Astragalus*
(B) Host fidelity	Moderate	High	?	High	High	Low-moderate
(C) Egg morphology	*L. anna*	intermediate	?	similar to *L. anna*	similar to *L. anna*	*L. melissa*
(D) Egg adhesion	Yes	No	Yes	Yes	No	Yes
(E) Male morphology	*L. anna*	Intermediate	*L. anna*	Intermediate	Intermediate	*L. melissa*
(F) Wing pattern	*L. anna*	*L. melissa*	Reduced	*L. melissa*	*L. melissa*	*L. melissa*

(A) Larval host plant is the genus of host plant used by females of each lineage for oviposition in the wild (CCN, MLF, JAF, ZG, LKL, CAB, pers. obs.) and on which larvae feed. (B) Host fidelity was assessed by oviposition preference experiments (Nice et al. 2002; Gompert et al. 2006; Fordyce et al. 2011). "Low" indicates little to no significant female preference for their natal host (i.e., the host listed as "Larval host plant"); "moderate" indicates significant preference exhibited by the females, but a substantial proportion of ovipositions observed on alternative hosts; "high" host fidelity indicates results of experiments where females laid > 80 % of their total eggs on the natal host plant. (C) Egg morphology was examined using electron microscopy (Forister et al. 2008). (D) Egg adhesion refers to whether eggs are strongly adhered to the larval host plant when they are laid. These data are a summary of the results of experiments that measured the force required to dislodge eggs from plants (Fordyce and Nice 2003) and from direct observations of egg adhesion during oviposition preference experiments (Nice et al. 2002; Gompert et al. 2006; Fordyce et al. 2011) and direct field observations. (E) Male genitalic morphology was assessed through morphometric analyses (Nice et al. 2005; Lucas et al. 2008) and reanalyzed here. (F) Wing morphology was assessed through principle component analysis of wing pattern character areas (Fig. S2). Wing patterns and the form of the male genitalia were originally used to delineate the parental species (Nabokov 1949) and were classified here based on similarity to the parental species. For details on ecological and morphological variation, see Materials and Methods.

variation reflective of ecological differentiation and possibly local adaptation.

Our large, multilocus DNA sequence dataset demonstrates that three of the four putative hybrid lineages in *Lycaeides* have admixed genomes composed of alleles from *L. anna* and *L. melissa*. Tests of alternative models of bifurcating and reticulating (hybrid) speciation events provide support for the hypothesis of hybrid speciation in these three admixed lineages. Parameter estimates under the model of hybrid speciation in combination with ecological and morphological measurements suggest that the hybrid lineages differ in their genomic structure and are morphologically and ecologically differentiated. We discuss the possibility of independent evolution in each of the lineages that might be consistent with the hypothesis of local adaptation in each mountain range. This work contributes to our understanding of hybrid speciation and constitutes the first, explicit test of the hypothesis of HHS based on a modeling approach and using a large, multilocus dataset.

Methods
MOLECULAR DATA AND GENOME-WIDE POPULATION DIFFERENTIATION

Sampling and molecular methods used to generate genome-wide sequence data, and analytical details for next-generation data analysis, are described in more detail elsewhere (Forister et al. 2011; Gompert et al. 2010a; Gompert and Buerkle 2011). Sequence data in this manuscript are for DNA samples from seven *Lycaeides* populations (Table S1, Fig. S1), each generated by pooling genomic DNA from 15 individuals. We used a restriction fragment-based procedure to amplify a subset of the genome for high-throughput sequencing (the approach we used is similar to the CRoPS method [van Orsouw et al. 2007]). We used the restriction enzymes EcoRI and MseI to create fragments to which unique 10 bp barcodes or multiplex identifier sequences (MIDs) were ligated along with sequencing adaptors. At the time of library preparation, the number of available MID sequences was limited, thus pooling individual DNA samples from populations was necessary because individual indexing was not possible. (Current methods include individual indexing; see Gompert et al. 2012 and Parchman et al. 2012.) The fragments were then amplified with PCR using iProof high-fidelity DNA polymerase (BioRad Laboratories Life Science, Hercules, CA) to reduce PCR error and primers complimentary to the adaptors. PCR products were separated on a 2% agarose gel and fragments between 400 and 550 bp were excised from the gel. Fragments were purified using the GENECLEAN Turbo DNA purification kit (MP Biomedicals, LLC, Santa Ana, CA). (For more details, see Gompert et al. 2010a.) This template was sequenced by SeqWright DNA Technology Services (Houston, TX) using the 454 GS XLR70 Titanium platform. Assembly of the resulting sequence reads was performed using the SeqMan Ngen sequence assembler v2.0.0 (DNASTAR) software.

All sites with two or more observed states were treated as variable. This approach is clearly overly simplistic (see Li et al. 2009), however, in the absence of systematic errors, this assumption is unlikely to bias results. We estimated genome-wide genetic differentiation between each pair of populations (measured as the

proportion of genetic variation partitioned between populations, ϕ_{ST} [Excoffier et al. 1992]), using a hierarchical Bayesian model (Gompert et al. 2010a; Gompert and Buerkle 2011). This model assumes that the population-level haplotype counts for each locus are distributed as a multivariate Pólya distribution (i.e., a Dirichlet compound multinomial distribution) with a vector of parameters comprised of the population haplotype (allele) frequencies. The use of this distribution, while integrating over all possible values of allele frequencies, allows estimation of allele frequencies while also accounting for the uncertainty associated with sampling sequences from individuals and sampling individuals from populations. Locus-specific ϕ_{ST} values are assumed to be drawn from a normal distribution with a mean and variance equal to the genome-level ϕ_{ST} (see eqs. 2–4 in Gompert et al. 2010a). (Software to implement the hierarchical, Bayesian analysis of molecular variance, BAMOVA, as described by other studies [Gompert et al. 2010a; Gompert and Buerkle 2011] is available at http://www.uwyo.edu/buerkle/software/.) This analysis was based on 1570 sequence loci, each with greater than $10 \times$ coverage across all populations and at minimum $1\times$ coverage per population per locus (see Results). We include low coverage loci, rather than restricting analyses to only loci with some arbitrarily higher coverage, to improve the efficiency and accuracy of inferences. Higher coverage loci clearly contribute relatively more information to estimates of population parameters, but there is still important information from loci with low coverage and discarding the low coverage loci reduces the overall sample size and, therefore, the accuracy and precision of these estimates (see Buerkle and Gompert 2012 for a full discussion of this issue).

We used an ordination method, nonmetric multidimensional scaling (NMDS), to visualize population differentiation. Unlike tree-based clustering methods (e.g., neighbor-joining), NMDS does not force a bifurcating relationship among populations. We employed the matrix of pairwise ϕ_{ST} values to conduct NMDS with two dimensions using the *MASS* package in R (Venables and Ripley 2002; R Development Core Team 2010).
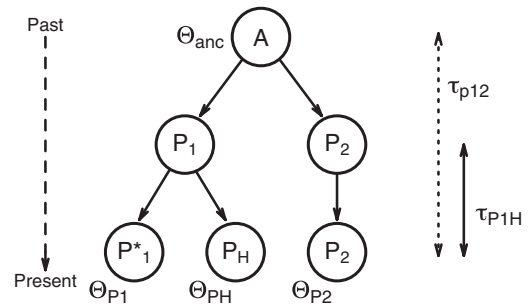
MODELING SPECIATION

We tested three competing hypotheses for the origin of each of the alpine *Lycaeides* populations and each alpine population was modeled independently. The hybrid speciation model (HS) posits the divergence of *L. anna* (P_1) and *L. melissa* (P_2) from a common ancestor (A) at some time in the past (τ_{P12}; Fig. 2a). Then, at a more recent time (τ_{admix}), *L. anna* and *L. melissa* contribute to a single admixture event in which some proportion (α) of the hybrid genome is derived from P_1, and P_2 contributes $1 - \alpha$. The hybrid lineage (P_H, an alpine population) and the parental species are isolated following the admixture event. We define population mutation rates ($4N_e\mu$) for the ancestral population (θ_A), parental population 1 (θ_{P1}), parental population 2 (θ_{P2}), and
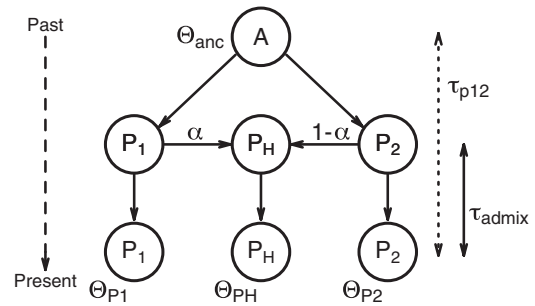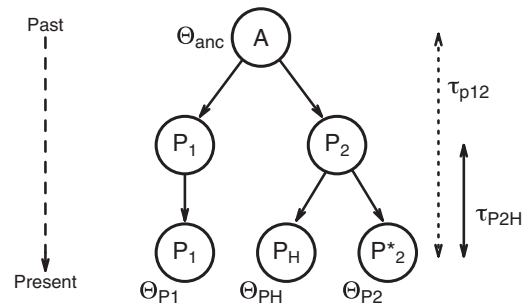

Figure 2. Graphical representation of (A) the hybrid speciation model, (B) the *L. anna*—alpine sister species model, and (C) the *L. melissa*—alpine sister species model. Parameters are described in the text.

the hybrid lineage (θ_H). Alternatively, the *L. anna*—alpine sister species model (AS) (Fig. 2b) assumes that, following the initial divergence between (P_1) and (P_2), population P_1 experienced a bifurcating speciation event that gave rise to two extant lineages P_1^* (i.e., *L. anna*) and an alpine population (P_H). This model replaces the admixture time (τ_{admix}) with the time of the more recent speciation event (τ_{P1H}) and does not include the admixture proportion, but is otherwise identical to the HS model. Finally, the *L. melissa*—alpine sister species model (MS) (Fig. 2c) is a mirror image of the AS model with a speciation event at time τ_{P2H} that gives rise to P_2^* (i.e., *L. melissa*) and the alpine population (P_H).

As is generally true with all modeling approaches, the models examined here are a simplification of what is more likely a complex evolutionary process, and this modeling approach could

be extended to incorporate more parameters. However, given limits on computational time and the low-coverage sequence data, these speciation models capture the basic dimensions of the hypotheses being tested. One specific simplification of the models examined here that requires comment is the exclusion of gene flow following the admixture event from the hybrid speciation model. Although gene exchange among lineages of hybrid origin, or between the hybrids and parental species, is a possibility, we do not consider cases with continuing gene flow to represent hybrid speciation. In other words, hybrid speciation is considered complete when the hybrid lineage is no longer exchanging genes with the parental taxa or any other incipient lineage (Buerkle and Rieseberg 2008). Thus, with respect to gene flow parameters, our simplified model comports with a rigorous definition of hybrid speciation (Buerkle and Rieseberg 2008). Another consideration is that continuing gene flow following admixture would simply have the effect of pushing the admixture time (τ_{admix}) forward in the model (i.e., more recent admixture). Furthermore, the effect of postadmixture gene flow would be to reduce the amount of differentiation observed between lineages. In this sense, the exclusion of postadmixture gene flow is a conservative approach because models accounting for such gene flow would presumably also be more likely to favor hybrid speciation compared to a model lacking this parameter if postadmixture gene flow had actually occurred. Consequently, although these models are relatively simple, we believe they adequately capture the dynamics of the hypotheses being tested.

We used an Approximate Bayesian Computation (ABC) approach (Beaumont et al. 2002; Beaumont 2008; Csillery et al. 2010) to estimate the posterior probability of each of these models for each alpine population. These estimates were based on sequence data from both of the putative parental species and each hybrid lineage. Each analysis included a subset of loci from the sequence data described in the previous section that had at least one sequence for each population in the analysis and were compatible with an infinite sites mutation model (660 loci for the Sierra Nevada, 490 loci for the Siskiyou Mts., 715 loci for the Warner Mts., or 691 loci for the White Mts.). We briefly describe the procedure used to estimate model posterior probabilities. We specified a prior probability for each of the models, $P(\gamma = j) = 1/3$, and priors for the following parameters conditional on the model (i.e., only a subset of parameters are relevant for a given model): effective population size (N_e) for P_A, N_e for P_1 or P_1^*, N_e for P_2 or P_2^*, N_e for P_H, the time (number of generations in the past) of admixture t_{admix}, or bifurcating speciation t_{P1H} or t_{P2H}, the time of the ancestral speciation event t_{P12}, the mean per locus mutation rate μ, the standard deviation of the mutation rate σ_μ, and the admixture proportion α (Table S2). Note, we placed priors on individual parameters, rather than the composite parameters $\theta = 4N_e\mu$ and $\tau = \frac{t}{4N_e}$ used to describe the models. We generated

10^6 simulated datasets per alpine population by first sampling a model (HS, AS, or MS) and then sampling model parameters from the appropriate priors. We used the software ms (Hudson 2002) to simulate sequence data with the sampled model and parameters following an infinite sites mutation model. The data were simulated to match the sampling and coverage of the observed data. Specifically, we simulated a sample of 30 gene copies per locus per population (i.e., 15 individuals) and then re-sampled the simulated sequences with replacement to match the sequence coverage for the locus.

We calculated 26 summary statistics for the observed genetic data and each of the simulated datasets: the number of alleles at a locus in each population (mean and variance among loci), nucleotide diversity in each population (π_i; mean, variance, and skew among loci), the net number of nucleotide differences between each population pair (π_{ij}^{net}; mean, variance, and skew among loci), the m_Y admixture coefficient estimator (Excoffier et al. 2005), and the admixture coefficient estimator given by equation (6) from Chakrabory et al. (1992). We scaled and rotated the summary statistics using a principal component analysis (PCA) using the *prcomp* function in R. This is similar to the approach of Bazin et al. (2010) and this orthogonal transformation of the summary statistics is performed to reduce their dimensionality. We then calculated the Euclidean distance between the transformed summary statistics for the observed data and each simulated dataset. We calculated the correlation between each of the transformed summary statistics and the model used to simulate the data (e.g., HS, AS, or MS) for the 50,000 (5%) simulated datasets with the smallest Euclidean distance. We used the two transformed summary statistics (S^*) most correlated with the speciation model to estimate model posterior probabilities. As proposed by Beaumont (2008), we treated the evolutionary model as a categorical variable and used a multinomial logit model to estimate $P(\gamma = j|S^* = s^*)$. We performed model parameter estimation using a weighted local regression (i.e., using the 10,000, 1%, simulations closest to s^*), which was conducted using the *postpr* function in the R package *abc* (Csillery et al. 2010). Bayes factors were used to compare among the three models of speciation.

ESTIMATING INITIAL ADMIXTURE PROPORTIONS AND SPECIATION TIMES

We estimated model parameters for the HS evolutionary model to quantify initial admixture proportions and determine whether hybrid, alpine lineages varied in their initial admixture proportions, times of admixture, and speciation times. We used an ABC framework to estimate model parameters using a similar approach as we described in the previous section. However, for parameter estimation, all datasets were simulated under the HS evolutionary model, and we conducted 10^6 simulations per alpine population. As

described previously, we scaled and rotated the summary statistics using a PCA, but for parameter estimation, we calculated the correlation between each rotated summary statistic and seven model parameters: θ_{PA}, θ_{P1}, θ_{P2}, θ_{PH}, τ_{P12} τ_{admix}, and α. The rotated summary statistic most correlated with each parameter was used to estimate the posterior probability distribution for all model parameters, except for α, for which the two most closely correlated summary statistics were used. If the same rotated summary statistic was most correlated with multiple parameters, it was included a single time. We computed the Euclidean distance between the retained and rotated summary statistic for each simulated dataset and the observed data. We obtained posterior estimates of parameters using weighted local multivariate regression (Beaumont et al. 2002) using 1% of the simulated datasets with the lowest Euclidean distances. We logit-transformed α and log-transformed all other model parameters prior to posterior density estimation. We assessed differences in initial admixture proportion among the hybrid lineages by estimating the posterior probability distribution for the difference in initial admixture proportion between pairs of populations and examining the proportion of times that the difference in initial admixture proportion was greater than zero in the postburnin Markov chain Monte Carlo (MCMC) for a given pairwise comparison. These proportions provide an estimate of the probability that the initial admixture proportions are different in each pairwise comparison.

EVALUATING THE ABC APPROACH

We evaluated the efficacy of the ABC approach we used to estimate speciation model posterior probabilities by applying the same procedure to 1000 of the simulated datasets, as the evolutionary model for each is known. We then calculated the mean posterior probability of each evolutionary model as a function of the true model.

We also assessed the precision and efficacy of the specific ABC approach we used to estimate demographic parameters by applying the same approach to 500 simulated datasets with known parameter values. We used root mean square error (RMSE) and relative RMSE (rRMSE; RMSE divided by the true parameter value) to measure the difference between each true parameter value and the estimated parameter value (median of the posterior probability distribution). We also quantified the proportion of the time that the true parameter value was contained within the 95% credible interval (CI) of the estimated parameter value (95% COV).

ESTIMATING CONTEMPORARY ADMIXTURE PROPORTIONS

We used a simple Bayesian model to estimate contemporary admixture proportions. For each putative hybrid lineage, we assumed that allele frequencies are a mixture of allele frequencies in

the parental species. Under this assumption, the contemporary admixture proportion, a, describes the proportion of hybrid genome contributed by *L. anna* (parental population 1). $1 - a$ describes the contribution of *L. melissa* (parental population 2). Thus, we must estimate the allele frequencies in the two parental populations and a for each hybrid population. Following Gompert and Buerkle (Gompert et al. 2010a; Gompert and Buerkle 2011), we model the likelihood of the observed allele counts given the population frequencies using a multivariate Pólya (multinomial-Dirichlet) distribution, which incorporates uncertainty that arose by sampling individuals from a population and sampling sequences (reads) from individuals:

$$
P(\mathbf{x}|\mathbf{p}, \nu) = \prod_i \prod_j \frac{n_{ij}!}{\prod_k (x_{ijk}!)} \frac{\Gamma\left(\sum_k \nu_j p_{ijk} + 1\right)}{\Gamma\left(n_{ij} + \sum_k \nu_j p_{ijk} + 1\right)} \quad (1)
$$
$$
\times \prod_k \frac{\Gamma(x_{ijk} + \nu_j p_{ijk} + 1)}{\Gamma(\nu_j p_{ijk} + 1)}.
$$

Here Γ is the gamma function, $\nu_j = 30$ is the number of gene copies (i.e., twice the number of diploid individuals) sampled from population $j \in \{0, 1\}$, p_{ijk} and x_{ijk} are the frequency and observed count of allele k in population j for locus i, and n_{ij} is the number of reads from population j for locus i. The likelihood of the data for the admixed population is computed likewise, but with $p_{ijk} = a p_{i1k} + (1 - a) p_{i0k}$. We placed uniform priors on the allele frequencies and a. To avoid loss of information in ancestry associated with derived haplotypes in the current model framework, we restricted this analysis to a single variable site within each of the contigs (the first variable nucleotide was selected). We obtained the marginal posterior probability distribution for a and parental population allele frequencies using MCMC. We implemented the MCMC algorithm in C++ using the Gnu Scientific Library (Galassi et al. 2009). For each admixed population, we sampled the MCMC for 47,500 steps following a 2500-step burnin. We summarized posterior estimates of a based on the median and 95% CIs of the postburnin samples. As with intial admixture (α), we assessed differences in contemporary admixture proportion (a) among the hybrid lineages by estimating the posterior probability distribution for the difference in contemporary admixture proportion between pairs of populations and examining the proportion of times that the difference in initial admixture proportion was greater than zero in the postburnin MCMC for each pairwise comparison.

LARVAL HOST PLANTS AND HOST FIDELITY

The larval hosts for each of the lineages has been determined from field observations of ovipositing *Lycaeides* females. At several localities, including the high-altitude sites in the Sierra Nevada,

White Mts., and Warner Mts., potential alternative hosts are available. However, in such cases, females in the field consistently oviposit on a single host, referred to as the "natal host" (Nice and Shapiro 1999; Nice et al. 2002; Fordyce and Nice 2003). Nevertheless, the presence of potential alternative larval host plants at some *Lycaeides* sampling localities prompted a series of cafeteria style oviposition preference experiments (choice trials) to estimate host fidelity. Wild females were captured at sampling localities, transported to a greenhouse and placed in choice arenas that included fresh-cut stems of two to four potential host plants and including the natal host. The results of these experiments (Nice et al. 2002; Gompert et al. 2006; Fordyce et al. 2011) were reanalyzed to estimate preference for the natal host in each of the lineages using a hierarchical Bayesian model specifically designed for count data (Fordyce et al. 2011).

EGG MORPHOLOGY

The external morphology of *Lycaeides* eggs was examined using scanning electron microscopy. Morphometric analyses were used to describe the variation among populations and lineages (Forister et al. 2006). The morphology of eggs from the Siskiyou Mts. populations has not yet been examined. We reanalyzed these data for six of the seven populations considered here. We contrasted the first principal component between lineages using analysis of variance (ANOVA). Post hoc tests were employed to test for variation among populations.

EGG ADHESION

While conducting oviposition preference experiments with *Lycaeides* females (Nice et al. 2002; Gompert et al. 2006; Fordyce et al. 2011), we discovered an unusual trait in the hybrid lineages in the Sierra Nevada (Fordyce and Nice 2003). Females from this lineage do not firmly adhere their eggs to the natal host plant. As a consequence, in this population, eggs laid by females on the natal host plant readily fall off of the plant and come to rest at the base of the plant. We recorded observations of egg adhesion in the other lineages during the course of conducting host preference experiments (see above) and from several years of field observations.

MALE MORPHOLOGY

Measurements of male genitalic morphology and wing pigment patterns were used originally by Nabokov (1943, 1949) to delineate the parental species *L. anna* and *L. melissa*. We used Nabokov's measurements to examine variation in male morphology across North America, including the populations sampled here (Nice and Shapiro 1999; Nice et al. 2005; Lucas et al. 2008). We reanalyzed these data for the seven focal populations (314 males, Table S4) in this study. PCA and an ANOVA on PC1 scores were used to compare population means.

WING PIGMENT PATTERNS

We measured eight wing pattern characters from 108 male and 61 female *Lycaeides* (Table S5). We removed the wings from each butterfly and photographed them using a digital camera. We focused on the underside of the hindwings because this is the surface of the wing that is visible when *Lyaceides* are at rest and is therefore important during courtship, and because this is what Nabokov used in his original descriptions of *Lyaceides* in North America (Nabokov 1943, 1949). Two features of the wing pigment patterns are obvious: black wing pattern spots on the interior of the wing, and aurorae, which are orange and iridescent elements along the wing margins (Fig. S2). We identified eight wing pattern elements of both the black spots and aurorae that were used in these analyses: M, Sc, M2, Cu1, a2, b2 (the orange portion) of a2, a6, and b2 of a6 (Fig. S2). We drew outlines around each wing and the eight characters using the IMAGEJ software to calculate areas of each element. We standardized each measurement by dividing it by total wing area. We used PCA to explore variation in the wing pattern characters. First, to examine and illustrate the patterns of variation among lineages within the sexes, separate PCAs were performed for females and males. We then used ANOVA on PC1 and PC2 scores to test for differences among lineages within the sexes. ANOVA was also used on the total dataset, with sex was included as a factor the analysis, to test whether the hybrid entities were morphologically distinct from each other and the parental species by contrasting the first principal component for wing pattern (which explained 68.9% of the variation) for all samples (i.e., females and males combined).

Results

MOLECULAR DATA AND GENOME-WIDE POPULATION DIFFERENTIATION

We assembled the 341,045 sequences into 15,262 contigs or loci. Files including sequences and quality scores have been deposited in the NCBI Short Read Archive (accession SRA010351). Coverage for individual loci varied substantially and we confined our analyses to loci with greater than $10\times$ coverage. Population genetic analyses of these 1570 DNA sequence loci confirm the intermediate nature of the four lineages, consistent with the hypothesis that these high-altitude populations arose via hybridization. Ordination of pairwise genetic distances between all populations confirmed previous findings from other data (Gompert et al. 2008b) that the populations in the four mountain ranges are differentiated from *L. anna* and *L. melissa*, but have intermediate genomic compositions, relative to the putative parental species (Fig. 3, Table S3). The four hybrid lineages also appear to be genomically differentiated from each other, particularly in the second dimension of the ordination.
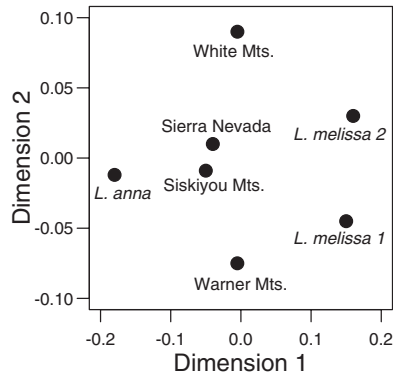
Figure 3. Scatter plot depicting ordination-based distances from NMDS of molecular variation among populations.

MODELING SPECIATION

We used an ABC approach to estimate the posterior probabilities of three speciation models (Fig. 2) for each putative hybrid lineage. The HS model had the highest posterior probability for all alpine populations, which ranged from 0.447 to 0.570 (Table 2). Similarly, the HS model was approximately one to four times as likely as the other models (Table 2). For the *Lycaeides* lineage from the Siskiyou Mts., our ability to discriminate among the three models was limited and the posterior probabilities for the HS and AS models were similar.

ESTIMATING INITIAL ADMIXTURE PROPORTIONS AND SPECIATION TIMES

We used ABC methods for parameter estimation under the HS model (Fig. 2a). The posterior probability distributions for each parameter are summarized in Table 3. Estimates of initial admixture proportions, α (the proportion of the admixed genome derived from *L. anna*), for the Sierra Nevada, Warner, and White mountains are consistent with the hypothesis of hybrid speciation. Each of these lineages has an estimate of initial admixture proportion with a 95% CI that does not include zero or one, indicating genomes consisting of elements of both of the putative parental species, *L. anna* and *L. melissa* (Fig. 4). The estimate of initial admixture for the population in the Siskiyou Mts. (Siskiyou Mts., $\alpha = 0.85$, 95% CI $0.10 - 1.00$) did not provide strong evidence of admixture and is consistent with either the hybrid speciation model or the bifurcating speciation model involving *L. anna* (see above). Among the three other lineages, there was variation in the admixture proportion (Sierra Nevada, $\alpha = 0.55$, 95% CI $0.24 - 0.82$; Warner Mts., $\alpha = 0.40$, 95% CI $0.05 - 0.91$; White Mts., $\alpha = 0.28$, 95% CI $0.02 - 0.91$). We made pairwise comparisons of estimates of initial admixture proportions by estimating the posterior probability of the difference in estimates and calculating the proportion of times that the difference was greater than zero in the postburnin MCMC for each pair of lineages. Although the CIs on these estimates were relatively wide, there were clear differences
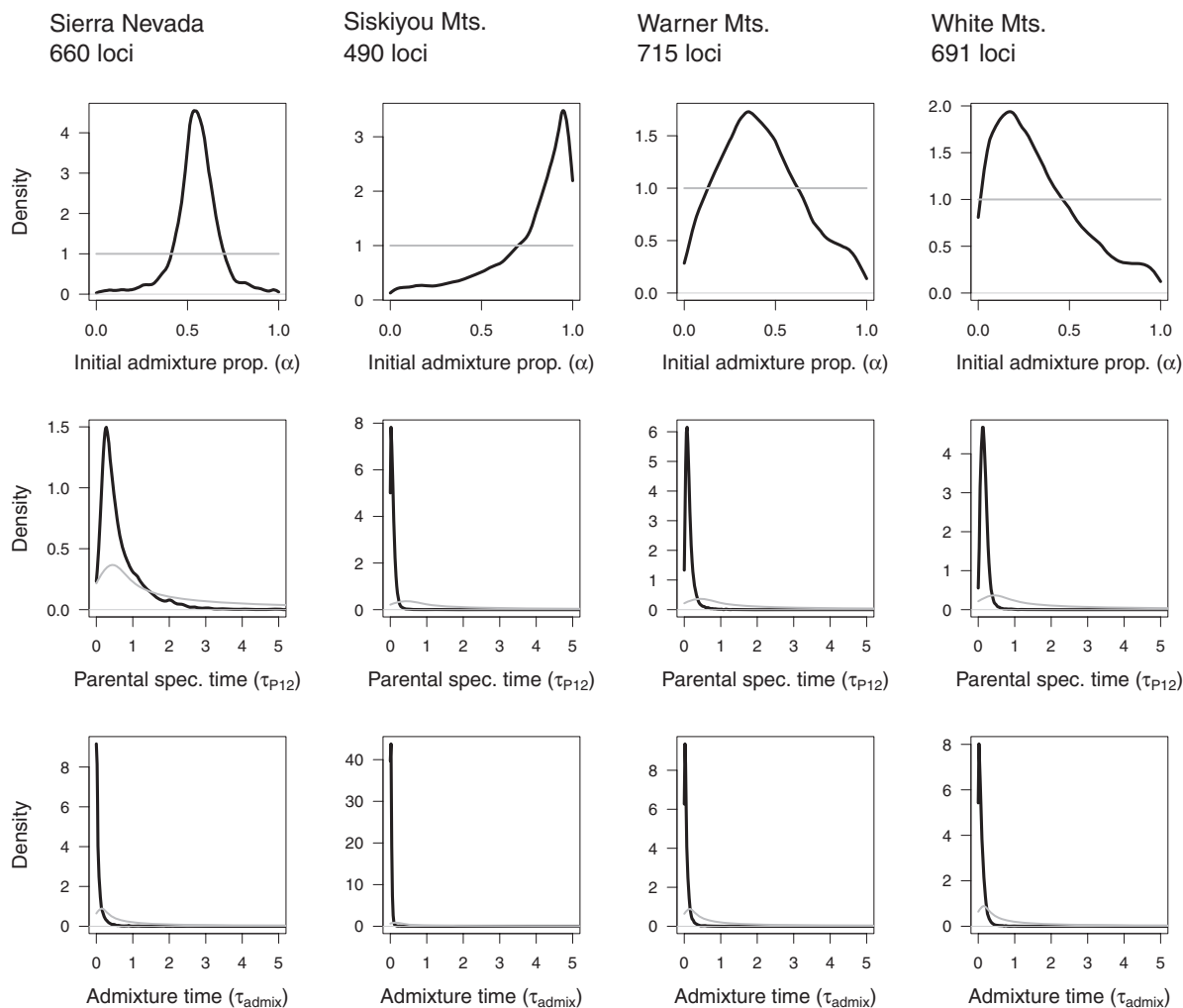
Table 2. Evolutionary model posterior probabilities and Bayes factors. Probabilities are given for the hybrid speciation (HS), *L. anna*—alpine sister species (AS) and *L. melissa*—alpine sister species (MS) models for each of the four hybrid lineages. Bayes factor for the HS model (numerator) relative to each model (denominator).

Population	Posterior probabilities			Bayes factors		
	HS	AS	MS	HS	AS	MS
Sierra Nevada	0.542	0.147	0.311	1.00	3.69	1.74
Siskiyou Mts.	0.447	0.442	0.111	1.00	1.01	4.02
Warner Mts.	0.570	0.192	0.238	1.00	2.97	2.40
White Mts.	0.526	0.339	0.135	1.00	1.55	3.99

Table 3. Median and 95% credible intervals for model parameters.

Population	θ_{PA}	θ_{P1}	θ_{P2}	θ_{PH}	τ_{P12}	τ_{admix}	α
Sierra Nevada	1.28	1.00	1.08	1.74	0.46	0.03	0.55
	$(0.20 - 3.14)$	$(0.28 - 2.02)$	$(0.21 - 3.51)$	$(0.04 - 10.4)$	$(0.06 - 2.37)$	$(0.00 - 0.49)$	$(0.24 - 0.82)$
Siskiyou Mts.	0.78	0.79	1.09	0.23	0.05	0.01	0.85
	$(0.09 - 2.57)$	$(0.08 - 5.58)$	$(0.08 - 9.13)$	$(0.01 - 3.09)$	$(0.00 - 0.26)$	$(0.00 - 0.08)$	$(0.10 - 1.00)$
Warner Mts.	1.63	0.91	1.99	1.15	0.10	0.04	0.40
	$(0.47 - 4.49)$	$(0.12 - 4.36)$	$(0.30 - 8.30)$	$(0.12 - 6.55)$	$(0.01 - 0.50)$	$(0.00 - 0.30)$	$(0.05 - 0.91)$
White Mts.	0.60	0.81	2.49	1.76	0.15	0.06	0.28
	$(0.15 - 1.39)$	$(0.23 - 2.74)$	$(0.55 - 9.19)$	$(0.22 - 9.28)$	$(0.02 - 0.45)$	$(0.00 - 0.36)$	$(0.02 - 0.91)$

Figure 4. Posterior probability distribution for historical model parameters (black lines). Prior distributions (gray lines) are also shown. Inferences were based on 490–715 loci, as indicated below population labels.

in the point estimates of initial admixture proportions (Fig. S3 and Table 4).

Estimates of speciation time of the parental species (*L. anna* and *L. melissa*) were similar across the four analyses as expected (Fig. 4). The speciation event for the parental species appears to be a relatively recent divergence and, although we do not have a genome-level mutation rate estimate for *Lycaeides* and cannot translate our speciation time estimates into absolute time, this result comports with previous investigations with other marker systems (Nice et al. 2005; Gompert et al. 2006, 2008a). However, the recency of the divergence between *L. anna* and *L. melissa* obviously constrains the time to admixture, which are also short (i.e., recent events) for all four hybrid lineages (Fig. 4).

TESTS OF THE ABC APPROACH

To evaluate the efficacy of the ABC approach for modeling speciation, we calculated the mean posterior probability of each evolutionary model as a function of the true model for 1000 of the

simulated datasets. The results are given in Table S6. Our ability to discriminate among models was modest. However, the most common incorrect inference was classification of HS as either AS or MS, thus our results are likely conservative and might underestimate the probability of HS for the observed data. An inherent limitation is that as admixture proportions tend toward 0 or 1, the HS model converges on either the AS or MS model.

We also assessed our ability to use ABC methods for parameter inference. The results of this assessment are reported in Table S7. rRMSE was on the order of 0.25 for most parameters (i.e., the difference between the estimated and true parameter value was about 25% of the true value), but higher for θ_{PA} and only 0.11 for α. Moreover, the mean RMSE for α was only 0.05, indicating that on average the estimated value of α was within $\pm 0.05\%$ of the true value. This is rather encouraging, as α is the parameter, we were most interested in because we consider this parameter to be most informative about the hypothesis of admixture.

Table 4. Comparisons of estimates of admixture proportions among hybrid lineages. Values are the probability that the admixture proportion of lineage (population) in that row is greater than the admixture proportion of the column lineage for (A) initial admixture proportions (α) and (B) contemporary admixture proportions (*a*).

Lineage:	Siskiyou Mts.	Warner Mts.	Sierra Nevada	White Mts.
(A) Initial admixture				
Siskiyou Mts.	0.00	0.83	0.79	0.87
Warner Mts.	0.17	0.00	0.31	0.63
Sierra Nevada	0.21	0.69	0.00	0.79
White Mts.	0.13	0.37	0.21	0.00
(B) Contemporary admixture				
Siskiyou Mts.	0.00	0.97	0.85	0.99
Warner Mts.	0.03	0.00	0.09	0.95
Sierra Nevada	0.15	0.91	0.00	0.99
White Mts.	0.01	0.05	0.01	0.00

Table 5. Estimates of contemporary admixture proportion (*a*) and 95% credible intervals for each hybrid lineage.

Population	*a*	95% credible interval
Sierra Nevada	0.672	(0.592–0.737)
Siskiyou Mts.	0.732	(0.623–0.824)
Warner Mts.	0.602	(0.530–0.674)
White Mts.	0.513	(0.438–0.589)
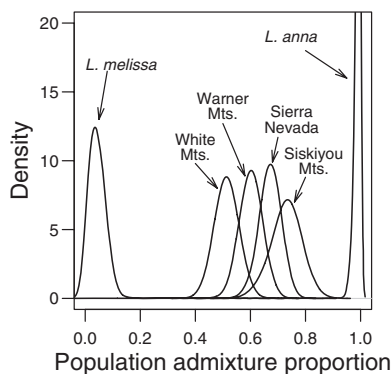
Figure 5. Posterior probability distributions for contemporary admixture proportions in each of the four hybrid lineages: White Mts., Warner Mts., Sierra Nevada, and Siskiyou Mts. For comparison, the posterior distributions for the parental taxa are included: *L. anna* and *L. melissa*.

ESTIMATING CONTEMPORARY ADMIXTURE PROPORTION

Posterior estimates of *a* based on the median and 95% CIs of the postburnin samples are summarized in Table 5 and Figure 5.

Examination of differences in estimates of contemporary admixture proportions between lineages (using the same approach used to examine pairwise differences in intial admixture proportions) demonstrate substantial differences in all but two of the pairwise comparisons between hybrid lineages (Sierra Nevada vs. Siskiyou Mts. and Sierra Nevada vs. Warner Mts.) (Fig. S4 and Table 4).

LARVAL HOST PLANTS AND HOST PREFERENCE

Female host preference was assessed in cafeteria-style choice experiments (Nice et al. 2002; Gompert et al. 2006; Fordyce et al. 2011). There is considerable variation among lineages in the level of female preference for their natal host (Fig. S5). Females from populations of the hybrid lineages had relatively strong preference for their natal hosts compared to populations of *L. anna* and *L. melissa*. This was especially true of the hybrid populations from the Sierra Nevada. Female oviposition preference, considered here as a population-level trait (see Fordyce et al. 2011), is coupled in *Lycaeides* with the potential for extrinsic or ecological reproductive isolation because mating also occurs on, or very near, the natal host plant (Nice and Shapiro 1999; Nice et al. 2002).

EGG MORPHOLOGY

Morphological measurements of egg characteristics were made from scanning electron microscopy images (Forister et al. 2006). We contrasted the first principal component (which explained 47.8% of the variation for egg morphology) among lineages using ANOVA. Post hoc tests revealed variation between populations (Fig. S6) with the hybrid populations having egg morphology somewhat intermediate between the parental species, though more similar to *L. anna* compared to *L. melissa*.

EGG ADHESION

Lack of egg adhesion was observed in two of the putative hybrid lineages: in the Sierra Nevada and White Mts. To the best of our knowledge, these are the only *Lycaeides* populations that exhibit this trait. Experiments used to quantify the strength of egg adhesion demonstrated that this is not a consequence of a plant trait. That is, eggs from the Sierra Nevada and White Mts. populations showed significantly less egg adhesion compared to other populations, regardless of which plant the egg was laid on (Fordyce and Nice 2003). We interpreted this to be adaptive, and potentially a mechanism that contributes to isolation of these populations, because the above-ground biomass of the host plants senesces near the end of the growing season and is subsequently removed by winter weather conditions. Overwintering eggs that fall from the host land near the site of new growth the following spring (Fordyce and Nice 2003).

Strong egg adhesion was observed in all other populations, including the populations of the parental species as well as the

populations in the Siskiyou and Warner mountains. The egg adhesion in the Warner Mts. is somewhat puzzling given our adaptive scenario for the lack of adhesion observed in the Sierra Nevada and White Mts. because the high-altitude Warner Mts. populations also use an alpine *Astragalus* as a larval host. However, there may be important ecological differences between these populations or their respective host plants. Alternatively, these different character states may represent underlying differences in the standing variation created by the hybridization event(s) giving rise to these lineages.

MALE MORPHOLOGY

Measurements of Nabokov's (1943, 1949) male genitalic characters for 314 males from the sampled populations (Lucas et al. 2008) (Table S4) were reanalyzed using PCA and an ANOVA on PC1 scores (which explained 90.8% of the variation). Comparison of population means form the basis of the data reported in Table 1 and Figure S7. Males from the Sierra Nevada, White mountains, and Warner mountains showed various degrees of intermediacy. Males from the Siskiyou mountains were not different from *L. anna*.

WING PIGMENT PATTERNS

Eight wing pigment pattern elements (Fig. S2) were measured from digital photos of wings of both males and females (169 total individuals) from the focal populations (Table S5). Female and male wing patterns exhibited similar patterns of variation across all lineages (Fig. S8). Most variation in wing pattern (80%) was explained by the first principal component. All eight measurements of wing pattern had positive loadings on PC1 suggesting that PC1 corresponded to wing pattern character size. This is consistent with previous analyses of wing pattern variation in *Lycaeides* (Fordyce et al. 2002; Gompert et al. 2010b). Wing pattern PC2 had positive loadings for black spots, but negative loadings for wing aurorae and accounted for 12% of wing pattern variation. Based on PC1, we detected significant differences in wing patterns between males and females ($F = 46.6$, $P < 0.0001$) and among populations ($F = 253.4$, $P < 0.0001$). All pairwise comparisons between hybrid entities were significant except between the White Mts. and the Sierra Nevada (Fig. S9). Based on PC2, we detected significant differences in wing pattern between the Warner Mts. and all other lineages ($F = 23.1$, $P < 0.0001$) and between males and females ($F = 5.9$, $P < 0.016$). Combined ANOVA on PC1 scores with population (i.e., lineage), sex, and the interaction as effects was performed to further test for significant among-lineage differences (Table S8). Post hoc tests revealed substantial variation among populations (Fig. S9), including a wide range of variation among the hybrid lineages compared to each other and the parental species.

Discussion

The combination of molecular, ecological, and morphological data presents a picture of a complex mosaic of differentiated lineages in western North American *Lycaeides*. Data from 1570 DNA sequence loci confirm the admixed and intermediate genomic composition of lineages in the Sierra Nevada, Warner, and White mountain ranges. There is less evidence of admixture for the lineage in the Siskiyou mountains. Evaluation of three alternative speciation models using ABC methods supported the hybrid speciation model over bifurcating speciation models in three of the lineages. Again, the lineage in the Siskiyou mountains was the exception with the hybrid speciation (HS) model and the *L. anna*-alpine sister species (AS) model having nearly equal probabilities and this lineage does not appear to be of hybrid origin. The Siskiyou population is perhaps best described as a sister taxon to *L. anna*. Although the strength of evidence supporting the HS model in each of the other three lineages was modest, the molecular data are clearly consistent with the hypothesis of hybrid speciation. Despite the evidence of genomic admixture and support for hybrid speciation, these lineages are not homogenous and exhibit distinct ecological, morphological, and genetic characters.

The lineages are similar in that they all occupy high-altitude sites in the four mountain ranges, consistent with the hypothesis that hybrid speciation is facilitated by escape from the influence of the parental species by colonization of novel habitats (Buerkle et al. 2000; Gompert et al. 2006). Morphometrics of the external morphology of *Lycaeides* eggs showed that the hybrid lineages have intermediate egg morphology relative to *L. anna* and *L. melissa*, although the morphology of eggs from the Siskiyou mountains has not yet been examined (Fig. S6). However, beyond these similarities, there are unique combinations of traits that distinguish these lineages from each other and the putative parental species, with both intermediate and extreme trait values exhibited in the hybrids, relative to the parental species (Table 1). Three of the four lineages use a high altitude, endemic *Astragalus* as the larval host plant. The high-altitude lineages tend to have much higher fidelity to their natal larval host plant, as measured by female oviposition preference, than either of the parental species (Fig. S5). Two of the lineages (in the Sierra Nevada and White Mts.) exhibit an unusual lack of egg adhesion that causes eggs to fall off of the host plant following oviposition. This trait appears to be adaptive in the alpine habitat where the above-ground biomass of the host plant senesces at the end of the high-altitude growing season (Fordyce and Nice 2003). Populations in the Sierra Nevada, Warner, and White mountain ranges are also intermediate in terms of the form of the male genitalia (Lucas et al. 2008), whereas males from the Siskiyou mountains are not significantly different from *L. anna* male morphology (Fig. S7). Wing patterns for both males and females also vary among the four lineages and

the parental species, but while the Sierra Nevada, Warner, and White lineages have wing patterns that tend to be intermediate, they are not significantly different from *L. melissa*. The Siskiyou lineage has unusual, reduced wing patterns that appear to be transgressive with respect to both of the parental species (Fig. S9). However, as discussed above, the lineage in the Siskiyou mountains might be a sister taxon of *L. anna* and their unusual wing patterns might not, therefore, be the product of admixture.

In accord with ecological and morphological differences, the three putatively hybrid lineages are distinct genetically. Ordination of pairwise ϕ_{ST} values illustrates the differentiation among populations in the Sierra Nevada, Warner, and White ranges (Fig. 3). Estimation of contemporary admixture proportions also revealed substantial differences in the contribution of each of the parental species to the genomes of the admixed lineages (Fig. 5, Table 4). These comparisons of the admixed lineages strongly suggest that there is little gene flow among them, either as a function of geographic isolation or strong independent selective regimes in each mountain range, or both. Thus, the contemporary patterns of ecological, morphological, and genetic differentiation among the admixed *Lycaeides* lineages are consistent with a history of independent evolution. There is little indication of cohesive or coordinated evolution among the lineages.

Estimates of initial admixture proportions (at the time of hybridization) also varied among the lineages (Figs. 4, S3; Table 4). Although estimates of these parameters clearly exhibit substantial uncertainty, the bulk of the posterior densities suggests that these populations differed in their initial admixture proportions. Estimates of initial admixture for three of the lineages show considerable genomic contribution from both parental species; however, the 95% CI for the initial admixture estimate for the Siskiyou Mts. lineage includes 1.0 (i.e., no contribution from *L. melissa*). If the Siskiyou Mts. lineage arose by hybridization, then the evidence suggests that the initial admixture event was asymmetrical with a majority of the genomic contribution derived from *L. anna*; however, as noted, the *L. anna*—alpine sister species model cannot be rejected for the Siskiyou lineage.

The variation among the estimates of initial admixture proportion might indicate that the lineages experienced independent evolutionary histories following a common hybridization event. That is, subsequent to hybridization, recombination, selection and drift acted upon the admixture of parental genomes independently resulting in the observed variation in estimates of admixture proportions. Alternatively, the variation in initial admixture might be evidence of repeated hybridization events leading to multiple hybrid lineages.

Distinguishing between these competing hypotheses is difficult. If the admixed lineages are the product of multiple hybridization events, one could expect variation in the estimates of the time to admixture. However, estimated admixture times were very

similar and obviously constrained by the relatively recent divergence time for the parental species, *L. anna* and *L. melissa* (Fig. 4, Table 3). Thus, even with 1570 sequence loci, we lack the resolution to identify differences in the timing of multiple potential admixture events. However, there is clear evidence of independent evolutionary histories for each lineage. As noted above, pairwise ϕ_{ST} values and estimates of contemporary admixture proportions revealed substantial differences among the four lineages. Differences between initial and contemporary admixture proportions in each lineage (Figs. 4, 5, Tables 3, 4) also indicate that the hybrid lineages experienced further independent evolution following the initial admixture. Consequently, our analyses are consistent with the hypothesis of independent evolution in these lineages following hybridization, but the current data cannot distinguish between one hybrid origin for all admixed lineages, or multiple independent hybrid origins.

The combination of genome-level genetic analyses and quantification of ecological and morphological variation reveals a mosaic of differentiation among the four lineages of *Lycaeides* butterflies in high-altitude habitats of mountain ranges in western North America. There is little evidence for hybrid speciation in the Siskiyou Mts. However, in the other three ranges (in the Sierra Nevada, Warner, and White Mountains), our analyses support the hypothesis that these lineages constitute species of hybrid origin that have experienced independent evolutionary histories. Although the available data do not allow us to distinguish between the hypotheses of a single or multiple origins, these admixed lineages represent well-documented cases of hybrid speciation and provide the foundation for comparative analyses of the genetic architecture of hybrid species and an opportunity to study in detail the role of stochastic and selective factors in shaping the genomes of hybrid lineages.

LITERATURE CITED
Abbott, R. J., M. J. Hegarty, S. J. Hiscock, and A. C. Brennan. 2010. Homoploid hybrid speciation in action. Taxon 59:1375–1386.

Arnold, M. L. 2006. Evolution through genetic exchange. Oxford Univ. Press, Oxford, UK.

Arnold, M. L., and N. H. Martin. 2010. Hybrid fitness across time and habitats. Trends Ecol. Evol. 25:530–536.

Bazin, E., K. J. Dawson, and M. A. Beaumont. 2010. Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. Genetics 185:587–602.

Beaumont, M. A. 2008. Joint determination of topology, divergence time and immigration in population trees. Pp. 135–154 *in* S. Matsumura, P. Forster, and C. Renfrew, eds. Simulations, genetics and human prehistory. McDonald Institute for Archaeological Research, Cambridge, UK.

Beaumont, M. A., W. Y. Zhang, and D. J. Balding. 2002. Approximate Bayesian Computation in population genetics. Genetics 162:2025–2035.

Buerkle, C. A., and Z. Gompert. 2012. Population genomics based on low coverage sequencing: how low should we go? Mol. Ecol. *In press*.

Buerkle, C. A., R. J. Morris, M. A. Asmussen, and L. H. Rieseberg. 2000. The likelihood of homoploid hybrid speciation. Heredity 84:441–451.

Buerkle, C. A., and L. H. Rieseberg. 2008. The rate of genome stabilization in homoploid hybrid species. Evolution 62:266–275.

Buerkle, C. A., D. E. Wolf, and L. H. Rieseberg. 2003. The origin and extinction of species through hybridization. Pp. 117–141 *in* C. A. Brigham and M. W. Schwartz, eds. Population viability in plants: conservation, management, and modeling of rare plants. Springer Verlag, New York.

Chakraborty, R., M. I. Kamboh, M. Nwankwo, and R. E. Ferrell. 1992. Caucasian genes in American Blacks—new data. Am. J. Hum. Genet. 50:145–155.

Csillery, K., M. G. B. Blum, O. E. Gaggiotti, and O. Francois. 2010. Approximate Bayesian Computation (ABC) in practice. Trends Ecol. Evol. 25:410–418.

Excoffier, L., A. Estoup, and J. M. Cornuet. 2005. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. Genetics 169:1727–1738.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491.

Fitzpatrick, B. M., and H. B. Shaffer. 2007. Hybrid vigor between native and introduced salamanders raises new challenges for conservation. Proc. Natl. Acad. Sci. 104:15793–15798.

Fordyce, J. A., Z. Gompert, M. L. Forister, and C. C. Nice. 2011. A hierarchical Bayesian approach to ecological count data: a flexible tool for ecologists. PLoS ONE 6:e26785.

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., J. A. Fordyce, C. C. Nice, Z. Gompert, and A. M. Shapiro. 2006. Egg morphology varies among populations and habitats along a suture zone in the *Lycaeides idas-melissa* species complex (Lepidoptera: Lycaenidae). Ann. Entomol. Soc. Am. 99:933–937.

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.

Forister, M. L., C. C. Nice, J. A. Fordyce, Z. Gompert, and A. M. Shapiro. 2008. Considering evolutionary processes in the use of single-locus genetic data for conservation, with examples from the Lepidoptera. J. Insect Conserv. 12:37–51.

Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi. 2009. GNU Scientific Library: reference manual. Network Theory Ltd, Godalming, UK.

Gompert, Z., and C. A. Buerkle. 2011. 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. 2006. Homoploid hybrid speciation in an extreme habitat. Science 314:1923–1925.

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

Grant, P. R., and B. R. Grant. 1992. Hybridization of bird species. Science 256:193–197.

Gross, B. L., A. E. Schwarzbach, and L. H. Rieseberg. 2003. Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). Am. J. Bot. 90:1708–1719.

Gross, B. L., K. G. Turner, and L. H. Rieseberg. 2007. Selective sweeps in the homoploid hybrid species *Helianthus deserticola*: evolution in concert across populations and across origins. Mol. Ecol. 16:5246–5258.

Guppy, C., and J. Shepard. 2001. Butterflies of British Columbia. UBC Press, Vancouver, Canada.

Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338.

Kunte, K., C. Shea, M. L. Aardema, J. M. Scriber, T. E. Juenger, L. E. Gilbert, and M. R. Kronforst. 2011. Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. PLoS Genet. 7:e1002274.

Li, J., Q. Li, W. Hou, K. Han, Y. Li, S. Wu, Y. Li, and R. Wu. 2009. An algorithmic model for constructing a linkage and linkage disequilibrium map in outcrossing plant populations. Genet. Res. 91:9–21.

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.

Mallet, J. 2007. Hybrid speciation. Nature 445:279–283.

Martin, N. H., A. C. Bouck, and M. L. Arnold. 2006. Detecting adaptive trait introgression between *Iris fulva* and *I*. brevicaulis in highly selective field conditions. Genetics 172:2481–2489.

Mavárez, J., C. A. Salazar, E. Bermingham, C. Salcedo, C. D. Jiggins, and M. Linares. 2006. Speciation by hybridization in *Heliconius* butterflies. Nature 441:868–871.

Nabokov, V. 1943. The nearctic forms of *Lycaeides* Hub. (Lycaenidae, Lepidoptera). Psyche 50:87–99.

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

Nice, C. C., N. Anthony, G. Gelembiuk, D. Raterman, and R. ffrench Constant. 2005. The history and geography of diversification within the butterfly genus *Lycaeides* in North America. Mol. Ecol. 14:1741–1754.

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.

van Orsouw, N. J., R. C. J. Hogers, A. Janssen, F. Yalcin, S. Snoeijers, E. Verstege, H. Schneiders, H. van der Poel, J. van Oeveren, H. Verstegen, et al. 2007. Complexity reduction of polymorphic sequences (CRoPS^{TM}): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS ONE 2:e1172.

Parchman, T. L., A. J. M. Z. Gompert, F. D. Shilkey, C. W. Benkman, and C. A. Burkle. 2012. Genome-wide association of genetics of an adaptive trait in lodgepole pine. Mol. Ecol. 21:2991–3005.

R Development Core Team. 2010. R: a language and nnvironment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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.

Schwarzbach, A. E., and L. H. Rieseberg. 2002. Likely multiple origins of a diploid hybrid sunflower species. Mol. Ecol. 11:1703–1715.

Scott, J. 1986. The butterflies of North America: a natural history and field guide. Stanford Univ. Press, Stanford, CA.

Taylor, S. J., M. Arnold, and N. H. Martin. 2009. The genetic architecture of reproductive isolation in Louisiana irises: hybrid fitness in nature. Evolution 63:2581–2594.

Venables, W. N., and B. Ripley. 2002. Modern applied statistics with S. 4th ed. Springer Verlag, New York.

Wang, X.-R., A. E. Szmidt, and O. Savolainen. 2001. Genetic composition and diploid hybrid speciation of high mountain pine, *Pinus densata*, native to the Tibetan Plateau. Genetics 159:337–346.

Supporting Information

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

Table S1. Sampling data for specimens used for 454 pyrosequencing.

Table S2. Prior distributions for model parameters.

Table S3. Median (lower triangle) and 95% credible intervals (upper triangle) of the posterior probability distributions of pairwise genome-level ϕ_{ST} .

Table S4. Numbers of males sampled for male genitalic morphology (see Lucas et al. 2008).

Table S5. Sampling for wing pigment pattern analysis.

Table S6. Mean posterior probability of each model (column) as a function of the true evolutionary model (row).

Table S7. Statistics describing the efficacy of the ABC approach to parameter estimation: mean RMSE, rRMSE, and 95% COV for 500 simulated data sets.

Table S8. Analysis of variance of wing pigment patterns.

Figure S1. Map depicting sampling locations of putative hybrid taxa (colored circles) and parental taxa (colored and numbered squares) in western North America.

Figure S2. Wing pigment pattern elements used for morphometric analyses.

Figure S3. Posterior probability distributions of the difference in initial admixture proportions for each pairwise comparison of the hybrid lineages.

Figure S4. Posterior probability distributions of the difference in contemporary admixture proportions for each pairwise comparison of the hybrid lineages.

Figure S5. Summary of oviposition preference experiment results.

Figure S6. Principal component scores (± 1 SD) of egg morphology for *Lycaeides* lineages: L. a. = *L. anna*, WA = Warner Mts., SN = Sierra Nevada, WH = White Mts., L. m. = *L. melissa*, (eggs from the lineage were not measured).

Figure S7. Principal component scores (± 1 SD) of male morphology for *Lycaeides* lineages: L. a. = *L. anna*, SY = Siskiyou Mts., WA = Warner Mts., SN = Sierra Nevada, WH = White Mts., L. m. = *L. melissa*.

Figure S8. Scatter plot depicting ordination of wing pattern area measurements of individual males (circles) and females (triangles) from among all lineages using principal coordinate axes.

Figure S9. Principal component scores (± 1 SD) of wing pigment patterns for *Lycaeides* lineages: L. i. = *L. anna*, SY = Siskiyou Mts., WA = Warner Mts., SN = Sierra Nevada, WH = White Mts., L. m. = *L. melissa*.