Molecular Ecology (2008) **17**, 5231–5244 doi: 10.1111/j.1365-294X.2008.03988.x

# Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*

ZACHARIAH GOMPERT,*† MATTHEW L. FORISTER,‡ JAMES A. FORDYCE§ and CHRIS C. NICE*
*Department of Biology, Population and Conservation Biology Program, Texas State University, San Marcos, TX 78666, USA, †Department of Botany, University of Wyoming, Laramie, WY 82071, USA, ‡Department of Biology, University of Nevada, Reno, NV 89557, USA, §Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA*

**Abstract**

**We investigated the extent and potential cause(s) of mitochondrial introgression within the polytypic North American *Lycaeides* species complex (Lepidoptera). By comparing population genetic structure based on mitochondrial DNA (COI and COII) and nuclear DNA (251 polymorphic amplified fragment length polymorphism markers), we detected substantial mito-nuclear discordance, primarily involving a single mitochondrial haplotype (h01), which is likely due to mitochondrial introgression between differentiated *Lycaeides* populations and/or species. We detected reduced mitochondrial genetic diversity relative to nuclear genetic diversity in populations where mitochondrial haplotype h01 occurs, suggesting that the spread of this haplotype was facilitated by selection. We found no evidence that haplotype h01 is associated with increased fitness (in terms of survival to eclosion, fresh adult weight, and adult longevity) in a polymorphic *Lycaeides melissa* population. However, we did find a positive association between mitochondrial haplotype h01 and infection by the endoparasitic bacterium *Wolbachia* in one out of three lineages tested. Linkage disequilibrium between mitochondrial haplotype h01 and *Wolbachia* infection status may have resulted in indirect selection favouring the spread of haplotype h01 in at least one lineage of North American *Lycaeides*. These results illustrate the potential for introgressive hybridization to produce substantial mito-nuclear discordance and demonstrate that an individual's mitochondrial and nuclear genome may have strikingly different evolutionary histories resulting from non-neutral processes and intrinsic differences in the inheritance and biology of these genomes.**

*Keywords*: hybridization, *Lycaeides*, mitochondrial introgression, selective sweep, *Wolbachia*

*Received 22 May 2008; revision received 25 September 2008; accepted 30 September 2008*

## Introduction

A proliferation of studies over the past few decades documenting geographical patterns of genetic variation has provided novel insights concerning evolutionary processes affecting natural populations (Avise 1994). Such studies have revealed that introgressive hybridization, defined as the spreading of genes via hybridization from one population into the gene pool of another, differentiated population, is a relatively common phenomenon both above and below the species level (e.g. Rieseberg 1991; Kim & Rieseberg 2001; Grant *et al*. 2004; Bull *et al*. 2006; Gompert *et al*. 2006a; Kronforst *et al*. 2006; Egger *et al*. 2007). While both nuclear and mitochondrial introgression occur, studies documenting mitochondrial introgression are particularly prevalent, which suggests the latter may be more common (e.g. Ruedi *et al*. 1997; Funk & Omland 2003; Chan & Levin 2005; Gompert *et al*. 2006a; Linnen & Farrell 2007; McGuire *et al*. 2007; Forister *et al*. 2008). Alternatively, mitochondrial introgression may simply be easier to detect, as mitochondrial DNA differentiation generally occurs more rapidly than nuclear DNA differentiation.

Correspondence: Zachariah Gompert, Fax: (307) 766-2851;
E-mail: zgompert@uwyo.edu

Numerous potential explanations for the prevalence of mitochondrial introgression have been proposed. The traditional explanation for extensive mitochondrial introgression is that, unlike nuclear alleles that are foreign to a population, foreign mitochondrial alleles are comparatively neutral and not linked to other alleles with deleterious affects in their novel genetic and/or external environment (Martinsen *et al*. 2001; Funk & Omland 2003). In addition, several hypotheses involving sexual selection have been proposed to account for mitochondrial introgression (summarized in Chan & Levin 2005). Wirtz (1999) suggested that, as females are generally the choosier sex, most hybrid mating should be between females of the less common species and males of the more common species, which would lead to increased mitochondrial introgression relative to nuclear introgression. Rieseberg *et al*. (1996) proposed a similar mechanism involving pollen competition in some plants. A theoretical study by Chan & Levin (2005) demonstrated that mitochondrial introgression occurs much more readily than nuclear introgression where hybridizing populations are primarily isolated by prezygotic barriers to gene flow. They found that mitochondrial introgression was particularly rapid when immigrants are rare relative to native individuals (Chan & Levin 2005).

Mitochondrial introgression might be facilitated by selective sweeps once a novel mitochondrial allele has entered a new population. A selective sweep can result from direct and/or indirect positive selection on the immigrant mitochondrial allele (Hurst & Jiggins 2005). Direct positive selection occurs when the immigrant mitochondrial allele increases fitness of individuals in the recipient population. A growing body of literature indicates that positive selection on mitochondrial DNA (mtDNA) might be more common than previously thought (Mishmar *et al*. 2003; Ballard & Whitlock 2004; Bensch *et al*. 2006; Grant *et al*. 2006). The lack of recombination within the mitochondrial genome means that hitch-hiking effects should be more common than in the nuclear genome, as selection on any portion of the mitochondrial genome will affect the entire mitochondrial genome. Indirect positive selection results from linkage disequilibrium between a foreign mitochondrial allele and other maternally inherited genetic material. For example, in species such as birds and butterflies where females are the heterogametic sex, meiotic drive involving a driving W chromosome can lead to an excess of female offspring and thus an increase in the frequency of any mitochondrial alleles that are in positive linkage disequilibrium with the driving W chromosome (Jaenike 2001; Hoekstra 2003; Ballard & Whitlock 2004).

Another common source of indirect selection on mtDNA in arthropods involves maternally inherited microorganisms (Hurst & Jiggins 2005). Parasitic maternally inherited microorganisms possess a number of mechanisms to ensure the survival and proliferation of the infected daughters of their host, often at the expense of infected males and uninfected individuals. Common mechanisms include producing female-biased sex ratios in the offspring of infected females (either by feminization or abortion of male offspring) and cytoplasmic incompatibility, which results in death during embryogenesis when infected males mate with uninfected females (Turelli *et al*. 1992; Hurst & Jiggins 2005). The most well-known endoparasitic microorganism infecting arthropods is the bacterium *Wolbachia*, which infects approximately 20% of insect species at any one time (Werren & Windsor 2000; Jiggins *et al*. 2001). *Wolbachia* infection has been implicated in cases of mitochondrial selective sweeps following introgressive hybridization (Jiggins 2003; Rasgon *et al*. 2003; Hurst & Jiggins 2005).

Here we examine geographical patterns of nuclear and mitochondrial genetic variation in North American *Lycaeides* (Lepidoptera), a group in which hybridization has been investigated under a variety of circumstances, including the formation of a hybrid species (Gompert *et al*. 2006a, b). Specifically, we investigate patterns of genetic differentiation and introgressive hybridization, and subsequently evaluate possible causes of widespread mitochondrial introgression in this group. *Lycaeides* is a holarctic genus with three nominal species occurring in North America: *L. idas*, *L. melissa*, and a recently described homoploid hybrid species that occupies the alpine region of the Sierra Nevada, hereafter referred to as *L.* sp. (alpine) (Nabokov 1949; Scott 1986; Brock & Kaufman 2003; Gompert *et al*. 2006b) (Fig. S1, Supporting information). Substantial variation exists among North American *Lycaeides* populations in larval host plant use (Scott 1986; Nice *et al*. 2002), female oviposition preference (Nice *et al*. 2002; C. C. Nice, unpublished data), egg adhesion properties (Fordyce & Nice 2003; Forister *et al*. 2006), habitat (Scott 1986; Gompert *et al*. 2006b), male genital morphology (Nabokov 1943, 1949; Lucas *et al*. 2008), wing pattern elements (Nabokov 1944, 1949; Fordyce *et al*. 2002; C. C. Nice & Z. Gompert, unpublished data), and male mate preference (Fordyce *et al*. 2002). While *L. idas* and *L. melissa* are not known to co-occur at a local scale, they are broadly sympatric over a large portion of their range, providing the opportunity for hybridization and introgression between these differentiated entities. Hybridization between these species has occurred in the past and was responsible for the origin of *L.* sp. (alpine) in the Sierra Nevada (Gompert *et al*. 2006b). Furthermore, morphological data indicate that hybridization may occur between *L. idas* and *L. melissa* in and around the Rocky Mountains (Nabokov 1949, 1952). Based on the degree of local specialization and prevalence of hybridization, North American *Lycaeides* is best considered a species complex. We employ the current taxonomic classification of North American *Lycaeides* throughout this manuscript to facilitate communication, but this does not imply that we believe these names describe distinct evolutionary lineages and/or reproductively isolated

entities. Rather, these names describe collections of populations that possess certain co-occurring morphological characters and may or may not be isolated from other groups of populations.

Populations of the US federally endangered *Lycaeides melissa samuelis* west of Lake Michigan possess the cytochrome oxidase c subunit I (COI)/cytochrome oxidase c subunit II (COII) mitochondrial haplotype h01, which they share with adjacent populations of *L. m. melissa* (Gompert *et al*. 2006a). *Lycaeides m. samuelis* populations east of Lake Michigan possess a different, divergent mitochondrial haplotype. As *L. m. samuelis* east and west of Lake Michigan are similar to each other and differentiated from *L. m. melissa* based on nuclear DNA, Gompert *et al*. (2006a) concluded that the presence of mitochondrial haplotype h01 in the western *L. m. samuelis* populations was likely due to introgression from nearby *L. m. melissa* populations. This mitochondrial haplotype also occurs in some *L. idas* populations in the western USA, *L. idas* populations from the Rocky Mountains, Alberta, and Alaska, and an unnamed *Lycaeides* population in the Warner Mountains of northern California (Gompert *et al*. 2006a, 2008). Haplotype h01 is by far the most common mitochondrial haplotype in North American *Lycaeides* populations (Gompert *et al*. 2008). The ubiquity of this mitochondrial haplotype and its presence in multiple taxonomic entities raises the possibility that the current distribution of this haplotype is the result of geographically widespread mitochondrial introgression facilitated by direct and/or indirect positive selection.

In this study, we describe mitochondrial and nuclear population genetic structure in the North American *Lycaeides* species complex and ask the following questions: (i) To what extent are geographical patterns of mtDNA and nDNA variation concordant at a continental scale? (ii) Is there evidence of introgressive hybridization? (iii) Is there evidence of selective sweeps of mtDNA associated with mitochondrial haplotype h01?, (iv) Is there evidence of direct and/or indirect selection associated with haplotype h01? The last question is addressed with both population genetic data and ecological experiments.

## Materials and methods

### *Collection and DNA extraction*

Adult *Lycaeides* were collected from 28 populations in North America (Table 1). Both males and females were collected from most *Lycaeides* populations, but only males were collected from *Lycaeides melissa samuelis* populations (with the exception of two females collected at Saratoga Springs, New York) in accord with US Fish and Wildlife Service permit PRT842392. DNA was isolated following the methods of Hillis *et al*. (1996) and Brookes *et al*. (1997). In no cases were multiple taxonomic entities encountered at a single locality.

### *Molecular markers*

We sequenced portions of the mitochondrial genes COI and COII for 9–11 individuals from each population sampled. Polymerase chain reaction (PCR) was performed using the primer pair C1-J-1751/C1-N-2191 for *COI* (Simon *et al*. 1994) and Pierre/Eva for COII (Caterino & Sperling 1999). This yielded fragments of approximately 450 bp and 550 bp for COI and COII, respectively. Fluorescently labelled dideoxy terminators were used for single-stranded sequencing reactions for both COI and COII. Labelled amplicons were separated and visualized using a Beckman 8800 automated sequencer (Beckman Coulter Inc.). Sequences were aligned using Sequencher 4.2.2., or by eye. The mtDNA data used in this manuscript were originally published in Gompert *et al*. (2008).

Amplified fragment length polymorphism (AFLP) marker profiles were produced for 7–32 individuals from each of the 28 populations sampled. AFLP profiles were generated using three selective primer pairs: *Eco*RI-ACA and *Mse*I-CTTG, *Eco*RI-ACA and *Mse*I-CTTA, *Eco*RI-AGT and *Mse*I-CTTA. Amplicons were separated and visualized on 6% denaturing polyacrylamide gels, using an ABI PRISM 377 DNA sequencer. GeneScan was used to visualize AFLP bands, which were sized by comparison to a standard ladder (ROX standard, Applied Biosystems Inc.) added to each lane. Band selection and quality control were performed following previously described methods (Gompert *et al*. 2006a), which were shown to yield highly reproducible results. This procedure generated 251 polymorphic AFLP markers.

### *Population genetic structure*

To visualize mitochondrial sequence variation for the sampled *Lycaeides*, we constructed a maximum-parsimony haplotype network using TCS 1.2.1 (Clement *et al*. 2000), which employs the statistical algorithms of Templeton *et al*. (1992). We used spatial analysis of molecular variance (SAMOVA) (Dupanloup *et al*. 2002) to identify natural discontinuities in the geographical distribution of mtDNA variation among North American *Lycaeides* populations. SAMOVA assigns populations to a specified number of geographically continuous clusters in an effort to maximize that proportion of molecular variation partitioned among clusters ($\phi_{CT}$). We estimated $\phi_{CT}$ with the number of clusters ($k$) set from 2 to 12. We plotted $k$ vs. $\phi_{CT}$ to aid in selecting the number(s) of clusters that best explained the mtDNA data.

The program STRUCTURE version 2.2 (Pritchard *et al*. 2000; Falush *et al*. 2007) was used to cluster individuals based on their AFLP banding profiles, which allowed us to identify groups or populations of genomically similar individuals. STRUCTURE version 2.2 employs a model-based Bayesian

**Table 1** Population data. Population number, nominal taxon, population name, mtDNA haplotypes, mtDNA cluster from SAMOVA for $k = 3$ and $k = 7$ respectively, and AFLP cluster with highest assignment probability under the admixture model for $k = 3$ and $k = 7$, respectively from Structure version 2.2. Sample size ($n$) for mtDNA and AFLP data is given

| Population no. | Taxa | Population | mtDNA ($n$) | Mitochondrial DNA clusters | AFLP clusters ($n$) |
|---|---|---|---|---|---|
| 1 | *Lycaeides idas anna* | Donner Pass, CA | h18(1) h19(1) h20(5) h03(3) | m3.1 m7.1 | n3.1 n7.4 (18) |
| 2 | | Leek Springs, CA | h18(8) h29(2) | m3.1 m7.1 | n3.1 n7.1 (28) |
| 3 | | Trap Creek, CA | h20(5) h21(1) h43(1) h44(1) h45(2) | m3.1 m7.1 | n3.1 n7.1 (24) |
| 4 | | Yuba Gap, CA | h18(8) h29(2) | m3.1 m7.1 | n3.1 n7.1 (29) |
| 5 | *Lycaeides idas azureus* | Indian Valley, CA | h23(4) h24(2) h25(1) h26(3) | m3.1 m7.1 | n3.1 n7.2 (7) |
| 6 | *Lycaeides idas ricei* | Cave Lake, CA | h01(10) | m3.2 m7.2 | n3.1 n7.2 (15) |
| 7 | | Deadfall Mdw., CA | h16(6) h17(4) | m3.1 m7.1 | n3.1 n7.1 (17) |
| 8 | | Marble Mts., CA | h31(2) h32(1) h33(7) | m3.1 m7.3 | n3.1 n7.2 (16) |
| 9 | | Mt. Ashland, OR | h01(10) | m3.2 m7.2 | n3.1 n7.2 (17) |
| 10 | *Lycaeides melissa anneta* | Alta, UT | h01(2) h46(8) | m3.2 m7.2 | n3.2 n7.6 (18) |
| 11 | *Lycaeides melissa inyoensis* | Big Pine, CA | h01(10) | m3.2 m7.2 | n3.2 n7.5 (24) |
| 12 | *Lycaeides melissa melissa* | Beckwourth Pass, CA | h01(1) h02(8) | m3.3 m7.4 | n3.2 n7.4 (22) |
| 13 | | Brandon, SD | h01(5) h04(2) h05(1) h06(1) h07(1) | m3.2 m7.2 | n3.2 n7.6 (24) |
| 14 | | Cedarville, CA | h01(1) h08(8) h09(1) | m3.2 m7.2 | n3.2 n7.6 (24) |
| 15 | | Garderville, CA | h01(3) h02(7) | m3.3 m7.5 | n3.2 n7.5 (27) |
| 16 | | Montague, CA | h01(9) h34(1) | m3.2 m7.2 | n3.2 n7.4 (19) |
| 17 | | Sierravalley, CA | h01(10) | m3.2 m7.2 | n3.2 n7.5 (27) |
| 18 | | Spring Creek, SD | h01(8) h04(1) h39(1) | m3.2 m7.2 | n3.2 n7.6 (28) |
| 19 | | Verdi, NV | h02(10) h47(1) | m3.3 m7.5 | n3.2 n7.5 (23) |
| 20 | *Lycaeides Melissa samuelis* | Fish Lake, WI | h01(5) | m3.2 m7.2 | n3.3 n7.7 (20) |
| 21 | | Fort McCoy, WI | h01(5) | m3.2 m7.2 | n3.3 n7.7 (19) |
| 22 | | Indiana Dunes, IN | h22(5) | m3.2 m7.6 | n3.3 n7.7 (22) |
| 23 | | Necedah, WI | h01(5) | m3.2 m7.2 | n3.3 n7.7 (23) |
| 24 | | Saratoga, NY | h22(5) | m3.2 m7.6 | n3.3 n7.7 (27) |
| 25 | *Lycaeides* sp. (alpine) | County Line Hill, CA | h02(3) h10(1) h11(2) h12(2) h13(2) | m3.1 m7.7 | n3.1 n7.3 (20) |
| 26 | | Carson Pass, CA | h03(8) h14(1) h15(1) | m3.1 m7.7 | n3.1 n7.3 (32) |
| 27 | | Eagle Pk., CA | h01(10) | m3.2 m7.2 | n3.1 n7.3 (28) |
| 28 | | Mt. Rose, NV | h03(1) h35(9) | m3.1 m7.7 | n3.1 n7.3 (18) |

clustering algorithm to assign individuals probabilistically to clusters to minimize deviations from linkage equilibrium. Analyses were conducted using a model allowing for recessive alleles, which is ideal for dominant molecular markers such as AFLPs (Falush *et al*. 2007). Both admixture and no-admixture models were used, and runs were conducted using a Markov chain Monte Carlo (MCMC) with 500 000 generations with an initial burn-in of 50 000 generations. Prior information regarding the population or species from which an individual was sampled was ignored. The number of clusters ($k$) was evaluated from 1 to 28 for both the admixture and no admixture models and 10 independent MCMC runs were conducted for each $k$ for both models. We then plotted $k$ vs. the mean log-likelihood for each $k$ to aid in selecting the number(s) of clusters that best explained the AFLP data.

The log-likelihoods of the data for STRUCTURE version 2.2 analyses conducted under both the admixture and no-admixture models were compared using Bayes factors for three and seven clusters (see Results) to determine whether the AFLP data were better explained when the genome of each individual was assumed to be drawn from a single cluster (no-admixture model) or proportionally from $k$ clusters (admixture model). Statistical tests using Bayes factors are similar to likelihood-ratio tests; however, unlike likelihood-ratio tests, the models compared need not be nested (Kass & Raftery 1995). The Bayes factor (B10) gives the strength of evidence against model 0 (in this case the no-admixture model) relative to model 1 (in this case the admixture model) and is calculated as the ratio of the model likelihoods (Kass & Raftery 1995). For each assumed number of clusters, the run with the best log-likelihood was selected for computation of the Bayes factor.

While methods exist to quantify discordance among phylogenetic trees (e.g. Shimodaira & Hasegawa 1999; Goldman *et al*. 2000), similar methods are lacking for population genetic data. Therefore, we employed a novel model-based Bayesian technique using the software program STRUCTURE version 2.2 to quantitatively test for discordance between patterns of mitochondrial and nuclear genetic variation. Specifically, we computed the log-likelihood of our AFLP data under a no-admixture model with three

and seven clusters assumed (see Results) as well as under a no-admixture model with three and seven clusters assumed, but with individuals constrained to clusters consistent with SAMOVA results based on the mtDNA data. Constrained runs for three and seven clusters were conducted as described above, but individual's assignments were constrained via the USEPOPINFO command. We then tested for discordance by comparing the likelihood of the AFLP data under unconstrained and constrained models using Bayes factors.

### *Introgressive hybridization and the distribution of h01*

The presence of mitochondrial haplotype h01 in multiple distinct nuclear backgrounds (see Results) may be due to retained ancestral polymorphism for the mitochondrial genome and/or the spread of haplotype h01 via introgressive hybridization. If the latter is true, populations possessing haplotype h01 should form a geographically contiguous group, whereas, such a geographical pattern would not be predicted due to ancestral polymorphism. As SAMOVA identifies geographically contiguous groupings of genetically similar populations, SAMOVA was used to address this question. Based on the SAMOVA analysis described previously (see Population genetic structure), we asked whether populations possessing haplotype h01 were assigned to a single cluster.

Two *Lycaeides melissa melissa* populations (Gardnerville, California and Beckwourth Pass, California), which possess both mitochondrial haplotype h01 and an additional divergent haplotype, provide the opportunity for an additional test regarding the distribution of mitochondrial haplotype h01. Both mitochondrial and nuclear genetic diversity are expected to be proportional to a population's effective population size. However, at least initially, introgression of a divergent allele could lead to an increase in mitochondrial genetic diversity relative to a population's nuclear genetic diversity. Thus, if the Gardnerville, California and Beckwourth Pass, California *L. m. melissa* populations possess haplotype h01 due to recent intrgression, as opposed to ancestral polymorphism, they should have a higher ratio of mitochondrial genetic diversity to nuclear genetic diversity than other *L. m. melissa* populations. To test this hypothesis we compared the ratio of mitochondrial to nuclear diversity for the eight *L. m. melissa* populations sampled. As a measure of mtDNA nucleotide diversity, $\pi$ (Tajima 1983, 1993) was calculated for each population using Arlequin 2.0 (Schneider *et al.* 2000). To estimate nuclear genetic diversity, we employed the software AFLP-SURV version 1.0 (Vekemans *et al*. 2002) to calculate unbiased expected heterozygosity (Nei 1973) based on the AFLP data for each population following the method described by Lynch & Milligan (1994) for dominant genetic markers. These estimates were made using a Bayesian method with a non-uniform prior distribution of allele frequencies and assuming no deviation from Hardy–Weinberg genotype frequencies. The ratio of mtDNA diversity to nuclear DNA diversity for each population was simply calculated by dividing $\pi$ for the mtDNA by unbiased expected heterozygosity for the AFLP marker data. We then tested if this mitochondrial–nuclear diversity ratio was higher for the Gardnerville, California and Beckwourth, California populations than the six other *L. m. melissa* populations using a Wilcoxon rank-sum test with the Stats package implemented in R (R Development Core Team 2008).

### *Test for selective sweep*

Positive selection favouring a specific mitochondrial allele, as opposed to demographic factors, will reduce mitochondrial variation with little effect on the nuclear genome. Thus, if mitochondrial haplotype h01 was spread by selective sweeps, its presence in populations should be associated with a reduction in the ratio of mitochondrial to nuclear diversity. As mentioned in the previous paragraph, this reduction in mitochondrial diversity should be preceded by an initial increase in mitochondrial diversity, but this should have little effect overall, as most populations do not show evidence of very recent introgression (i.e. the presence of highly divergent haplotypes). To test this hypothesis, we compared the ratio of mitochondrial to nuclear diversity for the 28 sampled populations. This diversity ratio was calculated as described in the preceding paragraph. We then tested if this mitochondrial–nuclear diversity ratio was lower for populations where haplotype h01 was present than populations where haplotype h01 was absent using a Wilcoxon rank-sum test with the Stats package implemented in R (R Development Core Team 2008).

### *Direct selection and h01*

In order to evaluate the potential influence of mitochondrial haplotype h01 on the performance and fitness of *Lycaeides*, we collected and genotyped (as described above) females from a population (Beckwourth Pass, California) known to have a mix of mitochondrial haplotypes h01 and h02. Sixteen females were confined individually to oviposition arenas with sprigs of host plants, and offspring from each female were split into two groups and reared on a native host plant, *Astragalus canadensis*, and a non-native host plant, *Medicago sativa*, that is also used by this population and many others throughout western North America. Larvae were reared individually in microcentrifuge tubes (in early instars), and finally in Petri dishes, and given fresh leaves *ad libitum* throughout. Fresh adult weight was recorded to the nearest hundredth of a milligram on a Mettler Toledo microbalance, and the fraction of individuals surviving to eclosion in each family

was recorded, as well as development time. As an additional metric of performance, adult longevity was measured by confining adults to small, clear plastic drinking cups with a supply of artificial nectar in a greenhouse at 30 °C. Longevity was recorded as the number of days butterflies survived. Analyses of variances (ANOVA) using restricted maximum likelihood (REML) mixed models (Littell *et al.* 1996) were used with JMP-IN software, version 7.0, to analyse data from performance experiments. In these models, haplotype was a fixed effect, as was host plant and sex; random factors included family and interactions with family (family was nested within haplotype).

### *Indirect selection and h01*

We tested for evidence of sex ratio distortion in terms of excess female offspring associated with mitochondrial haplotype h01, which could result from a driving W chromosome or parasitic infection. We counted the number of female and male offspring from the wild-caught Beckwourth Pass females following eclosion as adults (we were unable to determine the sex of larvae) (see above for details on larval rearing). First, we tested if the proportion of female offspring from females with haplotype h01 and/or females with haplotype h02 was significantly greater than 0.5 with a one-sample Wilcoxon rank sum test using the Stats package implemented in R (R Development Core Team 2008). We then tested if the proportion of female offspring differed between females with haplotype h01 and h02 using a $\chi^2$ goodness of fit test with the JMP-IN software version 7.0 (SAS Institute).

We used a PCR-based test to determine whether there was a positive association between haplotype h01 and *Wolbachia* infection. Such an association would be expected if the spread of mitochondrial haplotype h01 is facilitated by linkage disequilibrium with *Wolbachia* infection. We PCR-assayed a total of 109 *Lycaeides* individuals with known mitochondrial haplotypes. Specifically, we assayed 39 *L. m. samuelis* individuals from four populations including two that were fixed for haplotype h01 (Necedah, Wisconsin and Fort McCoy, Wisconsin) and two that did not possess haplotype h01 (Indiana Dunes, Indiana and Saratoga Springs, New York), 41 *Lycaeides idas ricei* individuals from four populations including two that were fixed for haplotype h01 (Cave Lake, California and Mt. Ashland, Oregon) and two that did not possess haplotype h01 (Marble Mountains, California and Deadfall Meadow, California), and 29 *L. m. melissa* individuals from two populations (Gardnerville, California and Beckwourth Pass, California) with a mixture of h01 and other mitochondrial haplotypes. Primers for the amplification of *Wolbachia* 16 s rDNA were used for PCR, and universal arthropod primers for 28 s rDNA were used to verify that PCR inhibitors were not present in the reactions. Standard PCR protocols were used for the amplification of *Wolbachia* and arthropod rDNA. Products were visualized on 1% agarose gel. We then tested for a positive association between the mitochondrial haplotype h01 and *Wolbachia* infection using Fisher's exact test as implemented in the R Stats package (R Development Core Team 2008). This test was performed separately for *L. m. samuelis*, *L. i. ricei*, and *L. m. melissa*.

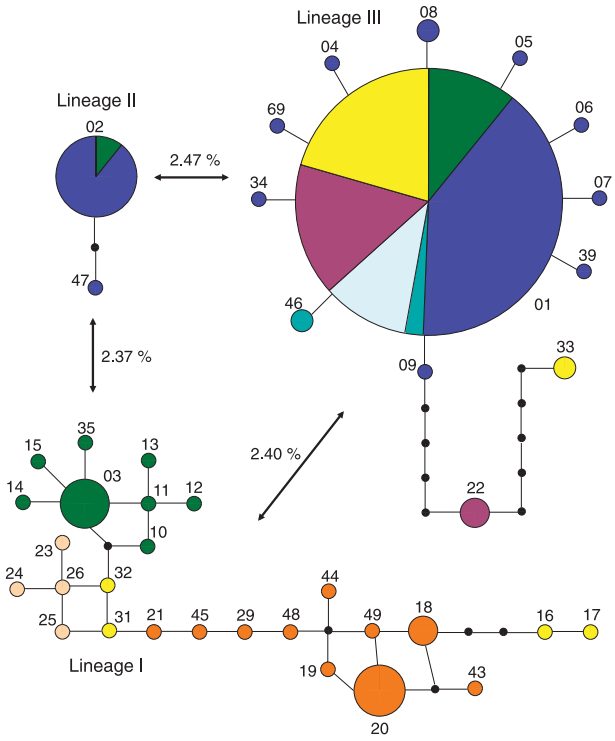


**Fig. 1** Maximum parsimony network for mtDNA haplotypes. A circle represents each haplotype and the area of each circle is approximately proportional to the frequency of the haplotype. Each haplotype is shaded according to the proportion of individuals classified as *Lycaeides i. anna* (orange), *L. i. azureus* (tan), *L. i. ricei* (yellow), *L. m. anneta* (aqua), *L. m. melissa* (dark blue), *L. m. inyoensis* (light blue), *L. m. samuelis* (violet), *L.* sp. (alpine) (green). The mean pair wise sequence divergence between each pair of networks is given. Lineage designations correspond to the major mitochondrial lineages detected by Gompert *et al.* (2008).

## Results

### *Population genetic structure*

We identified 40 unique mitochondrial haplotypes for the combined COI (408 bp) and COII (507 bp) sequence data (GenBank Accession nos EU409326–EU409354; Table 1). A single haplotype network could not be constructed at the 95% confidence level. Instead, three separate haplotype networks were produced; mean pairwise percent sequence divergence between each pair of haplotype networks was similar (2.37–2.47%) (Fig. 1). These three haplotype

networks correspond to the three major mitochondrial lineages detected in *Lycaeides* by Gompert *et al*. (2008). Mitochondrial haplotype h01, which was by far the most common haplotype with an overall frequency of 0.37, was sampled from *L. idas*, *L. melissa*, and the unnamed Warner Mountain population. Haplotypes from different networks (mitochondrial lineages) were detected in four populations (Gardnerville, California, Beckwourth Pass, California, County Line Hill, Nevada, and Marble Mountains, California); Gardnerville and Beckwourth Pass possessed haplotypes h01 (lineage III) and h02 (lineage II), the County Line Hill population possessed h02 and several less common haplotypes from mitochondrial lineage I, and the Marble Mountains population possessed haplotypes from lineages I and III (Fig. 1; Gompert *et al*. 2008). Of the 28 populations sampled, 13 were either fixed for haplotype h01 or contained haplotype h01 and a subset of the relatively uncommon haplotypes one mutational step from haplotype h01 (Fig. 1).

A significant proportion of molecular variance was partitioned among groups for all $k$ examined, and $\phi_{CT}$ generally increased with increasing values of $k$ (Fig. 2a). Here we report results for three and seven groups, as these values capture population genetic structure for mtDNA at coarse-grain and fine-grain levels, respectively. No values of $k$ cluster populations in a manner that contradicts the clustering scheme at other values of $k$. That is, the clusters observed for $k = 7$ consisted of divisions of clusters observed at $k = 3$ and no novel combinations of individuals not observed at lower $k$ were detected. When three groups were assumed ($\phi_{CT} = 81.16$, $P < 0.00001$), populations were clustered in accordance with the three major mitochondrial lineages described above (Fig. 3a). When seven groups were assumed ($\phi_{CT} = 87.01$, $P < 0.00001$), each of the clusters from $k = 3$ was further divided in accord with finer-grain population structure (Fig. 4a). For seven clusters a subset of *L. idas* and *L. melissa* populations and the Warner Mountain population, which span a large geographical range, were assigned to a single mitochondrial cluster (cluster m7.2) dominated by haplotype h01 (Fig. 4a).

For both the admixture and no-admixture Structure clustering analyses, the mean log-likelihood of the data generally increased with increasing values of $k$, as did the variance in the log-likelihood of the data (Fig. 2b, c). Here we report results for three and seven clusters, as these values capture population genetic structure based on nuclear DNA at coarse-grain and fine-grain levels, respectively, and are easily comparable to the mitochondrial data. Bayes factor analysis indicated that there is decisive evidence that the data are better explained under the admixture model than the no-admixture model for both three ($\ln L_{admix} = -38534.3$, $\ln L_{no\text{-}admix} = -38741.1$, $\ln B_{admix/no\text{-}admix} = 206.8$; see Table 2 for interpretation) and seven clusters ($\ln L_{admix} = -36302.4$, $\ln L_{no\text{-}admix} = -36525$, $\ln B_{admix/no\text{-}admix} = 222.6$; see

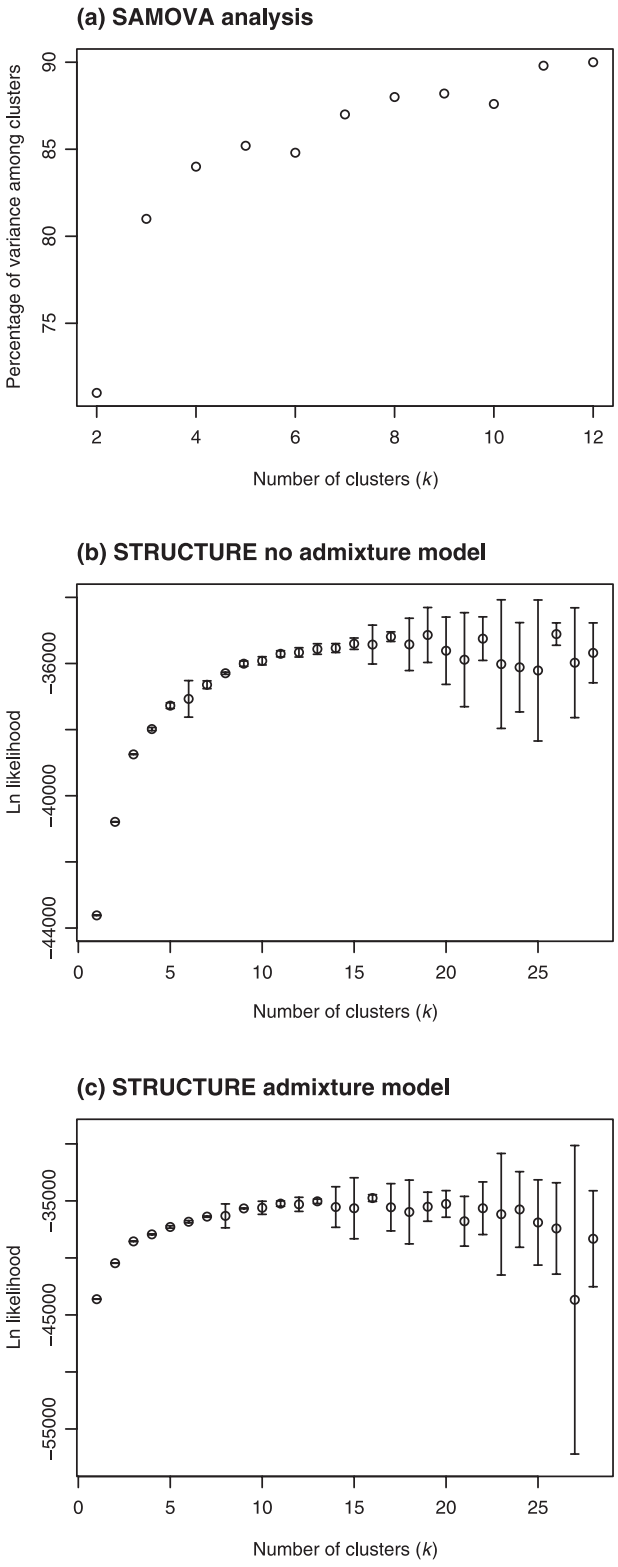


**Fig. 2** Plots of $k$ versus $\phi_{CT}$ for the mtDNA data (a), $k$ versus ln likelihood for the AFLP data under the no-admixture model (b), and $k$ versus ln likelihood for the AFLP data under the admixture model (c). For (b) and (c) values plotted represent means from 10 MCMC runs and vertical bars denote one standard deviation.

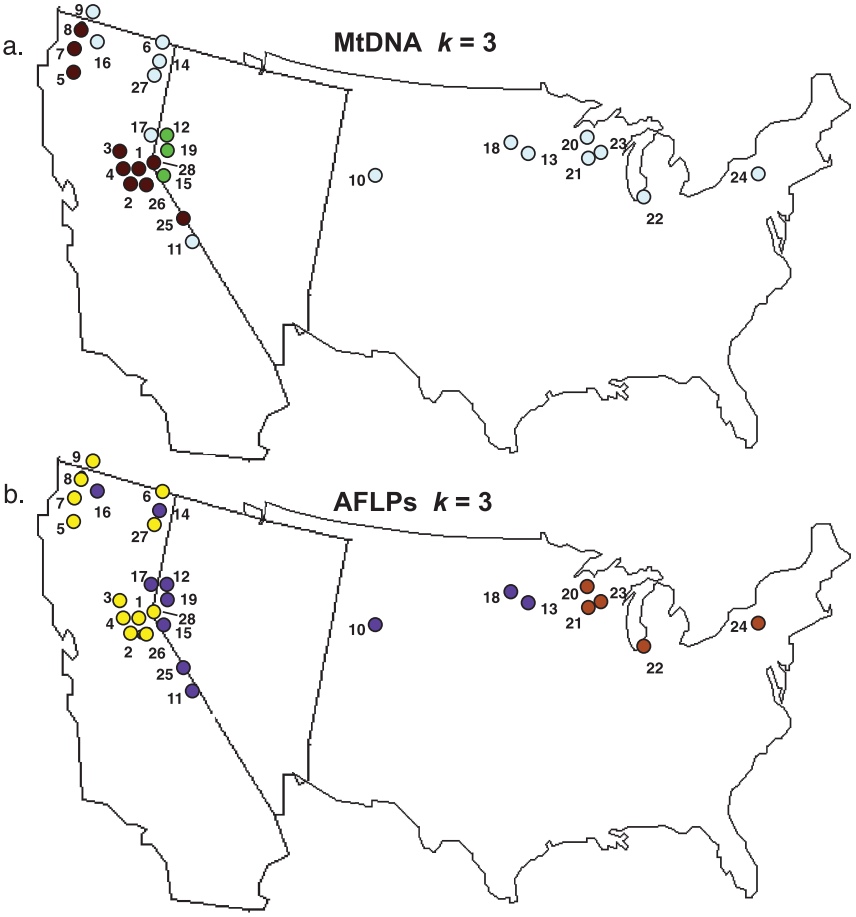


**Fig. 3** Comparison of patterns of mitochondrial (a) and nuclear (AFLP) (b) genetic variation when three clusters are assumed. The map depicts the USA with enlarged outlines of California and Nevada. Coloured circles represent populations, and each population is labelled with a population number from Table 1. Population circles are coloured based on the mtDNA cluster for each population: m3.1 (dark red), m3.2 (light turquoise), and m3.3 (bright green) (a) or the AFLP cluster with the highest assignment probability averaged across all individuals for that population: n3.1 (yellow), n3.2 (midnight blue), and n3.3 (brown) (b). Populations assigned to m3.2 have mitochondrial haplotype h01 and a wide variety of nuclear backgrounds.

**Table 2** Interpretation of Bayes factors adapted from Kass and Raftery (1995). $B_{10}$ designates the Bayes factor for model 1 ($M_1$) relative to model 0 ($M_0$)

| $B_{10}$ | $\ln(B_{10})$ | Strength of evidence against $M_0$ |
|---|---|---|
| 1 to 3 | 0 to 1.1 | Not worth more than a bare mention |
| 3 to 20 | 1.1 to 2.5 | Positive |
| 20 to 150 | 2.5 to 5.0 | Strong |
| > 150 | > 5.0 | Decisive |

Table 2 for interpretation), thus, only results from the admixture model are presented. When three clusters were assumed *L. idas* and *L.* sp. (alpine) individuals were generally assigned to one cluster (cluster n3.1), although some *Lycaeides idas ricei*, *L.* sp. (alpine), and Warner Mountain individuals were assigned with moderate to high probabilities to cluster n3.2 (Fig. 3b and Fig. S2, Supporting information). *Lycaeides melissa anneta*, *Lycaeides m. inyoensis*, and *L. m. melissa* individuals were assigned to another cluster (cluster n3.2), and *Lycaeides melissa samuelis* individuals were assigned to a third cluster (cluster n3.3) (Fig. 3b and S2). When seven clusters were assumed, *Lycaeides idas anna*, with the exception of some individuals from Donner Pass, California, were assigned to one cluster (cluster n7.1), while *L. i. ricei* and *Lycaeides idas azureus* individuals were assigned to another cluster (cluster n7.2) (Fig. 4b and S2). *Lycaeides* sp. (alpine) individuals and the Warner Mountain population were assigned to a third cluster (cluster n7.3) (Fig. 4b and S2). *Lycaeides m. melissa* individuals were assigned with high probability to one of three clusters in a manner consistent with geographical isolation (clusters n7.4, n7.5, and n7.6) (Fig. 4b and S2). *Lycaeides m. inyoensis* individuals were assigned with several *L. m. melissa* populations from California and Nevada to cluster n7.5, while *L. m. anetta* individuals were assigned with one *L. m. melissa* population from California and two from South Dakota to cluster n7.6 (Fig. 4b and S2). *Lycaeides m. samuelis* individuals were assigned to a seventh cluster (cluster n7.7) (Fig. 4b and S2). The Donner Pass, California *L. i. anna* population was assigned with moderate probability to both cluster n7.1 (with other *L. i. anna* individuals) and cluster n7.4 (with some California *L. m. melissa* populations) (Fig. 4b and S2). The *L. i. ricei* individuals from Deadfall Meadow, California were assigned with moderate probability to both clusters n7.2 (with other *L. i. ricei*) and n7.1 (with *L. i. anna* populations) (Fig. 4b and S2).

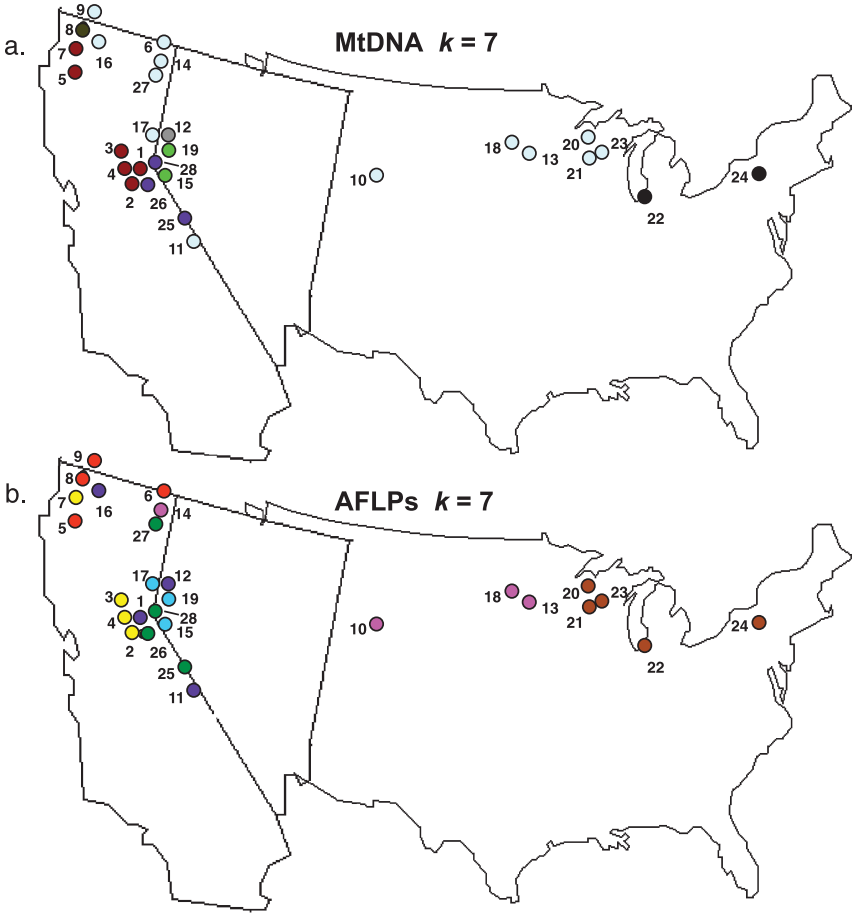


**Fig. 4** Comparison of patterns of mitochondrial (a) and nuclear (AFLP) (b) genetic variation when seven clusters are assumed. The map depicts the USA with enlarged outlines of California and Nevada. Coloured circles represent populations, and each population is labelled with a population number from Table 1. Population circles are coloured based on the mtDNA cluster for each population: m7.1 (dark red), m7.2 (light turquoise), m7.3 (olive green), m7.4 (grey), m7.5 (bright green), m7.6 (black), and m7.7 (black) (a), or the AFLP cluster with the highest assignment probability averaged across all individuals for that population: n7.1 (yellow), n7.2 (red), n7.3 (green), n7.4 (midnight blue), n7.5 (sky blue), n7.6 (pink), and n7.7 (brown) (b). Populations assigned to m7.2 have mitochondrial haplotype h01 and a wide variety of nuclear backgrounds.

Our Bayesian analyses of mito-nuclear discordance indicated that there was decisive evidence against a model where the AFLP data were constrained to be consistent with the mitochondrial data relative to an unconstrained no-admixture model for both three ($\ln L_{constrained} = -40\,663$, $\ln L_{unconstrained} = -38741.1$, $\ln B_{constrained/nconstrained} = 1921.9$; see Table 2 for interpretation) and seven clusters ($\ln L_{constrained} = -38743.8$, $\ln L_{unconstrained} = -36\,525$, $\ln B_{constrained/nconstrained} = 2218.8$). For an interpretation of Bayes factors based on Kass & Raftery (1995) see Table 2.

### *Introgressive hybridization and the distribution of h01*

With the exception of two populations (Gardnerville, California and Beckwourth Pass, California), all populations where haplotype h01 was present were identified as a single cluster using SAMOVA. This was true for $k = 1$–12. The Gardnerville and Beckwourth Pass populations contained haplotype h01 and a highly divergent haplotype, h02, with the latter occurring at a higher frequency (Table 1). These two populations generally clustered with the Verdi population, which also possessed haplotype h02 at a high frequency. The ratio of mitochondrial diversity relative to nuclear (AFLP) diversity was significantly greater for the Gardnerville, California and Beckwourth Pass, California populations than for six other *L. m. melissa* populations (Wilcoxon rank sum test, $W = 12$, $P = 0.03571$). Thus, the mitochondrial diversity for these two populations is greater relative to their nuclear genetic diversity than would be predicted based on the other sampled *L. m. melissa* populations.

### *Test for selective sweep*

The ratio of mitochondrial diversity relative to nuclear (AFLP) diversity was significantly greater for populations where haplotype h01 was absent than it was for populations where haplotype h01 was present (Wilcoxon rank sum test, $W = 58$, $P = 0.03299$) (Fig. 5). Thus, the presence of mitochondrial haplotype h01 is associated with reduced mitochondrial genetic variation relative to nuclear genetic variation. Four outlier populations were detected (Gardnerville, California, Beckwourth Pass, California, County Line Hill, Nevada, and Marble Mountains, California), which possessed mitochondrial haplotypes from two divergent mitochondrial lineages (probably because of recent/current introgression, see evidence regarding Gardnerville and Beckwourth Pass above). The removal of these populations does not alter the results of this analysis.

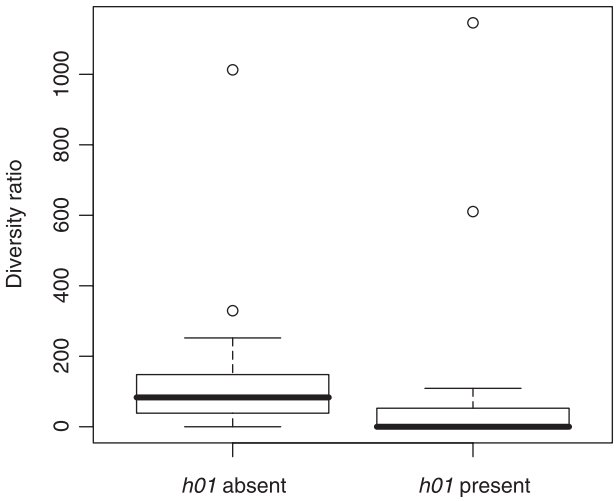


**Fig. 5** Boxplot showing median (with quantiles) mitochondrial-nuclear genetic diversity ratio with populations grouped by the presence or absence of mitochondrial haplotype *h01*. Wilcoxon rank sum test indicates a significantly lower diversity ratio for populations where haplotype *h01* is present than populations where *h01* is absent ($W = 58$, $P = 0.033$).

### *Direct selection and h01*

Females were genotyped and of the 16 larval families used in the performance experiments, 10 were found to possess haplotype h02, and six had haplotype h01. These 16 families included a total of 270 larvae, of which 106 survived to eclosion (of these, 59 were haplotype h02 and 47 were haplotype h01). We did not detect an effect of haplotype on any aspect of performance that we studied. Average survival was 37.14% (± 5.8, SE) across h01-haplotype families, and 40.1% (± 3.5, SE) across h02-haplotype families ($F_{1,14} = 0.14$, $P = 0.72$). Average adult weight was 20.73 mg (± 1.52, SE) for h01-haplotype individuals, and 20.19 mg (± 1.37, SE) for h02-haplotype individuals ($F_{1,9.8} = 1.52$, $P = 0.25$). There were no significant haplotype-by-host plant interactions, nor was there a significant effect of haplotype on development time or adult longevity.

### *Indirect selection and h01*

The mean proportion of female offspring from females with haplotypes h01 and haplotype h02 was 0.499 ($n = 4$ females) and 0.615 ($n = 10$ females), respectively. Neither of these proportions was significantly greater than 0.5 (h01: Wilcoxon rank-sum test, $V = 5$, $P = 0.565$; h02: Wilcoxon rank-sum test, $V = 31$, $P = 0.1708$). Moreover the proportion of female offspring did not differ between females with mitochondrial haplotypes h01 and h02 ($\chi^2 = 0.98$, d.f. = 1, $P = 0.32$).

All 20 *L. m. samuelis* with haplotype h01 tested positive for *Wolbachia* infection, while only one of the 19 *L. m. samuelis* that did not possess haplotype h01 tested positive for *Wolbachia* infection. This resulted in a positive association between the presence of haplotype h01 and *Wolbachia* infection in *L. m. samuelis* (Odds ratio = ∞, $P = 3.05 \times 10^{-10}$). For *L. i. ricei*, all 20 individuals with haplotype h01 and all but two of the 21 individuals that did not have haplotype h01 tested positive for *Wolbachia* infection. Similarly, all 29 *L. m. melissa* individuals tested positive for *Wolbachia* infection. Thus, we failed to detect a positive association between the presence of haplotype h01 and *Wolbachia* infection for both *L. i. ricei* (Odds ratio = ∞, $P = 0.2683$) and *L. m. melissa* (odds ratio = 0, $P = 1.00$).

## Discussion

There is strongly supported discordance in patterns of nuclear and mitochondrial variation across the North American *Lycaeides* species complex. Reduced genetic variation associated with one of the most widespread haplotypes strongly suggests that introgression and spread of mitochondrial variation between populations has been driven by natural selection. Furthermore, an association between this haplotype and *Wolbachia* infection in one lineage provides a possible mechanism by which selection has acted in at least one case.

### *Mito-nuclear discordance and introgressive hybridization*

Geographical patterns of variation for the mitochondrial and nuclear genomes are discordant in North American *Lycaeides* at both coarse and fine grains, as demonstrated by our Bayesian analysis, which found decisive evidence against a model where patterns of genetic variation for the AFLP data were constrained to be consistent with patterns of mitochondrial genetic variation. At this coarse grain ($k = 3$), regions of mito-nuclear discordance are most apparent in the vicinity of the Sierra Nevada, where all three mitochondrial clusters and nuclear clusters n3.1 and n3.2 adjoin, and in eastern North America where a major nuclear boundary is apparent that does not correspond to a boundary between mitochondrial groups (Fig. 3). Regions of mito-nuclear discordance are most apparent in the vicinity of the Sierra Nevada and near Lake Michigan at a fine grain ($k = 7$) as well. In the Sierra Nevada mitochondrial clusters m7.2, m7.3, m7.4, m7.5, and m7.7 and nuclear clusters n7.2, n7.3, n7.4, n7.5 and n7.6 adjoin, and near Lake Michigan mitochondrial clusters m7.2 and m7.6 and nuclear clusters n7.6 and n7.7 connect (Fig. 4).

We detected discordance between patterns of genetic variation for the mitochondrial and nuclear genomes in North American *Lycaeides*. The lack of recombination in the mitochondrial genome allows us to estimate genetic structure for the mitochondrial genome using a single marker. We did not attempt to determine if specific nuclear markers

exhibit patterns of genetic variation that are also discordant relative to the entire nuclear and/or mitochondrial genome. We expect that patterns of genetic variation for a random subset of nuclear markers would be discordant from the nuclear genome because of a combination of stochastic processes (e.g. genetic drift) and non-neutral forces (e.g. natural selection). However, the presence of nuclear markers discordant with the overall nuclear genome would not detract from our results and is beyond the scope of this paper, as the goal our current study was to compare overall patterns of variation between the mitochondrial genome and the nuclear genome.

The fact that populations possessing haplotype h01 form a geographically contiguous group is consistent with the hypothesis this haplotype has spread into multiple distinct nuclear genetic backgrounds via introgressive hybridization. A similar pattern might be produced by parapatric divergence with a low level of ongoing gene flow at some loci, including the mtDNA. Either scenario would predict mito-nuclear discordance at or near the boundaries of mitochondrial and nuclear entities (Funk & Omland 2003). Additionally, we found evidence that the presence of haplotype h01 in the Gardnerville and Beckwourth Pass populations, which are polymorphic for two divergent haplotypes, h01 and h02, is the result of recent introgression, as these populations possess a higher ratio of mitochondrial to nuclear genetic diversity than other sampled *Lycaeides m. melissa* populations. We cannot completely rule out the possibility that some of the observed mito-nuclear discordance is the result of incomplete lineage sorting. However, it is unlikely that incomplete lineage sorting would result in the presence of haplotype h01 in populations that form a geographically contiguous group, but have multiple distinct nuclear backgrounds.

The AFLP data indicate that nuclear hybridization may be prevalent in some locations as well. The intermediate nuclear assignments of *L.* sp. (alpine) individuals when three clusters are assumed are consistent with the hybrid origin of this species (Fig. S2) (Gompert *et al*. 2006b). Also, there is some evidence, when seven clusters are assumed, of introgression of the *L. m. melissa* nuclear genome into the *Lycaeides i. anna* population at Donner Pass, California, and of introgression of the *L. i. anna* genome into the *Lycaeides i. ricei* population at Deadfall Meadows, California (Fig. S2). However, details of nuclear introgression are difficult to determine on the basis of AFLP data alone and beyond the scope of this paper, and thus, will not be discussed further here.

### *Selection and the spread of haplotype h01*

We detected a reduction in mitochondrial genetic diversity relative to nuclear genetic diversity in populations where mitochondrial haplotype h01 was present, which is consistent with the hypothesis that the spread of this haplotype has been facilitated by natural selection (Fig. 5). Such a pattern cannot be accounted for by demographic events, as these would affect both mitochondrial and nuclear genetic diversity (Takahata 1995; Cann 2001; Ramos-Onsins & Rozas 2002; Bensch *et al*. 2006). This haplotype occurs in many distinct nuclear backgrounds and geographical localities (Table 1; Figs 3 and 4). This fact suggests that the possible selective advantage of haplotype h01 may not be associated with a specific habitat or nuclear background. The extensive geographical range over which mitochondrial haplotype h01 is found highlights the potential power of selection to structure patterns of genetic variation at individual loci.

While our results are consistent with the hypothesis that the current distribution of mitochondrial haplotype h01 is the result of hybridization followed by positive selection, we were unable to unambiguously identify specific selective pressures favouring haplotype h01. We found no evidence that haplotype h01 increased fitness (in terms of survival to eclosion, fresh adult weight, and adult longevity) relative to haplotype h02 in the *L. m. melissa* population at Beckwourth Pass, California, which is polymorphic for these divergent haplotypes. However, this does not mean that the spread of haplotype h01 has not been, in part, due to direct positive selection. We obtained relatively few families for this analysis, thus our power to detect haplotype associated fitness effects was limited. Furthermore, mitochondrial haplotype h01 may have a positive fitness-related effect on a trait or traits not measured in this study. In addition, natural selection is often dependent on current abiotic and biotic conditions and is unlikely to be constant through time (e.g. Grant & Grant 2002). Thus, mitochondrial haplotype h01 may be favoured by selection only periodically and/or under conditions different from those in the laboratory.

We found no evidence that Beckwourth Pass, California females with haplotype h01 produced offspring with a female-biased sex ratio. Therefore, we have no evidence of a driving W chromosome associated with haplotype h01. As was true above, our power to detect a female-biased sex ratio was limited by a small sample size. We found that *Lycaeides m. samuelis* individuals with mitochondrial haplotype h01 were more likely to be infected with *Wolbachia* than *L. m. samuelis* with different mitochondrial haplotypes. However, this relationship did not hold for *L. m. melissa* and *L. i. ricei* individuals, nearly all of which were infected with *Wolbachia*. The linkage disequilibrium between haplotype *h01* and *Wolbachia* infection may have contributed to the spread of haplotype *h01* among *L. m. samuelis* populations in eastern North America. *Wolbachia* facilitates its own spread through a host population by causing infected female hosts to produce female biased offspring and/or by causing uninfected females fertilized by infected males to

produce inviable offspring (Jiggins *et al*. 2001; Hurst & Jiggins 2005). Assuming the same *Wolbachia* strain is present in *L. m. samuelis* and the *L. m. melissa* population at Beckwourth pass, the latter mechanism seems more likely in this case, as females with haplotype *h01* did not have offspring with a female biased sex ratio.

Based on our findings it seems unlikely that indirect selection stemming from linkage disequilibrium with *Wolbachia* is the sole cause of the current high frequency of mitochondrial haplotype h01. Additional direct and/or indirect selective forces (as discussed above) may have contributed to the spread of haplotype h01. Alternatively, the lack of an association between the presence of haplotype h01 and *Wolbachia* infection in *L. m. melissa* and *L. i. ricei* may result from occasional horizontal spread of *Wolbachia* infection. While generally rare, horizontal spread of *Wolbachia* does occur (Hurst & Jiggins 2005). Additionally, different *Lycaeides* populations or even individuals may be infected with distinct strains of *Wolbachia* with varying affects on *Lycaeides* reproductive biology. Multiple *Wolbachia* strains have been found within single butterfly species in the past (Jiggins *et al*. 2001). Further study of the prevalence and affect of *Wolbachia* infection in *Lycaeides* will be necessary to determine the extent to which *Wolbachia* infection has driven the non-neutral introgression of mitochondrial haplotype h01.

Selection-mediated mitochondrial introgression in *Lycaeides* (and butterflies in general) is particularly interesting and unusual, regardless of the mechanism of selection, due to two idiosyncrasies of Lepidopteran biology. In *Lycaeides* and most other Lepidoptera, females are the heterogametic sex. This means that like the mitochondrial genome, the W sex chromosome is maternally inherited. Thus, if a mitochondrial allele becomes fixed in a foreign population, such as haplotype h01 in many *Lycaeides* populations, the same should be true of the W chromosome associated with the invading mitochondrial genome. Paternal leakage of mtDNA has been reported in butterflies, and if common, would alter this expectation (Andolfatto *et al*. 2003). This general expectation holds whether selection favours the invading mitochondrial allele, the invading W chromosome allele, or some other maternally inherited genetic material. In some instances, the Lepidopteran W chromosome is degenerate and composed almost entirely of heterochromatin, however, in other cases the W chromosome contains functional genes. For example, the gene responsible for wing melanism, which produces a mimetic wing pattern, is found on the W chromosome of *Papilio glaucus* (Scriber *et al*. 1996). At present, little is known regarding the W chromosome of *Lycaeides*, however, this study raises the interesting possibility that non-neutral variation on the W chromosome may have been swept along with mitochondrial haplotype h01. Second, because recombination does not occur in female butterflies (Turner & Sheppard 1975), even a few alleles that are deleterious in the recipient genetic background should be sufficient to slow or even prevent large portions of chromosomes from introgressing. Therefore, in Lepidoptera, we expect that selection mediated mitochondrial introgression will result in the spread of mtDNA and the W chromosome, with little or no autosomal introgression, which is consistent with the data for mitochondrial introgression in *Lycaeides*. Similarly, selection-mediated introgression of the W chrosomosme will result in the spread of associated mtDNA allele(s), and is an alternative possible explanation for the results reported in this manuscript.

## Conclusions

We detected discordance in patterns of nuclear and mitochondrial variation across the North American *Lycaeides* species-complex, probably due, at least in part, to mitochondrial introgression. A reduced ratio of mitochondrial to nuclear genetic variation associated with haplotype h01 suggests that introgression and spread of mitochondrial variation among *Lycaeides* populations has likely been facilitated by natural selection. The evolutionary history of the mitochondrial genome of *Lycaeides*, while discordant with the evolutionary history of the nuclear genome, is important in its own right. By comparing the unique evolutionary histories of different genes and genomes, particularly those with different modes of inheritance and biological properties, we gain a better, more complete picture of the evolutionary processes that underlie biological diversity. Although we were unable to unambiguously identify the specific selective force(s) that may have led to the spread of mitochondrial haplotype h01 in *Lycaeides*, we have found some evidence that *Wolbachia* infection might be involved, and thus, laid the foundation for future studies of selection and mitochondrial introgression in the North American *Lycaeides* species complex, a system that has already served as a good model for studying natural hybridization (e.g. Gompert *et al*. 2006a, b).

## Acknowledgements

We would like to thank A. Shapiro for continuing discussion and help in collecting specimens. We wish to thank the following people who assisted with specimen collection: S. Fuller, M. Amaral, T. McCabe, J. Oliver, K. Prudic, E. Runquist, P. Opler, G. Austin, K. Anderson, D. Underwood, D. Ganser, G. Gelembuik, N. Anthony, G. Marrone, M. Zubrik, S. Davkov. Thanks to A. Endo and H. Zama for their assistance rearing larvae in the laboratory. Also thanks to C. Jiggins, L. Lucas, A. Buerkle, C. Weinig, and the Buerkle/Weinig laboratory group at the University of Wyoming for comments on earlier versions of this manuscript. This work was funded by a NSF graduate research fellowship to ZG, the University of Tennessee, and a Research Enhancement Grant from Texas State University to CCN.

## References


Andolfatto PM, Scriber M, Charlesworth B (2003) Lack of association between mtDNA haplotypes and a female limited mimicry locus in *Papilio glaucus*. *Evolution*, **57**, 305–316.

Avise JC (1994) *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Bensch S, Irwin DE, Irwin JH, Kvist L, Akesson S (2006) Conflicting patterns of mitochondrial and nuclear DNA diversity in *Phylloscopus* warblers. *Molecular Ecology*, **15**, 161–171.

Brock JP, Kaufman K (2003) *Butterflies of North America*. Hillstar Editions L. C., Tucson, Arizona.

Brookes MI, Graneau YA, King P, Rose OC, Thomas CD, Mallet JLB (1997) Genetic analysis of founder bottlenecks in the rare British butterfly *Plebejus argus*. *Conservation Biology*, **11**, 648–661.

Bull V, Beltran M, Jiggins CD, McMillan WO, Bermingham E, Mallet J (2006) Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biology*, **21**, 4–11.

Cann RL (2001) Genetic clues to dispersal in human populations: retracing the past from the present. *Science*, **291**, 1742–1748.

Caterino MS, Sperling FAH (1999) *Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes. *Molecular Phylogenetics and Evolution*, **11**, 122–137.

Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, **59**, 720–729.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.

Egger B, Koblmuller S, Sturmbauer C, Sefc KM (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. *BMC Evolutionary Biology*, **7**, 137.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Fordyce JA, Nice CC (2003) Variation in butterfly egg adhesion: adaptation to local host plant senescence characteristics? *Ecology Letters*, **6**, 23–27.

Fordyce JA, Nice CC, Forister ML, Shapiro AM (2002) The significance of wing pattern diversity in the Lycaenidae: mate discrimination by two recently diverged species. *Journal of Evolutionary Biology*, **15**, 871–879.

Forister ML, Fordyce JA, Nice CC, Gompert Z, Shapiro AM (2006) Egg morphology varies among populations and habitats along a suture zone in the *Lycaeides idas-melissa* species complex (Lepidoptera: Lycaenidae). *Annals of the Entomological Society of America*, **99**, 933–937.

Forister ML, Nice CC, Fordyce JA, Gompert Z, Shapiro AM (2008) Considering evolutionary processes in the use of single-locus genetic data for conservation, with examples from the Lepidoptera. *Journal of Insect Conservation*, **12**, 37–51.

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics*, **34**, 397–423.

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Systematic Biology*, **49**, 652–670.

Gompert Z, Fordyce JA, Forister ML, Shapiro AM, Nice CC (2006a) Homoploid hybrid speciation in an extreme habitat. *Science*, **314**, 1923–1925.

Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM (2006b) Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology*, **15**, 1759–1768.

Gompert Z, Fordyce JA, Forister ML, Nice CC (2008) Recent colonization and radiation of North American *Lycaeides* (*Plebejus*) inferred from mitochondrial DNA. *Molecular Phylogenetics and Evolution*. doi: 10.1016/j.ympev.2008.04.035.

Grant PR, Grant BR (2002) Unpredictable evolution in a 30-year study of Darwin's finches. *Science*, **296**, 707–711.

Grant PR, Grant BR, Markert JA, Keller LF, Petren K (2004) Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*, **58**, 1588–1599.

Grant WS, Spies IB, Canino MF (2006) Biogeographic evidence for selection on mitochondrial DNA in North Pacific walleye pollock *Theragra chalcogramma*. *Journal of Heredity*, **97**, 571–580.

Hillis DM, Mable BK, Larson A, Davis SK, Zimmer EA (1996) Nucleic acids IV: sequencing and cloning. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), pp. 321–381. Sinauer & Associates, Sunderland, Massachusetts.

Hoekstra HE (2003) Unequal transmission of mitochondrial haplotypes in natural populations of field mice with XY females (genus *Akodon*). *American Naturalist*, **161**, 29–39.

Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 1525–1534.

Jaenike J (2001) Sex chromosome meiotic drive. *Annual Review of Ecology and Systematics*, **32**, 25–49.

Jiggins FM (2003) Male-killing *Wolbachia and* mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, **164**, 5–12.

Jiggins FM, Hurst GDD, Schulenburg J, Majerus MEN (2001) Two male-killing *Wolbachia* strains coexist within a population of the butterfly *Acraea encedon*. *Heredity*, **86**, 161–166.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kim S-C, Rieseberg LH (2001) The contribution of epistasis to species differences in annual sunflowers. *Molecular Ecology*, **10**, 683–690.

Kronforst MR, Young LG, Blume LM, Gilbert LE (2006) Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution*, **60**, 1254–1268.

Linnen CR, Farrell BD (2007) Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution*, **61**, 1417–1438.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) *SAS System of Mixed Models*. SAS Institute, Cary, North Carolina.

Lucas LK, Fordyce JA, Nice CC (2008) Patterns of genitalic morphology around suture zones in North American *Lycaeides* (Lepidoptera: Lycaenidae): Implications for taxonomy and historical biogeography. *Annals of the Entomological Society of America*, **101**, 172–180.

Lynch M, Milligan BG (1994) Analysis of population genetic-structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**, 1325–1335.

McGuire JA, Linkem CW, Koo MS *et al.* (2007) Mitochondrial introgression and incomplete lineage sorting through space and time: Phylogenetics of crotaphytid lizards. *Evolution*, **61**, 2879–2897.

Mishmar D, Ruiz-Pesini E, Golik P *et al*. (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences, USA*, **100**, 171–176.

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

Nabokov V (1944) Notes on the morphology of the genus *Lycaeides* (Lycaenidae, Lepidoptera). *Psyche*, **51**, 104–138.

Nabokov V (1949) The nearctic members of the genus *Lycaeides* Hubner (Lycaenidae, Lepidoptera). *Bulletin of the Museum of Comparative Zoology*, **101**, 479–541.

Nabokov V (1952) Butterfly collecting in Wyoming (1952). *The Lepidopterists' News*, **6**, 41.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA*, **70**, 3321–3323.

Nice CC, Fordyce JA, Shapiro AM, ffrench-Constant R (2002) Lack of evidence for reproductive isolation among ecologically specialized Lycaenid butterflies. *Ecological Entomology*, **27**, 702–712.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rasgon JL, Styer LM, Scott TW (2003) *Wolbachia*-induced mortality as a mechanism to modulate pathogen transmission by vector arthropods. *Journal of Medical Entomology*, **40**, 125–132.

Rieseberg LH (1991) Homoploid reticulate evolution in *Helianthus* (Asteraceae) – evidence from ribosomal genes. *American Journal of Botany*, **78**, 1218–1237.

Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM (1996) Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science*, **272**, 741–745.

Ruedi M, Smith MF, Patton JL (1997) Phylogenetic evidence of mitochondrial DNA introgression among pocket gophers in New Mexico (family Geomyidae). *Molecular Ecology*, **6**, 453–462.

Schneider S, Kueffer JM, Roessli D, Excoffier L (2000) *Arlequin version 2.0. A Software Program for Population Genetic Data Analysis*. Genetics and Biometry Laboratory, Department of Anthropology and Ecology, University of Geneva, Geneva.

Scott JA (1986) *The Butterflies of North America: A Natural History and Field Guide*. Stanford University Press, Stanford, California.

Scriber JM, Hagen RH, Lederhouse RC (1996) Genetics of mimicry in the tiger swallowtail butterflies, *Papilio glaucus* and *P. canadensis* (Lepidoptera: Papilionidae). *Evolution*, **50**, 222–236.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Tajima F (1983) Evolutionary relationships of DNA-sequences in finite populations. *Genetics*, **105**, 437–460.

Tajima F (1993) Statistical-analysis of DNA polymorphism. *Japanese Journal of Genetics*, **68**, 567–595.

Takahata N (1995) A genetic perspective on the origin and history of humans. *Annual Review of Ecology and Systematics*, **26**, 343–372.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data III. Cladogram estimation. *Genetics*, **132**, 619–633.

Turelli M, Hoffmann AA, McKechnie WS (1992) Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila-simulans* populations. *Genetics*, **132**, 713–723.

Turner JRG, Sheppard PM (1975) Absence of crossing-over in female butterflies (*Heliconius*). *Heredity*, **34**, 265–269.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

Werren JH, Windsor DM (2000) *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1277–1285.

Wirtz P (1999) Mother species-father species: unidirectional hybridization in animals with female choice. *Animal Behaviour*, **58**, 1–12.

---

Zach Gompert is a graduate student in the program in ecology at the University of Wyoming with research interests in hybridization and the genetic and ecological basis of reproductive isolation. Matt Forister is an assistant professor at the University of Nevada, Reno. In addition to speciation and specialization in Lycaenid butterflies, his research involves the quantitative genetics of host use in herbivorous insects and the ecological impacts of anthropogenic habitats. James Fordyce is assistant professor at the University of Tennessee with research interests in ecological factors that promote population differentiation and maintain variation. Chris Nice is associate professor at Texas State University with interests in evolutionary ecology and genetics.

---

## Supporting information

Additional supporting information may be found in the online version of this article:

**Fig. S1** Range map for North American Lycaeides. The approximate ranges of *L. idas*, *L. melissa*, and *L.* sp. (alpine) are shown in dark gray, light gray, and black respectively. The range maps for *L. idas* and *L. melissa* follow Nabokov (1949), and Brock and Kaufman (2003). The range map of the hybrid species is based on field-work conducted by ZG, JAF, and CCN during the summer of 2006.

**Fig. S2** Barplot for three (a) and seven clusters (b). Each vertical bar represents a single individual and is colored in accordance with the Bayesian estimate of the proportion of that individuals genome that originated in a given cluster or population based on STRUCTURE v2.2 under the admixture model. The color scheme corresponds to that given for the AFLP data in figure 4, thus, n3.1 (yellow), n3.2 (dark blue), and n3.3 (brown) (a), and n7.1 (yellow), n7.2 (red), n7.3 (green), n7.4 (dark blue), n7.5 (light blue), n7.6 (pink), and n7.7 (brown) (b). Numbers on the x-axis are population number from Table 1.

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