EVOLUTIONARY CONSEQUENCES OF HYBRIDIZATION IN THE *LYCAEIDES* SPECIES COMPLEX

THESIS

Presented to the Graduate Council of
Texas State University–San Marcos
in Partial Fulfillment
of the Requirements

for the Degree

Master of SCIENCE

by

Zachariah Gompert

San Marcos, Texas
November 2006

ACKNOWLEDGEMENTS

I would like to thank my research advisor Chris Nice and my committee members Jim Ott, Mike Forstner, Garland Upchurch, and Dan Bolnick. I would also like to thank my collaborators on this research: James Fordyce, Matt Forister, Art Shaprio, and Lauren Lucas. Thanks to those who have helped by reviewing versions of this manuscript, including Ben Fitzpatrick, Chris Jiggins, Sergey Gavrillets, the EEB discussion group at Texas State University-San Marcos, and several anonymous reviewers. Thanks to Allana Welsh and Ditmar Hahn for assistance in the laboratory, and thanks to all those who helped with collections. This work was funded by an NSF graduate research fellowship to ZG and Texas State University-San Marcos research enhancement grants to CCN.

This manuscript was submitted on 14 November 2006.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iv

ABSTRACT vi

CHAPTER I 1

Introduction 2

Methods 7

Results 11

Discussion 13

References 18

APPENDIX A 24

Figures 25

Tables 28

CHAPTER II 30

Introduction 31

Demonstration of Hybrid Speciation 33

Reproductive Isolation 36

Conclusions 37

Methods 38

References 47

APPENDIX B 53

Figures 54

Tables 64

ABSTRACT

EVOLUTIONARY CONSEQUENCES OF HYBRIDIZATION

IN THE *LYCAEIDES* SPECIES COMPLEX

by

Zachariah Gompert

Texas State University–San Marcos

December 2006

SUPERVISING PROFESSOR: CHRIS C. NICE

Hybridization—here defined as mating between differentiated populations—has several evolutionary consequences (Arnold 1997, Mallet 2005). If hybrids have reduced fitness compared to the parental populations, then selection may lead to the evolution of increased assortative mating, and thus, decrease the rate of hybridization (Mallet 2005). This process is referred to as reinforcement (Mayr 1942). Alternatively, if the populations are not sufficiently differentiated, hybridization may lead to their fusion (Arnold 1997). In both of these examples hybridization can be viewed as a transient phenomenon. This is

not always the case, as hybrid zones may be long lasting. The maintenance of a hybrid zone is likely when 1) the hybrid zone occurs along an environmental cline, or 2) when the hybrid zone is maintained by a selection-dispersal balance (in this case the hybrid zone is referred to as a tension zone) (Arnold 1997). Hybrid zones may allow alleles to pass from one of the parental populations into the other via backcrosses between the hybrids and parentals. Finally, the hybrids may come to occupy a novel niche and become self-sustaining and reproductively isolated from their parental populations (e.g. Reiseberg 2003).

Here I examine the effects of hybridization within the North American *Lycaeides idas-L. melissa* species complex (Lepidoptera: Lycaenidae). A previous phylogeographic study based on variation in the control region of the mitochondrial genome demonstrated that populations of *Lycaeides* were relegated to at least three refugia during Pleistocene glacial maxima (Nice *et al.* 2005). Secondary contact has occurred among these refugial populations following post Pleistocene range expansion (Nice *et al.* 2005). Based on discordance between the mitochondrial and morphological (male genitalic measurements and wing patterns) characters, Nice *et al.* (2005) concluded that introgressive hybridization had occurred at two of these contact zones, one near the Great Lakes and one along the Sierra Nevada.

My first study investigated the contact zone near the Great Lakes, which involved the North American endangered Karner Blue Butterfly (*Lycaeides melissa samuelis*) and the closely related *L. m. melissa*. These butterflies can be distinguished based on differences in life history and morphology. Western populations of *L. m. samuelis* share mitochondrial haplotypes with *L. m. melissa* populations, while eastern populations of *L.*

m. samuelis have divergent haplotypes (Nice *et al.* 2005). I tested two hypotheses concerning the presence of *L. m. melissa* mitochondrial haplotypes in western *L. m. samuelis* populations: (i) mitochondrial introgression has occurred from *L. m. melissa* populations into western *L. m. samuelis* populations, or (ii) western populations of the nominal *L. m. samuelis* are more closely related to *L. m. melissa* than to eastern *L. m. samuelis* populations yet are phenotypically similar to the latter. A Bayesian algorithm was used to cluster 190 *L. melissa* individuals based on 143 AFLP loci. This method clearly differentiated *L. m. samuelis* and *L. m. melissa*. Thus, genomic divergence was greater between western *L. m. samuelis* populations and *L. m. melissa* populations than it was between western and eastern populations of *L. m. samuelis*. These findings support the hypothesis that the presence of *L. m. melissa* mitochondrial haplotypes in western *L. m. samuelis* populations is the result of mitochondrial introgression. These data provide valuable information for conservation and management plans for the endangered *L. m. samuelis*, and illustrate the risks of using data from a single locus for diagnosing significant units of biodiversity for conservation.

The second study I conducted involved the contact zone between *L. idas* and *L. melissa* along the Sierra Nevada. Here I investigated the possibility that alpine adapted *Lycaeides* populations in the Sierra Nevada arose via hybrid speciation, following hybridization between *L. idas* and *L. melissa*. Theory predicts that homoploid hybrid speciation is facilitated by adaptation to a novel or extreme habitat; heretofore this has not been documented in animals (Buerkle *et al.* 2000). Using molecular data and data from ecological experiments, I demonstrated that the alpine-adapted butterflies in the genus *Lycaeides* are the product of hybrid speciation. I showed that the alpine

populations possess a mosaic genome derived from both *L. melissa* and *L. idas*. These alpine populations are differentiated from, and have a more recent origin than, their putative parental species. Adaptive traits allow persistence in the environmentally extreme alpine habitat and reproductively isolate these populations from their parental species, as predicted by theory. These studies demonstrate that hybridization has several important evolutionary consequences, some of which may be far from transient. Indeed hybridization may represent a mechanism driving biological diversification.

The first chapter presented has been published in *Molecular Ecology*, and thus, is formatted for that journal. The second chapter has been accepted at *Science*. It is formatted for that journal with some minor modifications.

References

Arnold ML (1997) *Natural Hybridization and Evolution* Oxford University Press, Oxford.

Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. *Heredity* **84**, 441-451.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**, 229-237.

Mayr E (1942) *Systematics and the origin of species* Columbia University Press, New York.

Nice CC, Anthony N, Gelembiuk G, Raterman D, ffrench-Constant R (2005) The history and geography of diversification within the butterfly genus *Lycaeides* in North America. *Molecular Ecology* **14**, 1741-1754.

CHAPTER I

IDENTIFYING UNITS FOR CONSERVATION USING MOLECULAR SYSTEMATICS: THE CAUTIONARY TALE OF THE KARNER BLUE BUTTERFLY

Has been published as: Gompert, Z., C. C. Nice, J. A. Fordyce, M. L. Forister, A. M. Shapiro. 2006. Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology* **15**, 1759-1768.

Introduction

The geographic distribution of genetic variation within and among taxa provides information on historical and contemporary demographic and evolutionary processes (Avise 1994, 2000; Knowles 2000, 2001). This information can also inform conservation efforts, both in terms of identifying units for conservation and designing management plans (Moritz 1994; Meffe & Carroll 1997; Primack 2004). The quest to identify appropriate biological units for conservation has a long an arduous history (Crandall *et al.* 2000). At present, a consensus has not been reached on how to best delineate units for conservation (Crandall *et al.* 2000; Moritz 2002). Defining units for conservation based on any single character, whether mitochondrial sequence data (e.g. Hebert *et al.* 2003) or morphology, may be problematic. Multiple characters must be examined and the processes that influence those characters must be understood to accurately delineate units for conservation. Here we examine patterns of genetic variation based on both mitochondrial DNA (mtDNA) and Amplified Fragment Length Polymorphism (AFLP) (Vos *et al.* 1995) markers to test alternative hypotheses concerning the history and current status of the North American endangered Karner Blue Butterfly (*Lycaeides melissa samuelis*).

Lycaeides melissa samuelis and its close relative, the Melissa Blue Butterfly (*L. m. melissa*), are small blue butterflies belonging to the family Lycaenidae. *Lycaeides melissa samuelis* has experienced a 99 percent range-wide decline in population size over the past century most of which occurred in the last 25 years (US Fish and Wildlife

Service 1992). This decline led to the listing of *L. m. samuelis* as a federally endangered species in the United States in 1992 (US Fish and Wildlife Service 1992, 2001). Remnant populations of *L. m. samuelis* are restricted to oak savannahs, pine prairies, and lakeshore sand dunes in Minnesota, Wisconsin, Indiana, Michigan, New York, and New Hampshire (Scott 1986; US Fish and Wildlife Service 1992, 2001). The closely related *Lycaeides m. melissa* is not considered endangered or threatened and is found in dry prairies and alfalfa fields over a large expanse of western North America from Minnesota to California (Lane & Weller 1994; Brock & Kaufman 2003). Both *L. m. samuelis* and *L. m. melissa* use papilionaceous legumes (Fabaceae) as larval host plants (Scott 1986; Brock & Kaufman 2003). However, *L. m. samuelis* uses only wild lupine (*Lupinus perennis*), while *L. m. melissa* uses a number of legume genera including *Astragalus, Medicago, Glycorrhiza,* and *Lupinus*—but not *L. perennis* (Scott 1986; US Fish and Wildlife Service 1992; Lane & Weller 1994). *Lycaeides melissa samuelis* populations are bivoltine, while *L. m. melissa* populations are variable but generally have more than two generations per year (US Fish and Wildlife Service 1992; Nice & Shapiro 1999). These two butterflies also differ morphologically both in wing patterns (Nabakov 1949; Opler & Kizek 1984; Lane & Weller 1994) and in the size and shape of the male genitalia (Nabakov 1949; Lane & Weller 1994; C. C. Nice unpublished data).

Nice *et al.* (2005) examined the geographic distribution of genetic variation for the AT-rich region of the mitochondrial genome in North American *Lycaeides*. Western populations of *L. m. samuelis* (i.e. populations in the state of Wisconsin) shared haplotypes with *L. m. melissa* populations in western North America; in fact, there were

no haplotypes in the Wisconsin *L. m. samuelis* populations that were not shared with *L. m. melissa* populations. In contrast, eastern *L. m. samuelis* populations (i.e. populations east of Lake Michigan) contained different haplotypes not found in any other *Lycaeides* populations (Nice *et al.* 2005). Thus there is discord between the traditional taxonomic boundary between *L. m. samuelis* and *L. m. melissa* based on morphological and ecological characters (Nabakov 1949; Lane & Weller 1994) and the patterns of genetic variation observed for mitochondrial DNA (mtDNA). Packer *et al.* (1998) surveyed allozyme variation in one *L. m. melissa* population from Minnesota and two *L. m. samuelis* populations, one from Wisconsin and one from New York. They found low levels of genetic divergence and concluded that *L. m. samuelis* and *L. m. melissa* were not clearly differentiated (Packer *et al.* 1998).

Phylogeographic evidence suggests that *L. m. melissa* and *L. m. samuelis* populations were confined to different glacial refugia during the Pleistocene, and that they may have experienced secondary contact following post-Pleistocene range expansion (Nice *et al.* 2005). A similar phylogeographic boundary has been observed in other organisms and is attributed to Pleistocene refugia southeast and southwest of Lake Michigan (Austin *et al.* 2002). Secondary contact may have facilitated gene exchange between *L. m. samuelis* and *L. m. melissa* in which case Lake Michigan may have served as a geographical barrier against mitochondrial introgression into the eastern *L. m. samuelis* populations. Alternatively, populations in Wisconsin that are nominally *L. m. samuelis* may be more closely related to *L. m. melissa* populations than to *L. m. samuelis* populations east of Lake Michigan. This may be because *L. m. samuelis* is paraphyletic or

the ecological and morphological similarity of western *L. m. samuelis* populations to eastern *L. m. samuelis* populations may be the result of convergent evolution under similar selective pressures. Multiple studies have suggested that lineages of lycaenids diversify rapidly and respond to natural selection acting on ecological traits (Nice & Shapiro 1999; Nice *et al.* 2002; Fordyce & Nice 2003). Host-associated selection, in particular, could be expected to produce convergent patterns in populations that do not share an immediate common ancestor (Shapiro 1991; Nice & Shapiro 2001). For example, molecular data and ecological studies suggest that host plant use has driven convergent evolution of adult phenology and wing patterns in populations of the nominal butterfly species *Mitoura muiri* in the Coast Range and Sierra Nevada of California (Nice & Shapiro 2001; Forister 2004).

The two scenarios presented above have different implications for the management and conservation of *L. m. samuelis*. If Wisconsin *L. m. samuelis* populations possess *L. m. melissa* mitochondrial haplotypes as the result of mitochondrial introgression, then all *L. m. samuelis* populations can continue to be managed as a single unit. However, if *L. m. samuelis* populations on opposite sides of Lake Michigan are not each other's closest relatives, then it may be necessary to manage *L. m. samuelis* populations east and west of Lake Michigan as separate units. In particular, if the latter scenario is correct, it is important that translocations do not cross Lake Michigan. This concern is pertinent, as translocations have been proposed for reintroduction of *L. m. samuelis* individuals into areas where populations no longer exist and for supplementing current populations (US Fish and Wildlife Service 2001).

The two hypothesized scenarios of the biogeographic history of *L. melissa* in North America can be distinguished by examining the overall pattern of relatedness among *L. melissa* populations based on the nuclear genome. Two clear predictions exist. If Wisconsin *L. m. samuelis* populations have *L. m. melissa* mitochondrial haplotypes as a result of mitochondrial introgression, the nuclear genome of Wisconsin *L. m. samuelis* individuals should be more similar to the nuclear genome of *L. m. samuelis* individuals east of Lake Michigan than to the nuclear genome of *L. m. melissa* individuals (e.g. Funk & Omland 2003). Conversely, if the Wisconsin *L. m. samuelis* populations are more closely related to *L. m. melissa* populations than to *L. m. samuelis* populations east of Lake Michigan, the nuclear genome of Wisconsin *L. m. samuelis* individuals should be more similar to the nuclear genome of *L. m. melissa* individuals. In this case, patterns of variation observed in mtDNA and nuclear markers would both conflict with the current taxonomy.

In order to accurately assess overall genomic divergence, a large number of presumed neutral nuclear markers are needed, as individual gene genealogies are subject to stochastic events and take time to reflect the true population or species phylogeny (Funk & Omland 2003; Machado & Hey 2003). The AFLP technique (Vos *et al.* 1995) is an ideal choice for such an undertaking for a number of reasons. This technique can generate a large number of nuclear markers (>100) in a short amount of time with only a modest start up cost (Bensch & Akesson 2005). This technique is especially useful in non-model organisms, as no prior knowledge of the genome is required (Bensch &

Akesson 2005). AFLP markers have been used successfully to detect genetic structure (e.g. Reineke *et al.* 1999; Mock *et al.* 2004; Wang *et al.* 2004; Irwin *et al.* 2005), and identify cases of introgression (e.g. Sullivan *et al.* 2004) in wild populations.

Here we use data from mtDNA sequences and AFLP markers to test two alternative hypotheses regarding the biogeographic history of the federally endangered *L. m. samuelis*: (i) mitochondrial introgression from *L. m. melissa* populations into Wisconsin *L. m. samuelis* populations has led to the presence of *L. m. melissa* mitochondrial haplotypes in the Wisconsin populations of *L. m. samuelis*, or (ii) Wisconsin *L. m. samuelis* populations are more closely related to *L. m. melissa* populations than to *L. m. samuelis* populations east of Lake Michigan.

Methods

Sample collection

Butterflies were collected from five *L. m. samuelis* populations and three *L. m. melissa* populations (Figure 1B, Table 1). Both males and females were collected from *L. m. melissa* populations, while only males were collected from *L. m. samuelis* populations (with the exception of two females collected at Saratoga Springs, NY) in accord with USFWS permit PRT842392. DNA was isolated following the methods of Hillis *et al.* (1996) and Brookes *et al.* (1997).

Mitochondrial DNA

We sequenced portions of the mitochondrial genes cytochrome oxidase *c* subunit 1 (COI) and cytochrome oxidase *c* subunit II (COII) for 5 individuals from each of the 8 populations. 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 and 550 base pairs for COI and COII respectively. Fluorescently labeled dideoxy terminators were used for single stranded sequencing reactions for both COI and COII according to Applied Biosystems Inc. specifications. Labeled amplicons were separated and visualized on 6% denaturing polyacrylamide gel using an automated DNA sequencer (Applied Biosystems model 377). Sequences were aligned using the program Sequencher 4.2.2. or by eye. A partition homogeneity test was performed using PAUP* v.4.0b10 (Swofford 2003) to determine if the COI and COII sequence data sets possessed conflicting phylogenetic signals. A maximum parsimony haplotype network was constructed for the combined dataset using TCS 1.2.1 (Clement *et al.* 2000), which employs the statistical algorithms of Templeton *et al.* (1992).

Analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) was employed to determine the proportion of the total genetic variation for COI and COII that was distributed according to the taxonomic boundary between *L. m. samuelis* and *L. m. melissa*. Spatial analysis of molecular variance (SAMOVA) (Dupanloup *et al.* 2002) was then used to identify the two geographically continuous groups of populations that maximized ϕ_{CT}. We compared ϕ_{CT} from AMOVA performed with populations grouped

according to subspecies to ϕ_{CT} based on the regional groups identified by SAMOVA in order to quantify the degree to which the current taxonomic boundary between *L. m. samuelis* and *L. m. melissa* is incongruent with the pattern of genetic structure observed in the mtDNA data. For AMOVA and SAMOVA, Tamura and Nei (1993) genetic distances were used; these distances were selected as a result of the DNA sequence evolution model selection procedure implemented in Modeltest 3.7 (Posada & Crandall 1998).

Amplified Fragment Length Polymorphism markers

In order to estimate overall genomic divergence and diversity within and between *L. m. samuelis* and *L. m. melissa*, AFLP marker profiles were produced for 19-28 individuals from each of the eight populations (190 individuals in total) following a modified version of the procedures introduced in Vos *et al.* (1995). AFLP markers 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 gel, using an ABI Prism 377 DNA Sequencer (Applied Biosystems Inc.). GeneScan was used to visualize AFLP bands, which were sized by comparison to a size standard ladder (ROX standard, Applied Biosystems Inc.) added to each lane. Bands with low peak heights (less than 150 relative fluorescent units) were not scored. Bands that were present in less than 5% of the individuals surveyed were not included for subsequent analysis. Because almost all *L. m. samuelis* individuals collected were male, a single AFLP marker that appeared to be sex-

linked was also excluded from all further analyses. AFLP banding patterns were highly reproducible. Twenty arbitrarily chosen individuals underwent a second amplification. For the twenty individuals 95.5% of scored bands were detected in both amplifications.

The program STRUCTURE (Pritchard *et al.* 2000) was used to cluster individuals based on their AFLP banding profiles. STRUCTURE employs a model-based Bayesian clustering algorithm to assign individuals probabilistically to clusters to minimize deviations from linkage equilibrium. The admixture model was run for 500,000 generations with an initial burnin of 50,000 generations. Prior information regarding the population or taxon from which an individual was sampled was ignored. STRUCTURE was also used to estimate the number of clusters (k) that best explained the data. The method of Evanno *et al.* (2005) was used to infer k. This procedure identifies the appropriate number of clusters using the *ad hoc* statistic Δk, which is based on the second order rate of change in the log probability of the data between successive values of k. Evanno *et al.* (2005) demonstrated that this method is able to detect the appropriate number of clusters for simulated data sets under a number of gene exchange models. It is not possible to evaluate Δk for k=1 (Evanno *et al.* 2005). We explored the probability of the data for 2-9 clusters. Ten simulations were run for each k, multiple runs of the same k produced highly consistent individual assignment probabilities. Multiple runs for each k were used to compute the variance in STRUCTURE estimates of the log probability of data for each k; these variance estimates were used in the calculation of Δk as described by Evanno *et al.* (2005).

Results

Mitochondrial DNA

Sequence was obtained for 410-bp of COI and 510-bp of COII for all 40 individuals examined. Conflicting phylogenetic signal between these gene regions was not detected using a partition homogeneity test (p=1.000), thus, COI and COII sequences were combined for all analyses. Three unique haplotypes were detected for the combined sequence data (Table 1). A single most parsimonious haplotype network was produced (Figure 1A). Haplotypes A and B differed by a single base, while these haplotypes differed from haplotype C by six or seven bases, respectively. The Sierraville, CA and Spring Creek, SD *L. m. melissa* populations were fixed for haplotype A (Figure 1B). A single individual from the Brandon, SD *L. m. melissa* population had haplotype B, while the other four individuals had haplotype A (Figure 1B). All three Wisconsin *L. m. samuelis* populations (Fish Lake, Fort McCoy, and Necedah) were also fixed for haplotype A; however, the Indiana Dunes, IN, and Saratoga Springs, NY *L. m. samuelis* populations were fixed for haplotype C (Figure 1B). Sequence divergence between *L. melissa* and Wisconsin *L. m. samuelis* populations (haplotypes A and B) and *L. m. samuelis* populations east of Lake Michigan (haplotype C) was 0.65-0.76%. Based on data from COI and COII, the Wisconsin *L. m. samuelis* populations are indistinguishable from the *L. m. melissa* populations.

AMOVA partitioned approximately 12% of the total genetic variation for COI and COII between subspecies (ϕ_{CT}=11.64, p<0.001, Table 2A). SAMOVA was able to partition approximately 99% of the total genetic variation for COI and COII between the following two regional groups: (i) all three *L. m. melissa* populations and the Wisconsin *L. m. samuelis* populations, (ii) the *L. m. samuelis* east of Lake Michigan (ϕ_{CT}=99.16, p<0.001, Table 2B). The groups identified by SAMOVA explained an additional 87% of the total genetic variation beyond that explained by groups based on subspecies identification.

Amplified Fragment Length Polymorphism markers

The three primer pairs generated a total of 143 AFLP bands ranging in size from 71 to 481-bp, all of which were polymorphic among all individuals. All three primer pairs produced similar numbers of AFLP bands. A total of 130 (90.91%) bands were polymorphic within *L. m. melissa* and a total of 124 (86.71%) bands were polymorphic within *L. m. samuelis*. Twenty AFLP bands were present exclusively in *L. m. melissa* populations and 11 AFLP bands were found only in *L. m. samuelis* populations.

Two clusters best explained the AFLP data (Figure 2). *Lycaeides melissa melissa* individuals were assigned with high probability to one cluster (cluster 1), and no *L. m. melissa* individuals had an assignment probability to cluster 1 less than 0.645 (Figure 1B; Figure 3). Nearly all *L. m. samuelis* individuals were assigned with high probability to another cluster (cluster 2), and no *L. m. samuelis* individuals had an assignment

probability to cluster 2 less than 0.455 (Figure 1B; Figure 3). Under the admixture model an individual's assignment probability to each cluster can be interpreted as the proportion of that individual's genome that originated in each cluster. The mean assignment probability of *L. m. melissa* populations to cluster 1 ranged from 0.935 at Brandon, SD to 0.992 at Sierraville, CA (Table 1). The mean assignment probability of *L. m. samuelis* populations to cluster 1, which equals one minus their mean assignment probability to cluster 2, ranged from 0.005 at Indiana Dunes, IN to 0.121 at Fish Lake, WI (Table 1). The lowest assignment probability to cluster 1 for a *L. m. melissa* population (Brandon, SD) and the highest assignment probability to cluster 1 for a *L. m. samuelis* population (Fish Lake, WI) occurred nearest the boundary between these taxa. However, even at these locations AFLP markers clearly distinguish between *L. m. melissa* and *L. m. samuelis* individuals (Figure 3). AFLP data, which provides a metric of genomic divergence, support the nominal taxonomic boundary between these taxa, which was based on ecological and morphological data (Nabakov 1949; Lane & Weller 1994).

Discussion

Phylogeographic history of L. m. samuelis

Mitochondrial DNA (COI and COII) and AFLP makers identified different boundaries between *L. m. samuelis* and *L. m. melissa.* All three *L. m. melissa* populations and the Wisconsin *L. m. samuelis* populations were fixed, or nearly fixed (as in Brandon, SD), for the same mitochondrial haplotype (haplotype A), while *L. m. samuelis*

populations east of Lake Michigan were fixed for a different divergent haplotype (haplotype C). Thus, COI and COII mitochondrial DNA data partitions these populations into two groups: (i) *L. m. melissa* and Wisconsin *L. m. samuelis*, and (ii) *L. m. samuelis* east of Lake Michigan, which are separated by 0.65-0.76% sequence divergence. The geographic pattern of genetic variation for COI and COII is very similar to the pattern identified by Nice *et al.* (2005) based on the AT-rich region of the mitochondrial genome. There is an apparent phylogeographical boundary between mitochondrial clades at or near Lake Michigan.

Unlike the mitochondrial data, AFLP data provided no evidence for a genetic boundary near Lake Michigan. Bayesian clustering of individuals based on AFLP marker profiles partitioned individuals into two clusters: one that consisted of *L. m. samuelis* individuals and one that consisted of *L. m. melissa* individuals. This pattern is in accord with patterns of variation in habitat and host plant use (Scott 1986; US Fish and Wildlife Service 1992; Lane & Weller 1994; Brock & Kaufman 2003), phenology (US Fish and Wildlife Service 1992; Nice & Shapiro 1999), wing morphology (Nabakov 1949; Lane & Weller 1994), male genitalic morphology (Nabakov 1949; Lane & Weller 1994), and allozyme data (Packer *et al.* 1998), and thus, corresponds to the pattern expected based on taxonomic designations.

The incongruent patterns of genetic variation observed in mtDNA and nuclear AFLP markers support the hypothesis that the presence of mitochondrial haplotypes in the Wisconsin *L. m. samuelis* populations that are identical to haplotypes found in *L. m.*

melissa populations is the result of mitochondrial introgression from *L. m. melissa* populations into the Wisconsin *L. m. samuelis* populations (Figure 1B). This mitochondrial introgression appears to have progressed as far as Lake Michigan. However, we cannot rule out the possibility of ancestral polymorphism. For example, the *L. m. melissa* lineage may have became fixed for one mitochondrial variant while *L. m. samuelis* continued to be polymorphic until selective sweeps or genetic drift fixed different mitochondrial haplotypes in the eastern and western portions of their range. This scenario implies that there has been insufficient time for significant sequence divergence to accumulate between *L. m. melissa* and western *L. m. samuelis*. These possibilities seem unlikely given homogeneity in terms of habitat, host plant use, morphology, and the AFLP data presented here over the entire range of *L. m. samuelis*.

Despite extensive mitochondrial introgression from *L. m. melissa* into the Wisconsin *L. m. samuelis* populations, there has been little nuclear introgression. This lack of nuclear introgression is evidenced by the fact that there are only six individuals with moderate assignment probabilities to both cluster 1 and cluster 2, most of which are from Fish Lake, WI (Figure 3). Many more individuals would be expected to have moderate assignment probabilities to both clusters if nuclear introgression were prevalent. There are two likely explanations for the lack of nuclear introgression in combination with widespread mitochondrial introgression. First, natural selection against *L. m. melissa* x *L. m. samuelis* hybrids and backcrosses may be sufficiently strong to limit nuclear introgression, while still allowing for neutral mitochondrial alleles to pass almost freely from *L. m. melissa* populations to the Wisconsin *L. m. samuelis* populations. It is

not uncommon to see unidirectional introgression in such cases (Chan & Levin 2005). This would provide evidence that at least some of the morphological and/or ecological differences between *L. m. melissa* and *L. m. samuelis* are important reproductive isolating barriers involved in maintaining the boundary between these taxa. Dissimilarity in wing pattern and/or male genitalic structure between *L. m. melissa* and *L. m. samuelis* may preclude hybrid and backcross individuals from mating. There is evidence that wing pattern is important for mate recognition and preference for other *Lycaeides* populations (Fordyce *et al.* 2002). Such prezygotic barriers are especially permeable to introgression of maternally inherited genes (Chan & Levin 2005). Additionally, differences in habitat and host-plant use between *L. m. melissa* and *L. m. samuelis* may reduce the fitness of individuals of mixed descent in either of the parental habitats. A second explanation for the lack of nuclear introgression between *L. m. samuelis* and *L. m. melissa* populations despite substantial mitochondrial introgression is a mitochondrial selective sweep. Because animal mitochondrial genomes usually do not undergo recombination (but see Eyre-Walker *et al.* 1999), a selective advantage for the *L. m. melissa* mitochondrial genome at a single locus may have been sufficient to drive a selective sweep of the entire mitochondrial genome. Such non-neutral variation in mitochondrial alleles has been postulated to explain other phylogeographic patterns (Levin 2000; Brumfield *et al.* 2001). At present we are unable to discriminate between these two possibilities. It would be possible to detect a mitochondrial selective sweep by comparing effective population size estimated from a number of nuclear gene sequences to an estimate based on mtDNA for the Wisconsin *L. m. samuelis* populations. A significantly lower effective population size estimate for mtDNA than for nuclear DNA would be indicative of a selective sweep

(Galtier *et al.* 2000). However, at present nuclear sequence data from several loci is not available for *L. m. samuelis*.

Conservation implications

We conclude, based on our data and the available morphological and ecological data that *L. m. samuelis* is a unique entity, distinct from *L. m. melissa*. This study finds no evidence for separate origins of the *L. m. samuelis* populations on different sides of Lake Michigan. As a result, our data do not suggest the need to treat populations east and west of Lake Michigan as separate units for conservation and management purposes. This does not mean that we can say for certain that translocations between different *L. m. samuelis* populations could take place without negative consequences, as population level local adaptation may still be present within *L. m. samuelis*, which could lead to reduced fitness of interpopulation hybrids and potentially lower the mean fitness of the recipient population of the translocation (i.e. outbreeding depression). Further investigation is needed prior to undertaking interpopulation translocations. However, it is clear in this case that mtDNA data does not accurately reflect the evolutionary relationships of this group.

The findings of this study highlight a potential problem regarding the recent trend to rely primarily on DNA sequence data, especially from the mitochondrial genome, to identify units of biodiversity (e.g. Moritz 1994, Hebert *et al.* 2003, 2004, Tautz *et al.* 2003). Data from a single locus such as mtDNA should be used with caution. In this case,

mtDNA incorrectly identifies the Wisconsin populations of the federally endangered species *L. m. samuelis* as populations of the widespread *L. m. melissa*. This is a case in which DNA taxonomy would fail to identify the appropriate units of biodiversity for conservation purposes. Such techniques would not support the conservation status of the Wisconsin *L. m. samuelis* populations, which is clearly warranted based on the strong correlations between patterns of genomic divergence, morphological characters and ecological data.

References

Austin JD, Lougheed SC, Neidrauer L, Chek AA, Boag PT (2002) Cryptic lineages in a small frog: the post—glacial history of the spring peeper, *Pseudacris crucifer* (Anura : Hylidae). *Molecular Phylogenetics and Evolution* **25**, 316—329.

Avise JC (1994) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer Associates, Sunderland MA.

Avise JC (2000) *Phylogeography: The History and Formation of Species* Harvard University Press, Cambridge MA.

Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology* **14**, 2899—2914.

Brock JP, Kaufman K (2003) *Butterflies of North America* Houghton Mifflin, Boston MA.

Brookes MI, Graneau YA, King P, *et al.* (1997) Genetic analysis of founder bottlenecks in the rare British butterfly *Plebejus argus*. *Conservation Biology* **11**, 648—661.

Brumfield RT, Jernigan RW, McDonald DB, Braun MJ (2001) Evolutionary implications of divergent clines in an avian (Manacus : Aves) hybrid zone. *Evolution* **55**, 2070—2087.

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—1659.

Crandall KA, Bininda—Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution* **15**, 290—295.

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

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611—2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes-application to human mitochondrial-DNA restriction data. *Genetics* **131**, 479—491.

Eyre—Walker A, Smith NH, Smith JM (1999) How clonal are human mitochondria? *Proceedings of the Royal Society of London Series B—Biological Sciences* **266**, 477—483.

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 (2004) Oviposition preference and larval performance within a diverging lineage of lycaenid butterflies. *Ecological Entomology* **29**, 264—272.

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.

Galtier N, Depaulis F, Barton NH (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**, 981—987.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B—Biological Sciences* **270**, 313—321.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14812—14817.

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). Sinauer Associates, Sunderland, MA.

Irwin DE, Bensch S, Irwin JH, Price TD (2005) Speciation by distance in a ring species. *Science* **307**, 414—416.

Knowles LL (2000) Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution* **54**, 1337—1348.

Knowles LL (2001) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology* **10**, 691—701.

Lane CP, Weller SJ (1994) A review of *Lycaeides* Hubner and Karner Blue butterfly taxonomy. *Minn. Agric. Exp. Stn. Misc. Publ. Ser.* **84**, 5—21.

Levin DA (2000) *The Origin, Expansion, and Demise of Plant Species* Oxford University Press, New York NY.

Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proceedings of the Royal Society of London Series B—Biological Sciences* **270**, 1193—1202.

Meffe GK, Carroll CR (1997) *Principles of Conservation Biology*, 2nd edn. Sinauer Associates, Sunderland MA.

Mock KE, Brim—Box JC, Miller MP, Downing ME, Hoeh WR (2004) Genetic diversity and divergence among freshwater mussel (*Anodonta*) populations in the Bonneville Basin of Utah. *Molecular Ecology* **13**, 1085—1098.

Moritz C (1994) Defining evolutionary significant units for conservation. *Trends in Ecology and Evolution* **9**, 373—376.

Moritz C (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology* **51**, 238—254.

Nabakov V (1949) The neartic members of *Lyaceides* Hubner (Lycaenidae, Lepidotera). *Bulletin of the Museum of Comparative Zoology* **101**, 479—540.

Nice CC, Anthony N, Gelembiuk G, Raterman D, Ffrench—Constant R (2005) The history and geography of diversification within the butterfly genus *Lycaeides* in North America. *Molecular Ecology* **14**, 1741—1754.

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.

Nice CC, Shapiro AM (1999) Molecular and morphological divergence in the butterfly genus Lycaeides (Lepidoptera : Lycaenidae) in North America: evidence of recent speciation. *Journal of Evolutionary Biology* **12**, 936—950.

Nice CC, Shapiro AM (2001) Population genetic evidence of restricted gene flow between host races in the butterfly genus *Mitoura* (Lepidoptera : Lycaenidae). *Annals of the Entomological Society of America* **94**, 257—267.

Opler P, Krizek G (1984) *Butterflies East of the Great Plains: An Illustrated Natural History* John Hopkins University Press, Baltimore MD.

Packer L, Taylor JS, Savignano DA, *et al.* (1998) Population biology of an endangered butterfly, *Lycaeides melissa samuelis* (Lepidoptera; Lycaenidae): genetic variation, gene flow, and taxonomic status. *Canadian Journal of Zoology—Revue Canadienne De Zoologie* **76**, 320—329.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817—818.

Primack RB (2004) *A Primer of Conservation Biology*, 3rd edn. Sinauer Associates, Sunderland MA.

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

Reineke A, Karlovsky P, Zebitz CPW (1999) Amplified fragment length polymorphism analysis of different geographic populations of the gypsy moth, *Lymantria dispar* (Lepidoptera : Lymantriidae). *Bulletin of Entomological Research* **89**, 79—88.

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

Simon C, Frati F, Beckenbach A, *et al.* (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.

Shapiro AM (1991) Convergent evolution in Western North American and Patagonian skippers (Hesperiidae). *Journal of Research on the Lepidoptera* **30**, 162—174.

Sullivan JP, Lavoue S, Arnegard ME, Hopkins CD (2004) AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea : Teleostei). *Evolution* **58**, 825—841.

Swofford DL (2003) PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), Sunderland MA.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial—DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512—526.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology & Evolution* **18**, 70—74.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic—analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA—sequence data .3. cladogram estimation. *Genetics* **132**, 619—633.

US Fish and Wildlife Service (1992) Endangered and threatened wildlife and plants: proposed endangered status for the Karner Blue butterfly, pp. 2241—2247.

US Fish and Wildlife Service (2001) Karner Blue butterfly (*Lycaeides melissa samuelis*) recovery plan, draft October 2001.

Vos P, Hogers R, Bleeker M, *et al.* (1995) AFLP — a new technique for DNA—fingerprinting. *Nucleic Acids Research* **23**, 4407—4414.

Wang ZS, Baker AJ, Hill GE, Edwards SV (2003) Reconciling actual and inferred population histories in the house finch (*Carpodacus mexicanus*) by AFLP analysis. *Evolution* **57**, 2852—2864.

APPENDIX A

CHAPTER 1 FIGURES AND TABLES

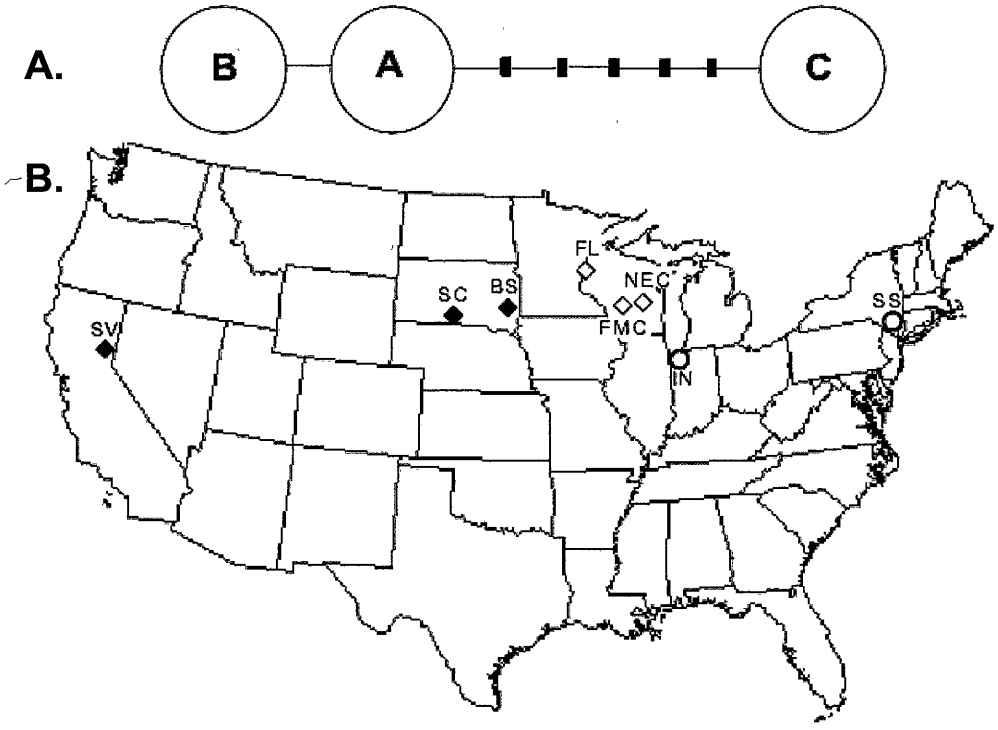

Figure 1. Mitochondrial DNA haplotype network and population map. (A) Mitochondrial DNA haplotype network showing the single most parsimonious haplotype network for the three haplotypes identified. Each circle represents a haplotype. Black squares represent missing haplotypes. Haplotype C is separated from haplotypes A and B by six and seven mutations, respectively. (B) Population map. Diamonds represent populations either fixed for mtDNA haplotype A or with both haplotypes A and B (which differ by a single base pair), circles represent populations fixed for haplotype C. Empty shapes represent populations with a high probability of assignment to cluster 1 based on AFLP data, filled shapes represent populations with a high probability of assignment to cluster 2 (Figure 3). The pattern of molecular variation is discordant between mtDNA data and AFLP markers.

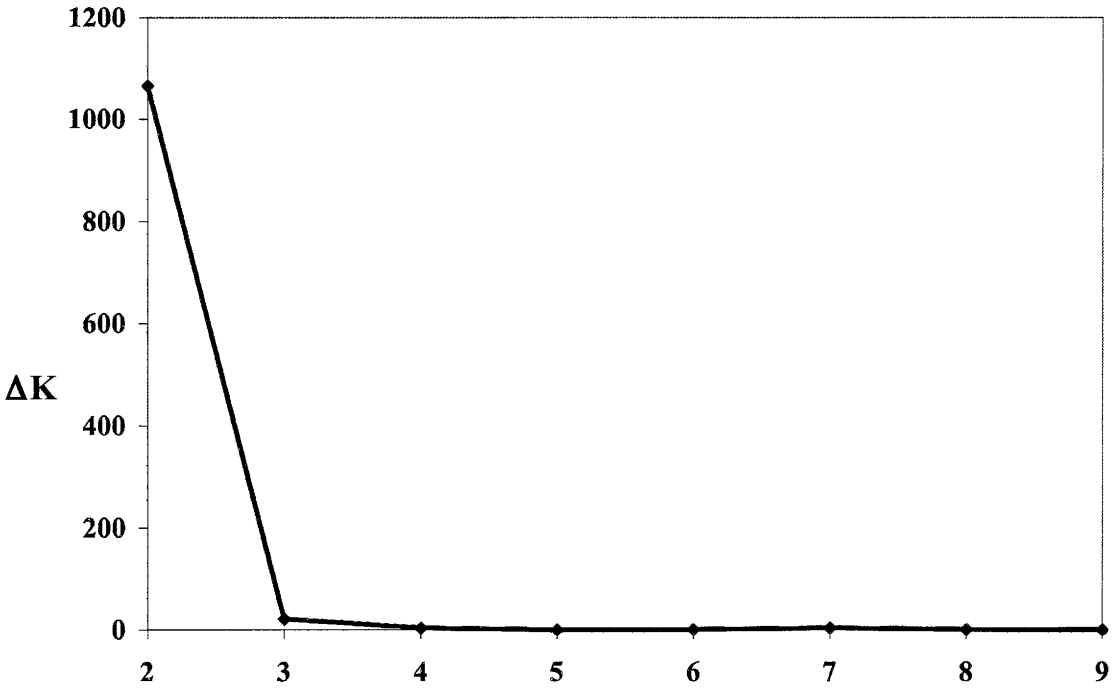

Figure 2. The number of clusters (k) versus the second order rate of change in k (Δk). The clear maximum for Δk at k=2 indicates that two clusters best explain the AFLP data for the sampled *Lycaeides melissa* populations.

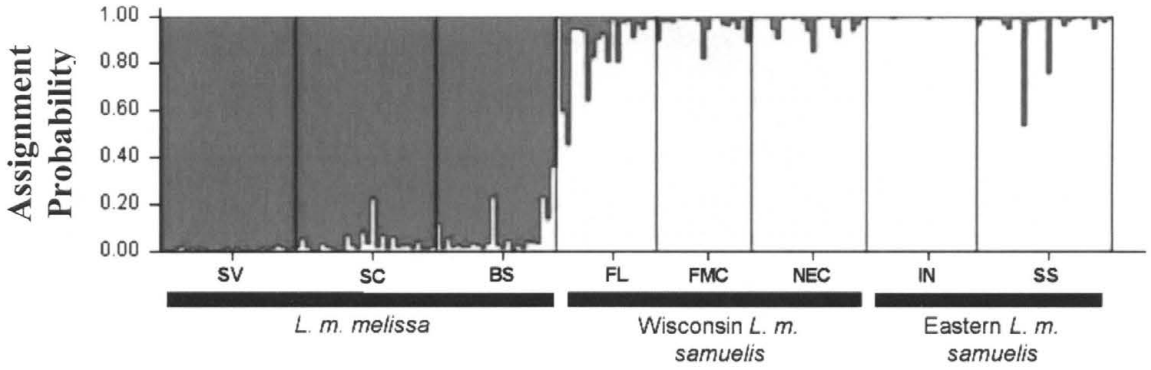

Figure 3. Bayesian assignment probabilities for k=2. Each vertical bar corresponds to one individual. The proportion of each bar that is dark gray represents an individual's assignment probability to cluster 1, the proportion of each bar that is light gray represents an individual's assignment probability to cluster 2. Most *L. m. melissa* have high assignment probabilities to cluster 1, while most *L. m. samuelis* (including the individuals from Wisconsin) have high assignment probabilities to cluster 2.

Table 1: Population data. Populations are labeled with their nominal taxonomic designations. The term p(cluster 1) refers to the mean assignment probability of the individuals from each population to cluster 1 based on AFLP loci (see text, Figure 3). N refers to the sample size for AFLP data.

ID	Population	Taxon	mtDNA Haplotype (no. of individuals)	p(cluster 1)	N
SV	Sierraville, CA	*L. m. melissa*	A(5)	0.992	27
SC	Spring Creek, SD	*L. m. melissa*	A(5)	0.968	28
BS	Brandon, SD	*L. m. melissa*	A(4), B(1)	0.935	24
FL	Fish Lake, WI	*L. m. samuelis*	A(5)	0.121	20
FMC	Fort McCoy, WI	*L. m. samuelis*	A(5)	0.040	19
NEC	Necedah, WI	*L. m. samuelis*	C(5)	0.032	23
IN	Indiana Dunes, IN	*L. m. samuelis*	C(5)	0.005	22
SS	Saratoga Springs, NY	*L. m. samuelis*	C(5)	0.044	27

Table 2: AMOVA for mitochondrial gene regions in COI and COII. (A) AMOVA for populations grouped according to subspecific identifications based on morphological and ecological differences. (B) AMOVA for populations grouped according to regions identified by SAMOVA to maximize ϕ_{CT}.

A. Source of Variation	d.f.	Sum of squares	Variance component	% of total	P-value
Among subspecies	1	9.111	0.162	11.64	<0.001
Among populations/ within subspecies	6	36.391	1.208	86.56	<0.001
Within populations	32	0.802	0.025	1.80	0.096

B. Source of Variation	d.f	Sum of squares	Variance component	% of total	P-value
Among regional groups	1	45.335	3.020	99.16	<0.001
Among populations/ within groups	6	0.167	0.001	0.02	1.000
Within populations	32	0.802	0.025	0.82	0.041

CHAPTER II

HOMOPLOID HYBRID SPECIATION IN AN EXTREME HABITAT

To be published as: Gompert, Z., J. A. Fordyce, M. L. Forister, A. M. Shapiro, C. C. Nice. Homoploid hybrid speciation in an extreme habitat. *Science, in press*.

Introduction

Of the proposed mechanisms of speciation, homoploid hybrid speciation is among the least explored, especially in animals (Coyne and Orr 2004). Homoploid hybrid speciation occurs via hybridization between parental species, giving rise to a derivative hybrid species without a change in chromosome number (Grant 1981, Arnold 1997, Dowling and Secor 1997, Riesberg 1997, Gross and Riesberg 2005). The paucity of research on homoploid hybrid speciation in animals may be explained by these factors: 1) a history of bias among zoologists against hybridization as a mechanism generating biodiversity (i.e. the prevailing notion that hybridization leads to offspring with reduced fitness), 2) the difficulty of identifying diploid hybrid species, or 3) the rarity of this mode of speciation in animals (Arnold 1997, Dowling and Secor 1997). However, a growing list of possible examples (e.g. African cichlids (Schelly *et al.* 2006), cyprinid fishes (Demarais and Minckley 1992), *Rhagoletis* fruit flies (Schwarz *et al.* 2005), *Heliconius* butterflies (Mavarez *et al.* 2006), and swallowtail butterflies (Scriber and Ording 2005)) suggest that homoploid hybrid speciation in animals may be more common than previously thought. Models predict that homoploid hybrid speciation is facilitated by ecological isolation, especially in novel or even extreme habitats (compared to the parental habitats). The colonization of a novel habitat may allow incipient hybrid species to avoid: 1) assimilation via introgression by the parental species and 2) extinction through competition with the parental species (Dowling and Secor 1997, Buerkle *et al.* 2000). While these predictions have been borne out in plants (Riesberg 2003), no examples of homoploid hybrid speciation described in animals have involved adaptation to a novel

habitat, although a switch to a novel host plant species has been implicated in one case of hybrid speciation in animals (Schwarz *et al.* 2005).

Here we demonstrate that homoploid hybrid speciation has occurred in the butterfly genus *Lycaeides* and was facilitated by adaptation to an environmentally extreme habitat. The ecologically, morphologically, and behaviorally distinct species *L. melissa* and *L. idas* (Fordyce *et al.* 2002, Nice *et al.* 2002, Fordyce and Nice 2003, Forister *et al.* 2006) have come into secondary contact in the Sierra Nevada of western North America (Nice *et al.* 2005). Unnamed populations of *Lycaeides* occur in the alpine habitat (i.e. above tree-line) of the Sierra Nevada, which is an environmentally extreme habitat not occupied by *L. melissa* or *L. idas*. The alpine habitat is a comparatively extreme habitat characterized by a short growing season and severe fluctuations in ambient temperature and relative humidity on a daily and seasonal basis (Lomolino *et al.* 2006). These alpine populations have male genitalia that are intermediate in size and shape compared to *L. melissa* and *L. idas* (Nice *et al.* 2005, Lucas *et al.* in review), and they have wing pattern elements that are qualitatively similar to those of *L. melissa* (Fordyce *et al.* 2002). However, our analysis of mitochondrial DNA (mtDNA) variation demonstrates that the alpine populations' mitochondrial haplotypes share a more recent common ancestor with the haplotypes of *L. idas* than with those of *L. melissa* (Fig. S1) (*see methods*). These discordant patterns indicate that hybridization may have played a role in the evolutionary history of alpine *Lycaeides* populations.

Demonstration of Homoploid Hybrid Speciation

If the alpine *Lycaeides* populations are a hybrid species they should possess a mosaic genome that is a blend of alleles derived from both *L. melissa* and *L. idas*. This was tested using a large multilocus genomic dataset, which consisted of 128 Amplified Fragment Length Polymorphism (AFLP) markers, three microsatellite markers (*Msat201*, *Msat4*, and *MsatZ12-1*), and sequence data from three nuclear genes (*Nuc1*, *Nuc3*, and *Ef1α*) and two mitochondrial gene regions (*COI* and *COII*) (*see methods*). To assess the overall genomic composition of the alpine *Lycaeides* populations we used the Bayesian assignment algorithm of Pritchard *et al.* to cluster *L. melissa*, *L. idas*, and alpine individuals based on their multilocus genotypes. (*see methods*) (Pritchard *et al.* 2000). *Lycaeides melissa* individuals were assigned to one cluster with high probability, *L. idas* individuals were assigned to the alternative cluster with high probability, and alpine *Lycaeides* individuals were assigned to both clusters with moderate probabilities (Fig. 1A, Table S1). This pattern is inconsistent with a bifurcating mode of speciation and is evidence of the mosaic nature of the alpine genome. Alleles shared between *L. melissa* and the alpine populations but not with *L. idas*, and alleles shared between *L. idas* and the alpine populations but not with *L. melissa* provide additional evidence of the mosaic nature of the alpine genome. Twelve AFLP fragments were unique to *L. melissa* and the alpine populations and 16 AFLP fragments were unique to *L. idas* and the alpine populations, but just five AFLP fragments were found only in *L. melissa* and *L. idas*. *Lycaeides melissa* and *L. idas* were fixed for different alleles at the *Nuc1* locus (Fig. S2, Table S1), while the alpine populations shared three *Nuc1* alleles with *L. melissa* and

three *Nuc1* alleles with *L. idas*. These results demonstrate that the alpine populations possess a mosaic genome, sharing alleles with both of the parental species.

While the alpine populations' mosaic genome is consistent with the hypothesis of homoploid hybrid speciation, a similar pattern could have arisen if the alpine populations constitute a hybrid swarm where there is continuous gene flow between the hybrid populations and *L. melissa* and/or *L. idas*. If the alpine populations are a hybrid swarm they should not be genetically differentiated from *L. melissa* and *L. idas* and there should be evidence of gene flow with these species. When *Lycaeides* individuals were assigned to three clusters using the Bayesian assignment algorithm of Pritchard *et al.* (2000), *L. melissa* and *L. idas* individuals were still assigned to their respective clusters, but the alpine *Lycaeides* individuals were assigned to a distinct, third cluster (Fig 1B, Table S1) (21). Additionally, the alpine populations are fixed for unique alleles for the mitochondrial genes COI and COII as well as the nuclear gene *Nuc3* (Figs. S1, S3, Table S1). These data indicate that the alpine populations are differentiated from *L. melissa* and *L. idas*. We did not detect excess heterozygosity or deviations from linkage equilibrium for any of the microsatellite markers or nuclear gene sequences (*see methods*); such deviations would be indicative of substantial current gene flow between the alpine populations and either *L. melissa* or *L. idas*. We employed the Bayesian assignment algorithm of Anderson and Thompson (2002) to assess the probability that an individual was an F_1 between *L. melissa* and the alpine populations or between *L. idas* and the alpine populations (*see methods*). No F_1 individuals were identified (Fig. S4). Thus, we conclude that the alpine populations do not constitute a hybrid swarm, as they are

genetically differentiated from *L. melissa* and *L. idas* and we find no evidence of gene flow between the alpine populations and either of these species.

The genetic patterns documented above might have arisen if *L. melissa*, *L. idas*, and the alpine populations all arose rapidly from a single ancestral species distributed along a geographic cline with the alpine populations originating from the center of the cline. This scenario is unlikely for two reasons. Phylogeographic data suggest that the current distribution of *L. melissa* and *L. idas* is the result of post-Pleistocene range expansion, and thus does not reflect the distribution of the ancestor of these species (Nice *et al.* 2005). The alpine populations also have a more recent origin than either *L. melissa* and *L. idas*. We calculated a coalescent-based estimate of the time to the most recent common ancestor (TMRCA) for mitochondrial variation for each of the three putative species: the alpine populations, *L. melissa*, and *L idas* (see methods). The estimated TMRCA for the alpine populations, 442,579 years before present (ybp), is substantially younger than that of either *L. melissa* or *L. idas,* 1,902,995 ybp and 1,267,885 ybp, respectively (20). Furthermore, pairwise estimates of τ (species divergence time x mutation rate) based on nuclear and mitochondrial sequence data were approximately four times greater for the divergence of *L. melissa* and *L. idas* (0.006576, 95% CI 0.002823-0.009855) than for the divergence of the alpine populations and either *L. melissa* (0.001318, 95% CI 0.000638-0.002233) or *L. idas* (0.001468, 95% CI 0.000763- 002454) (20) (Fig. S5). These multilocus genetic data demonstrate that the alpine populations possess a mosaic genome, are differentiated from both *L. melissa* and *L. idas*, and are

younger than either of their putative parental species. We conclude that the alpine *Lycaeides* populations are the result of homoploid hybrid speciation.

Reproductive Isolation

Homoploid hybrid speciation is more likely when the hybrid species colonizes a novel habitat (Rieseberg 1997, Rieseberg *et al.* 2003, Gross and Rieseberg 2005). Colonization of the alpine habitat by *Lycaeides* is consistent with this model. Reproductive isolation between the hybrid and the parental taxa may currently be maintained by several behavioral and ecological adaptations to the alpine habitat that are specifically associated with the alpine host plant. Females from the alpine populations have near perfect host fidelity for their host plant, the alpine endemic, *Astragalus whineyi* (Fig. 2) (*see methods*). In fact, alpine females have stronger host fidelity than has been recorded in any other *Lycaeides* population. Because mating in *Lycaeides* takes place on or near the host plant (Nice *et al.* 2002), the strong fidelity for the alpine endemic host may serve as a strong pre-zygotic barrier to gene flow between alpine *Lycaeides* populations and their parental species. This strong host fidelity is coupled with another unique host plant-related trait of the alpine populations involving egg adhesion. While *Lycaeides* females from non-alpine populations "glue" their eggs to their host plant when they oviposit, the alpine populations lack egg adhesion and eggs fall off the plant after oviposition (Fordyce and Nice 2003). This is a relevant character because *Lycaeides* overwinters as diapausing eggs (Fordyce and Nice 2003). During the winter, the senesced above ground biomass of the alpine host plant is blown away by strong winds (Fordyce

and Nice 2003), which means any eggs attached to the alpine host plant will be carried far from the site of new host plant growth, and thus, neonate larvae will likely die. Any females from non-alpine populations that oviposit on the alpine host plant would suffer a major reduction in fitness. Eggs from the alpine populations fall to the substrate at the base of the host plant and remain near the site of new plant growth in the following spring (Fordyce and Nice 2003). These differences in egg adhesion constitute a strong ecological barrier to gene flow between alpine *Lycaeides* populations and their parental species. Together, these alpine-associated adaptive traits, strong host fidelity for an alpine endemic host plant and the loss of egg adhesion, may act as an effective barrier to gene flow. Consequently, the same characters that allow these populations to flourish in the extreme alpine environment are effective at limiting gene flow from their parental species.

Beyond the habitat isolation associated with adaptation to the alpine environment, at least two other mechanisms may also contribute to reproductive isolation. Color pattern differences on the underside of the wings operate as species recognition cues, isolating the alpine populations from *L. idas* (Fordyce *et al.* 2002). Differences in male genital morphology have been documented and may operate in a similar manner to limit gene flow, although this was not explicitly tested.

Conclusions

This study indicates that homoploid hybrid speciation is an important mechanism for generating biodiversity in these butterflies. Our genetic evidence demonstrates that

the genomes of alpine populations of *Lycaeides* are of hybrid origin (Figs. 1, S2). Furthermore, these hybrid populations are differentiated and reproductively isolated from the parental species (Figs. S1, S3-S4). They are also younger than the parental populations, as expected of hybrid species. Reproductive isolation of the hybrid species arises, at least in part, from specific adaptations to the extreme alpine environment, which include high host fidelity for the alpine endemic host plant combined with a loss of egg adhesion, and morphological differences affecting mate choice. In addition to providing the most convincing molecular evidence for homoploid hybrid speciation in animals to date, this is the first case in animals in which adaptation to an extreme, novel habitat creates reproductive isolation between the hybrid species and its parental species.

Methods

Collection and DNA Extraction

Both male and female *Lycaeides* were collected from eight populations in California and Nevada between 2001 and 2004. Collection localities were: Carson Pass, California (38° 42' 47.48" N, 120° 01' 17.44" W), Mt. Rose, Nevada (39° 19' 20.71" N, 119° 55' 48.36" W), Trap Creek, California (39° 22' 43.17" N, 120° 40' 27.42" W), Yuba Gap, California (39° 19' 23.75" N, 120° 35' 39.15" W), Leek Springs, California (38° 37' 55.88" N, 120° 14' 23.64" W), Verdi, Nevada (39° 03' 01.08" N, 119° 55' 48.36" W), Sierraville, California (39° 37' 47.61" N, 120° 21' 40.14" W), and Gardnerville, Nevada (38° 48' 54.44" N, 119° 46' 44.27" W) (Fig. S6). Individuals were

stored at -80°C prior to DNA extraction. DNA was isolated from approximately 0.5 g of thoracic tissue following the methods of Hillis *et al.* (1996) and Brookes *et al.* (1997).

Mitochondrial DNA Amplification and Phylogeny Construction

We sequenced portions of the mitochondrial genes *cytochrome oxidase c subunit I* (*COI*; 408 bp) and *cytochrome oxidase c subunit II* (COII; 507 bp) for 9-11 individuals from each of the eight populations. PCR and sequencing were performed using the primer pairs C1-J-1751/C1-N-2191 (Simon *et al.* 1994) and Pierre/Eva (Caterino and Sperling 1999) for *COI* and *COII* respectively. Labeled 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. Fourteen unique haplotypes were identified (Table S1) (GenBank Accessions EF090312-EF090347). Phylogenetic hypotheses for the relationship among the sampled haplotypes were constructed using both Bayesian maximum likelihood methods and maximum parsimony. *Plebejus icarioides*, *P. shasta*, and *P. saepiolus* were used as the outgroup. For the Bayesian maximum likelihood analysis the sequence data were partitioned by gene region and codon position. Sequence evolution models for each partition were selected using Modeltest 3.7 based on AIC (Posada and Crandall 2000). The models selected were HKY, F81, TrN+I, K81uf+I, TrN, and TIM+γ for *COI* 1st codon position, *COI* 2nd codon position, *COI* 3rd codon position, *COII* 1st codon position, *COII* 2nd codon position, and *COII* 3rd codon position respectively. Bayesian analysis was conducted using MrBayes ver 3.1.2 (Huelsenbeck and Ronquist 2001) with one cold chain and three hot chains run for 2 x 10^6 generations

with a burnin of 5 x 10^5 generations. Maximum parsimony tree construction was implemented using PAUP* version 4.0b10 (Swofford 2002) using the heuristic search option with TBR branch swapping. Statistical support for bifurcations was assessed via 1000 bootstrap replicates (Fig. S1).

Amplified Fragment Length Polymorphism (AFLP) Data

AFLP marker profiles were produced for 208 individuals (Table S1) sampled from each of the eight populations using three selective primer pairs: *EcoRI*-ACA and *MseI*-CTTG, *EcoRI*-ACA and *MseI*-CTTA, *EcoRI*-AGT and *MseI*-CTTA. Amplicons were separated and visualized on 6% denaturing polyacrylamide gels, using an ABIPRISM 377 DNA sequencer. GeneScan was used to visualize AFLP bands, which were sized by comparison to a standard ladder (ROX standard, Applied Biosystems) added to each lane. Band selection and quality control was performed following previously described methods (Gompert *et al.* 2006), which were shown to yield highly reproducible results. 128 polymorphic AFLP markers were generated.

Co-dominant Nuclear Loci

Non-coding nuclear genes were amplified using the following primer pair combinations: *Nuc1* (nuc1f 5'ACGCGTGATAAGGAACTTCG 3' and nuc1r 5'CTCTGCTTGCAATTCATTTTTG 3') and *Nuc3* (nuc3f 5'TTTTCCATAGCTTGACAAGAGC 3' and nuc3r

5'TTGAATGTACTGTAACTTCTTGGTG 3'). Primers for non-coding nuclear genes were designed following methods modified from Jennings & Edwards (2005). A portion of the coding gene *elongation factor 1 subunit alpha* (*Ef1α*) was amplified using the following primer pair: (E234f 5' GTCACCATCATYGACGC 3' and LEf1ar 5' ACTTGCCCTCGGCCTTAC 3'). Sequencing was performed as described for mtDNA methods. Ambiguous heterozygous individuals were confirmed by cloning (between 2 and 20 clones were examined). The sequenced products for *Nuc1*, *Nuc3*, and *Ef1α* were 364 bp, 291 bp, and 308 bp in length respectively. Microsatellites *Msat201*, *Msat4*, and *MsatZ12-1* were amplified with primers from Anthony *et al.* (2001). The distribution of genetic variation for all six of these loci is given in Fig. S7, Table S1. A maximum-parsimony haplotype network was constructed to visualize the genealogical relationships among the sampled alleles for each of the sequence loci using TCS 1.2.1 (Clement *et al.* 2000), which employs the statistical algorithms of Templeton *et al.* (1992) (Figs. S2-S3, S8).

Bayesian Assignment Analyses Using Structure

The Bayesian assignment analysis of Pritchard *et al.* (2000) was performed using STRUCTURE 2.1 with a burnin of 50,000 generations and a markov chain of 500,000 generations. Two clusters were assumed to test the hypothesis that the alpine populations possessed a mosaic genome; three clusters were assumed to determine if the alpine populations were genetically differentiated from *L. melissa* and *L. idas* (Fig. 1, Table S1). Two and three clusters best explained the data when we ran these analyses with the

number of clusters set from one to nine. When more than three clusters were assumed, individual populations of *L. melissa* began to be assigned to their own clusters, but the *L. idas* and alpine clusters remain intact (data not shown).

Tests of Hardy-Weinberg and Linkage Equilibrium

Gene flow among populations will only cause excess heterozygosity and deviations from linkage equilibrium if the populations are differentiated at the loci being examined. To determine if the eight *Lycaeides* populations differed at the co-dominant loci we examined, we computed pair wise F_{ST} values for all pairs of populations based on all six co-dominant loci using the software Arlequin ver. 2.000 (Schneider *et al.* 2000). Significance of F_{ST} was assessed via 1000 permutations of alleles between populations. We also performed the test of population differentiation for all pairs of populations for all six co-dominant loci using Popgene ver. 3.4, which employs the methods of Raymond and Rousset (1995). Values of α followed the suggestion of Moran (2003). The degree of population differentiation varied among loci (Table S2). Overall, populations were sufficiently differentiated to make tests of excess heterozygosity and linkage equilibrium meaningful.

Tests for excess heterozygosity were performed with the software FSTAT ver. 2.9.3.2 (Goudet 1995); the significance of F_{IS} was assessed via a permutation test (1000 permutations were performed). Tests of linkage equilibrium were performed with the software POPGENE ver. 1.31 (Yeh and Boyle 1997) following the methods of Weir (Weir 1979). All six co-dominant loci (*Msat201*, *Msat4*, *MsatZ12-1*, *Nuc1*, *Nuc3*, and

Ef1α) were included in these analyses. We did not detect a significant (at α=0.05) excess of heterozygostiy for any of the loci in any of the eight populations. No significant (at α=0.05) deviations from linkage equilibrium were detected for any of the eight populations.

Assignment Analysis to Identify F_1 Individuals

The Bayesian assignment analysis of Anderson and Thompson (2002) was used to classify *Lycaeides* individuals based on their genotypes at all six co-dominant loci (*Msat201*, *Msat4*, *MsatZ12-1*, *Nuc1*, *Nuc3*, and *Ef1α*). Two separate analyses were performed. For one analysis *L. melissa* and alpine individuals were classified as pure *L. melissa*, pure alpine, or F_1's between *L. melissa* and the alpine populations. For the other analysis *L. idas* and alpine individuals were classified as pure *L. idas*, pure alpine, or F_1's between *L. idas* and the alpine populations. For both analyses the markov chain was run for 10^6 generations with a burnin of 500,000 generations. Half of the individuals were used to set priors for the allele frequencies for *L. melissa*, *L. idas*, and the alpine populations, but these individuals were excluded from the estimation of the mixing proportions. The results of these analyses are shown in Figure S4. We did not attempt to classify individuals as F_2's and backcrosses as these categories would be difficult to distinguish from the hybrid species.

Estimation of the Time to the Most Recent Common Ancestor (TMRCA)

The TMRCAs for *L. melissa*, *L. idas*, and the alpine populations were estimated based on genetic variation at the mitochondrial loci *COI* and *COII* using the program Genetree (Griffiths and Tavare 1994). The maximum likelihood estimates of θ necessary for estimating TMRCAs were obtained using the program Fluctuate ver. 1.4 (Kuhner *et al.* 1995) (Table S3). A single rare haplotype was pruned from *L. idas* to make the data set compatible with the infinite sites model, which is required by Genetree. Coalescent time (T) was converted into years (t) using $t=2N_{ef}Tg$, where N_{ef} is the female effective population size based on θ and g is the generation time in years. The arthropod mutation rate proposed by Brower (1994) was assumed to calculate N_{ef} from estimates of θ. While this mutation rate may not be accurate for *Lycaeides*, this does not affect the relative ages of *L. melissa*, *L. idas*, and the alpine populations. Thus, the choice of this mutation rate does not affect the interpretation of our results. Dates were also calculated using the slower *Papilio* mutation rate (Zakharov *et al.* 2004) for comparison. *Lycaeides melissa* is generally bivoltine, while *L. idas* and the alpine populations are univoltine, thus generation times of 0.5 years, 1 year, and 1 year were used for *L. melissa*, *L. idas*, and the alpine populations respectively. Coalescent times and times in years are given in Table S3.

Estimation of species divergence times (τ)

We conducted a Bayesian based coalescent analysis of sequence data for *COI/COII*, *Nuc1*, *Nuc3*, and *Ef1α* using the software MCMCcoal (Rannala and Yang 2003) to estimate pairwise τ (divergence time x mutation rate) for *L. melissa*, *L. idas*, and

the alpine populations. τ was estimated using three separate pairwise analyses instead of a single analysis with all three species because MCMCcoal requires a known bifurcating species phylogeny for all species that are included in a single analysis (Rannala and Yang 2003). MCMCcoal requires user specified gamma priors for θ and τ. Identical gamma priors (α=4, β=2000) for θ were used for all three species. The same gamma prior (α=4.5, β=600) for τ was used for all species pairs. While the prior used for τ affected the magnitude of estimates of τ, the order of these estimates remained constant. The prior selected was relatively broad, and thus was chosen to minimize the influence of the prior on the posterior probability estimates. MCMCcoal analyses were conducted using a burn-in of 10,000 generations and a Markov chain of 100,000 generations. Multiple runs with different random seeds were conducted to assess convergence.

Host plant preference

Lycaeides populations vary in host plant use; the natal host plant of each population is given in Table S1. To assess female host plant fidelity individual females were caught and caged with a single male (individual caged females do not lay eggs; interaction among individuals is required to stimulate oviposition). Caged females were presented with the four host plants used by the eight populations in this study: *Astragalus whitneyi*, *Lotus nevadensis*, *Lupinus polyphyllus*, and *Medicago sativa*. After 48 hours the number of eggs a female had laid on each of the host plants was recorded. Host plant fidelity was measured as the proportion of eggs a female laid on her population's natal host plant. This experiment was performed from July 23-August 1 2006. Sample sizes for

each population were as follows: Carson Pass (n = 17), Mt. Rose (n = 7), Verdi (n = 13), Gardnerville (n = 8), Trap Creek (n = 8), Yuba Gap (n = 11), Leek Springs (n = 14). A recent change in the ownership of the property in Sierraville where the *M. saltiva* field exists prohibited us from assessing oviposition preference at that locality. Previous experiments demonstrate that this population does not show preference for their natal host plant (Quade Test, t=6.61, p=0.02 unpublished data). Additional oviposition preference data is provided in (Nice et al. 2001).

The rank-based Kruskal Wallis test was used to detect differences among populations in fidelity for the natal host plant. Implementation of the Kruskal Wallis test followed Conover (1999) using the test statistic *T*. Post-hoc comparisons among populations used the procedure described in Conover (1999). Natal host plant preference was significantly greater for both alpine populations (Carson Pass and Mt. Rose) than for any other populations (Fig. 2).

References

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217—1229 .

Anthony N, Gelembiuk G, Raterman D, Nice C, ffrench-Constant R (2001) Isolation and characterization of microsatellite markers from the endangered Karner blue butterfly *Lycaeides melissa samuelis* (Lepidoptera). *Hereditas* **134**, 271—273.

Arnold ML (1997) *Natural Hybridization and Evolution* Oxford University Press, Oxford.

Brookes MI, Graneau YA, King P, et al. (1997) Genetic analysis of founder bottlenecks in the rare British butterfly *Plebejus argus*. *Conservation Biology* **11**, 648—661.

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius-erato* inferred from patterns of mitochondrial-DNA Evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 6491—6495.

Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. *Heredity* **84**, 441—451.

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

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

Conover WJ (1999) *Practical nonparametric statistics* John Wiley and Sons, New York.

Coyne JA, Orr HA (2004) *Speciation* Sinauer Associates.

Demarais BD, Dowling TE, Douglas ME, Minckley WL, Marsh PC (1992) Origin of *Gila-Seminuda* (Teleostei, Cyprinidae) through Introgressive Hybridization-Implications for Evolution and Conservation. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 2747—2751.

Dowling TE, Secor CL (1997) The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics* **28**, 593—619.

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 recognition 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.

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

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485—486.

Grant V (1981) *Plant Speciation* Columbia University Press, New York.

Griffiths RC, Tavare S (1994) Simulating Probability-Distributions in the Coalescent. *Theoretical Population Biology* **46**, 131—159.

Gross BL, Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. *Journal of Heredity* **96**, 241—252.

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, Sunderland, MA.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754—755.

Jennings WB, Edwards SV (2005) Speciational history of Australian grass finches (*Poephila*) inferred from thirty gene trees. *Evolution* **59**, 2033—2047.

Kuhner MK, Yamato J, Felsenstein J (1995) Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* **140**, 1421—1430.

Lomolino MV, Riddle BR, Brown JH (2006) *Biogeography*, Third edn. Sinauer Associates, Sunderland, Massachusetts.

Lucas LK, Fordyce JA, Nice CC. Patterns of genitalic morphology around suture zones in North American *Lycaeides* (Lepidoptera: Lycaenidae): implications for taxonomy and historical biogeography. *In review.*

Mavarez J, Salazar CA, Bermingham E, et al. (2006) Speciation by hybridization in Heliconius butterflies. *Nature* **441**, 868—871.

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos* **100**, 403—405.

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

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

Nice CC, Anthony N, Gelembiuk G, Raterman D, ffrench-Constant R (2005) The history and geography of diversification within the butterfly genus *Lycaeides* in North America. *Molecular Ecology* **14**, 1741—1754.

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

Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645—1656.

Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution* **49**, 1280—1283.

Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics* **28**, 359—389.

Rieseberg LH, Raymond O, Rosenthal DM, et al. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211—1216.

Schelly R, Salzburger W, Koblmuller S, Duftner N, Sturmbauer C (2006) Phylogenetic relationships of the lamprologine cichlid genus *Lepidiolamprologus* (Teleostei : Perciformes) based on mitochondrial and nuclear sequences, suggesting introgressive hybridization. *Molecular Phylogenetics and Evolution* **38**, 426—438.

Schneider S, Kueffer JM, Roessli D, Excoffier L (2000) ARLEQUIN ver 2.0. A software program for population genetic data analysis. Genetics and Biometry Laboratory, Dept. of Anthropology and Ecology, Univ. Geneva, Geneva.

Schwarz D, Matta BM, Shakir-Botteri NL, McPheron BA (2005) Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* **436**, 546—549.

Scott JA (1986) *The butterflies of North America, a natural history and field guide* Stanford University Press, Stanford, California.

Scriber JM, Ording GJ (2005) Ecological speciation without host plant specialization; possible origins of a recently described cryptic *Papilio* species. *Entomologia Experimentalis et Applicata* **115**, 247—263.

Simmons RB, Weller SJ (2001) Utility and evolution of cytochrome b in insects. *Molecular Phylogenetics and Evolution* **20**, 196—210.

Stanford RE, Opler PA (1996) *Atlas of western USA butterflies, including adjacent parts of Canada and Mexico.*, Denver and Fort Collins, Colorado.

Swofford DL (2002) PAUP*: Phylogenetic analysis using parsimony (and other methods). Ver. 4.0b10. Sinauer, Sunderland, MA.

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.

Weir BS (1979) Inferences about linkage disequilibrium. *Biometrics* **35**, 235—254.

Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany* **129**, 157.

Zakharov EV, Caterino MS, Sperling FAH (2004) Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera : Papilionidae). *Systematic Biology* **53**, 193—215.

APPENDIX B

CHAPTER 2 FIGURES AND TABLES

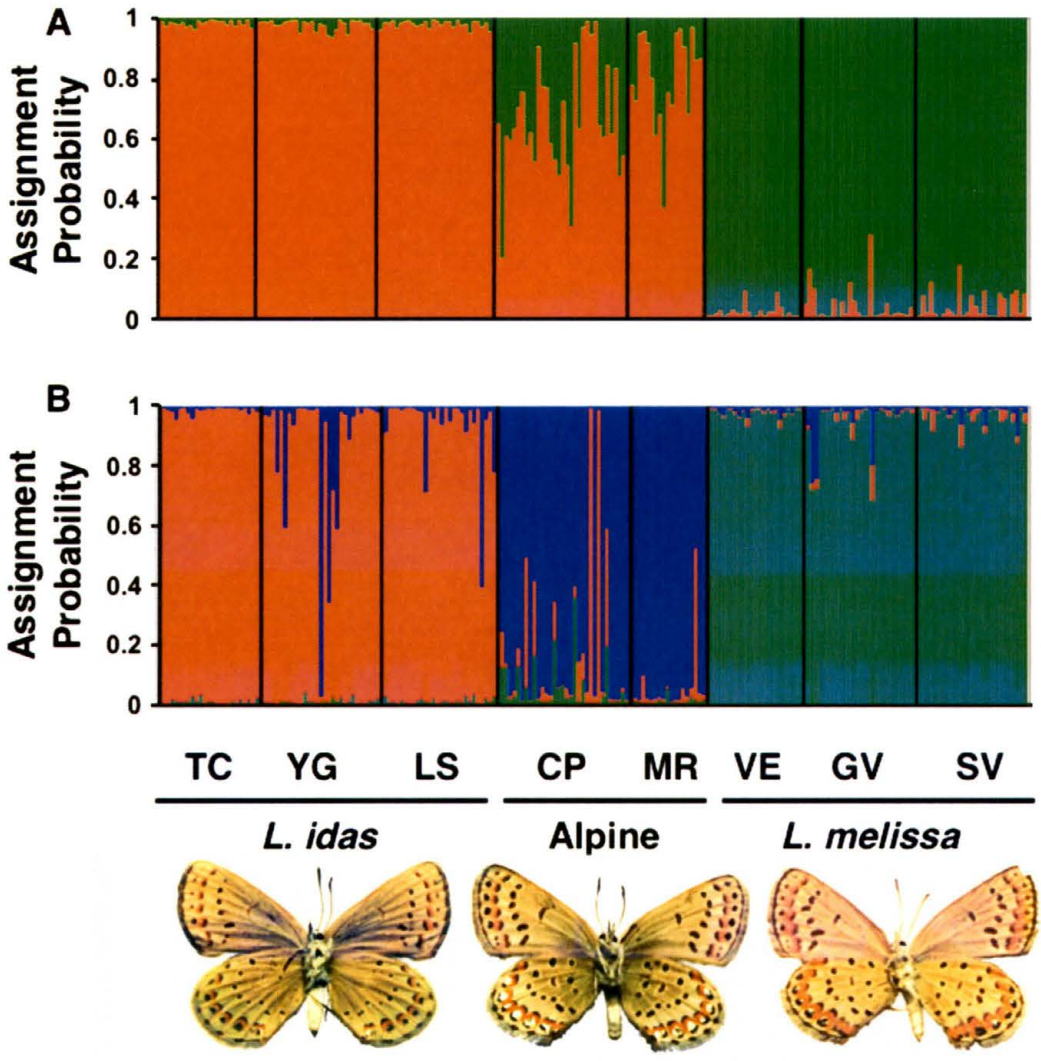

Figure 1. Barplots showing Bayesian assignment probabilities from the software STRUCTURE 2.1 (22) for two (Pritchard *et al.* 2000) and three (B) clusters. Each vertical bar corresponds to one individual. The proportion of each bar that is red, green, and blue represents an individuals assignment probability to clusters one, two, and three respectively. See Table S1 for mean population assignment probabilities. (CP = Carson Pass, MR = Mt. Rose, VE = Verdi, GV = Gardnerville, TC = Trap Creek, YG = Yuba Gap, LS = Leek Springs).

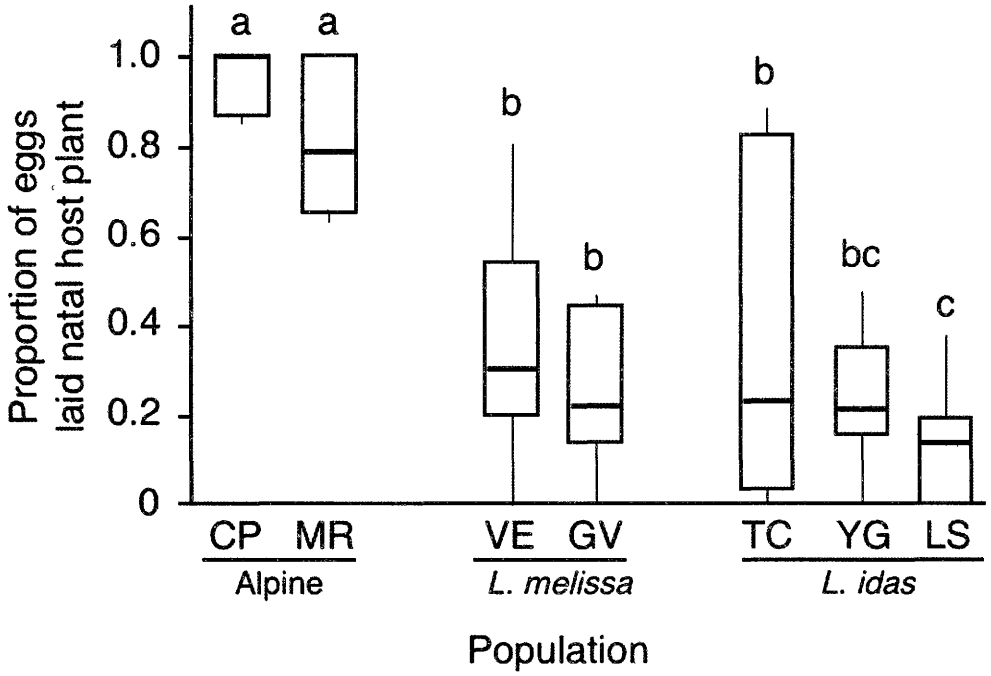

Figure 2. Natal host plant fidelity from seven focal populations. Box plots show the median proportion of eggs laid on the natal host plant for each population. Kruskal-Wallis test indicates significant differences in natal host plant preference among populations ($T = 46.67$; $p < 0.0001$). Different letters indicate differences in strength of preference for natal host among populations ($\alpha < 0.05$). Natal host plants are listed in Table S1. (CP = Carson Pass, MR = Mt. Rose, VE = Verdi, GV = Gardnerville, TC = Trap Creek, YG = Yuba Gap, LS = Leek Springs).

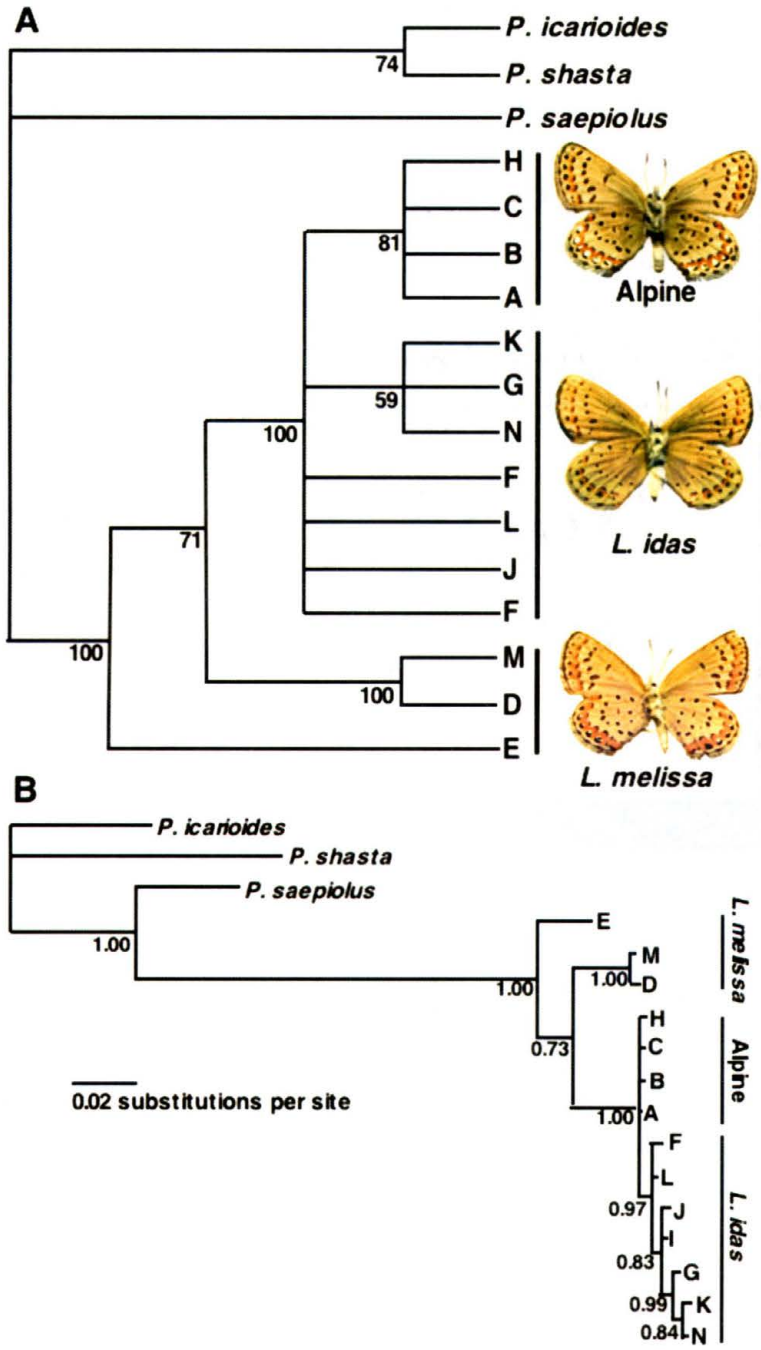

Figure S1. Phylogenetic relationships among mitochondrial haplotypes for *COI* and *COII*. Maximum parsimony phylogeny, numbers correspond to bootstrap support (A). Bayesian maximum likelihood phylogeny, numbers correspond to Bayesian posterior probabilities (B). Population allele frequencies are given in Table S1.

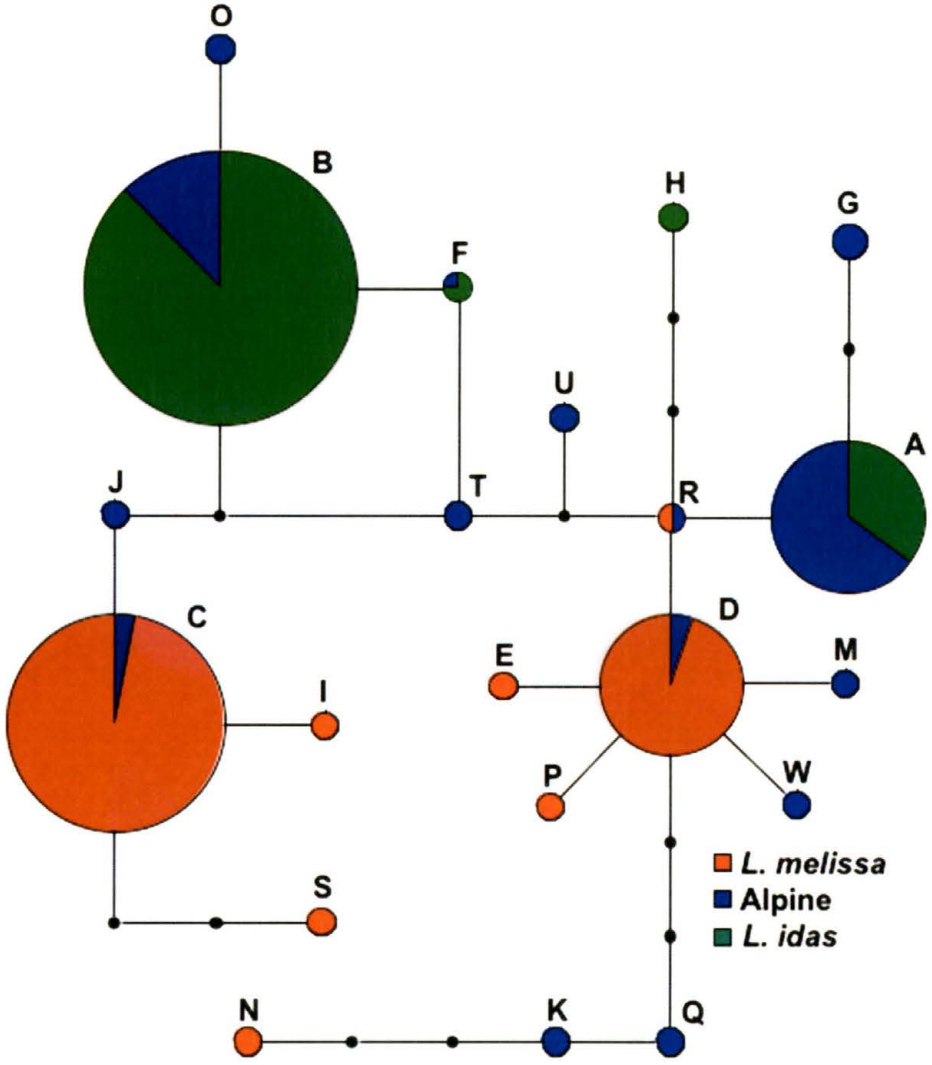

Figure S2. Maximum parsimony network for *Nuc1*. Each haplotype is represented by a circle and the size of each circle is approximately proportional to the frequency of the haplotype. Each haplotype is shaded green, red, and blue according to the proportion of the individuals that belong to *L. idas*, *L. melissa*, and alpine *Lycaeides* respectively.

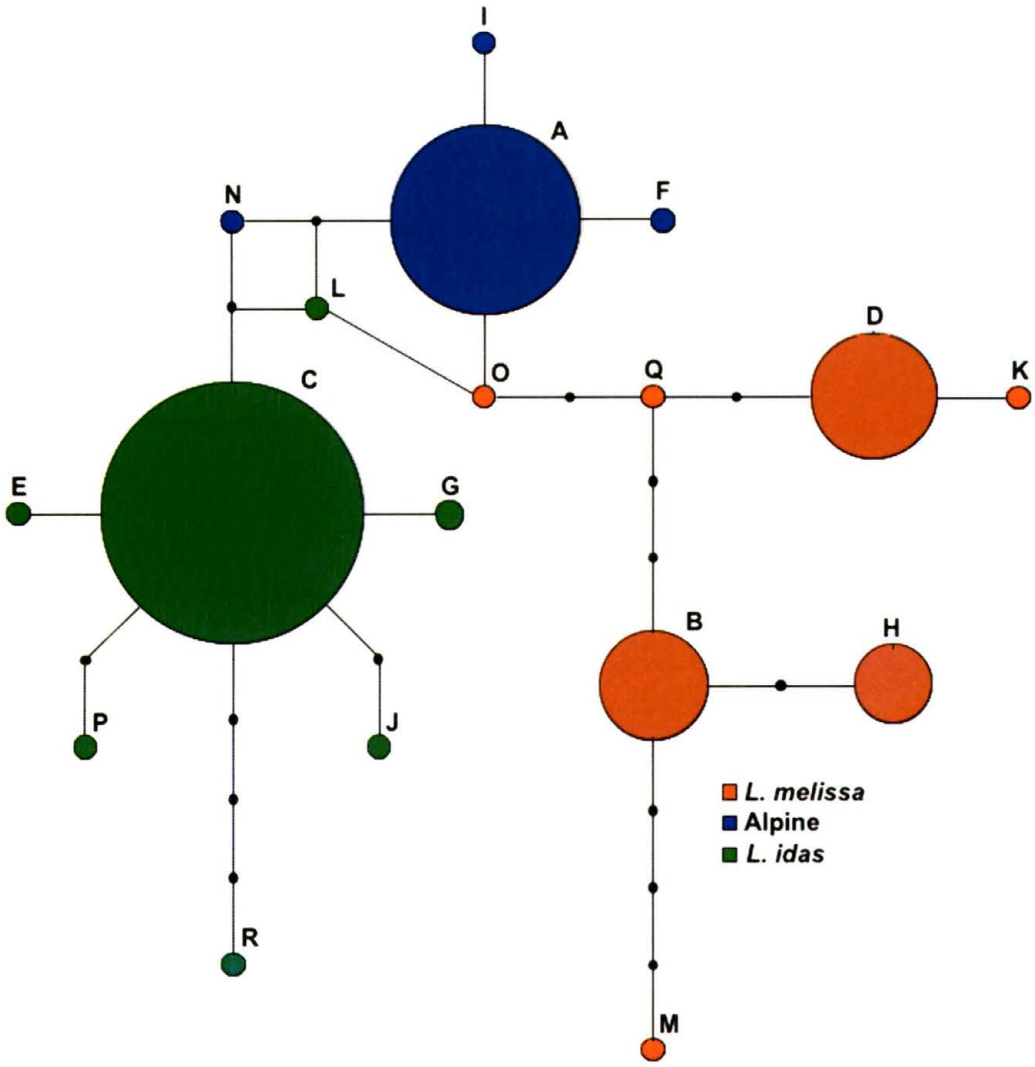

Figure S3. Maximum parsimony network for *Nuc3*. Each haplotype is represented by a circle and the size of each circle is approximately proportional to the frequency of the haplotype. Each haplotype is shaded green, red, and blue according to the proportion of the individuals that belong to *L. idas*, *L. melissa*, and alpine *Lycaeides* respectively.

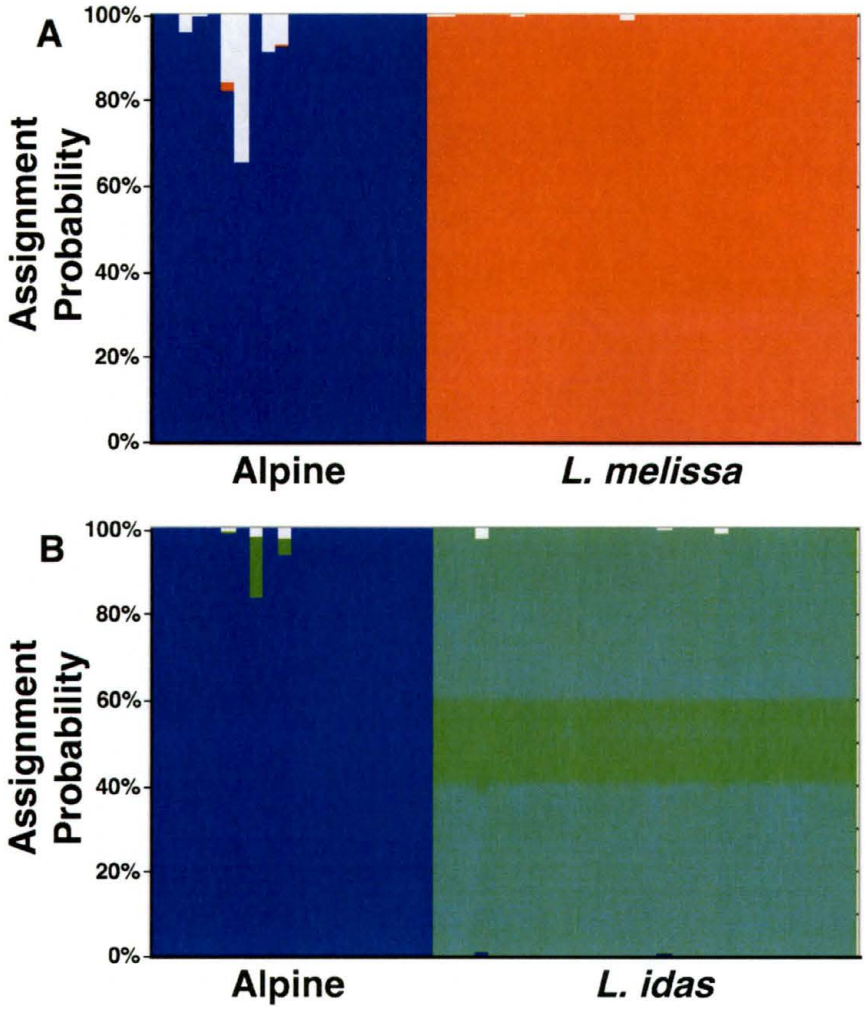

Figure S4. Barplots showing Bayesian assignment probabilities from the software NewHybrids (Anderson and Thompson 2002) for *L. melissa* and the alpine populations (A) and *L. idas* and the alpine populations (B). Each vertical bar represents an individual and the proportion of each bar represents an individual's assignment probability as pure *L. melissa* (red), pure *L. idas* (green), pure alpine *Lycaeides* (blue), and F_1 (white).

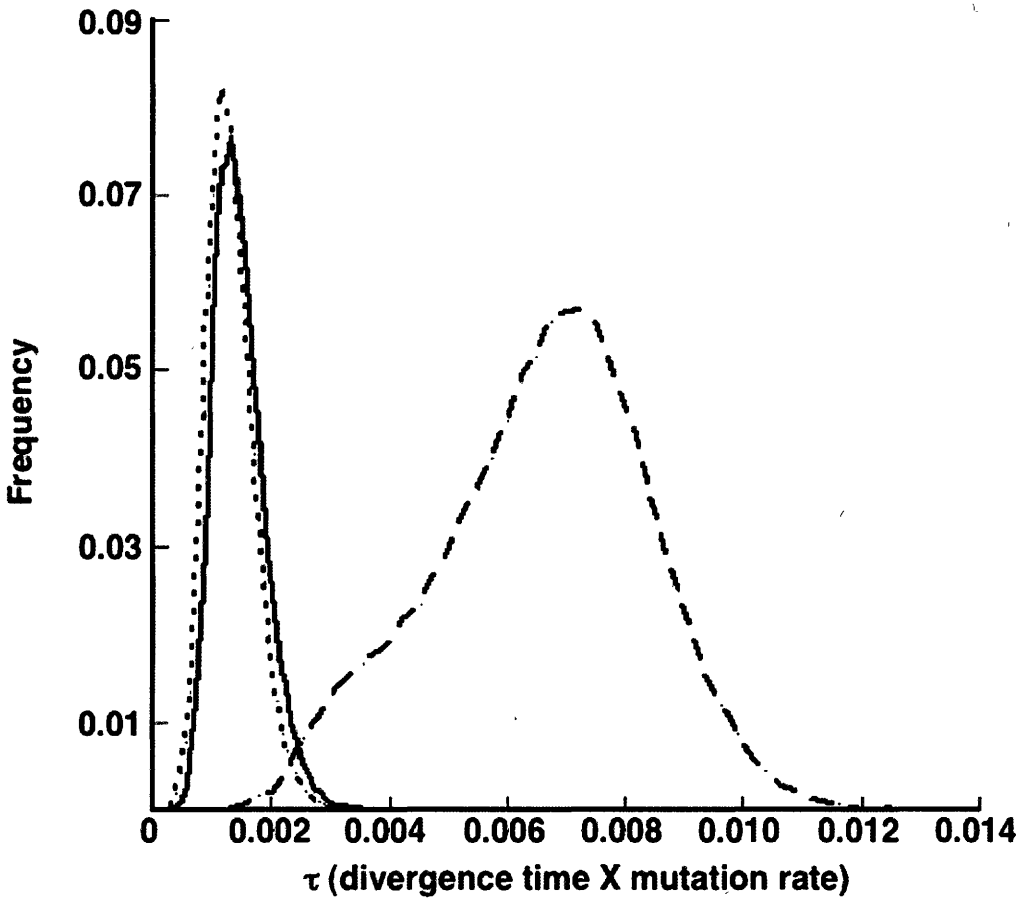

Figure S5. Posterior probability distribution for τ (divergence time x mutation rate) for *L. melissa-L. idas* (dashed line), *L. melissa*-alpine (dotted line), and *L. idas*-alpine (solid line). Mean (95% CI) estimates of τ for *L. melissa-L. idas*, *L. melissa*-alpine, and *L. idas*-alpine were 0.006576 (0.002823-0.009855), 0.001318 (0.000638-0.002233), and 0.001468 (0.000763-0.002454) respectively.

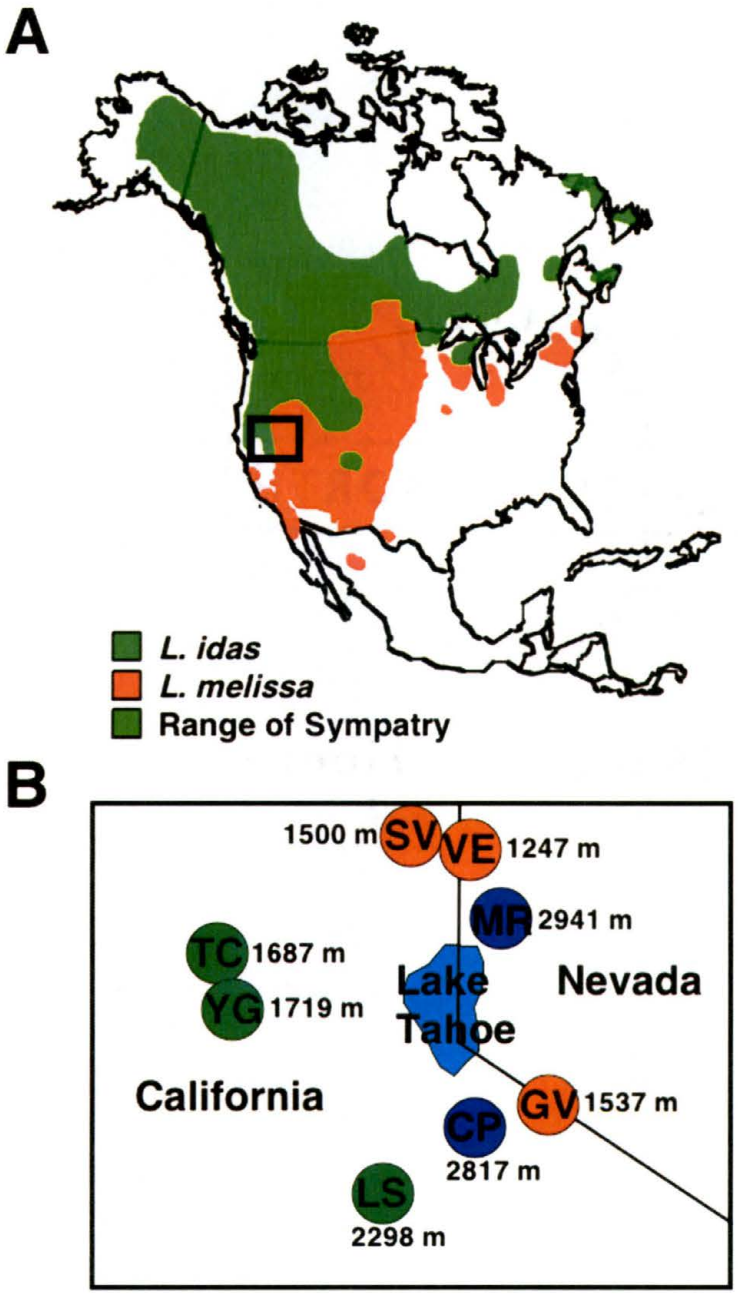

Figure S6. Approximate range of *Lycaeides melissa* and *L. idas* in North America (A) and sampling localities for this study (B). The range map follows Nabokov (1949), Stanford and Opler (1996) and Scott (1986). *Lycaeides idas* is shown in green, *L. melissa* is shown in red, and locations where these two species are sympatric are shown in dark green. The box denotes the focal region for this study. Sampling localities for *L. melissa*, *L. idas*, and alpine *Lycaeides* are shown in red, green, and blue respectively. (CP = Carson Pass, MR = Mt. Rose, VE = Verdi, GV = Gardnerville, TC = Trap Creek, YG = Yuba Gap, LS = Leek Springs).

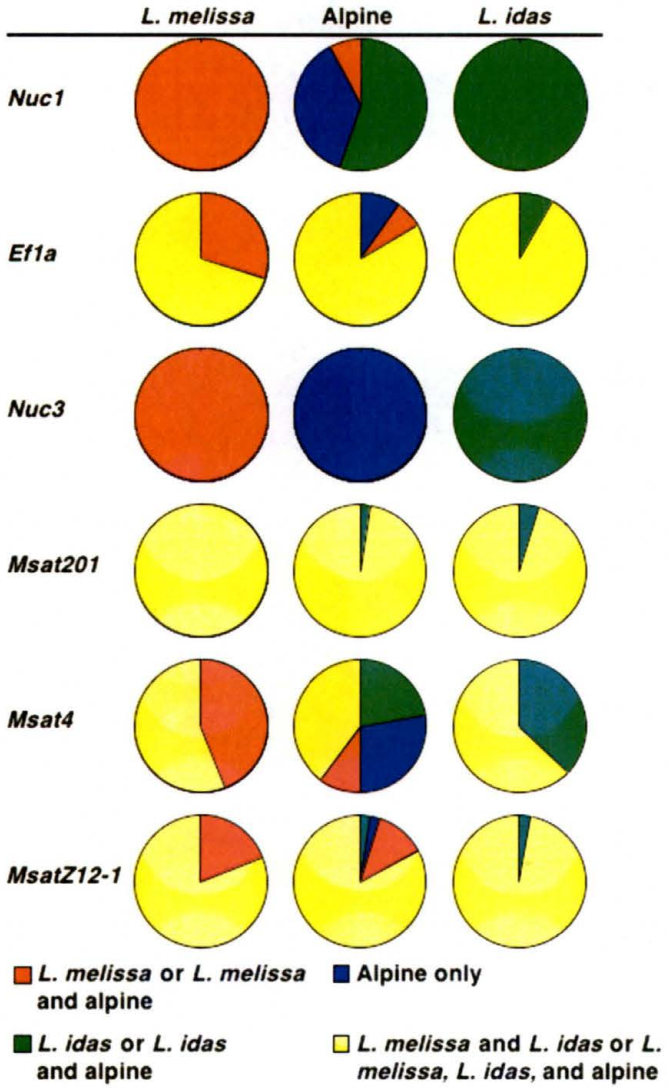

Figure S7. Distribution of allelic variation of co-dominant markers among *L. idas*, *L. melissa*, and the putative hybrid alpine populations. The proportion of each circle that is colored denotes the proportion of individuals with alleles that are unique to *L. idas* or *L. idas* and the alpine (green), unique to *L. melissa* or *L. melissa* and the alpine (red), unique to the alpine populations (blue), or shared between the parental species or the parental species and the alpine respectively (yellow). These loci demonstrate three general patterns. For *Nuc1* considerable allelic variation is shared between the alpine populations and the parental species. For *Nuc3* all three entities possess private alleles, while for *Ef1α*, *Msat201, Msat4*, and *MsatZ12-1* most of the allelic variation is shared among *L. idas*, *L. melissa*, and the alpine populations. Haplotype frequencies and sample sizes are given in Table S1.

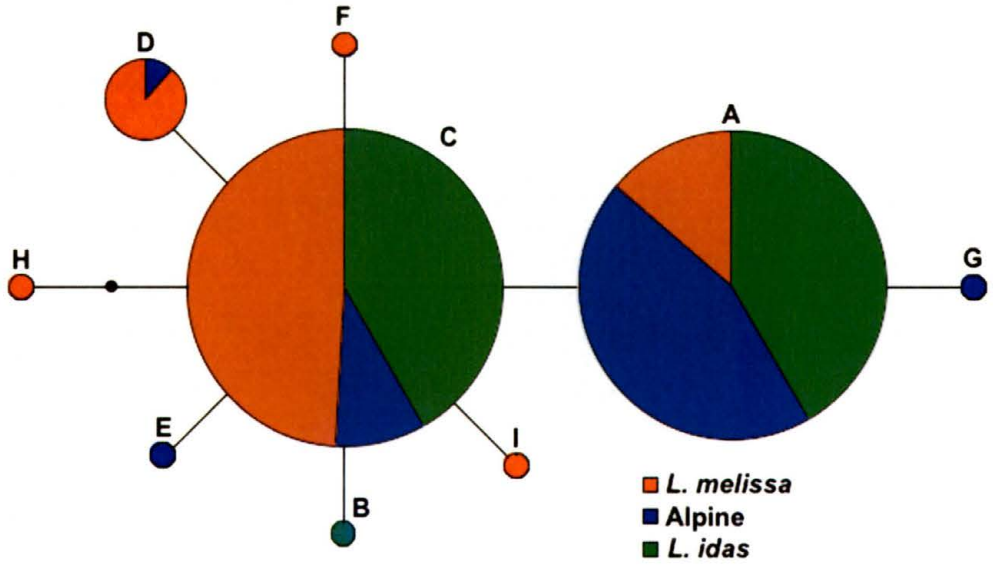

Figure S8. Maximum parsimony network for *Ef1α*. Each haplotype is represented by a circle and the size of each circle is approximately proportional to the frequency of the haplotype. Each haplotype is shaded green, red, and blue according to the proportion of the individuals that belong to *L. idas*, *L. melissa*, and alpine *Lycaeides* respectively.

Table S1. Summary data for all populations. Sample size and assignment probabilities are for the AFLP data. Sample sizes and allele frequencies are also given for all co-dominant markers and *COI/COII*.

Population (abbrev):	Trap Creek (TC)	Yuba Gap (YG)	Leek Springs (LS)	Carson Pass (CP)	Mt. Rose (MR)	Verdi (VE)	Sierra-ville (SV)	Gardner-ville (GV)
Nominal Taxonomic Designation	*L. idas*	*L. idas*	*L. idas*	Alpine	Alpine	*L. melissa*	*L. melissa*	*L. melissa*
Natal Host Plant	*Lotus nevadensis*	*Lotus nevadensis*	*Lupinus Polyphyllus*	*Astragalus whitneyi*	*Astragalus whitneyi*	*Medicago sativa*	*Medicago sativa*	*Medicago sativa*
Elevation (m)	1687	1719	2298	2817	2914	1247	1500	1537
Sample size	24	28	29	32	18	23	27	27
AFLP k=2								
Cluster 1	0.965	0.976	0.978	0.691	0.814	0.029	0.045	0.051
Cluster 2	0.020	0.024	0.022	0.309	0.186	0.971	0.955	0.949
AFLP k=3								
Cluster 1	0.944	0.916	0.895	0.141	0.058	0.010	0.019	0.017
Cluster 2	0.036	0.071	0.094	0.779	0.931	0.016	0.015	0.037
Cluster 3	0.020	0.013	0.011	0.080	0.011	0.974	0.965	0.946
Nuc1	10	10	10	10	10	11	10	10
A	0.20	0.15	0.05	0.05	0.70	0	0	0
B	0.65	0.85	0.90	).20	0.05	0	0	0
C	0	0	0	).05	0	0.59	0.25	0.75
D	0	0	0	0.05	0	0.36	0.50	0.10
E	0	0	0	0	0	0	0.10	0
F	0.15	0	0	0.05	0	0	0	0
G	0	0	0	0	0.25	0	0	0
H	0	0	0.05	0	0	0	0	0
I	0	0	0	0	0	0	0	0.10
K	0	0	0	0.15	0	0	0	0
N	0	0	0	0	0	0	0.05	0
P	0	0	0	0	0	0.05	0	0
Q	0	0	0	0.10	0	0	0	0
R	0	0	0	0.05	0	0	0	0.05
S	0	0	0	0	0	0	0.05	0
V	0	0	0	0	0	0	0.05	0
*	0	0	0	0.30	0	0	0	0
H_{obs}	0.30	0.10	0.20	0.90	0.60	0.66	0.50	0.40
H_{exp}	0.55	0.36	0.28	0.95	0.47	0.54	0.78	0.43
Ef1α	10	10	10	10	10	11	10	10
A	0.20	0.45	0.70	0.50	0.95	0.32	0	0.15
B	0.15	0.10	0	0	0	0	0	0
C	0.65	0.45	0.30	0.25	0.05	0.55	0.70	0.40
D	0	0	0	0.10	0	0.09	0.25	0.40
E	0	0	0	0.10	0	0	0	0
F	0	0	0	0	0	0.05	0	0
G	0	0	0	0.05	0	0	0	0
H	0	0	0	0	0	0	0.05	0
I	0	0	0	0	0	0	0	0.05

H_{obs}	0.50	0.50	0.40	0.80	0.10	0.55	0.60	0.50
H_{exp}	0.54	0.69	0.68	0.70	0.57	0.66	0.68	0.70
Nuc3	10	10	10	10	10	11	10	10
A	0	0	0	0.65	1.00	0	0	0
B	0	0	0	0	0	0.77	0.15	0
C	0.70	0.70	0.95	0	0	0	0	0
D	0	0	0	0	0	0.14	0.40	0.60
E	0	0	0.05	0	0	0	0	0
F	0	0	0	0.20	0	0	0	0
G	0.15	0.10	0	0	0	0	0	0
H	0	0	0	0	0	0.09	0.40	0.20
I	0	0	0	0.05	0	0	0	0
J	0.10	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0.05
L	0.05	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0.05	0
N	0	0	0	0.10	0	0	0	0
O	0	0	0	0	0	0	0	0.10
P	0	0.05	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0.05
R	0	0.15	0	0	0	0	0	0
H_{obs}	0.20	0.60	0.10	0.30	0.00	0.09	0.40	0.40
H_{exp}	0.58	0.50	0.19	0.63	0.00	0.48	0.74	0.62
Msat201	10	10	10	10	10	11	10	10
109	0.05	0.20	0.10	0.20	0	0.14	0.45	0.35
113	0.85	0.80	0.80	0.80	0.95	0.86	0.55	0.65
123	0.05	0	0	0	0.05	0	0	0
137	0	0	0.10	0	0	0	0	0
H_{obs}	0.20	0.50	0.40	0.20	0.10	0.27	0.30	0.30
H_{exp}	0.36	0.42	0.36	0.42	0.19	0.25	0.62	0.54
Msat4	10	10	10	10	10	11	10	10
204	0	0	0	0.05	0	0	0	0
210	0	0	0	0	0	0	0.05	0
212	0	0.05	0.05	0	0	0	0	0
214	0	0	0	0	0	0.05	0	0
218	0.15	0	0.15	0	0	0	0	0
220	0.45	0.45	0.10	0.15	0	0	0.05	0.20
222	0.05	0.05	0.10	0.15	0.40	0.14	0.20	0.65
224	0	0	0.05	0	0	0.05	0.20	0
226	0	0	0	0.15	0.10	0	0	0
228	0	0	0.05	0	0.40	0	0	0
230	0	0.10	0	0	0	0	0	0
232	0	0.05	0.05	0	0	0.05	0	0
234	0.05	0.10	0.25	0	0.10	0.09	0.15	0
236	0.05	0	0	0	0	0	0	0
238	0	0	0.05	0	0	0.05	0.10	0
240	0	0	0	0.15	0	0.05	0	0
242	0	0	0	0	0	0.05	0	0
244	0	0	0.05	0	0	0.18	0	0
246	0	0	0	0	0	0	0.10	0
248	0	0	0	0.05	0	0.32	0.15	0
250	0.05	0	0.05	0	0	0	0	0
252	0	0.05	0	0	0	0	0	0

256	0	0	0	0	0	0	0	0.15
258	0	0	0	0.10	0	0	0	0
260	0	0.15	0	0	0	0	0	0
262	0.05	0	0	0	0	0	0	0
264	0.05	0	0	0.05	0	0	0	0
268	0.05	0	0	0	0	0	0	0
270	0	0	0	0.15	0	0	0	0
292	0.05	0	0	0	0	0	0	0
H_{obs}	0.50	0.70	0.90	0.40	0.50	0.73	0.60	0.20
H_{exp}	0.84	0.83	0.89	0.93	0.70	0.89	0.92	0.61
MsatZ12-1	9	10	10	10	10	11	10	10
163	0.06	0	0	0	0	0	0	0
195	0	0	0	0	0	0.05	0	0
197	0.61	0.85	0.70	0.30	0.20	0.50	0.40	0.45
199	0.22	0.10	0.25	0.30	0.70	0	0.10	0
201	0.11	0	0	0.15	0	0.18	0.25	0.35
203	0	0	0	0.10	0.10	0.14	0.15	0.10
205	0	0	0.05	0	0	0.09	0	0.10
207	0	0	0	0.05	0	0	0	0
211	0	0.05	0	0.05	0	0	0	0
213	0	0	0	0.05	0	0.05	0	0
215	0	0	0	0	0	0	0.05	0
219	0	0	0	0	0	0	0.05	0
H_{obs}	0.78	0.30	0.40	0.70	0.20	0.45	0.60	0.50
H_{exp}	0.66	0.28	0.54	0.85	0.57	0.78	0.78	0.73
COI/COII	10	9	10	10	9	11	10	10
A	0	0	0	0.80	0.11	0	0	0
B	0	0	0	0.10	0	0	0	0
C	0	0	0	0.10	0	0	0	0
D	0	0	0	0	0	0.91	0	0.70
E	0	0	0	0	0	0	1.00	0.30
F	0	0	0.20	0	0	0	0	0
G	0	0	0.80	0	0	0	0	0
H	0	0	0	0	0.89	0	0	0
I	0.10	0	0	0	0	0	0	0
J	0.10	0	0	0	0	0	0	0
K	0.60	0.89	0	0	0	0	0	0
L	0.20	0	0	0	0	0	0	0
M	0	0	0	0	0	0.09	0	0
N	0	0.11	0	0	0	0	0	0

*This includes six alleles unique to Carson Pass each at a frequency of 0.05

Table S2. Pair-wise test of population differentiation and F_{ST}. P-values for Raymond and Rousset (1995) test of population differentiation are above the diagonal, pair-wise F_{ST} values are below the diagonal. Significant values at α=0.05 are in bold.

Nuc1	TC	YG	LS	CP	MR	VE	SV	GV
TC	-	0.22456	0.07251	**0.00618**	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**
YG	0.02931	-	0.60596	**0.00020**	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**
LS	0.09032	-0.01793	-	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**
CP	**0.12641**	**0.26092**	**0.30146**	-	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**
MR	**0.38941**	**0.56783**	**0.63959**	**0.25875**	-	**<0.00001**	**<0.00001**	**<0.00001**
VE	**0.45840**	**0.59094**	**0.62634**	**0.22159**	**0.49424**	-	0.05561	0.05256
SV	**0.37632**	**0.51316**	**0.55000**	**0.14149**	**0.41316**	0.07146	-	**0.00138**
GV	**0.51579**	**0.65263**	**0.68947**	**0.28079**	**0.55263**	0.03622	**0.23283**	-

Nuc3	TC	YG	LS	CP	MR	VE	SV	GV
TC	-	**0.020289**	**0.03484**	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**
YG	-0.01010	-	**0.02746**	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**
LS	0.10448	**0.10448**	-	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**	**<0.00001**
CP	**0.47368**	**0.47368**	**0.67368**	-	**0.00816**	**<0.00001**	**<0.00001**	**<0.00001**
MR	**0.75000**	**0.75000**	**0.95000**	**0.21053**	-	**<0.00001**	**<0.00001**	**<0.00001**
VE	**0.55461**	**0.55461**	**0.74737**	**0.52899**	**0.79501**	-	**0.00028**	**<0.00001**
SV	**0.40526**	**0.40526**	**0.60526**	**0.37895**	**0.65526**	**0.32062**	-	**<0.00001**
GV	**0.44211**	**0.44211**	**0.64211**	**0.41579**	**0.69211**	**0.44208**	0.04025	-

Ef1α	TC	YG	LS	CP	MR	VE	SV	GV
TC	-	0.24160	**0.00231**	**0.00399**	**<0.00001**	0.14808	**0.00236**	**0.00255**
YG	0.03909	-	0.17858	0.12737	**0.00159**	0.26437	**0.00001**	**0.00245**
LS	**0.25999**	0.03828	-	0.27392	0.09139	**0.03767**	**<0.00001**	**0.00024**
CP	**0.15789**	0.00695	0.00686	-	**0.01565**	0.15265	**<0.00001**	**0.01557**
MR	**0.58707**	**0.34928**	0.15296	**0.21951**	-	**0.00002**	**<0.00001**	**<0.00001**
VE	0.00182	-0.00992	0.13381	0.05207	**0.45399**	-	**0.01268**	0.05490
SV	**0.07291**	**0.20861**	**0.42372**	**0.26974**	**0.70548**	0.08578	-	0.07005
GV	0.13269	0.13272	**0.26995**	**0.11498**	**0.52867**	0.06268	0.06621	-

Msat201	TC	YG	LS	CP	MR	VE	SV	GV
TC	-	0.82543	0.53568	0.66360	0.73908	0.82373	**0.03077**	0.12565
YG	-0.02167	-	0.42147	1.00000	0.10617	0.53702	0.17417	0.48288
LS	-0.02716	-0.01801	-	0.41839	0.10647	0.39461	**0.01851**	0.07532
CP	-0.02632	-0.04532	-0.02167	-	0.10596	0.69037	0.18162	0.48201
MR	0.00277	0.09511	0.04605	0.08991	-	0.23309	**0.00110**	**0.00810**
VE	-0.04175	-0.01830	-0.02257	-0.03492	0.03186	-	**0.03841**	0.15246
SV	0.17949	0.05845	0.14665	0.08735	**0.34968**	0.17571	-	0.74608
GV	0.08134	-0.01100	0.05973	0.00513	**0.24321**	0.07416	-0.03093	-

Msat4	TC	YG	LS	CP	MR	VE	SV	GV
TC	-	0.15654	0.05449	**0.00195**	**<0.00001**	**<0.00001**	**0.00010**	**<0.00001**
YG	0.00067	-	**0.01393**	**0.00066**	**<0.00001**	**0.00010**	**0.00010**	**<0.00001**
LS	0.07172	**0.08254**	-	**0.00094**	**0.00106**	**0.01482**	0.10634	**<0.00001**
CP	0.07289	**0.08108**	**0.06674**	-	**0.00018**	**0.00036**	**0.00183**	**0.00004**
MR	**0.23617**	**0.23766**	**0.12666**	**0.12945**	-	**0.00033**	**0.00027**	**0.00002**
VE	**0.15976**	**0.15922**	**0.06682**	**0.06921**	**0.16559**	-	0.13941	**<0.00001**

SV	**0.12007**	**0.11866**	0.02215	0.05208	**0.12184**	0.01954	-	**0.00003**
GV	**0.23827**	**0.24426**	**0.21484**	**0.16453**	**0.16430**	**0.22494**	**0.16463**	-

MsatZ12-1	TC	YG	LS	CP	MR	VE	SV	GV
TC	-	0.13973	0.46605	0.23548	**0.00287**	**0.04488**	0.17546	**0.01700**
YG	**0.04967**	-	0.31887	**0.00234**	**<0.00001**	**0.00206**	**0.00251**	**0.00026**
LS	-0.02837	0.01662	-	**0.03174**	**0.00185**	**0.00407**	**0.00828**	**0.00056**
CP	0.03711	**0.23445**	**0.10195**	-	0.05649	0.10489	0.34756	**0.00186**
MR	**0.25442**	**0.49792**	**0.30465**	0.09722	-	**<0.00001**	**0.00039**	**<0.00001**
VE	0.02486	**0.12855**	0.07887	0.04683	**0.31895**	-	0.50392	0.80120
SV	0.02932	**0.19028**	**0.10640**	0.00346	**0.24677**	-0.01628	-	0.47623
GV	0.06350	**0.21585**	**0.14861**	0.06214	**0.34795**	-0.02245	-0.01962	-

Table S3. Coalescent-based estimates of the TMRCA for sampled mitochondrial haplotypes from *L. melissa*, *L. idas*, and the alpine populations. Coalescent time was converted into years before present (YBP) using *Nef* estimates from θ, a generation time of six months for *L. melissa* and one year for both *L. idas* and the alpine populations, and a mutation rate of *a*) 1.1×10^{-8}(Brower 1994) or *b*) 7.8×10^{-9} (Zakharov *el al.* 2004) substitutions per site per year.

	L. melissa	*L. idas*	Alpine
Coalescent Time	6.6949	3.7107	2.7406
SD	4.7257	1.4893	1.1751
θ	0.003127	0.001879	0.0008882
YBP *a*	1,902,995	1,267,885	442,579
YBP *b*	2,683,710	1,788,044	624,149

VITA

Zachariah Gompert was born in San Antonio, Texas, on February 12, 1981. He received his B.S. from Texas State University-San Marcos in 2004. He entered the population and conservation biology M.S. program at Texas State University-San Marcos in August of 2004.

Permanent address: 372 W. Nacogdouches St.

New Braunfels, TX 78130

This thesis was typed by Zachariah Gompert.