Molecular Ecology (2006) **15**, 1759–1768 doi: 10.1111/j.1365-294X.2006.02905.x

Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly

ZACHARIAH GOMPERT,* CHRIS C. NICE,* JAMES A. FORDYCE,† MATTHEW L. FORISTER‡ and ARTHUR M. SHAPIRO§

**Department of Biology, Population and Conservation Biology Program, Texas State University, San Marcos, TX 78666, USA, †Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA, ‡Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, NY 11794, USA, §Section of Evolution and Ecology, University of California, Davis, CA 95616, USA*

Abstract

The federally endangered North American Karner blue butterfly (*Lycaeides melissa samuelis*) and the closely related Melissa blue butterfly (*L. m. melissa*) can be distinguished based on 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. Here we test 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 informative amplified fragment length polymorphism (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*. This supports 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.

Keywords: AFLP, conservation genetics, DNA barcoding, genomic divergence, *Lycaeides melissa samuelis*, mitochondrial introgression

Received 5 October 2005; revision accepted 14 December 2005

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 history (Crandall *et al*. 2000). At present, 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

Correspondence: Zachariah Gompert, Fax: (512) 245–8713; E-mail: zg1002@txstate.edu

American endangered Karner blue butterfly (*Lycaeides melissa samuelis*).

The Karner blue butterfly (*L. m. samuelis*) and its close relative, the Melissa blue butterfly (*L. m. melissa*), are small butterflies belonging to the family Lycaenidae. *L. m. samuelis* has experienced a 99% 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 an endangered species in the United States in 1992 (US Fish and Wildlife Service 1992, 2003). 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, 2003). The closely related *L. 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 *Lupinus perennis* (Scott 1986; US Fish and Wildlife Service 1992; Lane & Weller 1994). *L. m. 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 & Krizek 1984; Lane & Weller 1994) and in the size and shape of the male genitalia (Nabakov 1949; Lane & Weller 1994; C. C. Nice, unpublished).

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 characteristics (Nabakov 1949; Lane & Weller 1994) and between the patterns of genetic variation observed for 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 geographic 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 2003a). 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 2003).

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 can be made. 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; Wang *et al*. 2003; Mock *et al*. 2004; Irwin *et al*. 2005) and to 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 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 *Lycaeides melissa samuelis* populations and three *Lycaeides melissa melissa* populations (Fig. 1B, Table 1). Both males and females were collected

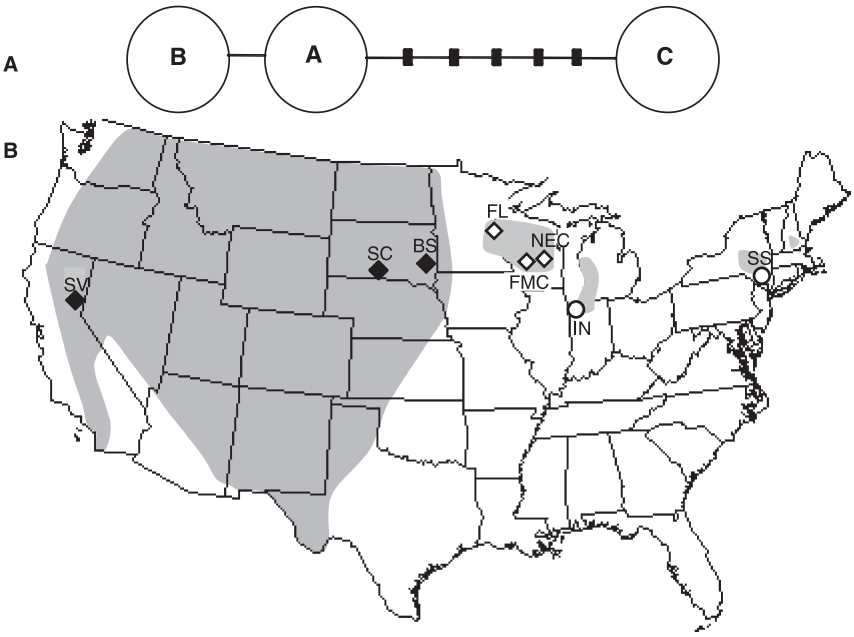

Fig. 1 Mitochondrial DNA haplotype network and population map. (A) Mitochondrial DNA haplotype network showing the single most parsimonious haplotype network for the three haplotypes identified. Each circle represents a haplotype. Black squares represent missing haplotypes. Haplotype C is separated from haplotypes A and B by six and seven mutations, respectively. (B) Population map. Dark grey shading marks approximate range of *Lycaeides melissa melissa* and light grey shading marks the approximate range of *Lycaeides melissa samuelis*. Population abbreviations are given in Table 1. 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 (Fig. 3). The pattern of molecular variation is discordant between mtDNA data and AFLP markers.

Table 1 Population data. Populations are labelled 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, Fig. 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

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 accordance with USFWS permit PRT842392. DNA was isolated from about 0.5 g of thoracic tissue following the methods of Hillis *et al*. (1996) and Brookes *et al*. (1997).

Mitochondrial DNA

We sequenced portions of the mitochondrial gene cytochrome oxidase *c* subunit 1 (COI) and cytochrome oxidase *c* subunit II (COII) for five individuals from each of the eight populations. Polymerase chain reaction (PCR) was performed using the primer pair C1-J-1751/C1-N-2191 for COI (Simon *et al*. 1994) and Pierre/Eva for COII (Caterino & Sperling 1999). This yielded fragments of approximately 450 and 550 base pairs (bp) for COI and COII, respectively. Fluorescently labelled dideoxy terminators were used for single-stranded sequencing reactions for both COI and COII according to Applied Biosystems specifications. Labelled amplicons were separated and visualized on 6% denaturing polyacrylamide gels 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* version 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 data set 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 & 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 gels, using an ABI PRISM 377 DNA sequencer (Applied Biosystems). GeneScan was used to visualize AFLP bands, which were sized by comparison to a size standard ladder (ROX standard, Applied Biosystems) 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

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

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 burn-in 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

Sequences were obtained for 410 bp of COI and 510 bp of COII for all 40 individuals examined (GenBank Accessions DQ234691–DQ234697). A 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 (Fig. 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 *Lycaeides melissa melissa* populations were fixed for haplotype A (Fig. 1B). A single individual from the Brandon, SD *L. m. melissa* population had haplotype B, while the other four individuals had haplotype A (Fig. 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 (Fig. 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 ($\Phi_{CT} = 11.64$, $P < 0.001$, 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 and (ii) the *L. m. samuelis* east of Lake Michigan ($\Phi_{CT} = 99.16$, $P < 0.001$, 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.

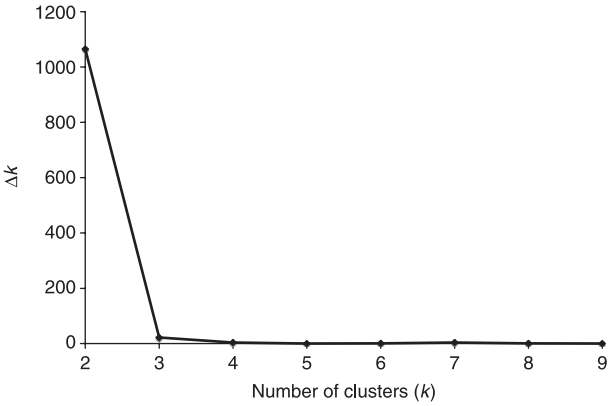

Fig. 2 The number of clusters (k) vs. the second order rate of change in k (Δk). The clear maximum for Δk at $k = 2$ indicates that two clusters best explain the AFLP data for the sampled *Lycaeides melissa* populations.

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 (Fig. 2). 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. *L. m. 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 (Figs 1B and 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 (Figs 1B and 3). This includes the Wisconsin *L. m. samuelis* populations that were grouped with *L. m. melissa* based on mtDNA. 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 (Fig. 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 Lycaeides melissa samuelis

Mitochondrial DNA (COI and COII) and AFLP markers identified different boundaries between *Lycaeides melissa samuelis* and *Lycaeides melissa 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*

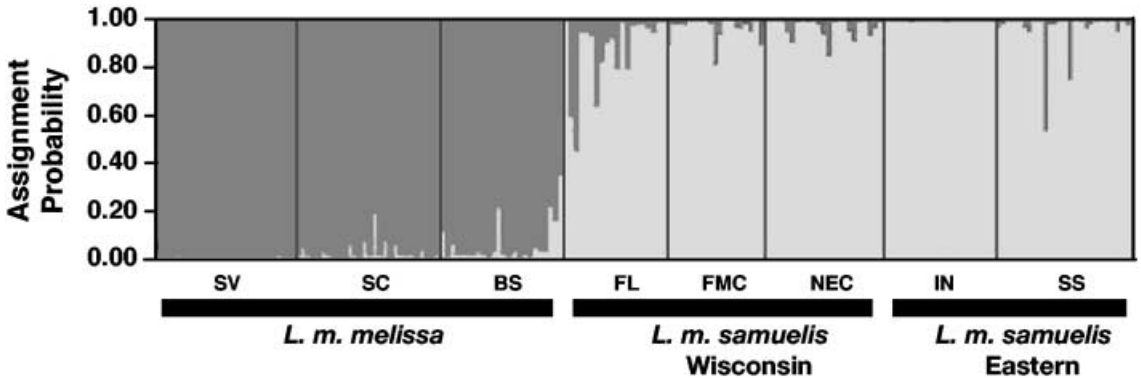

Fig. 3 Bayesian assignment probabilities for $k = 2$. Each vertical bar corresponds to one individual. The proportion of each bar that is dark grey represents an individual's assignment probability to cluster 1; the proportion of each bar that is light grey represents an individual's assignment probability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

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. This degree of sequence divergence is typical of other butterfly subspecies (e.g. Nice & Shapiro 2001; Fordyce & Nice 2003b). 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 phylogeographic 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 (Fig. 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 become 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 (Fig. 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* × *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 little 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 the evolutionary history of the mitochondrial genome is not indicative of the history of the nuclear genome, which means that the mtDNA data do 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; Holland & Hadfield 2002; Hebert *et al*. 2003, 2004; Tautz *et al*. 2003). This trend has met with a number of criticisms (e.g. Will & Rubinoff 2004; Prendini 2005; Wheeler 2005; Will *et al*. 2005). As stated by some of these critics, data from a single locus such as mtDNA should be used with caution. In this case, mtDNA incorrectly identifies the Wisconsin populations of the endangered species *L. m. samuelis* as populations of the widespread *L. m. melissa*. This is a case in which DNA systematics 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. This does not mean that mtDNA data should be ignored in general, as mtDNA has been used effectively to identify units for conservation (e.g. Holland & Hadfield 2002); however, we recommend obtaining corroborating evidence from nuclear markers to support conclusions from mtDNA.

Acknowledgements

We thank N. Anthony, G. Gelembiuk, R. ffrench-Constant, S. Fuller and M. Amaral for assistance in the field. J. Ott, B. Fitzpatrick, M. Forstner, B. Holland, an anonymous reviewer and the Texas State Evolution, Ecology and Behaviour discussion group provided helpful comments and discussion. C. Carnes and the USFWS assisted with logistical support and permitting. This work was funded by a Research Enhancement Grant from Texas State University to C. Nice.

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, Massachusetts.

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

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, Massachusetts.

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

Fordyce JA, Nice CC (2003b) Contemporary patterns in a historical context: phylogeographic history of the pipevine swallowtail *Battus philenor*, (Papilionidae). *Evolution*, **57**, 1089–1099.

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, USA*, **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, Massachusetts.

Holland BS, Hadfield MG (2002) Islands within an island: phylogeography and conservation genetics of the endangered Hawaiian tree snail *Achatinella mustelina*. *Molecular Ecology*, **11**, 365–375.

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. *Minnesota Agricultural Experiment Station Miscillaneous Publication Series*, **84**, 5–21.

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

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 Massachusetts.

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 & 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 Maryland.

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.

Prendini L (2005) Comment on 'Identifying spiders through DNA barcodes.' *Canadian Journal of Zoology*, **83**, 498–504.

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

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 California.

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

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.

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 (2002) PAUP* *Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0b10. Sinauer Associates, Sunderland Massachusetts.

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: determination of endangered status for the Karner blue butterfly. Proposed rule. *Federal Register*, **57**, 2241–2246.

US Fish and Wildlife Service (2003) Final recovery plan for the Karner blue butterfly (*Lycaeides melissa samuelis*). US Fish and Wildlife Service, Fort Snelling, Minnesota.

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.

Wheeler QD (2005) Losing the plot: DNA 'barcodes' and taxonomy. *Cladistics*, **21**, 405–407.

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.

Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics*, **20**, 47–55.

Zachariah Gompert is a Masters student at Texas State University with research interests in evolutionary biology, speciation, and genetics. Chris Nice is Assistant Professor at Texas State University with research interests in evolutionary and ecological genetics. James Fordyce is Assistant Professor at University of Tennessee with research interests in factors that promote population differentiation and plant insect interactions. Matt Forister, a postdoctoral research associate at the State University of New York, Stony Brook, is interested in speciation, specialization, and the genetics of host use in herbivorous insects. Arthur M. Shapiro is a professor at the University of California, Davis. He studies the biogeography, evolution, and phenology of butterflies in western North America and the Andean-Patagonian region.