Molecular Ecology (2005) **14**, 1741–1754 doi: 10.1111/j.1365-294X.2005.02527.x

The history and geography of diversification within the butterfly genus *Lycaeides* in North America

CHRIS C. NICE,* NICOLA ANTHONY,† GREG GELEMBIUK,‡ DENISE RATERMAN§ and RICHARD FFRENCH-CONSTANT¶

**Department of Biology, Texas State University–San Marcos, 601 University Dr., San Marcos, Texas 78666, USA, †Department of Biological Sciences, University of New Orleans, Louisiana, 70148, USA, ‡Department of Botany, University of Wisconsin, Madison, Wisconsin 53706, USA, §Department of Biology, University of California, Riverside, California 92521, USA, ¶Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK*

Abstract

The *Lycaeides* butterfly species complex in North America consists of two nominal, morphologically defined species. These butterflies are ecologically diverse and appear to be distributed as a geographically complex mosaic of locally differentiated populations that may be undergoing adaptive radiation. We asked whether patterns of molecular genetic variation within the species complex are congruent with currently recognized morphological species and whether the distribution of molecular variation is consistent with the hypothesis that Pleistocene climate changes contributed to the process of differentiation within the genus. Variation in the form of the genitalia from 726 males from 59 populations clearly distinguishes both species with only six populations containing morphologically intermediate or ambiguous individuals. However, partitioning of molecular variance in a 236 bp section of the mitochondrial AT-rich region from 628 individuals (57 populations) surveyed using single strand conformation polymorphism analysis (SSCP) indicates that only 26% of the total genetic variation is distributed along nominal species boundaries as defined by morphology. Instead, three phylogeographical groups were detected, represented by three major haplotype clades, which account for 90% of the total genetic variance. Pleistocene glaciations appear to have fostered divergence during glacial maxima, while postglacial range expansions created opportunities for gene exchange and reticulation along suture zones between geographical groups. Data presented here allow us to make inferences about the history of the species complex. However, evidence of ancestral polymorphism and reticulation limit our ability to define species boundaries based on mitochondrial DNA sequence variation.

Keywords: *Lycaeides*, mtDNA, nested clade analysis, North American phylogeography, Pleistocene refuges, suture zones

Received 13 October 2004; revision received 18 January 2005; accepted 7 February 2005

Introduction

Detailed information about the geographical distribution of population genetic variation within and among diverging taxa is critical for analyses of the evolutionary processes driving divergence and speciation (Hewitt 1996, 2000; Avise & Walker 1998; Klicka & Zink 1999; Knowles 2000, 2001). Analysis of phylogeographical structure is particularly important for organisms with extensive ranges and complex geographical patterns of polymorphism (Irschick & Shaffer 1997; Starkey *et al*. 2003). Phylogeographical data can be used to infer the biogeographical and demographic history of such organisms (Avise 1994). When integrated with data on the geographical distribution of morphological or ecological variation, such inferences can be used to test alternative hypotheses on the nature of the speciation process. Here we report the results of a continental-scale phylogeographical investigation of the *Lycaeides* species complex (Lepidoptera: Lycaenidae). This polytypic group of butterflies from North America exhibits extensive local morphological, ecological and behavioural diversity and

Correspondence: Chris C. Nice, Fax: 512-245-8713; E-mail: ccnice@txstate.edu

© 2005 Blackwell Publishing Ltd

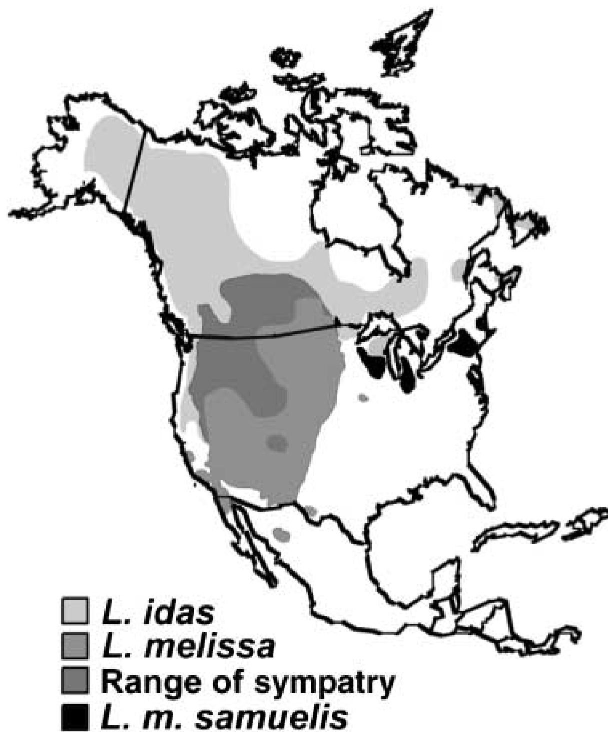

Fig. 1 Approximate ranges of *Lycaeides idas* and *Lycaeides melissa* in North America following Nabokov (1949), Opler (1992), Stanford & Opler (1996) and Scott (1986). The species' ranges overlap in west-central North America. The easternmost portion of the range of *L. melissa*, shaded black, represents the historical range of the endangered *Lycaeides melissa samuelis*.

represent an excellent opportunity to study mechanisms of evolutionary divergence and speciation.

Vladimir Nabokov revised the classification of North American members of the butterfly genus *Lycaeides* (Lepidoptera: Lycaeinidae) (Nabokov 1943, 1949). Using wing colour patterns and male genital morphology, Nabokov recognized two species: *Lycaeides idas*, the northern blue, and *Lycaeides melissa*, the Melissa blue. These species' ranges overlap in western North America (Fig. 1), occurring in sympatry or parapatry in a few geographically restricted areas (Nabokov 1949; Layberry *et al*. 1998; C.C. Nice, personal observation). Despite potential opportunities for gene flow between species where their ranges closely interdigitate, the morphological characters described by Nabokov appear to reliably distinguish both species throughout most of their ranges.

Besides morphological distinction, differences in behaviour and ecology exist between the species and among populations within species. Recent investigations of *L. idas* and *L. melissa* populations in the northern Sierra Nevada of California and surrounding areas indicate that some of the relatively subtle differences in wing pigment pattern used to distinguish populations are also recognized by the butterflies and appear to play a role in mate recognition (Fordyce *et al*. 2002). There is also variation in host use within the nominal species. For example, populations of *L. idas* occurring in close proximity in the Sierra Nevada prefer different host plant species, all of which differ from the hosts used by nearby *L. melissa* populations (Nice *et al*. 2002). Eggs from neighbouring populations may also differ in adhesion properties. In most *Lycaeides* populations, eggs are 'glued' to host plants during oviposition and remain there throughout winter until hatching. In populations of *L. melissa* above treeline in the Sierra Nevada, however, eggs are poorly adhered and fall off of host plants soon after oviposition. This behaviour may be an adaptation to the senescence characteristics of the preferred larval host plant species (Fordyce & Nice 2003). Ecotypic differentiation within *Lycaeides* may be occurring, driven by selection favouring habitat or host plant specialization. Our first step in exploring this possibility is an investigation of the geography and history of differentiation.

The investigation of the origins of locally adapted forms within the genus *Lycaeides* may help explain why these butterflies and members of the family Lycaenidae seem prone to endangerment. The genus *Lycaeides*, for example, includes several subspecies that are imperiled. The Karner blue butterfly, *Lycaeides melissa samuelis*, is federally endangered and occupies only a small fraction of its former range in northeastern United States (Andow *et al*. 1994). The Lotis blue, *Lycaeides idas lotis*, was endangered but is now presumed to be extinct (Arnold 1993). Nabokov's blue, *Lycaeides idas nabokovi*, is considered endangered by the state of Wisconsin. Understanding the processes of differentiation may illuminate the causes of endangerment and inform management decisions.

The goals of this study were to investigate morphological and molecular patterns of variation in these polytypic butterflies at a continental scale and evaluate the extent to which historical factors may have contributed to differentiation. Specifically, do these butterflies constitute a complex of numerous, locally differentiated and independent lineages, or, do they represent two evolutionarily significant units that have undergone limited introgression at their range limits, as the current taxonomy suggests? A survey of variation in the form of the male genitalia provided a basis for assessing the extent of morphological differentiation between the nominal species. We specifically address three main questions: (i) Is genetic variation within North American *Lycaeides* congruent with morphology and/or geographically structured? (ii) Are patterns of mtDNA sequence variation consistent with the hypothesis that fragmentation during the Pleistocene had significant impacts on the process of differentiation within the genus? (iii) Is phylogeographical partitioning congruent with patterns of differentiation observed in behavioural or ecological characters? Taken together, answers to these

© 2005 Blackwell Publishing Ltd, *Molecular Ecology*, 14, 1741–1754

questions will allow us to characterize phylogeographical structure within the species complex and assess the relative impacts of historical and ecological processes on the differentiation of these butterflies.

Materials and methods

Butterfly biology

North American members of the genus *Lycaeides* are relatively small butterflies which primarily use papilionaceous legumes (Fabaceae) as larval host plants. Mark–release–recapture studies of the endangered Karner blue, *Lycaeides melissa samuelis*, indicate that dispersal distances for these relatively sedentary butterflies average less than 500 m, but may occasionally approach 1 km or more (King 1998). Nabokov (1943, 1949) revised the systematics of North American *Lycaeides*, describing *Lycaeides idas* (the northern blue) and *Lycaeides melissa* (the Melissa blue), on the basis of morphology. While these wing pattern and male genitalic characters formed the basis of differentiating *L. melissa* and *L. idas*, Nabokov (1949) also detailed a situation of morphological intermediacy in the Teton range of Wyoming. Other cases of morphological intermediacy also occur in the Warner Mountains in northeastern California, where *L. melissa* wing phenotypes occur in conjunction with *L. idas*-like male genitalia (Scott 1986; Nice & Shapiro 1999). The extinct Lotis blue (*Lycaeides idas lotis*), which formerly occupied boggy habitats along the Mendocino Coast in northern California (Arnold 1993), also possessed the *L. idas* wing phenotype and *L. melissa*-like male genitalia (Scott 1986).

Because initial investigations indicated that genetic differentiation may not be congruent with current classification (Nice & Shapiro 1999), use of Latin binomials does not imply that we believe they are correctly applied to reciprocally monophyletic taxa. We will treat the North American members of *Lycaeides* as a species complex (see Results) and refer to populations by locality name and number (Table 1), for example Concord, NH (1).

Sample collection

We obtained butterflies from 60 *Lycaeides* populations (Table 1) including both fresh and papered specimens. *Lycaeides idas* and *L. melissa* were sampled from much of North America and *L. idas* were also obtained from Europe and Asia. Samples of Asian endemic *Lycaeides argyrognomon* were obtained for use as an outgroup. Both males and females were collected except for populations of the endangered Karner blue, from which only males were collected under permit (USFWS permit PRT842392). A map of collection localities is included in the supplementary material.

Mitochondrial DNA

Data on the distribution of different mitochondrial DNA (mtDNA) AT-rich region haplotypes within and among populations were obtained using single strand conformation polymorphism (SSCP) analysis (Orita *et al*. 1989). Previous studies have demonstrated that the AT-rich region is a valuable phylogeographical marker for Lepidoptera and other insects (Vila & Björklund 2004). SSCP was used to rapidly assess AT-rich region variation and minimize sequencing required in large sample sets (e.g. Anthony *et al*. 1995; Sunnucks *et al*. 2000).

DNA was isolated following methods of Hillis *et al*. (1996) and Brookes *et al*. (1997). Initially, the AT-rich region was amplified from 20 individuals using primer 12S and primer Met20- from Taylor *et al*. (1993). These fragments were cloned into the pCR 2.1 Vector (TA Cloning Kit, Invitrogen) and sequenced (see succeeding section). From these sequences, primer F4 (5′-CATGATAATCCTATAATGTTC-3′) was designed. The F4 primer site is approximately 300 bp from the Met20- site and included the most variable part of the AT-rich region. Primers Met20- and F4 were then used to amplify partial AT-rich region fragments for all individuals.

The resulting partial AT-rich region fragments were digested with *Pac*I endonuclease into two smaller products of approximately 129 bp and 175 bp. Digested products were denatured and run on high resolution polyacrylamide gels prepared with FMC mutation detection enhancement (MDE) gel and buffered with 20 mM HEPES (Liu & Sommer 1998). Novel banding patterns were detected by running test samples against controls of known sequence. Once a novel banding pattern was detected, the underlying nucleotide polymorphisms were confirmed by sequencing. An additional 51 individuals representing a cross-section of all variants detected were sequenced to ensure the accuracy of haplotype identification by SSCP. Amplified PCR products were cloned into pCR 2.1 vectors and three to five clones from each individual were sequenced. Sequences were analysed with an ABI 373 DNA sequencer, and aligned using Lasergene99 (DNASTAR Inc.) and by eye. The 236 bp region used for analysis excluded an apparently invariant segment, near the Met20- primer, containing a homopolymeric region that could not be unambiguously sequenced.

Data analysis

To assess population genetic structure within North American *Lycaeides* and objectively group populations based on genetic similarity, ARLEQUIN version 2.0 software (Schneider *et al*. 2000) was used to calculate pairwise *F*-statistic estimators (ϕ-statistics) (Excoffier *et al*. 1992) between populations using Tamura and Nei distances (with maximum likelihood γ shape parameter estimate = 0.258, calculated using PAUP* version 4.0b10 (Swofford 2002)).

© 2005 Blackwell Publishing Ltd,

Table 1 Population data. Populations are grouped according to their nominal taxonomic designation. Taxa designated as 'AG' use cultivated alfalfa as larval host. Populations designated as '? ALPINE' are an unnamed subspecies of *L. melissa* occurring above treeline in the Sierra Nevada of California. *h* = unbiased haplotype diversity ± SE (Nei 1987). Regional groups are determined from clustering analysis (see text, Fig. 3)

Nominal taxonomic designation	Population	*N*	mtDNA haplotypes (number of individuals)	*h*	Regional group	Number of males measured
L. m. samuelis	1 Concord, NH	14	P(13), E(1)	0.143 ± 0.12	Eastern	5
	2 Saratoga, NY	30	P(30)	0	Eastern	25
	3 Pine Bush/Albany, NY	8	P(8)	0		2
	4 Allegan, MI	1	Q(1)			6
	5 Indiana Dunes IN	23	P(23)	0	Eastern	23
	6 Winona, MN	1	A(1)			2
	7 Fish Lake, WI	22	A(22)	0	Central	28
	8 Eau Claire, WI	22	A(21), D(1)	0.091 ± 0.081	Central	22
	9 Black River, WI	18	A(18)	0	Central	18
	10 Sandhill, WI	21	A(21)	0	Central	21
	11 Welch/Hartman,WI	10	A(9), C(1)	0.200 ± 0.15	Central	10
	12 Fort McCoy, WI	25	A(25)	0	Central	25
	13 Necedah, WI	25	A(25)	0	Central	27
	14 Wood Co., WI	1	A(1)			6
	15 Adams, WI	4	A(4)	0		12
L. m. melissa	16 Range, MN	18	A(9), D(9)	0.529 ± 0.04	Central	5
	17 Clay Co., MN	2	A(1), G(1)	1.00 ± 0.50		6
L. m. melissa AG	18 Spring Creek, SD	29	A(4), D(5), F(1), G(19)	0.539 ± 0.09	Central	7
	19 Ward Mt., NV	2	D(2)			6
	20 Montague, CA	6	A(6)	0		6
	21 Gazelle, CA	5	A(3), D(2)	0.600 ± 0.18		15
	22 Sierra Valley CA	5	A(5)	0		13
	23 Verdi, NV	2	J(2)			10
	24 Gardnerville, NV	22	D(2), G(1), J(19)	0.255 ± 0.12	Western	26
	25 Boyd Farm, CA	1	A(1)			12
L. m. ? ALPINE	26 Mount Rose, NV	20	K(17), N(3)	0.268 ± 0.11	Western	24
	27 Carson Pass, CA	20	K (20)	0	Western	36
	28 The Nipple, CA	5	K(5)	0		3
	29 County Line, CA	15	K(12), L(3)	0.343 ± 0.13	Western	19
	30 Jeff Davis Peak, CA	—	Not genotyped	—		12
L. m. psuedosamuelis	31 Pitkin, CO	4	A(4)	0		2
L. m. annetta	32 Alta, UT	18	A(16), D(2)	0.209 ± 0.12	Central	15
L. m. inyoensis	33 Big Pine, CA	5	A(5)	0		5
	34 Garner Valley, CA	2	A(2)			18
	35 San Diego Mts., CA	21	A(21)	0	Central	16
L. i. nabokovi	36 Waubee Lake, WI	23	A(3), D(1), K(15), L(4)	0.549 ± 0.10		22
	37 Marinette, WI	3	A(2), G(1)	0.667 ± 0.31		4
L. i. ricei	38 Mount Ashland, OR	19	A(18), D(1)	0.105 ± 0.09	Central	32
	39 Marble Mts., CA	19	H(15), K(4)	0.351 ± 0.11	Western	15
	40 Deadfall, CA	13	K(7), L(6)	0.539 ± 0.06	Western	9
L. i. azureus	41 Indian Valley, CA	13	K(6), L(7)	0.538 ± 0.06	Western	15
L. i. anna	42 Shovel Creek, CA	9	K(9)	0		9
	43 Trap Creek, CA	17	K(16), L(1)	0.118 ± 0.10	Western	9
	44 Yuba Gap, CA	20	K (20)	0	Western	25
	45 Leek Springs, CA	13	K(2), L(11)	0.282 ± 0.14	Western	21
L. i. alaskensis	46 Healy, AK	3	F(1), K(2)	0.667 ± 0.31		3
	47 Walker River, AK	1	K(1)			1
	48 Deadhorse, AK	1	A(1)			1
	49 Nebesna, AK	3	A(1), D(1), I(1)	1.00 ± 0.27		3
	50 Tok, AK	1	A(1)			1
	51 Dawson, YT	2	G(2)			2
	52 Whitehorse, YT	4	K(4)	0		4
L. i. aster	53 Newfoundland, NF	1	G(1)			1
L. i. degener	54 Lerida, Spain	3	K(2), M(1)	0.667 ± 0.31		6
L. idas	55 Shushenskoye, Russia	1	G(1)			1
	56 Slovakia	—	Not genotyped	—		1
Warner Mts. Entity	57 Cedar Pass, CA	5	D(3), J(2)	0.600 ± 0.18		4
	58 Eagle Peak, CA	25	A(23), B(2)	0.153 ± 0.09	Central	30
	59 Emerson Peak, CA	—	Not genotyped	—		20
L. argyrognomon	60 Shushenskoye, Russia	2	G(1), O(1)	0.667 ± 0.31		
	Total =	628			Total =	727

These values were calculated for the 27 population samples with 10 or more individuals and were used to construct a UPGMA dendrogram using PHYLIP (Felsenstein 1993). Pairwise ϕ-statistics were also employed in a second analysis using the minimum evolution method (Rzhetsky & Nei 1992). A heuristic search using PAUP* (Swofford 2002) was implemented with starting trees created from random stepwise addition (25 replicates), tree-bisection–reconnection (TBR) branch-swapping and the MulTrees option. Bootstrap values for UPGMA and minimum evolution trees were not obtained because PHYLIP and PAUP* do not calculate ϕ-statistics. Spatial analysis of molecular variance (SAMOVA) was performed using data from these 27 populations to identify groups of populations that are phylogeographically homogeneous using the software SAMOVA 1.0 (Dupanloup *et al*. 2002). SAMOVA finds the partitioning of populations that maximizes the ϕ_{CT} value when a particular number of groups are specified (Dupanloup *et al*. 2002). SAMOVA can be used to identify the most likely number of groups within the data set from repeated analyses specifying the number of groups and choosing the partitioning of populations that maximizes the ϕ_{CT} value (Dupanloup *et al*. 2002). We performed these repeated analyses for 2 to 10 groups. This analysis served as an independent confirmation of the clustering observed using pairwise ϕ-statistics.

To estimate basal divergence time, we conducted Bayesian coalescent analysis of sequence data under a bifurcating population model, using BATWING version 1.02 (Wilson *et al*. 2003) and MCMCCOAL (Rannala & Yang 2003). BATWING was used to estimate posterior probabilities for alternative ancestral haplotypes, θ ($\theta = 2N\mu$, where μ is equal to the mutation rate per generation and N is the effective population size) and time of basal population divergence ($\boldsymbol{T} = \delta/N$, where δ is the time in generations). Joint estimates of θ and $\boldsymbol{T}$ were used to calculate τ (time of divergence multiplied by mutation rate). BATWING accommodates differences in relative subpopulation size, allows for exponential population growth, and can integrate across alternative potential population trees. BATWING provides a restricted choice of mutational models. A stepwise model was used for sequence evolution, approximating a nucleotide substitution model when few mutations have occurred (e.g. Walsh 2001). For inferring ancestral haplotype probabilites, a unique event polymorphism model was employed. Truncated uniform priors were used for θ, population growth rate, and times of population divergence, while the default Dirichlet (2, 2, … , 2) prior was used for relative subpopulation size. This analysis was performed on the subset of 27 populations containing more than 10 sampled individuals. As an alternative method, the Bayesian program MCMCCOAL was also used to estimate time of basal divergence (specified in terms of τ). MCMCCOAL uses a nucleotide substitution model for sequence evolution, conditional on a given population tree. In this case, the UPGMA tree recovered from ϕ-statistics was specified. MCMCCOAL requires priors for θ for all current and ancestral populations, and priors for branch lengths in the population tree. Priors were generated via an empirical Bayes approach. An identical prior for θ was used for each population. The two parameters of this gamma prior were obtained by calculating the mean and variance of Watterson (1975) across all sampled populations. A gamma prior for branch length was likewise obtained from the mean and variance for branch lengths in a UPGMA population average pairwise difference tree. BATWING and MCMCCOAL produce Markov chain Monte Carlo samples from the posterior probability distribution. Posterior means were used as parameter estimates.

The arthropod mtDNA mutation rate estimate of Brower (1994) of 1.17% per million years (Myr) and the *Papilio* (Lepidoptera: Papilionidae) COI + COII gene estimate of 0.39–0.51% per Myr (Zakharov *et al*. 2004) were used to calculate divergence time. To compare the mutation rate of AT-rich sequences to that of mtDNA cytochrome oxidase subunit I (COI) sequences, *Lycaeides* data from the AT-rich region was compared to the smaller COI data set of Nice & Shapiro (1999). Specifically, resampling without replacement was performed from each data set to generate 100 reduced data sets, identical in populations included and sample sizes for both gene regions. Watterson's θ (1975) per site and π (nucleotide diversity) were estimated for each data set and the mean was calculated across resamplings. For COI, $\theta = 0.0152 \pm 0.0047$ and $\pi = 0.0163 \pm 0.0086$ and for AT-rich region sequences, $\theta = 0.0170 \pm 0.0060$ and $\pi = 0.0232 \pm 0.0127$. Mutation rates appear to be comparable ($\mu = \theta/2\,N$, with N necessarily identical for the two mtDNA segments) supporting the use of the rate estimates of Zakharov *et al*. (2004) and Brower (1994) for the AT-rich region. Moreover, the findings of Vila & Björklund (2004) further confirm the similarity of nucleotide substitution rates between the AT-rich region and COI in butterflies. In sampling across nine species of *Erebia* and four species of *Coenonympha*, they found average interspecific distances within the two genera that were very similar for the AT-rich region and COI (average interspecific distance AT-rich region/COI: 0.087/0.067 in *Erebia* and 0.119/0.083 in *Coenonympha*) (Vila & Björklund 2004). In addition, extensive sampling within two species, *Erebia triaria* and *Erebia palarica*, provided similar values of θ for the AT-rich region and COI within each species ($\theta_{\text{per nucleotide}}$ AT-rich region/COI: 0.0047/0.0049 for *Erebia triaria* and 0.0018/0.0019 for *Erebia palarica*) (Vila & Björklund 2004).

We employed analyses of molecular variance (AMOVA) (Excoffier *et al*. 1992) to determine whether mtDNA variation was distributed according to current taxonomic boundaries within *Lycaeides*, or, alternatively, whether genetic variance is partitioned geographically as might be the case if morphospecies are not monophyletic groups. If morphospecies are independent evolutionary lineages, most of the

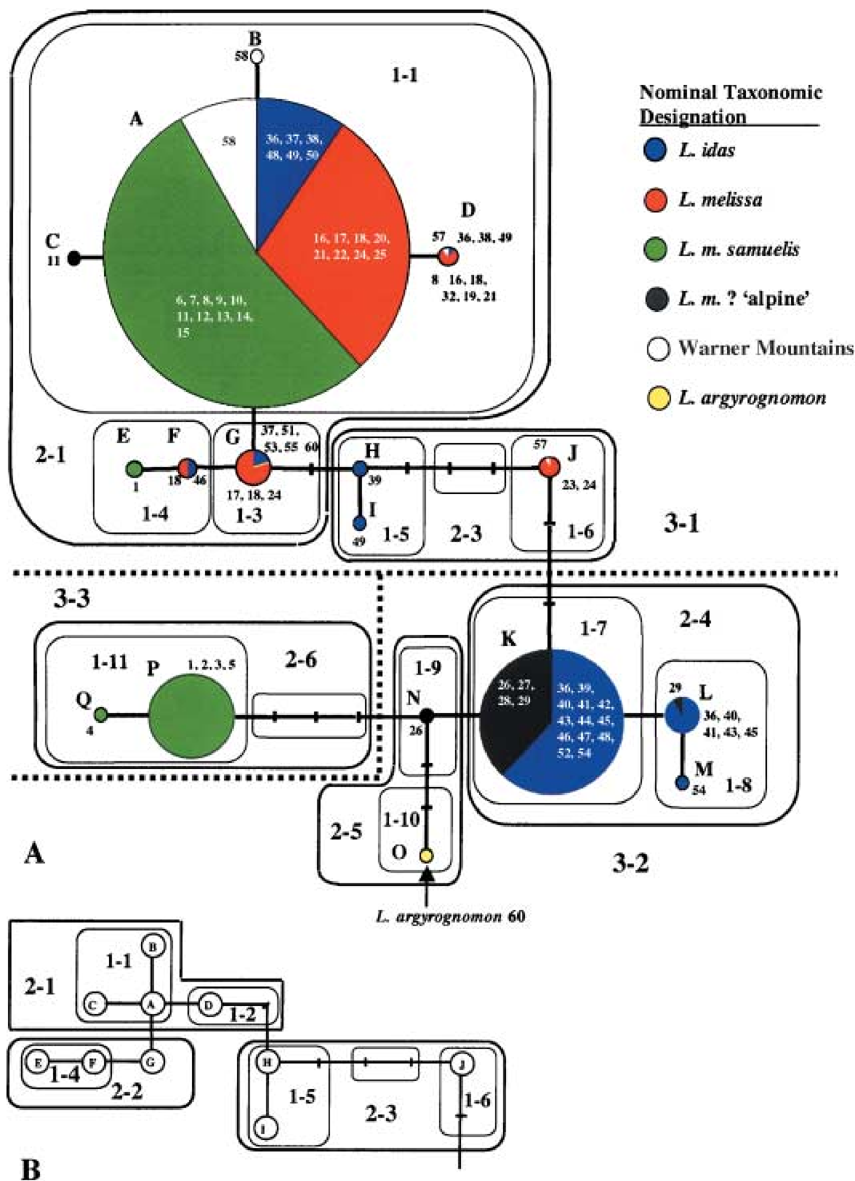

Fig. 2 The two equally parsimonious nested haplotype networks used for nested clade analysis. (A) The full nested network for configuration A. Each haplotype is represented by a circle. The size of each circle is proportional to the frequency of each haplotype in the entire data set. Each haplotype is divided into regions representing the proportion of individuals classified as *Lycaeides melissa*, *Lycaeides idas*, *Lycaeides melissa samuelis*, *L. m.*? 'alpine', or the Warner Mountains entity. Numbers indicate populations that contain that haplotype (see Table 1). (B) The simplified, partial nested network for configuration B. The two configurations differ with respect to connections between haplotypes A, D and H within clade 3-1.

genetic variance within our samples should be partitioned among groups of populations defined by morphology. If mtDNA sequence variation is distributed by geography, we expect to see a larger proportion of the genetic variance attributable to variation among regionally defined groups (i.e. phylogenetic groups with concordant geographical distributions). Nested AMOVA was performed using ARLEQUIN 2.000 (Schneider *et al*. 2000). In the first AMOVA analysis, populations were grouped into three nominal species based on morphology (Nabokov 1949): *L. idas*, *L. melissa*, *L. argyrognomon* (the Asia-endemic species). The Warner Mountains populations (of uncertain morphological affinity) were omitted. In the second AMOVA, populations were arranged into three geographical regions by two alternative methods that proved largely congruent: (1) clustering analyses using pairwise ϕ-statistics (Fig. 3) and (2) SAMOVA. The three regional groups were: (i) the eastern clade (populations east of Lake Michigan) (ii) the central clade (west of Lake Michigan to approximately the California–Nevada border), and (iii) the western clade (populations on the Pacific slope) (Fig. 3). Assignment of populations to each of these three regions based on clustering analyses of pairwise ϕ-statistics is provided in Table 1.

We also employed nested clade analysis (NCA) to test for significant phylogeographical structure. Maximum-parsimony haplotype networks were constructed using TCS 1.13 (Clement *et al*. 2000) which employs the statistical algorithms of Templeton *et al*. (1992). Haplotypes were then grouped into hierarchical clades according to the nesting rules of Templeton *et al*. (1987). Tests of geographical

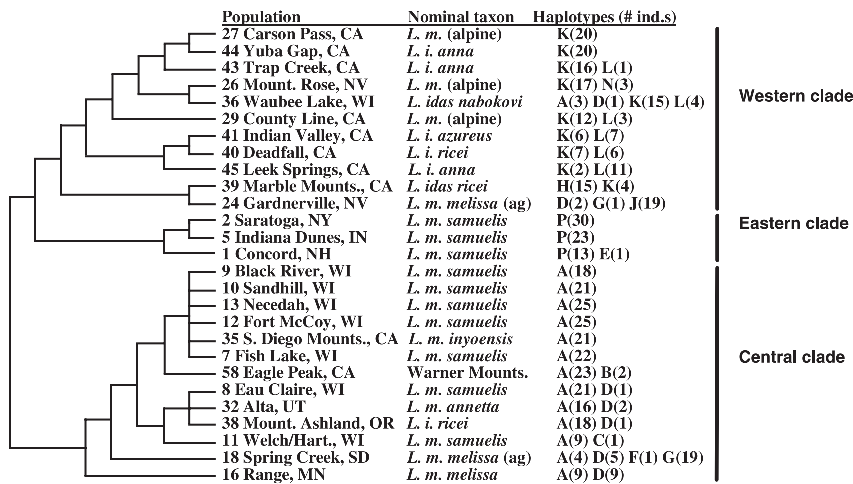

Fig. 3 Genetic relationships among the 27 populations with sample sizes of 10 or more individuals. Populations were clustered using pairwise ϕ_{ST} values using the UPGMA algorithm. Three geographical groups correspond to the distributions of three major clades identified by nested clade analysis (Fig. 2). The western group contains haplotypes from clade 3-2. The central group contains clade 3-1 haplotypes. The eastern group contains clade 3-3 haplotypes. Mixing of haplotypes from two of the three major clades occurs in three populations: Concord, NH (1), Waubee Lake, WI (36) and Marble Mountains, CA (39).

association and permutation tests were performed with GEODIS 2.0 (Posada *et al*. 2000) using 1000 random permutations. Inferences about historical and contemporary population genetic processes were made following the inference key of Templeton (2004).

Two AMOVAs were employed to determine whether mtDNA variation was distributed in accordance with ecological differences among well-studied populations from the Sierra Nevada and adjacent areas. In the first analysis, variance was partitioned among groups defined by female oviposition preference (a measure of host fidelity) using six populations that have been measured for this character (Nice *et al*. 2002; C.C. Nice & J.A. Fordyce, unpublished). Population locality numbers and preferred hosts were (26) *Astragalus whitneyi* (27) *A. whitneyi* (29) *A. whitneyi* (45) *A. whitneyi* (43) *Lotus nevadensis* and (44) *L. nevadensis*. Two alfalfa-feeding populations (23, 24) were excluded from analysis because they exhibit no clear host preference (Nice *et al*. 2002). In the second analysis, variance was partitioned among groups defined by egg adhesion characters (i.e. whether eggs are strongly adhered to host plants or not), a potential adaptation to habitat and/or host plant characteristics (Fordyce & Nice 2003). Eight populations were arranged into two groups using the data of Fordyce & Nice (2003): strong egg adhesion group (locality numbers: 23, 24, 43, 44, 45); weak adhesion group (locality numbers: 26, 27, 29). If mtDNA variation is distributed in accordance with either or both ecological characters, we expect significant among-group components of variance as assessed by permutation tests (1000 permutations).

Genitalic morphology

Morphological variation was quantified for all specimens examined in the mtDNA study plus an additional three populations (30, 56, 59 in Table 1) in order to confirm the nominal taxonomic status of individuals**.** Nabokov (1943, 1949) presented a morphometric approach for discriminating between *L. idas* and *L. melissa* based on five measurements of male genitalia and characteristics of wing patterns. Nice & Shapiro (1999) presented data from 310 male butterflies for these five genitalic measurements. We have added data from another 416 males yielding a total of 726 from 59 populations. Butterflies assigned to species by wing pigment patterns using Nabokov's (1949) characters were classified by discriminant function analysis of genitalic measurements using JMP version 3.15 (Sall and Lehman 1996). From this analysis, an error rate can be calculated that serves as an estimate of the proportion of males in which there is incongruence between wing patterns and genital morphology, or in which genitalic morphology is intermediate or ambiguous.

Results

Mitochondrial DNA

Alignment of partial AT-rich region sequences from *Lycaeides* (GenBank Accession nos AF338926–AF338948) revealed 23 unique haplotypes of which six contained sites at which complex dinucleotide repeat insertion/deletion (indel) events have occurred. Additionally, sequencing of multiple clones from individual butterflies revealed evidence of length heteroplasmy in two individuals. Such length variation in the mtDNA control region has been reported in other insects (Zhang & Hewitt 1997; Howell & Smejkal 2000), and led us to conclude that correct phylogenetic interpretation of variable copy numbers of repeated sequence motifs is difficult. Because of the heteroplasmy in two individuals, the difficulty in coding indel variation for phylogenetic analysis, and the possibility that such variation may be uninformative (Stewart & Baker 1994; Fumagalli *et al*. 1996), we chose to exclude indel variation from our analyses, collapsing our total haplotype number to 17.

Haplotype diversity for each population was calculated using the equation $h = n/n - 1(1 - \Sigma p_i^2)$, where p_i is the frequency of the ith haplotype and n is the sample size (Nei 1987). The majority of *Lycaeides* populations containing 1 or 2 haplotypes (mean number of haplotypes/population = 1.53) (Table 1). Thirty-three of the 57 population samples were fixed for a particular haplotype.

A maximum-parsimony network of the 17 unique haplotypes is presented in Fig. 2. Twenty sites are variable (excluding indels), 14 of which are informative. These 17 haplotypes are organized into three 3-step clades.

Three groups of populations are apparent in the UPGMA population dendrogram constructed from pairwise ϕ-statistics calculated among the 27 populations for which sample sizes were 10 or more individuals (Fig. 3). These groups correspond closely to the distribution of the major clades defined by NCA. The topology of the minimum evolution tree was nearly identical and is not presented. In SAMOVA, ϕ_{CT} asymptotically approached a maximum as the number of groups increased, attaining 95% of this maximal value with three groups. When a larger number of groups were specified, ϕ_{CT} values increased slightly, but groups began to consist of single populations. The SAMOVA specifying three geographical groups identified groups of populations that correspond to the three major clades observed in NCA. These three groups also correspond to the groups observed from the clustering based on pairwise ϕ-statistics (Fig. 3) with one exception: the Marble Mountains (39) population was placed in the central clade rather than the western clade. This difference results from the location of this population near to other populations with central clade haplotypes such as Mount Ashland (38) and from the fact that the Marble Mountain (39) population is one of the few populations containing haplotypes from two major clades. The patterns of population genetic variation (Figs 2 and 3) are surprising because of their incongruence with morphological data. Each major (3-step) clade occupies a large portion of North America with almost no overlap. The central clade (3-1) includes populations distributed from Wisconsin to eastern and southern California to Alaska. A clade 3-1 haplotype was also detected in one *Lycaeides melissa samuelis* population in New Hampshire [Concord (1)] and in *Lycaeides idas* from Newfoundland (53). The western clade (3-2) includes populations in the Sierra Nevada and northern Coast Ranges of California. These 3-2 haplotypes also occur in one *L. idas* population in northern Wisconsin [Waubee Lake (36)] and in Alaska. The eastern clade (3-3) is comprised of *L. m. samuelis* populations east of Lake Michigan. At the level of the genus, haplotypes are shared not only between subspecies but also among *L. idas* from North America, Asia and Europe, *L. melissa* from North America, and *Lycaeides argyrognomon* from Asia (Table 1, Fig. 2). These findings are congruent with Nice & Shapiro's (1999) analysis of sequence variation in *COI*.

Haplotype sharing among nominal species makes *L. argyrognomon* problematic as an outgroup, while AT-rich sequences from other lycaenid taxa resulted in satisfactory alignment of only 120 bp and did not clearly indicate a root (data not presented). Coalescent analysis using BATWING provides an alternative method of rooting and indicated a posterior probability of 90.0% for an ancestral haplotype within clade 2-3.

The basal divergence event among these populations appears to have occurred within a late-Pleistocene timeframe. Coalescent analysis using BATWING produced a posterior mean estimate for τ (mutation-scaled time of basal divergence) of 0.000602 (95% CI: 0.000335, 0.000969). Using a global arthropod mtDNA mutation rate of 1.17% per Myr (Brower 1994) yielded an estimate of 51 000 years for the time of the basal population split. Using an estimate based on the *Papilio* (Lepidoptera: Papilionidae) COI + COII genes of 0.39–0.51% per Myr (Zakharov *et al*. 2004) yielded an estimate of 118 000–154 000 years. The Bayesian program MCMCCOAL provided a second estimate of the time of basal divergence of $\tau = 0.000555$ (95% CI: 0.000284, 0.000911), which translated into an estimate of 47 000 years for the global arthropod mtDNA clock and 109 000–142 000 years for the *Papilio*-based clock. Although this result was somewhat sensitive to the choice of priors, all reasonable priors produced an estimate of under 200 000 years for the global arthropod mtDNA clock and under 600 000 years for the slower *Papilio*-based clock.

The AMOVA results indicate that mtDNA variation is strongly distributed along geographical rather than traditional taxonomic lines with the vast majority of populations being fixed for haplotypes from only one of the major clades. When groups were defined as morphological species, with the Warner Mountains populations excluded, 25.7% of the total variance was accounted for among species ($\phi_{CT} = 0.257$, $P = 0.001$, Table 2A). AMOVA partitions 90% of the overall genetic variation among the three geographical regions shown in Fig. 3 ($\phi_{CT} = 0.904$, $P < 0.0001$, Table 2B). The grouping of populations observed from SAMOVA, which was nearly identical to clustering analysis with the exception of the Marble Mountains (39), explained slightly less of the overall genetic variation ($\phi_{CT} = 0.900$, $P < 0.0001$).

There was no significant partitioning of molecular variance among populations for which we have female oviposition preference and egg adhesion data. When these populations were grouped by preferred host plant, there was no among-group component of variance ($\phi_{CT} = -0.1526$, $P = 0.587$). Similarly, there was no among-group component of variance for populations grouped by egg adhesion characteristics ($\phi_{CT} = -0.0223$, $P = 0.581$).

Populations of the endangered karner blue, *L. m. samuelis*, can be divided into two groups. Populations west of Lake Michigan (populations 6–15) contain haplotype A (part of clade 3-1) that is shared with many *L. melissa* populations

© 2005 Blackwell Publishing Ltd,

A. Source of Variation	d.f.	SSD	Variance component	% of total	*P* value
Among taxa	2	226.4	0.76	25.7	< 0.001
Among populations/within taxa	51	1109.9	1.96	66.4	< 0.001
Within populations	544	127.0	0.23	7.9	< 0.001

B. Source of Variation	d.f.	SSD	Variance component	% of total	*P* value
Among regional groups	2	2231.0	7.00	90.4	< 0.001
Among populations/within groups	24	194.2	0.40	5.1	< 0.001
Within populations	508	178.1	0.35	4.53	< 0.001

Table 2 Results of hierarchical analyses of molecular variance of the AT-rich region mtDNA sequence data. (A) Results grouping populations by morphological species and excluding the Warner Mountains populations. (B) Results grouping populations by region as determined by population clustering analysis (see text, Fig. 3)

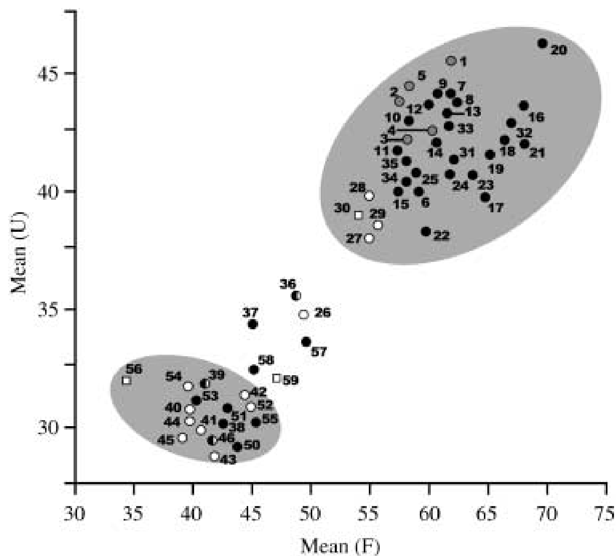

Fig. 4 Male genital morphology graph of population means for two measurements: mean falx length (F) and mean uncus length (U) (units of 10^{-5} m). These two measurements account for 83% of the variation between species. Numbers correspond to the population numbers given in Table. Black circles indicate populations containing haplotypes from clade 3-1. Open circles indicate populations containing clade 3-2 haplotypes. Grey circles indicate populations with clade 3-3 haplotypes. Squares for populations 30, 56 and 59 indicated populations that were not genotyped. The large shaded ovals contain populations unambigously identified as *Lycaeides melissa* and *Lycaeides idas*.

further west and *L. idas* populations in southern Oregon and northern Wisconsin. East of Lake Michigan, *L. m. samuelis* populations (1–5) contain haplotypes in clade 3-3. There is no corresponding distinction in morphological data (Fig. 4), nor are we aware of any corresponding division in ecology or behaviour (Andow *et al*. 1994).

Mixing of haplotypes from these major mtDNA lineages is relatively limited outside of Alaska/Yukon and occurs in only three non-Alaskan population samples: haplotypes H and K in the Marble Mountains (39) *Lycaeides idas ricei*; haplotypes P and E in Concord (1) *L. m. samuelis*; haplotypes from clades 3-1 and 3-2 in Waubee Lake (36) *Lycaeides idas nabokovi*. The presence of haplotypes from both clades 3-1 and 3-2 in the Alaska/Yukon, Waubee Lake, and Marble Mountains populations, in conjunction with a lack of sampling in intervening regions, raises the possibility that an area of substantial haplotype mixing extends throughout unsampled regions of Canada.

Results of NCA as well as the inferences for each clade are provided in the supplementary material. Both nested networks yielded the same results and inferences. Therefore, we only present analysis of network A (Fig. 2A). It should be noted that clade 2-2 exists only in network configuration B and is not present in network A.

Results of NCA are consistent with the hypothesis that the history of *Lycaeides* was strongly influenced by vicariance events that may have been associated with Pleistocene glaciations, followed by post-Pleistocene range expansions. The results from NCA allowed us to infer a chronology of events for this species complex in North America. That is, the age of nested clades increases as the hierarchical level increases. A pattern of past fragmentation is indicated at the highest level of the total cladogram. The inference at this level is that three groups of haplotypes (clades 3-1, 3-2 and 3-3) were fragmented into allopatric groups which we hypothesize were isolated in separate refuges. The results of NCA also indicate range expansion for clades 3-1, 3-3 and 2-1. Within clade 3-2, range expansion is indicated for clade 1-8 after an initial fragmentation event. Thus, the distribution of haplotypes from large portions of North America exhibit a pattern of expansion that has occurred through contiguous range increase (as opposed to long-range colonization) and supports the hypothesis of post-Pleistocene range expansions in *Lycaeides*. Large areas of relatively low genetic diversity conform to the predictions of leptokurtic range expansion (Ibrahim *et al*. 1996).

Within the three major geographical groups of populations, there is no evidence of further partitioning of genetic variance in accordance with the morphological, ecological or behavioural differences observed at a local scale. In the Sierra Nevada, for example, where local differentiation has

been intensively studied, populations of both morphological species (i.e. population 26–30 and 43–45), using several different host plant species and exhibiting habitat and behavioural differences, are nearly uniformly fixed for clade 3-2 haplotypes (Table 1, Fig. 3). There is similar genetic homogeneity within the other two major clades.

Genitalic morphology

We used the discriminant function of Nice & Shapiro (1999) to classify an additional 416 male butterflies. The overall error rate (comparing DFA results with designations based on wing patterns) from the combined 673 males (excluding 54 males from the Warner Mountains) was 2.5%. Multivariate analyses demonstrated that two measurements, falx length (F) and uncus length (U) (Nabokov 1949), accounted for 83% of the variation between species (Nice & Shapiro 1999). Figure 4 is a morphospace diagram representing the major differences among populations. Means of measures F and U are plotted for each population.

Nabokov's (1949) original delineation of *L. idas* and *L. melissa* based on genitalic morphology remains valid with the exception of six populations containing individuals of intermediate or uncertain morphology. Three of these populations are from the Warner Mountains, previously identified as having male genitalic morphology in conflict with wing patterns. The Mount Rose (26) population in the Sierra Nevada has intermediate morphology despite being ecologically indistinguishable from other Sierran *L. melissa* populations (i.e. populations 27–30) (Nice *et al*. 2002). The other populations with indeterminate morphology are of *L. i. nabokovi* in Wisconsin (populations 36, 37). Despite these cases, 97.5% of specimens examined for genitalic morphology were correctly matched to their wing pattern morphology.

Discussion

Phylogeographical and systematic hypotheses

Our extensive sampling design allowed us to examine in detail the distribution of mtDNA variation in North American *Lycaeides*. Little of the mtDNA variation in the genus *Lycaeides* is consistent with morphologically defined taxonomic boundaries (Table 2). Many haplotypes are shared between at least two morphospecies; haplotype G is shared among *Lycaeides idas*, *Lycaeides melissa* and *Lycaeides argyrognomon*. Despite haplotype sharing, large portions of North America contain relatively low genetic diversity and, within much of North America, major mtDNA lineages are confined to specific areas with little overlap. This pattern suggests that historical factors have played an important role in determining the geographical distribution of haplotypes, with coalescent analysis indicating that divergence began approximately 50 000–150 000 years ago.

Climate amelioration, at the end of the Pleistocene, led to rapid range expansions from refuges (Pielou 1991). This rapid increase in range size is hypothesized to have occurred by leading edge expansion by long distance dispersers (Hewitt 2000). Simulation studies show that such leptokurtic dispersal can result in large areas of low genetic diversity (Hewitt 1996, 2000; Ibrahim *et al*. 1996). Data reported here conforms to this pattern and NCA suggests the following history for *Lycaeides*. After an initial colonization of North America (Fig. 5A), three groups (each predominantly associated with one of the 3-step clades) became allopatrically isolated and confined into refuges (Fig. 5B). This period of allopatry was followed by contiguous range expansion, presumably involving recolonization of formerly glaciated

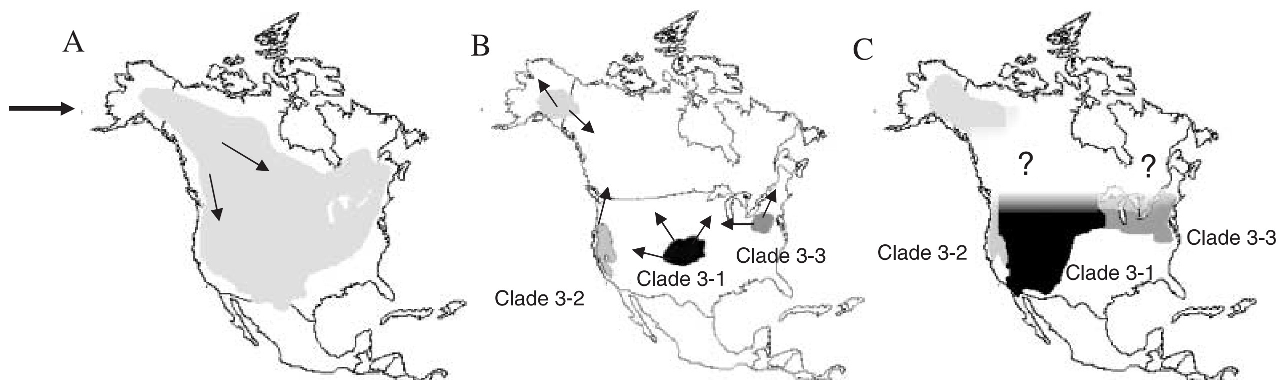

Fig. 5 A hypothetical geographical history of the major mtDNA lineages as inferred from nested clade analysis. (A) The initial colonization of North America and subsequent range expansion. (B) Fragmentation of the range of *Lycaeides* and hypothetical Pleistocene refuges in which lineage sorting occurred. Fragmentation in this manner presumably occurred repeatedly throughout Pleistocene glacial cycles. The Alaskan refuge contained haplotypes from at least clades 3-1 and 3-2. The directions of post-Pleistocene expansions are indicated with arrows. (C) The distribution of major mtDNA lineages following post-Pleistocene range expansion and prior to hypothesizes Holocene-aged introgression. Question mark indicates areas that have not been sampled. Range areas are approximations. See text for details.

or uninhabitable areas from at least three refuges (represented by clades 3-1, 3-2, 3-3; Fig. 5C).

Expansion of the central clade 3-1 may have been augmented recently by the apparent host switch to agricultural alfalfa, *Medicago sativa*, which was introduced 200 years ago and has since spread over large portions of North America (Michaud *et al*. 1988). All of the alfalfa-feeding populations sampled (populations 18–25) contain clade 3-1 haplotypes.

A fourth Beringeal refuge may also have existed in Alaska. The presence of haplotypes from two of the 3-step clades in population samples from Alaska and Yukon Territory suggests that the genetic diversity there is indicative of polymorphism that existed in ancestral populations. *Lycaeides* presumably colonized North America from Asia across a Beringian land bridge. The presence of clade 2-1, 2-4 and 2-5 haplotypes in *L. idas* from Spain and Russia also supports this interpretation. According to the NCA results, this ancestral polymorphism was subsequently partitioned into at least three southern glacial refuges. This interpretation comports with findings from mtDNA sequence variation found in Pleistocene-age brown bears (*Ursus arctos*) preserved in permafrost (Leonard *et al*. 2000; Barnes *et al*. 2002).

It seems likely that post-Pleistocene range expansions have brought formerly isolated clades into contact. If the association of *L. idas* with clade 3-2 seen in California is widespread outside of this region, gene flow might explain the presence of clade 3-1 haplotypes in *L. idas* in Waubee Lake, WI (36), Mount Ashland, OR (38) and the Marble Mountains, CA (39). Range expansion, contact between two clades and subsequent gene flow may also explain the presence of clade 3-2 haplotypes in *L. melissa* populations occurring at high altitudes in the Sierra Nevada of California, and clade 3-1 haplotypes in Wisconsin populations west of the phylogeographical divide at Lake Michigan that separates otherwise indistinguishable populations of *L. m. samuelis*.

In his classic paper, Charles Remington (1968) discussed 'suture zones' as areas of secondary contact between biotas that often include hybridizing taxa of varying systematic rank. Remington supposed that suture zones represented the confluence of post-Pleistocene range expansions similar to those inferred here for *Lycaeides*. Some phylogeographical boundaries in *Lycaeides* correspond to Remington's suture zones. Specifically, the boundary along the eastern edge of the Sierra Nevada corresponds to Remington's Pacific–Rocky Mountain suture zone VI. This area was also identified by Austin & Murphy (1987) as a 'centre of differentiation' where the edges of butterfly species ranges coincide along a biogeographical boundary, and by Swenson & Howard (2004) as a 'hotspot of hybrid zone clustering'. However, several of Remington's proposed suture zones are not evident in the *Lycaeides* data. There is no evidence, for example, for a major suture in the interior of the continent [Remington's (1968) zone IV]. Furthermore, some phylogeographical boundaries in *Lycaeides* do not conform to any of Remington's zones. The boundary at Lake Michigan cannot be easily reconciled with Remington's scheme. However, Austin *et al*. (2002) found a similar boundary in their survey of mtDNA sequence variation in spring peepers, *Pseudacris crucifer*. These authors concluded that this unusual pattern resulted from specific colonization dynamics of postglacial range expansions from multiple refuges including two proposed refuges southeast and southwest of Lake Michigan.

This phylogeographical boundary, separating eastern and western *Lycaeides melissa samuelis*, may indicate introgression of haplotype A into western *L. m. samuelis* populations from *Lycaeides melissa melissa* in Minnesota, displacing the presumed original *L. m. samuelis* haplotype P. Data from nuclear microsatellites and a single copy nuclear gene used to test this hypothesis confirm the genetic integrity of all *L. m. samuelis* populations (N. Anthony and G. Gelembiuk, unpublished data), suggesting that *L. m. samuelis* populations have a history of isolation with mtDNA gene flow occurring only recently.

Hypotheses of introgression among closely related, sympatric or parapatric taxa need to be interpreted carefully. Introgression of mtDNA followed by selective sweeps or lineage sorting can obscure evolutionary histories (e.g. Shaw 2002). Machado & Hey (2003) have noted that mtDNA is unlinked to loci associated with the development of reproductive isolation, and may readily cross species boundaries, complicating phylogeny reconstruction. Furthermore, different loci may exhibit different patterns of divergence in cases of recent differentiation (Machado & Hey 2003). Thus, mtDNA-based phylogenetic or phylogeographical hypotheses must be corroborated by data from multiple loci, especially in cases of recent divergence. In the case of *Lycaeides*, a further note of caution should be expressed regarding the cases of presumed introgression discussed above. These may be instances of ancestral polymorphism, or even combinations of gene flow and ancestral polymorphism. Indeed, in certain cases where suitable neighbouring donor populations are absent, retention of ancestral polymorphism appears to provide the most parsimonious explanation for the co-occurance of deeply diverged haplotypes [e.g. haplotypes P and E in Concord, NH (1); haplotypes D, G, and J in Gardnerville, CA (24)].

Despite the evidence for introgression discussed above, NCA provided no evidence for an inference of gene flow. This appears to result from the paucity of population samples containing haplotypes from more than one clade. Most populations are fixed or nearly so, presumably because population sizes are relatively small, which may be related to the ecological specialization observed in these butterflies. These two aspects of the current data set (i.e. low within-population diversity and few cases of admixture between major clades) appear to limit the ability of NCA to detect gene flow (Templeton 2004). NCA also fails to provide a means for evaluating the degree of statistical confidence

associated with alternative inferences (Knowles & Maddison 2002) and assumes that the observed sequence variation is not structured by selection. This assumption is most relevant when comparisons are made between the distributions of interior clades, which are assumed to be ancestral, and tip clades, which are assumed to be derived. Nevertheless, the results of NCA are supported by the AMOVA and clustering techniques used in this study. Namely, that the pronounced regional structure and large areas of genetic homogeneity are consistent with a model of Pleistocene fragmentation and leading-edge expansion.

The lack of congruence between morphology and molecular variation demonstrated in AMOVAs might represent the product of some ancestral polymorphism, gene flow between *L. idas*, *L. melissa* and *L. m. samuelis* along the contact zones between expanding ranges, and our bias toward sampling in these zones of gene exchange. In other words, excluding suture zones, Alaska, and parts of Canada, molecular and morphological variation is predominantly distributed in three distinct phylogeographical lineages. The large area occupied by clade 3-1 haplotypes in the middle of the continent generally corresponds to the area inhabited by *L. melissa* populations (16–25 and 31–35). Ignoring presumed intergrade populations, clade 3-2 includes *L. idas* populations along the west coast (40–45). Clade 3-3 haplotypes are confined in *L. m. samuelis* populations east of Lake Michigan (1–5), having been recently replaced by 3–1 haplotypes to the west. If *L. m. samuelis* does indeed constitute a distinct evolutionary unit, the evolutionary divergence between *L. m. samuelis* and other *L. melissa* populations is not reflected in the morphology of these butterflies. Alternatively, convergent morphological evolution has occurred in these two lineages. Future investigations of unsampled areas in the Rocky Mountains and much of Canada, and data from nuclear markers will be necessary to confirm these interpretations.

Implications for conservation

Our investigations of *Lycaeides* show that ecologically and evolutionarily distinct lineages may not be reliably diagnosed with a small-scale survey. If our survey of mtDNA variation had been confined to the eight Wisconsin populations of *L. m. samuelis*, we would have failed to detect clade 3-3 haplotypes which are present in other populations of this subspecies east of Lake Michigan.

In cases of rapid and recent divergence where ancestral polymorphism and/or recent gene flow may retard lineage sorting, identifying significant units for conservation may be difficult (Crandall *et al*. 2000). Monophyletic taxonomic groups and concordance among different data sets are expected only after a rather lengthy period of divergence (Neigel & Avise 1986; Shaw 1998; Edwards & Beerli 2000). If divergent selection on ecological and morphological characters is occurring, as appears to be the case in *Lycaeides*, portions of the genome may evolve significantly more rapidly than neutral loci (Orr & Smith 1998; Whitlock 1999). This may be especially true if traits such as wing pigment patterns and male genital morphology are under sexual selection (Eberhard 1996; Ting *et al*. 2000) and contribute to reproductive isolation. Consequently, in these situations, surveys of neutral variation could fail to detect distinct evolutionary lineages in which significant adaptive differences already exist. Differences in habitat use, life history traits and host plant choice between species should be considered to determine the evolutionary significance of differentiating taxa (Crandall *et al*. 2000). In the case of both the Karner blue (*L. m. samuelis*) and Nabokov's blue (*L. i. nabokovi*), ecological non-exchangeability may also provide sufficient justification for their consideration as valid evolutionary significant units under the strategy proposed by Crandall *et al*. (2000).

Acknowledgements

We wish to thank the following people who assisted for this project: J. Fordyce, A.M. Shapiro, C. Carnes, K. O'Brien, M. Boyce, S. Fuller, M. Amaral, T. McCabe, M.S. Bergh, C. Grutzmacher, D. Bowen, T. Rocheleau, R. Knutson, R. King, C. Bleser, J. Maxwell, S. Krauth, D. Young, and A. Swengel. The Museum of Entomology at the University of Wisconsin and the Bohart Museum of Entomology at the University of California, Davis kindly provided specimens. Comments from three anonymous reviewers greatly improved this manuscript.

Supplementary material

The following material is available from http://www.blackwellpublishing.com/

Figure S1. Sampling localities for the survey of mtDNA sequence variation. The key to population numbers is Table 1 Haplotypes contained within populations are designated by major mtDNA lineage identified by their nested clade number following Fig. 2(A).

Figure S2. Alignment of 23 unique haplotypes from mtDNA AT-rich region sequences from *Lycaeides* showing variable sites only. Gray boxes contain insertion/deletion variation that was excluded from analyses (see text). As a result of this exclusion, haplotypes A1, A2 and A3 become identical to haplotype A and were collapsed (i.e. renamed 'A' in Table 1), D1 was collapsed to D, G1 was collapsed to G and N1 was collapsed to N leaving 17 unique haplotypes (A through Q) when indel variation was omitted. Total counts and frequencies of haplotypes in the entire data set are provided in the last two columns, respectively. Numbers in brackets indicate counts and frequencies of collapsed haplotypes.

Figure S3. Results of nested clade analysis for network A (Fig. 2A). Clade numbers correspond to Fig. 2(A). Tip clades are shaded. Inference chains for each higher level clade are reported in the cells immediately to the left of the clade numbers. Results from analysis of network configuration B (Fig. 2B) did not substantially differ from those of network A.

References

Andow DA, Baker RJ, Lane CP (1994) *Karner Blue Butterfly: A symbol of a Vanishing Landscape*. Minnesota Agricultural Experiment Station, University of Minnesota, St Paul.

Anthony NM, Brown JK, Markham PG, ffrench-Constant RH (1995) Molecular analysis of cyclodiene resistance associated mutations among populations of the sweetpotato whitefly *Bemisia tabaci*. *Pesticide Biochemistry and Physiology*, **51**, 220–228.

Arnold RA (1993) The Lotis Blue, *Lycaeides idas lotis* (Lintner). In: *Conservation Biology of the Lycaenidae (Butterflies)* (ed. New TR), pp. 143–144. Occasional Paper of the IUCN Species Survival Commission no. 8.

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

Austin GT, Murphy DD (1987) Zoogeography of Great Basin butterflies: patterns of distribution and differentiation. *Great Basin Naturalist*, **47**, 186–201.

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

Avise JC, Walker D (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 457–463.

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science*, **295**, 2267–2270.

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.

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

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.

Eberhard WG (1996) *Female Control: Sexual Selection by Cryptic Female Choice*. Princeton University Press, Princeton, New Jersey.

Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

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.

Felsenstein J (1993) *PHYLIP (Phylogeny Inference Package)*, Version 3.5c. Department of Genetics, Washington University, Seattle.

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.

Fumagalli L, Taberlet P, Favre L, Hausser J (1996) Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Molecular Biology and Evolution*, **13**, 31–46.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature (London)*, **405**, 907–913.

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

Howell N, Smejkal CB (2000) Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction. *American Journal of Human Genetics*, **66**, 1589–1598.

Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation by different forms of dispersal during range expansion. *Heredity*, **77**, 282–291.

Irschick DJ, Shaffer HB (1997) The polytypic species revisited: Morphological differentiation among tiger salamanders (*Ambystoma tigrinum*) (Amphibia: Caudata). *Herpetologica*, **53**, 30–49.

King RS (1998) Dispersal of karner blue butterflies (*Lycaeides melissa samuelis* Nabokov) at Necedah National Wildlife Refuge. *Transactions of the Wisconsin Academy of Sciences, Arts and Letters*, **86**, 101–110.

Klicka J, Zink RM (1999) Pleistocene effects on North American songbird evolution. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **266**, 695–700.

Knowles LL (2000) Test 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.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Layberry RA, Hall PW, Lafontaine JD (1998) *The Butterflies of Canada*. University of Toronto Press, Toronto.

Leonard JA, Wayne RK, Cooper A (2000) Population genetics of ice age brown bears. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 1651–1654.

Liu Q, Sommer SS (1998) The SSCP phenomenon: addition of HEPES buffer dramatically affects electrophoretic mobility. *BioTechniques*, **25**, 50–56.

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.

Michaud R, Lehman WF, Rumbaugh MD (1988) World distribution and historical developments. In: *Alfalfa and Alfalfa Improvement, Agronomy Monograph No. 29* (eds Hanson AA, Barnes DK, Hill RR), pp. 25–56. ASA-CSSA-SSSA, Madison, Wisconsin.

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

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

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation.

In: *Evolutionary Processes and Theory* (eds Nevo E, Karlin S). Academic Press, New York.

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

Opler PA (1992) *A field guide to eastern butterflies*. Houghton Mifflin Co., Boston, New York, London.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single strand conformation polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 2766–2770.

Orr MR, Smith TB (1998) Ecology and speciation. *Trends in Ecology & Evolution*, **13**, 502–506.

Pielou EC (1991) *After the Ice Age*. University of Chicago Press, Chicago.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic analysis of geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

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

Remington CL (1968) Suture zones of hybrid interaction between recently joined biotas. *Evolutionary Biology*, **2**, 321–428.

Rzhetsky A, Nei M (1992) A simple method for estimating and testing minimum-evolution trees. *Molecular Biology and Evolution*, **9**, 945–967.

Sall J, Lehman A (1996) *JMP Start Statistics. A Guide to Statistical Data Analysis Using JMP and LMP in Software*. Duxbury Press, Wadsworth Publishing Company, London.

Schneider S, Kueffer JM, Roessli D, Excoffier L (2000) *ARLEQUIN version 2.000: A software program for population genetic data analysis*. Genetics and Biometry Laboratory, Department of Anthropology and Ecology, University of Geneva, Switzerland.

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

Shaw KL (1998) Species and the diversity of natural groups. In: *Endless Forms, Species and Speciation* (eds Howard DJ, Berlocher SH). Oxford University Press, New York.

Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16122–16127.

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

Starkey DE, Shaffer HB, Burke RL *et al.* (2003) Molecular systematics, phylogeography, and the effects of Pleistocene glaciation in the painted turtle (*Chrysemys picta*) complex. *Evolution*, **57**, 119–128.

Stewart DJ, Baker AJ (1994) Evolution of mtDNA D-loop sequences and their use in phylogenetic studies of shrews in the subgenus *Otisorex* (Sorex: Siricidae: Insectivora). *Molecular Phylogenetics and Evolution*, **3**, 38–46.

Sunnucks P, Wilson ACC, Beheregaray LB *et al.* (2000) SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Molecular Ecology*, **9**, 1699–1710.

Swenson NG, Howard DJ (2004) Do suture zones exist? *Evolution*, **58**, 2391–2397.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (and other methods)*, Version 4 0b10. Sinauer Associates, Sunderland, Massachusetts.

Taylor MFJ, McKechnie SW, Pierce N, Kreitman M (1993) The lepidopteran mitochondrial control region: structure and evolution. *Molecular Biology and Evolution*, **7**, 1259–1272.

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 789–809.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, **117**, 343–352.

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.

Ting C-T, Tsaur S-C, Wu C-I (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 5313–5316.

Vila M, Björklund M (2004) The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). *Journal of Molecular Evolution*, **58**, 280–290.

Walsh B (2001) Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals. *Genetics*, **158**, 897–912.

Watterson G (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Whitlock MC (1999) Neutral additive genetic variance in a metapopulation. *Genetical Research*, **74**, 215–221.

Wilson IJ, Weale ME, Balding DJ (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **166**, 155–188.

Zakharov EV, Caterino SM, 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.

Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology*, **25**, 99–120.

Chris Nice is Assistant Professor at Texas State University with research interests in evolutionary and ecological genetics. Nicola Anthony is Assistant Professor at the University of New Orleans with research interests in molecular ecology and conservation genetics. Gregory Gelembiuk is a Ph.D. student in Zoology and Botany at University of Wisconsin–Madison, with research interests in ecological genetics and in evolutionary genomics of niche expansion in colonizing species. Denise Raterman is a Ph.D. student at University of California-Riverside studying mammalian fertilization proteins. Professor Ffrench-Constant is interested in the molecular basis of natural selection. His research focuses on insecticide resistance, mimicry and parasitism and uses insects as model systems. Research work includes a population level analysis of DDT resistance in *Drosophila* and positional cloning of wing pattern genes in *Heliconius* butterflies.

This is an electronic version of an article published in Molecular Ecology: complete citation information for the final version of the paper, as published in the print edition of Molecular Ecology, is available on the Blackwell Synergy online delivery service, accessible via the journal's website at http://www.blackwellpublishing.com/mec or http://www.blackwell-synergy.com.