Heredity (2009) 102, 590–599

www.nature.com/hdy

ORIGINAL ARTICLE

Introgression as a likely cause of mtDNA paraphyly in two allopatric skippers (Lepidoptera: Hesperiidae)

EV Zakharov[1], NF Lobo, C Nowak[2] and JJ Hellmann
Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA

Gene transfer between species during interspecific hybridization is a widely accepted reality in plants but is considered a relatively rare phenomenon among animals. Here we describe a unique case of mitochondrial DNA (mtDNA) paraphyly in the skipper genus, *Erynnis*, that involves well-diverged allopatric species. Using molecular evidence from both mitochondrial and nuclear genomes, we found high levels of intraspecific divergence in the mitochondrial genome within *E. propertius* (over 4% pair-wise sequence divergence) but no such differentiation in the nuclear genome. Sequence comparisons with related *Erynnis* suggest that past, but recent and infrequent introgression between *E. propertius* and *E. horatius* is the most reasonable explanation for the observed pattern of mtDNA paraphyly. This example of putative introgression highlights the complexity of mtDNA evolution and suggests that similar processes could be operating in other taxa that have not been extensively sampled. Our observations reinforce the importance of involving multiple genes with different modes of inheritance in the analysis of population history of congeneric taxa.
Heredity (2009) **102,** 590–599; doi:10.1038/hdy.2009.26; published online 18 March 2009

Keywords: allopatric distribution; *Erynnis horatius*; *Erynnis propertius*; organelle capture; introgression; mtDNA

Introduction

For more than two decades, mitochondrial DNA (mtDNA) has been used as a primary marker to reveal geographic patterns in the genetic structure of species (Avise *et al*., 1987). Since then, several authors have pointed out its numerous advantages, including its stable and relatively short circular structure, lack of recombination due to uniparental inheritance and multicopy status in most cells (Avise, 2000). Moreover, due to relatively higher substitution rates in the mitochondrial genome, mtDNA is generally less conserved than many nuclear genes, providing higher resolution for lower level phylogenies and species identification through DNA bar coding (Hebert *et al*., 2003a).

Although the application of organelle DNA in the above-mentioned fields is widely accepted, discrepancies between phylogenies inferred from mtDNA and nuclear markers or morphological traits have repeatedly raised questions concerning the credibility of mtDNA for systematics and phylogeography (Shaw, 2002; Babik *et al*., 2005; Chan and Levin, 2005; Rubinoff and Holland, 2005; Wiemers and Fiedler, 2007). Owing to the lack of recombination (but see Slate and Phua, 2003; Rokas *et al*., 2003), mtDNA genes are inherited as a single linkage group, but the mitochondrial genome does not necessarily reflect the entirety of a species' evolution or population history (Ballard and Whitlock, 2004). Examples for such exceptions include paralog gene evolution due to the appearance of nuclear pseudogenes with mitochondrial origin (Bensasson *et al*., 2001; Thalmann *et al*., 2004); the finding that mtDNA may not evolve in a pattern consistent with a strictly neutral equilibrium model because mtDNA variation influences organismal fitness (Ballard and Rand, 2005); and many species have mtDNA lineages that are phylogenetically intermixed with other species (Wahlberg *et al*., 2003; Peters *et al*., 2007).

In a literature survey, Funk and Omland (2003) documented that 23% of 2319 assayed animal species had mtDNA that was not monophyletic. Given this prevalence they concluded that the occurrence of species with non-monophyletic mtDNA is statistically supported, taxonomically widespread and far more common than generally recognized. The authors listed a number of possible explanations including inadequate phylogenetic information, poor taxonomy or inaccurate species limits, interspecific hybridization, incomplete lineage sorting (ILS) and unrecognized parology.

Gene transfer between species during interspecific hybridization is a widely accepted reality in plants but is considered relatively rare among animal species (Dowling and Secor, 1997). Yet, speciation is often a slow

Correspondence: Dr JJ Hellmann, Department of Biological Sciences, 107 Galvin Life Sciences Building, University of Notre Dame, Notre Dame, IN 46556, USA.
E-mail: hellmann.3@nd.edu
[1]*Current address: Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada N1G 2W1.*
[2]*Current address: Department of Limnology and Conservation, Research Institute Senckenberg, Gelnhausen, Germany.*
Received 16 September 2008; revised 15 December 2008; accepted 13 February 2009; published online 18 March 2009

evolutionary process and diverging species may remain in genetic contact for millions of years (Prager and Wilson, 1975; Avise and Walker, 1998). This creates an opportunity for introgressive hybridization. Such hybridization is not simply an interesting phenomenon that occasionally creates reticular pattern of genome evolution, but it can be a mechanism for interspecific transfer of adaptive traits (Scriber and Ording, 2005). Hybridization can be important in evolution by widening a species' niche (Schwarz *et al*., 2005). For example, hybrid genotypes sometimes show enhanced fitness (hybrid vigor) in novel environments as in the cases of hybrid speciation in some butterflies (Scriber and Ording, 2005; Gompert *et al*., 2006; but see Haldane, 1922).

In this study, we describe a case of mtDNA paraphyly in a genus of skipper butterflies (*Erynnis*) that involves well-diverged allopatric species. We explain the possible source, cause and mechanism that likely contributed to the origin of the observed paraphyly. Using molecular evidence from both mitochondrial and nuclear genomes, we conclude that past, but recent and infrequent introgression is the most likely explanation. This example of putative introgression in *Erynnis* highlights the complexity of mtDNA evolution and suggests that similar processes could be operating in other taxa that have not been extensively sampled.

Materials and methods

The genus *Erynnis* Schrank, 1801 is composed of a large number of species, totaling 17 taxa in the United States and Canada (Pelham, 2008). In this study we follow *Erynnis* classification given by Burns (1964), whose revision was based on morphological, geographical and ecological examination of ~11 000 adults of New World *Erynnis* species. For further details on the classification and species delimitations of Burns (1964), see the online Supplementary materials.

We previously reported divergent patterns of mtDNA in *Erynnis propertius* that were identified in a phylogeographic survey of the species across its range in western North America (Zakharov and Hellmann, 2008). The main set of 527 specimens was collected in California, Oregon, Washington and southern Canada (Figures 1 and 2). In this initial data set, 28 mtDNA haplotypes were identified with one haplotype being highly diverged from the rest (see Results; Zakharov and Hellmann, 2008). To investigate this outlying haplotype further, we gathered 85 additional specimens including nine species of *Erynnis* and two out-group taxa, *Thorybes pylades* and *Epargyreus clarus* (for the full list of specimens see the Supplementary Appendix). For the additional specimens, we sequenced the same 894 bp fragment of the mtDNA gene, *ND5*, that was used in the previous study (Zakharov and Hellmann, 2008). Sequences for *E. montanus* and two additional out-group species (*Pyrgus communis* and *Hylephila phyleus*) were downloaded from GenBank. For selected individuals of *E. propertius* (see Results), nucleotide sequences of two other mtDNA genes (cytochrome oxidase I and II; COI and COII) were produced for comparison. Where possible, nucleotide sequence data were also obtained for a 403 bp fragment of a nuclear gene, *wingless* (*wg*). In total, the complete data set included sequences for 105 individuals from ten *Erynnis* species and four out-groups.

Total genomic DNA was extracted using a Qiagen DNeasy tissue kit. Material of varying quality was used for DNA extraction, ranging from freshly caught samples preserved in alcohol to dried specimens from museum and amateur collections. Although we attempted to recover DNA sequence for the nuclear gene from every specimen, four of them (DNA samples 377, 858, 789 and 791, see Supplementary Appendix) failed to amplify resulting in a reduced data set.

Polymerase chain reactions (PCR) were performed in an Eppendorf Mastercycler thermocycler using the following temperature profile: 95 °C for 2 min, followed by 35 cycles of 94 °C for 60 s, 49 °C for 45 s and 72 °C for 60 s, with the final extension at 72 °C for 7 min. Each reaction was run using Taq DNA Polymerase in Storage Buffer B kit (M1661; Promega Inc., Madison, WI, USA) following the manufacturer's recommendations. PCR products were visualized by electrophoresis in 2% agarose gel stained with GelRed (no. 41003; Biotium Inc., Haywood, CA, USA) and subsequently cleaned using AMPure PCR purification kit (Agencourt Inc., Beverly, MA, USA). Purified PCR products were directly sequenced using Applied Biosystems BigDye terminator cycle sequencing (ABI, Foster City, CA, USA) as per the manufacturer's recommendations. Fluorescently labeled sequencing products were purified using Agencourt CleanSEQ system and fractionated on an ABI 3730 automated sequencer. All fragments were sequenced in both directions using the same primers that were used for PCR (Table 1). Sequences were assembled into contiguous arrays using Sequencher, version 4.6 (GeneCode Corp., Ann Arbor, MI, USA). Clear double peaks in chromatograms of *wg* sequences were attributed to the presence of heterozygote individuals. Those polymorphic states were coded using the conventional format established by the International Union of Pure and Applied Chemistry (IUPAC) format.

Initial analysis of sequence similarity was performed using a neighbor-joining clustering algorithm with uncorrected *p*-distances calculated from pair-wise sequence comparisons in PAUP* 4.10b (Swofford, 1998). Further, phylogenetic relationships among sequenced taxa and individuals were inferred by a Markov chain Monte Carlo (MCMC) procedure implemented in MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001). Specifically, we used the substitution model with the best fit to the data chosen with Akaike information criterion (Akaike, 1974) in Modeltest version 3.7 (Nylander, 2004).

Bayesian analyses included two separate concurrent MCMC runs, each composed of four chains, three heated and one cold (see Huelsenbeck and Ronquist, 2001). Each Markov chain was started from a random tree and run for up to 5×10^6 generations, sampling the chains every 100th cycle. After discarding burn-in samples (Huelsenbeck and Bollback, 2001), the data were used to generate a majority-rule consensus tree where the percentage of samples recovering any particular clade of the consensus tree represented the clade's posterior probability (Huelsenbeck and Ronquist, 2001). For additional details on Bayesian analysis and testing of alternative topologies using Bayes factor, see Supplementary materials.

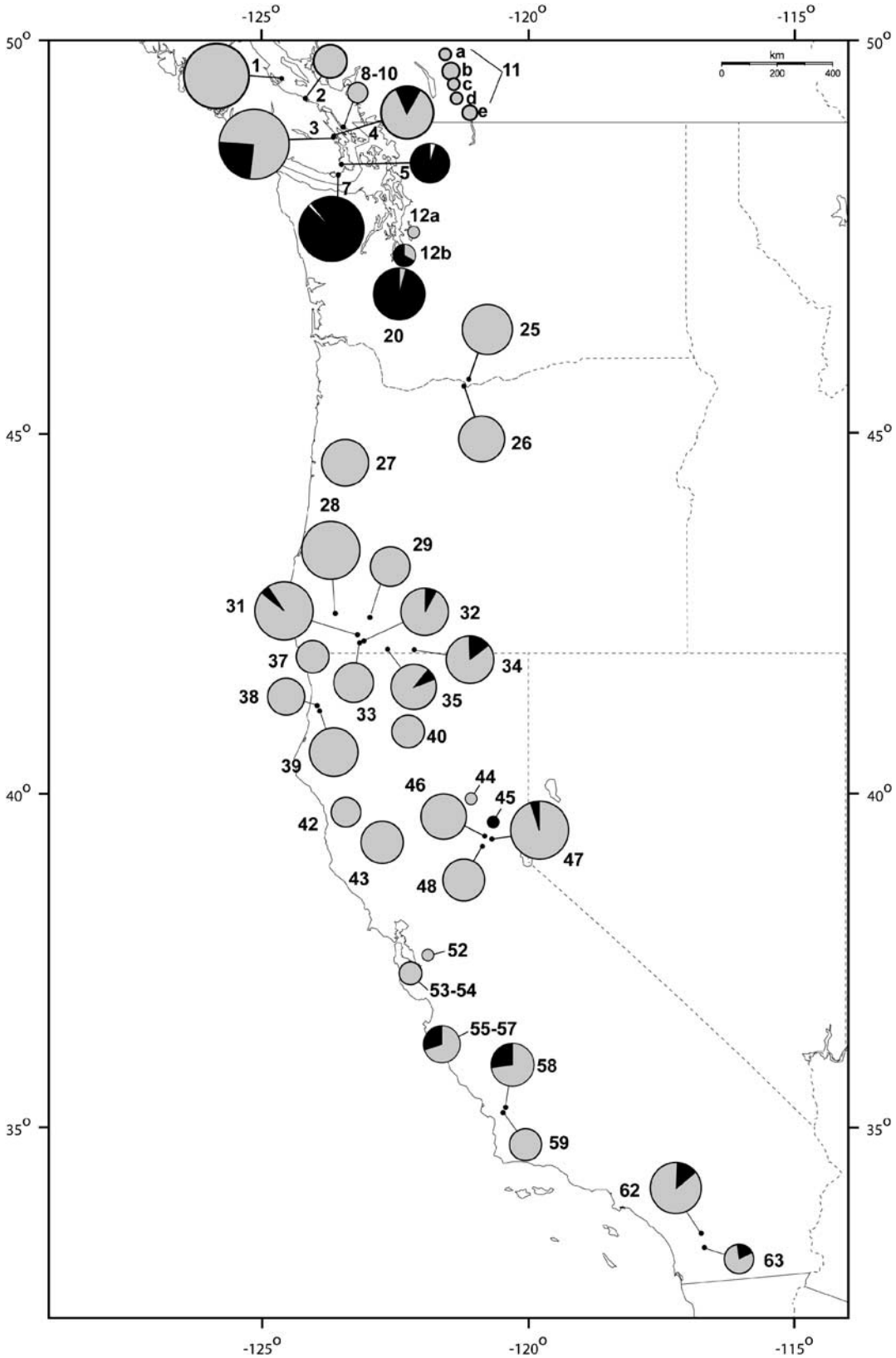

Figure 1 Distribution of *Erynnis propertius* and *E. horatius* mtDNA haplogroups in populations of *E. propertius*. Each sampled population is represented by a pie chart displaying the relative abundance of two haplogroups observed for sequenced fragments of the *ND5* gene. The *E. propertius* haplogroup (shown as gray) includes 27 closely related haplotypes with less than 0.5% sequence divergence. The *E. horatius* haplogroup (shown as black) is represented by a single haplotype that has about 5% sequence divergence from other *E. propertius* haplotypes. Numbers next to the pie charts indicate sampling locality as described previously (see Zakharov and Hellmann, 2008). The size of the pie chart is proportional to the sample size. For example, locality 52 refers to a single specimen, locality 3—42 specimens. Adapted from Zakharov and Hellmann (2008).

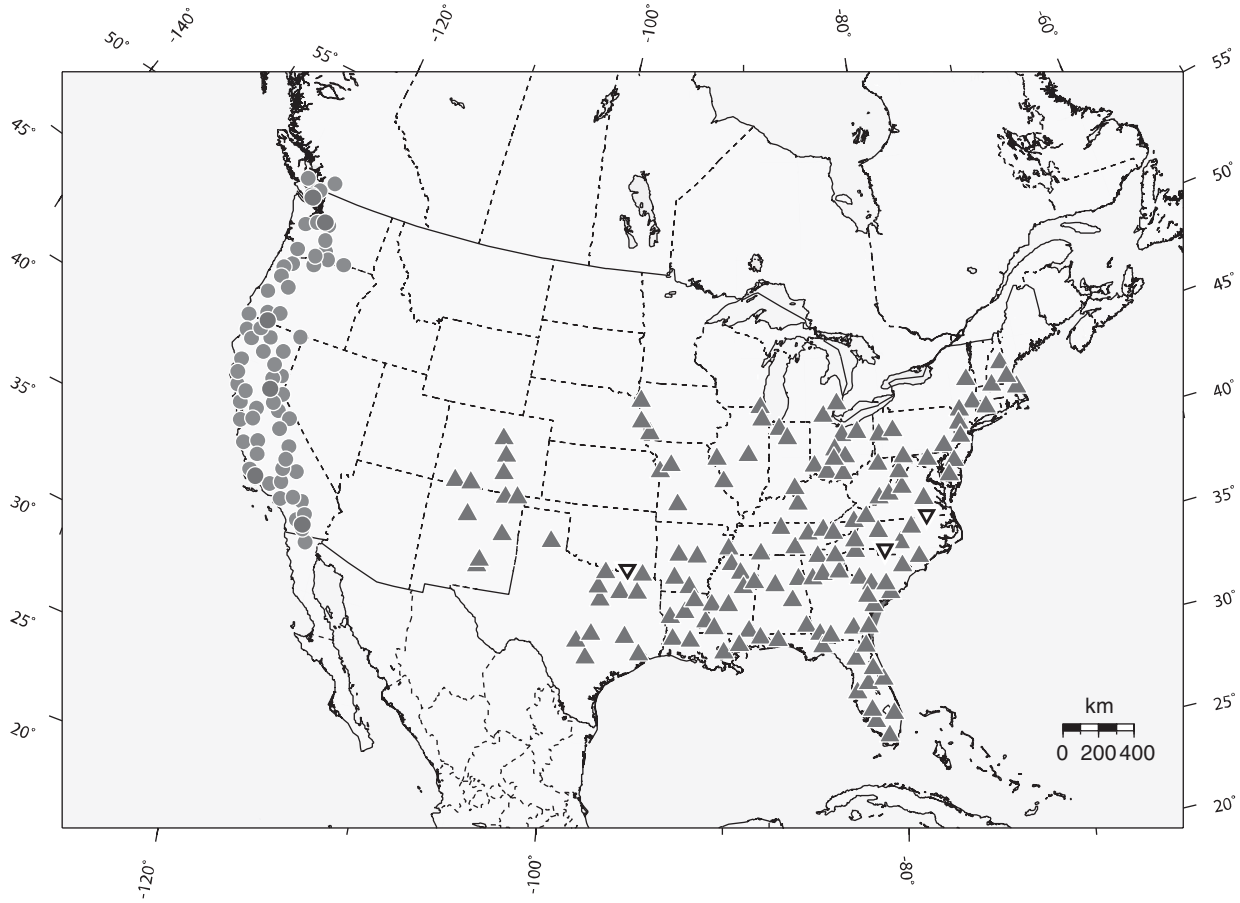

Figure 2 Allopatric distribution of *Erynnis propertius* (grey or blue circles) and *E. horatius* (grey or red triangles). Black open triangles indicate geographic origin of analyzed specimens of *E. horatius*. Dark grey or red circles indicate populations of *E. propertius* that carry mtDNA of *E. horatius*. The map was generated using online service available at http://www.aquarius.ifm-geomar.de and edited in Adobe Illustrator CS3. Range information was combined from Burns (1964) and Opler *et al*. (2006). The color reproduction of this figure is available on the html full text version of the manuscript.

Table 1 List of primers used in this study

Position in reference sequence[a]	*Gene*	*Primer*	*Source*	*Direction*	*Sequence (5′–3′)*
2183	*COI*	Jerry	Simon *et al*. (1994)	Forward	CAACATTTATTTTGATTTTTTGG
2578	*COI*	K741	Simon *et al*. (1994)	Reverse	TGGAAATGTGCAACTACATAATA
3038	*COII*	Patrick	Simon *et al*. (1994)	Forward	CTAATATGGCAGATTATATGTAATGGA
3782	*COII*	Eva	Simon *et al*. (1994)	Reverse	GAGACCATTACTTGCTTTCAGTCATCT
6659	*ND5*	ND5-A1	Morinaka *et al*. (2000)	Forward	AATATDAGGTATAAATCATAT
7080	*ND5*	ND5-3F	Simon *et al*. (1994)	Forward	ATCYTTWGAATAAAATCCAGC
7172	*ND5*	ND5-7R	Morinaka *et al*. (2000)	Reverse	ATAARTGATAWTCARGATATT
7553	*ND5*	ND5-3R	Simon *et al*. (1994)	Reverse	TAACTAAAAGWGCWCAAATTCC
—	*Wg*	LepWg1	Brower and DeSalle (1998)	Forward	GARTGYAARTGYCAYGGYATGTCTGG
—	*Wg*	ModLepWg2	Brower and DeSalle (1998)	Reverse	ACTICGCARCACCARTGGAATGTRCA

[a]Position relative to *Drosophila yakuba* mtDNA sequence (Clary and Wolstenholme, 1985).

To test if shared polymorphism observed in mtDNA between *E. propertius* and its congener(s) is not reflected in the nuclear genome, we analyzed microsatellite allele frequencies and distribution among three groups of individuals: *E. propertius* with the regular mitotype, *E. propertius* with the divergent mitotype and *E. horatius*, the species with the mtDNA haplotype most similar to the divergent form of *E. propertius* (see below). Microsatellite genotyping for available individuals of *E. horatius* was carried out using 14 highly variable loci identified for *E. propertius* (Zakharov *et al*., 2007). To differentiate clusters of genetically related individuals or populations using multi-locus microsatellite data, we used a Bayesian clustering algorithm implemented in the program Structure version 2.2 (Pritchard *et al*., 2000) as described previously in Zakharov and Hellmann (2008).

Results

All sequence data were easily aligned by eye due to lack of indels. For the *ND5* gene, low-quality sequence bases were trimmed from the ends of the alignment resulting

Table 2 Intra- and interspecific divergence based on uncorrected *p*-distances calculated for *ND5* data

	Species	*Number of sequences compared*	*Maximum intraspecific divergence, %*	*Distance covered by sampling*	*Average interspecific distances (below diagonal) with corresponding minimum values (above diagonal), %*											
					1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1	*E. propertius*	397[a]	0.58	2000	—	4.94	4.82	4.82	4.64	7.76	5.41	7.73	6.70	7.76	6.98	7.52
2	*ND5* haplotype EPRP	130[a]	0.00	2000	5.07	—	**0.00**	3.88	5.77	7.52	5.52	7.92[a]	7.52	8.58	7.28	7.52
3	*E. horatius*	5	0.24	1800	5.13	**0.07**	—	3.76	5.62	7.40	5.41	7.76	7.40	8.46	7.42	7.40
4	*E. tristis*	35	0.35	450	5.22	3.92	3.96	—	5.18	6.58	5.29	8.49	7.64	7.99	6.95	7.76
5	*E. pacuvius*	13	0.12	550	5.18	6.11	6.13	5.88	—	6.34	4.71	8.48	6.35	6.89	5.94	5.05
6	*E. funeralis*	3	0.00	—	4.33	6.56	6.63	6.40	5.98	—	5.76	9.55	8.11	9.05	5.94	5.03
7	*E. persius*	18	4.47	3000	6.09	5.72	5.75	5.20	5.41	6.21	—	6.84	6.58	7.40	6.42	7.34
8	*E. martialis*	1	—	—	8.33	8.28	8.23	8.75	8.76	9.40	7.36	—	9.93	9.64	7.54	9.09
9	*E. icelus*	11	0.59	5500	6.99	7.50	7.53	7.61	6.48	7.45	6.54	9.97	—	4.23	5.88	5.64
10	*E. brizo*	3	0.00	0	8.12	8.58	8.62	8.00	7.35	8.70	7.58	9.40	4.65	—	5.94	5.05
11	*E. montanus*	2	0.26	—	8.16	8.06	8.18	7.65	7.45	8.36	7.61	8.50	6.28	6.03	—	4.02
12	*E. tages*	2	0.00	0	7.78	7.52	7.55	7.78	7.33	8.09	7.57	8.95	5.92	5.05	4.82	—

[a]Sequence data set was reconstructed from haplotype sequences and number of those haplotype occurrences.
Values shown in bold illustrate high similarity of the *E. propertius* divergent haplotype with *E. horatius* sequences.

in a concatenated sequence data set of 851 bp that contained 563 constant, 87 noninformative and 201 parsimony informative characters, with 3.34% of missing data. The *wg* data set included 277 invariant characters, 43 variable noninformative and 83 parsimony informative characters. Twenty-seven sequences contained double peaks in at least one position and those were treated as polymorphic states as described in the Materials and methods. The proportion of missing/ambiguous data was 4.96%. Additional information on data matrix parameters is given in online Supplementary materials.

Comparison of 28 described mtDNA haplotypes of *E. propertius* indicates the presence of two distinct genetic lineages (Zakharov and Hellmann, 2008). One lineage is represented by a group of 27 haplotypes with less than 0.6% sequence divergence among them and 0.27% average uncorrected *p*-distance (Table 2). Another lineage is completely invariant across the whole range of *E. propertius* and constitutes a single haplotype that, on average, is 5.1% divergent from the first lineage with uncorrected pair-wise distance ranging from 4.9 to 5.3%. This haplotype was the second most commonly observed haplotype out of all screened individuals of *E. propertius*. Nucleotide sequences obtained for fragments of other mtDNA genes, COI and COII (GenBank accession numbers FJ041310–FJ041317), indicate a nearly identical amount of divergence and geographic patterns of haplotype distribution as revealed by the *ND5* gene (Supplementary materials). This suggests the presence of a highly different mitotype within *E. propertius*.

DNA sequence comparison of this divergent haplotype against congeners of *E. propertius* that co-occur with it along the western coast of North America showed *E. tristis* as the nearest related species with about 4% sequence divergence. However, comparison of sequence data with all sampled species from North America placed the divergent haplotype from *E. propertius* as a close match to another, geographically distant species, *E. horatius*, with 99.5–100% sequence similarity. In addition, *E. horatius* was distant from any other analyzed species of *Erynnis* at a minimum of 3.94% sequence divergence. Overall, the average sequence variation for pair-wise species comparisons in our data set ranged from ~4 to ~10% (Table 2).

Phylogenetic relationships based on mtDNA sequence data confirm that every *Erynnis* species investigated forms a monophyletic entity (except for *E. propertius*), as does the genus itself (Figure 3a). Moreover, there is a clear differentiation of two major lineages within *Erynnis* (Figure 3) that correspond to the traditional subdivision of two subgenera within *Erynnis* (Burns, 1964; online Supplementary materials).

Bayesian analysis confirms the affinity of the divergent haplotype of *E. propertius* to the *E. horatius* clade with high posterior probability. Notably, *E. horatius* and *E. propertius* are not each other's closest relatives (that is, sister species); instead, there is a very strong support (0.99 posterior probability) that *E. horatius* is more closely related to *E. tristis* (Figure 3a). Again, this agrees with previous taxonomy (Burns, 1964).

Nucleotide sequence data from the nuclear gene, *wg*, did not fully resolve interspecific relationships (Figure 3b). Nevertheless, individuals of *E. propertius* carrying mtDNA of both types appeared as a weakly supported, but monophyletic clade. *E. tristis* and *E. horatius* retain a close relationship and are represented by a well-supported monophyletic clade with one exception: a specimen of *E. horatius* with incomplete sequence (266 bp from the 3′ end) had basal placement relative to the clade (*E. propertius* (*E. horatius*, *E. tristis*)).

The overall sequence divergence is lower for *wg* data compared to mtDNA most likely due to slower substitution rates in the nuclear genome. Average intraspecific distance was 0.13% for *E. propertius*, 0.46% for *E. tristis* and 0.08% for *E. horatius*. Average interspecific divergence was ~0.5% for *E. tristis* and *E. horatius*, 0.65% for *E. horatius* and *E. propertius*, and 1.16% for *E. propertius* and *E. tristis*.

A Bayesian clustering algorithm implemented in the program Structure differentiated all individuals of *E. propertius* that were genotyped with 17 microsatellite markers based on their geographic origin, regardless of

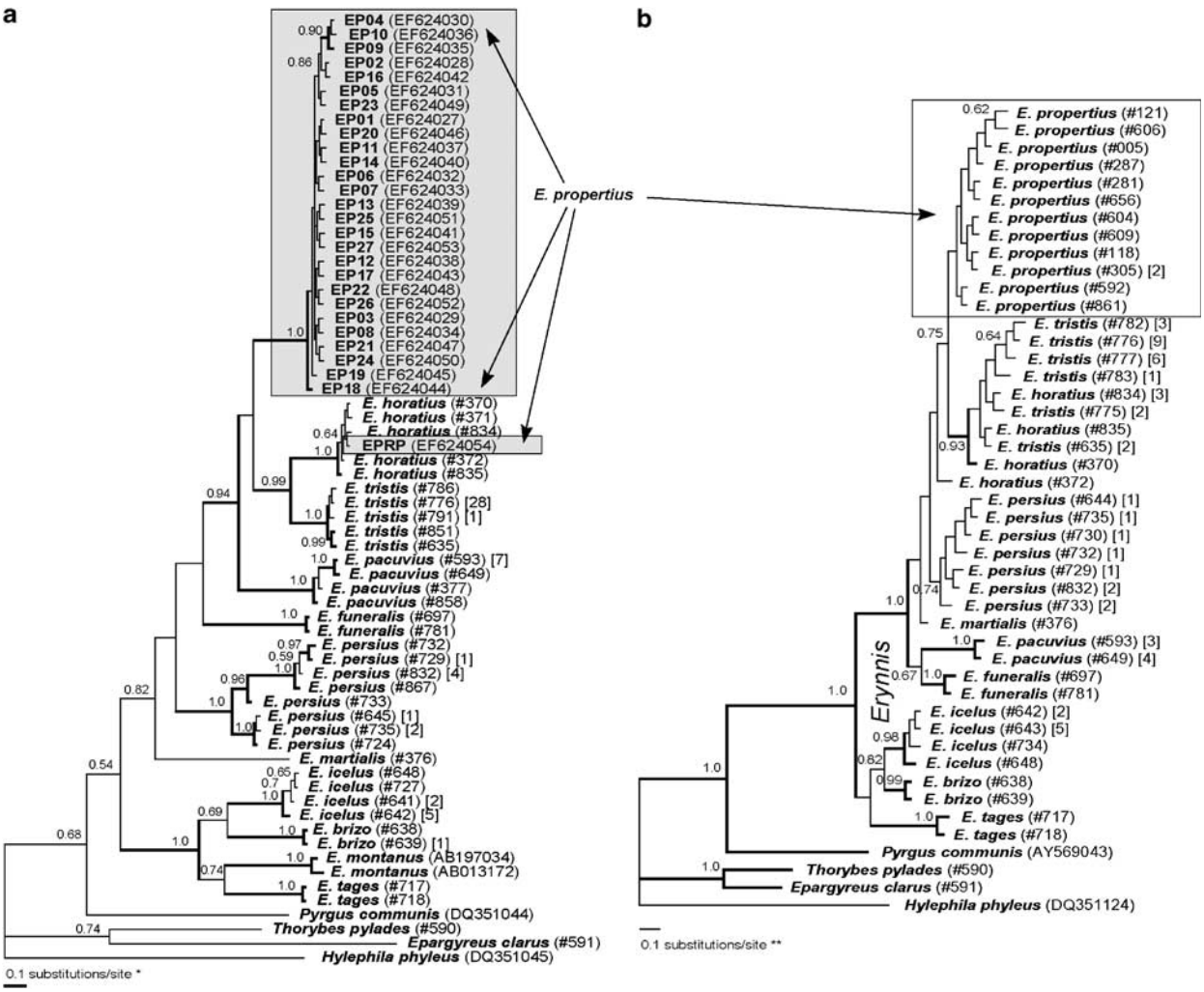

Figure 3 Bayesian phylogeny of *Erynnis* based on nucleotide sequence data. (**a**) An 851 bp fragment of mtDNA gene *ND5*, including 28 haplotypes of *E. propertius* (shown in gray boxes) described in Zakharov and Hellmann (2008). (**b**) A 403 bp fragment of the nuclear gene, *wingless* (*wg*). Numbers in parenthesis refer to sample DNA number (Supplementary Appendix). Numbers in square brackets indicate number of other specimens with identical sequence (see Supplementary Appendix). Numbers above nodes indicate clade posterior probabilities (pp). Nodes with strong support (pp > 0.9) are shown as thick lines. *Note*: There appears to be an error in scale bar on the Bayesian phylogenies. The mean pair-wise path lengths (sum of branch lengths between tips) among species in the trees produced by MrBayes are 6–8 times larger compared to the *D* score values inferred in PAUP* under the same substitutions model. *Tree length, $TL_{MrBayes} = 12.9$; $TL_{PAUP} = 2.03$; **$TL_{MrBayes} = 8.83$; $TL_{PAUP,\ p} = 1.14$.

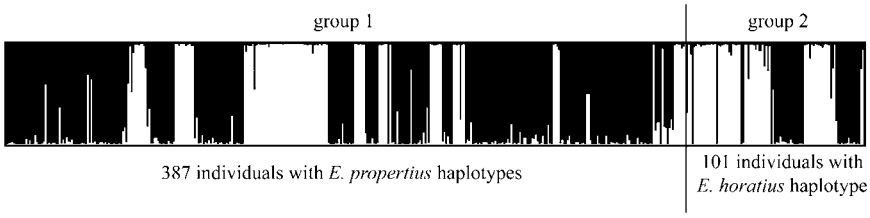

Figure 4 Bar plots of Bayesian assignment probabilities inferred from Structure (Pritchard *et al*., 2000) for $K = 2$ population (species) clusters. Group 1, individuals with *E. propertius* haplotypes; group 2, individuals with *E. horatius haplotype*. The proportion of each line is coded as black or white to represent individual assignment probability to corresponding inferred clusters. Individuals with both mitotypes are clustered according their geographic origin irrespective of the affinities of their mitochondrial genomes. For further details on geographic associations of *E. propertius* haplotypes, see Zakharov and Hellmann (2008).

their mtDNA haplotype (Figure 4; see Zakharov and Hellmann, 2008 for details). Inclusion of data sampled for *E. horatius* in Structure analysis did not affect individual clustering and average assignment probabilities for *E. propertius*. All five individuals of *E. horatius* were assigned to a mainland cluster(s) of *E. propertius* under a range of *K* values (assumed populations) from 3 to 8.

Discussion

Although mtDNA remains a popular marker for analysis of phylogeographic patterns in natural populations, a variety of new evidence suggests that it has complex patterns of variation within and among species. Populations that have been isolated recently are likely to share some of the mtDNA haplotype diversity that existed before their separation. For this reason, newly evolving species may not appear monophyletic with respect to mitochondrial genomes until their mtDNA gene pools have sorted to reciprocal monophyly by lineage sorting. Another concern is the possibility of phylogenetic relationships being confounded by the repeated introgression of mtDNA. To a certain degree, the same pitfalls apply to nuclear genes and because of these complications, many recent studies use a multi-locus, multi-genome or whole-genome approach to avoid biased phylogenetic inference from single-gene data (Begun *et al*., 2007; Nazari *et al*., 2007; Winterton *et al*., 2007; Regier *et al*., 2008). Such a multi-locus strategy does not work for mtDNA as it is inherited as a single linkage group.

Source of mtDNA paraphyly in *Erynnis*

Recent studies estimate the proportion of non-monophyletic mtDNA in animal species at over 20% (Funk and Omland, 2003; Wiemers and Fiedler, 2007). This number may be an overestimate because some cases of apparent paraphyly result from poor taxonomy that has failed to reveal differentiation (Hebert *et al*., 2004; Borisenko *et al*., 2008). Yet, in Lepidoptera specifically, a review of 147 publications reported 31 studies in which genotypic information was perceived to be in conflict with nominal taxonomic boundaries (Forister *et al*., 2008). This suggests that our findings of a complex evolutionary history involving interspecific hybridization and introgression in *Erynnis* may not be unique.

To ensure that interpretation of our results has not been obscured by methodological errors, we (1) critically examined our study material for species identification, (2) used a variety of analysis methods to perform phylogenetic reconstruction and (3) attempted to minimize the impact of limiting sampling where possible.

Aside from methodological error, there are three major biological causes of paraphyly. These are introgressive hybridization among divergent lineages (species), ILS of mtDNA gene trees relative to species trees and selection on mtDNA that might retain ancestral haplotypes within divergent lineages. We can exclude the possibility of retained ancestral polymorphism in our case as the *E. horatius* haplotype appears invariant across the whole range of *E. propertius*. *Erynnis horatius* and *E. propertius* otherwise show high levels of haplotype diversity.

Although a rigorous statistical framework is generally required to distinguish ILS from introgression (Peters *et al*., 2007), three main approaches have been commonly used to delineate the competing hypotheses when the requirements for a statistical test cannot be met (for example, due to limited sample size). First, examination of mtDNA tree topology serves as a heuristic method to reveal the cause of paraphyly (Baker *et al*., 2003). Shallow interspecific divergence with low resolution and lack of distinct clades is an indicator of recent species divergence that precluded mtDNA lineages from sorting to reciprocal monophyly. Divergence patterns observed in our study, however, show clear divergence gaps among species of *Erynnis*. Moreover, the species with paraphyletic mtDNA are not each other's immediate relatives. On the basis of this line of evidence, we can conclude that lineage sorting for this group of species is completed.

Second, a comparison of nuclear DNA and mtDNA trees generally supports ILS if similar patterns of shared DNA polymorphism and paraphyly are observed between genomes (Ballard, 2000). Although our data from the nuclear gene *wg* demonstrate some shared polymorphism between *E. horatius* and *E. tristis*, these do not involve *E. propertius*. Estimation of Bayes factors to differentiate between lack of resolution (too low substitution rates) and possible ILS in the *E. horatius–E. tristis* clade indicated no support for either hypothesis. It appears that due to much lower substitution rates, *wg* provides little resolution among closely related species (see below). Regardless, the nuclear sequence data for *E. propertius* appear monophyletic whereas the mtDNA are not. Thus, ILS between *E. propertius* and *E. horatius* is not supported.

Third, hybridization is favored over ILS if a sympatric distribution provides opportunities for interspecific mating (Donnelly *et al*., 2004). Allopatry does not refute the possibility of introgression, especially if the species are capable of active dispersal or species ranges came in contact. A scenario of past population contact between *E. propertius* and *E. horatius* is plausible, particularly in the southwestern United States where their current distributions come closest to contact (Figure 2). However, the two species have clearly disjunct ranges with little, recent gene flow as evidenced by their high amount of sequence divergence. Lack of any polymorphism within the divergent haplotype of *E. propertius*, as well as no evidence for interspecific transfer of nuclear genes between *E. propertius* and *E. horatius*, suggests organelle capture after the species have evolved apart. Thus, it is likely that during a relatively recent but brief contact, *E. propertius* acquired mtDNA of *E. horatius* through introgressive hybridization.

On the basis of inferred phylogenetic relationships and current distribution of species in the *E. juvenalis* group, to which *E. propertius* and *E. horatius* belong (Supplementary materials), it is likely that the southernmost parts of these two species' ranges were in a secondary contact millions of years ago. It is also interesting that even though the introgressed haplotype was present in the genetically unique populations of *E. propertius* in southern California, it is most widespread in the northern parts of the species' range. It also appears not to have changed by mutation during or after northward expansion. Therefore, it could be argued that adaptive traits associated with the presence of this haplotype enhanced its spread (Andrews *et al*., 1998; Melo-Ferreira *et al*., 2005). A detailed comparative phylogeographic study of *E. horatius* and its other close relatives is required to draw conclusions on the location and approximate time frame of the possible zone of interspecific contact.

Because interspecific gene flow is usually heterogeneous across the genome and introgressive hybridization is often cryptic, questions about its frequency and role in speciation are of substantial importance to evolutionary

biology. In accordance with Haldane's rule, hybrid forms of the heterogametic sex (females in Lepidoptera) are more susceptible to inviability and sterility (Haldane, 1922). Therefore, the introgression of maternally inherited mtDNA in most lepidopteran species should become negligible once postzygotic incompatibility is established and thus the case presented here could be uncommon.

In some cases, however, interspecific relationships and species delimitation using mtDNA in arthropods are complicated by the presence of a symbiotic bacteria, *Wolbachia*, that may influence mtDNA diversity in a number of ways (see Ballard and Rand, 2005). For example, there is at least one well-documented case of introgression of mtDNA between sibling species of Lepidoptera (*Acraea encedana* to *A. encedon*) due to infection by male-killing *Wolbachia* (Jiggins, 2003; Hurst and Jiggins, 2005). Generally, maternally inherited, *Wolbachia*-induced infections tend to carry along the mtDNA genotype that was initially associated with the spread of the infection (Turelli *et al*., 1992; Narita *et al*., 2006). This causes discrepancy between mtDNA-based phylogenies and inferences based on nuclear genes (Dean and Ballard, 2004). *Wolbachia* can directly affect the fitness of infected individuals through increased or reduced fertility and/or longevity (Hariri *et al*., 1998; Snook *et al*., 2000), but it also may protect a selectively disadvantageous mtDNA type from extinction through inducing inviability in offspring from females with a more successful mitotype (de Stordeur, 1997; Ballard and James, 2004).

We have not tested the hypothesis of symbiont association with *E. propertius* as it appears to be an infrequent phenomenon in nature (Jiggins, 2003; Hurst and Jiggins, 2005). Yet, if this mechanism underlies our observations, it implies that symbiotic organisms can overcome the obstacles of Haldane's rule and that, at least in Lepidoptera, interspecific mitochondrial genome transfer could be a common evolutionary process. Further studies will be required to understand patterns of mtDNA evolution in *Erynnis* and its applicability to other animals.

Use of mtDNA in phylogenetics

As demonstrated by past hybridization of *E. propertius* and *E. horatius*, a potential problem of mtDNA is its inability to assign individuals to particular species. Our mtDNA data for many *Erynnis* do, however, reveal species groupings that support previous taxonomic classifications. Hebert *et al*. (2003b) proposed the 3% nucleotide divergence level at the mtDNA COI gene for differentiating species in Lepidoptera. On the basis of this criterion, our mtDNA sequence data support accepted species delimitations for *E. propertius*, *E. horatius* and their congeners with an average interspecific divergence in nucleotide sequences of $\sim$4 to 10%. It has been demonstrated for some butterflies (for example, Lycaenidae) that there is an upper bound of intraspecific divergence in mtDNA of 3.2% for at least 95% of studies species, but there is no lower limit (Wiemers and Fiedler, 2007). This suggests that closely related species can have very similar mtDNA.

Maximum values of intraspecific divergence for *E. horatius* and *E. propertius* also are eight to twenty times less than the minimum values of interspecific divergences between the two species. With the exception of *E. persius*, the amount of intraspecific divergences in *ND5* in all of our sampled species of *Erynnis* is much less than the highest value of intraspecific divergence. *E. persius*, in contrast, is highly polymorphic and is likely to comprise at least two independent species (Forbes, 1936). These results suggest that the mtDNA genome is sufficient to identify species groupings within a large sample for *Erynnis*.

Despite paraphyly in mtDNA, we do not advocate relying solely on nuclear markers. First, slower substitution rates in the commonly used coding regions of the nuclear genome limit the power of these loci to discern relationships among closely related species. Second, direct sequencing of nuclear markers in diploid organisms inevitably leads to complications with inference of haplotypes from polymorphic sequences in heterozygous individuals. Although there are computational methods for estimating individual haplotypes from population data (Clark, 1990; Stephens *et al*., 2001; Stephens and Donnelly, 2003), it is not always possible to fully determine gametic phase without additional experimental data. The solution to the problem is costly and involves isolation of alleles before sequencing through cloning into a vector. The most common alternate approach is using IUPAC format to encode the heterozygous base calls as ambiguous states that are treated as polymorphic characters. This, however, creates another danger for data interpretation because not every program contains algorithms for proper handling of ambiguous data.

Faster-evolving nuclear markers, such as microsatellites, are very useful for population genetic studies within species, but have certain limitations with application across even closely related species. Specifically, our inability to differentiate *E. horatius* from *E. propertius* based on microsatellite data can be related to limited sampling or indicates that primers that were designed for *E. propertius* microsatellites are not effective in *E. horatius*. Although shared versus fixed allelic differences between the two species appear a valid explanation for the lack of *E. horatius* differentiation in Structure analysis, there is another possibility related to the methodological problems with lepidopteran microsatellites. Specifically, it is possible that although the primers developed to amplify microsatellite repeats in *E. propertius* do yield PCR products for *E. horatius*, those fragments may not be homologous and can represent a different tandem repeat within a microsatellite family. This phenomenon is believed to exist in many Lepidoptera (Meglécz *et al*., 2004; Zhang, 2004).

Thus, it appears that in the debate on which DNA marker is better, there is no a simple answer to who is right and wrong. Both mtDNA and nuclear genes have number of advantages and disadvantages, and their utility and success rate are strongly dependent on the area of application. As noted by Zink and Barrowclough (2008), mtDNA 'will prove to be robust indicators of patterns of population history and species limits,' and nuclear genes as rather lagging indicators of changes in population structure 'will prove important for quantitative estimates of the depth of haplotype trees, rates of population growth and values of gene flow.'

Acknowledgements

We thank Arthur M Shapiro (University of California Davis), Cheryl B Barr (Essig Museum of Entomology, University of California Berkeley) and Michael A Wall (San Diego Natural History Museum) for invaluable help with specimen procurement and/or specimen loans and donations; and Jason Dzurisin and Christian Brown for help at the lab bench. We also thank four anonymous reviewers whose comments and suggestions significantly improved the paper. This research was supported by the Office of Science (BER), US Department of Energy, grant nos. DE-FG02-05ER 64023 and DE-FG02-06ER64263, and by the California Energy Commission, CIEE award no. MEX-06-04 to JJH.

References

Akaike H (1974). A new look at the statistical model identification. *IEEE Trans Autom Contr* **19**: 716–723.

Andrews TD, Jermiin LS, Easteal S (1998). Accelerated evolution of cytochrome *b* in simian primates: adaptative evolution in concert with other mitochondrial proteins? *J Mol Evol* **47**: 249–257.

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

Avise JC, Arnold RM, Ball RM, Bermingham E, Lamb T, Neigel JE *et al*. (1987). Instraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* **18**: 489–522.

Avise JC, Walker D (1998). Pleistocene phylogeographic effects on avian populations and the speciation process. *Proc R Soc Lond B Biol Sci* **265**: 457–463.

Babik W, Branicki W, Crnobrnja-Isailovic J, Cogalniceanu D, Sas I, Olgun K *et al*. (2005). Phylogeography of two European newt species—discordance between mtDNA and morphology. *Mol Ecol* **14**: 2475–2491.

Baker JM, López-Medrano E, Navarro-Sigüenza AG, Rojas-Soto OR, Omland KE (2003). Recent speciation in the Orchard Oriole group: divergence of *Icterus spurius spurius* and *Icterus spurius fuertesi*. *Auk* **120**: 848–859.

Ballard JWO (2000). When one is not enough: introgression of mitochondrial DNA in *Drosophila*. *Mol Biol Evol* **17**: 1126–1130.

Ballard JWO, James AC (2004). Differential fitness of mitochondrial DNA in perturbation cage studies correlates with global abundance and population history in *Drosophila simulans*. *Proc R Soc Lond B Biol Sci* **271**: 1197–1201.

Ballard JWO, Rand DM (2005). The population biology of mitochondrial DNA and its phylogenetic implications. *Annu Rev Ecol Syst* **36**: 621–642.

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

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW *et al*. (2007). Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* **5**: e310. doi:10.1371/journal.pbio.0050310.

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001). Mitochondrial pseudogenes: evolution's misplaced witness. *Trends Ecol Evol* **16**: 314–321.

Borisenko AV, Lim BK, Ivanova NV, Hanner RH, Hebert PDN (2008). DNA barcoding in surveys of small mammal communities: a field study in Suriname. *Mol Ecol Resour* **8**: 471–479.

Brower AVZ, DeSalle R (1998). Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of *wingless* as a source of characters for phylogenetic inference. *Insect Mol Biol* **7**: 1–10.

Burns JM (1964). Evolution in Skipper Butterflies of the genus *Erynnis*. *Univ Calif Publ Entomol* **37**: 214pp.

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

Clark AG (1990). Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol* **7**: 111–122.

Clary DO, Wolstenholme DR (1985). The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J Mol Evol* **22**: 252–271.

de Stordeur E (1997). Nonrandom partition of mitochondria in heteroplasmic *Drosophila*. *Heredity* **79**: 615–623.

Dean MD, Ballard JWO (2004). Linking phylogenetics with population genetics to reconstruct the geographic origin of a species. *Mol Phylogenet Evol* **32**: 998–1009.

Donnelly P, Pinto J, Girod R, Besansky NJ, Lehmann T (2004). Revisiting the role of introgression vs shared ancestral polymorphisms as key processes in shaping genetic diversity in the recently separated sibling species of the *Anopheles gambiae* complex. *Heredity* **92**: 61–68.

Dowling TW, Secor CL (1997). The role of hybridization and introgression in the diversification of animals. *Annu Rev Ecol Syst* **28**: 593–619.

Forbes WTM (1936). The persius group of Thanaos (Lepidoptera, Hesperiidae). *Psyche* **43**: 104–113.

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

Funk DJ, Omland KE (2003). Species-level paraphyly and polyphyly: frequency, causes, consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Syst* **34**: 397–423.

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

Haldane JBS (1922). Sex ratio and unisexual sterility in hybrid animals. *J Genet* **12**: 101–109.

Hariri AR, Werren JH, Wilkinson GS (1998). Distribution and reproductive effects of *Wolbachia* in stalk-eyed flies (Diptera: Diopsidae). *Heredity* **81**: 254–260.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a). Biological identification through DNA barcodes. *Proc R Soc Lond B Biol Sci* **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*. *Proc Natl Acad Sci USA* **101**: 14812–14817.

Hebert PDN, Ratnasingham S, deWaard JR (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B* **270**(Suppl): S96–S99.

Huelsenbeck JP, Bollback JP (2001). Empirical and hierarchical Bayesian estimation of ancestral states. *Syst Biol* **50**: 351–366.

Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754–755.

Hurst GDD, Jiggins FM (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc R Soc Lond B Biol Sci* **272**: 1525–1534.

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

Meglécz E, Petenian F, Danchin E, D'Acier AC, Rasplus JY, Faure E (2004). High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Mol Ecol* **13**: 1693–1700.

Melo-Ferreira J, Boursot P, Suchentrunk F, Alves PC (2005). Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into

three other hare species in northern Iberia. *Mol Ecol* **14**: 2459–2464.
Morinaka S, Minaka N, Sekiguchi M, Erniwati E, Prijono SN, Ginarsa IK *et al*. (2000). Molecular phylogeny of birdwing butterflies of the tribe Troidini (Lepidoptera: Papilionidae)—using all species of the genus *Ornithoptera*. *Biogeography* **2**: 103–111.
Narita S, Numura M, Kato Y, Fukatsu T (2006). Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. *Mol Ecol* **15**: 1095–1108.
Nazari V, Zakharov EV, Sperling FAH (2007). Phylogeny, historical biogeography, and taxonomic ranking of Parnassiinae (Lepidoptera, Papilionidae) based on morphology and seven genes. *Mol Phylogenet Evol* **42**: 131–156.
Nylander JAA (2004). *MrModeltest v2. Program distributed by the author*. Evolutionary Biology Centre: Uppsala University.
Opler PA, Pavulaan H, Stanford RE, Pogue M coordinators (2006). Butterflies and moths of North America. Bozeman, MT: Mountain Prairie Information Node.http://www.butterfliesandmoths.org/.
Pelham JP (2008). A catalogue of the Butterflies of the United States and Canada. *J Res Lepid* **40**: xiv +658 pp.
Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland K (2007). Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). *Evolution* **61**: 1992–2006.
Prager EM, Wilson AC (1975). Slow evolutionary loss of the potential for interspecific hybridization in birds: a manifestation of slow regulatory evolution. *Proc Natl Acad Sci USA* **72**: 200–204.
Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
Regier JC, Cook CP, Mitter C, Hussey A (2008). A phylogenetic study of the bombycoid complex (Lepidoptera) using five protein-coding nuclear genes, with comments on the problem of macrolepidopteran phylogeny. *Syst Entomol* **33**: 175–189.
Rokas A, Ladoukakis E, Zouros E (2003). Animal mitochondrial DNA recombination revisited. *Trends Ecol Evol* **18**: 411–417.
Rubinoff D, Holland BS (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst Biol* **54**: 952–961.
Schwarz D, Matta BM, Shakir-Botteri NL, McPheron BA (2005). Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* **436**: 546–549.
Scriber JM, Ording GJ (2005). Ecological speciation without host plant specialization; possible origins of a recently described cryptic *Papilio* species. *Entomol Exp Appl* **115**: 247–263.
Shaw KL (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proc Natl Acad Sc USA* **99**: 16122–16127.
Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved PCR primers. *Ann Entomol Soc Am* **87**: 651–701.
Slate J, Phua SH (2003). Patterns of linkage disequilibrium in mitochondrial DNA of 16 ruminant populations. *Mol Ecol* **12**: 597–608.
Snook RR, Cleland SY, Wolfner MF, Karr TL (2000). Offsetting effects of *Wolbachia* infection and heat shock on sperm production in *Drosophila simulans*: analyses of fecundity, fertility and accessory gland proteins. *Genetics* **155**: 167–178.
Stephens M, Donnelly P (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* **73**: 1162–1169.
Stephens M, Smith NJ, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**: 978–989.
Swofford DL (1998). *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods)*. Sinauer Associates: Sunderland, MA.
Thalmann O, Hebler J, Poinar HN, Pääbo S, Vigilant L (2004). Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Mol Ecol* **13**: 321–335.
Turelli M, Hoffmann AA, Mckechnie W (1992). Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics* **132**: 713–723.
Wahlberg N, Oliveira R, Scott J (2003). Phylogenetic relationships of *Phyciodes* butterfly species (Lepidoptera: Nymphalidae): complex mtDNA variation and species delimitations. *Syst Entomol* **28**: 257–273.
Wiemers M, Fiedler K (2007). Does the DNA barcoding gap exist?—A case study in blue butterflies (Lepidoptera: Lycaenidae). *Front Zool* **4**: 8.
Winterton SL, Wiegmann BM, Schlinger EI (2007). Phylogeny and Bayesian divergence time estimations of small-headed flies (Diptera: Acroceridae) using multiple molecular markers. *Mol Phylogenet Evol* **43**: 808–832.
Zakharov EV, Hellmann JJ (2008). Genetic differentiation across a latitudinal gradient in two co-occurring butterfly species: revealing population differences in a context of climate change. *Mol Ecol* **17**: 189–208.
Zakharov EV, Hellmann JJ, Romero-Severson J (2007). Microsatellite loci in the *Propertius* duskywing, *Erynnis propertius* (Lepidoptera: Hesperiidae), and related species. *Mol Ecol Notes* **7**: 266–268.
Zhang DX (2004). Lepidopteran microsatellite DNA: redundant but promising. *Trends Ecol Evol* **19**: 507–509.
Zink RM, Barrowclough GF (2008). Mitochondrial DNA under siege in avian phylogeography. *Mol Ecol* **17**: 2107–2121.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)