

Considering evolutionary processes in the use of single-locus genetic data for conservation, with examples from the Lepidoptera

Matthew L. Forister · Chris C. Nice ·
James A. Fordyce · Zachariah Gompert ·
Arthur M. Shapiro

Received: 9 October 2006 / Accepted: 4 December 2006 / Published online: 1 February 2007
© Springer Science+Business Media B.V. 2007

Abstract The increasing popularity of molecular taxonomy will undoubtedly have a major impact on the practice of conservation biology. The appeal of such approaches is undeniable since they will clearly be an asset in rapid biological assessments of poorly known taxa or unexplored areas, and for discovery of cryptic biodiversity. However, as an approach for diagnosing units for conservation, some caution is warranted. The essential issue is that mitochondrial DNA variation is unlikely to be causally related to, and thus correlated with, ecologically important components of fitness. This is true for DNA barcoding, molecular taxonomy in general, or any technique that relies on variation at a single, presumed neutral locus. Given that natural selection operates on a time

scale that is often much more rapid than the rates of mutation and allele frequency changes due to genetic drift, neutral genetic variation at a single locus can be a poor predictor of adaptive variation within or among species. Furthermore, reticulate processes, such as introgressive hybridization, may also constrain the utility of molecular taxonomy to accurately detect significant units for conservation. A survey of published genetic data from the Lepidoptera indicates that these problems may be more prevalent than previously suspected. Molecular approaches must be used with caution for conservation genetics which is best accomplished using large sample sizes over extensive geography in addition to data from multiple loci.

Keywords DNA barcoding · Molecular taxonomy · COI · Lineage sorting · Hybridization

Matthew L. Forister, Chris C. Nice and James A. Fordyce contributed equally to this paper.

M. L. Forister (✉)
Department of Natural Resources and Environmental
Science, University of Nevada, Mail Stop 186, 1000 Valley
Road, Reno, NV 89512, USA
e-mail: mforister@cabnr.unr.edu

C. C. Nice · Z. Gompert
Department of Biology, Population and Conservation
Biology Program, Texas State University, San Marcos, TX
78666, USA

J. A. Fordyce
Department of Ecology and Evolutionary Biology,
University of Tennessee, Knoxville, TN 37996, USA

A. M. Shapiro
Section of Evolution and Ecology, Center for Population
Biology, Storer Hall, University of California, Davis, CA
95616, USA

Introduction

A fundamental challenge for conservation biology is to accurately identify units of biodiversity and to diagnose those units that merit conservation concern (Vane-Wright et al. 1991; Moritz 1994, 2002; Vogler and DeSalle 1994; Bowen 1998, 1999; Haig 1998; Grady and Quattro 1999; Paetkau 1999; Crandall et al. 2000; Goldstein et al. 2000; Fraser and Bernatchez 2001; Agapow et al. 2004). Identification or circumscription of taxa consists of two different, yet equally important, aspects. First, identification can refer to the accurate demarcation of a taxon previously described, such as a species or population with a known conservation status. Second, identification can include the discovery of

a new taxon and determination of whether it merits conservation attention. Traditionally, identification has been the responsibility of the taxonomist, who would often combine morphological, behavioral, life history, ecological, and genetic data to determine taxon identity. However, the number of taxonomic specialists, those with an intimate knowledge of a group of organisms, is declining. Funding for alpha taxonomy is increasingly scarce and even museums, the hearts of taxonomy, are facing financial and philosophical challenges (Noss 1996; Wheeler 2004). Conversely, the field of molecular taxonomy (or molecular systematics) is growing in popularity.

Molecular taxonomy uses genetic variation, usually from one or more genes, to construct a phylogeny with the implicit assumption that the gene genealogies reflect the phylogeny of the species sampled (Pamilo and Nei 1988; Harrison 1989; Brower and DeSalle 1994; Brower et al. 1996; Degnan and Rosenberg 2006). Molecular taxonomy has the advantage that it is relatively easily applied, does not require the same expertise as traditional taxonomy, and, importantly, provides an evolutionary framework for the taxa in question (Tautz et al. 2002, 2003). The pinnacle of enthusiasm for molecular approaches to taxonomy is DNA barcoding, which proposes the use of a 648 base pair (bp) fragment of the mitochondrial cytochrome c oxidase subunit one (COI) gene to identify and delineate species (Hebert et al. 2003a, b). This approach has some advantages, including the ability to obtain taxonomic information from portions or fragments of organisms, its applicability at any life stage (e.g. Sperling et al. 1994; Greenstone et al. 2005; Miller et al. 2005), and the speed with which data can be obtained. The concept of using DNA to identify an organism is not new, and routinely employs a BLAST (Basic Local Alignment Search Tool) search on NCBI's (National Center for Biotechnology) GenBank or similar databases (Tautz et al. 2003; Hebert et al. 2004b). DNA barcoding is simply an attempt to standardize which piece of DNA is used.

The arguments put forth by proponents of DNA barcoding strongly suggest that molecular approaches and a common database will be increasingly useful for identification. In essence, the molecular database will act as a "molecular field guide," facilitating identifications. They also argue that discovery of cryptic species will be enhanced with this approach. Discovery of cryptic variation has long been the purview of molecular phylogenetics and the utility of molecular approaches for this aspect of identification is undeniable (Donnellan and Aplin 1989; Good 1989). In fact, some molecular phylogeneticists have

criticized the DNA barcoding agenda as being a misapplication and simplification of molecular techniques (Sperling 2003a; Rubinoff 2006). Our concern here is not with all of molecular phylogenetics, but with the utility of molecular taxonomic approaches for demarcation, especially when they utilize single-locus genetic data (as emphasized by proponents of DNA barcoding) and are applied to taxa of conservation interest. What might the consequences be of using DNA barcodes, or any single locus sequence, in conservation?

Applying molecular taxonomy to conservation certainly has advantages for screening for illegal trafficking of endangered species (Baker et al. 1996; DeSalle and Birstein 1996; Palumbi and Cipriano 1998; Ludwig 2006) and revealing cryptic species or lineages previously undescribed (e.g. Brower 1996; Omland et al. 2000; Witt et al. 2006; but see Irwin 2002) (for other applications of phylogenetics in conservation see Purvis et al. 2005). However, the utility of phylogenetic approaches for identifying closely related taxa (particularly those of recent origin) or for diagnosing populations or lineages that merit protection is unresolved (Sperling 2003a). The term evolutionary significant unit (ESU) has been used to refer to populations or lineages that represent unique, significant adaptive variants within species (Avice 1989; Moritz 1994, 2002; Vogler and DeSalle 1994; Bowen 1998, 1999; Grady and Quattro 1999; Paetkau 1999; Crandall et al. 2000; Goldstein et al. 2000; Fraser and Bernatchez 2001; Agapow et al. 2004). Increasingly, molecular taxonomy and phylogenetic analyses have been used to identify ESUs, based upon the assumption that historically distinct populations have the greatest potential to contain distinct adaptive variation. However, earlier incarnations of the ESU included the use of other information in addition to molecular data: "Identification of ESUs within a species was recognized as a difficult task, requiring the use of natural history information, morphometrics, range and distribution data, as well as protein electrophoresis, cytogenetic analysis, and restriction mapping of nuclear and mitochondrial DNA" (Ryder 1986). For a population to be designated as an ESU based solely on sequence variation at a single marker locus ignores the fact that adaptive evolution and population differentiation can proceed at a rapid pace compared to rates of molecular evolution at a presumably neutral marker. Thus, historical isolation would trump morphological, life history, and ecological discontinuities discordant with molecular data. Despite the claim that molecular phylogenetic approaches will provide a more accurate view of biodiversity (Hebert et al. 2003a, b; but see

Brower 2006), a number of processes occurring at or below the species level, including recent differentiation and hybridization, will result in an underestimation of biodiversity.

Our principal objectives are to provide researchers and conservation managers with a discussion of patterns and processes at or near the species level that can create problems associated with molecular taxonomy based on single-locus genetic data. The relevant literature from genetic studies of the Lepidoptera is examined, which provides a context in which future conservation genetics work may be evaluated. We urge that employment of molecular genetics data for identification of units for conservation be undertaken with healthy skepticism and an informed opinion about both the utility and the limitations of these techniques.

We focus here on the Lepidoptera for several reasons. The order has a long history of study and is well described compared to other invertebrate taxa. The ecologies and life histories of the Lepidoptera are relatively well known; this is especially true of the butterflies, which enables us to carefully examine the efficacy of molecular taxonomy for this group. This group is also charismatic and important in conservation. Disproportionately more Lepidoptera appear on the endangered species list than any other group of insects and the Lepidoptera have been used as bioindicator species for identification of diversity hotspots or as indicators of ecosystem health (Kremen 1992; Gaston and David 1994; Beccaloni and Gaston 1995; Cremene et al. 2005; Werner and Buszko 2005). Furthermore, several recent examples of the DNA barcoding approach involve investigations of Lepidoptera (e.g. Hebert et al. 2004a; Hajibabaei et al. 2006). Lepidoptera also serve as an important test case for the use of mitochondrial DNA (mtDNA) in species identification because female Lepidoptera are heterogametic, and Haldane's rule predicts reduced viability of the heterogametic sex in hybrids (which should limit introgression of a maternally inherited genetic element; discussed further below) (Sperling 2003b).

Survey of the literature on the Lepidoptera

We searched the genetic literature involving Lepidoptera for any study in which genetic data were found to be in conflict in some way with nominal taxonomic designations. Specifically, we looked for examples in which researchers set out to study taxa (species, subspecies or "races") believed to be distinct based on morphological, ecological, or behavioral characteristics, but concluded that mtDNA sequences alone could not be used to define

or identify the taxa in question. Studies which simply surveyed genetic variation within ecologically or morphologically polymorphic species without any particular expectation of reproductive isolation or genetic divergence among taxa were not considered.

The following key words generated 437 studies in ISI Web of Science: "Lepidoptera* OR butterfly* OR moth* NOT mother*" combined with "mtdna OR mitochondrial." Of those 437 studies, 147 examined mtDNA sequences from two or more Lepidopteran taxa; these studies ranged from phylogeographic studies of mtDNA variation among races and subspecies to systematic studies including only one or two specimens per taxon. For the sake of brevity in the results reported here, we focus on studies which used direct sequencing, as opposed to restriction-site analysis (in a very small number of cases, we have included studies using restriction-site analysis if they also included extensive sequencing to verify the identity of haplotypes). We avoided duplication in that list of 147 by removing a small number of studies in which similar conclusions were reached with the same taxa using different data sets.

We found 31 studies in which genotypic information was perceived to be in conflict with nominal taxonomic boundaries (Table 1). That is 21% of the total number (147) of studies using mtDNA sequence data. It should be noted that many of the 147 studies focused on higher-level systematics, and had little to no chance of discovering taxonomic-molecular discrepancies among closely related taxa. Again for the sake of brevity in the results reported here, we did not report cases in which one or two individuals (out of many dozens or hundreds sequenced from individual taxa) were found to be a mismatch between taxonomic designation and mtDNA genotype (e.g. Dasmahapatra et al. 2002; Ounap and Viidalepp 2005; Kronforst et al. 2006; Mullen 2006).

Nine of the 31 studies documented in Table 1 involve taxa that are either threatened, endangered, or rare. In at least 20 of the 31 studies, the discrepancy between genetic data and taxonomy was caused by the sharing of mtDNA haplotypes among focal taxa. In other words, individuals assigned a priori to different taxa possessed identical sequences of mtDNA. The population genetic processes underlying this and other sources of genetic-taxonomic conflict are discussed in detail below.

The percentage of studies reported above should not be taken as an estimate of the proportion of sister taxa in the Lepidoptera that can or cannot be identified with single-locus genetic data. There is an inherent bias, particularly within taxonomically controversial groups,

Table 1 A summary of published studies in which patterns from the analysis of mitochondrial DNA (mtDNA) sequences were not consistent with nominal taxonomic boundaries. Designations of rank or status used below (i.e. races, subspecies, or species) are those used by the authors of each study; “species complex” generally refers to a group of taxa of controversial taxonomic status containing at least some members that are distinct and possibly genetically isolated. Only the source of mtDNA data is reported below (although nuclear DNA was studied in many cases)

Source	Taxa	Conservation issue	Characters used in the recognition of taxa	mtDNA data	Summary
1 Aagaard et al. (2002)	<i>Aricia agestis-antaxeres</i> species complex (Lycaenidae)	Putative species status of distinct populations in the United Kingdom. Both nominal species are on European red data lists.	Wing pattern and voltinism.	Cyt-B, 325 bp ($n = 179$)	Private haplotypes in focal populations are at most distinguished by a single substitution, thus elevated status (from subspecies to species) is not recommended.
2 Bogdanowicz et al. (2000)	<i>Lymantria dispar</i> races and subspecies (Lymantriidae)	None	Morphology, including larval and adult characters, as well as developmental abnormalities in some hybrids.	COI, 378 bp ($n = 72$); COII, 345 bp ($n = 72$); NDI, 329 bp ($n = 72$)	Haplotype distributions indicate broad geographic structure but a single haplotype is shared throughout a diversity of described taxa in Asia.
3 Brower and Jeansonne (2004)	<i>Danaus plexippus</i> geographic races (Nymphalidae)	Legislation of commercial trafficking in live <i>D. plexippus</i> .	Populations have divergent migratory behavior in the wild, and different susceptibilities to a protozoan parasite. Mimetic wing pattern.	COI-tRNA ^{Leu} -COII-tRNA ^{Ala} s, 1,532 bp ($n = 38$)	Eastern and western races are relatively homogeneous (haplotypes are shared), which (the authors suggest) may support a ban on trafficking to reduce pathogen or parasite spread.
4 Brower (1994)	<i>Heliconius erato</i> races (Nymphalidae)	None	Mimetic wing pattern.	COI, leucine tRNA, COII, 950 bp ($n = 52$) COI, 665 bp ($n = 18$)	Haplotypes are shared among races with highly divergent wing morphologies.
5 Eastwood and Hughes (2003)	<i>Acrodipsas cuprea</i> races (Lycaenidae)	Considered rare.	Wing pattern.	COI, 665 bp ($n = 18$)	Individuals from three of the four taxa have haplotypes in common.
6 Fordyce and Nice (2003)	<i>Battus philenor</i> subspecies (Papilionidae)	None	Adult morphology and ecological traits, primarily clutch size.	COII, 422 bp ($n = 71$)	Subspecific designations explain very little (approximately 3%) of genetic variation.
7 Forister et al. (2004)	<i>Hesperia comma</i> complex (Hesperiidae)	Populations in some parts of the range considered rare and declining.	Wing pattern, phenology, and habitat associations.	COII, 474 bp ($n = 151$)	Some nominal boundaries correlate with mtDNA lineages (and private haplotypes), others do not.
8 Gompert et al. (2006)	<i>Lycaeides melissa</i> and <i>L. m. samuelis</i> (Lycaenidae)	<i>L. m. samuelis</i> is a federally endangered species in the U.S.	Wing pattern, male genital morphology, host plant and habitat associations.	COI, 410 bp; COII, 510 bp ($n = 40$)	MtDNA haplotypes of the endangered taxon from a portion of the range are shared with the more widespread taxon.
9 Grapputo et al. (2005)	<i>Dahlica</i> and <i>Siederia</i> species (Psychidae)	Some taxa are considered rare.	Ecological, reproductive, and morphological differences (primarily wing scale shape).	COI-tRNA-COII-12 S, 942 bp ($n = 73$)	Identification of specimens by morphology is in many cases in disagreement with phylogenetic clustering by haplotype.
10 Grill et al. (2006)	<i>Maniola</i> species (Satyridae)	<i>M. chia</i> , a geographically isolated island taxon, is considered rare.	Wing pattern and color.	.COI, 486 bp; Cytb, 502 bp ($n = 16$)	<i>M. chia</i> and its closest relative are not reciprocally monophyletic (and are thus inferred to be conspecific); low levels of gene flow inferred between other taxa based on similarity of haplotypes.

Table 1 continued

Source	Taxa	Conservation issue	Characters used in the recognition of taxa	mtDNA data	Summary
11 Groman and Pellmyr (2000)	<i>Prodoxus quinquepunctellus</i> ^a host races (Prodoxidae)	None	Body size, ovipositor morphology, and phenology.	COI, 686 bp (<i>n</i> = 78)	A common haplotype is found in every population (of both host-associated taxa).
12 Jiggins (2003)	<i>Acraea encedon</i> and <i>A. encedana</i> (Nymphalidae)	None	Authors do not describe diagnostic characters.	COI, 430 bp (<i>n</i> = 55)	Introgression associated with <i>wolbachia</i> infection has homogenized mitochondrial haplotypes in infected individuals of both species.
13 Kaila and Stahls (2006)	<i>Elachista</i> morphospecies (Elachistidae)	None	Host association, larval mime architecture, and adult morphology.	COI, 705 bp (<i>n</i> = 49)	Identical haplotypes found in multiple morphospecies; in other cases, distinct haplotypes within taxa do not cluster together in phylogenetic analyses.
14 Kato and Yagi (2004)	<i>Parides</i> (<i>Byasa</i>) <i>alcinous</i> subspecies (Papilionidae)	None	Wing pattern, pupal diapause, egg size, host-plant use, and poor hybrid development (between two of the subspecies).	NADH dehydrogenase subunit 5, 778 bp (<i>n</i> = 47)	The majority of individuals from two of the taxa possessed identical haplotypes.
15 Kruse and Sperling (2001)	<i>Archips argyrospila</i> species complex (Tortricidae)	None	Primarily pheromone type and larval host, but also wing pattern.	COI, 475 bp (<i>n</i> = 97)	Molecular variation (in terms of lack of monophyly and patterns of haplotype sharing) corresponds to species distinction of some but not all of the taxa.
16 Landry et al. (1999)	<i>Argyrotaenia franciscana</i> species group	None	Primarily wing pattern and habitat.	COI, 799 bp (<i>n</i> = 48)	Based on close relationships among haplotypes, it is suggested that four of the species in the complex be treated as one.
17 Lushai et al. (2005)	<i>Danaus chryseippus</i> subspecies (Nymphalidae)	None	Wing pattern, migration behavior, mate choice.	12 S, 347 bp; COI, 537 bp (<i>n</i> = 126)	Haplotypes shared among all four taxa in areas of sympatry.
18 Narita et al. (2006)	Two unnamed sibling species, nominally one, <i>Eurema hecabe</i> (Pieridae)	None	Host association, wing coloration, mate choice.	ND5, 716 bp (<i>n</i> = 62)	<i>Wolbachia</i> infection appears to facilitate the extensive interspecific spread of mtDNA.
19 Nice and Shapiro (2001b)	<i>Mitoura</i> species complex (Lycaenidae)	<i>M. muiri</i> is considered a species of conservation concern by the Nature Conservancy.	Wing pattern, host association, habitat, and phenology.	COII, 480 bp (<i>n</i> = 58)	In one half of its range, <i>M. muiri</i> populations are fixed for the haplotype of the more common taxon.
20 Nice and Shapiro (2001b)	<i>Oeneis chryxus</i> complex (Satyridae)	None	Wing pattern.	COII, 480 bp (<i>n</i> = 60)	A common haplotype is found in populations of the three focal taxa.
21 Rubinoff and Sperling (2004)	<i>Hemiteuca maia</i> complex (Saturniidae)	The bog buckmoth (lacking binomial) has a small range and is legally protected.	Habitat association and host plant use.	COI, 624 np (<i>n</i> = 16 plus outgroups)	An absence of reciprocal monophyly indicates that the protected taxon is not genetically isolated.
22 Segreaves et al. (2005)	<i>Tegeticula intermedia</i> and <i>T. elatella</i> (Prodoxidae)	None	Genital morphology and pollination habits.	COI, 779 bp (<i>n</i> = 146)	Haplotypes shared between the “cheater” and “pollinator” species.

Table 1 continued

Source	Taxa	Conservation issue	Characters used in the recognition of taxa	mtDNA data	Summary
23 Sperlting and Hickey (1994)	<i>Choristoneura fumiferana</i> species complex (Tortricidae)	None	Primarily larval host and voltinism, morphological variation is less distinctive.	COI, tRNA, and COII, 1,573 bp ($n = 10$); COII, 470 bp ($n = 37$)	Three of five taxa share haplotypes; it is concluded that they are recently evolved races not meriting species status.
24 Sperlting et al. (1996)	Pheromotypes of <i>Feltia jaculifera</i> (Noctuidae)	None	Mate recognition (pheromones) and flight period.	COI, COII, and tRNA, 2,300 bp ($n = 3$); COI, 820 bp ($n = 74$)	Haplotypes are shared among three of the four pheromotypes.
25 Sperlting et al. (1995)	<i>Yponomeuta</i> species (Yponomeutidae)	None	Host association and slight morphological differences.	COI, COII, and tRNA, 2,294 bp ($n = 44$)	One haplotype shared among two of three taxa (other haplotypes very similar in the same two taxa).
26 Sperlting et al. (1999)	<i>Lambdina fuscilaria</i> subspecies and related species (Geometridae)	None	Host association, mate recognition (pheromones), and slight morphological differences.	COI, COII, and tRNA, 2,300 bp ($n = 3$); COI, 441 bp ($n = 28$) & 518 bp ($n = 145$)	Haplotypes are shared among species (including outgroups) and subspecies of the focal taxon.
27 Svensson et al. (2005)	<i>Prodoxus quinquepunctellus</i> host races (Prodoxidae)	None	Body size, ovipositor morphology, and phenology.	COI, 618 bp ($n = 95$)	Out of seven haplotypes, three are shared between the taxa.
28 Vandewoestijne et al. (2004)	<i>Aglais urticae</i> subspecies (Nymphalidae)	None	Wing pattern.	COI, 1,216 bp, control region, 715 bp ($n = 68$)	Geographic distribution of haplotypes indicates high levels of ongoing gene flow among regions and subspecies.
29 Wahlberg et al. (2003)	<i>Phyciodes</i> species (Nymphalidae)	None	Wing pattern, genitalia, ecology.	COI, 1,450 bp ($n = 140$)	Introgression is inferred between two sympatric taxa (based on a lack of reciprocal monophyly).
30 Weingartner et al. (2006)	<i>Pararge</i> species (Nymphalidae)	None	Wing color.	COI, 1,450 bp ($n = 71$)	One haplotype is shared among individuals of two distinct subspecies.
31 Whinnett et al. (2005)	<i>Melinaea isocomma</i> , <i>M. marsaeus</i> , <i>M. satevis</i> , and <i>M. menophilus</i>	None	Authors do not describe diagnostic characters.	COI/COII, 1,619 bp ($n = 172$)	Species have identical or very similar haplotypes.

^a *P. quinquepunctellus* appears twice in the table (see the study by Svensson et al. (2005)). However, populations studied by Groman and Pellmyr (2000) are considered by some authors to be *P. decipiens*, and are, in any event, allopatric to those studied by Svensson et al. (2005)

which prohibits such a tally: a researcher studying a suite of closely related taxa may decide which category of names to use (species vs. subspecies, for example) only after the genetic data has been analyzed. Rather than put a number to the potential error rate in molecular taxonomy, our review of the literature highlights cases in which reliance on single locus data for taxonomic identification may present serious problems. In most of the cases in our survey of the Lepidopteran literature, the authors had some experience with the natural history of the studied organisms. Consequently, failures of molecular taxonomy were readily identified. As molecular taxonomy is applied to less well known taxa, it may become nearly impossible to detect such shortcomings.

Hajibabaei et al. (2006) examined the efficacy of DNA barcoding in three Lepidopteran families and concluded that 97.9% of the 521 species were accurately distinguished. This figure is truly impressive, however, it is not clear how these results apply to assessing the utility of DNA barcoding for conservation at the species or subspecies levels. Hajibabaei et al. (2006) used relatively few individuals per species and all specimens were collected from a single locality (Area de Conservación Guanacaste, Costa Rica). Consequently, no information about geographical genetic variation within species was assayed. This feature of their data introduces a bias that prevents a clear assessment of the utility of barcoding for conservation (Sperling 2003a). This is because sampling from a single locality provides little information on the efficacy of DNA barcoding to distinguish units for conservation from closely related lineages which may often be allopatric (e.g. Hall and Harvey 2002). Furthermore, data from a single locality provide no insight into geographical phenomena such as barriers between taxa, clinal variation, or isolation by distance. A more comprehensive test of DNA barcoding was done by Meyer and Paulay (2005) using marine gastropods. They surveyed genetic variation across the ranges of gastropod species and found considerable error rates (of species identification), as high as 34% for some groups (Meyer and Paulay 2005), because intraspecific variation overlapped with interspecific variation (see also Funk and Omland 2003).

Patterns and processes

There are several important mechanisms that may cause single-locus molecular data to misdiagnose the true evolutionary relationships among populations, subspecies and species. Here we briefly outline these

mechanisms and describe several relevant methodological approaches.

Incomplete lineage sorting and the problem of recency

Perhaps one of the most important causes of taxonomic misdiagnosis by single locus genetic data arises from the simple fact that selection can change trait values more rapidly than the neutral processes of drift and mutation can create lineage divergence. Natural selection can be extremely effective over just a few generations (Lande and Arnold 1983; Schluter and Smith 1986; Grant and Grant 1993; Reznick and Ghalambor 2001) and can play a direct and important role in speciation (Funk 1998; Nagel and Schluter 1998; Orr and Smith 1998; Jiggins et al. 2001; Ramsey et al. 2003; Forister 2005, Rundle and Nosil 2005). Estimates of the strength of selection across multiple studies also demonstrate that in certain situations selection associated with mate recognition and mate choice (sexual selection) may be especially effective (Hoekstra et al. 2001; Kingsolver et al. 2001; Naisbit et al. 2001).

Neutral processes, on the other hand, operate more slowly. For Lepidoptera, estimates of rates of sequence divergence range from 1.7% divergence per lineage per million years based on the general arthropod mitochondrial mutation rate (Brower 1994) to slower estimates of 0.39–0.51% per lineage per million years estimated in the swallowtail butterfly genus *Papilio* (Zakharov et al. 2004). Adaptive, quantitative traits (such as morphological or behavioral traits) are likely to evolve at rates considerably faster than genes evolving by neutral processes. To put these rates in perspective for the 648 bp fragment of COI recommended for DNA barcoding, we can expect one nucleotide change to occur every 130,000–390,000 years on average (this is of course only a very rough estimate and it should be noted that mutations occur stochastically rather than regularly as this kind of calculation misleadingly implies). The rate disparity between the evolution of adaptive traits and presumably neutral genes becomes particularly important for molecular taxonomy in adaptive radiations. Here, selection operating on morphological, ecological and/or behavioral traits can far outpace neutral changes at single loci such as mtDNA markers. (Neutrality of mtDNA sequence variation is assumed throughout this paper; for a discussion of potential non-neutrality of mtDNA variation under certain conditions and the consequences for population and phylogenetic analyses of such non-neutrality see Ballard and Whitlock (2004) and Hurst and Jiggins (2005)).

A consequence of the different rates of evolution associated with adaptive traits and neutral genetic variation is the pattern of incomplete lineage sorting, in which some alleles within one taxon may be most closely related to alleles in another taxon. Put another way, for cases of rapid divergence, there is some time in which the genetic variation that existed in the ancestral species (ancestral polymorphism) is not sorted into monophyletic gene trees (Pamilo and Nei 1988; Doyle 1992; Maddison 1997; Paetkau 1999) (Fig. 1, based on Avise (1994)). Variation in each descendant lineage is thus a sample of the ancestral polymorphism. Gene trees in new species can therefore be polyphyletic or unresolved for some time after a speciation event (Figs. 1 and 2) (Neigel and Avise 1986; Brower et al. 1996). This situation might continue until reproductive isolation, mutation and drift eventually create reciprocal monophyly and complete lineage sorting between the sister taxa, at which point the gene tree might be congruent with the species tree (Figs. 1 and 2) (Maddison 1997). Assuming neutrality, Neigel and Avise (1986) and Takahata (1989) demonstrated that this period of incongruence between gene trees and species trees for mtDNA will last on average for a time (in generations) roughly equal to four times the effective population size ($4N_e$ generations). Consequently, genealogical relationships among mtDNA haplotypes, even if correctly reconstructed, are unlikely to reflect the true phylogenetic relationships of the two lineages for a considerable period of time until mutation and drift eventually “sort the lineages” and the two descendant taxa become reciprocally monophyletic (Pamilo and Nei 1988; Takahata 1989; Maddison 1997).

The studies reported in Table 1 provide numerous illustrations of the early stages of the diversification and lineage sorting process as depicted in Figs. 1 and 2. For example, the butterfly *Acrodipsas cuprea* in southeastern Australia is composed of four geographic races characterized by differences in male wing color pattern that apparently evolved recently and in allopatry (study 5 in Table 1). Six mitochondrial haplotypes were found throughout the four races: three of them were “private” haplotypes (found in only one taxon), while the other three were shared among taxa (compare to Fig. 2A immediately after the gene pools are divided). Similarly, a mix of private and common haplotypes were found among species in the *Mitoura* complex (study 19), and among morphologically and phenologically divergent host races of *Prodoxus quinquepunctellus* yucca moths (study 11). Further along in the process of differentiation, shared haplotypes may

no longer exist, but sequences sampled from different taxa do not yet form reciprocally monophyletic lineages. This has caused many authors to infer that focal taxa should not be considered genetically distinct taxa or species (e.g. studies 16, 21, 29). This problem is not unique to the Lepidoptera. Funk and Omland (2003) found incomplete lineage sorting, introgression and other processes resulting in non-monophyletic relationships in 23% of the 2319 metazoan species they surveyed.

While the idea of incomplete lineage sorting conflicts with common intuitions about the utility of phylogenetics for delineating species, it is worth noting that mitochondrial genes (and chloroplast genes in plants) represent a tiny fraction of the total genome. Natural selection operates on specific, ecologically relevant gene loci. The tiny genome fraction that consists of mitochondrial genes usually has nothing to do with adaptive differentiation driven by selection. In a very real sense, the genome can be considered a mosaic of genes, some of which undergo rapid allele frequency changes due to selection; the rest (the majority), including mitochondrial genes, experience the effects of genetic drift at a relatively slower pace.

The phenomenon of incomplete lineage sorting should not itself be considered erroneous. While it can create difficulties for diagnosing taxonomic boundaries, lineage sorting represents the natural genealogical processes that occur during divergence and speciation. A solution is to examine genealogies across multiple loci (Wu 1991; Doyle 1992; Moore 1995, 1997; Maddison 1997; Hoelzer 1997). Complete consensus across multiple gene trees can rule out incomplete lineage sorting. Likewise, conflict between gene trees can be used to detect the presence of ancestral polymorphism. A phylogeny that accurately reflects evolutionary relationships among taxa may be estimated from the combination of data from multiple loci. As Maddison (1997) eruditely suggested, the species tree “can be visualized like a fuzzy statistical distribution, a cloud of gene histories.”

Hybridization

The second important cause of misdiagnoses of taxa is hybridization (i.e. interbreeding between distinct populations). Introgressive hybridization occurs when genes from one lineage “invade” the gene pool of another lineage through hybrid matings (Fig. 1D) (Funk and Omland 2003, Mallet 2005). Hybridization, followed by backcrossing, can lead to the establishment of the introgressing allele(s) at high frequency in the

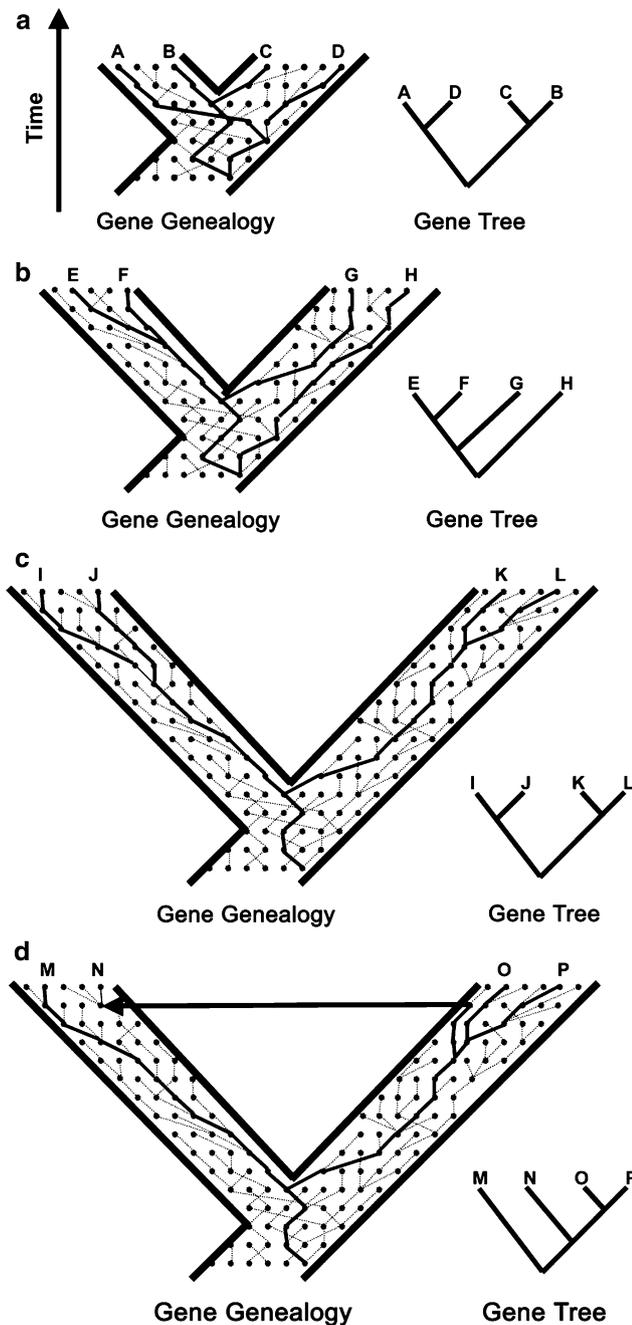


Fig. 1 Gene genealogies of diverging taxa sampled over time. Circles indicate genes in each generation (horizontal row). Thick lines indicate the species tree. Thin solid lines indicate the genealogy of the sampled genes labeled with letters. Dotted lines indicate genealogical relationships of unsampled genes. (a) Shortly after speciation, the gene tree exhibits polyphyly with respect to the two species. (b) Later, the gene tree exhibits paraphyly with respect to the two species. (c) Even later, the gene tree exhibits monophyly with respect to the two species. (d) Reticulate genealogy caused by gene flow (introgressive hybridization). The gene tree exhibits paraphyly with respect to the two species

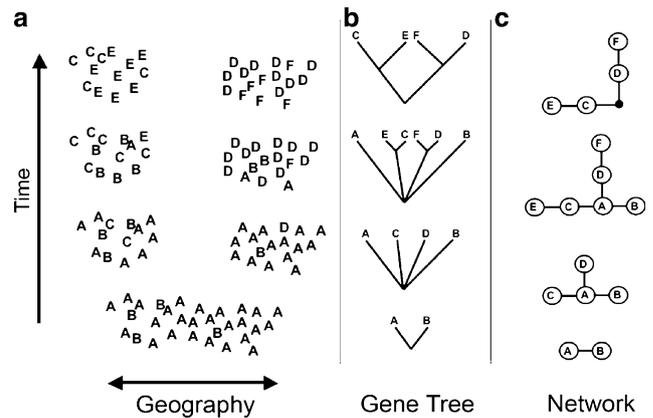


Fig. 2 A schematic representation of genealogical evolution. The geographical distribution of haplotypes (sequence alleles) of a hypothetical speciation event and corresponding gene trees and networks of sampled genes over time. (a) An ancestral population (at the bottom of the figure) containing two haplotypes (a and b) undergoes allopatric divergence. Over time, new haplotypes (i.e. c–f) evolve by mutation, some become extinct, and allele frequencies change by genetic drift. (b) Gene trees of the haplotypes change over time. Shortly after speciation, the gene tree is unresolved, exhibiting a basal polytomy. Later, new haplotypes evolve (i.e. c–f) although the ancestral haplotypes (a and b) are still extant. Even later, ancestral haplotypes become extinct and the last gene tree (top) exhibits monophyly with respect to the two species. Note that the ancestor-descendant relationships between haplotypes are obscured. (c) Haplotype networks representing the genealogical relationships among haplotypes over time. Circles with letters indicate haplotypes. A closed circle represents a missing but inferred haplotype that is presumed to be extinct (i.e. haplotype a in this example).

new gene pool by chance (genetic drift) or by selection if the introgressing allele is positively selected (selective sweep).

The likelihood of this gene flow occurring is reduced if the introgressing gene is linked to loci under selection or is itself disfavored by selection (Barton 1985). Thus presumed neutral loci unlinked to nuclear genes, such as mtDNA and chloroplast DNA (cpDNA), are the most likely candidates for introgression. However, while cytoplasmic DNA (mtDNA and cpDNA) may move readily across species boundaries, nuclear markers may also show introgression (e.g. Strieff et al. 2005). Chan and Levin (2005) considered introgression in the context of models of frequency-dependent assortative mating. For reasonable parameter values (frequency of hybridization, strength of selection, etc.), they found that extensive introgression can occur quite readily, even without prolonged contact between the species. Sympatry was not required for introgression.

Occasional immigrants, including rare, long-distance dispersers, can lead to introgression.

In a phylogenetic context, introgressed genes clearly do not represent the true evolutionary relationships among the taxa involved (Fig. 1) and therefore present a serious problem for accurate diagnoses of units for conservation. Gompert et al. (2006; study 8 in Table 1) recently described this phenomenon in the case of the endangered Karner blue butterfly (*Lycaeides melissa samuelis*) in which mtDNA sequences are shared between this endangered species and its close relative the Melissa blue butterfly, *L. m. melissa* (although mtDNA haplotypes are introgressed, nuclear loci are diagnostic for the two taxa). Other examples in which introgression has been implicated include studies 12, 16, and 21 in Table 1.

For the Lepidoptera, mitochondrial introgression should be less frequent because the females are the heterogametic sex (Sperling 1993, 2003b). Haldane's rule predicts that the heterogametic sex should exhibit reduced viability as hybrids compared to the homogametic sex (Haldane 1922; Turelli and Orr 2000). Since mtDNA is maternally inherited, Haldane's rule should restrict introgression in the Lepidoptera (Sperling 1993). Despite this, several clear cases of mitochondrial introgression in the Lepidoptera have been reported (examples from Table 1, cited above; also, Jiggins et al. 1997), and introgression may be even more prevalent in animals with male heterogamety.

Detecting introgression can be difficult but is most easily recognized by discordant patterns of variation across multiple loci. mtDNA introgression is most easily detected by comparison to data from multiple nuclear markers (e.g. Gompert et al. 2006). It is clear that data from multiple markers (i.e. several loci) with varying modes of inheritance offer the best possibility for not only detecting discordant patterns arising from incomplete lineage sorting and/or introgressive hybridization, but also for accurately capturing the evolutionary history of closely related taxa (Maddison 1997; Chan and Levin 2005; Gompert et al. 2006).

Taxonomic rank

Perhaps the mechanisms discussed above are irrelevant to molecular taxonomy, and DNA barcoding in particular, because data from mtDNA can not be expected to diagnose taxa below the species level (for precisely the reasons discussed above, such as lineage sorting). Thus any conflict between genetic data and taxonomy can be dismissed as a misapplication of technique. Such an argument is unsatisfactory for a number of reasons. Despite the fact that species are generally regarded as

the fundamental unit of biodiversity, there is no general agreement on what a species is (for a review of species concepts see Harrison 1998; Templeton 1989; Hey 2001; Coyne and Orr 2004). This problem extends to subspecies, and for molecular taxonomy the designation of a species or subspecies rank does not necessarily predict sequence divergence (Cognato 2006). For example, the percent sequence divergence for COI between the Melissa blue, *L. m. melissa*, and the endangered Karner blue, *L. m. samuelis*, averages 2.0% (not including introgressed haplotypes; data from Nice and Shapiro 1999 and Gompert et al. 2006). Despite the presence of shared haplotypes, average molecular divergence between the distinct haplotypes in the two subspecies exceeds, or is comparable, to the divergence observed at the same locus between many recognized species pairs, for example *Papilio canadensis* and *P. rutulus* (Papilionidae) are 2.0% divergent (Caterino and Sperling 1999), *Ithomia salapia* and *I. iphianassa* (Nymphalidae) are 0.3% divergent (Mallarino et al. 2005) and *Chrysoritis palmus* and *C. nigricans* (Lycaenidae) are 1.3% divergent (Rand et al. 2000). Furthermore, matings between some butterfly species pairs that exceed 7% divergence at COI can produce viable hybrid offspring (for review of Lepidoptera hybrid viability, see Presgraves (2002)).

Some authors have suggested that the Karner Blue deserves full species status (e.g. Cech and Tudor 2005). From that observation, one might conclude that the sequence divergence between the Melissa blue and Karner blue noted above (comparable to other species pairs) is not surprising. However, whether species or subspecies, approaches based on mtDNA fail (due to shared haplotypes) to recognize the two taxa, which are divergent in wing pattern, male genital morphology, life history and host plant use (Gompert et al. 2006). Thus the question of formal taxonomic rank must be viewed separately from the question of whether or not sequence-based approaches are able to distinguish taxa that are recognizable on ecological, morphological, or behavioral grounds.

It could be argued by the proponents of DNA barcoding that it is inappropriate to use the barcoding approach below the species level because the rank of subspecies implies that taxa are not clearly differentiated. From this position it logically follows that discovery of cryptic variation is impossible. The problem is that cryptic species cannot exist or be discovered if taxonomy is assumed to be correct a priori. If a taxon is already identified as one species, then it cannot be two (one being cryptic). Similarly, it cannot be possible to discover that a subspecies warrants species status. This tautology is irrelevant when we (accurately) treat

phylogenetic hypotheses as hypotheses. Taxonomic rank will not necessarily, therefore, predict the utility or accuracy of DNA sequence-based approaches for identification because population and evolutionary processes are not influenced by Latin binomials.

Analytical advances

Population genetic phenomena occurring at or near the species level, such as lineage sorting and hybridization, not only present problems for DNA barcoding approaches, they require their own analytical tools, such as the implementation of coalescent, modeling-based approaches (discussed below). Methods of DNA sequence analysis employed at higher taxonomic levels (i.e. all of the statistical machinery used for phylogenetic reconstruction) may be inappropriate at lower taxonomic levels (Posada and Crandall 2001). There are several reasons for this:

1. Phylogenetic algorithms assume that evolutionary relationships arise by bifurcations of ancestral species to produce two new species. However, hybridization and hybrid speciation events (as well as recombination generally) can create reticulating patterns that violate this assumption (Posada and Crandall 2001) (Fig. 1 D).

2. Taxa (and/or sequences) at the tips of phylogenetic trees represent extant taxa (and/or sequences), whereas all “ancestral” taxa (and/or sequences) are, by definition, extinct (Swofford et al. 1996; Felsenstein 2004, Freeland 2005). Variation in sequence data from within populations or species represents the genealogical relationships among the sequences and it is clearly possible, and indeed likely, that “ancestral” sequences still exist. Thus the convention in phylogenetics of representing all extant sequences at the tips of trees fails to illustrate the genealogical patterns of descent among sequences within populations and species (Fig. 2).

3. Because of the inherent stochastic nature of lineage sorting and mutation, phylogenetic relationships across multiple loci are unlikely to be congruent until well after divergence is complete (see Incomplete lineage sorting and the problem of recency above). Phylogenetic analyses are not designed to deal with the “noise” created by the discordance among gene trees (Degnan and Rosenberg 2006). At best, combined analyses of multiple loci in these situations will create phylogenetic trees with little or no support which would highlight the recency of divergence. At worst, important information about population-level processes would be obscured or ignored.

Instead of phylogenetic methods, alternative tools have been used by population geneticists over the last decade to handle the astounding increase in population-level molecular data and the associated problems (Wakeley 2007). In Fig. 2C we present a hypothetical case using just one example of an alternative method: a haplotype network, which is a graphical representation of the genealogical relationships among sequences. Networks allow for “ancestral” sequences within the data set and are free of the assumption that relationships among sequences must be bifurcating. The following references are just a few of the sources which contain comprehensive introductions to networks and other tools available to population geneticists today: Posada and Crandall (2001), Lowe et al. (2004), Freeland (2005), Excoffier and Heckel (2006).

The most significant conceptual advance (relevant to the issues which motivated the present paper) may be the fact that certain patterns that are considered “noise” within traditional phylogenetic analyses actually represent valuable data which can be used to reconstruct and model evolutionary histories and processes. Coalescent theory (Hudson 1990, Kingman 2000) uses genealogical relationships among sequences (i.e. gene trees) to make inferences about demographic processes and histories of populations. For example, the sharing of mtDNA haplotypes among taxa or populations, which is simply inconvenient to analyses which assume bifurcating patterns of descent, can be used to estimate a class of population parameters, such as migration rate and ancestral population sizes, which get to the very heart of dynamic processes of diversification. Continuing advances in coalescent theory enable population geneticists to do more with sequence data than simply address taxonomic hypotheses (Hein et al. 2005).

Conclusions

Our goal here has been to point out that there are very important places where molecular data can be perceived to be in conflict with taxonomy, and that this conflict is most likely to occur in situations where identification is the most challenging: at or near the species level, and among closely related taxa. In addition, significant discrepancies between molecular, morphological, and ecological characteristics in some cases, at least in the Lepidoptera, involve taxa of conservation concern. It is important to note the variety of ways that the taxa included in Table 1 differ within each study. The taxa are divergent in various characteristics, from wing patterns (in a majority of the studies) to genetically based differences in phenology

(e.g. studies 11 and 23). Those divergent characteristics caused the researchers to suppose a priori that the taxa would be genetically distinct; that expectation was not borne out by the mtDNA markers analyzed.

The frequent conflict between single-locus genetic data and taxonomy ultimately teaches us that diagnosis of units for conservation should be based on a recognition of the multiple forces that drive the evolution of molecular, ecological, morphological and behavioral characters and the differential in rates of evolution for these characters (Rubinoff 2006). Non-molecular data often provide critical information on adaptive differences between recently diverged lineages that may not be detected with DNA data. The combination of data from multiple genetic markers with more traditional taxonomic data is the most effective approach to avoid taxonomic misdiagnosis. Discrepancies between the two types of data highlight an underappreciated facet of biodiversity: namely that a certain portion of the taxa we observe and may wish to conserve are of relatively recent origin. Consequently, all population-level processes should be considered when attempting to identify units for conservation, not solely the presumed neutral dynamics that underlie mitochondrial DNA evolution.

Acknowledgments An early version of this paper was improved by comments from N. J. Sanders, F. A. H. Sperling, an anonymous reviewer, and the EEB discussion group at Texas State.

References

- Aagaard K, Hindar K, Pullin AS, James CH, Hammarstedt O, Balstad T, Hanssen O (2002) Phylogenetic relationships in brown argus butterflies (Lepidoptera: Lycaenidae: *Aricia*) from north-western Europe. *Biol J Linn Soc* 75:27–37
- Agapow PM, Bininda-Emonds ORP, Crandall KA, Gittleman JL, Mace GM, Marshall JC, Purvis A (2004) The impact of species concept on biodiversity studies. *Q Rev Biol* 79:161–179
- Avise JC (1989) A role for molecular-genetics in the recognition and conservation of endangered species. *Trends Ecol Evol* 4:279–281
- Avise JC (1994) Molecular markers, natural history and evolution. Chapman & Hall, New York
- Baker CS, Cipriano F, Palumbi SR (1996) Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan. *Mol Ecol* 5:671–685
- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Mol Ecol* 13:729–744
- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annu Rev Ecol Syst* 16:113–148
- Beccaloni GW, Gaston KJ (1995) Predicting the species richness of neotropical forest butterflies: Ithomiinae (Lepidoptera: Nymphalidae) as indicators. *Biol Conserv* 71:77–86
- Bogdanowicz SM, Schaefer PW, Harrison RG (2000) Mitochondrial DNA variation among worldwide populations of gypsy moths, *Lymantria dispar*. *Mol Phylogenet Evol* 15:487–495
- Bowen BW (1998) What is wrong with ESUs? The gap between evolutionary theory and conservation principles. *J Shellfish Res* 17:1355–1358
- Bowen BW (1999) Preserving genes, species, or ecosystems? Healing the fractured foundations of conservation policy. *Mol Ecol* 8:S5–S10
- Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial-DNA evolution. *Proc Natl Acad Sci USA* 91:6491–6495
- Brower AVZ (1996) A new mimetic species of *Heliconius* (Lepidoptera: Nymphalidae), from southeastern Colombia, as revealed by cladistic analysis of mitochondrial DNA sequences. *Zool J Linn Soc* 116:317–332
- Brower AVZ (2006) Problems with DNA barcodes for species delimitation: “ten species” of *Astrartes fulgerator* reassessed (Lepidoptera: Hesperiiidae). *System Biodivers* 4:127–132
- Brower AVZ, DeSalle R (1994) Practical and theoretical considerations for choice of a DNA sequence region in insect molecular systematics, with a short review of published studies using nuclear gene regions. *Ann Entomol Soc Am* 87:702–716
- Brower AVZ, Jeansonne MM (2004) Geographical populations and “subspecies” of new world monarch butterflies (Nymphalidae) share a recent origin and are not phylogenetically distinct. *Ann Entomol Soc Am* 97:519–523
- Brower AVZ, DeSalle R, Vogler A (1996) Gene trees, species trees and systematics: a cladistic perspective. *Annu Rev Ecol Syst* 27:423–450
- Caterino MS, Sperling FAH (1999) *Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes. *Mol Phylogenet Evol* 11:122–137
- Cech R, Tudor G (2005) Butterflies of the East Coast: an observer’s guide. Princeton University Press, Princeton NJ
- Chan KMA, and Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* 59:720–729
- Cognato AI (2006) Standard percent DNA sequence difference for insects does not predict species boundaries. *J Econ Entomol* 99:1037–1045
- Coyne JA, Orr HA (2004) Speciation. Sinauer Associates
- Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends Ecol Evol* 15:290–295
- Cremene C, Groza G, Rakosy L, Schileyko AA, Baur A, Erhardt A, Baur B (2005) Alterations of steppe-like grasslands in Eastern Europe: a threat to regional biodiversity hotspots. *Conserv Biol* 19:1606–1618
- Dasmahapatra KK, Blum MJ, Aiello A, Hackwell S, Davies N, Bermingham EP, Mallett T (2002) Inferences from a rapidly moving hybrid zone. *Evolution* 56:741–753
- Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *Pub Lib Sci Genet* 2:762–768
- DeSalle R, Birstein VJ (1996) PCR identification of black caviar. *Nature* 381:197–198
- Donnellan SC, Aplin KP (1989) Resolution of cryptic species in the New Guinean lizard *Sphenomorphus jobiensis* (Scincidae) by electrophoresis. *Copeia* 1:81–88
- Doyle JJ (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst Bot* 17:144–163
- Eastwood R, Hughes JM (2003) Phylogeography of the rare myrmecophilous butterfly *Acrodipsas cuprea* (Lepidoptera: Lycaenidae) from pinned museum specimens. *Aust J Zool* 51:331–340

- Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nat Rev Genet* 7:745–758
- Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Inc Sunderland MA
- Fordyce JA, Nice CC (2003) Contemporary patterns in a historical context: phylogeographic history of the pipevine swallowtail, *Battus philenor* (Papilionidae). *Evolution* 57:1089–1099
- Forister ML (2005) Independent inheritance of preference and performance in hybrids between host races of *Mitoura* butterflies (Lepidoptera: Lycaenidae). *Evolution* 59:1149–1155
- Forister ML, Fordyce JA, Shapiro AM (2004) Geological barriers and restricted gene flow in the holarctic skipper *Hesperia comma* (Hesperiidae). *Mol Ecol* 13:3489–3499
- Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Mol Ecol* 10:2741–2752
- Freeland JR (2005) *Molecular ecology*. John Wiley & Sons Ltd, West Sussex England
- Funk DJ (1998) Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution* 52:1744–1759
- Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Syst* 34:397–423
- Gaston KJ, David R (1994) Hotspots across Europe. *Biodivers Lett* 2:108–116
- Goldstein PZ, DeSalle R, Amato G, Vogler AP (2000) Conservation genetics at the species boundary. *Conserv Biol* 14:120–131
- Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM (2006) Identifying units for conservation using molecular taxonomy: the cautionary tale of the Karner blue butterfly. *Mol Ecol* 15:1759–1768
- Good DA (1989) Hybridization and cryptic species in *Dicamptodon* (Caudata: Dicamptodontidae). *Evolution* 43:728–744
- Grady JM, Quattro JM (1999) Using character concordance to define taxonomic and conservation units. *Conserv Biol* 13:1004–1007
- Grapputo A, Kumpulainen T, Mappes J (2005) Phylogeny and evolution of parthenogenesis in Finnish bagworm moth species (Lepidoptera: Psychidae: Naryciinae) based on mtDNA-markers. *Ann Zool Fenn* 42:141–160
- Grant BR, Grant PR (1993) Evolution of Darwin's finches caused by a rare climatic event. *Proc R Soc Lond B Biol Sci* 251:111–117
- Greenstone MH, Rowley DL, Heimbach U, Lundgren JG, Pfannenstiel RS, Rehner SA (2005) Barcoding generalist predators by polymerase chain reaction: carabids and spiders. *Mol Ecol* 14:3247–3266
- Grill A, Gkiokia E, Alvarez N (2006) Evolutionary history and patterns of differentiation among European *Maniola* butterflies (Lepidoptera: Satyrinae). *Eur J Entomol* 103:613–618
- Groman JD, Pellmyr O (2000) Rapid evolution and specialization following host colonization in a yucca moth. *J Evol Biol* 13:223–236
- Haig SM (1998) Molecular contributions to conservation. *Ecology* 79:413–425
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc Nat Acad Sci USA* 103:968–971
- Haldane JBS (1922) Sex-ratio and unisexual sterility in hybrid animals. *J Genet* 12:101–109
- Hall JPW, DJ Harvey (2002) The phylogeography of Amazonia revisited: new evidence from Riodinid butterflies. *Evolution* 56:1489–1497
- Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Tree* 4:6–11
- Harrison RG (1998) Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. In: Howard DJ, Berlocher SH (eds) *Endless forms: species and speciation*. Oxford University Press, Oxford, pp 19–31
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. *Proc R Soc Lond B Biol Sci* 270:313–321
- 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 Biol* 270:S96–S99
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) 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, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. *Pub Lib Sci Biol* 2:1657–1663
- Hein J, Schierup MH, Wiuf C (2005) Gene genealogies, variation and evolution: a primer in coalescent theory. Oxford University Press, New York
- Hey J (2001) The mind of the species problem. *Trends Ecol Evol* 16:326–329
- Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hoang A, Hill CE, Beerli P, Kingsolver JG (2001) Strength and tempo of directional selection in the wild. *Proc Natl Acad Sci USA* 98:9157–9160
- Hoelzer GA (1997) Inferring phylogenies from MTDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution* 51:622–626
- Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Surveys Evol Biol* 7:1–44
- 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* 272:1525–1534
- Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution* 56:2383–2394
- Jiggins FM (2003) Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics* 164:5–12
- Jiggins CD, McMillan WO, King P, Mallet J (1997) The maintenance of species differences across a *Heliconius* hybrid zone. *Heredity* 79:495–505
- Jiggins CD, Naisbit RE, Coe RL, Mallet J (2001) Reproductive isolation caused by colour pattern mimicry. *Nature* 411:302–305
- Kaila L, Stahls G (2006) DNA barcodes: evaluating the potential of COI to differentiate closely related species of *Elachista* (Lepidoptera: Gelechioidea: Elachistidae) from Australia. *Zootaxa*:1–26
- Kato Y, Yagi T (2004) Biogeography of the subspecies of *Parides* (*Byasa*) *alcinous* (Lepidoptera: Papilionidae) based on a phylogenetic analysis of mitochondrial ND5 sequences. *Syst Entomol* 29:1–9
- Kingman JFC (2000) Origins of the coalescent: 1974–1982. *Genetics* 156:1461–1463
- Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P (2001) The strength of phenotypic selection in natural populations. *Am Nat* 157:245–261

- Kremen C (1992) Assessing the indicator properties of species assemblages for natural areas monitoring. *Ecol Appl* 2:203–217
- Kronforst MR, Young LG, Blume LM, Gilbert LE (2006) Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution* 60:1254–1268
- Kruse JJ, Sperling FAH (2001) Molecular phylogeny within and between species of the *Archips argyrospila* complex (Lepidoptera: Tortricidae). *Ann Entomol Soc Am* 94:166–173
- Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution* 37:1210–1226
- Landry B, Powell JA, Sperling FAH (1999) Systematics of the *Argyrotaenia franciscana* (Lepidoptera: Tortricidae) species group: evidence from mitochondrial DNA. *Ann Entomol Soc Am* 92:40–46
- Lushai G, Allen JA, Goulson D, MacLean N, Smith DAS (2005) The butterfly *Danaus chrysippus* (L) in East Africa comprises polyphyletic, sympatric lineages that are, despite behavioural isolation, driven to hybridization by female-biased sex ratios. *Biol J Linn Soc* 86:117–131
- Lowe A, Harris S, Ashton P (2004) Ecological genetics: design, analysis, and application. Blackwell Science Ltd, Malden MA
- Ludwig A (2006) A sturgeon view on conservation genetics. *Eur J Wildl Res* 52:3–8
- Maddison WP (1997) Gene trees in species trees. *Syst Biol* 46:523–536
- Mallarino R, Bermingham E, Willmott KR, Whinnett A, Jiggins CD (2005) Molecular systematics of the butterfly genus *Ithomia* (Lepidoptera: Ithomiinae): a composite phylogenetic hypothesis based on seven genes. *Mol Phylogenet Evol* 34:625–644
- Mallet J (2005) Hybridization as an invasion of the genome. *Trends Ecol Evol* 20:229–237
- Meyer CP, G Paulay (2005) DNA barcoding: error rates based on comprehensive sampling. *Pub Lib Sci Biol* 3:2229–2238
- Miller KB, Alarie Y, Wolfe GW, Whiting MF (2005) Association of insect life stages using DNA sequences: the larvae of *Philodytes umbrinus* (Motschulsky) (Coleoptera: Dytiscidae). *Syst Entomol* 30:499–509
- Mullen SP (2006) Wing pattern evolution and the origins of mimicry among North American admiral butterflies (Nymphalidae: Limenitids). *Mol Phylogenet Evol* 39:747–758
- Moore WS (1995) Inferring phylogenies from MTDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49:718–726
- Moore WS (1997) Mitochondrial-gene trees versus nuclear-gene trees, a reply to Hoelzer. *Evolution* 51:627–629
- Moritz C (1994) Defining evolutionarily significant units for conservation. *Trends Ecol Evol* 9:373–375
- Moritz C (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Syst Biol* 51:238–254
- Nagel L, Schluter D (1998) Body size, natural selection, and speciation in sticklebacks. *Evolution* 52:209–218
- Naisbit RE, Jiggins CD, Mallet J (2001) Disruptive sexual selection against hybrids contributes to speciation between *Heliconius cydno* and *H. melpomene*. *Proc R Soc Lond B* 268:1849–1854
- Narita S, Nomura M, Kato Y, Fukatsu T (2006) Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. *Mol Ecol Online*
- Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: Nevo E, Karlin S (eds), *Evolutionary processes and theory*. 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. *J Evol Biol* 12:936–950
- Nice CC, Shapiro AM (2001a) Patterns of morphological, biochemical, and molecular evolution in the *Oeneis chryxus* complex (Lepidoptera: Satyridae): a test of historical biogeographical hypotheses. *Mol Phylogenet Evol* 20:111–123
- Nice CC, Shapiro AM (2001b) Population genetic evidence of restricted gene flow between host races in the butterfly genus *Mitoura* (Lepidoptera: Lycaenidae). *Ann Entomol Soc Am* 94:257–267
- Noss RF (1996) The naturalists are dying off. *Conserv Biol* 10:1–3
- Omland KE, Tarr CL, Boarman WI, Marzluff JM, Fleischer RC (2000) Cryptic genetic variation and paraphyly in ravens. *Proc R Soc Lond B Biol Sci* 267:2475–2482
- Orr MR, Smith TB (1998) Ecology and speciation. *Trends Ecol Evol* 13:502–506
- Ounap E, Viidalepp J, Saarma U (2005) Phylogenetic evaluation of the taxonomic status of *Timandra griseata* and *T. comae* (Lepidoptera: Geometridae: Sterrhinae). *Eur J Entomol* 102:607–615
- Paetkau D (1999) Using genetics to identify intraspecific conservation units: a critique of current methods. *Conserv Biol* 13:1507–1509
- Palumbi SR, Cipriano F (1998) Species identification using genetic tools: the value of nuclear and mitochondrial gene sequences in whale conservation. *J Hered* 89:459–464
- Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Mol Biol Evol* 5:568–583
- Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol* 16:37–45
- Presgraves DC (2002) Patterns of postzygotic isolation in Lepidoptera. *Evolution* 56:1168–1183
- Purvis A, Gittleman JL, Brooks T (eds) (2005) *Phylogeny and conservation*. Cambridge University Press, Cambridge
- Ramsey J, Bradshaw HD, Schemske DW (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* 57:1520–1534
- Rand DB, Heath A, Suderman T, Pierce NE (2000) Phylogeny and life history evolution of the genus *Chrysoritis* within the Aphnaeini (Lepidoptera: Lycaenidae), inferred from mitochondrial cytochrome oxidase I sequences. *Mol Phylogenet Evol* 17:85–96
- Reznick DN, Ghalambor CK (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* 112–113:183–198
- Rubinoff D (2006) Utility of mitochondrial DNA barcodes in species conservation. *Conserv Biol* 20:1026–1033
- Rubinoff D, Sperling FAH (2004) Mitochondrial DNA sequence, morphology and ecology yield contrasting conservation implications for two threatened buckmoths (*Hemileuca*: Saturniidae). *Biol Conserv* 118:341–351
- Rundle HD, Nasil P (2005) Ecological speciation. *Ecol Lett* 8:336–352
- Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. *Trends Ecol Evol* 1:9–10
- Schluter D, Smith JNM (1986) Natural selection on beak and body size in the song sparrow. *Evolution* 40:221–231
- Segraves KA, Althoff DM, Pellmyr O (2005) Limiting cheaters in mutualism: evidence from hybridization between mutualist and cheater yucca moths. *Proc R Soc Lond B Biol Sci* 272:2195–2201

- Sperling FAH (1993) Mitochondrial DNA variation and Haldane's rule in the *Papilio glaucus* and *Papilio troilus* species groups. *Heredity* 71:227–233
- Sperling FAH (2003a) DNA barcoding: Deus ex machina. *Newsl Biol Survey Can* 22:50–53
- Sperling FAH (2003b) Butterfly molecular systematics: from species definitions to higher-level systematics. In: Boggs CL, Watt WB, Ehrlich PR (eds) *Butterflies, ecology and evolution taking flight*. Univ. Chicago Press, Chicago, pp 431–458
- Sperling FAH, Hickey DA (1994) Mitochondrial DNA sequence variation in the spruce budworm species complex (*Choristoneura*: Lepidoptera). *Mol Biol Evol* 11:656–665
- Sperling FAH, Anderson GS, Hickey DA (1994) A DNA-based approach to the identification of insect species used for postmortem interval estimation. *J Forensic Sci* 39:418–427
- Sperling FAH, Byers R, Hickey D (1996) Mitochondrial DNA sequence variation among pheromotypes of the dingy cutworm, *Feltia jaculifera* (Gn) (Lepidoptera: Noctuidae). *Can J Zool* 74:2109–2117
- Sperling FAH, Landry JF, Hickey DA (1995) DNA-based identification of introduced ermine moth species in North America (Lepidoptera, Yponomeutidae). *Ann Entomol Soc Am* 88:155–162
- Sperling FAH, Raske AG, Otvos IS (1999) Mitochondrial DNA sequence variation among populations and host races of *Lambdina fiscellaria* (Gn) (Lepidoptera: Geometridae). *Insect Mol Biol* 8:97–106
- Streiff R, Veyrier R, Audiot P, Meusnier S, Brouat C (2005) Introgression in natural populations of bioindicators: a case study of *Carabus splendens* and *Carabus punctatoauratus*. *Mol Ecol* 14:3775–3786
- Svensson GP, Althoff DM, Pellmyr O (2005) Replicated host-race formation in bogus yucca moths: genetic and ecological divergence of *Prodoxus quinquepunctellus* on yucca hosts. *Evol Ecol Res* 7:1139–1151
- Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) *Molecular systematics*. Sinauer Associates, Sunderland Mass, pp 407–514
- Takahata N (1989) Gene genealogy in three related populations: consistency probability between gene and population trees. *Genetics* 122:957–966
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2002) DNA points the way ahead of taxonomy—In assessing new approaches, it's time for DNA's unique contribution to take a central role. *Nature* 418:479–479
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends Ecol Evol* 18:70–74
- Templeton AR (1989) The meaning of species and speciation: a genetic perspective. In: Otte D, Endler JA (eds), *Speciation and its consequences*. Sinauer Associates Inc, Sunderland MA, pp 2–27
- Turelli M, Orr HA (2000) Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* 154:1663–1679
- Vandewoestijne S, Baguette M, Brakefield PM, Saccheri IJ (2004) Phylogeography of *Aglaia urticae* (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and control region. *Mol Phylogenet Evol* 31:630–646
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect? Systematics and the agony of choice. *Biol Conserv* 55:235–254
- Vogler AP, Desalle R (1994) Diagnosing units of conservation management. *Conserv Biol* 8:354–363
- Wahlberg N, Oliveira R, Scott JA (2003) Phylogenetic relationships of *Phyciodes* butterfly species (Lepidoptera: Nymphalidae): complex mtDNA variation and species delimitations. *Syst Entomol* 28:257–273
- Wakeley J (2007) *Coalescent theory: an Introduction*. Roberts & Company Publishers, Colorado
- Weingartner E, Wahlberg N, Nylin S (2006) Speciation in *Pararge* (Satyrinae: Nymphalidae) butterflies—North Africa is the source of ancestral populations of all *Pararge* species. *Syst Entomol* 31:621–632
- Werner U, Buszko J (2005) Detecting biodiversity hotspots using species-area and endemics-area relationships: the case of butterflies. *Biodiv Conserv* 14: 1977–1988
- Wheeler QD (2004) Taxonomic triage and the poverty of phylogeny. *Philos Trans R Soc Lond B Biol Sci* 359:571–583
- Whinnett A, Zimmermann M, Willmott KR, Herrera N, Mallarino R, Simpson F, Joron M, Lamas G, Mallett J (2005) Strikingly variable divergence times inferred across an Amazonian butterfly “suture zone”. *Proc R Soc Lond B Biol Sci* 272:2525–2533
- Witt JDS, Threlloff DL, Hebert PDN (2006) DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Mol Ecol* 15:3073–3082
- Wu CI (1991) Inferences of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics* 127:429–435
- Zakharov EV, Caterino MS, Sperling FAH (2004) Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera : Papilionidae). *Syst Biol* 53:193–215