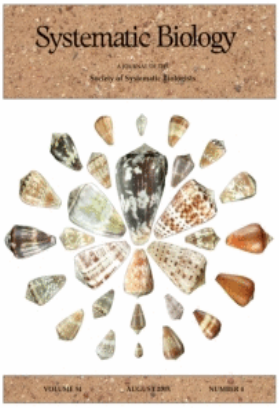

Systematic Biology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713658732

Delimiting Species in Recent Radiations

H. Bradley Shaffer [a]; Robert C. Thomson [a]
[a] Section of Evolution and Ecology, and Center for Population Biology, University of California, Davis, CA, USA

First Published on: 01 December 2007
To cite this Article: Shaffer, H. Bradley and Thomson, Robert C. (2007) 'Delimiting Species in Recent Radiations', Systematic Biology, 56:6, 896 - 906
To link to this article: DOI: 10.1080/10635150701772563
URL: http://dx.doi.org/10.1080/10635150701772563

Syst. Biol. 56(6):896–906, 2007

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150701772563

Delimiting Species in Recent Radiations

H. BRADLEY SHAFFER AND ROBERT C. THOMSON

Section of Evolution and Ecology, and Center for Population Biology, University of California, Davis, CA 95616, USA; E-mail: hbshaffer@ucdavis.edu (H.B.S.), rcthomson@ucdavis.edu (R.C.T.)

Abstract.—Despite considerable effort from the systematics community, delimiting species boundaries in recent radiations remains a daunting challenge. We argue that genealogical approaches, although sometimes useful, may not solve this important problem, because recently derived species often have not had sufficient time to achieve monophyly. Instead, we suggest that population genetic approaches that rely on large sets of informative markers like single nucleotide polymorphisms (SNPs) provide an alternative framework for delimiting very recently derived species. We address two major challenges in applying such markers to species delimitation: discovering markers in nonmodel systems and using them to delimit recently derived species. Using turtles as a test case, we explore the utility of a single, relatively low-coverage genomic resource as an aid in gene and marker discovery. We exploit an end-sequenced bacterial artificial chromosome (BAC) library from an individual painted turtle (*Chrysemys picta*) and outline a novel protocol that efficiently identifies primer pairs that amplify homologous sequences across the tree of living turtles. Preliminary data using this library to discover SNPs in *Emydura macquarii*, a species that diverged from *C. picta* ~210 million years ago, indicate that sequences identified from the *Chrysemys* BAC library provide useful SNPs even in this very distantly related taxon. Several recent methods in wide use in the population genetics literature allow one to discover potential species, or test existing species hypotheses, with SNP data and may be particularly informative for very recently derived species. As BAC and other genomic resources become increasingly available for scattered taxa across the tree of life, we are optimistic that these resources will provide abundant, inexpensive markers that will help delimit boundaries in problematic, recent species radiations. [BAC end sequence; *Chrysemys*; comparative genomics; *Emydura*; recent speciation; species delimitation; turtle.]

Species criteria, definitions, and delimitations occupy one of the most contentious and least resolvable debates in systematics. It seems fair to say that many systematists agree that species are real, important, and sometimes extremely difficult to identify (de Queiroz, 1998; Hey et al., 2003). There is also a growing sense that the virtually limitless empirical data available from emerging genomic databases may help solve the problem of delimiting difficult, recently derived species (Shaw and Danley, 2003; Liti et al., 2006). Presumably, it must be the case that more data are better than less, and that information from multiple, independent genetic loci should help in delimiting species, even if those species are very recently formed and therefore have few or no unique features. However, it is also true that more data do not always help to resolve difficult problems in historical biology. Identifying species trees among rapidly speciating taxa is a case in point; situations exist where the most probable gene tree does not match the actual species tree, and adding additional data does not help (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007).

One point that seems incontrovertible is that it is getting easier to acquire genomic-level data for an ever-widening set of taxa. The NCBI Entrez Genome Project Web page currently lists 236 eukaryote genomes and some 1100 prokaryotic genomes that are completed or in progress and many more partial genomic resources, including end-sequenced bacterial artificial chromosome (BAC) libraries and expressed sequence tag (EST) databases. These resources are produced for single species (often for single individuals), but they generally include hundreds of thousands to millions of base pairs of sequence data for that individual. For these taxa, and presumably many of their relatives, the availability of such genomic tools implies that the information contained in very large quantities of DNA sequence data can now be brought to bear on species delimitation problems. Although these tools may not solve every problem in species identification and delimitation, they should represent a major step forward over the previous generation of genetic analyses. In addition, these multigene data sets should allow systematists to test species lineages that have been proposed based on single gene analyses, including mitochondrial DNA (Avise and Walker, 1999).

In this paper, we focus on two important issues in the use of genomic data in recent species delimitation. First, can genomic-level tools, which will always be available for only a small fraction of the total diversity of life, be applied to related species that lack such tools? More specifically, how can we find hundreds of markers quickly and efficiently in nonmodel systems? Second, how can we use those markers to delimit species in a rigorous, satisfying way? For this second question, we consider several population genetically oriented approaches that may be used in conjunction with more phylogenetically oriented strategies to identify and delimit recently derived species.

In pursuing these questions, we do not discuss the vast literature on species concepts except to say that we fundamentally agree with Mayden (1997) and de Queiroz (1998, 2005) that species are segments of evolutionary lineages. Under this view, which goes back at least to Simpson (1951), species delimitation boils down to the identification of "metapopulation lineages" (de Queiroz, 2005), with the most difficult problems often represented by the youngest species pairs (Fig. 1). As species persist over time, the attributes that they take on, and therefore the ease with which lineages can be discovered and delimited, changes (de Queiroz, 1998). In practice, older species have often had time to accumulate a range of features, including reproductive isolation, fixed apomorphies, and gene-tree monophyly,

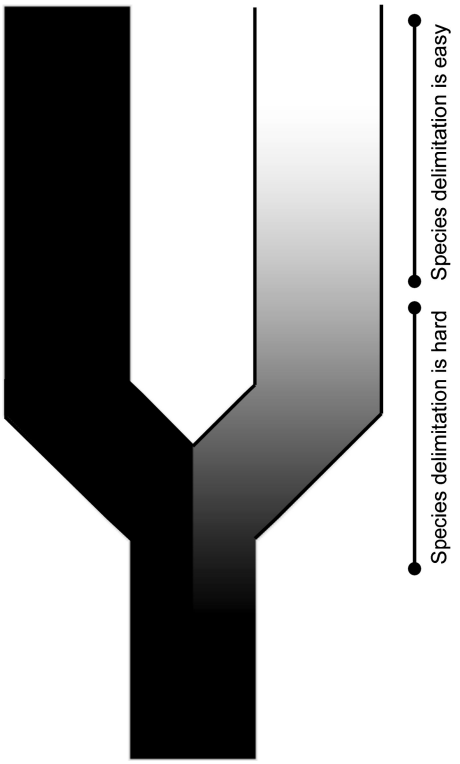

FIGURE 1. A schematic diagram of the divergence of species through time. We focus our attention on the early part of the process, since that is when species delimitation is most difficult. Figure modified from de Queiroz (1998).

that allow for straightforward lineage (= species) delimitation. However, very young taxa that are early in the process of speciation often lack these features, making them particularly challenging for systematists (Fig. 1). We therefore focus our attention on the value of genomic data for delimiting young species early in the process of diversification.

Sites and Marshall (2003, 2004) provide recent summaries of advances in the actual criteria for delimiting species boundaries. These authors place species delimitation methods into two broad categories: tree-based and non–tree-based. Tree-based methods tend to require demonstrable monophyly, whereas non–tree-based methods more often focus on the discovery of differentiation associated with the earliest stages of lineage diversification. Both are clearly valuable in species delimitation research, particularly when sufficient time has elapsed for monophyly to evolve (Hudson and Coyne, 2002). Maddison and Knowles recently examined the related problem of phylogeny reconstruction among known species that have not had time to sort to monophyly, and point out that some limited phylogenetic information still resides in these incompletely sorted gene trees (Maddison and Knowles, 2006). Presumably, the same is true for species delimitation, and methods developed to extract this "fuzzy" information using tree-based approaches are being developed. Here, we concentrate our discussion on approaches derived from population genetics rather than those that require monophyly for species identification. In doing so, we fully recognize that tree-based methods (Baum and Shaw, 1995; Wiens and Penkrot, 2002) may also identify recently derived species; our hope is that both approaches can be brought to bear on difficult but often critically important recent species radiations.

SPECIES BOUNDARIES IN RECENT RADIATIONS

One of the more disturbing realizations to emerge from the speciation literature is that many of the most widely cited and widely studied examples of recent species radiations lack clear evidence on actual species boundaries. Consider the following textbook examples of recent species radiations:

Darwin's Finches

Petren et al. (2005) present the most complete summary to date of species boundaries in several groups of Darwin's finches, including data on 1428 birds from 74 populations for 16 microsatellite loci for the cactus finches (*Geospiza scandens* and *conirostris*), the sharp-beaked ground finch (*G. difficilis*), and the warbler finches (*Certhidea olivacea* and *fusca*). Although these and other species in the group have been extensively analyzed with mitochondrial and nuclear markers (Freeland and Boag, 1999), species boundaries remain fuzzy and ill-defined for many taxa, making these animals "the poster child for foggy species boundaries" (Petren, personal communication, 2006). The most probable explanations for this lack of resolution include hybridization (Grant et al., 2005) and recency of speciation coupled with rapid morphological divergence (Petren et al., 2005). In resolving species boundaries, these authors conclude that additional individual sampling will not help, but additional nuclear genetic data may help clarify these difficult species boundaries.

Cichlid Fishes

Cichlid fishes constitute one of the most famous and widely publicized cases of adaptive speciation and extreme rates of lineage diversification on earth. Yet, cichlid species boundaries have proven remarkably difficult to verify with molecular tools. Recent analyses of mitochondrial and nuclear DNA sequences (Seehausen, 2004) and SINES (Takahashi et al., 2001) as well as older mtDNA analyses (Moran and Kornfield, 1993) indicate that morphologically and ecologically defensible species are often not monophyletic and are frequently characterized by extremely recent divergence. Although hybridization may be further clouding the historical reconstruction of species boundaries (Seehausen, 2004), it remains the case that molecular analyses have yet to clarify species boundaries in these rapidly diversifying fishes.

Galápagos Tortoises

As an icon for speciation, adaptation, and conservation, the Galápagos tortoises are among the most

intensively studied species on earth. With 1611 sequences currently lodged in GenBank (release 161.0, release date 15 August, 2007) for the species *Geochelone nigra* and its 14 subspecies, one might think that the debate over species delimitation would finally be resolved for this endangered tortoise complex. Although currently considered a single species (Pritchard, 1996), recent work utilizing mitochondrial DNA and microsatellites for hundreds of tortoises from across the archipelago (Caccone et al., 2002) and from more restricted sets of localities (Beheregaray et al., 2003; Ciofi et al., 2006) indicated that additional evolutionary lineages may exist. Even after this enormous effort by a talented team of investigators, species boundaries among Galápagos tortoise populations remain controversial and unsettled (Russello et al., 2005).

Tiger Salamanders

Taking an example from our own work, the tiger salamander complex stands out as a recent radiation (estimates range from 3.5 to 5 million years old) with important conservation (Shaffer et al., 2004) and phylogeographic (Shaffer and McKnight, 1996) implications. The ~15-species complex is of interest as a focus of conservation (IUCN, 2006), speciation biology, and because it contains the axolotl (*Ambystoma mexicanum*, see http://salamander.uky.edu/), an important model system in vertebrate developmental biology (Shaffer, 1993). A concerted effort with allozymes (Shaffer, 1984a), morphometrics (Shaffer, 1984b; Irschick and Shaffer, 1997), and mitochondrial DNA sequences (Shaffer and McKnight, 1996) has revealed a complex pattern of extremely rapid speciation events characterized by short internodes and poorly resolved species boundaries. Our most recent analysis of eight nuclear loci derived from an extensive expressed sequence tag (EST) library (Weisrock et al., 2006) demonstrated that whereas some taxa exhibit monophyly for a majority of genes, others fail to do so, leaving species boundaries still very much up in the air.

These and many other examples demonstrate a simple point. Frequently, the molecular data and analytical approaches that have been used to delimit species boundaries in recently radiating lineages have highlighted the difficulty in recovering what appear to be good species, rather than providing unambiguous guidelines for species delimitation. Although this empirical result may stem from several sources, ranging from incorrectly identified species to hybridization to an inappropriate choice of molecular markers, we feel that two issues may plague many studies in the recent literature. First, the reliance on phylogenetic tree–oriented approaches and species concepts (Cracraft, 1989; Shaw, 2001) often requires gene-tree monophyly and complete lineage sorting to recognize species (Baum and Donoghue, 1995). However, recent theoretical work has demonstrated that reciprocal monophyly for a pair of species under the neutral coalescent takes from 4 to 7 N_e generations for 50% of nuclear genes sampled (where N_e is the historically effective population size of each descendant taxon) to 9 to 12 N_e generations for 95% of sampled genes (Hudson and Coyne, 2002). Although historical and current population sizes are seldom known, it seems likely that reciprocal monophyly for most genes will require tens of thousands of generations (and millions may be more common), implying a long waiting time for monophyly to evolve even in completely isolated lineages. A recent review suggested that this long waiting time may be causing a lack of species monophyly even for the rapidly evolving metazoan mitochondrial DNA (Funk and Omland, 2003), and it will be a much more severe problem for nuclear genes because of their lower levels of variation coupled with their larger effective population size (Hudson and Coyne, 2002). Second, most studies have relied on one or a few genes or, in studies that use mitochondrial DNA only, a single linkage block. Given the extremely idiosyncratic nature of gene tree histories (Rosenberg and Nordborg, 2002; Maddison and Knowles, 2006), particularly when populations are structured (Wakeley, 2000; Irwin, 2002), accurate pictures of recent speciation histories will only be possible when integrating over the historical information available from multiple, independent gene histories (Jennings and Edwards, 2005).

Recent advances in population-genetically oriented methodologies combined with the increasingly wide availability of genomic resources including expressed sequence tag (EST), end-sequenced bacterial artificial chromosome (BAC) libraries, and low-to-moderate coverage genome sequences now afford the opportunity to collect large amounts of nucleotide variation data that can be effectively aimed at species delimitation problems across a broad spectrum of species. Obviously, some groups have better developed genomic resources than others, and lineages with few resources may require considerable marker development. However, our own empirical work in developing PCR primers, DNA sequences, and single nucleotide polymorphisms (SNPs) suggests that relatively modest genomic resources may provide a rich source of data, even for taxa that are distantly related to a model species.

Why SNPs?

Single nucleotide polymorphisms are a promising marker-type for species delimitation. SNPs have low rates of substitution (10^{-8} to 10^{-9}) and therefore extremely low levels of homoplasy relative to microsatellites (which have mutation rates on the order of 10^{-4}) at similar levels of divergence (Brumfield et al., 2003). In addition, the mutation rate of microsatellite loci varies across alleles within loci and across loci to a much higher degree than is likely to be observed in SNP loci (Macaubas et al., 1997), which can make it difficult to fit models of molecular evolution to microsatellite data. One limitation of SNPs compared to microsatellites is that they are generally biallelic and therefore contain less information per locus than most microsatellites (Morin et al., 2004). However, this limitation is somewhat offset by the relative ease of scoring many loci for SNPs relative to microsatellites.

SNPs are similar to amplified fragment length polymorphisms (AFLPs) and related marker-types in that they are usually biallelic. However, SNPs are codominant markers whereas AFLPs are dominant (homozygous and heterozygous individuals cannot be distinguished), rendering SNPs more informative per locus, on average, than dominant markers. Most biallelic marker types, aside from SNPs, are anonymous, which makes homoplasy difficult to assess, as loci and alleles are often difficult to distinguish. Some estimates of levels of homoplasy have been made in AFLP data sets. These estimates range from 5% in potatoes (*Solanum*) (vanderVoort et al., 1997) to 100% among species from different subtribes of *Carduinae* thistles (O'Hanlon and Peakall, 2000), with a range of estimates falling between these extremes (O'Hanlon and Peakall, 2000; Mechanda et al., 2004). Finally, reproducibility has been a concern with certain anonymous marker types, particularly randomly amplified polymorphic DNA (RAPD) and, to a lesser extent, AFLPs (Jones et al., 1997). Our feeling is that the relative ease and certainty with which SNPs can be developed, genotyped, and analyzed, coupled with their low cost and low levels of homoplasy, make them an ideal marker choice for species boundary research.

Developing Markers in Nonmodel Systems

The traditional strategy for DNA marker discovery in nonmodel systems has been to use the power of comparative analysis to predict primer sequences in taxa for which there are no available sequence data. This "universal primer" approach (Kocher et al., 1989; Lyons et al., 1997) has been widely applied across the tree of life. In theory, the approach is straightforward—if two or more taxa share a homologous DNA sequence for a 20– to 30–base pair fragment, then the most parsimonious prediction is that their most recent common ancestor, and all descendants of that MRCA, will also share that sequence. When a pair of such conserved sequences is identified, they then comprise a candidate pair of primers that should amplify a homologous DNA fragment for all descendants of that MRCA, and potentially for additional related taxa. This universal primer strategy has been a powerful one for primer discovery for organellar genes, particularly for mitochondrial DNA in metazoans (Simon et al., 1994), chloroplast genes (Taberlet et al., 1991; Demesure et al., 1995), and, to a lesser extent, mitochondrial genes in plants (Demesure et al., 1995; Dumolin-Lapegue et al., 1997).

The universal primer strategy has also been used for nuclear genes, particularly in efforts to locate putative exon-primed, intron-crossing (EPIC) primers. Introduced more than a decade ago (Palumbi and Baker, 1994), and developed for a variety of organisms (Hassan et al., 2002; Creer et al., 2005), the universal EPIC approach has been heralded by some as a strategy that may yield usable sequence across both narrow and broad phylogenetic distances (Aitken et al., 2004), although others have felt that such nuclear primers are likely to be taxonomically restricted and "may not be applicable even in closely related species" (Zhang and Hewitt, 2003).

Here, we explore a somewhat different approach to primer design for gene and marker discovery. One of the limitations of the universal primer approach is that it requires having at least two taxa that span the target species and that have both been sequenced for a homologous region of the genome. In many groups, such comparative data simply do not exist across enough of the genome to allow for the development of many independent markers. As an alternative approach, we explore a strategy of developing primers directly from a single species, rather than identifying conserved primer sequences across divergent taxa. We then ask how those primers perform in increasingly divergent taxa sampled from a known tree. This very direct approach has not been systematically explored, at least to our knowledge, for any clade. We explore this strategy for two reasons. First, it is a simple, straightforward approach to primer discovery that can be accomplished with a single genomic resource for a single species. Second, the possibility exists that primers and the sequences that they span that were discovered based on a universal primer approach may be biased toward relatively conserved sequences simply because locus selection proceeds by picking parts of the genome (primers) that contain conserved sequences. If primers that have long been conserved tend to span genomic regions that are also relatively conservative, this could mean that universally primed sequences are on average somewhat less variable than might be otherwise expected. Because highly variable sites and sequences should be the most useful for delimiting very recent speciation events, we reasoned that primers derived from a single species might avoid such a low-variation bias if it exists.

An Example with Turtles

Turtles are a reasonable case study for marker development with limited genomic information. As a clade, the crown group of living turtles is low-diversity (about 320 living species), deeply divergent (living members span a phylogenetic split of approximately 210 million years (Near et al., 2005)), and has limited genomic resources available. A recent literature review (Engstrom et al., in press) identified 202 mitochondrial primers that span the turtle mitochondrial genome and an additional 181 microsatellite primers but primers for only 11 nuclear loci. In addition, for problems at the within/between species boundary, many of these 11 nuclear loci have been found to be largely uninformative (Caccone et al., 2004; Spinks and Shaffer, 2005).

Currently, a single genome-level resource exists for turtles. Initially constructed as part of the "100 BAC" NSF initiative (http://www.nsf.gov/bio/pubs/awards/bachome.htm), this end-sequenced BAC library consists of 3461 end sequences, each approximately 500 to 800 base pairs in length, for a single, field-collected, western painted turtle (*Chysemys picta*) from the State of Washington. In total, the end-sequenced resource yielded 2,432,811 base pairs of usable sequence

information for this individual turtle. Such a resource base may be paltry by the standards of many model organisms, but it represents a tremendous amount of genomic information for phylogenetic analysis (2.5 megabases of data for one individual). Because the major features of turtle phylogeny are becoming reasonably well known (Krenz et al., 2005; Near et al., 2005; Parham et al., 2006), we were able to choose representative taxa that span recent-to-ancient divergence times with respect to *C. picta* and ask how well primers derived from *Chrysemys* sequence alone work as one moves increasingly large phylogenetic distances from this single species.

We break the problem of going from this BAC resource to a panel of usable markers into the following steps: (1) primer discovery, (2) SNP discovery, and (3) genotyping. Finally, we demonstrate the utility of our strategy by developing a small data set for SNP discovery with an application in a distantly related turtle species.

Primer discovery.—Using the *Chrysemys* BAC end-sequences, we sought to develop a set of 96 potential primer pairs that would provide orthologous DNA sequence information for species delimitation studies across turtles. Each of the 3416 end-sequences was screened for repeat sequences (e.g., known short interspersed elements [SINEs], long interspersed elements [LINEs], and simple repeats) with RepeatMasker (Smit et al., 2004), using *Gallus gallus* as the query species (RepBase update 10.04) and sensitivity/processing time set to "slow" in order to maximize the number of repeats identified. Regions of sequence that were identified as repetitive were masked with Ns and were subsequently removed using a Perl script (available from R.C.T). The aim of this screening step was to isolate sequences that were single copy and thus would allow us to prime orthologous regions of the genome. The remaining end-sequences were selected at random and compared to existing GenBank sequences using BLASTX (http://www.ncbi.nlm.nih.gov/BLAST/). From this group, we chose 48 sequences that showed high similarity (defined as e-value $<10^{-5}$) over at least ~350 bp with existing GenBank sequences, 24 that showed low similarity (e-value $>10^{-5}$), and 24 that showed no similarity (no BLASTX matches), resulting in 96 sequences for primer design. We designed a single primer pair for each of these end-sequences using the program Primer3, setting the optimal primer size to 20 bp and optimal annealing temp to 60°C (Rosen and Skaletsky, 2000). In addition, the "product size ranges" setting was set to favor the largest possible primed region, with the constraint that the entire primed region (including both primer sequences) had to fall within the region of similarity revealed in the BLAST search. All primer sets were designed under the same conditions so that subsequent PCRs could be run as high-throughput batches. Thus, although we used comparative BLASTX results to organize end-sequence products into high, low, and no similarity categories, all actual primers were perfect matches to the *C. picta* BAC turtle.

We screened all 96 primer pairs in 25-μL PCR reactions using standard PCR conditions, AmpliTaq DNA polymerase, and an annealing temperature of 60°C across a panel of five species that provide phylogenetic coverage across the turtle tree of life (Fig. 2). These five taxa included the original painted turtle (*C. picta*) from which the BAC library was constructed (serving as a

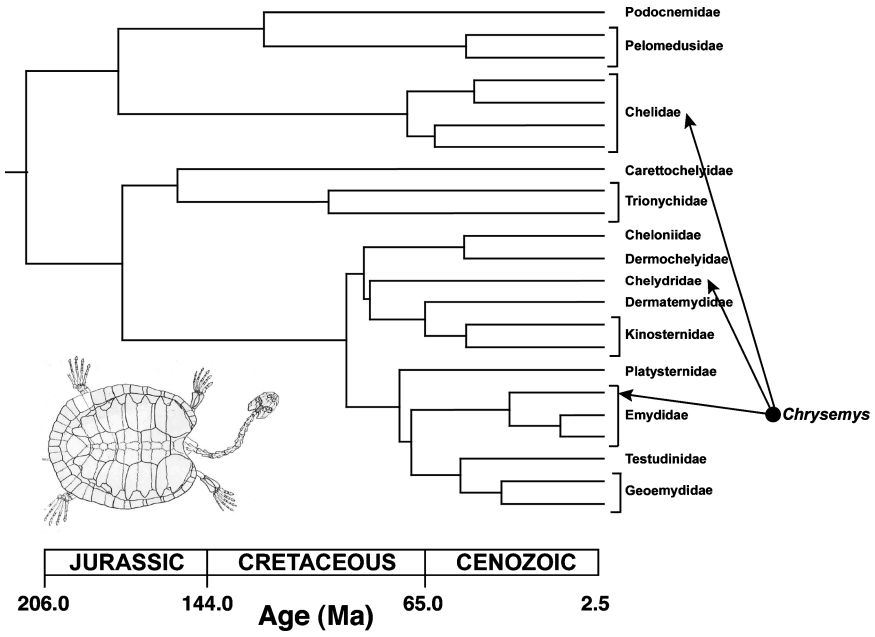

FIGURE 2. A phylogeny of 14 families (23 species) representing the major clades of living turtles. The phylogenetic placement of the BAC turtle, *Chrysemys picta* (Emydidae), and the families to which it is compared in Figure 3 are shown. Figure modified from Near et al. (2005).

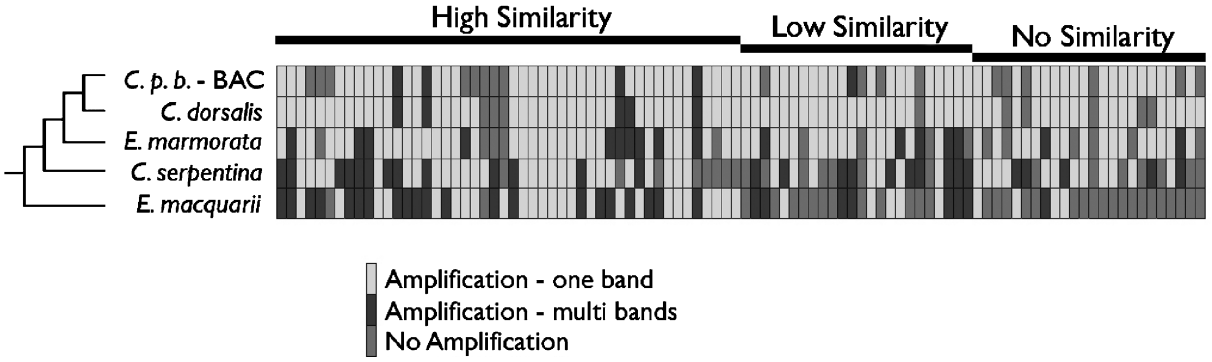

FIGURE 3. Results of 480 PCR reactions for high, low, and no similarity sequences to a phylogenetic range of turtle taxa. Primers were all derived from the BAC turtle, and all amplifications were tried only once at 60°C. Amplifications that yielded a single band of the predicted size are shown in white, amplifications that yielded multiple bands are shown in black, and failed amplifications yielding no visible bands are shown in gray.

positive control, and hereafter referred to as the BAC turtle), a southern painted turtle (variously considered a subspecies of *Chrysemys picta*, *C. p. dorsalis*, or a full species, *C. dorsalis*), a western pond turtle (*Emys* [previously *Clemmys*, now referred to *Actinemys* by some authors] *marmorata*), a snapping turtle (*Chelydra serpentina*), and an Australian side-necked turtle (*Emydura macquarii*) (Turtle Taxonomy Working Group, in press). Based on a reasonably well-understood phylogeny of the major clades of turtles (Krenz et al., 2005; Parham et al., 2006), these taxa span phylogenetic nodes at roughly 2-, 50-, 100-, and 200-million-year intervals with respect to the BAC turtle (Fig. 2). *Chrysemys picta* and *C. dorsalis* have been variously considered conspecific subspecies or distinct species and are estimated to have diverged approximately 2 to 3.5 million years ago (Starkey et al., 2003); other divergence dates are derived from Near et al. (2005).

Each PCR reaction was run once, visualized on 1% agarose gels, and scored as amplifying a single product of the predicted size, multiple products one of which was the predicted size, or no products (Fig. 3). Although limited to a single specimen per species and a single PCR with no optimization, the results suggest that primers chosen from a single BAC individual routinely amplify the predicted product across turtles (Table 1). For 48 "high-similarity" sequences, there was a gradual fall-off from roughly 80% of the primers producing a single amplicon within *Chrysemys* to 50% in *Emydura*. Most of the remaining high similarity primers produced multiple bands, rather than no bands, suggesting that many more sequences could be recovered with additional optimization. For the 24 "low-similarity" sequences, the falloff from *Chrysemys dorsalis* (96%) to *E. macquarii* (21%) for primers that yielded single bands was considerably steeper, with what appears to be a qualitative shift in efficiency between the *Emys marmorata* (a confamilial emydid with *Chrysemys*) and *Chelydra serpentina* (family Chelydridae). The "no-similarity" primers yielded a slightly smaller set of successful amplifications from *Chrysemys* (79%) to *Emydura* (17%). In addition, the fraction of PCR experiments that yielded no amplification increased dramatically for the no-similarity primers as a function of phylogenetic divergence (Fig. 3, Table 1).

In summary, these experiments suggest that primers developed from a single-species genomic resource like a BAC, EST, or whole genomic library can be a rich source of comparative data across large phylogenetic distances. Perhaps not surprisingly, this appears to be a particularly strong result for sequences with high sequence similarity to outgroup taxa. However, even sequences with no BLAST hits to Genbank, which presumably represent either noncoding or unique turtle sequences, produced a single amplicon in 50% or more of primer pairs out to the snapping turtle, suggesting that this is also a reasonably efficient strategy for marker discovery in less conserved genomic regions.

TABLE 1. The percentage of primers designed from a single-specimen BAC library that amplify across the turtle tree of life. Entries for time are the estimated age of the most recent common ancestor of the species and the BAC turtle (Near et al., 2005). Table entries are the percentage of primers, developed from the BAC turtle, that yielded single, multiple, or no amplicons across species. For definitions of high, low, and no similarity, see text.

		High similarity			Low similarity			No similarity		
	Time (Ma)	Single	Multiple	No	Single	Multiple	No	Single	Multiple	No
C. picta (BAC)	0	75	8	17	79	4	17	75	0	25
C. dorsalis	2–3.5	83	10	6	96	0	4	79	0	21
E. marmorta	34	69	21	10	62	21	17	67	8	25
C. serpentina	94	56	27	17	29	42	29	50	21	29
E. macquarii	210	52	44	4	21	29	50	17	4	79

SNP discovery.—To convert primers that amplify at the PCR level into SNP markers that may be useful for species delimitation, we employed a discovery panel approach in which SNP loci are found by resequencing a subset of the total number of individuals in the study, followed by sequence alignment and "tagging" (identification) of polymorphic sites. The makeup of a discovery panel will vary from study to study and depends on the specific questions being asked. If a study seeks to test the concordance of species taxa with real lineages (Hey and Machado, 2003), then the discovery panel should include individuals broadly representative of the current species taxonomy (at least one member of every named species, subspecies, or other hypothesis being tested). Alternatively, for studies that seek to delimit species without respect to current taxonomy (or when no species level taxonomy exists), we favor discovery panels composed of natural sampling units (e.g., individuals from all allopatric parts of the range) that are also geographically representative of the overall range of the populations of interest. Although increasing the number of individuals included in the discovery panel will increase the number of polymorphic sites that are identified, the cost savings provided by the SNP approach (relative to sequencing all individuals) will be reduced. Also, because multiple SNPs derived from the same sequence are expected to be tightly linked, generally one need only include enough individuals on the discovery panel to discover one SNP per sequence locus when methods call for unlinked marker data. The number of individuals required to accomplish this will vary with the level of variation present in the sequence data. When multiple SNPs are identified in a single sequence, we favor choosing one at random to reduce the bias that can result if, for example, maximally differentiated SNPs are chosen instead.

Several software packages exist that can assist in the tagging of polymorphic sites in sequence data. Given the challenge of distinguishing true heterozygotes from ambiguous/poor quality reads, this step is actually more challenging than one might expect. PolyPhred is one package that integrates with the Phred/Phrap/Consed genome assembly package and that can recognize the heterozygous positions that are routinely encountered when sequencing diploid loci (Gordon et al., 1998; Stephens et al., 2006). This software can also rank heterozygous sites according to their "quality," as measured by the fit of putatively heterozygous sites to the ideal pattern for a SNP in that sequence data. By using these and similar tools, researchers can quickly locate the highest quality (and therefore, most likely) polymorphic positions in their sequence data, as well as judge whether they have included enough individuals in their discovery panel to confidently call heterozygous positions.

Genotyping.—Genotyping strategies tend to center on two issues: accuracy and cost. At one extreme, one can simply sequence each individual and treat haplotypes, or contained polymorphisms, as characters for species delimitation. Sequences tend to be accurate but expensive (currently averaging several dollars per sequence at most facilities). At the other extreme are tiling microarrays (Mockler and Ecker, 2005), where tens of thousands of SNPs can be scored from an individual for pennies per genotype, assuming that the genomic resources are available for the target species (unlikely for most species delimitation problems). In between are strategies ranging from low-throughput, relatively high-cost approaches like single-base extension of individual SNPs from PCR products (Hsu et al., 2001) to the Illumina Beadstation that can provide >1000 simultaneous SNP genotypes/individual for a fraction of the cost. Our goal is not to review these platforms but to simply point out that a range of options, costs, and levels of automation exist that should allow most research groups to accurately genotype large numbers of individuals for multiple SNPs at a reasonable cost. Even the conversion of SNPs into restriction fragment length polymorphisms has its place as a low-tech genotyping strategy (Fitzpatrick and Shaffer, 2004), although it is relatively inefficient compared to most other approaches.

SNP discovery in a distantly related clade.—As a test of our BAC-derived primer design strategy, we amplified, sequenced, and identified SNPs for four loci, across two populations, for the distantly related pleurodire *Emydura macquarii* (family Chelidae). This species complex occurs throughout eastern, central, and northern Australia and southern New Guinea and presents a challenging case for species delimitation. Two recent analyses of the Australian populations have recognized a single species (Georges and Adams, 1996) or a set of at least seven species and numerous additional subspecies (Cann, 1998). Because the primers in Figure 3 were designed from the BAC *Chrysemys picta* sequence data, one might expect them to effectively identify polymorphic markers for this and closely related species. Our interest here was to ask whether PCR primers derived from a single species yield informative markers for species delimitation even in distantly related taxa.

Our sequencing panel included five individuals from each of two adjacent river drainages in central Australia, the Cooper and Murray-Darling. *Emydura* from these two drainages are currently referred to the subspecies *E. macquarii emmotti* and *E. m. krefftii*, respectively, although considerable uncertainty over species delimitation exists (Turtle Taxonomy Working Group, in press). We attempted to sequence these 10 individuals for four loci that had yielded single bands at the PCR level. Of these four loci, three yielded useable sequence, each of which contained a SNP every ~250 bp (Table 2). Although clearly limited and very preliminary, these results demonstrate that our strategy appears to effectively identify loci, even in distantly related taxa, that can provide useful variation for defining species boundaries. For example, locus TB73 has a potentially fixed difference between *Emydura* from these two river systems at positions 112 and 481, and potential SNP frequency differences and/or river-specific alleles exist at most other variable positions.

TABLE 2. SNP loci identified by resequencing five *Emydura macquarii* from each of two river systems. Locus refers to the primer pairs used for sequencing and position refers to the base pair position of each SNP locus within that sequencing read. Alternative bases (e.g., A/G) refer to heterozygous individuals and question marks denote missing data for one individual due to low sequence quality.

Locus	TB69		TB73			TB81	
Position	339	397	112	183	481	393	415
Cooper system							
HBS 101421	A	?	?	?	C	C/A	G/A
HBS 101243	G	C/T	G	A	C	C/A	G/A
HBS 101363	G	T	G	A	C	A	G
HBS 101238	G	C/T	G	A	C	C/A	G/A
HBS 101365	G	C/T	G	A	C	C	G/A
Murray-Darling system							
HBS 101483	A	C	A	A	G	C	A
HBS 101494	A	C	A	A	G	C	A
HBS 101489	G/A	C	A	G/A	G	C	A
HBS 101491	G	C	A	G/A	G	C	A
HBS 101495	G/A	C	A	A	G	C	A

SPECIES DELIMITATION

Even if researchers agree on a species concept, there invariably are multiple strategies for delimiting species using the same data sets (Wiens and Penkrot, 2002; Sites and Marshall, 2003, 2004), and these strategies may lead different researchers to different conclusions (Marshall et al., 2006). Our own position is that when species are old and well-differentiated (the "easy" part of Fig. 1), most methods and most data will lead most researchers to the same species delimitation conclusions. When species are young, delimitation is more difficult, both conceptually and empirically. For this part of the empirical spectrum, Hey et al. (2003) summarized the situation nicely. "In short, species entities are very difficult to study, for they are evolutionarily and demographically dynamic. They will often not be very distinct, and the degree to which they are distinct can change over time…" (Hey et al., 2003: 599). How, then, can one use SNPs or other genotypic data to delimit recently derived, closely related species?

We follow de Queiroz (1998, 2005) in defining species as separately evolving metapopulation lineages. That is, a species is a demographically and genetically interconnected set of populations (metapopulation) that has continuity through time (Simpson, 1951; Mayden, 1997). Although this definition does not provide an unambiguous cutoff for when speciation has occurred, it emphasizes that the primary goal of species delimitation research is lineage discovery and delimitation. Such lineage discovery is neither straightforward nor precise in the earliest stages of speciation, and we must expect some fuzziness and uncertainty in delimiting young species in the early stages of differentiation (de Queiroz, 1998; Hey et al., 2003). Under the neutral coalescent the expected time to monophyly for even a single nuclear gene is often long (Hudson and Coyne, 2002), suggesting that many real species exist that have not yet achieved monophyly for even a few genes.

The metapopulation lineage species definition naturally leads to operational species delimitation approaches that recognize, across multiple loci, sets of populations that freely exchange genes in nature, and have severely restricted or no gene exchange with other sets of populations. Whether these conditions exist due to intrinsic features of organisms (behavioral or physiological reproductive isolating mechanisms) or extrinsic features of their distribution across the landscape (an island population separated from its mainland counterparts) is immaterial, since both produce distinct lineages. Rather, the critical point is that evidential support exists for species, and (in our opinion) that there is enough redundancy in that evidence to ensure that metapopulation lineages, rather than allele frequencies shaped by natural selection gradients or idiosyncratic histories of individual loci, are the best explanation of the data in hand. If two or more loci indicate that lineages are distinct from each other, those metapopulation lineages become candidates for species recognition. A single locus, even if fixed for different states, provides a testable hypothesis regarding species distinction but not, in our minds, unambiguous evidence for species recognition. This reliance on multiple markers implies that species in their very earliest stages will probably not be recognized, but it is the price one pays for the taxonomic stability that follows from only recognizing comparatively well-differentiated species.

Two fundamentally different strategies, aimed at very different research agendas, exist within what we refer to as species delimitation research. The first is the relatively exploratory phase where one asks whether a set of samples comprise one or many species, with no a priori hypothesis of what those species might be. Examining all possible species among a set of individuals is complex combinatorial problem, and we think that a satisfying and successful approach is implemented in STRUCTURE (Pritchard et al., 2000; Falush et al., 2003). STRUCTURE seeks to cluster individuals without regard to their population-of-origin based on rough conformity to Hardy-Weinberg genetic expectations. The strategy employed by STRUCTURE is straightforward and matches reasonably well the properties of metapopulation lineages. The primary assumptions of the model are Hardy-Weinberg equilibrium within populations (or metapopulations) and linkage equilibrium between loci, although the latter assumption has been relaxed (Falush et al., 2003). The program then seeks sets of individuals that are in approximate multilocus Hardy-Weinberg equilibrium, searching parameter space sequentially across an arbitrary number of K taxa. A Markov chain Monte Carlo (MCMC) search strategy simultaneously estimates allele frequencies and population of origin for each potential K (Pritchard et al., 2000; Falush et al., 2003), and the method can identify admixed (that is, mixed-origin) individuals between otherwise differentiated populations (Falush et al., 2003). Although originally developed for population identification, STRUCTURE has performed well in the identification of domestic strains of chickens, suggesting that it may serve as a useful tool for the initial stages of lineage diversification (Rosenberg et al., 2001). Our sense is

that STRUCTURE may be most useful in determining the lower bound of potential species; if units are not identified by this approach, they lack the evidential support to be even candidates for species recognition. Additional programs that are relevant to identifying the early phases of lineage diversification continue to be developed, and should provide additional insights into the initial diversification of species (Corander et al., 2004; Guillot et al., 2005).

Alternatively, a priori named species (species taxa) may be tested with genotypic data (Wiens and Penkrot, 2002; Hey et al., 2003; Sites and Marshall, 2004). Here, the delimitation problem is much more prescribed, since one is asking whether two or more sets of individuals should be considered separate species. When species are very young, the hybrid zone barrier (HZB) approach (Porter, 1990) is appealing in its simplicity—essentially it calculates Wright's F_{st} values between sets of species taxa, converts them to Nm values, and argues from population genetic theory that when $0.0 < Nm < 0.5$, two populations are strong potential candidates for species recognition. Like STRUCTURE, the HZB approach may recognize taxa that are insufficiently differentiated for many practitioners to feel that they warrant species recognition, and the suggested cutoff for Nm is somewhat arbitrary. The extent to which STRUCTURE and HZB tend to identify the same sets of potential species has not been tested, although such a test would be very interesting.

Among the most recently proposed methods, we feel that the emerging parametric approaches that simultaneously estimate ancestral population size, migration rate and divergence times are particularly appealing for testing species taxa. The isolation-with-migration (Nielsen and Wakeley, 2001; Hey and Nielsen, 2004) model considers an important set of conditions for the earliest phases of the species delimitation problem—populations that have diverged but may still incur limited, ongoing gene flow (or gene flow in the recent past). Based on MCMC sampling of parameter space, IM works with multilocus data (sequences and microsatellites) from a pair of species taxa and provides important insights into the earliest stages of differentiation and speciation. This approach attempts to fit a six-parameter model of the population divergence process to the data and can provide information about time since divergence, ongoing migration, and population size. Limitations of this parameter-rich model include the requirement of a large amount of data in order to obtain robust parameter estimates. Additionally, run times can be extremely long (on the order of months) and the current software implementation of the IM model does not include a model of molecular evolution for SNPs. Although it provides no prescribed cutoff value for species recognition, Hey and Nielsen (2004) provide an interesting example of how IM can be used to gain deeper insights into the history of speciation in *Drosophila pseudoobscura* and *D. persimilis*. We view this approach as an important melding of population genetics and speciation biology that is central to the metapopulation lineage concept.

CONCLUSIONS

Although the empirical study of species boundaries is enjoying renewed interest in the systematics community (Wiens and Penkrot, 2002; Sites and Marshall, 2004), significant challenges remain in the empirical delimitation of species boundaries in recent radiations of closely related species (Hey et al., 2003). This stems from several sources, including the historical reliance on a few genomic regions, the limited number of available markers for most taxa, and a paucity of appropriate methods for the analysis of these data. The approach that we have outlined appears to be an effective strategy for quickly identifying many markers, at least for clades where some genomic resources are available. Both tree-based and non–tree-based methods of analysis can and should contribute to further empirical progress in delimitation of recent species, and we discuss several population genetic techniques that can help identify species lineages that have not had time to sort into monophyletic groups. As genomic resources continue to become available for diverse taxa drawn from the tree of life, we look forward to finally bringing the appropriate data to bear on "the species problem" in recent species radiations that have long fascinated, and frustrated, evolutionary biologists.

ACKNOWLEDGMENTS

We thank Scott Edwards and Andy Shedlock for access to the *Chrysemys* BAC end-sequence data; Bob Macey and the JGI for DNA from the BAC turtle; Phil Spinks for help with sequencing and troubleshooting; Arthur Georges, Louise Whitaker, and Rob Shaffer for field assistance in Australia; and John Wiens for organizing the symposium and critically reading the paper. This material is based upon work supported in part by the NSF, the University of California Agricultural Experiment Station, and the UC Davis Center for Population Biology.

REFERENCES

Aitken, N., S. Smith, C. Schwarz, and P. A. Morin. 2004. Single nucleotide polymorphism (SNP) discovery in mammals: A targeted-gene approach. Mol. Ecol. 13:1423–1431.

Avise, J. C., and D. E. Walker. 1999. Species realities and numbers in sexual vertebrates: Perspectives from an asexually transmitted genome. Proc. Natl. Acad. Sci. USA 96:992–995.

Baum, D. A., and M. J. Donoghue. 1995. Choosing among alternative "phylogenetic" species concepts. Syst. Bot. 20:560–573.

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pages 289–303 *in* Monographs in systematic botany; Experimental and molecular approaches to plant biosystematics, 53 (P. C. Hoch and A. G. Stephenson, eds.). Missouri Botanical Garden, St. Louis, Missouri.

Beheregaray, L. B., C. Ciofi, A. Caccone, J. P. Gibbs, and J. R. Powell. 2003. Genetic divergence, phylogeography and conservation units of giant tortoises from Santa Cruz and Pinzon, Galapagos Islands. Conserv. Genet. 4:31–46.

Brumfield, R. T., P. Beerli, D. A. Nickerson, and S. V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. Evol. 18:249–256.

Caccone, A., G. Gentile, C. E. Burns, E. Sezzi, W. Bergman, M. Ruelle, K. Saltonstall, and J. R. Powell. 2004. Extreme difference in rate of mitochondrial and nuclear DNA evolution in a large ectotherm, Galapagos tortoises. Mol. Phylogen. Evol. 31:794–798.

Caccone, A., G. Gentile, J. P. Gibbs, T. H. Fritts, H. L. Snell, J. Betts, and J. R. Powell. 2002. Phylogeography and history of giant Galapagos tortoises. Evolution 56:2052–2066.

Cann, J. 1998. Australian freshwater turtles. Beaumont Publishing, Singapore.

Ciofi, C., G. A. Wilson, L. B. Beheregaray, C. Marquez, J. P. Gibbs, W. Tapia, H. L. Snell, A. Caccone, and J. R. Powell. 2006. Phylogeographic history and gene flow among giant Galapagos tortoises on southern Isabela Island. Genetics 172:1727–1744.

Corander, J., P. Waldmann, P. Marttinen, and M. J. Sillanpaa. 2004. BAPS 2: Enhanced possibilities for the analysis of genetic population structure. Bioinformatics 20:2363–2369.

Cracraft, J. 1989. Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. Pages 28–59 *in* Speciation and its consequences (D. Otte, and J. A. Endler, eds.). Sinauer Associates, Sunderland, Massachusetts.

Creer, S., A. Malhotra, R. S. Thorpe, and C. E. Pook. 2005. Targeting optimal introns for phylogenetic analyses in non-model taxa: Experimental results in Asian pitvipers. Cladistics 21:390–395.

de Queiroz, K. 1998. The general lineage concept of species. Pages 57–75 *in* Endless forms: Species and speciation (D. J. Howard, and S. H. Berlocher, eds.). Oxford University Press, New York.

de Queiroz, K. 2005. Ernst Mayr and the modern concept of species. Proc. Natl. Acad. Sci. USA 102:6600–6607.

Degnan, J. H., and N. A. Rosenberg. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:762–768.

Demesure, B., N. Sodzi, and R. J. Petit. 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4:129–131.

Dumolin-Lapegue, S., M. H. Pemonge, and R. J. Petit. 1997. An enlarged set of consensus primers for the study of organelle DNA in plants. Mol. Ecol. 6:393–397.

Engstrom, T. N., T. Edwards, M. F. Osentoski, and E. M. Myers. 2007. A compendium of PCR primers for mtDNA, microsatellite and other nuclear loci for freshwater turtles and tortoises. *in* Defining turtle diversity: Proceedings of a workshop on genetics, ethics, and taxonomy of freshwater turtles and tortoises. (H. B. Shaffer, A. Georges, N. N. FitzSimmons, and A. G. J. Rhodin, eds.). Chelonian Research Foundation, Lunenberg, Massachusetts.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Fitzpatrick, B. M., and H. B. Shaffer. 2004. Environment-dependent admixture dynamics in a tiger salamander hybrid zone. Evolution 58:1282–1293.

Freeland, J. R., and P. T. Boag. 1999. The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. Evolution 53:1553–1563.

Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 34:397–423.

Georges, A., and M. Adams. 1996. Electrophoretic delineation of species boundaries within the short-necked freshwater turtles of Australia (Testudines: Chelidae). Zool. J. Linn. Soc. 118:241–260.

Gordon, D., C. Abajian, and P. Green. 1998. Consed: A graphical tool for sequence finishing. Genome Res. 8:195–202.

Grant, P. R., B. R. Grant, and K. Petren. 2005. Hybridization in the recent past. Am. Nat. 166:56–67.

Guillot, G., F. Mortier, and A. Estoup. 2005. GENELAND: A computer package for landscape genetics. Mol. Ecol. Notes 5:712–715.

Hassan, M., C. Lemaire, C. Fauvelot, and F. Bonhomme. 2002. Seventeen new exon-primed intron-crossing polymerase chain reaction amplifiable introns in fish. Mol. Ecol. Notes 2:334–340.

Hey, J., and C. A. Machado. 2003. The study of structured populations—New hope for a difficult and divided science. Nat. Rev. Genet. 4:535–543.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760.

Hey, J., R. S. Waples, M. L. Arnold, R. K. Butlin, and R. G. Harrison. 2003. Understanding and confronting species uncertainty in biology and conservation. Trends Ecol. Evol. 18:597–603.

Hsu, T. M., X. Chen, S. Duan, R. D. Miller, and P. Y. Kwok. 2001. Universal SNP genotyping assay with fluorescence polarization detection. BioTechniques 31:560–568.

Hudson, R. R., and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565.

Irschick, D. J., and H. B. Shaffer. 1997. The polytypic species revisited: Morphological differentiation among tiger salamanders (*Ambystoma tigrinum*) (Amphibia: Caudata). Herpetologica 53:30–49.

Irwin, D. E. 2002. Phylogeographic breaks without geographic barriers to gene flow. Evolution 56:2383–2394.

IUCN. 2006. IUCN Red List of Threatened Species.www.iucnredlist.org.

Jennings, W. B., and S. V. Edwards. 2005. Speciational history of Australian grass finches (*Poephila*) inferred from thirty gene trees. Evolution 59:2033–2047.

Jones, C. J., K. J. Edwards, S. Castaglione, M. O. Winfield, F. Sala, C. vandeWiel, G. Bredemeijer, B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda, R. Linacero, A. Vazquez, and A. Karp. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breed. 3:381–390.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Paabo, F. X. Villablanca, and A. C. Wilson. 1989. Dynamics of mitochondrial-DNA evolution in animals—Amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86:6196–6200.

Krenz, J. G., G. J. P. Naylor, H. B. Shaffer, and F. J. Janzen. 2005. Molecular phylogenetics and evolution of turtles. Mol. Phylogen. Evol. 37:178–191.

Kubatko, L. S., and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Liti, G., D. B. H. Barton, and E. J. Louis. 2006. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. Genetics 174:839–850.

Lyons, L. A., T. F. Laughlin, N. G. Copeland, N. A. Jenkins, J. E. Womack, and S. J. Obrien. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat. Genet. 15:47–56.

Macaubas, C., L. Jin, J. Hallmayer, A. Kimura, and E. Mignot. 1997. The complex mutation pattern of a microsatellite. Genome Res. 7:635–641.

Maddison, W. P., and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.

Marshall, J. C., E. Arevalo, E. Benavides, J. L. Sites, and J. W. Sites. 2006. Delimiting species: Comparing methods for mendelian characters using lizards of the *Sceloporus grammicus* (Squamata: Phrynosomatidae) complex. Evolution 60:1050–1065.

Mayden, R. L. 1997. A hierarchy of species concepts: The denouement in the saga of the species problem. Pages 381–424 *in* Species (M. F. Clardge, H. A. Dawah, and M. R. Wilson, eds.). Chapman and Hall, London.

Mechanda, S. M., B. R. Baum, D. A. Johnson, and J. T. Arnason. 2004. Sequence assessment of comigrating AFLP (TM) bands in *Echinacea*—Implications for comparative biological studies. Genome 47:15–25.

Mockler, T. C., and J. R. Ecker. 2005. Applications of DNA tiling arrays for whole-genome analysis. Genomics 85:1–15.

Moran, P., and I. Kornfield. 1993. Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei, Cichlidae) of Lake Malawi. Mol. Biol. Evol. 10:1015–1029.

Morin, P. A., G. Luikart, and R. K. Wayne. 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19:208–216.

Near, T. J., P. A. Meylan, and H. B. Shaffer. 2005. Assessing concordance of fossil calibration points in molecular clock studies: An example using turtles. Am. Nat. 165:137–146.

Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. Genetics 158:885–896.

O'Hanlon, P. C., and R. Peakall. 2000. A simple method for the detection of size homoplasy among amplified fragment length polymorphism fragments. Mol. Ecol. 9:815–816.

Palumbi, S. R., and C. S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol. Biol. Evol. 11:426–435.

Parham, J. F., C. R. Feldman, and J. L. Boore. 2006. The complete mitochondrial genome of the enigmatic bigheaded turtle (*Platysternon*): Description of unusual genomic features and the reconciliation of

phylogenetic hypotheses based on mitochondrial and nuclear DNA. BMC Evol. Biol. 6.
Petren, K., P. R. Grant, B. R. Grant, and L. F. Keller. 2005. Comparative landscape genetics and the adaptive radiation of Darwin's finches: The role of peripheral isolation. Mol. Ecol. 14:2943–2957.
Porter, A. H. 1990. Testing nominal species boundaries using gene flow statistics—The taxonomy of 2 hybridizing Admiral butterflies (*Limenitis*, Nymphalidae). Syst. Zool. 39:131–147.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.
Pritchard, P. C. A. 1996. The Galapagos tortoises: Nomenclatural and survival status. Chelonian Res. Monogr. 1:1–85.
Rosen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. Pages 365–386 *in* Bioinformatics methods and protocols: Methods in molecular biology (S. Krawetz, and S. Misener, eds.). Humana Press, Totowa, New Jersey.
Rosenberg, N. A., T. Burke, K. Elo, M. W. Feldmann, P. J. Freidlin, M. A. M. Groenen, J. Hillel, A. Maki-Tanila, M. Tixier-Boichard, A. Vignal, K. Wimmers, and S. Weigend. 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159:699–713.
Rosenberg, N. A., and M. Nordborg. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. Genet. 3:380–390.
Russello, M. A., S. Glaberman, J. P. Gibbs, C. Marquez, J. R. Powell, and A. Caccone. 2005. A cryptic taxon of Galapagos tortoise in conservation peril. Biol. Lett. 1:287–290.
Seehausen, O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.
Shaffer, H. B. 1984a. Evolution in a paedomorphic lineage. I. An electrophoretic analysis of the Mexican ambystomatid salamanders. Evolution 38:1194–1206.
Shaffer, H. B. 1984b. Evolution in a paedomorphic lineage. II. Allometry and form in the Mexican ambystomatid salamanders. Evolution 38:1207–1218.
Shaffer, H. B. 1993. Phylogenetics of model organisms: The laboratory axolotl, *Ambystoma mexicanum*. Syst. Biol. 42:508–522.
Shaffer, H. B., and M. L. McKnight. 1996. The polytypic species revisited: Genetic differentiation and molecular phylogenetics of the tiger salamander *Ambystoma tigrinum* (Amphibia: Caudata) complex. Evolution 50:417–433.
Shaffer, H. B., G. B. Pauly, J. C. Oliver, and P. C. Trenham. 2004. The molecular phylogenetics of endangerment: cryptic variation and historical phylogeography of the California tiger salamander, *Ambystoma californiense*. Mol. Ecol. 13:3033–3049.
Shaw, K. L. 2001. The genealogical view of speciation—Commentary. J. Evol. Biol. 14:880–882.
Shaw, K. L., and P. D. Danley. 2003. Behavioral genomics and the study of speciation at a porous species boundary. Zoology 106: 261–273.
Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Ann. Entomol. Soc. Am. 87:651–701.
Simpson, G. G. 1951. The species concept. Evolution 5:285–298.
Sites, J. W., and J. C. Marshall. 2003. Delimiting species: A Renaissance issue in systematic biology. Trends Ecol. Evol. 18:462–470.
Sites, J. W., and J. C. Marshall. 2004. Operational criteria for delimiting species. Annu. Rev. Ecol. Evol. Syst. 35:199–227.
Smit, A. F. A., R. Hubley, and P. Green. 2004. RepeatMasker Open-3.0.
Spinks, P. Q., and H. B. Shaffer. 2005. Range-wide molecular analysis of the western pond turtle (Emys marmorata): Cryptic variation, isolation by distance, and their conservation implications. Mol. Ecol. 14:2047–2064.
Starkey, D. E., H. B. Shaffer, R. L. Burke, M. R. J. Forstner, J. B. Iverson, F. J. Janzen, A. G. J. Rhodin, and G. R. Ultsch. 2003. Molecular systematics, phylogeography, and the effects of Pleistocene glaciation in the painted turtle (*Chrysemys picta*) complex. Evolution 57:119–128.
Stephens, M., J. S. Sloan, P. D. Robertson, P. Scheet, and D. A. Nickerson. 2006. Automating sequence-based detection and genotyping of SNPs from diploid samples. Nat. Genet. 38:375–381.
Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. Plant Mol. Biol. 17:1105–1109.
Takahashi, K., Y. Terai, M. Nishida, and N. Okada. 2001. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. Mol. Biol. Evol. 18:2057–2066.
Turtle Taxonomy Working Group. [Bickham, J. W., Iverson, J. B., Parham, J. F., Philippen, H. D., Rhodin, A. G. J., Shaffer, H. B., Spinks, P. Q., and van Dijk, P. P.] 2007. An annotated list of modern turtle terminal taxa, with comments on areas of taxonomic instability and recent change. *in* Defining turtle diversity: Proceedings of a workshop on genetics, ethics, and taxonomy of freshwater turtles and tortoises (H. B. Shaffer, A. Georges, N. FitzSimmons, and A. G. J. Rhodin, eds.). Chelonian Research Foundation, Lunenberg, Massachusetts.
vanderVoort, J., P. vanZandvoort, H. J. vanEck, R. T. Folkertsma, R. C. B. Hutten, J. Draaistra, F. J. Gommers, E. Jacobsen, J. Helder, and J. Bakker. 1997. Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol. Gen. Genet. 255:438–447.
Wakeley, J. 2000. The effects of subdivision on the genetic divergence of populations and species. Evolution 54:1092–1101.
Weisrock, D. W., H. B. Shaffer, B. L. Storz, S. R. Storz, and S. R. Voss. 2006. Multiple nuclear gene sequences identify phylogenetic species boundaries in the rapidly radiating clade of Mexican ambystomatid salamanders. Mol. Ecol. 15:2489–2503.
Wiens, J. J., and T. A. Penkrot. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Syst. Biol. 51:69–91.
Zhang, D. X., and G. M. Hewitt. 2003. Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. Mol. Ecol. 12:563–584.

First submitted 15 December 2006; reviews returned 19 March 2007; final acceptance 14 August 2007
Guest Associate Editor: John Wiens