Molecular Ecology (2006) **15**, 2489–2503 doi: 10.1111/j.1365-294X.2006.02961.x

Multiple nuclear gene sequences identify phylogenetic species boundaries in the rapidly radiating clade of Mexican ambystomatid salamanders

DAVID W. WEISROCK,* H. BRADLEY SHAFFER,† BRIAN L. STORZ,†‡ SHONNA R. STORZ†§ and S. RANDAL VOSS*

*Department of Biology, University of Kentucky, Lexington, KY 40506, USA, †Section of Evolution and Ecology and Center for Population Biology, University of California, Davis, CA 95616, USA, ‡Department of Biological Sciences, Florida State University, Tallahassee, FL 32306, USA, §Department of Nutrition, Food and Exercise Sciences, Florida State University, Tallahassee, FL 32306, USA

Abstract

Delimiting the boundaries of species involved in radiations is critical to understanding the tempo and mode of lineage formation. Single locus gene trees may or may not reflect the underlying pattern of population divergence and lineage formation, yet they constitute the vast majority of the empirical data in species radiations. In this study we make use of an expressed sequence tag (EST) database to perform nuclear (nDNA) and mitochondrial (mtDNA) genealogical tests of species boundaries in *Ambystoma ordinarium*, a member of an adaptive radiation of metamorphic and paedomorphic salamanders (the *Ambystoma tigrinum* complex) that have diversified across terrestrial and aquatic environments. Gene tree comparisons demonstrate extensive nonmonophyly in the mtDNA genealogy of *A. ordinarium*, while seven of eight independent nuclear loci resolve the species as monophyletic or nearly so, and diagnose it as a well-resolved genealogical species. A differential introgression hypothesis is supported by the observation that western *A. ordinarium* localities contain mtDNA haplotypes that are identical or minimally diverged from haplotypes sampled from a nearby paedomorphic species, *Ambystoma dumerilii*, while most nDNA trees place these species in distant phylogenetic positions. These results provide a strong example of how historical introgression can lead to radical differences between gene trees and species histories, even among currently allopatric species with divergent life history adaptations and morphologies. They also demonstrate how EST-based nuclear resources can be used to more fully resolve the phylogenetic history of species radiations.

Keywords: adaptive radiation, *Ambystoma*, EST, gene tree, introgression, lineage sorting, paedomorphosis

Received 22 November 2005; revision accepted 2 March 2006

Introduction

When large numbers of related species originate over a short time period it is called a radiation (Simpson 1953; Gittenberger 1991; Schluter 2000). Although widely (and correctly) viewed as one of the cornerstones of evolutionary biology, species radiations offer special challenges in phylogenetic species delimitation. Because of the rapid and often recent nature of speciation, intraspecific gene trees are often resolved as nonmonophyletic, complicating the delimitation of phylogenetic species. Darwin's finches (Freeland & Boag 1999; Sato *et al*. 1999), East African cichlids (Parker & Kornfield 1997; Nagl *et al*. 1998; Ruber *et al*. 2001), and tiger salamanders (Shaffer 1984a, b; Shaffer & McKnight 1996) are all examples of well-studied radiations where species are often resolved as nonmonophyletic when one or a few loci are used to reconstruct evolutionary history. While no consensus has yet emerged on how to deal with rampant nonmonophyly of rapidly radiating species, there is a clear need to average the signal across multiple genes to gain insights into species boundaries using phylogenetic criteria (Avise & Ball 1990; Baum & Shaw 1995).

Correspondence: David W. Weisrock, Fax: (859) 257-1717; E-mail: weisrock@uky.edu

There are several nonmutually exclusive explanations for the nonmonophyly of gene genealogies within species drawn from radiations (Patton & Smith 1994; Maddison 1997; Funk & Omland 2003). Here, we focus on the two most commonly encountered problems: incomplete lineage sorting and hybridization. Lineage sorting is the stochastic process of eliminating ancestral allelic copies in descendent species. The shorter the time between speciation events, the greater the probability that nonmonophyletic alleles will be carried into populations of a newly formed species (Pamilo & Nei 1988; Maddison 1997; Poe & Chubb 2004). If speciation events are rapid and ancestral population sizes are large, incomplete lineage sorting can lead to apparent nonmonophyly of good species (including good genealogical or biological species). Hybridization (current or ancient) between species may also yield nonmonophyletic genealogies within species. For recently diverged species this may be especially important, since sexual isolation takes time to evolve, particularly in allopatric taxa (Coyne & Orr 1989). Although reproductive isolation among *adaptively* radiated species may evolve more quickly (Schluter 2001; Fitzpatrick 2002), even infrequent episodes of successful hybridization can be sufficient for horizontal movement of both neutral and selected alleles (Chan & Levin 2005).

The *Ambystoma tigrinum* complex is a recently derived and diverse group of salamanders found throughout North America from southern Canada to central Mexico. Ecological differentiation across aquatic and terrestrial habitats has yielded extreme phenotypic variation in the group, especially in life history characteristics. Some species exhibit complete metamorphosis to a terrestrial adult stage, others attain sexual maturity as larval-like aquatic adults (paedomorphosis), and still others can facultatively express either state (Shaffer 1984a; Shaffer & Voss 1996). Whereas metamorphic and facultative taxa are capable of dispersal among isolated, aquatic breeding habitats, obligatorily paedomorphic taxa are generally confined to single bodies of water. Work over the last two decades has demonstrated that paedomorphosis has evolved multiple times in the complex (Shaffer 1984a; Shaffer & Voss 1996), and that the 15 contained species have evolved within the last 5 million years (Shaffer *et al*. 2004). Finally, it is clear from both laboratory (Voss & Shaffer 1996) and field (Riley *et al*. 2003; Fitzpatrick & Shaffer 2004) studies that even the most divergent members of the clade are capable of producing fertile hybrids.

Here, we focus on the molecular determination of species boundaries in *Ambystoma ordinarium*, a distinctive, facultative paedomorph found in high-elevation streams of the central Mexican Plateau (Fig. 1). Metamorphic and paedomorphic *A. ordinarium* have been found in some populations, indicating the possibility of terrestrial dispersal (Anderson & Worthington 1971), although our recent fieldwork indicates that paedomorphosis dominates extant populations. *A. ordinarium* was described as a species by

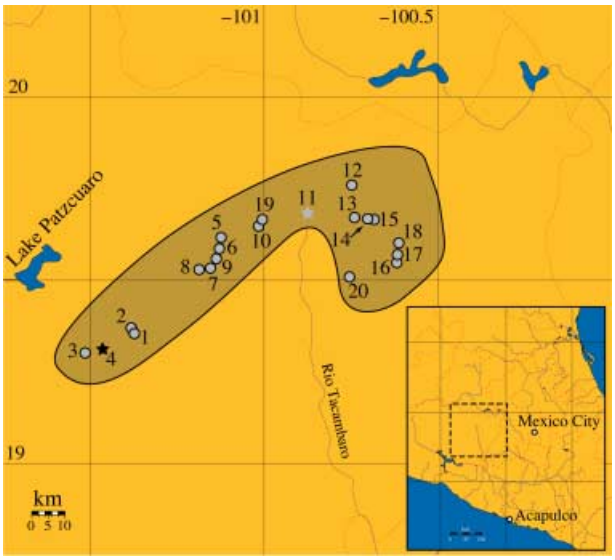

Fig. 1 Geographic distribution and collecting localities for *Ambystoma ordinarium*. The shaded area gives the approximate distribution of *A. ordinarium* based on Anderson & Worthington (1971), Anderson (1975), and the personal collection information of the authors. Locality numbers correspond to those listed in Table 1. Stars denote the localities used in the mtDNA study of Shaffer & McKnight (1996). The white star denotes the locality used in the allozyme study of Shaffer (1984a).

Taylor (1939) based on its distinctive morphology and a unique paedomorphic occupation of high-elevation stream habitats (most paedomorphic tiger salamanders are lacustrine). Subsequent analyses revealed a relatively specialized morphology (Shaffer 1984b) and prey capture biomechanics (Shaffer & Lauder 1985) that are indicative of a well-defined, ecologically and functionally differentiated species.

In contrast, molecular analyses have indicated a different evolutionary history. A mitochondrial DNA study of all members of the *A. tigrinum* species complex (Shaffer & McKnight 1996) revealed two well-differentiated *A. ordinarium* haplotypes (uncorrected 'p' distance = 3.61%) placed in divergent positions of the mtDNA gene tree (Fig. 2; Shaffer & McKnight 1996). The haplotype sampled from the western portion of *A. ordinarium*'s range was minimally diverged (uncorrected 'p' distance = 0.63%) from a haplotype sampled from the geographically proximal Lake Patzcuaro paedomorphic species, *Ambystoma dumerilii* (Fig. 1). These patterns have led to speculation that *A. ordinarium* is not a monophyletic taxon (Highton 2000) and the intriguing possibility that '*A. ordinarium*' is comprised of two (or more) lineages that have convergently evolved this unique ecology, life history and associated morphology.

In this study we used DNA sequence data from a greatly expanded mtDNA data set, in combination with multiple nuclear expressed sequence tag-based (EST) loci, to obtain a detailed genealogical perspective on the evolutionary

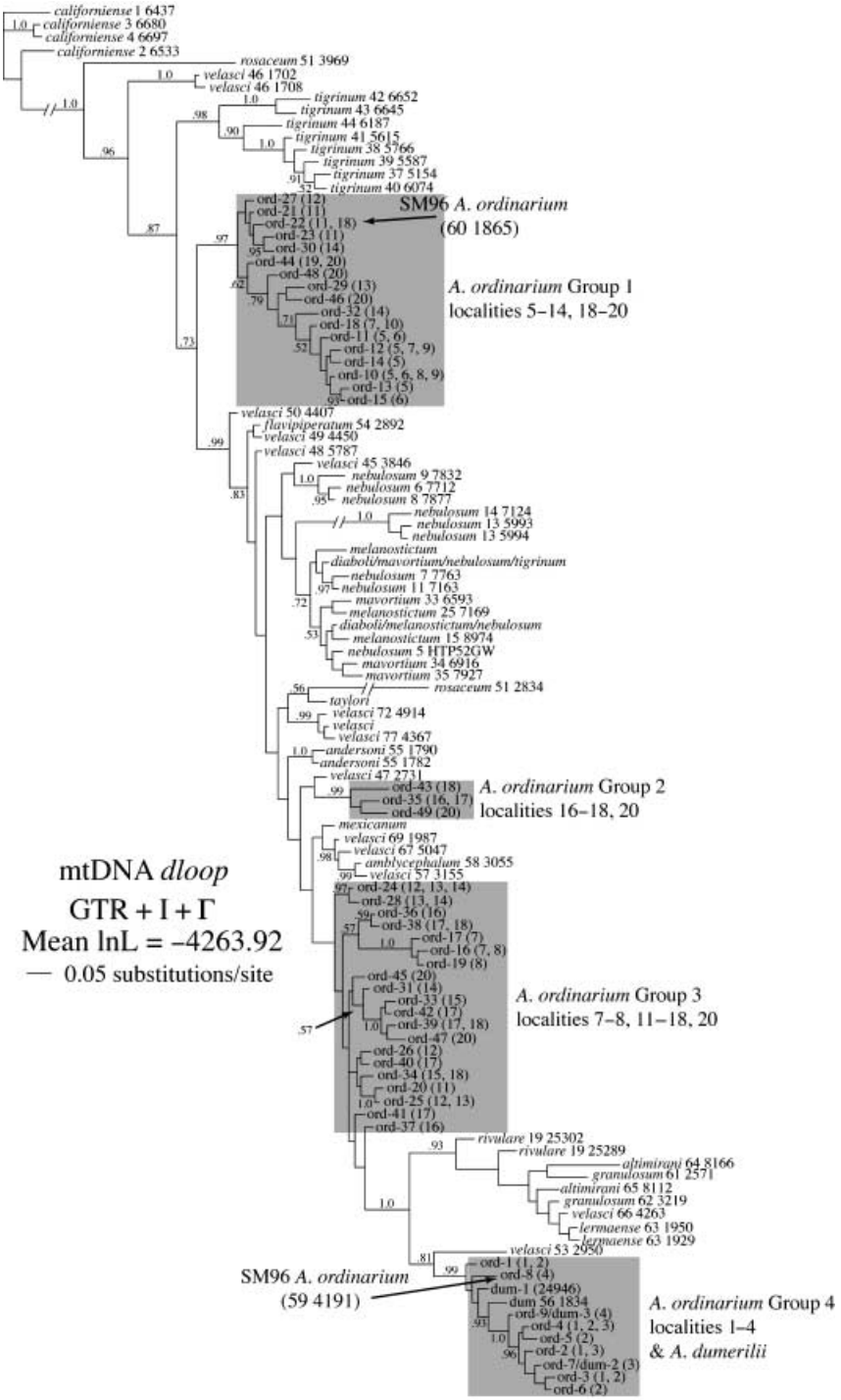

Fig. 2 Majority-rule consensus phylogram resulting from Bayesian analysis of the mtDNA *dloop* haplotypes with indels coded as binary characters. Posterior probabilities > 0.50 are given for branches. Tree tips corresponding to the haplotype sequences of Shaffer & McKnight (1996) are labelled according to the species, locality, and individual that they were sampled from. *Ambystoma ordinarium* haplotypes fell into four groups spread throughout the tree, which are encompassed by shaded boxes. The two *A. ordinarium* haplotypes identified in Shaffer & McKnight (1996) are indicated with a SM96 label. The localities where each *A. ordinarium* haplotype was found are given in parentheses.

history and species boundaries of *A. ordinarium*. We assessed, using genealogical exclusivity criteria, whether *A. ordinarium* represents a single phylogenetic species or is comprised of multiple independent and genealogically exclusive sets of populations as suggested by initial mtDNA inferences (Highton 2000). As has recently been emphasized by several authors, species delimitation studies require explicit criteria by which species are recognized (de Queiroz 1999; Sites & Marshall 2004). The criterion of genealogical exclusivity (Avise & Ball 1990; Baum & Shaw 1995) is particularly focused on genealogical patterns across unlinked loci, such that a set of populations that has undergone species-level divergence should yield concordant patterns of monophyly across the genome. Although we advocate this

concept for allopatrically distributed taxa like the Mexican ambystomatids, we also note that the original derivation of this criterion does not recognize the stochasticity of genetic drift, which results in variation in the time to monophyly across loci (Hudson & Coyne 2002), nor does it account for differential patterns of gene flow across markers (Funk & Omland 2003; Chan & Levin 2005). Consequently, not all loci are expected to exhibit exclusive monophyly, even after substantial time since divergence. A reasonable solution which we use in this study is to recognize phylogenetic species as lineages exhibiting monophyletic patterns in a majority of sampled loci, which are not contradicted by phylogenetic patterns at other loci (Dettman *et al*. 2003).

Materials and methods

Identification of polymorphic nuclear loci

The strategy used to identify intraspecific, polymorphic nuclear markers was described in Putta *et al*. (2004). Six polymorphic EST markers were used in conjunction with two additional nuclear loci (*col1a1* and *dlx3*) shown to be polymorphic across the *Ambystoma tigrinum* complex (Voss *et al*. 2001). Six of the nuclear markers correspond to protein-coding genes that have a putative function based on strong BLAST hits: (1) Collagen, type I, alpha-1 (*col1a1*), (2) Distal-less Homeobox 3 (*dlx3*), (3) *ctg1908* exhibits similarity to *Homo sapiens* DEAD box polypeptide 5, (4) *g1d6* exhibits similarity to *Homo sapiens* karyopherin alpha 6, (5) *g1c12* flanks an intron boundary in a gene exhibiting similarity to *Homo sapiens* myosin regulatory light chain MRCL2, and (6) *g3d7* flanks an intron boundary in a gene exhibiting similarity to *Homo sapiens* solute carrier family 25, member 5. Two additional markers (*ctg1506* and *g1f1*) do not have significant GenBank BLAST hits. These latter two markers probably correspond to 3′ untranslated regions of genes. The genomic positions of *ctg1506* (linkage group 11), *g1c12* (linkage group 2), *g1d6* (linkage group 6), *colla1* (linkage group 11), and *dlx3* (linkage group 11) have been determined (www.ambystoma.org); based on single PCR amplicons, the three unmapped ESTs also correspond to single loci.

Sampling

Ambystoma ordinarium has a relatively narrow distribution both geographically and ecologically in the high pine forest habitat of eastern Michoacan, Mexico (Anderson & Worthington 1971; Anderson 1975). The species is distinct morphologically, and the combination of larval body colour, gill structure, and head shape distinguish both larvae and metamorphs from other Mexican ambystomatid salamanders (Shaffer 1984b). *A. ordinarium* is also a habitat specialist restricted to streams, rather than the more common pond/lake breeding of most members of the tiger salamander complex. Over a 3-month period of intensive fieldwork, we identified 20 sampling localities that span the geographic range of the species, and sampled 217 paedomorphic or young larval individuals of *A. ordinarium* (Fig. 1; Table 1). Throughout this field effort no metamorphic individuals were observed. We used 41 samples from

Table 1 Information for sampled localities of *Ambystoma ordinarium*

Locality number*	*n*†	Latitude N, longitude W	Locality description
1	14	19°22′10″, 101°22′53″	Small stream in Cruz de Plato, ~11 km W (by road) Villa Madero, 0.3 km W of paved road
2	15	19°22′12″, 101°22′57″	Small stream in Cruz de Plato, ~11 km W (by road) Villa Madero, 0.3 km W of paved road
3	9	19°18′06″, 101°30′54″	Spring-fed stream, in town of El Pedregoso, ~3.5–4 km W of Patzcuaro-Taucomaro Hwy
4	11	19°18′28″, 101°28′04″	Large stream passing under paved road, 10.2 km (by road), E of San Gregorio
5	10	19°37′02″, 101°07′27″	10 km SSE (straight line) of Morelia, S of San Miguel del Monte, in N flowing creek
6	16	19°35′14″, 101°07′43″	12.5 km SSE (straight line) of Morelia, in SW flowing stream
7	7	19°31′58″, 101°09′06″	18.75 km S (straight line) of Morelia, in S flowing creek
8	9	19°31′47″, 101°11′12″	17.5 km S (straight line) of Morelia, in town of Las Palomas, in WNW flowing stream
9	9	19°33′35″, 101°08′14″	~16.75 km S (straight line) of Morelia, in E flowing stream
10	10	19°38′53″, 101°00′57″	Small stream, 12.6 km E (by road), then 2.4 km S (by road) of Morelia at town of Pino Real
11	10	19°40′58″, 100°52′22″	SW flowing stream, S of Hwy. 15, 0.4 km W of San Jose Lagunillas between Morelia and Ciudad Hidalgo
12	10	19°45′38″, 100°44′52″	Small S flowing stream, 12 km S of Hwy. 126 and 51 intersection, then 2.6 km E of Hwy 51 on dirt road
13	10	19°40′19″, 100°44′24″	Small stream ~0.2 km N of intersection of Hwy. 51 and 15, where it crosses under Hwy. 51
14	10	19°40′04″, 100°42′08″	Small E flowing stream, 4.7 km E of intersection of Hwy 15 and 51
15	10	19°39′60″, 100°40′60″	Small S flowing stream, 10.7 km E of intersection of Hwy 51 and 15
16	10	19°32′54″, 100°37′06″	14.7 km S (by road) of Ciudad Hidalgo in small stream just E of road
17	15	19°34′12″, 100°36′58″	12.6 km S (by road) Ciudad Hidalgo, in E flowing stream
18	12	Not available	7.8 km S (by road) Ciudad Hidalgo, in main N flowing stream
19	10	19°39′57″, 101°00′19″	Small N flowing stream 23 km E (by road) of Morelia (Jose Maria Morelos Parque Nacional)
20	10	19°30′34″, 100°45′16″	Small SW flowing stream 1.7 km S (by road) of San Antonio Villalongin

*The locality number corresponds to the numbered localities in Fig. 1.
†The number of individuals sampled per locality.

12 additional Mexican and US species or subspecies of the *A. tigrinum* complex to place *A. ordinarium* within the phylogenetic context of the larger species radiation. This sampling included *Ambystoma californiense* (HBS8849), *Ambystoma amblycephalum* (HBS3055), *Ambystoma andersoni* (HBS1782 and 1790), *Ambystoma dumerilii* (HBS24946 and 24950-24951), *Ambystoma flavipiperatum* (HBS2892), *Ambystoma granulosum* (HBS2571), *Ambystoma lermaense* (HBS1950), *Ambystoma tigrinum melanostictum* (HBS6792), *Ambystoma rivulare* (HBS25288–25305), *Ambystoma rosaceum* (HBS2834 and 3969), *Ambystoma taylori* (HBS4892), and *Ambystoma velasci* (HBS1702, 1987, 2950, 3248, 3258, 4263, 4367, 4828, 4880, 4892, 4901, 4903, 4914, 4967, and 5787). The majority of these samples were used in Shaffer & McKnight (1996) and represent major clades identified in that study. Phylogenetic trees were rooted with *A. californiense* based on previous phylogenetic studies (Shaffer & McKnight 1996; Samuels *et al*. 2005).

Mitochondrial DNA sequence data were collected from newly sampled individuals of *A. ordinarium* and combined with previous mtDNA sequence data from other species of the *A. tigrinum* complex (Shaffer & McKnight 1996). New mtDNA sequence data were also collected from the above-listed samples of *A. dumerilii* and *A. rivulare*. Our new mtDNA sequence is approximately 1100 bp in length and covers the entire length of the noncoding insert region between $tRNA^{Thr}$ and $tRNA^{Pro}$, $tRNA^{Pro}$, and the *dloop* gene. We refer to this gene region throughout the paper as *dloop*. The newly sequenced mtDNA region contains approximately 240 additional base pairs from the 3′ end of *dloop* than was used by Shaffer & McKnight (1996); the missing 240 bp were coded as 'missing' in the new data matrix. Nuclear sequence data for all eight nuclear markers were collected from the newly sampled *A. ordinarium* individuals and the above-listed representatives of the *A. tigrinum* complex. Within *A. ordinarium*, we collected sequence data from all genomic markers for the majority of sampled individuals. However, polymerase chain reaction (PCR) in some individuals was difficult for some loci; consequently, data were not available for some *A. ordinarium* individuals. We collected sequence data from most of the available representatives of the *A. tigrinum* complex. We were unable to PCR amplify an orthologous fragment for *g1c12* in *A. californiense*. We were also unable to amplify an orthologous fragment for *g3d7* in *A. dumerilii*. In all we sequenced approximately 246 individual salamanders for eight nuclear and one mitochondrial genes, yielding a total of approximately 5200 bp of sequence data per individual. All sequence data have been deposited in GenBank. mtDNA *dloop* sequences [including the mtDNA sequence data of Shaffer & McKnight (1996)] are archived under accession numbers DQ240923–DQ241217. Nuclear sequences are archived under accession numbers: *col1a1* DQ252522–DQ252937; *dlx3* DQ248350–DQ248859; *ctg1505* DQ252938–DQ253449; *ctg1908* DQ254302–DQ254797; *g1c12* DQ254798–DQ255269; *g1d6* DQ255270–DQ255783; *g1f1* DQ253450–DQ253923; and *g3d7* DQ253924–DQ254301.

Table 2 PCR primer sequences for all loci used in this study

Locus	Primer name	Nucleotide sequence 5′→3′
dloop	THR	AAACATCGATCTTGTAAGTC
	DL1	AATATTGATAATTCAAGCTCCG
	DL4	GCCACTGGTTAAAATCTATG
col1a1	COL1A1.5.1	CACCGAAGCCTCCCAAAACATCAC
	COL1A1.3.1	GAGCCCTTCCATCTTAGTCGT
dlx3	DLX3f	GGCGAGGCGCACCTCTCCAACTGGTGA
	DLX3r	AGGCTCCCACCTTCTGAGTTGGGAAGG
ctg1506	EFC1506.5.1	AGGATATCCGCTCAGAAATATGAAG
	EFC1506.3.1	CTGACCACTTGCAAAACTTACTACCT
ctg1908	EFC1908.5.1	CTCATGACTTAATTGCTGTTCTTCG
	EFC1908.3.1	ATAACCATTCTGAGGTTTTGAGTTG
g1d6	G1-D6.5.1	CAGCGTGCCCACCCGATAGAA
	G1-D6.3.1	TCCCAAAAAGTAAAATGTGCAAAGAAAA
g1f1	G1-F1.5.1	TTAGTTTGGGTGCAGACAGGA
	G1-F1.3.1	GGTGCTCAACAACAAATCAACT
g1c12	G1-C12.5.1	CCCAAATCCAGGAGTTCAAA
	G1-C12.3.2	CAAGGCAGCCAAATTATCGT
g3d7	G3-D7.5.2	TCCTTTTCCCCAGTTTGTTG
	G3-D7.3.2	TATGAAACCCTGCTCCTTGG

DNA sequencing, haplotype phasing, and alignment

DNA isolation, PCR amplification, and sequencing were described previously (Samuels *et al*. 2005). Primer sequences are listed in Table 2. Alignment of DNA sequences was straightforward due to low divergences. Insertion-deletion (indel) events greater than 1 bp were scored as a single mutation. Heterozygous nucleotide positions were identified through dual peaks present in electropherograms (Brumfield *et al*. 2003). Heterozygous indel positions were identified by a sharp transition in the electropherogram from clean to garbled sequence, where the transition corresponded to the same position of a homozygous indel in other individuals (Bhangale *et al*. 2005). To identify haplotypes from heterozygotes we used a Bayesian approach implemented in PHASE version 2.1 (Stephens *et al*. 2001). To reduce the error of incorrect haplotypic phasing, phased positions not receiving a posterior probability (PP) of ≥ 0.90 in an individual was scored as missing data. In total, the proportion of nucleotide characters that could not be unambiguously phased was less than 5%.

Genetic diversity

Within *A. ordinarium*, the number of segregating sites, the average number of nucleotide differences per site between

two sequences (π), the proportion of segregating polymorphic sites ($\theta = 4N\mu$), and the minimum number of recombination events (Hudson & Kaplan 1985) within a locus were calculated using DNASP version 4.0 (Rozas *et al.* 2003). These calculations did not include indel events. We assessed departure from neutral evolution for individual loci using Tajima's *D* statistic (Tajima 1989) and Fu and Li's *D** and *F** statistics (Fu & Li 1993). Genotypic disequilibrium among loci was tested with GENEPOP version 3.4 (Raymond & Rousset 1995).

Gene tree reconstruction

We constructed gene trees using appropriate methodologies to resolve both hierarchical relationships among haplotypes and population-level haplotype networks (Morando *et al.* 2003). Haplotype networks were generated using statistical parsimony analysis performed in TCS version 1.21 (Clement *et al.* 2000) using gaps as a fifth character state. Hierarchical bifurcating gene trees were reconstructed using Bayesian analysis implemented in MRBAYES version 3.04 (Huelsenbeck & Ronquist 2001) with separate partitions for nucleotide polymorphism data and indels. Indels were scored as a presence/absence binary character and analysed using the Lewis Mk model (Lewis 2001). Evolutionary models for substitution data were chosen using the Akaike information criterion in MODELTEST version 3.6 (Posada & Crandall 1998). Five Markov chains were used with a temperature profile of 0.2, default substitution priors were used in all analyses and random trees were used to start each Markov chain. Chains were run for 10 million generations. Although all analyses reached stationarity (based on visual inspection) within the first 1 million generations, we conservatively discarded the first 5 million generations. Three replicate analyses were performed for each data set to insure that a stable posterior distribution was reached. For the Bayesian gene trees in which *A. ordinarium* haplotypes were not resolved as monophyletic we used the posterior distribution of trees to statistically test *A. ordinarium* monophyly. The program PAUP* version 4.0 (Swofford 2002) was used to filter the posterior distributions of trees to keep only those that resolved *A. ordinarium* as monophyletic. If less than 5% of the trees from the posterior distribution contained a monophyletic *A. ordinarium*, the data were considered to statistically reject *A. ordinarium* monophyly under a Bayesian criterion.

Results

Genetic diversity patterns

Haplotype and nucleotide diversity across *Ambystoma ordinarium* was highest in the mtDNA *dloop* sequence, with nucleotide diversity 2–21 times greater than that found in any nuclear locus (Table 3). Even so, all nuclear loci contained multiple single nucleotide polymorphisms (SNPs), and most contained segregating indels, yielding considerable levels of nucleotide diversity (Table 3). Most indels were 1 or 2 bp in length; however, the *g3d7* locus contained a 10-bp indel. Most loci exhibited patterns consistent with neutrality using all statistical tests. The *g1f1* locus significantly deviated from neutrality using Tajima's *D* statistic ($D = 2.4339$, $P < 0.05$) and Fu and Li's *F** statistic ($F^* = 2.0075$, $P < 0.05$), although these were not significant after a Bonferroni adjustment of the alpha level to 0.006.

We did not detect significant patterns of linkage disequilibrium among loci. While many of the nuclear loci are known to reside on different linkage groups, the loci *col1a1* and *dlx3* are separated by 3.9 centimorgans; these genes are separated from *ctg1506* by 22.5–26.4 centimorgans (Smith *et al.* 2005). However, our linkage disequilibrium tests indicated that *col1a1* and *dlx3* are in linkage equilibrium; we therefore treated each as providing independent estimates of phylogeny. Two nuclear loci (*col1a1* and *g1f1*) were each found to contain at least one intragenic recombination event. However, we analysed them each as single genes and found that the monophyletic patterns resolved at these loci are likely not influenced by recombination (see below).

Table 3 Nucleotide variation measures within *Ambystoma ordinarium* for all loci used in this study

Locus	Individuals sampled	Alignment length	Number of haplotypes	Segregating SNPs	Segregating indels	π	θ
mtDNA *dloop*	204	1059	49	63	3	0.0152	0.0101
col1a1	179	703	7	5	0	0.0013	0.0011
dlx3	212	150	4	3	0	0.0035	0.003
ctg1506	215	277	4	4	0	0.0036	0.0022
ctg1908	205	500	7	2	4	0.0007	0.0006
g1c12	200	981	9	11	3	0.0013	0.0017
g1d6	215	309	4	4	2	0.002	0.0019
g1f1	215	403	6	9	1	0.0074	0.0034
g3d7	189	749	14	13	3	0.0032	0.0027

Mitochondrial DNA genealogy

The data set of all new and previously published (Shaffer & McKnight 1996) *dloop* sequence contained 111 haplotypes (49 restricted to *A. ordinarium*) and 1074 aligned positions (Tables 3 and 4). Statistical parsimony reconstructed a number of network subsets that cannot be linked under the 95% criterion (results not shown). Bayesian analysis revealed an extensive nonmonophyletic pattern for *A. ordinarium* haplotypes (Fig. 2). A statistical test of *A. ordinarium* monophyly using Bayesian criteria strongly reject this hypothesis (P = 0.0). Two groups of *A. ordinarium* haplotypes (groups 1 and 4, Fig. 2) correspond to the divergent haplotypes recovered in Shaffer & McKnight (1996). Group 4 contains all haplotypes from localities 1–4, including the haplotype sampled from locality 59 of Shaffer & McKnight (1996). All haplotypes sampled from the Lake Patzcuaro paedomorphic species, *Ambystoma dumerilii*, are also found in this clade. Two haplotypes were found in both *A. dumerilii* and *A. ordinarium* (*ord-7/dum-2* and *ord-9/dum-3*). Two additional *A. dumerilii* haplotypes [*dum-1* and the previously reported *A. dumerilii* locality 56 of Shaffer & McKnight (1996)] are minimally diverged from *A. ordinarium* group 4 haplotypes.

Group 1 contains *A. ordinarium* haplotypes sampled from localities 5–14 and 18–20, including the haplotype from locality 60 of Shaffer & McKnight (1996). Two other regions of the *dloop* tree contain *A. ordinarium* haplotypes (Fig. 2): group 2 contains three haplotypes from localities 16–18 and locality 20, and group 3 comprises 20 haplotypes arranged along the backbone of the tree. These haplotypes were sampled from localities 7–8, 11–18, and 20 and form a poorly supported, paraphyletic group with respect to a clade containing haplotypes sampled from a number of *A. tigrinum* complex species and the *A. ordinarium/A. dumerilii* haplotype clade from localities 1–4.

Nuclear genealogical patterns within A. ordinarium

Nuclear genealogical relationships among *A. ordinarium* haplotypes were examined within the context of the tiger salamander complex. Alignment lengths and descriptions of polymorphism across the *Ambystoma tigrinum* complex at each locus can be found in Table 4; additional locality and specimen information for each gene can be found in Tables S1–S8 (Supplementary material). For all nuclear loci most haplotypes across the *A. tigrinum* complex could be linked into a single haplotype network under the 95% statistical parsimony criterion, reflecting the low divergences and relatively complete haplotype coverage of our samples. Statistical parsimony and Bayesian analysis yielded very similar gene trees, and we summarize the nuclear loci exhibiting *A. ordinarium* monophyly in the form of a generalized tree (Fig. 3). Nuclear loci exhibiting nonmonophyletic Bayesian phylogenetic trees are presented individually in Figs 4 and 5 with their corresponding posterior probabilities. Haplotype networks for the individual nuclear loci are presented in Figs S1–S4 (Supplementary material). The individual Bayesian nuclear gene trees resulting in *A. ordinarium* monophyly are presented in Figs S5–S6 (Supplementary material).

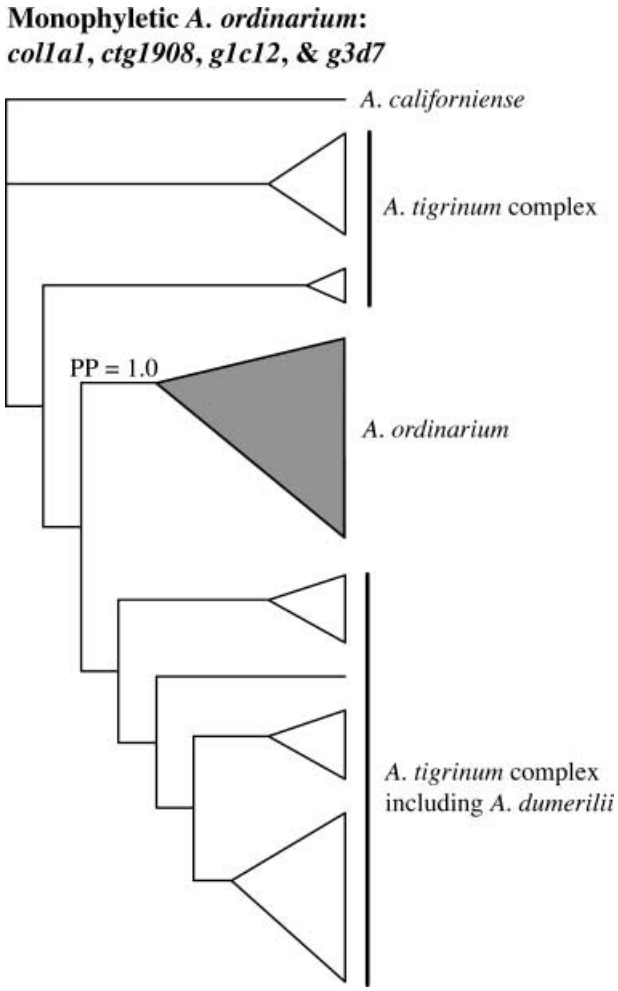

Fig. 3 Schematic phylogeny depicting the genealogical pattern of *Ambystoma ordinarium* monophyly recovered in Bayesian analyses of four nuclear loci (*col1a1*, *ctg1908*, *g1c12*, and *g3d7*). Phylogenetic structure is meant to be generalized; substantial discordance in species-level relationships exists across loci and no single tree can represent the branching structure for all loci. Details of phylogenetic/haplotype network structure can be found in supplementary Figs S1 to S6.

Four nuclear loci exhibit genealogical exclusivity for *A. ordinarium* haplotypes according to Bayesian criteria (Fig. 3; *col1a1*, *ctg1908*, *g1c12*, and *g3d7*), all with strong measures of branch support (PP = 1.0). One of these loci (*g3d7*) resolves *A. ordinarium* to be paraphyletic under statistical parsimony criteria (Fig. S2). Two additional nuclear loci (Fig. 4; *ctg1506* and *g1d6*) resolve paraphyletic groups of *A. ordinarium* haplotypes that are nearly exclusive. For these loci, the majority of *A. ordinarium* haplotypes were

Table 4 Nucleotide variation measures across the *Ambystoma tigrinum* complex (including *Ambystoma ordinarium*) for all loci used in this study

Locus	Alignment length	Number of haplotypes	Polymorphic sites*	Indel events	Average distance† (variance)	Average distance‡ (variance)
mtDNA *dloop*	1074	111	167	3	0.0583 (0.00001)	0.0224 (0.0001)
col1a1	703	50	60	2	0.0285 (0.00001)	0.0098 (0.00004)
dlx3	150	11	11	0	0.03 (0.00005)	0.0151 (0.00005)
ctg1506	277	20	27	0	0.0305 (0.00005)	0.0179 (0.00008)
ctg1908	508	27	60	12	0.0238 (0.00001)	0.013 (0.00004)
g1c12	1056	34	88	18	—	0.0127 (0.00003)
g1d6	315	22	32	4	0.0289 (0.00002)	0.0199 (0.00009)
g1f1	404	26	42	5	0.033 (0.00004)	0.0133 (0.00005)
g3d7	777	34	70	18	0.027 (0.00001)	0.011 (0.00003)

*Not including indel events.
†Average uncorrected pairwise distance between *A. californiense* haplotype and all remaining *A. tigrinum* complex haplotypes calculated using nucleotide polymorphism data (excluding indels).
‡Average uncorrected pairwise distance among all ingroup haplotypes using nucleotide polymorphism data (excluding indels).

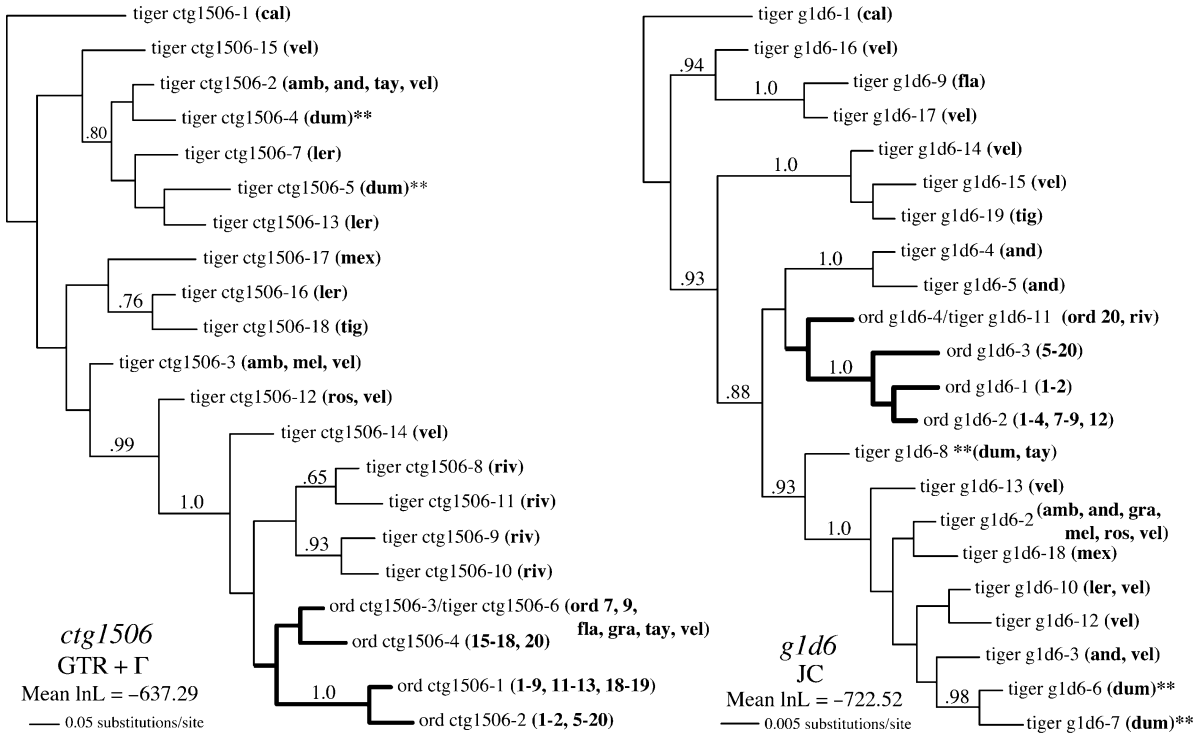

Fig. 4 Majority-rule consensus phylograms resulting from Bayesian analysis of the nuclear loci *ctg1506* and *g1d6*. The locus name along with the evolutionary model used, and the mean –ln L from the posterior distribution are presented below each tree. Thick branches lead to alleles sampled from *Ambystoma ordinarium*. Following each allele (bold) is a list of the *Ambystoma tigrinum* complex taxa where that allele was found (amb, *Ambystoma amblycephalum*; and, *Ambystoma andersoni*; cal, *Ambystoma californiense*; dum, *Ambystoma dumerilii*; fla, *Ambystoma flavipiperatum*; gra, *Ambystoma granulosum*; ler, *Ambystoma lermaense*; mav, *Ambystoma tigrinum mavortium*; mel, *Ambystoma tigrinum melanostictum*; mex, *Ambystoma mexicanum*; riv, *Ambystoma rivulare*; ros, *Ambystoma rosaceum*; tay, *Ambystoma taylori*; tig, *Ambystoma tigrinum tigrinum*; vel, *Ambystoma velasci*). The sampling localities where each *A. ordinarium* allele was found are also given in parentheses. *A. dumerilii* haplotypes are marked with double asterisks.

restricted to a strongly supported clade (PP = 1.0), while a smaller number of *A. ordinarium* haplotypes that were also present in other *A. tigrinum* complex species formed a sister group to this exclusive *A. ordinarium* clade. For *ctg1506*, three copies of the *ordctg1506-3* haplotype were found in two localities, and this haplotype was shared with four other species from across the Mexican plateau. In addition, 22 copies of *ordctg1506-4* were found across five localities;

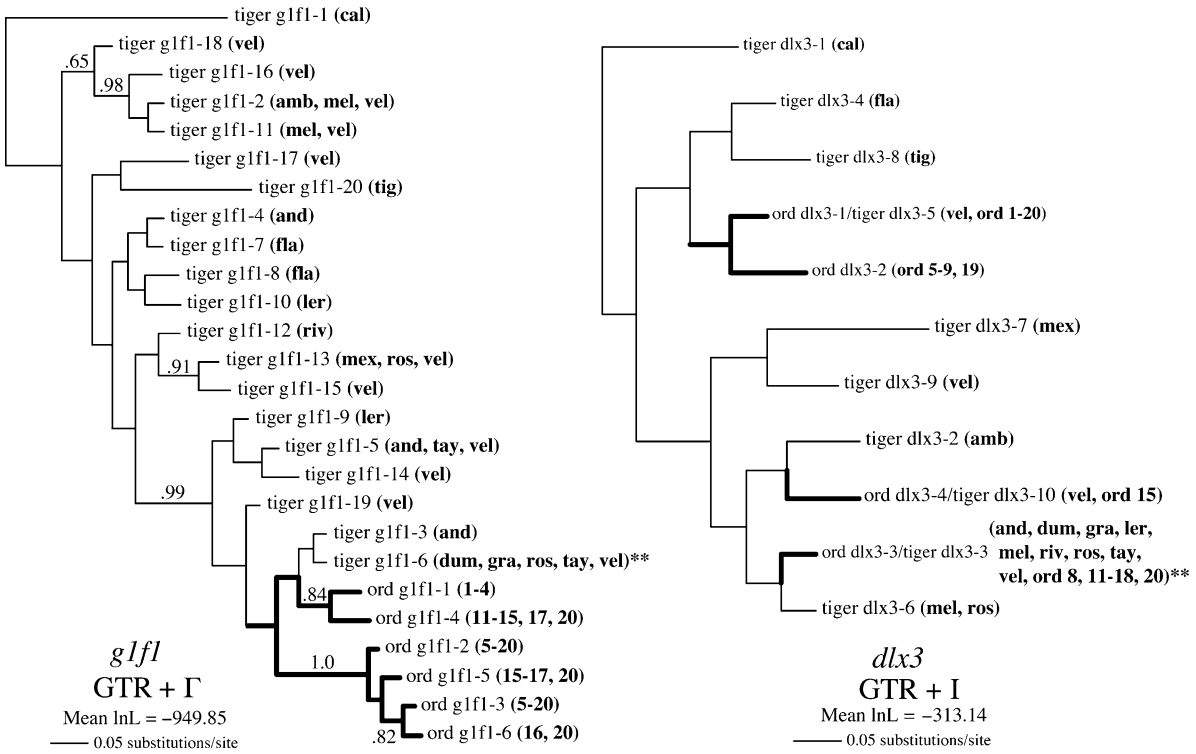

Fig. 5 Majority-rule consensus phylograms resulting from Bayesian analysis of the nuclear loci *g1f1* and *dlx3*. Tree descriptions as in Fig. 4.

together, these two haplotypes form a poorly supported sister group to the primary *A. ordinarium* clade. For *g1d6*, two heterozygotes from locality 20 each contain a copy of *ordg1d6-4*, a haplotype that is also found in all 18 individuals of *A. rivulare*. The *g1f1* locus also contains paraphyletic assemblages of *A. ordinarium* haplotypes (Fig. 5), but the distribution of haplotypes across taxa is quite different. No haplotypes are shared between *A. ordinarium* and other *A. tigrinum* complex species at this locus, but *A. ordinarium* is reconstructed as paraphyletic (with weak statistical support) with respect to a clade of haplotypes found in a number of other species. However, monophyly of *A. ordinarium g1f1* haplotypes cannot be rejected using Bayesian criteria (P = 0.164). Only *dlx3* exhibits extensive nonmonophyly for *A. ordinarium* haplotypes (Fig. 5). There is a high degree of shared haplotypes with other species at *dlx3*; only one of the four *A. ordinarium* haplotypes is not shared with other species of the *A. tigrinum* complex, and all relationships for this locus have weak statistical support. The *dlx3* Bayesian posterior distribution is the only nuclear gene that statistically rejects monophyly of *A. ordinarium* haplotypes (P = 0.0004).

There are no concordantly resolved genealogically exclusive geographic groups within *A. ordinarium*. The majority of nuclear loci contain haplotypes that are found in all or most of the *A. ordinarium* localities. The one exception to this is *g1f1*, which contains one haplotype found exclusively in all individuals from localities 1–4. However, this haplotype is closely related to a haplotype found in a number of additional localities.

Nuclear genealogical patterns across the A. tigrinum *complex*

Two patterns stand out from the gene trees reconstructed from different loci. First, there is inconsistency across loci in phylogenetic relationships among different populations and species of the complex (Figs 4 and 5; Figs S1–S6). For example, *Ambystoma amblycephalum* and *Ambystoma granulosum* contain identical haplotypes at two loci (*ctg1908* and *g1d6*), yet their haplotypes are placed in distant positions in the gene trees at four other loci (*dloop*, *ctg1506*, *g1f1*, and *g3d7*). Similarly, haplotypes from *A. granulosum* and *A. lermaense* are shared or phylogenetically similar at most loci except for two (*ctg1506* and *g1f1*), where they are phylogenetically distant. Second, there are frequent patterns of species and heterozygous individuals containing phylogenetically distant haplotypes. For example, at the *g1c12* locus the two haplotypes sampled from the single individual of *A. amblycephalum* are placed in divergent, well-supported clades. Finally, in contrast to the mtDNA gene tree, *A. dumerilii* haplotypes at five nuclear loci are placed in phylogenetically distant positions from all *A. ordinarium* haplotypes. A less extreme pattern of divergence between these species is found in the *g1f1* haplotype network where

the *A. dumerilii* haplotype is 2–5 mutational steps from *A. ordinarium* haplotypes. However, there is no sharing of haplotypes between these species at this locus. The exception to this pattern is the *dlx3* locus, where all sampled *A. dumerilii* contain a haplotype shared with nine other species, including *A. ordinarium*. However, *dlx3* is also the shortest sequence (150 bp), and it had essentially no statistical support (PP < 0.50) for all gene tree nodes.

Discussion

Our multilocus examination of population-level DNA sequence variation within *Ambystoma ordinarium* reveals the complexity of genealogical patterns that can exist across loci within recently evolved species. Based on comprehensive mtDNA sequence analysis, *A. ordinarium* is reconstructed as polyphyletic with respect to most of the Mexican ambystomatid species (Fig. 2), suggesting either an extreme mismatch between gene trees and species trees, or that *A. ordinarium* properly represents several undescribed, and relatively unrelated species of stream-dwelling paedomorphic salamanders (Highton 2000). Nuclear loci tell a very different story: four loci exhibit exclusive monophyly for *A. ordinarium* haplotypes, three loci exhibit *A. ordinarium* paraphyly with two of these loci on the brink of exclusive monophyly, and a single, short, and largely uninformative nuclear locus exhibits *A. ordinarium* polyphyly (Figs 3–5). Using a relatively stringent genealogical species criterion (Avise & Ball 1990; Baum & Shaw 1995; Sites & Marshall 2004), which delimits species as exclusive groups of individuals exhibiting concordant patterns of monophyly across unlinked genes, *A. ordinarium* can be diagnosed as a genealogical species based on its nuclear genealogical patterns. Unlike the mtDNA gene tree, the alternative patterns to exclusive *A. ordinarium* monophyly seen in the nuclear genes are neither concordant across genes nor well-supported within gene trees, suggesting that these genes either have not sorted completely or do not contain adequate signal to be phylogenetically informative.

A major challenge in interpreting intraspecific gene genealogies in species radiations is to understand the causes of nonmonophyly (e.g. Patton & Smith 1994; Nagl *et al*. 1998; Sato *et al*. 1999; Shaw 2002). It is often assumed, based on the neutral coalescent, that the effects of lineage sorting and introgression in an mtDNA genealogy will also be reflected in nuclear genealogies. Our results provide an important contrasting case in which nuclear gene trees resolve *A. ordinarium* as a genealogical species, while the mtDNA does not. The trend towards monophyly across nuclear loci indicates that variation in nuclear genealogies is attributable to a stochastic lineage sorting process that has gone to near completion. The contrasting lack of *A. ordinarium* monophyly in the mtDNA tree is surprising, and warrants further discussion. Although much of the mtDNA tree is incompletely resolved with low statistical support, two specific regions of the tree are well supported and at odds with the overall pattern in the nuclear data. First, three *A. ordinarium* haplotype groups, labelled 1, 2, and 4 on Fig. 2 are well-supported (PPs of 0.97–0.99) and not each other's closest relatives, with group 1 separated from most other members of the tiger salamander complex by a very well-supported node. Second, group 4 is itself not exclusive to *A. ordinarium*, but also contains all copies of the Lake Pazcuaro endemic paedomorphic species *Ambystoma dumerilii*. Although incomplete lineage sorting may account for a portion of *A. ordinarium*'s nonmonophyly in the mtDNA tree, our relatively complete examination of mitochondrial and nuclear genes indicates that hybridization is a likely source of discrepancy in group 4.

Nonmonophyly via incomplete lineage sorting

Nonmonophyly of mtDNA haplotypes sampled from *A. ordinarium* could result in part from the retention of ancestral allelic lineages whose coalescence pre-dates speciation. Retention of ancestral polymorphism in extant populations are especially prevalent in species involved in rapidly diversifying radiations, where incomplete lineage sorting preceding speciation is more likely to occur (Pamilo & Nei 1988; Maddison 1997; Albertson *et al*. 1999; Takahashi *et al*. 2001). Our results support the conclusions of Shaffer & McKnight (1996) that the *Ambystoma tigrinum* complex went through a period of rapid lineage diversification. Evidence for this comes from the short internal branch lengths across the mtDNA tree, and incongruence of interspecific relationships across loci. As the durations of branches separating speciation events approach a polytomy, the probability of lineage coalescence during these intervals becomes very low and the probability of gene tree concordance approaches zero (Poe & Chubb 2004).

The greatest challenge to invoking incomplete lineage sorting as a factor in the nonmonophyly of *A. ordinarium* mtDNA haplotypes is the monophyletic reconstruction of *A. ordinarium* haplotypes for many nuclear loci. The stochasticity of the lineage sorting process allows for the possibility that some nuclear loci will achieve monophyly before mtDNA monophyly (Hudson & Turelli 2003), although the average fourfold faster expected coalescence time for mtDNA (Moore 1995; Palumbi *et al*. 2001) makes this an unlikely outcome (Hudson & Coyne 2002). The resolution of monophyly or near-monophyletic paraphyly at seven of eight nuclear loci indicates that *A. ordinarium* has been a divergent evolutionary lineage for a substantial period of time (Tajima 1983; Neigel & Avise 1986; Rosenberg 2003). Under idealized conditions with purely neutral evolution and constant population size, it takes from about 4–9 N_e generations for 50% and 95% (respectively) of the nuclear genes in a sample to reach monophyly (Hudson & Coyne

2002), although geographic subdivision within species will increase these times (Wakeley 2000). The nuclear genealogical patterns of monophyly and paraphyly suggest that *A. ordinarium* divergence approaches the latter half of this temporal scale [the maintenance of *dlx3* polyphyly in *A. ordinarium* may be partially attributed to natural selection, which has been documented in field studies of this gene in other tiger salamander species (Riley *et al*. 2003; Fitzpatrick & Shaffer 2004)]. While a highly biased female sex ratio or reduced female dispersal could yield retained ancestral polymorphism in mtDNA but not nDNA (Hoelzer 1997; Navajas & Boursot 2003), we have no field evidence that such mechanisms are plausible for the *A. tigrinum* complex (Trenham *et al*. 2001). We suggest that incomplete mtDNA lineage sorting in the face of rapid diversification, in combination with uncertainty in tree reconstruction among clades, may be responsible for the apparent nonmonophyly of *A. ordinarium* groups 1, 2 and 3 (Fig. 2). However, this does not explain the sequence identity between *A. ordinarium* and *A. dumerilii* in group 4, and the fixation of *A. dumerilii*-like haplotypes in populations 1–4 of *A. ordinarium*.

Nonmonophyly via introgressive hybridization

Mitochondrial introgression is emerging as a frequent pattern in population-level studies (Funk & Omland 2003; Chan & Levin 2005). Our data provide strong evidence that the nonmonophyletic patterns in *A. ordinarium* group 4 (Fig. 2) can be attributed to mtDNA introgression of the Lake Patzcuaro paedomorphic species, *A. dumerilii*, into western populations (localities 1–4) of *A. ordinarium* (Fig. 1). This is indicated by the recovery of identical *dloop* haplotypes shared between these two species and the recovery of other shared *dloop* sequences that are minimally divergent across these two species. In contrast, nuclear haplotypes from *A. ordinarium* and *A. dumerilii* are frequently placed in distant phylogenetic positions, a pattern concordant with a previous allozyme study (Shaffer 1984a), indicating that these are not phylogenetically closely related taxa. These results indicate recent hybridization between these two species and the differential introgression of mtDNA alleles without substantial introgression of nuclear alleles.

Reproductive isolation in rapid species radiations may evolve either as a by-product of ecological differentiation in allopatrically distributed species or through direct selection for premating isolation in sympatrically distributed species (Schluter 2000, 2001). However, complete reproductive isolation likely requires considerable time to evolve as evidenced by hybridization among ecologically differentiated species in a variety of adaptive radiations (e.g. Ruber *et al*. 2001; Grant *et al*. 2003). Although hybridization is known to be possible among many members of the *A. tigrinum* complex in the laboratory (Voss & Shaffer 1996) and field (Riley *et al*. 2003; Fitzpatrick & Shaffer 2004), it is surprising to detect past hybridization patterns between *A. ordinarium* and *A. dumerilii* given their allopatric distributions, substantially different functional morphologies (Shaffer & Lauder 1985), and radically different habitat requirements (*A. dumerilii* is an obligate lacustrine paedomorph incapable of metamorphosis, whereas *A. ordinarium* is a specialist in flowing stream environments that is polymorphic for paedomorphosis and metamorphosis; Shaffer 1984b). A priori, one might assume that obligate paedomorphic species restricted to distinct hydrobasins would be ecologically incapable of contact and hybridization with other taxa. However, a small number of *A. ordinarium* metamorphs have been found in nature (Anderson & Worthington 1971) suggesting that limited reproductive contact and mtDNA introgression between these species may be facilitated through occasional metamorphosis and long-distance dispersal events. Although such contact is currently unlikely (Lake Patzcuaro is separated from the high-elevation stream habitat of *A. ordinarium* by approximately 30 km of flat, deforested habitat inappropriate for either species), during cooler glacial periods streams flowing from the western flanks of *A. ordinarium* habitat (that is, localities 1–4, Fig. 1) may have formed an ecological connection with Lake Patzcuaro (Watts & Bradbury 1982). mtDNA introgression through transient contact has been suggested in other systems (Glor *et al*. 2004; Weisrock *et al*. 2005), and we suggest that introgression between *A. ordinarium* and *A. dumerilii* offers an extreme example of how gene flow can be maintained between ecologically differentiated and allopatrically distributed species.

Application of EST-based nuclear markers to problems at the 'species boundary'

This study adds to a limited body of work demonstrating the importance of utilizing a multilocus genealogical approach in both phylogeographic and speciation research (e.g. Hare & Avise 1998; Machado & Hey 2002; Shaw 2002; Broughton & Harrison 2003; Dettman *et al*. 2003). A major limitation to collecting nuclear genealogical information relevant to the intra–interspecific boundary has been a lack of knowledge about the genomes of most non-model organisms. ESTs provide a tractable source of phylogenetic information (Theodorides *et al*. 2002); our study highlights a number of aspects that make them attractive as a source of nuclear variation for use in studies at the population–species boundary (Hare 2001; Zhang & Hewitt 2003). Our data demonstrate that ESTs can provide substantial intraspecific variation from numerous unlinked portions of the genome, particularly in the form of SNPs, which may be the marker of choice for future population-genetic studies (Brumfield *et al*. 2003; Morin *et al*. 2004). In addition, EST primers developed from one species can be applied to related, but relatively divergent taxa. Primers used in this

study were developed from *A. mexicanum* and *A. tigrinum*, but we successfully amplified loci from the overwhelming majority of sampled species from the *A. tigrinum* complex, including the most divergent member of the clade, *A. californiense*. Finally, large amounts of prior DNA sequence information are not necessary in an EST-based approach. From a set of just 123 primer combinations designed from orthologous ESTs in *A. mexicanum* and *A. tigrinum* we identified 20 polymorphic loci, suggesting that a moderate development effort can achieve significant empirical results. Small-scale EST projects targeting similar tissues at similar ontogenetic stages in one or a few relatively divergent species should be sufficient to identify orthologues for use in conserved primer development, and the subsequent identification of informative SNPs. Obviously, the issues associated with ascertainment bias may still apply (Brumfield *et al*. 2003), although they can be accommodated relatively easily at the planning stage of such multigene studies.

Acknowledgements

We thank J. J. Smith and J. A. Walker for laboratory assistance and Dolores Huacuz and the late Virginia Graue for invaluable assistance in the field. *Ambystoma ordinarium* samples were collected with a Permiso de Pesca de Fomento No. 150199-213-03. This publication utilized computing resources provided by the University of Kentucky subcontract on NIH 2P20RR016481-04 from the National Center for Research Resources. This material is based upon work supported by grants #NSF IOB-0242833, NSF DEB-0213155, and NIH NCRR-5R24RR16344-05.

Supplementary material

The supplementary material is available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2961/MEC2961sm.htm

Fig. S1 Statistical parsimony haplotype networks for the *col1a1* and *ctg1908* nuclear loci. Recovered haplotypes across the *Ambystoma tigrinum* complex are denoted as open circles or diamonds. Smaller filled circles represent undetected, but inferred, intermediate haplotypes. Single lines represent one mutational event. Haplotypes recovered from *A. ordinarium* are given an 'ord' notation and are lightly shaded. Populations from which a particular *Ambystoma ordinarium* haplotype was found are given in bold numbers inside parentheses. Haplotypes recovered from all remaining *A. tigrinum* complex species are given a 'tig' notation with a taxon label placed next to the haplotype. Species that share haplotypes are presented in boxes. Haplotypes recovered from *Ambystoma dumerilii* are presented as diamonds. The *col1a1* haplotypes tig1 (*Ambystoma californiense*) and tig18 (*A. t. tigrinum*) could not be linked to the network under the 95% criterion of statistical parsimony.

Fig. S2 Statistical parsimony networks for the *g1c12* and *g3d7* nuclear loci. Haplotype network descriptions are as in supplementary Fig. S1. The *g3d7* haplotypes tig1 (*Ambystoma californiense*) and tig6 (*Ambystoma tigrinum melanostictum*) could not be linked to the network under the 95% criterion of statistical parsimony.

Fig. S3 Statistical parsimony networks for the *ctg1506* and *g1d6* nuclear loci. Haplotype network descriptions are as in supplementary Fig. S1. The *ctg1506* haplotype tig18 (*Ambystoma tigrinum tigrinum*) and the *g1d6* haplotype tig17 (*Ambystoma mexicanum*) could not be linked to the networks under the 95% criterion of statistical parsimony.

Fig. S4 Statistical parsimony networks for the *g1f1* and *dlx3* nuclear loci. Haplotype network descriptions are as in supplementary Fig. S1. The *g1f1* haplotype tig1 (*Ambystoma californiense*) could not be linked to the network under the 95% criterion of statistical parsimony.

Fig. S5 Majority-rule consensus phylograms resulting from Bayesian phylogenetic analysis of the nuclear loci *col1a1* and *ctg1908*. The locus name along with the best-fit evolutionary model used, and the resulting mean lnL from the Bayesian posterior distribution are presented to the lower left of each tree. Thick branches lead to alleles sampled from *Ambystoma ordinarium*. The *Ambystoma tigrinum* complex species that a particular allele was sampled from is given in parentheses following the haplotype name (amb, *Ambystoma amblycephalum*; and, *Ambystoma andersoni*; cal, *Ambystoma californiense*; dum, *Ambystoma dumerilii*; fla, *Ambystoma flavipiperatum*; gra, *Ambystoma granulosum*; ler, *Ambystoma lermaense*; mav, *Ambystoma tigrinum mavortium*; mel, *Ambystoma tigrinum melanostictum*; mex, *Ambystoma mexicanum*; riv, *Ambystoma rivulare*; ros, *Ambystoma rosaceum*; tay, *Ambystoma taylori*; tig, *Ambystoma tigrinum tigrinum*; vel, *Ambystoma velasci*). The sampling localities that a particular *A. ordinarium* allele was found is also given in parentheses. *A. dumerilii* haplotypes are marked with double asterisks.

Fig. S6 Majority-rule consensus phylograms resulting from Bayesian phylogenetic analysis of the nuclear loci *g1c12* and *g3d7*. Tree descriptions are as in supplementary Fig. S5.

Table S1 Locality and specimen information for tiger salamander samples used in this study and the *col1a1* haplotypes sequenced from them. Bold text numbers in parentheses designate the locality number of the sample as given in Shaffer & McKnight (1996). Plain text numbers in parentheses are H. B. Shaffer specimen numbers.

Table S2 Locality and specimen information for tiger salamander samples used in this study and the *dlx3* haplotypes sequenced from them. Bold text numbers in parentheses designate the locality number of the sample as given in Shaffer & McKnight (1996). Plain text numbers in parentheses are H. B. Shaffer specimen numbers.

Table S3 Locality and specimen information for tiger salamander samples used in this study and the *ctg1506* haplotypes sequenced from them. Bold text numbers in parentheses designate the locality number of the sample as given in Shaffer & McKnight (1996). Plain text numbers in parentheses are H. B. Shaffer specimen numbers.

Table S4 Locality and specimen information for tiger salamander samples used in this study and the *ctg1908* haplotypes sequenced from them. Bold text numbers in parentheses designate the locality number of the sample as given in Shaffer & McKnight (1996). Plain text numbers in parentheses are H. B. Shaffer specimen numbers.

Table S5 Locality and specimen information for tiger salamander samples used in this study and the *g1c12* haplotypes sequenced from them. Bold text numbers in parentheses designate the locality number of the sample as given in Shaffer & McKnight (1996). Plain text numbers in parentheses are H. B. Shaffer specimen numbers.

Table S6 Locality and specimen information for tiger salamander samples used in this study and the *g1d6* haplotypes sequenced from them. Bold text numbers in parentheses designate the locality number of the sample as given in Shaffer & McKnight (1996). Plain text numbers in parentheses are H. B. Shaffer specimen numbers.

Table S7 Locality and specimen information for tiger salamander samples used in this study and the *g1f1* haplotypes sequenced from them. Bold text numbers in parentheses designate the locality number of the sample as given in Shaffer & McKnight (1996). Plain text numbers in parentheses are H. B. Shaffer specimen numbers.

Table S8 Locality and specimen information for tiger salamander samples used in this study and the *g3d7* haplotypes sequenced from them. Bold text numbers in parentheses designate the locality number of the sample as given in Shaffer & McKnight (1996). Plain text numbers in parentheses are H. B. Shaffer specimen numbers.

References

Albertson RC, Markert JA, Danley PD, Kocher TD (1999) Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences, USA*, **96**, 5107–5110.

Anderson JD (1975) *Ambystoma ordinarium*. *Catalogue of American Amphibians and Reptiles*, **164**, 1–2.

Anderson JD, Worthington RD (1971) The life history of the Mexican salamander *Ambystoma ordinarium* Taylor. *Herpetologica*, **27**, 165–176.

Avise JC, Ball RM (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology*, **7**, 45–67.

Baum DA, Shaw KL (1995) Genealogical perspectives on the species problem. In: *Experimental and Molecular Approaches to Plant Biosystematics* (eds Hoch PC, Stephenson AG), pp. 289–303. Missouri Botanical Garden, St. Louis, Missouri.

Bhangale TR, Rieder MJ, Livingston RJ, Nickerson DA (2005) Comprehensive identification and characterization of diallelic insertion–deletion polymorphisms in 330 human candidate genes. *Human Molecular Genetics*, **14**, 59–69.

Broughton RE, Harrison RG (2003) Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics*, **163**, 1389–1401.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.

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

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Coyne JA, Orr HA (1989) Patterns of speciation in *Drosophila*. *Evolution*, **43**, 362–381.

Dettman JR, Jacobson DJ, Taylor JW (2003) A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution*, **57**, 2703–2720.

Fitzpatrick BM (2002) Molecular correlates of reproductive isolation. *Evolution*, **56**, 191–198.

Fitzpatrick BM, Shaffer HB (2004) Environment-dependent admixture dynamics in a tiger salamander hybrid zone. *Evolution*, **58**, 1282–1293.

Freeland JR, Boag PT (1999) Phylogenetics of Darwin's finches: paraphyly in the tree-finches, and two divergent lineages in the warbler finch. *Auk*, **116**, 577–588.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics*, **34**, 397–423.

Gittenberger E (1991) What about nonadaptive radiation? *Biological Journal of the Linnean Society*, **43**, 263–272.

Glor RE, Gifford ME, Larson A *et al*. (2004) Partial island submergence and speciation in an adaptive radiation: a multilocus analysis of the Cuban green anoles. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **271**, 2257–2265.

Grant PR, Grant BR, Keller L, Markert JA, Petren K (2003) Inbreeding and interbreeding in Darwin's finches. *Evolution*, **57**, 2911–2916.

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution*, **16**, 700–706.

Hare MP, Avise JC (1998) Population structure in the American oyster as inferred by nuclear gene genealogies. *Molecular Biology and Evolution*, **15**, 119–128.

Highton R (2000) Detecting cryptic species using allozyme data. In: *The Biology of Plethodontid Salamanders* (eds Bruce RC, Jaeger R, Houck LD), pp. 215–241. Academic/Plenum Publishers, New York.

Hoelzer GA (1997) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution*, **51**, 622–626.

Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution*, **56**, 1557–1565.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Hudson RR, Turelli M (2003) Stochasticity overrules the 'three-times rule': genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, **57**, 182–190.

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

Lewis PO (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, **50**, 913–925.

Machado CA, Hey J (2002) The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 1193–1202.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial gene trees versus nuclear gene trees. *Evolution*, **49**, 718–726.

Morando M, Avila LJ, Sites JW Jr (2003) Sampling strategies for delimiting species: genes, individuals, and populations in the

Liolaemus elongates-kriegi complex (Squamata: Lioleamidae) in Andean-Patagonian South America. *Systematic Biology*, **52**, 159–185.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution, and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Nagl S, Tichy H, Mayer WE, Takahata N, Klein J (1998) Persistence of neutral polymorphisms in Lake Victoria cichlid fish. *Proceedings of the National Academy of Sciences, USA*, **95**, 14238–14243.

Navajas M, Boursot P (2003) Nuclear ribosomal monophyly versus mitochondrial DNA polyphyly in two closely related mite species: the influence of life history and molecular drive. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, S124–S127.

Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: *Evolutionary Processes and Theory* (eds Karlin S, Nevo E), pp. 515–534. Academic Press, New York.

Palumbi SR, Cipriano F, Hare MP (2001) Predicting nuclear gene coalescence from mitochondrial data: the three times rule. *Evolution*, **55**, 859–868.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568–583.

Parker A, Kornfield I (1997) Evolution of the mitochondrial DNA control region in the mbuna (Cichlidae) species flock of Lake Malawai, East Africa. *Journal of Molecular Evolution*, **45**, 70–83.

Patton JL, Smith MF (1994) Paraphyly, polyphyly, and the nature of species boundaries in pocket gophers (genus *Thomomys*). *Systematic Biology*, **43**, 11–26.

Poe S, Chubb AL (2004) Birds in a bush: five genes indicate explosive evolution of avian orders. *Systematic Biology*, **58**, 404–415.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Putta S, Smith JJ, Walker JA *et al.* (2004) From biomedicine to natural history research: EST resources for ambystomatid salamanders. *BMC Genomics*, **5**, 54.

de Queiroz K (1999) The general lineage concept of species and the defining properties of the species category. In: *Species: New Interdisciplinary Essays* (ed. Wilson RA), pp. 49–89. MIT Press, Cambridge, Massachusetts.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exacts test and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Riley SPD, Shaffer HB, Voss SR, Fitzpatrick BM (2003) Hybridization between a rare, native tiger salamander (*Ambystoma californiense*) and its introduced congener. *Ecological Applications*, **13**, 1263–1275.

Rosenberg NA (2003) The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution*, **57**, 1465–1477.

Rozas J, Sánchez-DelBarrio JC, Messegyer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Ruber L, Meyer A, Sturmbauer C, Verheyen E (2001) Population structure in two sympatric species of the Lake Tanganyika cichlid tribe Eretmodini: evidence for introgression. *Molecular Ecology*, **10**, 1207–1225.

Samuels AK, Weisrock DW, Smith JJ *et al.* (2005) Transcriptional and phylogenetic analysis of five complete ambystomatid salamander mitochondrial genomes. *Gene*, **349**, 43–53.

Sato A, Figueroa F, O'hUigin C *et al.* (1999) Phylogeny of Darwin's finches as revealed by mtDNA sequences. *Proceedings of the National Academy of Sciences, USA*, **96**, 5101–5106.

Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford, UK.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.

Shaffer HB (1984a) Evolution in a paedomorphic lineage. I. An electrophoretic analysis of the Mexican ambystomatid salamanders. *Evolution*, **38**, 1194–1206.

Shaffer HB (1984b) Evolution in a paedomorphic lineage. II. Allometry and form in the Mexican ambystomatid salamanders. *Evolution*, **38**, 1207–1218.

Shaffer HB, Lauder GV (1985) Patterns of variation in aquatic ambystomatid salamanders: kinematics of the feeding mechanism. *Evolution*, **39**, 83–92.

Shaffer HB, McKnight ML (1996) The polytypic species revisited: genetic differentiation and molecular phylogenetics of the tiger salamander *Ambystoma tigrinum* (Amphibia: Caudata) complex. *Evolution*, **50**, 417–433.

Shaffer HB, Pauley GB, Oliver JC, Trenham PC (2004) The molecular phylogenetics of endangerment: cryptic variation and historical phylogeography of the California tiger salamander, *Ambystoma californiense*. *Molecular Ecology*, **13**, 3033–3049.

Shaffer HB, Voss SR (1996) Phylogenetic and mechanistic analysis of a developmentally integrated character complex: alternate life history modes in ambystomatid salamanders. *American Zoologist*, **36**, 24–35.

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

Simpson GG (1953) *The Major Features of Evolution*. Columbia University Press, New York.

Sites JW Jr, Marshall JC (2004) Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 199–227.

Smith JJ, Kump DK, Walker JA, Parichy DM, Voss SR (2005) A comprehensive expressed sequence tag linkage map for tiger salamander and Mexican axolotl: enabling gene mapping and comparative genomics in *Ambystoma*. *Genetics*, **171**, 1161–1171.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Swofford DL (2002) PAUP: *Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Takahashi K, Terai Y, Nishida M, Okada N (2001) Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. *Molecular Biology and Evolution*, **18**, 2057–2066.

Taylor EH (1939) New salamanders from Mexico with a discussion of certain known forms. *University of Kansas Science Bulletin*, **26**, 407–439.

Theodorides K, De Riva A, Gomez-Zurita J, Foster PG, Vogler AP (2002) Comparison of EST libraries from seven beetle species: towards a framework for phylogenomics of the Coleoptera. *Insect Molecular Biology*, **11**, 467–475.

Trenham PC, Koenig WD, Shaffer HB (2001) Spatially autocorrelated demography and interpond dispersal in the pond breeding salamander *Ambystoma californiense*. *Ecology*, **82**, 3519–3530.

Voss SR, Shaffer HB (1996) What insights into the developmental biology of urodeles does the study of interspecific hybrids provide? *International Journal of Developmental Biology*, **40**, 885–893.

Voss SR, Smith JJ, Gardiner DM, Parichy DM (2001) Conserved vertebrate chromosome segments in the large salamander genome. *Genetics*, **158**, 735–746.

Wakeley J (2000) The effects of population subdivision on the genetic divergence of populations and species. *Evolution*, **54**, 1092–1101.

Watts RC, Bradbury JP (1982) Paleoecological studies at Lake Patzcuaro on the west-central Mexican Plateau and at Chalco in the basin of Mexico. *Quaternary Research*, **17**, 56–70.

Weisrock DW, Kozak KH, Larson A (2005) Phylogeographic analysis of mitochondrial gene flow and introgression in the salamander, *Plethodon shermani*. *Molecular Ecology*, **14**, 1457–1472.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

This work represents the latest collaborative effort between the Voss and Shaffer laboratories on the evolutionary genetics of speciation in the tiger salamander complex. David Weisrock is currently a postdoc in the Voss laboratory, with interests in multigene approaches to phylogenetics and phylogeography. Randal Voss's primary research interests are in amphibian genomics and quantitative, population, and conservation genetics. Brad Shaffer's laboratory group has research interests in amphibian genetics, conservation and systematics. Brian and Shonna Storz contributed to the field work in Mexico while they were undergraduates in the Shaffer laboratory.
