doi:10.1111/j.1558-5646.2011.01480.x

DELIMITING SPECIES USING MULTILOCUS DATA: DIAGNOSING CRYPTIC DIVERSITY IN THE SOUTHERN CAVEFISH, *TYPHLICHTHYS SUBTERRANEUS* (TELEOSTEI: AMBLYOPSIDAE)

Matthew L. Niemiller,[1,3] Thomas J. Near,[2] and Benjamin M. Fitzpatrick[1]

[1]*Department of Ecology and Evolutionary Biology, University of Tennessee, 569 Dabney Hall, Knoxville, Tennessee 37996*

[2]*Department of Ecology and Evolutionary Biology, Yale University, 370 ESC, New Haven, Connecticut 06520*

[3]*E-mail: mniemill@utk.edu*

Received June 17, 2011
Accepted September 25, 2011

A major challenge facing biodiversity conservation and management is that a significant portion of species diversity remains undiscovered or undescribed. This is particularly evident in subterranean animals in which species delimitation based on morphology is difficult because differentiation is often obscured by phenotypic convergence. Multilocus genetic data constitute a valuable source of information for species delimitation in such organisms, but until recently, few methods were available to objectively test species delimitation hypotheses using genetic data. Here, we use recently developed methods for discovering and testing species boundaries and relationships using a multilocus dataset in a widely distributed subterranean teleost fish, *Typhlichthys subterraneus*, endemic to Eastern North America. We provide evidence that species diversity in *T. subterraneus* is currently underestimated and that the picture of a single, widely distributed species is not supported. Rather, several morphologically cryptic lineages comprise the diversity in this clade, including support for the recognition of *T. eigenmanni*. The high number of cryptic species in *Typhlichthys* highlights the utility of multilocus genetic data in delimiting species, particularly in lineages that exhibit slight morphological disparity, such as subterranean organisms. However, results depend on sampling of individuals and loci; this issue needs further study.

KEY WORDS: **Bayesian, cave, conservation, phylogenetics, speciation, species tree, subterranean.**

A major problem facing the conservation and management of biodiversity is that a significant fraction of species remains unidentified and unknown to science (Wilson 2003). This is aggravated by lack of consensus on a definition of the term "species." However, most biologists agree that the "species phenomenon" is real. The species phenomenon is the fact that contemporary biological diversity is not a continuum, but rather shows consistent discontinuities along morphological, genetic, and ecological axes (Dobzhansky 1937; Sterelny 1999; Coyne and Orr 2004; Hausdorf 2011). Groups of organisms separated by these discontinuities (or a subset agreed not to include discontinuities between sexes or life stages) have traditionally been given taxonomic names and ranks. The continuities and discontinuities represented by the species phenomenon are among the most important emergent patterns in evolutionary biology (Hausdorf 2011). From the perspective of conservationists concerned with preserving the natural structure of biodiversity, the loss of an entire group of organisms (i.e., extinction) is a more significant and regrettable event than the death of individuals.

The identification and documentation of species and evolutionary significant units (ESUs; Ryder 1986, Waples 1991) can have significant effects on biodiversity assessments, conservation

programs, biological control, and ecological and evolutionary studies (Isaac et al. 2004; Beheregaray and Caccone 2007; Bickford et al. 2007; Bortolus 2008). However, our understanding of species delimitation is poor for several groups of organisms, particularly those that exhibit little if any morphological differentiation, and this could compromise our ability to study and conserve such taxa (Bickford et al. 2007).

In recent years, phylogeographic analyses have uncovered considerable levels of cryptic phylogenetic diversity (Highton et al. 1989; Gomez et al. 2002; Molbo et al. 2003; Pfenninger and Schwenk 2007; Hollingsworth and Near 2009), due, in large part, to the ever-increasing ease of generating molecular data. The discovery of cryptic diversity in subterranean ecosystems has been particularly prevalent, as several studies have revealed considerable genetic variation in wide-ranging, morphologically indistinct taxa (reviewed in Juan and Emerson 2010). The long-term stability of subterranean ecosystems compared to those on the surface and the highly fragmented hydrological structure of groundwater are thought to promote high endemism (Gibert and Deharveng 2002; Verovnik et al. 2003; Finston et al. 2007). Most widely distributed subterranean species are hypothesized to consist of multiple, unrecognized cryptic species (Barr and Holsinger 1985; Culver et al. 1995; Finston et al. 2007; Lefebure et al. 2006, 2007; Trontelj et al. 2009). Strong selective pressures and the isolated nature of subterranean ecosystems lead to the apparent discord between morphological and genetic differentiation in these taxa due to convergence or parallel evolution (Culver et al. 1995; Wiens et al. 2003; Finston et al. 2007; Culver and Pipan 2009). Among larger subterranean, aquatic macrofauna (e.g., crayfish, fish, and salamanders), few species have broad distributions. Recent molecular analyses of the only aquatic, obligate cave-dwelling vertebrate in Europe, the European Cave Salamander (*Proteus anguinus*), also revealed cryptic molecular diversity (Goricki and Trontelj 2006; Trontelj et al. 2009), as six distinct lineages were identified within the one previously described species.

How these patterns of genetic variation should be interpreted taxonomically depends on the troublesome question of what categories such as "species" and "ESU" are intended to represent. The question has no single correct answer, but it remains important because many conservation laws and regulations use taxonomic categories as operational units. Indeed, a pragmatic species definition for conservation is "a distinct group of organisms meriting independent legal status because extinction of such a group would constitute a substantial loss of biological diversity" (Pasachnik et al. 2010). In addition, even with an agreed upon definition of "species," it is not always straightforward to relate this idea to observable patterns in genetic data (Shaffer and Thompson 2007). Although many protocols for species delimitation have been proposed (e.g., Sites and Marshall 2004), they depend on specifying null and alternative hypotheses about predefined species designations. Recently, O'Meara (2010) proposed a method using multilocus data that does not require a priori designation of species. The method assumes that there is gene flow within species but little or no gene flow between species. Given this assumption, gene trees of unlinked loci should exhibit congruence on branches between species but not within species. Within species, gene flow and independent assortment make each gene tree independent, and if populations are large and there is no selection, each is a random draw from the neutral coalescent. Thus, incongruence among gene trees is best explained by common membership in a species gene pool. O'Meara's method uses a heuristic search for the set of delimited species trees that minimizes gene tree parsimony score and maximizes the number of species lineages consistent with a set of estimated gene trees.

Here, we delimit species and infer species relationships using newly developed methods to species delimitation within a widespread taxon for which morphology provides little other diagnostic information. In Eastern North America, the Southern Cavefish (*Typhlichthys subterraneus*) (family Amblyopsidae) has the largest known distribution of any subterranean fish in the world, spanning more than 5° of latitude and over 140,000 km^2 (Proudlove 2006; Niemiller and Poulson 2010) throughout caves and karst of the Interior Low Plateau and Ozark Highlands (Fig. 1). As many as four species of *Typhlichthys* have been described (*T. subterraneus*, *T. osborni*, *T. wyandotte*, and *T. eigenmanni*); however, all species were synonymized under *T. subterraneus* due to a lack of morphological variation (Woods and Inger 1957). Nevertheless, some populations have been noted to exhibit subtle morphological differences from typical *T. subterraneus* and potentially represent undescribed species (Cooper and Beiter 1972; Burr and Warren 1986; Niemiller and Poulson 2010). Because of its large distribution across several major hydrological units and documentation of cryptic diversity in other wide-ranging subterranean taxa, several authors have hypothesized that *T. subterraneus* is a species complex comprised of several morphologically cryptic species, possibly resulting from several parallel colonizations by a surface-dwelling common ancestor (Swofford 1982; Barr and Holsinger 1985; Holsinger 2000; Niemiller and Poulson 2010). The few studies that have examined genetic variation in *T. subterraneus* have found considerable genetic differentiation among morphologically similar populations structured among hydrological units (Swofford 1982; Bergstrom et al. 1995; Bergstrom 1997; Niemiller and Fitzpatrick 2008). The discovery of significant genetic variation across the range of *T. subterraneus* warrants further detailed investigations of species delimitation in these cavefish.

We take an integrated approach using a multilocus genetic dataset, extensive sampling across the range of nominal *T. subterraneus*, and distributional and hydrological data to investigate the

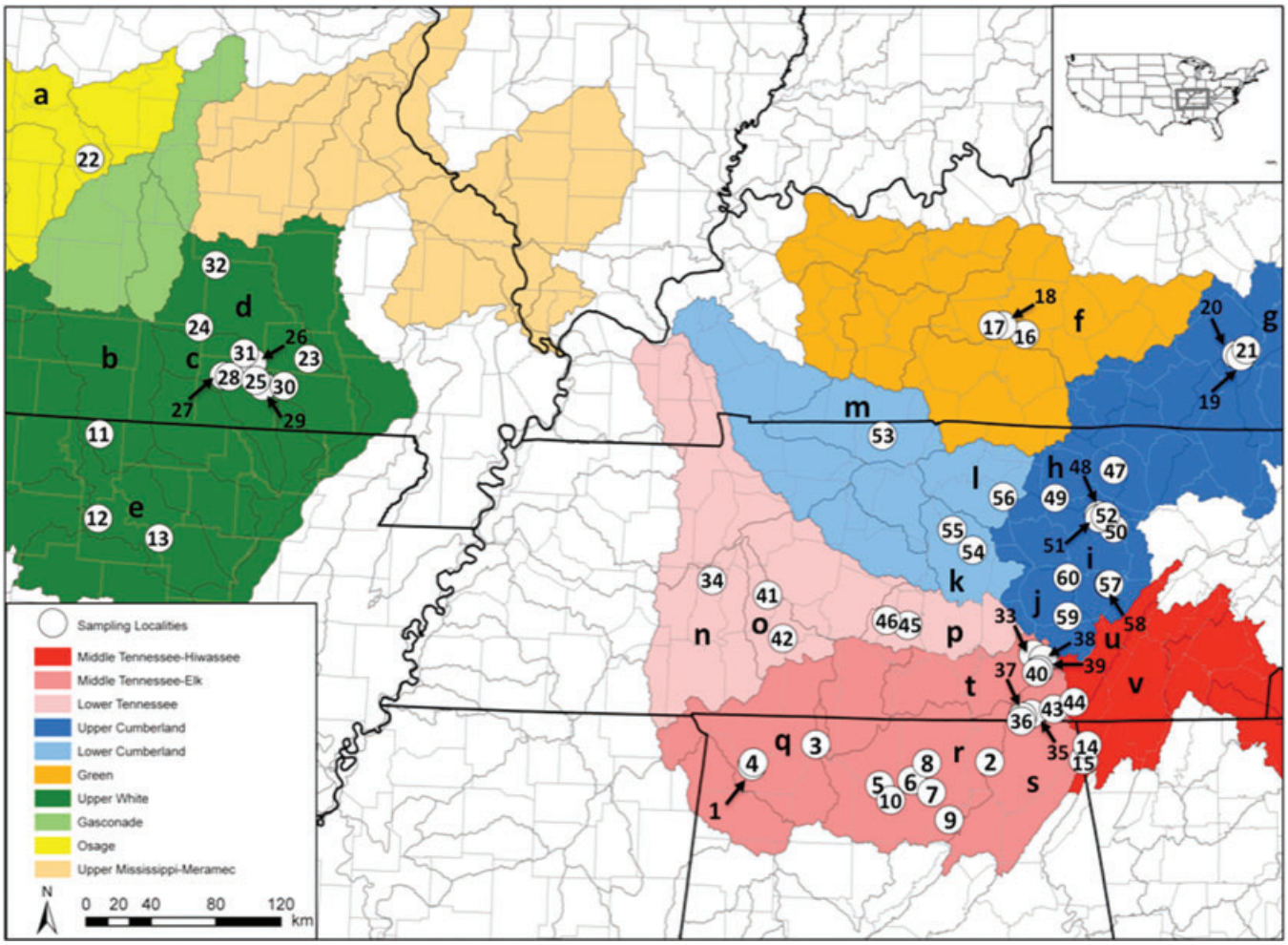

Figure 1. Map illustrating the distribution and sampling localities of *Typhlichthys*. Numbered localities correspond to populations listed in Table 1. Major river basins (HUC6 watersheds) are color-coded on the map. County borders also are outlined. Lower case letters identify subbasins: (a) Osage-Lake of the Ozarks, (b) North Fork White, (c) Eleven Point, (d) Current, (e) Middle White, (f) Upper Green, (g) Upper Cumberland-Lake Cumberland, (h) Upper Cumberland-Cordell Hull, (i) Caney Fork, (j) Collins, (k) Stones, (l) Lower Cumberland-Old Hickory Lake, (m) Red, (n) Lower Tennessee-Beech, (o) Buffalo, (p) Upper Duck, (q) Tennessee-Pickwick Lake, (r) Tennessee-Wheeler Lake, (s) Tennessee-Guntersville Lake, (t) Upper Elk, (u) Sequatchie, and (v) Middle Tennessee-Chickamauga.

phylogeography and diversity among populations in this broadly distributed species. Using newly developed methods to delimit species and species trees, we (1) infer whether *T. subterraneus* comprises morphologically cryptic yet genetically distinct lineages or a single widely distributed species; (2) estimate the number of and phylogenetic relationships among putative species; and (3) examine whether limited dispersal among hydrological drainages and ecoregions has promoted diversity within *Typhlichthys* by testing for an association of genetic divergence with hydrological structure. We first delimit putative species in *Typhlichthys* without making a priori designations using O'Meara's (2010) method while examining the effects of number of samples and number of loci on species delimitation. We then validate delimited species assignments using other recently developed approaches to delimit species (Yang and Rannala 2010), test for taxonomic distinctiveness (Cummings et al. 2008), and infer relationships among delimited species (Heled and Drummond 2010).

Materials and Methods

SPECIMEN AND TISSUE COLLECTION

Specimens and tissue samples (fin clips) were collected from 60 populations throughout the range of *T. subterraneus* in Alabama, Arkansas, Georgia, Kentucky, Missouri, and Tennessee (Table 1; Fig. 1). One to five samples were collected and analyzed from each locality, as this taxon is a species of conservation concern in several states throughout its distribution. Additionally, we collected representative samples from other amblyopsids to

Table 1. Locality information, including county, state, sample size, major hydrological basin, subbasin (in parentheses), and ecoregion, and delimited species assignments for 60 populations of *Typhlichthys*. Delimited species assignments are from O'Meara's (2010) method using three-, six-, and nine-gene datasets (number of individuals in parentheses). For the three-gene dataset, analyses were conducted using 135, 60, and 20 individuals. For the six-gene dataset, analyses were conducted using 60 and 20 individuals, and 20 individuals were used in the nine-gene analysis.

							Delimited species assignment		
No.	Locality	County	State	*n*	Basin (subbasin)	Ecoregion	three-gene (135, 60, 20)	six-gene (60, 20)	nine-gene (20)
1	McKinney Pit	Colbert	AL	4	Tennessee (TN-Pickwick)	Interior Low Plateau	F, F, G	I, F	F
2	Guess Creek Cave	Jackson	AL	1	Tennessee (TN-Wheeler)	Southwestern Appalachians	L, F, F	F, F	D
3	Davis Bat Cave	Lauderdale	AL	1	Tennessee (TN-Pickwick)	Interior Low Plateau	F, F	I	
4	Key Cave	Lauderdale	AL	2	Tennessee (TN-Pickwick)	Interior Low Plateau	R, F	J	
5	White Spring Cave	Limestone	AL	1	Tennessee (TN-Wheeler)	Interior Low Plateau	C, L, F	F, F	D
6	Bobcat Cave	Madison	AL	1	Tennessee (TN-Wheeler)	Interior Low Plateau	D, L	I	
7	Muddy Cave	Madison	AL	1	Tennessee (TN-Wheeler)	Interior Low Plateau	D, D	C	
8	Shelta Cave	Madison	AL	3	Tennessee (TN-Wheeler)	Interior Low Plateau	L, F	F	
9	Beech Spring Cave	Marshall	AL	1	Tennessee (TN-Wheeler)	Southwestern Appalachians	L, F	F	
10	Cave Spring Cave	Morgan	AL	1	Tennessee (TN-Wheeler)	Interior Low Plateau	F, F	I	
11	Norfolk Lake	Baxter	AR	1	White (North Fork White)	Ozark Highlands	Q, I	B	
12	Alexander Cave	Stone	AR	2	White (Middle White)	Ozark Highlands	B, B, B	B, B	B
13	Ennis Cave	Stone	AR	1	White (Middle White)	Ozark Highlands	B, B	B	
14	Limestone Caverns	Dade	GA	2	Tennessee (TN-Chickamauga)	Ridge and Valley	E, E	E	
15	Long's Rock Wall Cave	Dade	GA	3	Tennessee (TN-Chickamauga)	Ridge and Valley	E, E, E	E, E	E
16	L and N Railroad Cave	Barren	KY	4	Green (Upper Green)	Interior Low Plateau	G, G	G	
17	Mammoth Cave	Edmonson	KY	4	Green (Upper Green)	Interior Low Plateau	G, G	G	
18	Sander's Cave	Edmonson	KY	4	Green (Upper Green)	Interior Low Plateau	G, G	G	
19	Dave's Cave	Pulaski	KY	3	Cumberland (Cumberland-Lake Cumberland)	Southwestern Appalachians	J, J	J	
20	Drowned Rat Cave	Pulaski	KY	3	Cumberland (Cumberland-Lake Cumberland)	Southwestern Appalachians	J, J	J	

Continued.

Table 1. Continued.

No.	Locality	County	State	*n*	Basin (subbasin)	Ecoregion	Delimited species assignment		
							three-gene (135, 60, 20)	six-gene (60, 20)	nine-gene (20)
21	Well's Cave	Pulaski	KY	1	Cumberland (Cumberland-Lake Cumberland)	Southwestern Appalachians	J, J, A	J, F	A
22	Carroll Cave	Camden	MO	4	Osage (Osage-Lake of the Ozarks)	Ozark Highlands	P, I, B	B, B	B
23	Coalbank Cave	Carter	MO	3	White (Current)	Ozark Highlands	I, I	B	
24	Concolor Cave	Howell	MO	3	White (Current)	Ozark Highlands	I, I	B	
25	Bliss Camp Cave	Oregon	MO	2	White (Eleven Point)	Ozark Highlands	I, I	B	
26	Falling Spring Cave	Oregon	MO	2	White (Eleven Point)	Ozark Highlands	I, I	B	
27	Posy Spring Cave	Oregon	MO	4	White (Eleven Point)	Ozark Highlands	I, I	B	
28	Roaring Spring Cave	Oregon	MO	3	White (Eleven Point)	Ozark Highlands	I, I	B	
29	Turner Spring Cave	Oregon	MO	1	White (Eleven Point)	Ozark Highlands	I, I	B	
30	Panther Cave	Ripley	MO	2	White (Current)	Ozark Highlands	I, I	B	
31	Brawley Cave	Shannon	MO	1	White (Eleven Point)	Ozark Highlands	I, I	B	
32	Flying W Cave	Shannon	MO	2	White (Current)	Ozark Highlands	Q, I	B	
33	Blowing Springs Cave	Coffee	TN	4	Tennessee (Upper Elk)	Southwestern Appalachians	C, C	C	
34	Baugus Cave	Decatur	TN	4	Tennessee (TN-Beech)	Interior Low Plateau	S, F, G	F, F	F
35	Garner Spring Cave	Franklin	TN	4	Tennessee (TN-Guntersville)	Southwestern Appalachians	K, K	K	
36	Little Crow Creek Cave	Franklin	TN	2	Tennessee (TN-Guntersville)	Southwestern Appalachians	K, K	K	
37	Salt River Cave	Franklin	TN	5	Tennessee (TN-Guntersville)	Southwestern Appalachians	K, K	K	
38	Big Mouth Cave	Grundy	TN	4	Tennessee (Upper Elk)	Southwestern Appalachians	C, C, C	C, C	C
39	Crystal Cave	Grundy	TN	3	Tennessee (Upper Elk)	Southwestern Appalachians	C, C, C	C, C	C
40	Trussell Cave	Grundy	TN	1	Tennessee (Upper Elk)	Southwestern Appalachians	C, C	C	
41	Cave Branch Cave	Hickman	TN	1	Tennessee (Buffalo)	Interior Low Plateau	N, N, F	F, F	D
42	Allens Creek Cave	Lewis	TN	1	Tennessee (Buffalo)	Interior Low Plateau	N, N	F	
43	Lost Pig Cave	Marion	TN	1	Tennessee (TN-Guntersville)	Southwestern Appalachians	M, M, E	H, E	E
44	Pryor Cave Spring	Marion	TN	1	Tennessee (Sequatchie)	Southwestern Appalachians	M, M	H	
45	Gallagher Cave South	Marshall	TN	2	Tennessee (Upper Duck)	Interior Low Plateau	D, D	D	
46	Pompie Cave	Maury	TN	1	Tennessee (Upper Duck)	Interior Low Plateau	D, D, D	D, D	C

Continued.

Table 1. Continued.

No.	Locality	County	State	*n*	Basin (subbasin)	Ecoregion	Delimited species assignment		
							three-gene (135, 60, 20)	six-gene (60, 20)	nine-gene (20)
47	East Water Supply Cave	Overton	TN	1	Cumberland (Cumberland-Cordell Hull)	Interior Low Plateau	A, A	A	
48	Anderson Spring Cave	Putnam	TN	2	Cumberland (Caney Fork)	Interior Low Plateau	A, A, A	A, A	A
49	Bartlett Cave	Putnam	TN	2	Cumberland (Cumberland-Cordell Hull)	Interior Low Plateau	R, F	A	
50	Blind Fish Cave	Putnam	TN	2	Cumberland (Caney Fork)	Southwestern Appalachians	A, A	A	
51	Jacque's Cave	Putnam	TN	3	Cumberland (Caney Fork)	Southwestern Appalachians	A, A	A	
52	Stamp's Cave	Putnam	TN	2	Cumberland (Caney Fork)	Southwestern Appalachians	A, A	A	
53	Sinking Ridge Cave	Robertson	TN	2	Cumberland (Red)	Interior Low Plateau	G, G, F	G, F	B
54	Herring Cave	Rutherford	TN	3	Cumberland (Stones)	Interior Low Plateau	O, O, F	F, F	D
55	Patton's Cave	Rutherford	TN	4	Cumberland (Stones)	Interior Low Plateau	O, O	F	
56	Flat Rock Cave	Smith	TN	2	Cumberland (Cumberland-Old Hickory Lake)	Interior Low Plateau	D, D, D	D, C	C
57	Camps Gulf Cave	Van Buren	TN	1	Cumberland (Caney Fork)	Southwestern Appalachians	H, H, A	A, A	A
58	Camps Gulf Cave No. 2	Van Buren	TN	2	Cumberland (Caney Fork)	Southwestern Appalachians	H, H	A	
59	Blowing Cave	Warren	TN	1	Cumberland (Collins)	Interior Low Plateau	A, A, A	E, A	A
60	Jaco Spring Cave	Warren	TN	3	Cumberland Collins)	Interior Low Plateau	D, D, D	F, D	C

serve as outgroups: *Chologaster cornuta*, *Amblyopsis rosae*, and *Speoplatyrhinus poulsoni*.

DNA EXTRACTION, AMPLIFICATION, AND SEQUENCING

Genomic DNA was extracted using the Qiagen DNEasy Kit (Qiagen Inc., Valencia, CA). Polymerase chain reaction (PCR) was used to amplify one mitochondrial gene and eight nuclear genes (Table 2). The mitochondrial protein-coding gene NADH dehydrogenase 2 (*nd2*) was amplified using primers presented in Kocher et al. (1995). The nuclear encoded first intron of the ribosomal protein *s7* was amplified using primers presented in Chow and Hazama (1998) and exon 3 of the nuclear recombination activating gene 1 (*rag1*) was amplified using primers presented in Holcroft (2004). Six other nuclear protein-coding genes used in this study (zic family member 1, *zic1*; myosin heavy polypeptide 6, *myh6*; hypothetical protein LOC564097, *ptr*; T-box brain 1, *tbr1*; similar to SH3 and PX domain containing 3 gene, *sh3px3*; and pleiomorphic adenoma genelike 2, *plagl2*) were selected among putatively single-copy genes identified in Li et al. (2007). One hundred thirty-five *T. subterraneus* individuals from 60 populations were amplified for the *nd2*, *s7*, and *rag1* loci. We also amplified a single individual from each locality (60 individuals in total) for the nuclear loci *myh6*, *plagl2*, and *tbr1*. A subset of 20 *T. subterraneus* individuals was amplified for an additional three nuclear genes (*ptr*, *sh3px3*, and *zic1*) representing the major lineages and geographic cover of *Typhlichthys* identified from the *nd2*, *s7*, and *rag1* datasets (see below). Additionally, representative samples of all other amblyopsid species were amplified for all nine genes. PCR conditions followed protocols used in previous studies

Table 2. Loci and selected best-fit molecular evolutionary models for character partitions implemented in phylogenetic analyses.

Locus	Abbreviation	Length	Ploidy	Model of first codon	Model of second codon	Model of third codon	Model of intron
NADH dehydrogenase 2	*nd2*	1044	*n*	TVM+I+G	GTR+I+G	GTR+I+G	NA
Intron 1 of ribosomal protein S7	*s7*	841	2*n*	NA	NA	NA	HKY+G
Exon 3 of recombination activating gene 1	*rag1*	1446	2*n*	HKY+I	TVM+I	TVM+G	NA
Zic family member 1	*zic1*	855	2*n*	F81	F81	TVM	NA
Myosin heavy polypeptide 6	*myh6*	786	2*n*	HKY+I	HKY	TVM+I	NA
Hypothetical protein LOC564097	*ptr*	761	2*n*	TrN	TrN	TVM+G	NA
T-box brain 1	*tbr1*	705	2*n*	HKY	F81	HKY+I	NA
Similar to SH3 and PX domain containing 3 gene	*sh3px3*	760	2*n*	GTR	K81uf+I	TIM+I	NA
Pleiomorphic adenoma genelike 2	*plagl2*	603	2*n*	GTR	TVM	TVM	NA

NA = the gene does not contain the specified partition.

(Kocher et al. 1995; Holcroft 2004; Li et al. 2007). Clean PCR products were sequenced at the Molecular Systematics and Conservation Genetics Laboratory, Department of Ecology and Evolutionary, Yale University, New Haven, Connecticut, or the Molecular Biology Resource Facility, Division of Biology, University of Tennessee, Knoxville, Tennessee.

GENETIC ANALYSES

Forward and reverse sequences for each template were aligned and edited using SEQUENCHER version 4.5 (Gene Codes, Ann Arbor, MI) with ambiguous base calls verified manually by examining the electropherogram for each sequence. Resulting contigs were aligned using SEQUENCHER and MACCLADE version 4.07 (Maddison and Maddison 2005). Some individuals contained heterozygous genotypes for the sampled nuclear loci. Haplotype phase of nuclear sequences was inferred using PHASE version 2.1 (Stephens et al. 2001; Stephens and Scheet 2005). Unique DNA sequences generated for this study were accessioned into GenBank (HQ707644–HQ707817, HQ729504–HQ729677, JN592064–JN592443).

GENE TREE ESTIMATION

Gene trees for each locus were constructed using partitioned Bayesian analyses. Sequences for all but one gene (*s7*) represent coding regions. Therefore, each locus (except *s7*) was partitioned by codon. The best-fit models of molecular evolution for each partition were selected using the Akaike's Information Criterion (AIC) implemented in MODELTEST version 3.7 (Posada and Crandall 1998) and are listed for each partition in Table 2. Each locus was partitioned accordingly and unlinked allowing values for transition/transversion ratio, proportion of invariable sites, and among-site rate heterogeneity to vary across codon partitions during analyses. Bayesian posterior probabilities were estimated in MRBAYES 3.1 (Ronquist and Huelsenbeck 2003). Two independent runs using six Markov chains and temperature profiles at the default setting of 0.2 were conducted for 10 million generations, sampling every 100th generation. Random trees were used to begin each Markov chain and a molecular clock was not enforced. Approximately, the first two million generations (20%) were discarded as burn-in to ensure stationarity after examination of log-likelihood values for each Bayesian run using the program TRACER version 1.5 (Rambaut and Drummond 2007). Samples from the stationary distribution of trees were used to generate 50% majority-rule consensus trees for each locus.

"SPECIES" DELIMITATION

We used the nonparametric heuristic method described in O'Meara (2010) and implemented in the program BROWNIE version 2.1 (O'Meara et al. 2006) to jointly estimate the number of "species" and the species tree within *Typhlichthys*. This new approach apportions individuals into putative species and jointly estimates the species tree from a multigene dataset from multiple individuals. The key premise of the method is that there is ongoing sexual reproduction among the members of a species but no actual gene flow between species. This should result in recognizable genealogical patterns, given enough time. The method itself does not assume there are intrinsic biological barriers to gene exchange, nor does it specify that a certain fraction of gene trees must be monophyletic within species.

An important assumption of this and other coalescent-based, species delimitation methods is that there is no structure within species; that is, each delimited species is panmictic. If structure is present, however, and migration is low, lineages sampled from the same population are more likely to coalesce with each other than with lineages from other populations, increasing the time to coalescence of lineages from different populations. This results in gene trees that are more similar to each other than expected under neutral coalescence (O'Meara 2010), with long branches connecting populations within species and can result in over-splitting of the number of delimited species and inflate cryptic species diversity. Likewise, sample size will have an effect on species delimitation if within-species structure is present. The genealogical history of a sample taken from structured populations can be treated as a two-step process in which an initial burst of coalescent events occurs within populations with some migration events before the remaining lineages, each in separate populations, enters the unstructured coalescent process. These phases have been called the scattering phase and collecting phase, respectively (Wakeley 1999; Wakeley and Aliacar 2001). If migration is low among populations, all samples from a single population will coalesce into a single lineage during the scattering phase (Wilkins 2004). Many samples taken from the same population are likely to share or have similar alleles across multiple loci. As O'Meara's (2010) method attempts to minimize excess structure within each species while minimizing gene tree conflict, populations with many samples could potentially be inferred as individual species even if little genetic differentiation exists. For example, if 20 individuals were sampled from every population, each population could be inferred to be a distinct species. On the other hand, if single alleles are sampled from each population, the sample includes only the collecting phase, which can resemble a re-scaled neutral coalescent (Wakeley and Aliacar 2001).

We investigated the effects of sample size and number of loci on species delimitation using O'Meara's (2010) method in *Typhlichthys* by conducting multiple analyses varying both number of loci and number of samples included. In total, we conducted seven analyses. In the first six, we considered a single allele from each individual fish for 135 individuals for three genes (*nd2*, *s7*, and *rag1*), 60 individuals for three and six genes (*nd2*, *s7*, *rag1*, *myh6*, *plagl2*, and *tbr1*), and 20 individuals for three, six, and nine genes. The few heterozygotes included only closely related alleles and there was no difference, in practice, between randomly sampling one alternative or coding the heterozygous sites as uncertain. In the last analysis we included both alleles for each fish in the 60-individual, six-gene set. This approach doubles the sample size without accounting for the hierarchical structure of the sample, created by sampling pairs of alleles within individuals. There is currently no proper way to treat diploid samples in these analyses.

Heuristic searches were run using default settings with the following exceptions. For all datasets, the number of random starting species trees (NReps) was set to 100, all possible taxon reassignments on leaf splits were explored (Subsample = 1), and the minimum number of samples per species (MinSamp) was set to 2. The 50% majority-rule consensus gene trees were used as input trees. We conducted 10,000 independent runs for each dataset to find the optimal delimited species tree on the Newton cluster at the University of Tennessee, Knoxville.

ESTIMATING THE SPECIES TREE

To evaluate the consistency of the delimited species tree estimate, we used the species assignments from BROWNIE as input data for species tree estimation. Several methods recently have been proposed for inferring species trees from multiple gene trees given that assignments to species are known a priori (Carstens and Knowles 2007; Edwards et al. 2007; Liu and Pearl 2007). In the supermatrix approach (Rokas et al. 2003; Nylander et al. 2004; Rokas and Carroll 2005), sequences from multiple loci are concatenated and analyzed using traditional phylogenetic methods. However, this approach suffers from a number of limitations reviewed previously (Degnan and Rosenberg 2006; Kubatko and Degnan 2007). Therefore, we employed a recently developed Bayesian Markov Chain Monte Carlo (MCMC) method (Heled and Drummond 2010) that jointly estimates multiple gene trees embedded in a shared species tree under the multispecies coalescent. This method named *BEAST is implemented in the program BEAST version 1.6.1 (Drummond and Rambaut 2007) and assumes that incongruence among multiple gene trees is because of incomplete lineage sorting and not gene flow. *BEAST is considerably more accurate than supermatrix approaches and also offers advantages over another existing Bayesian method BEST (Liu and Pearl 2007; Liu et al. 2008), which, like *BEAST, estimates species tree topology, divergence times, and population sizes from multiple gene trees under the multiple coalescent. BEST assumes that population size is constant over a branch, the species tree prior is uniform, and also requires the designation of an outgroup. *BEAST offers greater flexibility of population size and species tree priors and does not require an outgroup (Heled and Drummond 2010). We conducted *BEAST analyses on all six datasets defining species a priori according to species delimitation inferred using the above methods in the program BROWNIE. Analyses were partitioned by locus and by codon position in protein-coding loci. Partition-specific models of nucleotide substitution (Table 2) were implemented, all parameters were unlinked across loci (not across data partition), and an uncorrelated lognormal (UCLN) model of rate variation was assumed for each partition. A Yule process speciation prior was used for the branching rates. Three independent runs were executed in BEAST with each run consisting of 200 million generations sampling every

20,000 generations. Resulting tree and log files from each run were combined using the program LogCombiner version 1.5.3 (http://beast.bio.ed.ac.uk/LogCombiner). Convergence of parameter values was assessed by plotting the marginal probabilities using the program Tracer version 1.5 (Rambaut and Drummond 2007). The first 50 million generations were discarded as burn-in. Pooled post-burn-in effective sample sizes for all parameters were >300, indicating that the pooled log file accurately represented the posterior distribution (Kuhner 2009).

VALIDATION OF DELIMITED SPECIES

We explored the validity of delimited species inferred using O'Meara's (2010) nonparametric method using two approaches. First, we conducted Bayesian species delimitation (Yang and Rannala 2010), a multilocus, coalescent-based method that includes prior information about population size and divergence times and uses reversible-jump Markov chain Monte Carlo (rjMCMC) to estimate the posterior distribution for different species delimitation models. This method accommodates the species phylogeny as well as lineage sorting due to ancestral polymorphism. We used the species tree inferred from *Beast analyses as the guide tree in each analysis.

Bayesian species delimitation was conducted using the program Bpp version 2.0 (Rannala and Yang 2003; Yang and Rannala 2010) for each delimited species dataset. The prior distributions on ancestral population size (θ) and root age (τ_0) were assigned gamma distributions of G(2,2000) and G(2,1000), respectively. Other divergence time parameters were assigned the Dirichlet prior (Yang and Rannala 2010: eq. 2). We used algorithm 0 with the fine-tuning parameter = 15.0, and each species delimitation model was assigned equal prior probability. Each rjMCMC analysis was run for 500,000 generations with a burn-in of 50,000 and run at least twice to confirm consistency between runs.

We also assessed the taxonomic distinctiveness of delimited species using the genealogical sorting index (*gsi*; Cummings et al. 2008) whereby a quantitative measure of the degree to which ancestry of delimited species is exclusive is generated for individual genes and for multilocus data. The relative degree of exclusive ancestry is on a scale from 0 to 1, where 1 indicates complete monophyly. Using this statistic, hypothesized delimited species can be tested against a null hypothesis of no divergence. We calculated an ensemble *gsi* (*egsi*) and *gsi* for each locus in each delimited species dataset using the Genealogical Sorting Index web server (http://www.genealogicalsorting.org). The 50% majority-rule consensus gene trees were used as input trees. The null hypothesis that the degree of exclusive ancestry is observed by chance alone (i.e., no divergence) was evaluated by estimating a *P* value using 10,000 permutations. Uneven sample sizes among groups can shift *P* values downward for smaller group sizes; therefore, significance was inferred at $P < 0.01$ (Polihronakis 2009; Gazis et al. 2011).

We also conducted a concatenated analyses on the 60-individual, six-gene dataset to compare the results of these newly developed methods to the traditional supermatrix approach. A fully partitioned analysis was run using MrBayes 3.1 where each locus was considered a partition (and codon position in each locus for protein-coding loci) and was assigned its own substitution model. All loci were assumed to have the same tree topology. We ran the MCMC analysis using the same conditions and generated the 50% majority-rule consensus tree as mentioned above.

POPULATION STRUCTURE

If interconnectivity of drainage basins influences patterns of phylogeographic structure in low vagility species, such as subterranean organisms, then cavefish populations located in different drainage subbasins should exhibit greater genetic divergence than those populations distributed within the same basin. Such a pattern has been observed in other subterranean (Niemiller et al. 2008) and surface species (Kozak et al. 2006) in the Interior Highlands of North America. Accordingly, we assessed spatial structure of genetic variation by conducting hierarchical analyses of molecular variance (AMOVAs; Excoffier et al. 1992) on uncorrected sequence divergences for the 135 individual datasets for the *nd2*, *s7*, and *rag1* loci in Arlequin 3.0 (Excoffier et al. 2005). Nesting was imposed in three ways. First, we grouped populations by major hydrological basins. Additionally, we examined the effects of grouping on genetic variance by hydrological subbasins, as several subbasins may exist within a single major hydrological basin (e.g., Tennessee River basin). Lastly, we also grouped populations by ecoregion (Omernik 1987). Significance of variance components was assessed by 10,000 permutations.

Results

SPECIES DELIMITATION

The nonparametric species delimitation approach implemented resulted in seven delimited species for all analyses involving 20 individuals (three-, six-, and nine-gene) (Table 3; Fig. 2); however, species assignment of some samples was problematic and only six delimited species were retained after generating the 50% majority-rule consensus tree. Several populations consistently grouped into the same species across the 20-individual analyses (e.g., populations 12 and 22, 15 and 43, and 48, 57, and 59; Fig. 2), whereas some populations (e.g., populations 2, 5, 21, and 53; Fig. 2) did not. Analyses involving 60 individuals (one sampled from each population) resulted in 15 and 11 delimited species for the three-gene and six-gene datasets, respectively (Table 3; Fig. 3). The primary differences between these two analyses included the splitting

Table 3. **Number of delimited species, number of best trees, and tree score for each *Typhlichthys* delimited species analysis using the nonparametric method of O'Meara (2010). The number of species used for subsequent analyses is indicated in parentheses after generating the 50% majority-rule consensus tree of the best delimited species trees. In the bottom set, the number of species after Bayesian species delimitation is listed for each delimited species analysis (see Figs. 4 and 5). The number of different species delimitation models with posterior probabilities >0.01 and the posterior probability of the model with the highest posterior probability is listed in parentheses.**

	20-individuals			60-individuals			135-individuals		
Loci	No. species	No. trees	Score	No. species	No. trees	Score	No. species	No. trees	Score
Three-gene	7 (7)	13	6.000	16 (16)	45	19.654	21 (19)	40	47.194
Six-gene	7 (6)	25	16.336	11 (11)	14	58.738			
Nine-gene	7 (6)	2	28.285						
Three-gene	4 (5, 0.53)			14 (2, 0.98)			15 (1, 1.00)		
Six-gene	4 (4, 0.46)			10 (2, 0.99)					
Nine-gene	6 (1, 1.00)								

of species B in the six-gene analysis into two species in the three-gene analysis (species B and I), splitting of species F (six-gene) into species F, N, and O (three-gene), and splitting of species A (six-gene) into species A and H (three-gene). The 135-individual, three-gene analysis resulted in the delimitation of an additional five species (21 species in total) although only three were included in subsequent analyses after generating the majority-rule consensus tree. These additional species included species P and Q split from species I in the six-gene analysis, as well as species R and S from species F. Several populations were problematic and did not consistently group with a particular set of populations in the 60- and 135-individual analyses, including populations 4, 6, 7, 10, and 49 (Fig. 3). Most of these "problematic" populations (except 49) are located in the same hydrological basin and ecoregion. The analysis using the six-gene, 60-individual dataset but with the nuclear loci phased resulted in 16 delimited species. Phasing the nuclear had the same effect as increasing sample size (increased number of delimited species), as the number of alleles in each locus was doubled.

SPECIES TREE ESTIMATION

The species trees estimated for each delimited species dataset in *BEAST are presented in Figures 4 and 5. All species trees show strong support for monophyly of *Typhlichthys* and several delimited species-level relationships. There also is strong support for monophyly of delimited species west of the Mississippi River in the Ozark Highlands of Missouri and Arkansas that is comprised of one to four species, depending on the delimited species analysis (species B in the 20-individual, species B and I in the 60-individual, and species B, I, P, and Q in the 135-individual).

BAYESIAN SPECIES DELIMITATION

The Bayesian species delimitation results for *Typhlichthys* for each delimited species dataset are shown in Figures 4 and 5. In general, Bayesian species delimitation supported the guide tree with speciation probabilities >0.95 for most nodes in all analyses. However, in all but the 20-individual, nine-gene analysis, Bayesian species delimitation supported fewer speciation events and delimited species. This was most pronounced in the three-gene datasets where four delimited species pairs were collapsed in the 135-individual data (Fig. 5), one pair in the 60-individual dataset (Fig. 5), and three pairs in the 20-individual dataset resulting in support for four rather than seven delimited species. For the six-gene datasets, 10 species are supported in the 60-individual dataset and four species in the 20-individual dataset out of the 11 and six species delimited using O'Meara's (2010) method, respectively (Table 3). In most cases, delimited species that were collapsed into a single species by Bayesian species delimitation consisted of closely related populations that were considered the same species in other nonparametric species delimitation analyses (e.g., species F and S, A and H, and I and Q in the 135-individual, three-gene dataset; L and D in the 60-individual, three-gene dataset; F and I in the 60-individual, six-gene dataset, and F and G in the 20-individual, three-gene dataset; Figs. 4 and 5). For analyses with 20 individuals, the posterior probability distributions of models support up to five different species delimitation models with posterior probability values >0.01 (Table 3), whereas fewer species delimitation models were supported for analyses with 60 and 135 individuals.

GENEALOGICAL TESTS OF DISTINCTIVENESS

Values of *gsi* and *egsi* indicate a high degree of exclusive ancestry within delimited species for the all delimited species datasets (Tables 4, 5, S1–S4), despite some levels of discordance among delimited species and loci. In the 135-individual, three-gene dataset (Table 4), most delimited species had high *egsi* values above 0.6 with six delimited species monophyletic at all three loci and all measures of exclusive ancestry were significant.

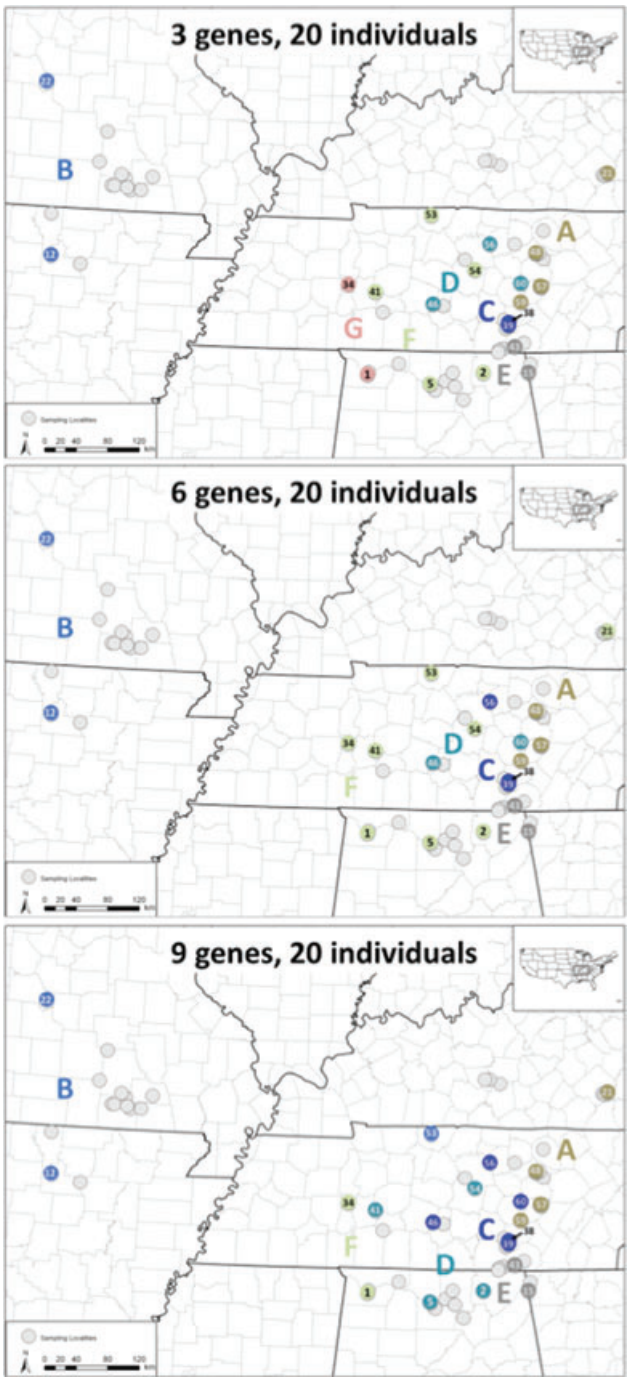

Figure 2. Geographic distribution of delimited species from the 20-individual, three-gene (top), 20-individual, six-gene (middle), and 20-individual, nine-gene (bottom) datasets. Numbered localities and delimited species correspond to populations listed in Table 1.

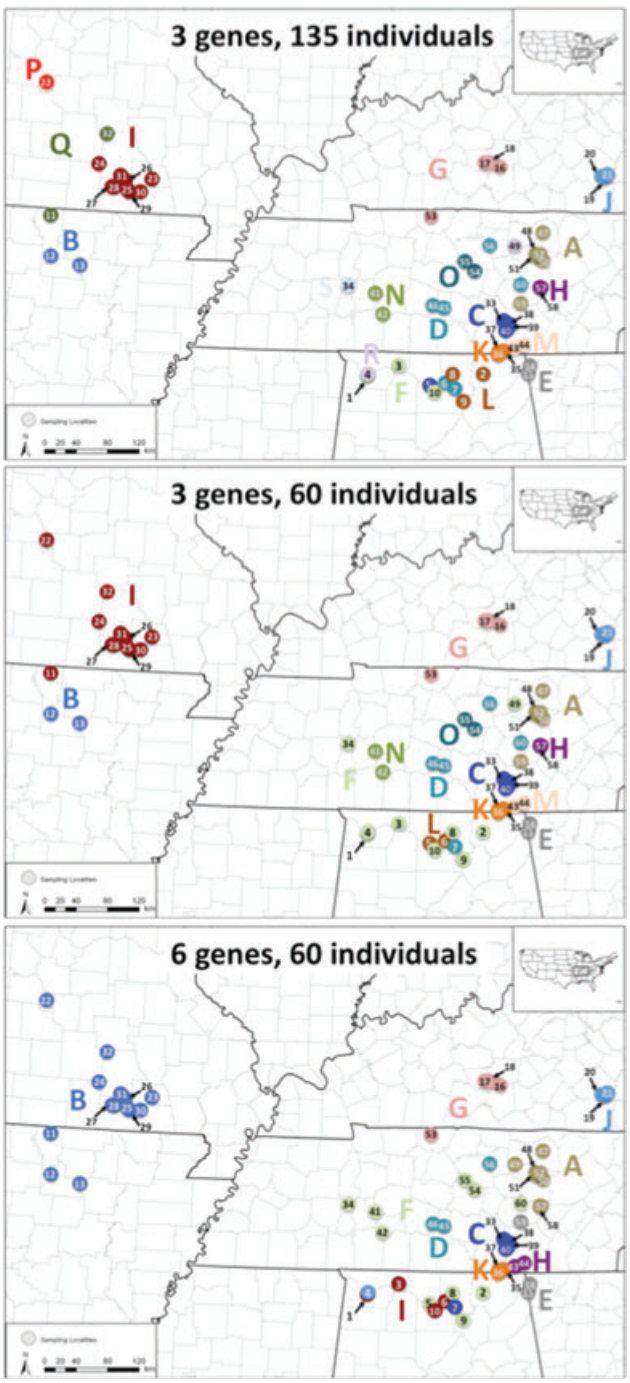

Figure 3. Geographic distribution of delimited species from the 135-individual, three-gene (top), 60-individual, three-gene (middle), and 60-individual, six-gene (bottom) datasets. Numbered localities and delimited species correspond to populations listed in Table 1.

Likewise, egsi values were high for the most delimited species of both the 60-individual, three-gene (Table S4) and 60-individual, six-gene (Table 5) datasets. Almost half (seven out of 15) of the delimited species in the 60-individual, three-gene dataset were monophyletic at all loci; however, no delimited species exhibited exclusive ancestry at all loci in the 60-individual, six-gene dataset and some *gsi* values for delimited species were not significant for the *myh6*, *plagl2*, and *tbr1* loci. *Gsi* values were generally lower for these loci, which are consistent with lower overall genetic variation and shared ancestry across delimited species in

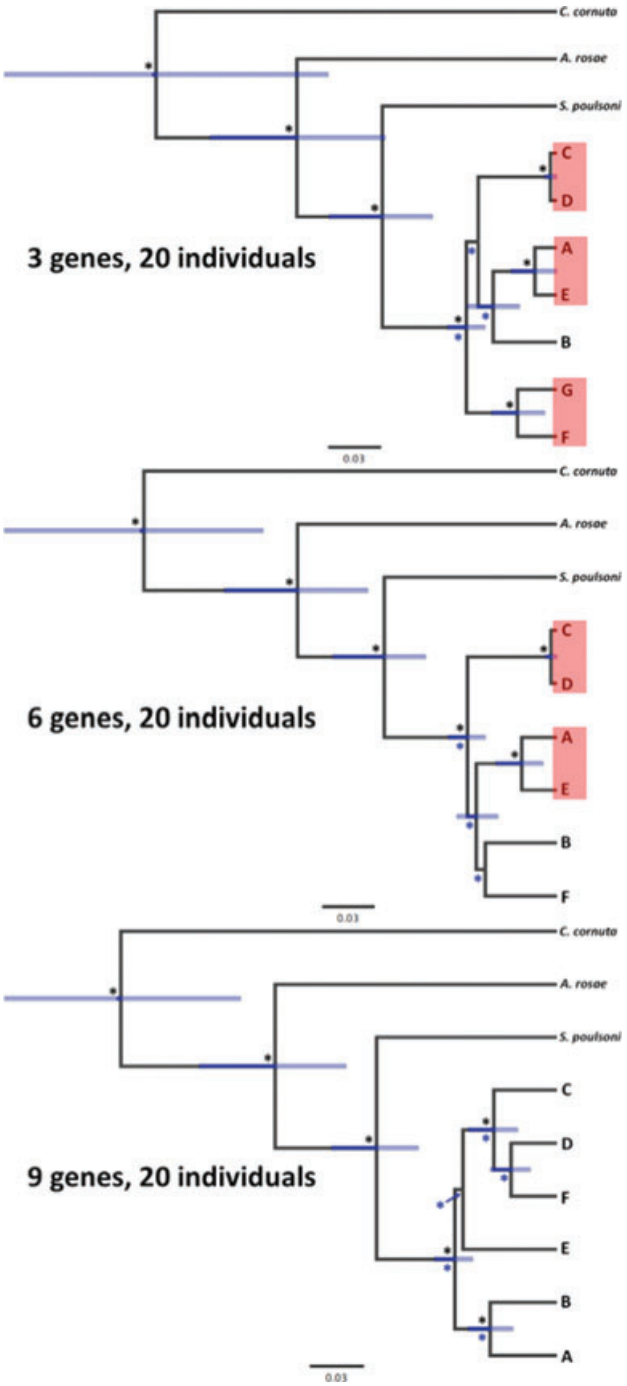

Figure 4. Species tree phylogenies based on the 20-individual datasets and delimited species assignments inferred using *beast: three-gene (top), six-gene (middle), and nine-gene (bottom). Clade posterior probabilities >0.95 are indicated above the branch with an asterisk in black and uncertainty in the relative divergence times are shown by bars on nodes with the length corresponding to the 95% highest posterior density (HPD) of the node ages. Nodes with speciation probabilities >0.95 under Bayesian species delimitation are denoted with an asterisk in blue. Scale bars represent substitutions per site. Delimited species in red boxes were collapsed into a single species under Bayesian species delimitation. Delimited species correspond to those labeled in Figure 2.

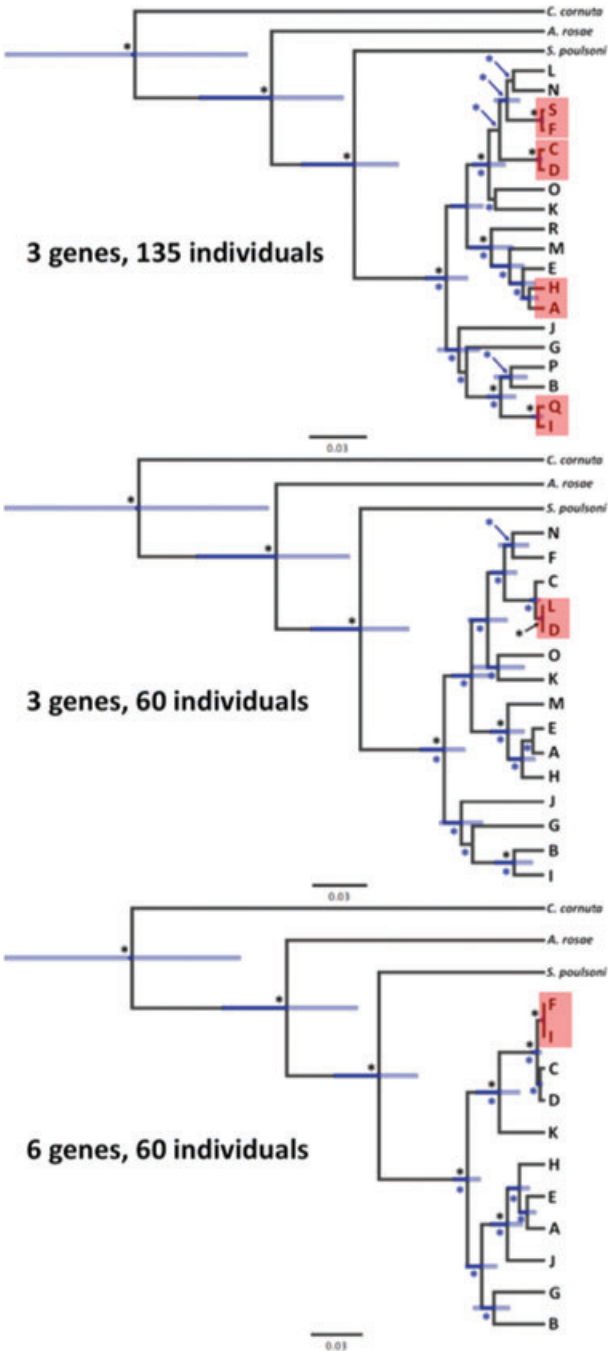

Figure 5. Species tree phylogenies based on the 135- and 60-individual datasets and delimited species assignments inferred using *beast: 135-individual, three-gene (top), 60-individual, three-gene (middle), and 60-individual, six-gene (bottom). Clade posterior probabilities > 0.95 are indicated above the branch with an asterisk in black and uncertainty in the relative divergence times is shown by bars on nodes with the length corresponding to the 95% highest posterior density (HPD) of the node ages. Nodes with speciation probabilities >0.95 under Bayesian species delimitation are denoted with an asterisk in blue. Scale bars represent substitutions per site. Delimited species in red boxes were collapsed into a single species under Bayesian species delimitation. Delimited species correspond to those labeled in Figure 3.

Table 4. Genealogical sorting index (*gsi*) and *P* values of 19 delimited *Typhlichthys* species for gene trees based on the 135-individual, three-gene dataset. *P* values are based on 10,000 permutations and are given in parentheses. Species in bold were monophyletic at all three loci.

Species	*nd2*	*s7*	*rag1*	All combined
S. poulsoni	**1.0000 (0.0045)**	**1.0000 (0.0045)**	**1.0000 (0.0032)**	**0.7500 (0.0004)**
A	1.0000 (<0.0001)	0.3704 (<0.0001)	1.0000 (<0.0001)	0.5926 (0.0001)
B	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**1.0000 (0.0003)**	**0.7500 (0.0001)**
C	0.5015 (<0.0001)	0.5959 (<0.0001)	0.5300 (<0.0001)	0.4068 (0.0001)
D	0.4646 (<0.0001)	0.4647 (<0.0001)	0.5315 (<0.0001)	0.3652 (0.0001)
E	1.0000 (<0.0001)	0.7939 (<0.0001)	1.0000 (<0.0001)	0.6985 (0.0001)
F	0.8270 (<0.0001)	0.5386 (<0.0001)	1.0000 (<0.0001)	0.5914 (0.0001)
G	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**0.7500 (0.0001)**
H	1.0000 (<0.0001)	0.6617 (0.0003)	1.0000 (<0.0001)	0.6654 (0.0001)
I	1.0000 (<0.0001)	1.0000 (<0.0001)	0.8934 (<0.0001)	0.7234 (0.0001)
J	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**0.7500 (0.0001)**
K	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**0.7500 (0.0001)**
L	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**0.7500 (0.0001)**
M	1.0000 0.0048)	0.4963 (0.0140)	1.0000 (0.0039)	0.6241 (0.0016)
N	1.0000 (0.0049)	0.4963 (0.0123)	0.4963 (0.0137)	0.4981 (0.0046)
O	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**1.0000 (<0.0001)**	**0.7500 (0.0001)**
P	1.0000 (<0.0001)	0.7444 (<0.0001)	0.7444 (<0.0001)	0.6222 (0.0001)
Q	0.6617 (<0.0001)	1.0000 (<0.0001)	0.3234 (0.0029)	0.4963 (0.0002)
R	0.7444 (<0.0001)	0.7444 (<0.0001)	1.0000 (<0.0001)	0.6222 (0.0001)
S	1.0000 (<0.0001)	0.4157 (<0.0001)	1.0000 (<0.0001)	0.6039 (0.0001)

these genes. *Egsi* values were significant for almost all delimited species of the 20-individual datasets (three-gene, six-gene, and nine-gene; Tables S1–S3), but a greater proportion of individual *gsi* values were insignificant, largely the result of small sample sizes within delimited species. Although some delimited species exhibited monophyly across all loci in each dataset, overall support for exclusive ancestry was mixed, with some delimited species showing strong signal while others showed low signal.

CONCATENATION

The partitioned Bayesian concatenated analysis of the 60-individual, six-gene dataset was not concordant with the delimited species assignments (Fig. S1), as several delimited species (e.g., D, E, F, and I; Fig. 3, Table 1) were not monophyletic in the concatenated majority-rule consensus tree. Additionally, the topology of the concatenated tree disagrees with the multilocus specie tree topology estimated in *BEAST.

POPULATION STRUCTURE

Hierarchical AMOVA of populations grouped by hydrological subbasin revealed that genetic structure in each locus (*nd2*, *s7*, and *rag1*) of the 135-individual dataset is highly correlated with hydrological boundaries, as the majority of variation was significantly partitioned among watersheds for each locus (66.7–79.5%; Table 6B). Likewise, hierarchical AMOVA of populations grouped by hydrological basin also revealed genetic structure associated with hydrological boundaries, but to a much lesser degree (35.7–37.9%; Table 6A). Genetic structure also is correlated with ecoregion but to a similar level observed for hydrological basins (27.2–30.9%; Table 6C).

Discussion

Our multilocus approach examining varying numbers of loci and individuals using O'Meara's (2010) method revealed the potential for as many as 19 or more species (based on the 135-individual, three-gene dataset) within a wide-ranging and morphologically invariable cavefish species. In addition to the multilocus genetic data, structuring of genetic variation with surface hydrological subbasins provides additional evidence for the recognition of multiple, genetically defined species and offers a hypothesis for a role of hydrological barriers in speciation of *Typhlichthys*. For many groups of organisms, such as subterranean taxa, data from multiple disciplines that could be used to identify species frequently are lacking. For instance, subterranean organisms often are morphologically cryptic and for many species little information is available regarding reproductive isolation among populations, behavior, life history, habitat preferences, and sometimes even exact distribution. A logical first step to delimit species boundaries in such taxa is to use information from multiple genetic loci to formulate hypotheses to be later tested with additional, independent datasets. The use of multiple loci lowers the

Table 5. Genealogical sorting index (*gsi*) and *P* values of 11 delimited *Typhlichthys* species for gene trees based on the 60-individual, six-gene dataset. *P* values are based on 10,000 permutations and are given in parentheses.

Species	*nd2*	*s7*	*rag1*	*myh6*	*plagl2*	*tbr1*	All combined
S. poulsoni	1.0000 (0.0115)	1.0000 (0.0106)	1.0000 (0.0106)	1.0000 (0.0097)	1.0000 (0.0117)	1.0000 (0.0081)	1.0000 (0.0004)
A	0.7490 (<0.0001)	0.7490 (<0.0001)	0.7490 (<0.0001)	0.5293 (<0.0001)	0.5892 (<0.0001)	0.8588 (<0.0001)	0.7040 (0.0001)
B	1.0000 (<0.0001)	1.0000 (<0.0001)	1.0000 (<0.0001)	0.7617 (<0.0001)	0.5159 (<0.0001)	0.4801 (<0.0001)	0.7930 (0.0001)
C	0.2152 (0.0463)	0.7860 (<0.0001)	0.6433 (<0.0001)	0.4649 (0.0002)	1.0000 (<0.0001)	0.1676 (0.1233)	0.5462 (0.0001)
D	1.0000 (0.0005)	0.6554 (0.0014)	0.4831 (0.0037)	0.4831 (0.0038)	1.0000 (0.0003)	0.2246 (0.0462)	0.6410 (0.0004)
E	0.2615 (0.0193)	1.0000 (<0.0001)	0.1384 (0.2464)	0.1729 (0.0906)	0.1959 (0.0617)	0.0601 (0.9079)	0.3048 (0.0038)
F	0.3826 (0.0003)	0.3826 (0.0002)	0.4135 (<0.0001)	0.2668 (0.0062)	0.4134 (<0.0001)	0.3548 (0.0003)	0.3690 (0.0001)
G	1.0000 (<0.0001)	1.0000 (<0.0001)	1.0000 (<0.0001)	0.2638 (0.0136)	1.0000 (<0.0001)	0.4741 (0.0007)	0.7897 (0.0001)
H	1.0000 (0.0126)	1.0000 (0.0098)	1.0000 (0.0116)	0.1867 (0.1110)	0.3222 (0.0476)	0.4917 (0.0292)	0.6668 (0.0058)
I	0.2638 (0.0114)	0.3427 (0.0040)	0.2638 (0.0111)	0.1736 (0.0938)	0.2989 (0.0066)	0.1455 (0.1947)	0.2480 (0.0057)
J	0.2989 (0.0071)	0.5793 (0.0002)	0.3990 (0.0020)	0.3990 (0.0028)	0.3427 (0.0034)	1.0000 (<0.0001)	0.5031 (0.0002)
K	1.0000 (0.0003)	1.0000 (0.0004)	1.0000 (0.0002)	0.2246 (0.0377)	0.6554 (0.0018)	0.6554 (0.0011)	0.7559 (0.0002)

risk of inaccurate species identification relative to that observed in single-locus datasets (Roe et al. 2010), particularly among closely related species, as sole dependence on a single locus can result in over- and underestimating species diversity (Will and Rubinoff 2004; Meyer and Paulay 2005; Roe and Sperling 2007; Roe et al. 2010). Delimiting species using multiple loci alone is a difficult problem, however, particularly for recently diverged taxa. O'Meara's (2010) method delimits the number of species and jointly estimates the species tree from a multilocus dataset that includes multiple individuals sampled from each lineage. Unlike the majority of other methods, this approach does not make any assumptions regarding species assignment a priori, species tree topology, or congruence between gene trees and the species tree.

Our results show that many delimited species of *Typhlichthys* using O'Meara's (2010) method are supported, but the number of individuals and loci sampled influences the number of delimited species. Increasing the number of *Typhlichthys* individuals sampled or phasing the nuclear loci yielded a greater number of delimited species, whereas increasing the number of loci yielded fewer delimited species. Although it would be premature to use the newly developed method described in O'Meara (2010) or other recently developed approaches (e.g., Yang and Rannala 2010) to conduct alpha taxonomy from multilocus genetic data alone, these methods offer a means to develop taxonomic and phylogenetic hypotheses in understudied groups or organisms with little morphological differentiation, such as many subterranean taxa. A potential criticism of interpreting the results of O'Meara's (2010) method alone is that the delimited species merely reflect structuring of genetic variation among populations within a single species. Indeed AMOVA results indicate significant structuring of genetic variation by hydrological drainages. Likewise, increasing the number of individual samples (i.e., from 20 to 60 to 135 in the three-gene dataset) resulted in increasing numbers of delimited species, whereas increasing loci (i.e., from three to six loci in the 60-individual datasets) resulted in fewer delimited species. Phasing the nuclear data also resulted in an increase in the number of delimited species (i.e., from 11 to 16 species in the six-gene, 60-individual dataset). Significant intraspecific structure might have resulted in inflation of the true number of cryptic species because population structure tends to result in more similar gene trees across loci than expected under neutral coalescence (O'Meara 2010). The neutral coalescent assumes panmixia with a constant, large, effective population size over time and no selection (Hudson 1983; Tajima 1983). Unfortunately, these assumptions likely are inappropriate for most organisms, including *Typhlichthys*, as subdivided populations represent most species. It seems most appropriate to use a single sequence from an individual rather than phasing nuclear data using the current methods of species delimitation, as other studies have also done (i.e., O'Meara 2010; Yang and Rannala 2010; Leache and Fujita 2010), to take

Table 6. Hierarchical analysis of molecular variance for each locus in the three-gene (135-individual) dataset grouped according to (A) hydrological basin, (B) hydrological subbasin, and (C) ecoregion (see Table 1).

(A)

Locus	Source of variation	df	Sum of squares	Variance component	% variance	ϕ-statistics
nd2	Among basins	4	2313.828	19.803	37.94	$\phi_{CT} = 0.379^{***}$
	Among populations within basins	55	3800.636	31.513	60.38	$\phi_{SC} = 0.973^{***}$
	Within populations	75	65.571	0.874	1.68	$\phi_{ST} = 0.983^{***}$
	Total	134	6180.034	50.286		
s7	Among basins	4	637.482	2.711	36.77	$\phi_{CT} = 0.368^{***}$
	Among populations within basins	55	1076.152	4.476	60.70	$\phi_{SC} = 0.960^{***}$
	Within populations	210	39.198	0.187	2.53	$\phi_{ST} = 0.975^{***}$
	Total	269	1752.831	7.373		
rag1	Among basins	4	487.905	2.067	35.74	$\phi_{CT} = 0.357^{***}$
	Among populations within basins	55	838.867	3.464	59.88	$\phi_{SC} = 0.932^{***}$
	Within populations	210	53.214	0.253	4.38	$\phi_{ST} = 0.956^{***}$
	Total	269	1379.986	5.784		

(B)

Locus	Source of variation	df	*SS*	*VC*	*V%*	ϕ-statistics
nd2	Among subbasins	21	5352.530	38.520	79.45	$\phi_{CT} = 0.794^{***}$
	Among populations within subbasins	38	761.934	9.091	18.75	$\phi_{SC} = 0.912^{***}$
	Within populations	75	65.571	0.874	1.90	$\phi_{ST} = 0.982^{***}$
	Total	134	6180.034	48.485		
s7	Among subbasins	21	1485.164	5.299	77.17	$\phi_{CT} = 0.772^{***}$
	Among populations within subbasins	38	228.470	1.381	20.11	$\phi_{SC} = 0.881^{***}$
	Within populations	210	39.198	0.187	2.72	$\phi_{ST} = 0.973^{***}$
	Total	269	1752.831	7.373		
rag1	Among subbasins	21	1070.860	3.584	66.71	$\phi_{CT} = 0.667^{***}$
	Among populations within subbasins	38	255.912	1.535	28.58	$\phi_{SC} = 0.858^{***}$
	Within populations	210	53.214	0.253	4.72	$\phi_{ST} = 0.953^{***}$
	Total	269	1379.986	5.373		

(C)

Locus	Source of variation	df	*SS*	*VC*	*V%*	ϕ-statistics
nd2	Among ecoregions	3	1743.117	15.960	30.85	$\phi_{CT} = 0.309^{*}$
	Among populations within ecoregions	56	4371.346	34.895	67.46	$\phi_{SC} = 0.976^{*}$
	Within populations	75	65.571	0.874	1.69	$\phi_{ST} = 0.983^{*}$
	Total	134	6180.034	48.485		
s7	Among ecoregions	3	442.676	1.967	27.17	$\phi_{CT} = 0.272^{*}$
	Among populations within ecoregions	56	1270.957	5.088	70.26	$\phi_{SC} = 0.965^{*}$
	Within populations	210	39.198	0.187	2.58	$\phi_{ST} = 0.974^{*}$
	Total	269	1752.831	7.242		
rag1	Among ecoregions	3	377.813	1.729	30.04	$\phi_{CT} = 0.300^{*}$
	Among populations within ecoregions	56	948.959	3.773	65.55	$\phi_{SC} = 0.937^{*}$
	Within populations	210	53.214	0.253	4.40	$\phi_{ST} = 0.956^{*}$
	Total	269	1379.986	5.531		

Significance is based on 10,000 permutations: <0.001.

a conservative approach and minimize oversplitting of delimited species. However, sequences should first be phased and analyses on phased data conducted to determine if any individuals are heterozygous for alleles from two delimited species. Future models that incorporate more complex speciation scenarios, such as geographic population structure, would be especially valuable to the accurate delimitation of cryptic species.

We further tested the hypothesis that *Typhlichthys* is comprised of multiple cryptic lineages by Bayesian species delimitation (Yang and Rannala 2010) and assessing genealogical patterns of divergence (Cummings et al. 2008). Bayesian species delimitation strongly supported most speciation events in each dataset, although fewer delimited species were supported in all datasets but one (20-individual, nine-gene). The oversplitting of closely related populations into separate species, particularly in the three-gene datasets, is caused by significant population structure that is subsequently obscured by incomplete lineage sorting in other less variable loci included in the six-gene or nine-gene datasets. Because the Bayesian species delimitation approach outlined in Yang and Rannala (2010) is challenging to implement, a user-specified guide tree is recommended to reduce computational space. The guide tree represents the phylogenetic relationships among the most subdivided possible delimitation of individuals into species (i.e., maximum number of delimited species) that are biological plausible based upon other datasets, such as morphology, geography, or geology (Yang and Rannala 2010). However, an accurate guide tree is critical to the outcome of the model (Leache and Fujita 2010; Yang and Rannala 2010), as errors in assignment of individuals to populations or in guide tree topology can lead to inference errors. In many study systems, including *Typhlichthys*, accurately defining a guide tree is challenging, as nonmolecular datasets (e.g., morphology, behavior, hydrology, and geology) often are not particularly useful in generating hypotheses of species boundaries. Although surface hydrological drainages often coincide with species boundaries in many freshwater fish and could be useful in generating a potential guide tree in our system, subterranean drainage patterns do not necessarily correlate with surface drainage patterns and our knowledge of subterranean hydrological connectivity is poor throughout most of the range of *Typhlichthys*. Our approach of first delimiting species using O'Meara's (2010) method then estimating the species tree using the multilocus Bayesian approach implemented in *BEAST to generate a guide tree is advantageous in conducting Bayesian species delimitation in such study systems.

From a phylogenetic perspective, speciation is a transitional process where gene genealogies of diverging lineages change from polyphyletic ancestral gene copies to monophyletic derived alleles (i.e., lineage sorting). This gradual process is influenced by both time and effective population size (Avise and Ball 1990; Maddison 1997; Avise 2004; Weisrock et al. 2010). For recently diverged sets of populations, reciprocal monophyly of all loci usually is not evident but characteristic topological patterns are expected that can be used to identify independently evolving lineages (Knowles and Carstens 2007; Cummings et al. 2008). Despite discordance among loci and the fact that most delimited species do not exhibit monophyly across all loci, most delimited cavefish species had significant patterns of genealogical exclusivity in their mitochondrial and nuclear genes, as measured by *gsi* values for individual genes and *egsi* values, which integrates genealogical patterns across loci (Tables 4, 5, S1–S4), for all datasets examined. The observed low resolution in nuclear loci for some delimited species could be the result of lack of genetic variation, male-biased gene flow, or recent divergence with retention of ancestral polymorphism. Sex-biased dispersal has not been documented in amblyopsid cavefish (Niemiller and Poulson 2010). The extensive distribution of *Typhlichthys* has led some to hypothesize that dispersal through subterranean channels is primarily responsible for the wide range of the species across multiple drainage basins (Holsinger 2005). Genetic divergence is low within hydrological subbasins, even among populations distributed on opposite sides of a river (Niemiller and Fitzpatrick 2008; Table 6B), suggesting these populations have either recently been isolated or that some dispersal occurs between them. However, the majority of genetic variation was partitioned among subbasins for each locus and is indicative of vicariance due to significant dispersal barriers across hydrological boundaries, as has been demonstrated for other aquatic, subterranean taxa (Lefebure et al. 2006, 2007; Finston et al. 2007; Carlini et al. 2009). Therefore, recent divergence and incomplete lineage sorting most likely explain lack of genealogical exclusivity for some delimited species at nuclear loci.

The results of Bayesian species delimitation and genealogical distinctiveness support most *Typhlichthys* species delimited using O'Meara's (2010) nonparametric method, but the question remains as to how many distinct lineages to recognize taxonomically. This question, in part, depends on the species concept used to recognize species. The biological species concept (Mayr 1942) is difficult to use for many species, including most subterranean organisms, as it may be impossible to test for reproductive isolation because individuals are difficult to collect and rear or for conservation reasons. However, the Bayesian species delimitation method adopts the biological species concept, recognizing groups that have not experienced recent gene flow and where discordance among loci is due to lineage sorting only (Yang and Rannala 2010). Few species would be recognized under a phylogenetic species concept that allows only monophyletic groups to be considered species, as few lineages exhibit monophyly across all loci examined depending on the dataset. However, invoking a genealogical species concept (Baum and Shaw 1995) or metapopulation lineage species concept (de Queiroz 1998, 1999, 2007)

would result in recognition of all delimited species of *Typhlichthys*, but the number of species recognized depends of both the number of loci and number of individuals considered. Based on our analyses, we diagnose from 10 to 15 population-level lineages (from the 135- and 60-individual datasets) and strongly suggest that diversity is vastly underestimated in *Typhlichthys*. Uncorrected mtDNA sequence divergence ranged from 3.6% to 12.2% among these lineages. However, we refrain from describing these lineages (with one exception below) until additional work in an integrative framework (Dayrat 2005; Rubinoff et al. 2006; Roe and Sperling 2007; Shaffer and Thompson 2007; Groeneveld et al. 2009; Roe et al. 2010) incorporating information from different fields of study (e.g., morphology, genetics, behavior, and geography) is conducted to assess validity of these putative lineages.

CONSERVATION AND TAXONOMIC IMPLICATIONS

Hidden diversity in groundwater habitats is not limited to our study, as several recent molecular studies have documented hidden diversity in groundwater fauna (Culver et al. 1995; Verovnik et al. 2003; Wiens et al. 2003; Finston et al. 2007; Zakšek et al. 2007; Buhay and Crandall 2009), although almost all of these studies involve invertebrate fauna. Our study and that on the European Cave Salamander (*P. anguinus*; Goricki and Trontelj 2006; Trontelj et al. 2009) have revealed significant, cryptic diversity in obligate, subterranean vertebrates. Almost all subterranean, aquatic macrofauna are endemic to small- to medium-sized (<200 km) groundwater basins, and only a very small fraction of species have large distributions (Trontelj et al. 2009). Rather, these species, including *T. subterraneus*, are actually cryptic species complexes comprised of morphologically similar species, each with considerably smaller ranges. The prevalence of cryptic species in groundwater taxa has implications in the assessment and conservation of groundwater biodiversity (Trontelj et al. 2009), including greater endemism and biodiversity at a regional scale but a decrease in faunal similarity among regions (e.g., groundwater basins).

The discovery of cryptic, distinct lineages, and putative species within the nominal *T. subterraneus* has obvious conservation implications. As currently recognized, *T. subterraneus* is considered secure, although the species is listed as "Vulnerable" by IUCN (IUCN 2007) and afforded protection in several states, including Alabama (listed as "Protected"), Arkansas (listed as an "Inventory Element"), Georgia (listed as "Endangered"), Kentucky (listed as "Special Concern"), and Tennessee (listed as "Deemed in Need of Management"). Because *T. subterraneus* is already a species of conservation concern in many parts of its range, the recognition of multiple, cryptic species likely would result in several species more rare than previously supposed. These species would have a much more restricted distribution comprised of fewer populations and, consequently, fewer individuals. Accordingly, the different species might require different conservation and management strategies.

Typhlichthys subterraneus was described from a well near Bowling Green, Warren County, Kentucky, in the Green River drainage (Girard 1859). Eigenmann (1905) later described both *T. osborni* and *T. wyandotte* based on differences in head width and eye diameter. Although the type locality of *T. subterraneus* is unknown, *T. subterraneus* and *T. osborni* likely are the same species, as *T. osborni* was described from nearby Horse Cave, Kentucky in the same hydrological basin. *Typhlichthys wyandotte* was described from a well near Corydon, far outside the known distribution of *Typhlichthys* but within the range of *A. spelaea* and represents this species. *Typhlichthys eigenmanni* was described as a fourth species in the genus from Camden County, Missouri, but synonomized along with all other species under *T. subterraneus* by Woods and Inger (1957). Recently, Parenti (2006) demonstrated that *T. eigenmanni* Charlton (1933) is an available name. In our analyses, populations west of the Mississippi River in the Ozark Highlands of Arkansas and Missouri showed strong support for phylogenetic distinctiveness with one to four lineages recognized depending on the dataset. These populations occur in a distinct ecoregion and are allopatric from populations east of the Mississippi River. Given biogeographical and phylogenetic evidence, we advocate resurrection of *T. eigenmanni* for Ozark Highland populations of *Typhlichthys*. Further work likely will result in recognition of additional species both east and west of Mississippi River.

Conclusions

Using newly developed methods to delimit species and infer the species tree from multilocus genetic data, we identified several (up to 15) putative cryptic species within the nominal *T. subterraneus*, a discovery also made by several other molecular investigations of subterranean taxa. Our approach presents a way to develop taxonomic and phylogenetic hypotheses in understudied groups or taxa that exhibit little morphological differentiation, such as subterranean organisms. The occurrence of multiple, isolated phylogenetic groups inferred from multiple loci and associated with hydrological basins indicates that *Typhlichthys* possesses reduced dispersal abilities and implicates a strong role for geographic isolation. However, the exact evolutionary history of *Typhlichthys* is difficult to surmise, as the timing of subterranean colonization is difficult to infer from molecular data. The discovery of cryptic species within *T. subterraneus*, a species of conservation concern across its range, has obvious implications for conservation and management agencies.

ACKNOWLEDGMENTS

We are grateful to B. Miller, N. Mann, G. Moni, A. Moni, J. Cooley, V. Walenta, J. Page, B. Gee, B. Walden, L. Simpson, B. Kuhajda,

J. Jensen, D. Fenolio, R. Long, J. Eberly, S. House, J. Buhay, B. Biddix, B. Walter, B. Glorioso, J. Todd, T. Niemiller, A. Romero, G. Graening, J. A. Miller, J. H. Miller, and G. Reynolds for assistance with field sampling or laboratory work. We thank B. O'Meara for assistance with using BROWNIE. We thank the Missouri Department of Conservation, Tennessee Department of Environment and Conservation, Tennessee Wildlife Resources Agency, National Park Service, U.S. Forest Service, Southeastern Cave Conservancy, Inc., Carroll Cave Conservancy, National Speleological Society, and private landowners for permitting access to caves on their properties. We thank the Tennessee Cave Survey, the Tennessee Department of Environment and Conservation, and the Missouri Department of Conservation for providing locality data and the state agencies of Alabama, Georgia, Kentucky, Missouri, and Tennessee, the National Park Service, and the U.S. Forest Service for permits. An earlier draft of this manuscript was reviewed by J. Fordyce, C. D. Hulsey, B. O'Meara, and G. Reynolds. This work was funded by the Tennessee Wildlife Resources Agency (contract nos. ED-06–02149-00 and ED-08023417–00; MLN and BMF), Kentucky Department of Fish and Wildlife Resources (contract no. PON2 660 1000003354; MLN and BMF), National Science Foundation (DDIG award no. DEB-1011216 to MLN; DEB-0716155, DEB-1061806, and ANT-0839007 to TJN), the Department of Ecology and Evolutionary Biology at the University of Tennessee (MLN and BMF), the Department of Biology at Middle Tennessee State University (MLN), National Speleological Society (MLN), and American Society of Ichthyologists and Herpetologists (MLN).

LITERATURE CITED

Avise, J. C. 2004. Molecular markers, natural history, and evolution, 2nd edn. Sinauer Associates Inc., Sunderland, MA.

Avise, J. C., and R. M. Ball. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. Oxford Surv. Evol. Biol. 7:45–67.

Barr, T. C., and J. R. Holsinger. 1985. Speciation in cave faunas. Annu. Rev. Ecol. Syst. 16:313–337.

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pp. 289–303 *in* P. C. Hoch and A. G. Stephenson, eds. Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, St. Louis, MO.

Beheregaray, L. B., and A. Caccone. 2007. Cryptic biodiversity in a changing world. J. Biol. 6:9.

Bergstrom, D. E. 1997. The phylogeny and historical biology of Missouri's *Amblyopsis rosae* (Ozark cavefish) and *Typhlichthys subterraneus* (southern cavefish). Master's thesis, University of Missouri, Columbia, MO.

Bergstrom, D. E., D. B. Noltie, and T. P. Holtsford. 1995. Ozark cavefish genetics: the phylogeny of Missouri's Ozark cavefish (*Amblyopsis rosae*) and southern cavefish (*Typhlichthys subterraneus*). Final Report, Endangered Species Project SE-01–27: improving the status of endangered species in Missouri Ozark cavefish genetics. Missouri Department of Conservation, Jefferson City, MO.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22:148–155.

Bortolus, A. 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. Ambio 37:114–118.

Buhay, J. E., and K. Crandall. 2009. Taxonomic revision of cave crayfish in the genus *Cambarus* subgenus *Aviticambarus* (Decapoda: Cambaridae) with descriptions of two new species, *C. speleocoopi* and *C. laconensis*, endemic to Alabama, USA. J. Crust. Biol. 29:121–134.

Burr, B. M., and M. L. Warren, Jr. 1986. A distributional atlas of Kentucky fishes. Vol. 4. Kentucky State Nature Preserves Commission Scientific and Technical Series, Frankfort, KY.

Carlini, D. B., J. Manning, P. G. Sullivan, and D. W. Fong. 2009. Molecular genetic variation and population structure in morphologically differentiated cave and surface populations of the freshwater amphipod *Gammarus minus*. Mol. Ecol. 18:1932–1945.

Carstens, B. C., and L. L. Knowles. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Syst. Biol. 56:400–411.

Charlton, H. H. 1933. The optic tectum and its related fiber tracts in blind fishes. A. *Troglichthys rosae* and *Typhlichthys eigenmanni*. J. Comp. Neurol. 57:285–325.

Chow, S., and K. Hazama. 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. Mol. Ecol. 7:1255–1256.

Cooper, J. E., and D. P. Beiter. 1972. The southern cavefish, *Typhlichthys subterraneus* (Pisces: Amblyopsidae), in the eastern Mississippian Plateau of Kentucky. Copeia 1972:879–881.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates Inc., Sunderland, MA.

Culver, D. C., T. C. Kane, and D. W. Fong. 1995. Adaptation and natural selection in caves: the evolution of *Gammarus minus*. Harvard Univ. Press, London, U.K.

Culver, D. C., and T. Pipan. 2009. The biology of caves and other subterranean habitats. Oxford Univ. Press, Oxford, U.K.

Cummings, M. P., M. C. Neel, and K. L. Shaw. 2008. A genealogical approach to quantifying lineage divergence. Evolution 62:2411–2422.

Dayrat, B. 2005. Toward integrative taxonomy. Biol. J. Linn. Soc. 85:407–415.

Degnan, J., and N. Rosenberg. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:e68.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. Pp. 57–75 *in* D. J. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford Univ. Press, Oxford, U.K.

———. 1999. The general lineage concept of species and the defining properties of the species category. Pp. 49–89 *in* R. A. Wilson, ed. Species: new interdisciplinary essays. MIT Press, Cambridge, MA.

———. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia Univ. Press, New York.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Edwards, S. V., L. Liu, and D. K. Pearl. 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA 104:5936–5941.

Eigenmann, C. H. 1905. Divergence and convergence in fishes. Biol. Lect. Mar. Biol. Lab. Woods Hole 8:59–66.

Excoffier, L., G. Laval, and S. Schneider. 2005. ARLEQUIN (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. 1:47–50.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491.

Finston, T., M. Johnson, W. Humphreys, S. M. Eberhard, and S. A. Halse. 2007. Cryptic speciation in two widespread subterranean amphipod genera reflects historical drainage patterns in an ancient landscape. Mol. Ecol. 16:355–365.

Gazis, R., S. Rehner, and P. Chaverri. 2011. Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. Mol. Ecol. 20:3001–3013.

Gibert, J., and L. Deharveng. 2002. Subterranean ecosystems: a truncated functional biodiversity. Bioscience 52:473–481.

Girard, C. F. 1859. Ichthyological notices. Proc. Acad. Nat. Sci. Phila. 1859:56–68.

Gomez, A., M. Serra, G. R. Carvalho, and D. H. Lunt. 2002. Speciation in ancient cryptic species complexes: evidence from the molecular phylogeny of *Brachionus plicatilis* (Rotifera). Evolution 56:1431–1444.

Goricki, S., and P. Trontelj. 2006. Structure and evolution of the mitochondrial control region and flanking sequences in the European cave salamander *Proteus anguinus*. Gene 387:31–41.

Groeneveld, L. F., D. W. Weisrock, R. M. Rasoloarison, A. D. Yoder, and P. M. Kappeler. 2009. Species delimitation in lemurs: multiple genetic loci reveal low levels of species diversity in the genus *Cheirogaleus*. BMC Evol. Biol. 9:30.

Hausdorf, B. 2011. Progress toward a general species concept. Evolution 65:923–931.

Heled, J., and A. J. Drummond. 2010. Bayesian inference of specie trees from multilocus data. Mol. Biol. Evol. 27:570–580.

Highton, R., G. C. Maha, and L. P. Maxson. 1989. Biochemical evolution in the slimy salamanders of the *Plethodon glutinosus* complex in the eastern United States. Illinois Biol. Monogr. 57:1–160.

Holcroft, N. I. 2004. A molecular test of alternative hypotheses of tetraodontiform (Acanthomorpha: Tetraodontiformes) sister group relationships using data from the RAG1 gene. Mol. Phylogenet. Evol. 32:749–760.

Hollingsworth, P. R., and T. J. Near. 2009. Temporal patterns of diversification and microendemism in Eastern Highland endemic barcheek darters (Percidae: Etheostomatinae). Evolution 63:228–243.

Holsinger, J. R. 2000. Ecological derivation, colonization, and speciation. Pp. 399–415 *in* H. Wilkens, D. C. Culver, and W. F. Humphreys, eds. Ecosystems of the world, Subterranean ecosystems,Vol. 30. Elsevier, Oxford, U.K.

———. 2005. Vicariance and dispersalist biogeography. Pp. 591–599 *in* D. C. Culver and W. B. White, eds. Encyclopedia of caves. Elsevier, Oxford, U.K.

Hudson, R. R. 1983. Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. 23:183–201.

IUCN. 2007. 2007 IUCN Red List of Threatened Species. Prepared by the Species Survival Commission, IUCN, Gland, Switzerland and Cambridge, U.K.

Isaac, N. J. B., J. Mallet, and G. M. Mace. 2004. Taxonomic inflation: its influence on macroecology and conservation. Trends Ecol. Evol. 9:464–469.

Juan, C., and B. C. Emerson. 2010. Evolution underground: shedding light on the diversification of subterranean insects. J. Biol. 9:17.

Knowles, L. L., and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56:887–895.

Kocher, T. D., J. A. Conroy, K. R. McKaye, J. R. Stauffer, and S. F. Lockwood. 1995. Evolution of the ND2 gene in East African cichlids. Mol. Phylogenet. Evol. 4:420–432.

Kozak, K. H., R. A. Blaine, and A. Larson. 2006. Gene lineages and eastern North American paleodrainage basins: phylogeography and speciation in salamanders of the *Eurycea bislineata* species complex. Mol. Ecol. 15:191–207.

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

Kuhner, M. K. 2009. Coalescent genealogy samplers: windows into population history. Trends Ecol. Evol. 24:86–93.

Leache, A. D., and M. K. Fujita. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proc. R. Soc. Lond. B 277:3071–3077.

Lefebure, T., C. J. Douady, M. Gouy, P. Trontelj, J. Briolay, and J. Gibert. 2006. Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. Mol. Ecol. 15:1797–1806.

Lefebure, T., C. J. Douady, F. Malard, and J. Gibert. 2007. Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod. Mol. Phylogenet. Evol. 42:676–686.

Li, C., G. Orti, G. Zhang, and G. Lu. 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol. Biol. 7:44.

Liu, L., and D. K. Pearl. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56:504–514.

Liu, L., D. K. Pearl, R. T. Brumfield, and S. V. Edwards. 2008. Estimating species trees using multiple-allele DNA sequence data. Evolution 62:2080–2091.

Maddison, W. P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Maddison, W., and D. R. Maddison. 2005. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer Associates Inc., Sunderland, MA.

Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

Meyer, C. P., and G. Paulay. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3:e422.

Molbo, D., C. A. Machado, J. G. Sevenster, L. Keller, and E. A. Herre. 2003. Cryptic species of fig-pollinating wasps: implications for the evolution of the fig-wasp mutalism, sex allocation, and precision of adaptation. Proc. Natl. Acad. Sci. USA 100:5867–5872.

Niemiller, M. L., and B. M. Fitzpatrick. 2008. Phylogenetics of the southern cavefish (*Typhlichthys subterraneus*): implications for conservation and management. Pp. 79–88 *in* Proceedings of the 18th National Cave and Karst Management Symposium. St. Louis, MO.

Niemiller, M. L., B. M. Fitzpatrick, and B. T. Miller. 2008. Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus*) inferred from gene genealogies. Mol. Ecol. 17:2258–2275.

Niemiller, M. L., and T. L. Poulson. 2010. Studies of the Amblyopsidae: past, present, and future. Pp. 169–280 *in* E. Trajano, M. E. Bichuette, and B. G. Kappor, eds. The biology of subterranean fishes. Science Publishers, Enfield, NH.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.

O'Meara, B. C. 2010. New heuristic methods for joint species delimitation and species tree inference. Syst. Biol. 59:59–73.

O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

Omernik, J. M. 1987. Ecoregions of the conterminous United States (map supplement). Ann. Assoc. Am. Geogr. 77:118–125.

Parenti, L. R. 2006. *Typhlichthys eigenmanni* Charlton, 1933, an available name for a blind cavefish (Teleostei: Amblyopsidae), differentiated on the basis of characters of the central nervous system. Zootaxa 1374:55–59.

Pasachnik, S. A., A. C. Echternacht, and B. M. Fitzpatrick. 2010. Gene trees, species and species trees in the *Ctenosaura palearis* clade. Conserv. Genet. 11:1767–1781.

Pfenninger, M., and K. Schwenk. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evol. Biol. 7:121.

Polihronakis, M. 2009. Hierarchical comparative analysis of genetic and genitalic geographic structure: testing patterns of male and female genitalic evolution in the scarab beetle *Phyllophaga hirticula* (Coleoptera: Scarabeaidae). Biol. J. Linn. Soc. 96:135–149.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Proudlove, G. S. 2006. Subterranean fishes of the world. International Society for Subterranean Biology, Moulis, France.

Rambaut, A., and A. J. Drummond. 2007. TRACER v1.4. Available at http://beast.bio.ed.ac.uk/Tracer. Accessed May 17, 2011.

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

Roe, A. D., A. V. Rice, S. E. Bromilow, J. E. K. Cooke, and F. A. H. Sperling. 2010. Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle. Mol. Ecol. Res. 6:946–959.

Roe, A. D., and F. A. H. Sperling. 2007. Population structure and species boundary delimitation of cryptic *Dioryctria* moths: an integrative approach. Mol. Ecol. 16:3617–3633.

Rokas, A., and S. B. Carroll. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol. Biol. Evol. 22:1337–1344.

Rokas, A., B. Williams, N. King, and S. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804.

Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Rubinoff, D., S. Cameron, and K. Will. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J. Hered. 97:581–594.

Ryder, O. A. 1986. Species conservation and systematics: the dilemma of subspecies. Trends Ecol. Evol. 1:9–10.

Shaffer, H. B., and R. C. Thompson. 2007. Delimiting species in recent radiations. Syst. Biol. 56:896–906.

Sites, J. W., and J. C. Marshall. 2004. Operational criteria for delimiting species. Annu. Rev. Ecol. Evol. Syst. 35:199–227.

Stephens, M., and P. Scheet. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. Am. J. Human Genet. 76:449–462.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Human Genet. 68:978–989.

Sterelny, K. 1999. Species as ecological mosaics. Pp. 119–138 *in* R. A. Wilson, ed. Species: new interdisciplinary essays. MIT Press, Cambridge, MA.

Swofford, D. L. 1982. Genetic variability, population differentiation, and biochemical relationships in the family Amblyopsidae. Master's thesis, Eastern Kentucky University, Richmond, KY.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

Trontelj, P., C. J. Douady, C. Fiser, J. Gibert, S. Goricki, T. LeFebure, B. Sket, and V. Zakšek. 2009. A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts? Freshwater Biol. 54:727–744.

Verovnik, R., B. Sket, S. Prevorcnik, and P. Trontelj. 2003. Random amplified polymorphic DNA diversity among surface and subterranean populations of *Asellus aquaticus* (Crustacea: Isopoda). Genetica 119:155–165.

Wakeley, J. 1999. Nonequilibrium migration in human history. Genetics 153:1863–1871.

Wakeley, J., and N. Aliacar. 2001. Gene genealogies in a metapopulation. Genetics 159:893–905.

Waples, R. S. 1991. Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. Mar. Fish. Rev. 53:11–22.

Weisrock, D. W., R. M. Rasoloarison, I. Fiorentino, J. M. Ralison, S. M. Goodman, P. M. Kappeler, and A. D. Yoder. 2010. Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. PLoS One 5:e9883.

Wiens, J. J., P. T. Chippindale, and D. M. Hillis. 2003. When are phylogenetic analyses misled by convergence? A case study in Texas cave salamanders. Syst. Biol. 52:501–514.

Wilkins, J. F. 2004. A separation-of-timescales approach to the coalescent in a continuous population. Genetics 168:2227–2244.

Will, K., and D. Rubinoff. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55.

Wilson, E. O. 2003. The encyclopedia of life. Trends Ecol. Evol. 18:77–80.

Woods, L. P., and R. F. Inger. 1957. The cave, spring, and swamp fishes of the family Amblyopsidae of central and eastern United States. Am. Midl. Nat. 58:232–256.

Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. USA 107:9264–9269.

Zakšek, V., B. Sket, and P. Trontelj. 2007. Phylogeny of the cave shrimp *Troglocaris*: evidence of a young connection between Balkans and Caucasus. Mol. Phylogenet. Evol. 42:223–235.

Associate Editor: L. Kubatko

Supporting Information

The following supporting information is available for this article:

Figure S1. Consensus tree from the partitioned Bayesian concatenated analysis based on the 60-individual, six-gene dataset.

Table S1. Genealogical sorting index (*gsi*) and *P* values of seven delimited *Typhlichthys* species for gene trees based on the 20-individual, three-gene dataset.

Table S2. Genealogical sorting index (*gsi*) and *P* values of six delimited *Typhlichthys* species for gene trees based on the 20-individual, six-gene dataset.

Table S3. Genealogical sorting index (*gsi*) and *P* values of six delimited *Typhlichthys* species for gene trees based on the 20-individual, nine-gene dataset.

Table S4. Genealogical sorting index (*gsi*) and *P* values of 15 delimited *Typhlichthys* species for gene trees based on the 60-individual, three-gene dataset.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.