Conservation Genetics
https://doi.org/10.1007/s10592-024-01640-8

RESEARCH

Population genomics and mitochondrial DNA reveal cryptic diversity in North American Spring Cavefishes (Amblyopsidae, *Forbesichthys*)

Roberto V. Cucalón[1] · Joel B. Corush[1] · Matthew L. Niemiller[2] · Amanda N. Curtis[2] · Pamela B. Hart[3] · Bernard R. Kuhajda[4] · Matthew R. Thomas[5] · Brian Metzke[6] · Mark A. Davis[1] · Milton Tan[1]

Received: 13 June 2024 / Accepted: 27 August 2024

Abstract

The North American freshwater genus *Forbesichthys* is composed of facultative cave-dwelling fishes restricted to springs and caves in southern Illinois, southeastern Missouri, southwestern Kentucky, and central Tennessee. These fishes were previously considered a single species, the Spring Cavefish (*F. agassizii*), but recent molecular evidence led to the recognition of the Shawnee Hills Cavefish (*F. papilliferus*). The Shawnee Hills Cavefish is hypothesized to be restricted to Illinois, Missouri, Kentucky and north-central Tennessee, whereas the Spring Cavefish is restricted to the Eastern Highland Rim of central Tennessee. However, the distributions of *Forbesichthys* are difficult to ascertain due to their intermittent appearance in surface springs, making sampling challenging. We assessed the species status, distribution, connectivity, and population sizes of the *Forbesichthys* spp. using Restriction-site Associated DNA sequencing (RADseq) and the mitochondrial NADH dehydrogenase 2 locus. Our results corroborate the recognition and hypothesized distributions of the Shawnee Hills Cavefish and Spring Cavefish. Furthermore, we suggest the recognition of three Evolutionary Significant Units (ESUs) and two Management Units (MUs) within the Shawnee Hills Cavefish. Although all populations analyzed appear to have reasonable genetic diversity and population stability over time, this regionalization has implications for both groundwater policy and management. Our study provides important information relevant to understanding potential population distributions and the identification of unique lineages that may deserve additional protection.

Keywords Fish · RADseq · BestRAD · Population genetics · Evolutionary significant unit · Management Unit

Introduction

Rare, Threatened, or Endangered (RTE) species are often characterized by shortfalls that can hamper conservation and management efforts (Hortal et al. 2015). Knowledge gaps about a species' complete distribution (Wallacean shortfall), abundance and population dynamics (Prestonian shortfall), or species interactions (Eltonian shortfall) can result in poorly informed conservation prioritization and misallocation of resources (Lomolino and Heaney 2004; Cardoso et al. 2011; Hortal et al. 2015). Furthermore, understanding whether species have restricted distributions and limited habitat breadth is crucial to assess their vulnerability to extinction (Davis et al. 2015; Chichorro et al. 2019). As the understanding of species boundaries develops, it is possible for single widespread species assumed to be of little conservation concern to comprise multiple lineages with narrow ranges and requiring conservation actions (Niemiller et al. 2013b; Gales et al. 2023). Simply put, managing a species

✉ Roberto V. Cucalón
rvcucalont@gmail.com

1 Illinois Natural History Survey, Prairie Research Institute, University of Illinois Champaign-Urbana, Champaign, IL, USA

2 Department of Biological Sciences, The University of Alabama in Huntsville, Huntsville, AL, USA

3 Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL, USA

4 Tennessee Aquarium Conservation Institute, Chattanooga, TN, USA

5 Kentucky Department of Fish and Wildlife Resources, Frankfort, KY, USA

6 Illinois Department of Natural Resources, Springfield, IL, USA

Published online: 11 September 2024

without basic knowledge of its range or population dynamics can result in allocation of resources to areas that do not maximize the effectiveness of said efforts (Arponen 2012). The modern genomics era has only expanded opportunities to gain critical insights and a finer-grained perspective to address some of these shortfalls, particularly with respect to RTE species (Avise 2010; Ouborg 2010; Ouborg et al. 2010; Supple and Shapiro 2018).

Incorporating genomics into conservation efforts has facilitated a more fulsome assessment of RTE species by increasing species and population-level resolution (for example, see Patton et al. 2019; Niemiller et al. 2013c, 2022a, b). Moreover, these same genomic data can be harnessed to delineate various levels of conservation units such as cryptic species, Evolutionary Significant Units (ESU), or Managements Units (MU), ultimately guiding better-informed management interventions (Funk et al. 2012; Stanton et al. 2019). This is particularly applicable in species occupying narrow environmental niches with disjunct ranges, as the assessment of the population dynamics of these species particularly benefits from genetic analyses. For example, species restricted to springs and spring-fed creeks such as the Barrens Topminnow (*Fundulus julisia*) have been found to have distinct ESUs and MUs (Hurt et al. 2017), cryptic reef-associated fish have been identified in association with discontinuity of reefs (Priest et al. 2016), and gobies that are restricted to mudflats have been found to exhibit high levels of genetic differentiation associated with distance between coastlines (Corush et al. 2022). However, despite the opportunities presented by genomics, it also poses challenges, primarily associated with analytical limitations, as well as sampling constraints due to species endangerment status or accessibility issues (Steiner et al. 2013; Shafer et al. 2015).

Forbesichthys spp. inhabit springs, spring runs, and associated surface waters but also occur in subterranean waters of southern Illinois, southeastern Missouri, western Kentucky, and central Tennessee (Weise 1957; Woods and Inger 1957; Smith and Welch 1978; Etnier and Starnes 1993; Niemiller and Poulson 2010). The genus *Forbesichthys* (former *Chologaster* and later *Forbesella*) has traditionally recognized a single species, Spring Cavefish (*Forbesichthys agassizii*); however, the Shawnee Hills Cavefish (*F. papilliferus*) (Fig. 1) also known as Northern Spring Cavefish or Karst Cavefish, was recently resurrected based on a limited mitochondrial and nuclear loci dataset from nine samples from nine localities (Niemiller et al. 2013a, c). This was subsequently supported via genomics based on six samples from four localities (Hart et al. 2020). Based on these studies, the Shawnee Hills Cavefish was hypothesized to be restricted to Illinois, Missouri, Kentucky, and Tennessee, although no samples of the Shawnee Hills Cavefish from Tennessee nor Missouri were sequenced. The species has been found east of the Mississippi River in southern Illinois, west of the Mississippi River in southeastern Missouri, in north-central Tennessee, and in the Western Pennyroyal

Fig. 1 Picture of the Spring Cavefish (*Forbesichthys agassizii*) taken from Clayborne Spring in Coffee County, Tennessee and the Shawnee Hills Cavefish (*Forbesichthys papilliferus*) taken from Snake Road in Union County, Illinois. Copyright: Matthew Niemiller

Karst in Kentucky. All of these populations are located in regions influenced by different geological formations. For instance, *F. papilliferus* in Illinois is found in the LaRue-Pine Hills with Devonian-age geology (Kolata and Nimz 2010) while the southeastern Missouri population is found in Ordovician-age geology (Thompson 1991), and the Kentucky and northern Tennessee populations in Mississippian-age geology (McDowell 1986). Similarly, the Spring Cavefish appears to be predominantly found in the Barrens Plateau section of the Eastern Highland Rim of central Tennessee (Starnes and Etnier 1986; Hart et al. 2020) which corresponds to Mississippian-age geology (Peterson 1962). Moreover, NatureServe (2024) includes populations from the Green River watershed in Kentucky southward into Tennessee as *F. agassizii* based on Adams et al. (2020). Yet, conclusions from these studies were based on samples from a limited number of locales and did not include the entirety of the species' distribution.

Troglobitic species in the family Amblyopsidae include many RTE species. These include the Hoosier Cavefish (*Amblyopsis hoosieri*), Northern Cavefish (*A. spelaea*), Southern Cavefish (*Typhlichthys subterraneus*), and the Ozark Cavefish (*Troglichthys rosae*), which are classified as 'Critically Imperiled' (S1), 'Vulnerable' (S3), 'Imperiled' (S2), and 'Vulnerable' (G3), respectively (Chakrabarty et al. 2014; Office of Kentucky Nature Preserves 2019; NatureServe 2024), and the Alabama Cavefish (*Speoplatyrhinus poulsoni*) listed federally as endangered (Federal Register 1988). Prior conservation assessments under IUCN Red List criteria, which considered all *Forbesichthys* populations as a single species (i.e., *F. agassizii*), evaluated the taxon as having a low extinction risk (i.e., Least Concern; NatureServe (2013). Moreover, earlier surveys to understand the distribution of the cavefish in Illinois have revealed what appear to be largely stable populations in southern Illinois (Metzke et al. 2016). The revised assessments of the status and distribution of *Forbesichthys* spp. may alter prioritization of resources and approaches to conserving species within the genus. More recently, given its restricted distribution within Illinois, the Spring Cavefish was listed as Threatened on the Illinois List of Endangered and Threatened Species in 2020 (Illinois Endangered Species Protection Board 2020). Also, following the recognition of the Shawnee Hills Cavefish, Missouri listed this species as endangered (Missouri Department of Conservation 2024). Although the most recent NatureServe conservation assessment recognizes both species, it does not assign a subnational status rank for Kentucky for either species while the Spring Cavefish is ranked as 'Apparently Secure' (S4) in Tennessee (NatureServe 2024). Given the restricted distributions of the Spring Cavefish and Shawnee Hills Cavefish, there is a pressing need to resolve their species boundaries and distributions to better understand their global and subnational conservation status.

Throughout much of their distribution, *Forbesichthys* spp. are only seasonally observed on the surface (Weise 1957; Smith and Welch 1978; Niemiller and Poulson 2010). Despite indications of stable populations, assessing their abundance is challenging due to sampling difficulties and stochastic hydrological shifts in their habitats (Metzke and Holtrop 2014; Metzke et al. 2016). Nonetheless, *Forbesichthys* spp. appear to be quite locally abundant where they occur. For instance, previous surveys in southern Illinois (Metzke et al. 2016) and our findings at Rich Pond in Warren County, Kentucky ($n = 30$), demonstrate this trend. Fishes have the ability to disperse through subterranean passages including across watershed boundaries (Ray et al. 2014), thus *Forbesichthys*, which regularly use groundwater habitats, may potentially be less restricted by surface watershed boundaries. By contrast, mark-recapture studies of *Forbesichthys* have thus far provided little evidence of dispersal among springs (Adams et al. 2000). In line with the Prestonian and Wallacean shortfalls, we know relatively little about groundwater connections and potential barriers to dispersal as well as the full potential range of this species. The Shawnee Hills Cavefish currently has a disjunct distribution with a population west of the Mississippi River in southeastern Missouri (McDonald and Pflieger 1979) and the southern Illinois range, but many questions remain, such as whether *Forbesichthys* found in Missouri, Illinois, and Kentucky represent discrete populations, and whether the Illinois range represents multiple populations. Given the above, conservation of the understudied *Forbesichthys* spp. can be informed by genetic and genomic data to fill knowledge gaps. Here, we aim to address the following questions: (1) Are the distributions of *Forbesichthys* spp. based on genomics congruent with previous reports based on ND2 mitochondrial DNA? (2) Do *Forbesichthys* spp. show evidence of population structure? and (3) Is there evidence of stable or decreasing population sizes in *Forbesichthys* spp. that would inform their potential extinction risk?

Materials & methods

Data collection

Forbesichthys spp. were collected primarily from springs and associated spring runs located in the states of Illinois, Missouri, Kentucky and Tennessee (Table S1). Locales with extant occurrence records were visited in spring when individuals were more likely to be present in surface waters. Fishes were collected with dip nets and sampling continued until targeted habitats were exhausted or ten individuals

were collected. We performed non-lethal tissue sampling of fin clips and preserved tissues in 95% EtOH. A total of 228 tissue samples from *Forbesichthys* spp. were collected from 25 localities.

DNA extraction

We extracted whole genomic DNA using DNeasy® Blood and Tissue kits (Qiagen©) with the following modification to the protocol: after the addition of 200 µl of AL buffer, samples were incubated at 70 ºC for 10 min and then 200 µl of EtOH were added. Samples were stored at 4 ºC overnight before processing via spin column filtration as directed in the manufacturer's protocol. Finally, the elution buffer was warmed to 60 ºC prior to the final elution step. After extraction, samples were quantified using a Qubit™ 3 Fluorometer (ThermoFisher).

Mitochondrial DNA sequencing

From the extracted whole genomic DNA, we amplified a 1,044 bp fragment of the NADH dehydrogenase 2 (ND2) mitochondrial locus using primers and protocols outlined previously (Kocher et al. 1995) and as used by Niemiller et al. (2013a) for amblyopsid cavefishes, including samples of both *Forbesichthys* species. Amplified DNA was sequenced using Applied Biosciences (ABI) 3730xl DNA Analyzers at Roy J. Carver Biotechnology Center at the University of Illinois at Urbana Champaign (UIUC) and Eurofins, Inc. (Louisville, Kentucky), chromatograms were checked visually, and contigs assembled using Geneious Prime version 11. All sequences were trimmed to the same length with no missing data. Existing ND2 sequences representing *Forbesichthys* spp. including outgroup sequences for *Amblyopsis spelaea* were downloaded from NCBI GenBank (Table S1).

Sequences were aligned using MAFFT v7.490 (Katoh and Standley 2013). Aligned sequences were analyzed using maximum-likelihood phylogenetic analysis in IQ-TREE 2 using ModelFinder to find the best substitution model and assessing support using 1,000 replicates for ultrafast bootstrapping (Kalyaanamoorthy et al. 2017; Hoang et al. 2018; Minh et al. 2020). Haplotype network was constructed using the median-joining method in the program PopART (Leigh and Bryant 2015).

RAD sequencing

RAD libraries were prepared following the BestRAD protocol (Ali et al. 2016) using restriction enzyme SbfI and NEBNext® Ultra™ DNA Library Prep Kit for Illumina® as detailed in Ackiss et al. (2020) and as follows: Genomic DNA concentrations were normalized to 20 ng/µl for library preparation. Samples were ordered based on concentration and plated with 76 samples per plate. For initial digestion, 10 µl of normalized DNA was added to 0.68 µl water, 1.2 µl 10x CutSmart® Buffer (NEB), and 0.12 µl of SbfI-HF (NEB) per sample. Samples were placed in the thermocycler at 37 °C for 60 min followed by 80 °C for 20 min. After digestion, an additional 1.44 µl of water, 0.4 µl of T4 DNA Ligase Buffer with rATP (NEB), 0.16 µl T4 DNA Ligase (NEB), and 2 µl of well-specific adapter was added to each well. To ligate adapters, samples were then placed in the thermocycler at 25 °C for 90 min followed by 65 °C for 20 min. After adapters were added to digested DNA, libraries were created for each plate by pooling 5 µl of each sample together, creating three libraries each containing labeled DNA from 76 samples. The 3 samples were then sonicated for a total of 3 min in increments of 30 s with 59 s between cycles. Sonication occurred at 25% amplitude using the Q500® sonicator (Qsonica). Sonicated DNA was visualized on 2% E-Gel™ EX Agarose Gels (Invitrogen™) to verify the bulk of the DNA was in the desired size range (200–500 bp). Samples were then purified using a double-sided (0.65X and 1.0X) purification with AMPure XP beads. To remove non-ligated DNA that fell within the target size range, an additional purification was done using Dynabeads™ M-280 Streptavidin (Invitrogen). To remove the SbfI enzyme liberated in the previous step, an additional 1.5X AMPure XP bead clean-up was completed. Following the "Protocol for use with NEBNext Ultra DNA Library Prep Kit for Illumina (E7370)" in conjunction with the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® and NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1), final library preparation was conducted with the following modifications. NEBnext Adaptor for Illumina® was used in a 1:10 dilution. Cleanup of Adapter-ligated DNA was done with size selection targeting 250 bp fragments. PCR enrichment of adapter-ligated DNA was modified to: 21 µl DNA, 2 µl of i5 primer and 2 µl of i7 primer. The final PCR was run for 12 cycles. Final libraries were sent to the Roy J. Carver Biotechnology Center at UIUC for pooling and sequencing on the Illumina® NovaSeq™ 6000 using the Sp or S4 flow cells.

SNP calling and filtering

Sequence read data were demultiplexed using process_radtags from STACKS 2.64 (Rochette et al. 2019) using the flag --bestrad specific for the data prepared with BestRAD libraries and specifying the enzyme SbfI. Each read was then aligned to a *Forbesichthys agassizii* draft genome assembly (GenBank accession GCA_026546735.1). We aligned reads using bwa 0.7.17 (Li and Durbin 2009) with default parameters followed by compression into sorted bam files

using samtools (Danecek et al. 2021). The bam files were used as input by the STACKS's module gstacks to generate a catalog of loci. Loci were then filtered with the following conditions. A locus was kept if it was present at least in 65% of the individuals and had a minimum allele count of 3, using the STACKS's module populations. Additional filtering steps were carried out with vcftools (Danecek et al. 2011) including, the removal of individuals with more that 30% missing data, keeping genotypes with more than 2 read depth, mean depth across sites of 5, less than 91 read depth, and genotype quality equal or greater than 20. To account for linkage disequilibrium among loci, we thinned the SNPs to one every 500 bp across the genome using the flag *--thin* from vcftools. Filtered SNPs were output in a variant call format (VCF) for downstream analyses (See Fig. S1 for a general overview of analyses).

Population structure

To study population structure, we ran multiple analyses. For fastSTRUCTURE (Raj et al. 2014) which assesses the number of *K* populations and admixture between them, we tested *K* values ranging from 1 to 25 (the maximum number of sampling localities) using the default convergence criterion (10^{-6}), simple priors, and a random starting seed. Three independent runs were performed across *K* values 1 to 25. The optimal *K* number of clusters in each run was assessed with the chooseK.py script. We performed Discriminant Analysis of Principal Components (DAPC) in adegenet (Jombart 2008; Jombart et al. 2010). We employed a *k*-means clustering algorithm to identify an optimal number of clusters from 1 to 25 with 1,000 randomly starting centroids in each *k*-means iteration, and compare the clustering solutions using Bayesian Information Criterion (BIC). To avoid overfitting of discriminant functions, we used α-score optimization to evaluate the optimal number of principal components (PCs) to retain in the DAPC. With population assignments, we estimated measures of genetic differentiation by F_{ST} (an indicator of reduced gene flow) using the R package hierfstat (Goudet 2005). Significance was evaluated by generating p-values using a permutation test with 1000 replicates in R. In addition, we estimated 95% confidence intervales from 100 bootstrap replicates over loci using the program boot.ppfst, from hierfstat. Confidence intervales that do not include zero were considered as significant values. Diversity summary statistics was carried out based on the ND2 sequences including nucleotide diversity (Paradis 2010), haplotype richness, effective number of haplotypes [equation: 1 / (Sum pi^2), (Brown and Weir 1983)], private haplotypes, unbiased haplotype diversity [equation: (N / (N-1)) * 1 - Sum pi^2, (Anon 1996)], where N is the number of sequences and pi is the frequency of the i^{th} allele for the population. For the RADseq data we calculated observed heterozygosity, expected heterozygosity, and inbreeding coefficient using the R packages, hierfstat (Goudet 2005) and dartR (Gruber et al. 2018; Mijangos et al. 2022).

Phylogenetic reconstruction was carried out in IQTREE2 version 2.2.0.6 (Minh et al. 2020). The input file alignment was generated by converting the VCF output to PHYLIP format using the script vcf2phylip.py (Ortiz 2019). By default, vcf2phylip.py writes heterozygotes with ambiguous base coding. In IQ-TREE, ambiguous constant sites (e.g. C vs. Y which indicates C or T) are considered invariant. Therefore, invariant sites from the resulting alignment were automatically removed by IQTREE2 to avoid violation of the use of the ascertainment bias correction flag (+ASC). To select the best model, we use ModelFinder with correction for ascertainment bias using the option *-m* MFP+ASC in IQTREE2, and branch support was calculated using ultrafast bootstrap (UFBS) for 1000 replicates (Kalyaanamoorthy et al. 2017; Hoang et al. 2018; Minh et al. 2020). UFBS was interpreted as strong support if greater than 95% (Hoang et al. 2018).

Effective population size

To better understand the population dynamics of *Forbesichthys* spp. and potential concerns for genetic bottlenecks, we estimated changes over time in effective population size (N_e), the size of an idealized population (as described in Liu & Fu [2015]). Note that effective population size does not indicate the true number of individuals (i.e., census population size), and estimates of effective population may be smaller or larger than this number; rather, it gives an indication of population genetic diversity, such as the capacity for the population to avoid inbreeding. We estimated effective population size using Stairway Plot 2 (Liu and Fu 2020) based on the site frequency spectrum (SFS) calculated using the script easySFS.py (Gutenkunst et al. 2009) from the VCF generated from STACKS 2 including variant and invariant sites. The easySFS script implements a down projection method proposed by Marth et al. (2004) to account for the presence of missing data, a feature common in RADseq datasets. This method consists of "projecting down" to a smaller number of samples and taking an "average over" all possible resamplings to create a complete data matrix. The projection values (i.e., reduced sample size to be used) were chosen by maximizing the number of segregating sites within each population as recommended in Gutenkunst et al. (2009). To convert estimates from coalescent units to population sizes and absolute time, we assumed a generation time of 1 based on the FishTraits database (Frimpong and Angermeier 2009; Xie et al. 2013) and a mean mutation rate of 5.97×10^{-9} mutations per generation across species of fish (Bergeron et al. 2023). Because this mean

rate is faster than observed in multiple independent studies in diverse fishes including in Atlantic Herrings (*Clupea harengus*) (2.0×10^{-9}; (Feng et al. 2017), Guppies (*Poecilia reticulata*) (2.89×10^{-9}; (Burda and Konczal 2023), Lake Malawi Cichlids (Cichlidae) (3.5×10^{-9}; (Malinsky et al. 2018), and Nine-spined Sticklebacks (*Pungitius pungitius*) ($4.29–4.83\times10^{-9}$; (Zhang et al. 2023), we also performed the analysis with half the mutation rate of 2.99×10^{-9} mutations per generation to span this range.

Results

Population structure and phylogenetic analysis

Mitochondrial ND2 dataset

We successfully amplified ND2 sequences from 222 individual cavefishes resulting in an alignment of 918 base pairs with 112 parsimony informative sites (Table S1). Our analysis recovered two main clades within *Forbesichthys*: one corresponding to the Shawnee Hills Cavefish and the other corresponding to the Spring Cavefish, with a mean 5% (SD = 0.09) uncorrected pairwise divergence between them (Fig. 2). We identified three haplotypes ($n=62$) within the Spring Cavefish restricted to Tennessee, two of which are shared across watershed boundaries (Haps1–3; Figs. 2, 3 and 4; Table S1–S2). Within the Shawnee Hills Cavefish, we identified 17 haplotypes (Table S2). One clade was formed by four haplotypes unique to Illinois and a single unique haplotype within Missouri (Hap8, $n=13$). Two of the haplotypes are from GenBank sequences from Illinois Cave Spring Cave (JX459215 and JX459216, INHS 37654). These Illinois haplotypes included one dominant haplotype (Hap4; $n=12$), and three unique haplotypes each including a single sample (Haps5–7; Figs. 2, 3 and 4, Table S2). The remaining 12 haplotypes from Kentucky and Tennessee differed from the Illinois-Missouri clade by two substitutions (Hap17, 2 samples) or a clade that differed by at least four substitutions including 11 haplotypes in Kentucky and one unique haplotype (Hap20) found in a single Tennessee sample from Clarksville Lake Cave in Montgomery County. Only Hap11 ($n=43$) was shared between the Green River and Cumberland River, while Hap17 ($n=2$), Hap18 ($n=7$), and Hap20 ($n=1$) were unique to the Cumberland River, while Hap9 ($n=14$), Hap10 ($n=46$), Hap12 ($n=2$), Hap13 ($n=2$), Hap14 ($n=2$), Hap15 ($n=3$), Hap16 ($n=2$), and Hap19 ($n=4$) were unique to the Green River. The Shawnee Hills Cavefish had greater mitochondrial genetic diversity in Kentucky than in the other states, and these haplotypes are distinct from those found in Illinois and Missouri. Nearly all sites had only one or two haplotypes, with the exception of Spring Creek which had five (Figs. 3 and 4).

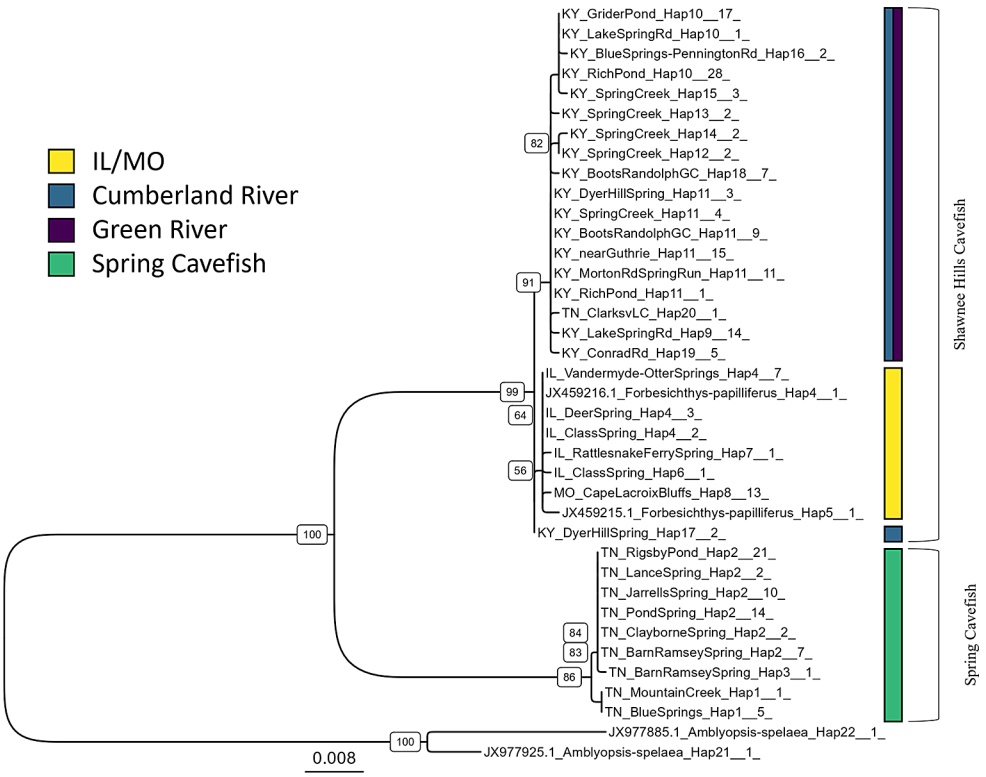

Fig. 2 Maximum likelihood phylogenetic analysis of the ND2 dataset inferred using IQ-TREE version 2.2.0.6. Ultrafast bootstrap support values over 50% are shown. Duplicated sequences per location were collapsed. Tip labels information contain: State_Location_Haplotype_Number of individuals

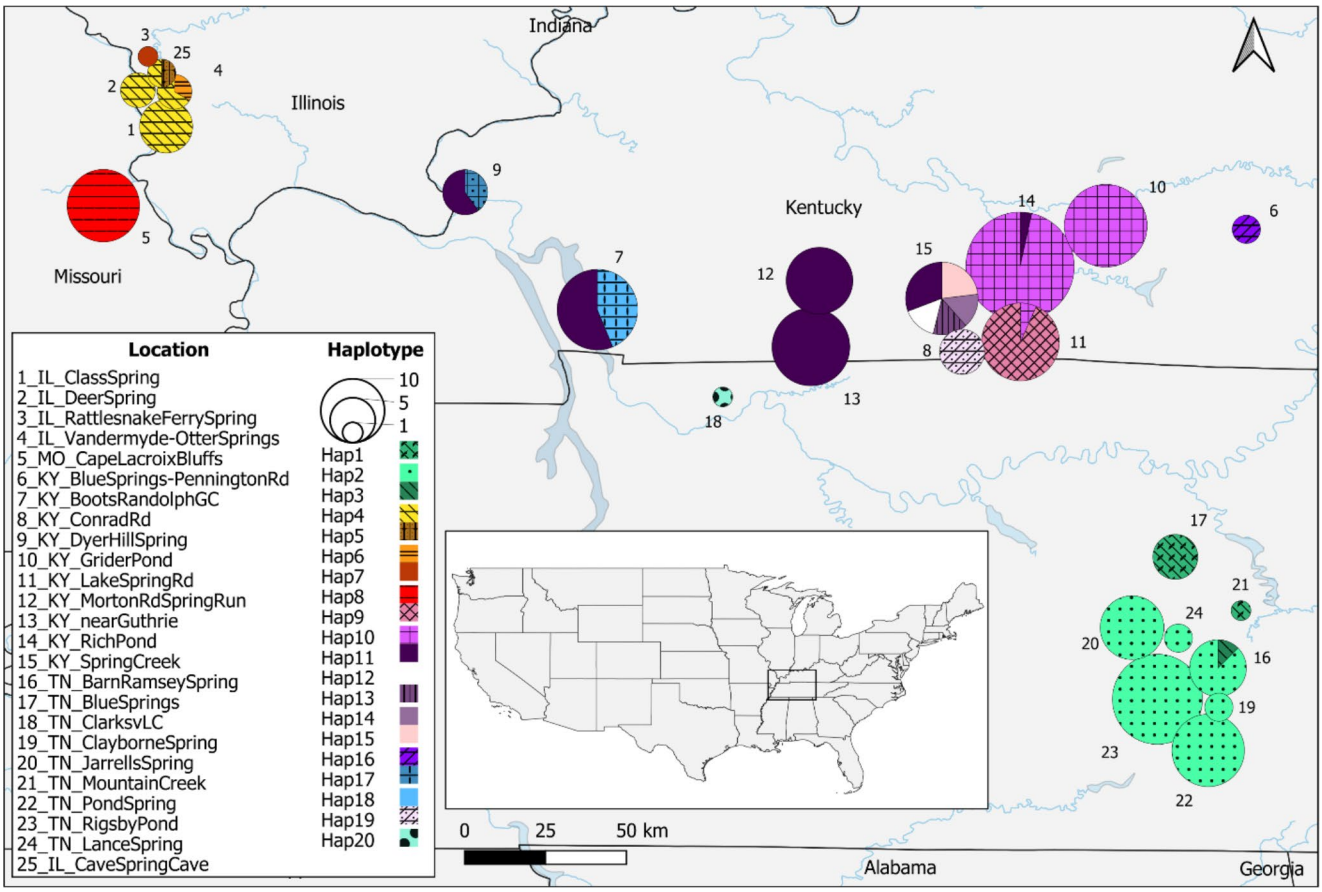

Fig. 3 Sites sampled in Illinois, Missouri, Kentucky, and Tennessee included in the ND2 analysis. Sites are color-coded by haplotype (see Fig. 1), with the sizes of the circle indicating the number of samples from each site. Inset map of the United States shows the study area enclosed in a square. Map generated in QGIS version 3.10.14-A Coruña

RAD SNP dataset

A total of 126 individual cavefishes were sequenced. SNP genotyping of RAD loci data resulted in a dataset of 11,289 variable SNPs with 8% missing data. Population structure analysis with DAPC (Fig. 5) and fastSTRUCTURE (Fig. 6) supported four clusters. Maximum likelihood phylogenetic analysis again recovered two main clades corresponding to the two species, with subclades mostly matching population clusters identified by DAPC and fastSTRUCTURE (Fig. 7). Twenty specimens from Tennessee corresponded to the Spring Cavefish (Fig. 8). The Shawnee Hills Cavefish clade was divided into three subclades. Two subclades were identified within Kentucky: one represented by 31 individuals in the Cumberland River, and one represented by 47 individuals from the Green River (Fig. 8). The Clarksville Lake Cave individual in Tennessee (Cumberland River) was grouped with the Kentucky Cumberland River population cluster in fastSTRUCTURE, but was reconstructed as sister to all other samples of Shawnee Hills Cavefish in the phylogeny. Furthermore, the Green River clade in Kentucky was more closely related to a clade including Illinois-Missouri samples (14 samples for each state) than they were to the Cumberland River clade, but the relationship was not strongly supported (80% UFBS for clade formed by Green River Kentucky and Illinois-Missouri clades). An Illinois clade was identified as sister to a Missouri clade (Fig. 7); however, despite the clearly distinct clades in the tree, these two populations were not strongly supported as separate clusters by DAPC or fastSTRUCTURE.

Nuclear BestRAD phylogenetic and population analyses provided greater resolution among specimens than mitochondrial ND2 due to the greater amount of genetic variation in the much larger genomic dataset. Mitochondrial and nuclear data support the genetic distinctiveness of Spring Cavefish and Shawnee Hills Cavefish. While a haplotype unique to Missouri compared to Illinois does indicate some minor genetic differentiation, the genomic RAD data provided far greater resolution in identifying two separate clades. In the Cumberland River and Green River, most haplotypes were unique to each watershed, but they did not form distinct clades, and Hap11 was found in both

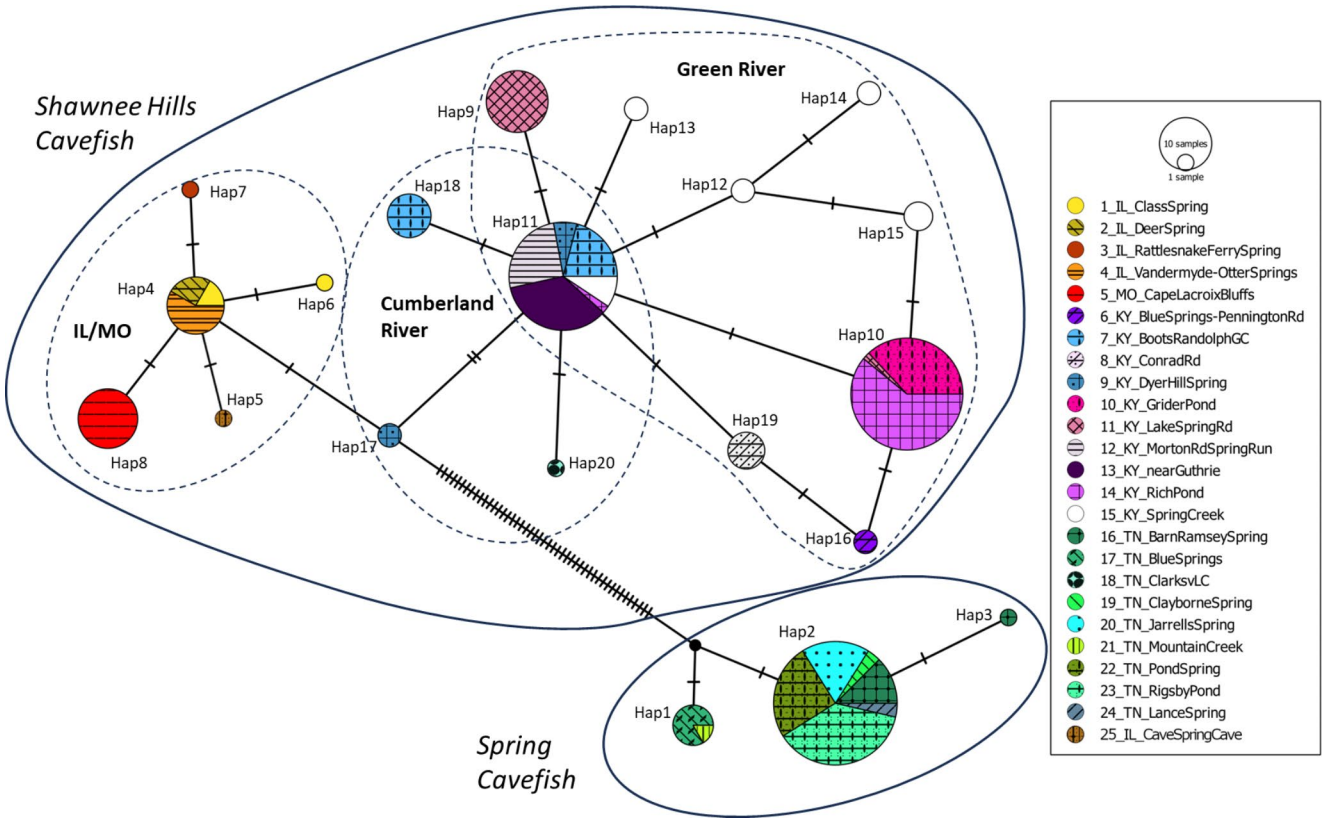

Fig. 4 Haplotype network of ND2 sequences. Circles indicate each haplotype, with the size of the circle indicating the number of samples identified with that haplotype. Each tick indicates the number of substitution differences between haplotypes. Circles are shown with pie charts showing the proportion of samples from each site. Haplotypes belonging to each species are indicated by solid ellipses, while haplotypes belonging to each Shawnee Hills Cavefish population cluster are indicated by dashed ellipses

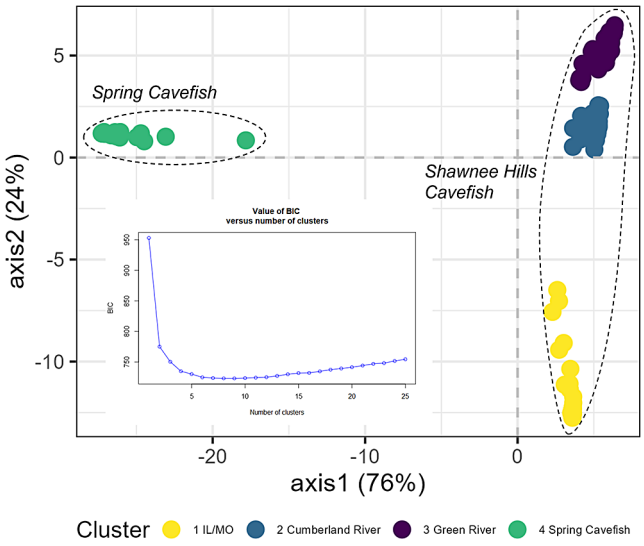

Fig. 5 Discriminant analysis of principal components (DAPC) of 11,289 SNPs. Inset shows the Bayesian Information Criterion (BIC) for the estimation for the number of clusters. Four clusters are evident based on the point where the "elbow" in BIC values occurs or starts to flatten out. Data points belonging to each species are included in dashed ellipses

the Cumberland River and Green River drainages. In the nuclear DNA RAD dataset, two main clades are identified that were split between each river drainage, with the exception of the Clarksville Lake Cave sample (as previously noted). Furthermore, while Cumberland River and Green River ND2 haplotypes almost all form a clade, in the RAD-seq data the Green River clade is more closely related to the Illinois-Missouri population rather than forming a clade with the Cumberland River clade (Fig. 7).

We also assessed genetic differentiation by computing pairwise F_{ST} among the five populations identified by nuclear genomic data (with Illinois and Missouri split as two separate populations) to understand the evidence for genetic connectivity among populations. Populations experiencing consistent gene flow are expected to be quite similar, and therefore have low values of genetic differentiation (near 0), while populations that are completely isolated will have genetic differentiation near 1. Pairwise F_{ST} between Illinois and Missouri were the smallest at 0.272, consistent with their relatively close relationship (Table 1), but still indicative of significant genetic differentiation (all p-values < 0.0001, permutation test). All other pairwise F_{ST} were much larger, and evidence of strong genetic differentiation

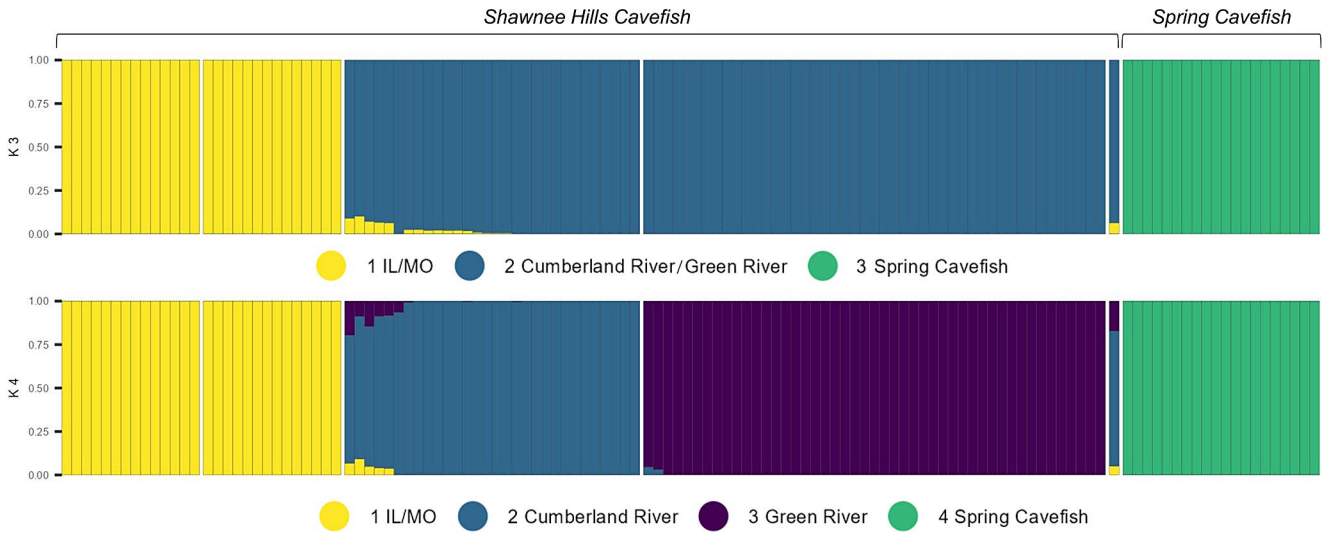

Fig. 6 Population assignment suggested by fastSTRUCTURE was K = 3 (Model complexity that maximizes marginal likelihood = 3; Table S5) and K = 4 (Model components used to explain structure in data = 4) based on 11,289 SNPs. K = 5 did not indicate a separation between Illinois and Missouri (not shown)

(0.436, 0.439, 0.542), and suggesting remarkably low connectivity among genetic clusters of Shawnee Hills Cavefish. On the other hand, shared haplotypes observed across localities within the Spring Cavefish in Tennessee suggests high connectivity. Nearly maximum genetic differentiation was estimated between Shawnee Hills Cavefish populations and Spring Cavefish (≥ 0.927), indicative of a long history of genetic isolation and reinforcing their designation as distinct species (Table 1).

Effective population size

Number of SNPs, segregating sites, and samples used based on selecting the optimal projection are reported in Table S3. From a medium-high population size in deep time, populations experienced a sharp decline followed by a rebound and finally by a more contemporary population decline. This relatively recent population decline, however, appears to have started 1,000–7,500 years ago (depending on the assumed mutation rate), prior to expected anthropogenic influences on groundwater (Figs. 9 and 10). Although population structure can affect estimates of effective population size, analyzing Illinois and Missouri together and separately did not greatly influence results for each population, which is consistent with recent divergence. There is a more recent stability in population size in the last 30–100 years (depending on the assumed mutation rate). Regardless of the mutation rate assumed, the current effective population sizes among populations were generally similar, ranging between 5,000 and 7,000 effective individuals.

Discussion

It can be particularly difficult to assess the distributions and population sizes of RTE species, which has consequences for their conservation. In particular, groundwaters harbor many endemic species and are an increasingly threatened habitat (Saccò et al. 2024); this would be expected to have consequences for the conservation of diverse species dependent on spring and groundwaters, such as *Forbesichthys*. The distribution of species in this genus is difficult to assess due to the intermittent appearance of individuals in surface springs. The population level mitochondrial and genomic investigation herein provides information relevant to understanding potential population breaks and the uniqueness of potential lineages that may deserve protections.

Distribution of the Shawnee Hills and Spring Cavefish

Here we confirm Shawnee Hills Cavefish and Spring Cavefish as distinct lineages consistent with their recognition as separate species. Rather than being a single, wider-ranging species, this restricts each one to reduced distributions. Indeed, given our current data and distributional knowledge, these species may well represent short-range endemics (Harvey et al. 2011; Davis et al. 2015). Given the above, we argue that, at a minimum, a re-evaluation of the global conservation status of each species is imperative. This includes additional surveys for intervening localities between genetically-distinct clusters and species, as well as monitoring of known populations. However, we will note that, to date,

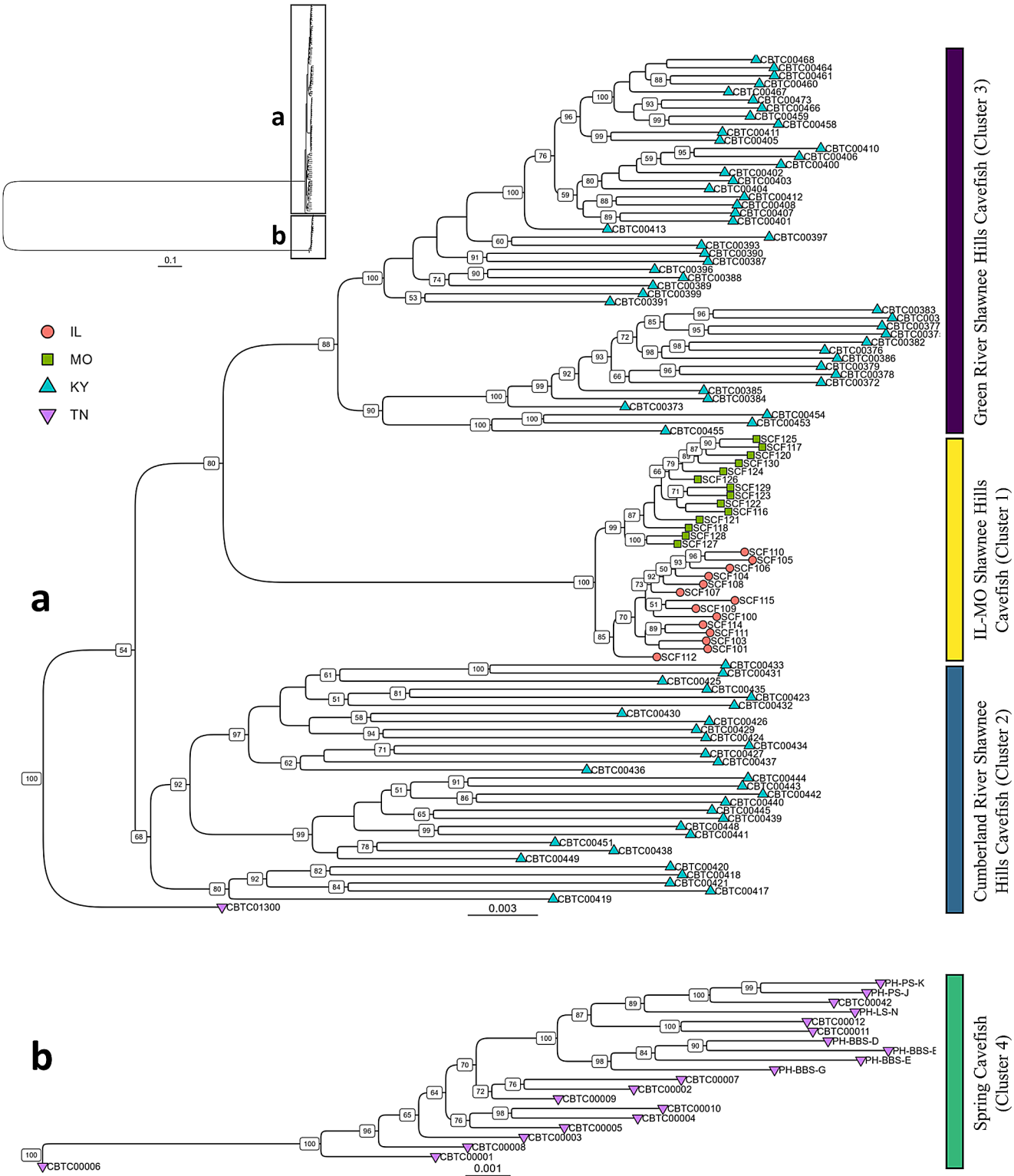

Fig. 7 Maximum likelihood phylogenetic analysis inferred using IQ-TREE version 2.2.0.6 on 10,262 unlinked SNPs from genomic RAD-seq. a, Shawnee Hills Cavefish, *Forbesichthys papilliferus*. b, Spring Cavefish, *Forbesichthys agassizii*. Ultrafast bootstrap support values over 50% are shown

Cluster	IL 1	MO 1	KY 2	KY 3	TN 4
IL 1	-	(0.232-0.319)	(0.416-0.453)	(0.516-0.556)	(0.974-0.977)
MO 1	0.272	-	(0.422-0.459)	(0.520-0.565)	(0.974-0.977)
KY 2	0.436	0.439	-	(0.276-0.306)	(0.924-0.929)
KY 3	0.536	0.542	0.288	-	(0.949-0.948)
TN 4	0.975	0.976	0.927	0.946	-

Table 1 Pairwise genetic differentiation based on 11,289 SNPs between major lineages (see Fig. 4). Estimated F_{ST} values are the low diagonal and lower and upper bound 95% confidence interval values are in the upper diagonal in parenthesis. IL 1 and MO 1 refer to the Illinois and Missouri populations in lineage 1 of Shawnee Hills Cavefish. TN 4 refers to the Spring Cavefish. KY 2 and KY 3 refer to the Cumberland River and Green River clusters of Shawnee Hills Cavefish respectively. F_{ST} values near 1 between the Shawnee Hills Cavefish and Spring Cavefish indicate virtually no genetic connectivity. High F_{ST} values between other lineages indicate limited gene flow, even between Illinois and Missouri. All F_{ST} values showed a p-value < 0.0001 based on permutation test

population distributions, at least in Illinois, appear stable (Metzke and Holtrop 2014).

Our present cluster delineations clarify the potential for geographic breaks that limit connectivity between species and among genetically-distinct populations. The separation between the Shawnee Hills Cavefish and Spring Cavefish is consistent with prior results (Niemiller and Poulson 2010; Niemiller et al. 2013a, c; Hart et al. 2020). Previous studies,

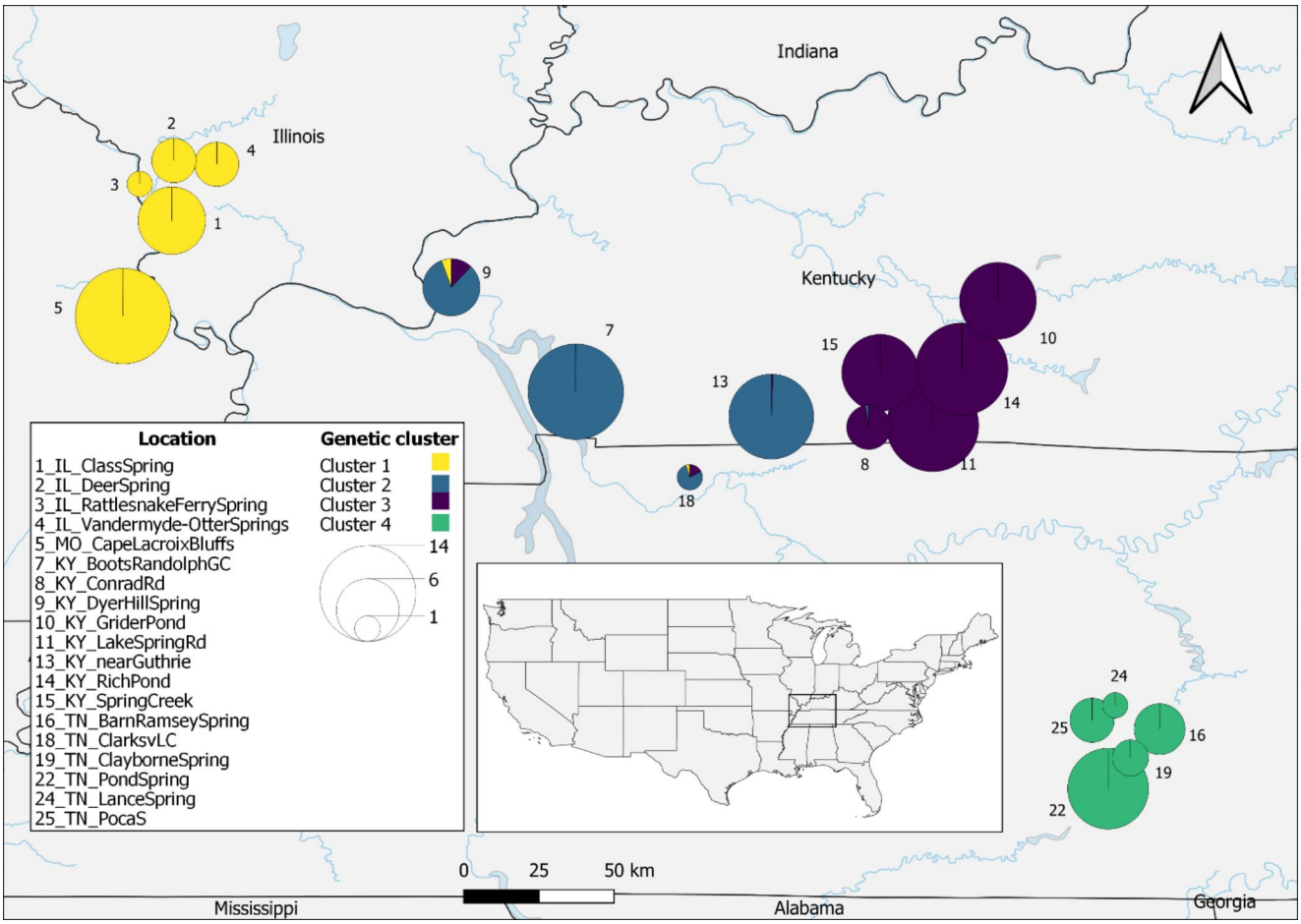

Fig. 8 Sites sampled in Illinois, Missouri, Kentucky, and Tennessee included in the genomic RADseq analysis. Sites are color-coded by the mean probability of cluster assignment across all samples within a site (see Fig. 4), with the sizes of the circle indicating the number of samples from each site. Inset map of the United States shows the study area enclosed in a square. Map generated in QGIS version 3.10.14-A Coruña

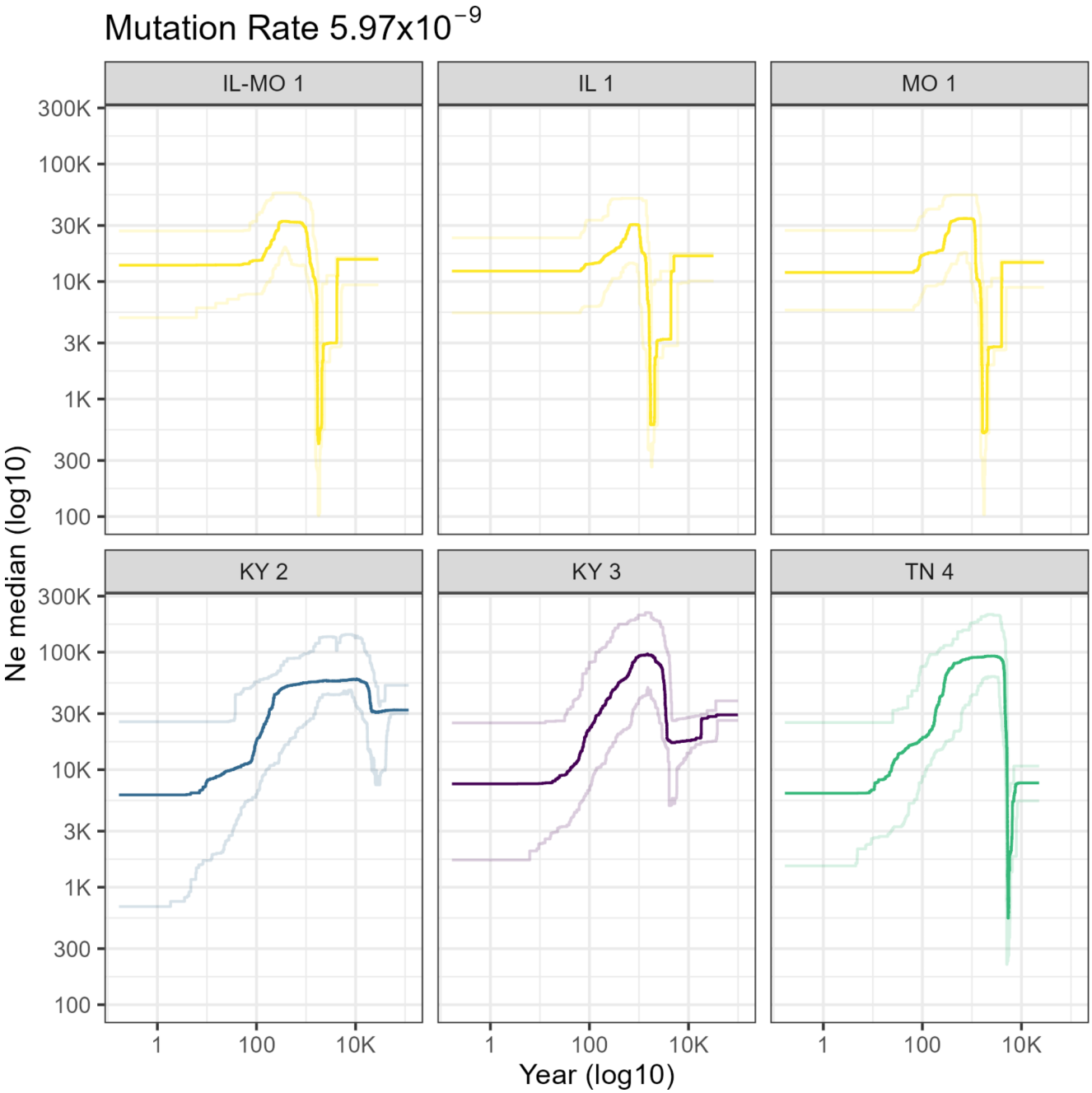

Fig. 9 Effective population size (N_e) estimates from Stairway Plot 2 based on population-specific analyses (see number of sequences and SNPs in Table S3), assuming a generation time of 1 year and a mutation rate of 5.97×10^{-9} per generation (Bergeron et al. 2023). The upper and lower transparent lines correspond to the 95% confidence interval

however, included relatively few samples from relatively few sites. The Shawnee Hills Cavefish is found in southeast Missouri, southern Illinois, southwestern Kentucky, and north-central Tennessee through associations with the Mississippi River, Ohio River, Green River, and Cumberland River. The separation between the Illinois-Missouri and Kentucky clades is explained by the Ohio River, which may act as a barrier to dispersal. The separation between Kentucky populations appears to be best explained by separation between populations found in the Cumberland River watershed and those in the Green River watershed. The clade separation between Illinois and Missouri populations is explained by the current position of the Mississippi River presenting a biogeographic barrier to dispersal. This is potentially consistent with the geological history of the Mississippi River and this population as previously

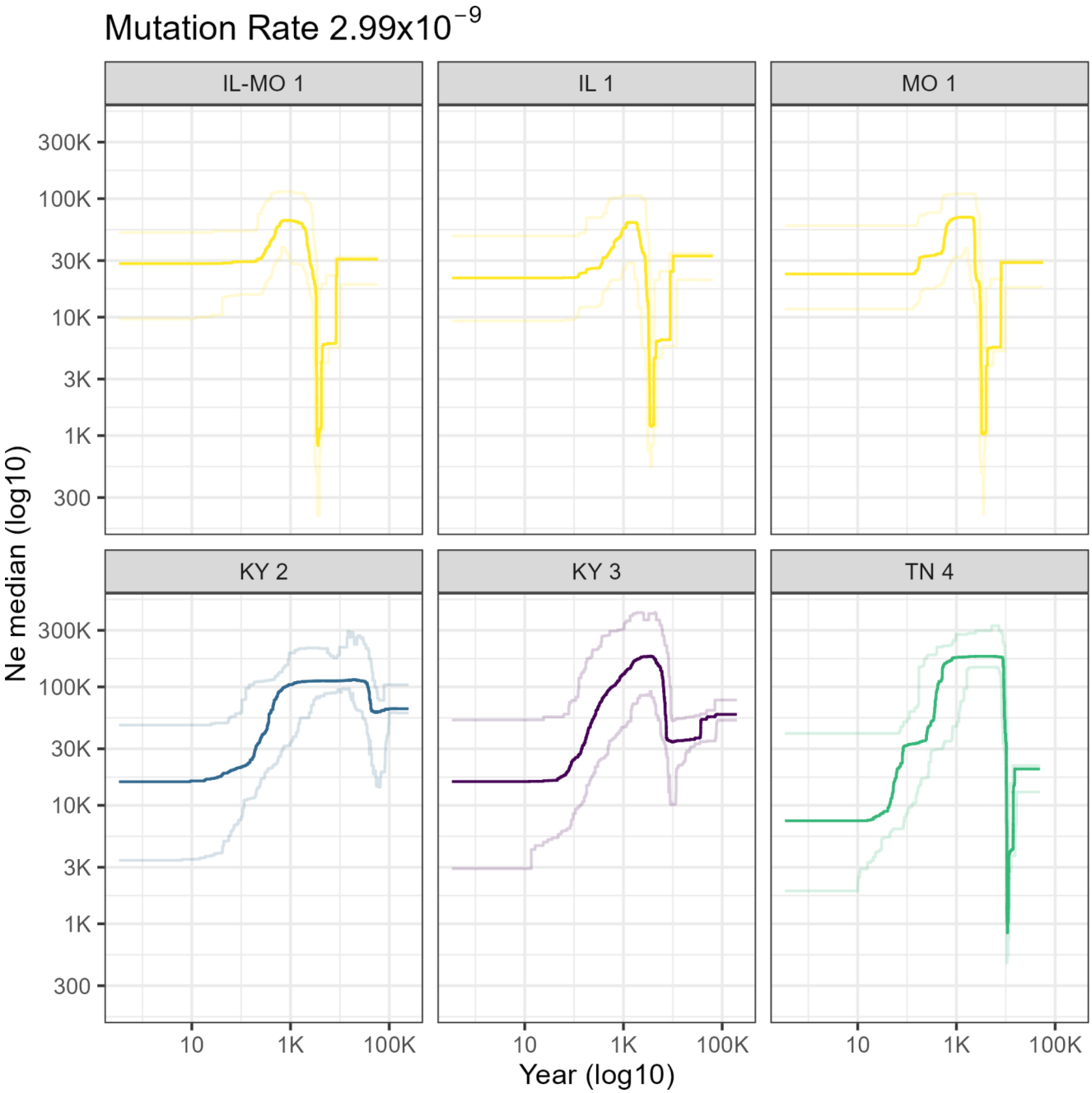

Fig. 10 Effective population size (N_e) estimates from Stairway Plot 2 based on population-specific analyses (see number of sequences and SNPs in Table S3), assuming a generation time of one year and a mutation rate of 2.99×10^{-9} per generation. The upper and lower transparent lines correspond to the 95% confidence interval

proposed (McDonald and Pflieger 1979), and recovered in other cave species (Katz et al. 2018). This Missouri population would have previously been continuous with the Illinois population, but the Mississippi River was diverted eastward roughly 2,000 years ago (Fisk 1944), isolating this population from most of the range in southern Illinois and Kentucky. Consequently, geographic breaks in *Forbesichthys* appear to be fairly well explained by major watershed boundaries.

Alternatively, while the Illinois cluster in the Pine Hills corresponds to Devonian-age geology, the Missouri cluster corresponds to Ordovician-age geology and the Kentucky and Tennessee clusters correspond to the Highland Rim in Mississippian-age geology (McDowell 1986; Starnes and Etnier 1986; Thompson 1991; Kolata and Nimz 2010). This

suggests that the distribution and lack of gene flow may also be influenced by geological strata. Thus, while the Mississippi River's role as a barrier is significant, geological factors should not be discounted in understanding the population dynamics of *Forbesichthys*.

Genetic connectivity patterns among populations of the Shawnee Hills Cavefish and Spring Cavefish

Some aspects of the population genomics of Shawnee Hills Cavefish remain to be resolved. The locality on the Kentucky side of the Lower Ohio River (Dyer Hill Spring, site 9) exhibits slight admixture with the Illinois-Missouri population, while other localities in Kentucky appear to have little to no contribution (Figs. 6 and 8). Despite the high genetic differentiation between the two lineages suggesting virtually no connectivity overall, this slight admixture could be indicative of limited connectivity between populations in close proximity on either side of the Ohio River, or perhaps ancestral polymorphism shared between populations. Shawnee Hills Cavefish have been intermittently found in the Lower Ohio-Bay HUC8 on the east side (Metzke and Holtrop 2014), although none were found in the course of the present study. Individuals here would be geographically close to the westernmost Kentucky populations, which may promote occasional introgression. Given that the introgressed signal is currently assigned to the sampled Illinois population, this suggests that Lower Ohio Illinois individuals are genetically similar to the western Illinois samples we obtained; in other words, that there is a single widespread population in Illinois rather than a unique population in the Lower Ohio-Bay HUC8. Although the present-day Ohio River may have intermittently isolated the Illinois and Kentucky populations, occasionally, individuals of the Illinois population may have crossed the Ohio River through either surface or subterranean waters. However, our lack of genomic data of the Shawnee Hill Cavefish from the Lower Ohio River renders it an open question as to whether individuals can move across the Ohio River, and whether gene flow is unidirectional (i.e. only from Illinois to Kentucky) or bidirectional (i.e. also from Kentucky to Illinois).

The relationship of Shawnee Hills Cavefish in Tennessee (Clarksville Lake Cave, site 18) to other populations of Shawnee Hills Cavefish in Kentucky also requires additional research. While this sample is assigned to the Cumberland River population, fastSTRUCTURE suggests some probability of admixture from the Illinois (5%) and Green River (17%) populations of the Shawnee Hills Cavefish, while phylogenetic analysis suggests it is sister to all other Shawnee Hills Cavefish (Fig. 6). Interpreting the fastSTRUCTURE result might suggest some form of introgression into this region from Illinois and Green River, but this is unlikely due to its greater geographic distance from Illinois. An alternative interpretation from the phylogeny is this sample represents a fourth lineage of Shawnee Hills Cavefish, and therefore some ancestral polymorphism is shared between populations, although the placement outside the remaining samples is poorly supported (54% UFBS for the clade formed by the remaining samples). Nevertheless, it is most likely that this sample is somewhat biased by missing data, which can both pull samples towards the root in phylogenetic analysis and increase uncertainty in population assignment in structure analyses. We find that this sample has 24% missing data, which is relatively high compared to the overall (mean) proportion of 8%, although it is also not the sample with maximum missing data (range 0.66–37%). The ND2 haplotype for this sample is unique, but groups within other Kentucky samples, consistent with this population not representing a separate lineage.

The distribution and habitat of the Spring Cavefish almost mirrors that of the Barrens Topminnow (*Fundulus julisia*) a critically endangered species restricted to springs and spring-fed creeks (Williams and Etnier 1982; Jelks et al. 2008), yet the genetic signature of the two species is very different. The topminnow has two ESUs between the Elk River and Caney Fork, and 2 MUs between two different river systems within the Caney Fork drainage (Hurt et al. 2017). However, no such pattern is shown in the Spring Cavefish between the Elk, Duck (*F. julisia*, historical records), and Caney Fork systems. This seems to suggest that the Spring Cavefish likely uses aquifer connections for gene flow that are not reflected in surface drainages.

Evolutionary significant units and management units within the Shawnee Hills Cavefish

Given unexpected and pronounced genetic diversity recovered within the Shawnee Hills Cavefish (Figs. 3, 4, 5, 6, 7 and 8), we conservatively suggest that, at a minimum, the three major lineages of Shawnee Hills Cavefish should be designated as Evolutionarily Significant Units (ESUs) (Moritz 1994). Specifically, an Illinois-Missouri clade, the Cumberland River drainage clade, and the Green River drainage clade all represent distinct ESUs and should be managed accordingly (Figs. 5, 6, 7 and 8). Despite finding similarity in the mitochondrial and nuclear genomes among Illinois and Missouri fish, suggesting these localities are closely related, the recovery of significant genetic differentiation between Missouri and Illinois suggests population fragmentation between the two states, providing support for the recognition of separate Management Units (MUs) for each state. This yields a total of three ESUs within the Shawnee Hills Cavefish, one of which is split into two MUs. The inclusion of the Cumberland River and Green River

populations within Shawnee Hills Cavefish contradicts the suggestion that Shawnee Hills Cavefish may be restricted to Illinois and Missouri (Adams et al. 2020). These populations of *Forbesichthys* are clearly more closely-related to the Illinois and Missouri populations than they are to Spring Cavefish, and we thus assign them to the Shawnee Hills Cavefish. Evolutionary divergence among lineages may indicate that ESUs are not interchangeable if future relocation is ever needed. Transplantation to Adams County of Shawnee Hills Cavefish has previously failed (Adams et al. 2020), which might be explained by differences in locally adaptive alleles among ESUs. This regionalization also has implications for policy of groundwater usage (Taylor et al. 2024), as populations occurring in particularly restricted regions (e.g. the Missouri clade) may be more threatened by continued intensification of groundwater usage. Moreover, prolonged droughts resulting from climate change, along with groundwater pollution stemming from agricultural activities, could exacerbate the threat of extirpation for certain localized populations.

Effective population size is a metric that facilitates exploration of the level of genetic diversity, which is important for understanding the potential that populations may experience elevated inbreeding risk (Lohmueller et al. 2008), or lack sufficient genomic diversity and thus capacity for adaptation (England et al. 2003). There is some uncertainty in estimates of effective population size due to both estimation uncertainty as well as uncertainty in an appropriate mutation rate. Generally, effective population sizes seem to have been relatively low compared to historic or prehistoric effective population sizes, but appear to have been stable between the last 30–100 years, between 5,000 and 7,000 individuals (with broader confidence intervales) may exist for each of the ESUs (i.e., Illinois-Missouri clade, Cumberland River, and Green River drainage clades). Nonetheless, even these estimates are above the 500 threshold considered as a good measure that facilitates the maintenance of genetic adaptability (Harmon and Braude 2010), or even the higher 1000 threshold (Frankham et al. 2014). These population size estimates stand in contrast to those of some other fish species associated with groundwater-fed systems, such as the Watercress Darter (*Etheostoma nuchale*) (N_e: 657–1,760) (Fluker et al. 2010), the Trispot Darter (*E. trisella*) (N_e: 33–208) (Fast et al. 2024), and the Arkansas Darter (*E. cragini*) (N_e: 208–1,360) (Baker et al. 2018). Similarly, many species across diverse taxa fail to meet the 50/500 rule thresholds, although freshwater fishes are more likely to meet these thresholds than mammals, amphibians, and plants, but less likely than marine fishes (Clarke et al. 2023). Given the replication across populations and species in changes in effective population size change, we suggest that the population decline and rebound may have been driven by ancient climatic events affecting all populations concurrently. Nevertheless, our estimates of effective population size so far suggest that there is sufficient genetic diversity (see also summary statistics Table S4) in all *Forbesichthys* spp. populations to avoid inbreeding and to adapt to changing environments.

The proposed Illinois and Missouri MUs presented similar effective population sizes (10,000+), both when considered as single and separate populations, which is reflective of their recent divergence. While counterintuitive because the sum of the effective population sizes estimated for each state is not the estimated effective population size of the samples analyzed together, this similarity in effective population size indicates just how recently the two populations have diverged, as estimated effective population sizes through time reflect when these two populations comprised a single ancestral population.

Considerations for future research of the Shawnee Hills Cavefish

Genetic monitoring of the Shawnee Hills Cavefish is important for assessing its distribution and genetic diversity. Specifically, determining if populations from the Ohio River-Bay HUC8 are extant, and if so, genetic data shed light on whether individuals from these watersheds are connected with individuals from other Illinois watersheds and if there is signal for introgression from the Cumberland River. Furthermore, repeated sampling could be useful for assessing changes in genetic diversity over time (both short- and long-term), which may help to confirm whether genetic diversity has indeed remained stable. We also foresee continued decreases in genome sequencing costs and technological shifts to allow for more complete investigations of genome-wide genetic diversity, which may provide further information in relation to conservation.

Presently available data are sufficient to test a variety of genomic questions with potential conservation significance. Evolutionary distinctiveness has been proposed as complementary information to prioritize species for conservation, as extinction of evolutionarily distinct lineages may lead to the irrevocable loss of tens of thousands to even millions of years of independent evolutionary history (Nee and May 1997; Veron et al. 2017); while there appears to be divergence within the Shawnee Hills Cavefish populations, it would be possible to estimate the timing of divergence between these ESUs to quantify their evolutionary distinctiveness. Further research could be aimed at determining whether genetic differentiation among populations may be due exclusively to genetic drift, or if there are signals of local adaptation, which as previously noted may have consequences in extreme cases where extirpation is likely and

propagation, reintroduction, or translocation may be a last resort (thought these measures should be taken with caution, see George et al. 2009). In addition, given the availability of whole genome sequence data for spring cavefish species (i.e., whole mitogenomes), it may be possible to design, optimize, and validate eDNA assays for further assessing population presence and seeking new sites, even when spring cavefishes are in subsurface habitats. Positive eDNA samples can also be used for deeper genome sequencing for understanding genetic diversity and population assignment.

Our genomic data address some of the Wallacean and Prestonian shortfalls relating to species of *Forbesichthys*. A clearer understanding of species distribution, evolutionary significant units, management units, population sizes, and genetic differentiation between springs and caves add to the general knowledge necessary to develop more effective and meaningful species management. Our findings of restrictive connectivity across the range of the Shawnee Hills Cavefish coupled with the identification of distinct three ESUs and two MUs leads to the conclusion that local management across its range is necessary to protect the diversity found within the Shawnee Hills Cavefish. Similarly, the much smaller range of the Spring Cavefish, including areas of heavy plant nursery agriculture with groundwater pumping (Federal Register 2019), suggests a need for increased conservation efforts.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10592-024-01640-8.

Acknowledgements We thank Stephanie Brandt (Kentucky Department of Fish & Wildlife Resources) for samples from Kentucky and Tennessee. We appreciate Jeffrey Haas for assistance with managing the abe server at the UIUC School of Integrative Biology, which we used for computational analyses. We are grateful to Conservation Genetics Editor Michael Russelo, and two anonymous reviewers for their thoughtful recommendations to improve this manuscript. Finally, we thank the Illinois Department of Natural Resources (contract no. 22NH0202 to MT), the USDA National Institute of Food and Agriculture, Hatch project 1026333 (ILLU-875-984 to RVC), and the National Science Foundation (award no. 2047939 to MLN) for the support for this work.

Author contributions RVC, JBC, MAD, BM, MLN, and MT contributed to the study conception and design. Sample collection was performed by BM, MLN, PBH, BRK, and MRT. Material preparation was performed by RVC, JBC, and ANC. Data collection and formal analysis were performed by RVC and JBC. The first draft of the manuscript was written by MT and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the Illinois Department of Natural Resources (contract no. 22NH0202 to MT), by the USDA National Institute of Food and Agriculture, Hatch project 1026333 (ILLU-875-984 to RVC), and by the National Science Foundation (award no. 2047939 to MLN). The authors have no relevant financial or non-financial interests to disclose.

Data availability The datasets generated during and/or analyzed during the current study are available at https://zenodo.org/records/13685671. Raw demultiplexed genomic data were deposited into NCBI SRA accession PRJNA1147188. The mtDNA sequences generated were uploaded to NCBI GenBank (PQ179047 - PQ179263).

Declarations

Competing interests The authors declare no competing interests.

References

Ackiss AS, Larson WA, Stott W (2020) Genotyping-by-sequencing illuminates high levels of divergence among sympatric forms of coregonines in the Laurentian Great Lakes. Evol Appl 13(5):1037–1054. https://doi.org/10.1111/eva.12919

Adams GL, Adams SR, Burr BM (2000) Early Natural History and Larval Habitat Utilization of the Spring Cavefish, Forbesichthves agassizi, in Southern Illinois. Illinois Department of Natural Resources. https://dnr.illinois.gov/content/dam/soi/en/web/dnr/grants/documents/wpfgrantreports/2000063w.pdf

Adams GL, Burr BM, Warren ML Jr. (2020) Amblyopsidae: Cavefishes. Freshwater fishes of North America: Characidae to Poeciliidae, vol 2. Johns Hopkins University

Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR (2016) Rad capture (rapture): flexible and efficient sequence-based genotyping. Genetics 202(2):389–400. https://doi.org/10.1534/genetics.115.183665

Anon A (1996) The evaluation of forensic DNA evidence. National Academies. https://doi.org/10.17226/5141

Arponen A (2012) Prioritizing species for conservation planning. Biodivers Conserv 21(4):875–893. https://doi.org/10.1007/s10531-012-0242-1

Avise JC (2010) Perspective: conservation genetics enters the genomics era. Conserv Genet 11(2):665–669. https://doi.org/10.1007/s10592-009-0006-y

Baker J, Wagner B, Wood R (2018) Gene flow and genetic structure of two of Arkansas's rarest darter species (Teleostei: Percidae), the Arkansas Darter, *Etheostoma cragini*, and the Least Darter, *E. Microperca*. J Ark Acad Sci 72(1):148–160. https://doi.org/10.54119/jaas.2018.7213

Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St. Leger J, Shao C, Stiller J, Gilbert MTP, Schierup MH, Zhang G (2023) Evolution of the germline mutation rate across vertebrates. Nature 615(7951) Article 7951. https://doi.org/10.1038/s41586-023-05752-y

Brown A, Weir B (1983) Measuring genetic variation in plant populations. In: Tanksley SD, Orton TJ (eds) Isozymes in Plant Genetics and breeding, part A. Elsevier Science Amsterdam, pp 219–239

Burda K, Konczal M (2023) Validation of machine learning approach for direct mutation rate estimation. Mol Ecol Resour 23(8):1757–1771. https://doi.org/10.1111/1755-0998.13841

Cardoso P, Erwin T, Borges P, New T (2011) The seven impediments in invertebrate conservation and how to overcome them. Biol Conserv 144:2647–2655. https://doi.org/10.1016/j.biocon.2011.07.024

Chakrabarty P, Prejean J, Niemiller M (2014) The Hoosier cavefish, a new and endangered species (Amblyopsidae, Amblyopsis) from the caves of southern Indiana. ZooKeys 412:41–57. https://doi.org/10.3897/zookeys.412.7245

Chichorro F, Juslén A, Cardoso P (2019) A review of the relation between species traits and extinction risk. Biol Conserv 237:220–229. https://doi.org/10.1016/j.biocon.2019.07.001

Clarke SH, Lawrence ER, Matte J-M, Gallagher BK, Salisbury SJ, Michaelides SN, Koumrouyan R, Ruzzante DE, Grant JWA, Fraser DJ (2023) Global assessment of effective population sizes: consistent taxonomic differences in meeting the 50/500 rule. bioRxiv 20230922558974. https://doi.org/10.1101/2023.09.22.558974

Corush JB, Pierson TW, Shiao J-C, Katayama Y, Zhang J, Fitzpatrick BM (2022) Amphibious mudskipper populations are genetically connected along coastlines, but differentiated across water. J Biogeogr 49(4):767–779. https://doi.org/10.1111/jbi.14345

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. Bioinformatics 27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

Davis MA, Douglas MR, Webb CT, Collyer ML, Holycross AT, Painter CW, Kamees LK, Douglas ME (2015) Nowhere to go but up: impacts of climate change on demographics of a short-range endemic (*Crotalus willardi obscurus*) in the Sky-Islands of Southwestern North America. PLoS ONE 10(6):e0131067. https://doi.org/10.1371/journal.pone.0131067

England PR, Osler GHR, Woodworth LM, Montgomery ME, Briscoe DA, Frankham R (2003) Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. Conserv Genet 4(5):595–604. https://doi.org/10.1023/A:1025639811865

Etnier DA, Starnes WC (1993) The fishes of Tennessee. University of Tennessee

Fast KM, Fluker BL, Kuhajda BR, O'Neil PE, McGregor SW, Piteo MS, Sandel MW (2024) Conservation genomics of the threatened Trispot Darter (*Etheostoma trisella*). Conserv Genet 25(2):291–304. https://doi.org/10.1007/s10592-023-01572-9

Federal Register (2019) Endangered and Threatened Wildlife and Plants; Endangered Species Status for Barrens Topminnow, proposed rule. Vol. 83 (3): 490–498. https://www.federalregister.gov/documents/2019/10/21/2019-22857/endangered-and-threatened-wildlife-and-plants-endangered-species-status-for-barrens-topminnow

Federal Register (1988) Reclassification of the Alabama Cavefish from threatened to endangered. Vol. 53, No. 188, Part IV (28 September 1988)

Feng C, Pettersson M, Lamichhaney S, Rubin C-J, Rafati N, Casini M, Folkvord A, Andersson L (2017) Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. eLife 6:e23907. https://doi.org/10.7554/eLife.23907

Fisk HN (1944) Geological investigation of the alluvial valley of the Lower Mississippi River [Technical Report]. War Department, Corps of Engineers. https://usace.contentdm.oclc.org/digital/collection/p266001coll1/id/10015/

Fluker BL, Kuhajda BR, Lang NJ, Harris PM (2010) Low genetic diversity and small long-term population sizes in the spring endemic watercress darter, *Etheostoma nuchale*. Conserv Genet 11(6):2267–2279. https://doi.org/10.1007/s10592-010-0111-y

Frankham R, Bradshaw CJA, Brook BW (2014) Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biol Conserv 170:56–63. https://doi.org/10.1016/j.biocon.2013.12.036

Frimpong EA, Angermeier PL (2009) Fish traits: a database of ecological and life-history traits of Freshwater fishes of the United States. Fisheries 34(10):487–495. https://doi.org/10.1577/1548-8446-34.10.487

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends Ecol Evol 27(9):489–496. https://doi.org/10.1016/j.tree.2012.05.012

Gales SM, Parsons KT, Biesack EE, Ready J, Siccha-Ramirez R, Rosa LC, Rosa R, Rotundo MM, Bills R, Rodrigues AES, Rodrigues-Filho LFS, McDowell J, Sales JBL (2023) Almost half of the Gymnura Van Hasselt, 1823 species are unknown: phylogeographic inference as scissors for cutting the hidden Gordian knot and clarify their conservation status. J Syst Evol 0(0):1–24. https://doi.org/10.1111/jse.13027

George AL, Kuhajda BR, Williams JD, Cantrell MA, Rakes PL, Shute JR (2009) Guidelines for propagation and translocation for Freshwater Fish Conservation. Fisheries 34(11):529–545. https://doi.org/10.1577/1548-8446-34.11.529

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186. https://doi.org/10.1111/j.1471-8278

Gruber B, Unmack PJ, Berry OF, Georges A (2018) Dartr: an r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour 18(3):691–699. https://doi.org/10.1111/1755-0998.12745

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet 5(10):e1000695. https://doi.org/10.1371/journal.pgen.1000695

Harmon LJ, Braude S (2010) Conservation of small populations: Effective Population sizes, Inbreeding, and the 50/500 rule. In: Braude S, Low BS (eds) An introduction to methods and models in Ecology, Evolution, and Conservation Biology. Princeton University Press, pp 125–138. https://doi.org/10.1515/9781400835454-017

Hart PB, Niemiller ML, Burress ED, Armbruster JW, Ludt WB, Chakrabarty P (2020) Cave-adapted evolution in the north American amblyopsid fishes inferred using phylogenomics and geometric morphometrics. Evolution 74(5):936–949. https://doi.org/10.1111/evo.13958

Harvey MS, Rix MG, Framenau VW, Hamilton ZR, Johnson MS, Teale RJ, Humphreys G, Humphreys WF (2011) Protecting the innocent: studying short-range endemic taxa enhances conservation outcomes. Invertebr Syst 25(1):1. https://doi.org/10.1071/is11011

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35(2):518–522. https://doi.org/10.1101/153916

Hortal J, de Bello F, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, Ladle RJ (2015) Seven shortfalls that Beset large-scale knowledge of Biodiversity. Annu Rev Ecol Evol Syst 46(46, 2015):523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hurt C, Kuhajda B, Harman A, Ellis N, Nalan M (2017) Genetic diversity and population structure in the barrens Topminnow (*Fundulus julisia*): implications for conservation and management of a critically endangered species. Conserv Genet 18(6):1347–1358. https://doi.org/10.1007/s10592-017-0984-0

Jelks HL, Walsh SJ, Burkhead NM, Contreras-Balderas S, Diaz-Pardo E, Hendrickson DA, Lyons J, Mandrak NE, McCormick F, Nelson JS, Platania SP, Porter BA, Renaud CB, Schmitter-Soto JJ, Taylor EB (2008) Conservation status of Imperiled North American Freshwater and Diadromous Fishes. Fisheries 33(8):372–407. https://doi.org/10.1577/1548-8446-33.8.372. Jr.

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11(1):94. https://doi.org/10.1186/1471-2156-11-94

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14(6):587–589. https://doi.org/10.1038/nmeth.4285

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment Software Version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780. https://doi.org/10.1093/molbev/mst010

Katz AD, Taylor SJ, Davis MA (2018) At the confluence of vicariance and dispersal: Phylogeography of cavernicolous springtails (Collembola: Arrhopalitidae, Tomoceridae) codistributed across a geologically complex karst landscape in Illinois and Missouri. Ecol Evol 8(20):10306–10325. https://doi.org/10.1002/ece3.4507

Kocher TD, Conroy JA, McKaye KR, Stauffer JR, Lockwood SF (1995) Evolution of NADH Dehydrogenase Subunit 2 in East African Cichlid Fish. Mol Phylogenet Evol 4(4):420–432. https://doi.org/10.1006/mpev.1995.1039

Kolata DR, Nimz CK (eds) (2010) Geology of Illinois. Illinois State Geological Survey. http://www.worldcat.org/oclc/706711791

Illinois Endangered Species Protection Board (2020) Checklist of Illinois endangered and threatened animals and plants. Illinois Endangered Species Protection Board, Springfield, Illinois, p. 10. https://dnr.illinois.gov/content/dam/soi/en/web/dnr/espb/documents/et-list-review-andrevision/illinoisendangeredandthreatenedspecies.pdf

Leigh JW, Bryant D (2015) POPART: full-feature software for haplotype network construction. Methods Ecol Evol 6(9):1110–1116. https://doi.org/10.1111/2041-210X.12410

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Liu X, Fu Y-X (2015) Exploring population size changes using SNP frequency spectra. Nat Genet 47(5):555–559. https://doi.org/10.1038/ng.3254

Liu X, Fu Y-X (2020) Stairway plot 2: demographic history inference with folded SNP frequency spectra. Genome Biol 21(1):280. https://doi.org/10.1186/s13059-020-02196-9

Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R, Clark AG, Bustamante CD (2008) Proportionally more deleterious genetic variation in European than in African populations. Nature 451(7181):994–997. https://doi.org/10.1038/nature06611

Lomolino M, Heaney L (2004) Frontiers of Biogeography: New Directions in the Geography of Nature M. V. Lomolino, L. R. Heaney. 2004. Frontiers of Biogeography: New Directions in the Geography of Nature. In *Ecoscience* (p. 425). Sinauer Associates. https://doi.org/10.2980/1195-6860(2006)13[424:FOBNDI]2.0.CO;2

Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat Ecol Evol 2(12):1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Marth GT, Czabarka E, Murvai J, Sherry ST (2004) World Populations Genet 166(1):351–372. https://doi.org/10.1534/genetics.166.1.351. The Allele Frequency Spectrum in Genome-Wide Human Variation Data Reveals Signals of Differential Demographic History in Three Large

McDonald EF, Pflieger WL (1979) The Spring Cavefish, *Chologaster agassizi* (Pisces: Amblyopsidae), in Southeastern Missori. Am Midl Nat 102(1):194–196. https://doi.org/10.2307/2425086

McDowell RC (1986) The geology of Kentucky: A text to accompany the geologic map of Kentucky. In *Professional Paper* (1151-H). U.S. Geological Survey. https://doi.org/10.3133/pp1151H

Metzke BA, Holtrop AM (2014) Survey of Spring Cavefish (Forbesichthys agassizii), Ozark minnow (Notropis nubilus) and Largescale Stoneroller (Campostoma oligolepis) Status in Illinois. Illinois Natural History Survey. http://hdl.handle.net/2142/47068

Metzke BA, Adams GL, Adams SR (2016) Status of Spring Cavefish (*Forbesichthys agassizii*) in Southern Illinois—ProQuest. Trans Ill State Acad Sci 109:35–39

Mijangos JL, Gruber B, Berry O, Pacioni C, Georges A (2022) dartR v2: an accessible genetic analysis platform for conservation, ecology and agriculture. Methods Ecol Evol 13(10):2150–2158. https://doi.org/10.1111/2041-210X.13918

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R, Teeling E (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015

Missouri Department of Conservation (2024) Missouri species and communities of conservation concernchecklist. Missouri Department of Conservation, Jefferson City, Missouri. https://mdc.mo.gov/sites/default/files/2024-03/SOCC%20Checklist%202024_ADA_0.pdf

Moritz C (1994) Defining evolutionarily significant units for conservation. Trends Ecol Evol 9(10):373–375. https://doi.org/10.1016/0169-5347(94)90057-4

NatureServe (2024) NatureServe Explorer. https://explorer.natureserve.org/

NatureServe (2013) Forbesichthys agassizii. The IUCN Red List of Threatened Species 2013: E.T202611A18230637. Https://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T202611A18230637.en. Accessed on 31 March 2024.

Nee S, May RM (1997) Extinction and the loss of Evolutionary History. Science 278(5338):692–694. https://doi.org/10.1126/science.278.5338.692

Niemiller ML, Poulson TL (2010) Subterranean fishes of North America: Amblyopsidae. Biology of Subterranean fishes. CRC

Niemiller ML, Fitzpatrick BM, Shah P, Schmitz L, Near TJ (2013a) Evidence for repeated loss of selective constraint in rhodopsin of Amblyopsid Cavefishes (Teleostei: Amblyopsidae). Evolution 67(3):732–748. https://doi.org/10.1111/j.1558-5646.2012.01822.x

Niemiller ML, Graening GO, Fenolio DB, Godwin JC, Cooley JR, Pearson WD, Fitzpatrick BM, Near TJ (2013b) Doomed before they are described? The need for conservation assessments of cryptic species complexes using an amblyopsid cavefish (Amblyopsidae: Typhlichthys) as a case study. Biodivers Conserv 22(8):1799–1820. https://doi.org/10.1007/s10531-013-0514-4

Niemiller ML, McCandless JR, Reynolds RG, Caddle J, Near TJ, Tillquist CR, Pearson WD, Fitzpatrick BM (2013c) Effects of climatic and geological processes during the Pleistocene on the Evolutionary history of the Northern Cavefish, *Amblyopsis spelaea* (Teleostei: Amblyopsidae). Evolution 67(4):1011–1025. https://doi.org/10.1111/evo.12017

Niemiller ML, Davis MA, Tan M, Apodaca JJ, Dooley KE, Cucalón RV, Benito JB, Niemiller KDK, Hardman RH, Istvanko D, Thames D (2022a) Mitochondrial DNA and Population Genomics reveal additional cryptic diversity in the Green Salamander (Subgenus, *Castaneides*) Species Complex. Front Conserv Sci 3. https://doi.org/10.3389/fcosc.2022.890859

Niemiller ML, Hardman R, Thames D, Istvanko D, Davis MA, Ogle C, Niemiller KDK, Dooley KE, Clark TM (2022b) The distribution and conservation status of the green salamander (*Aneides aeneus*) in Tennessee, USA. Herpetological Conserv Biology 17(2):249–265

Office of Kentucky Nature Preserves (2019) Endangered, Threatened, and Special Concern Plants, Animals, and Natural Communities of Kentucky. Office of Kentucky Nature Preserves, Kentucky Energy and Environment Cabinet, Frankfort, Kentucky. https://

eec.ky.gov/Nature-Preserves/biodiversity/Documents/Rare_species_of_Kentucky.pdf
Ortiz EM (2019) vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis. (v2.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.2540861
Ouborg NJ (2010) Integrating population genetics and conservation biology in the era of genomics. Biol Lett 6(1):3–6. https://doi.org/10.1098/rsbl.2009.0590
Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Kuke, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. Trends Genet 26(4):177–187. https://doi.org/10.1016/j.tig.2010.01.001
Paradis E (2010) Pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26(3):419–420. https://doi.org/10.1093/bioinformatics/btp696
Patton A, Apodaca J, Corser J, Wilson C, Williams L, Cameron A, Wake D (2019) A New Green Salamander in the Southern appalachians: evolutionary history of Aneides aeneus and implications for management and conservation with the description of a cryptic microendemic species. Copeia 107:748. https://doi.org/10.1643/CH-18-052
Peterson MNA (1962) The Mineralogy and Petrology of Upper Mississippian Carbonate rocks of the Cumberland Plateau in Tennessee. J Geol 70(1):1–31. https://doi.org/10.1086/626791
Priest MA, DiBattista JD, McIlwain JL, Taylor BM, Hussey NE, Berumen ML (2016) A bridge too far: dispersal barriers and cryptic speciation in an Arabian Peninsula grouper (Cephalopholis hemistiktos). J Biogeogr 43(4):820–832. https://doi.org/10.1111/jbi.12681
Raj A, Stephens M, Pritchard JK (2014) FastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197(2):573–589. https://doi.org/10.1534/genetics.114.164350
Ray CK, Tan M, Armbruster JW (2014) First Record of *Chrosomus erythrogaster* (Cypriniformes: Cyprinidae) in the Mobile Basin. Southeast Nat 13(4):N33–N36
Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stack 2: Analytical methods for paired-end sequencing improve RadSeq-based population genomics. Mol Ecol 28(21):4737–4754
Saccò M, Mammola S, Altermatt F, Alther R, Bolpagni R, Brancelj A, Brankovits D, Fišer C, Gerovasileiou V, Griebler C, Guareschi S, Hose GC, Korbel K, Lictevout E, Malard F, Martínez A, Niemiller ML, Robertson A, Tanalgo KC, Reinecke R (2024) Groundwater is a hidden global keystone ecosystem. Glob Change Biol 30(1):e17066. https://doi.org/10.1111/gcb.17066
Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, Colling G, Dalén L, De Meester L, Ekblom R, Fawcett KD, Fior S, Hajibabaei M, Hill JA, Hoezel AR, Höglund J, Jensen EL, Krause J, Kristensen TN, Zieliński P (2015) Genomics and the challenging translation into conservation practice. Trends Ecol Evol 30(2):78–87. https://doi.org/10.1016/j.tree.2014.11.009
Smith PW, Welch NM (1978) A Summary of the Life History and Distribution of the Spring Cavefish, *Chologaster agassizii*, Putnam, with Population Estimates for the Species in Southern Illinois. *Biological Notes; No. 104*. https://hdl.handle.net/2142/95773
Stanton DWG, Frandsen P, Waples RK, Heller R, Russo I-RM, Orozco-terWengel PA, Pedersen C-ET, Siegismund HR, Bruford MW (2019) More grist for the mill? Species delimitation in the genomic era and its implications for conservation. Conserv Genet 20(1):101–113. https://doi.org/10.1007/s10592-019-01149-5
Starnes WC, Etnier DA (1986) Drainage evolution and fish biogeography of the Tennessee and Cumberland rivers drainage realm. Zoogeography North Am Freshw Fishes, 325–361
Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation Genomics of Threatened Animal Species. *Annual Review of Animal Biosciences*, *1*(Volume 1, 2013), 261–281. https://doi.org/10.1146/annurev-animal-031412-103636
Supple MA, Shapiro B (2018) Conservation of biodiversity in the genomics era. Genome Biol 19(1):131. https://doi.org/10.1186/s13059-018-1520-3
Taylor RS, Manseau M, Wilson PJ (2024) Delineating conservation units should be independent of effective population size. Trends Ecol Evol 39(2):121–122. https://doi.org/10.1016/j.tree.2023.11.010
Thompson TL (1991) Paleozoic succession in Missouri Part 2 Ordovician System. Division of Geology and Land Survey. Missouri Department of Natural Resources. http://archive.org/details/1991MOPaleozoicSuccessionPart2
Veron S, Davies TJ, Cadotte MW, Clergeau P, Pavoine S (2017) Predicting loss of evolutionary history: where are we? Biol Rev 92(1):271–291. https://doi.org/10.1111/brv.12228
Weise JG (1957) The Spring Cave-Fish, *Chologaster papilliferus*. Ill Ecol 38(2):195–204. https://doi.org/10.2307/1931678
Williams JD, Etnier DA (1982) Description of a new species, *Fundulus julisia*, with a redescription of *Fundulus albolineatus* and a diagnosis of the subgenus *Xenisma* (Teleostei: Cyprinodontidae). Occasional Papers Museum Nat History Univ Kans 102:1–20
Woods LP, Inger RF (1957) The Cave, Spring, and swamp fishes of the Family Amblyopsidae of Central and Eastern United States. Am Midl Nat 58(1):232–256. https://doi.org/10.2307/2422371
Xie Z, Frimpong EA, Lee S (2013) FishTraits version 2: Integrating ecological, biogeographic and bibliographic information. Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, 447–448. https://doi.org/10.1145/2467696.2467791
Zhang C, Reid K, Sands AF, Fraimout A, Schierup MH, Merilä J (2023) De Novo Mutation Rates in Sticklebacks. Mol Biol Evol 40(9):msad192. https://doi.org/10.1093/molbev/msad192