

## RESEARCH ARTICLE

# Conservation genomics of urban populations of Streamside Salamander (*Ambystoma barbouri*)

N. Wade Hubbs<sup>1,2</sup>\*, Carla R. Hurt<sup>1</sup>\*, John Niedzwiecki<sup>3</sup>, Brian Leckie<sup>1</sup>, David Withers<sup>4</sup>

**1** Department of Biology, Tennessee Technological University, Cookeville, TN, United States of America, **2** Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States of America, **3** Belmont University, Nashville, TN, United States of America, **4** Tennessee Department of Environment and Conservation, Nashville, TN, United States of America

\* These authors contributed equally to this work.

\* [churt@tntech.edu](mailto:churt@tntech.edu)

## OPEN ACCESS

**Citation:** Hubbs NW, Hurt CR, Niedzwiecki J, Leckie B, Withers D (2022) Conservation genomics of urban populations of Streamside Salamander (*Ambystoma barbouri*). PLoS ONE 17(6): e0260178. <https://doi.org/10.1371/journal.pone.0260178>

**Editor:** Mark A. Davis, University of Illinois at Urbana-Champaign, UNITED STATES

**Received:** November 2, 2021

**Accepted:** June 14, 2022

**Published:** June 30, 2022

**Copyright:** © 2022 Hubbs et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All sequences generated for mitochondrial haplotypes and GBS sequencing are available on Genbank. The accession numbers are available in the [Supporting Information](#).

**Funding:** CH Tennessee Wildlife Resources Agency <https://twra.state.tn.us/> The funding agency did not participate in this research.

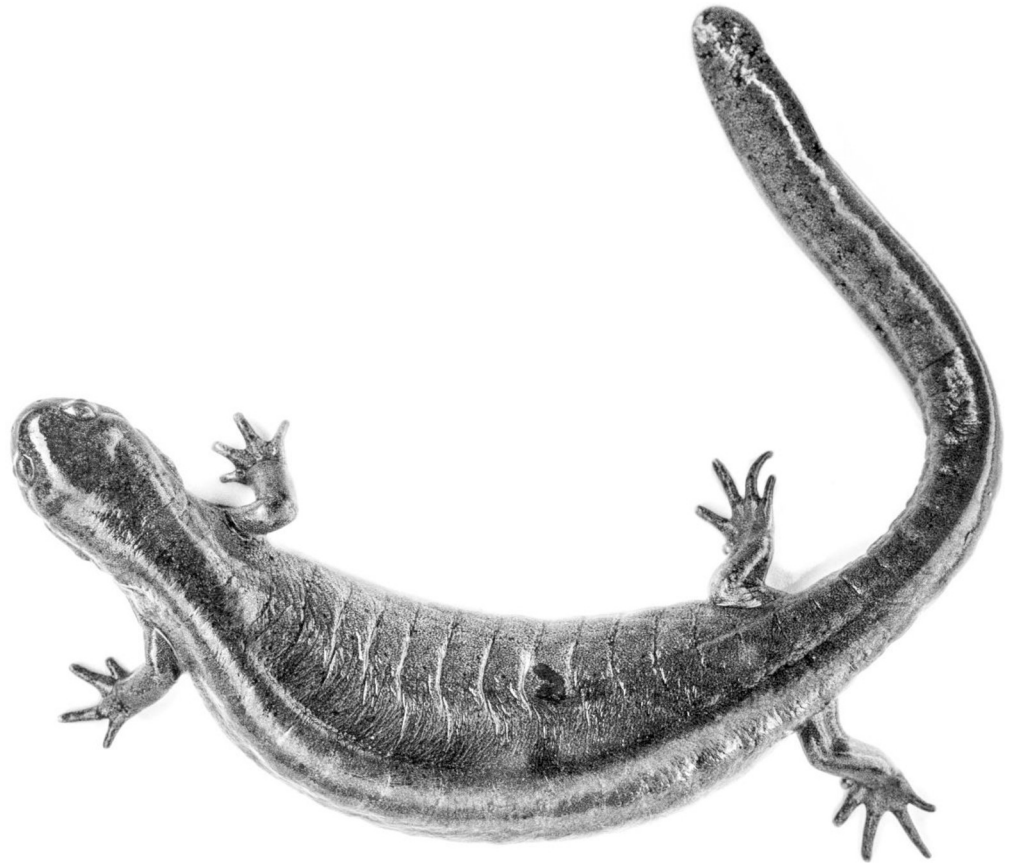
**Competing interests:** The authors have declared no competing interests.

## Abstract

In Tennessee, populations of the state endangered Streamside Salamander (*Ambystoma barbouri*) are in decline as their distribution lies mostly within rapidly developing areas in the Nashville Basin. Information regarding the partitioning of genetic variation among populations of *A. barbouri* and the taxonomic status of these populations relative to northern populations and their congener, the Small-mouthed Salamander (*A. texanum*), have important implications for management and conservation of this species. Here we combined mitochondrial sequencing and genome-wide single nucleotide polymorphism (SNP) data generated using Genotyping-by-Sequencing (GBS) to investigate patterns of genetic variation within Tennessee populations of *A. barbouri*, to assess their relationship to populations in Kentucky, and to examine their phylogenetic relationship to the closely related *A. texanum*. Results from phylogenetic reconstructions reveal a complex history of Tennessee *A. barbouri* populations with regards to northern populations, unisexual *A. barbouri*, and *A. texanum*. Patterns of mitochondrial sequence variation suggest that *A. barbouri* may have originated within Tennessee and expanded north multiple times into Kentucky, Ohio, Indiana, and West Virginia. Phylogenetic reconstructions based on genome-wide SNP data contradict results based on mitochondrial DNA and correspond to geographic and taxonomic boundaries. Variation in allele frequencies at SNP genotypes, as identified by multivariate analyses and Bayesian assignment tests, identified three evolutionary significant units (ESUs) for *A. barbouri* within Tennessee. Collectively, these results emphasize the need for prioritizing conservation needs for Tennessee populations of *A. barbouri* to ensure the long-term persistence of this species.

## Introduction

Genetic variation, population structure, and demographic history are increasingly recognized as important factors for the design of effective conservation strategies [1]. Streamside



**Fig 1. Image of Tennessee *Ambystoma barbouri*.** Photograph by David Pineros.

<https://doi.org/10.1371/journal.pone.0260178.g001>

Salamander (*Ambystoma barbouri*, Fig 1) populations in Middle Tennessee are declining due to rapid urbanization in and around the Nashville Basin and as a result, were reclassified from "deemed in need of management" to state endangered in 2018 by the Tennessee Wildlife Resources Agency (TWRA) [2–4]. Population fragmentation and loss of genetic variation that inevitably accompany loss of habitat threaten the long-term adaptive potential and persistence of *A. barbouri*, but can be mitigated by management actions aimed at maintaining genetic diversity. Information regarding *A. barbouri*'s taxonomy, population structure and patterns of genetic variation are currently needed to prioritize conservation needs and to efficiently allocate management resources. Genomic tools are increasingly being used to improve recovery and management planning in at-risk species, including salamanders, by informing taxonomic relationships, demographic histories, and biologically meaningful units of conservation that will preserve genetic diversity and the long-term adaptive potential of species [5–7].

Taxonomic uncertainties surrounding the evolutionary relationship between *A. barbouri* and its sister-species *A. texanum* (Small-mouthed Salamander) are a concern for precisely identifying targets of conservation. Adults of these two species are nearly indistinguishable based on external morphological features alone and were previously considered to be conspecific. They can be differentiated using scanning electron microscopy on the number and shape of maxillary and premaxillary teeth, larval life-history traits, and choice of breeding environment [8, 9]. In addition to their morphological ambiguity, phylogenetic reconstructions using mitochondrial and nuclear sequence data also produced conflicting results regarding the

relationship between *A. texanum* and *A. barbouri* [10]. Reconstructions based on mitochondrial sequences show that *A. texanum* and *A. barbouri* are not reciprocally monophyletic; *A. texanum* is recovered as a clade nested within *A. barbouri* from Tennessee, suggesting that pond-breeding *A. texanum* were more recently derived from a stream-breeding *A. barbouri* ancestor from Central Tennessee. However, reconstructions based on two nuclear gene sequences resulted in reciprocally monophyletic clades for the two species, consistent with a history where *A. barbouri* and *A. texanum* are older independent lineages derived from a shared common ancestor. Additional sequence data are needed to resolve unanswered phylogenetic questions as these earlier reconstructions were based on a limited number of molecular markers (two nuclear and one mitochondrial gene) and representative outgroups.

Mitochondrial evidence also informs on the origins of unisexual ambystomatid populations (family Ambystomatidae) that are common in the Great Lakes region of North America. Unisexual ambystomatids exhibit a unique reproductive mode whereby male sperm activates egg development, but contributes variable amounts of nuclear genetic material depending on compatibility with unrelated cytoplasmic DNA (termed kleptogenesis) [11]. While the nuclear genome of unisexuals is a mixture of different ambystomatid species, all known mitochondrial haplotypes across their range nest within *A. barbouri*. Mitochondrial haplotypes are most closely related to mitochondrial haplotypes found south of the Ohio River, and southwest of the Kentucky River [12]; however, the relationship of Tennessee *A. barbouri* to unisexual ambystomatids has not been explored.

At a finer scale, uncertainties also exist regarding the relationship among isolated populations of *A. barbouri* in Tennessee and discontinuous populations in Kentucky, Indiana, Ohio, and West Virginia. Phylogenetic reconstructions of *A. barbouri* populations based on mitochondrial sequence data (913 bp) from the D-loop and an adjacent intergenic spacer suggest that Tennessee *A. barbouri* populations are both ecologically and genetically distinct from more northern populations [13]. At hatching, *A. barbouri* from Tennessee were smaller and less developed than individuals from Kentucky. Also, laboratory behavioral assays show that Tennessee *A. barbouri* were similar to western Kentucky individuals, but less active than individuals from northern populations [10]. These differences were supported by genetic data; phylogenetic reconstructions recovered mitochondrial haplotypes from Tennessee as monophyletic and basal to haplotypes from Kentucky, Ohio, and West Virginia [13]. These results support an early divergence of Tennessee populations and raise questions regarding the geographic origin of *A. barbouri*. However, conclusions from this study were limited as these reconstructions are based on a single mitochondrial gene from only three individuals from a few populations within a kilometer of each other in Rutherford County.

Observations from field surveys suggest that *A. barbouri* populations in Tennessee are in decline, and the accompanying loss of genetic variation in small populations further threatens the long-term persistence of this species [14]. Maintenance of genetic variation is a fundamental priority for conservation planning and requires information regarding the partitioning of genetic variation within and between isolated populations. Data on patterns of genetic variation in Tennessee populations of *A. barbouri* are limited to a handful of mitochondrial and nuclear sequences used for phylogenetic studies [13, 10]. Genome-wide surveys of genetic variation from fine-scale population sampling across the state are critical for assessing genetic variation within populations, establishing units of conservation and maintaining historical patterns of gene flow between populations to improve the outcome of recovery efforts [15].

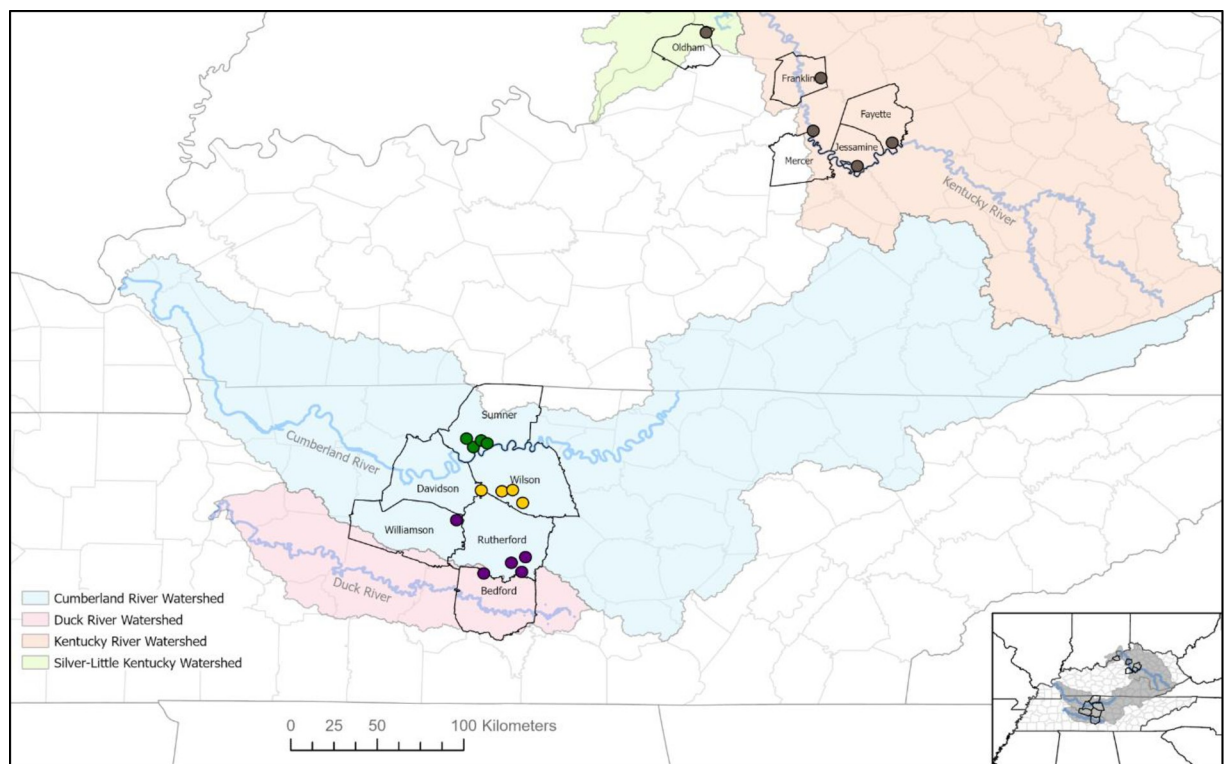
Here we investigate patterns of genetic variation within and among populations of Tennessee *A. barbouri* as well as the taxonomic relationship between *A. barbouri* and *A. texanum* in order to prioritize conservation needs and inform management practices aimed at maximizing long-term persistence of this species. Mitochondrial sequence data and genome wide SNP

genotypes were used to (1) further examine the phylogenetic relationship between *A. barbouri* and *A. texanum*, (2) review the taxonomic relationship of disjunct populations of *A. barbouri* in Tennessee, relative to northern populations, (3) evaluate patterns of genetic differentiation between geographically isolated populations of *A. barbouri* within the state of Tennessee and (4) estimate within-population genetic variation and examine demographic history of Tennessee *A. barbouri*. Taxonomic relationships, units of conservation, and geographic partitioning of genetic variation are discussed in the context of establishing conservation priorities and designing effective management strategies for this species.

## Methods

### Tissue collection and DNA extraction

Tissue samples of *A. barbouri* in the form of adult tail clips, eggs, and whole larvae were obtained from field surveys or from collaborators (Fig 2, Table 1). Whole larvae were euthanized with a solution of tricaine sulphonate (buffered MS-222) at a concentration of 3g/L. This protocol was approved by the Institutional Animal Care and Use Committee at Tennessee Tech University (protocol 17-18-006P) and scientific collection permits were obtained from Tennessee Wildlife Resources Agency (permit 1475). Historical and predicted sites of *A. barbouri* were surveyed from January 2018 to March 2018 and again from November 2018 to March 2019, coinciding with timing of oviposition as reported by Niemiller et al. [8]. At each site, *A. barbouri* adults and eggs were collected by turning over cover objects within and near seasonal streams. Additionally, pools were searched for free-swimming larvae; only one larva



**Fig 2. Map of surveyed populations of *A. barbouri* in Tennessee and Kentucky.** County lines are shown with black borders. Shaded regions indicate major watersheds as depicted in the legend. Tennessee populations are colored by ESU assignment where green circles indicate northern ESU populations, yellow circles indicate central ESU populations, and purple circles indicate southern ESU populations.

<https://doi.org/10.1371/journal.pone.0260178.g002>

**Table 1. Population IDs, sample sizes for mtDNA/GBS analyses, and map coordinates for all populations.** Population IDs for *A. barbouri* collections correspond to map locations in Fig 1. For outgroup samples, AAFB denotes Arnold Air Force Base and AKT1 denotes *A. texanum* from Arkansas. Population IDs by county are as follows: Bedford County (B6), Davidson County (D3), Rutherford County (R1, R7, and R9), Sumner County (S2, S5, S7, and S8), Wilson County (W1, W3, and W4), and Williamson County (Wil2).

Population ID (Map ID)	Watershed (HUC10)	Sample Size mtDNA	Sample Size GBS
<u><i>A. barbouri</i>—Tennessee</u>			
Sumner 2 (S2)	Lower Cumberland-Old Hickory Lake	3	–
Sumner 5 (S5)	Lower Cumberland-Old Hickory Lake	3	11
Sumner 7 (S7)	Lower Cumberland-Old Hickory Lake	5	37
Sumner 8 (S8)	Lower Cumberland-Old Hickory Lake	6	22
Davidson 3 (D3)	Stones	3	29
Wilson 1 (W1)	Stones	5	15
Wilson 3 (W3)	Stones	4	12
Wilson 4 (W4)	Stones	3	6
Williamson 2 (Wil2)	Lower Cumberland-Sycamore	3	5
Rutherford 1 (R1)	Stones	3	20
Rutherford 7 (R7)	Stones	12	28
Rutherford 9 (R9)	Stones	3	11
Bedford 6 (B6)	Upper Duck	4	8
<u><i>A. barbouri</i>—Kentucky</u>			
Kentucky FC (FC)	Lower Kentucky	3	3
Kentucky (RR)	Lower Kentucky	5	4
Kentucky SL (SL)	Sil Kentucky	4	3
Kentucky SW (SW)	Lower Kentucky	4	5
Kentucky JS (JS)	Lower Kentucky	2	2
<u>Outgroups</u>			
<i>A. mabeei</i>	—	2	–
<i>A. maculatum</i> (AAFB)	—	2	–
<i>A. talpoideum</i> (AAFB)	—	1	1
<i>A. texanum</i> (AAFB)	–	–	4
<i>A. texanum</i> (AKT1)	—	1	1

<https://doi.org/10.1371/journal.pone.0260178.t001>

was sampled from any single pool to avoid sampling related individuals from the same clutch. A total of 235 individuals were included for genetic analysis. Samples included 225 individuals from 13 populations of *A. barbouri* in Tennessee spanning six counties in the Nashville Basin: Bedford County (B6), Davidson County (D3), Rutherford County (R1, R7, and R9), Sumner County (S2, S5, S7, and S8), Wilson County (W1, W3, and W4), and Williamson County (Wil2). A total of five *A. texanum* individuals were sampled; these included four individuals from (Arnold Airforce Base, Coffee County, TN, AAFB) and one from Craighead Co., Arkansas (AKT1). Outgroups included two *A. mabeei* (North Carolina), two *A. maculatum* (AAFB), and one *A. talpoideum* (AAFB).

DNA was extracted from tail clips, eggs, and larvae using the EZNA Tissue DNA Mini kit (OMEGA BIO-TEK) following the manufacturer's protocol except that DNA was eluted in water. Approximately 1200 bp of the mitochondrial D-loop was targeted for PCR amplification using primers developed by Shaffer and McKnight [15] (Table 2). The mitochondrial D-loop has been shown to be informative for evaluating population structure and species relationships within the genus *Ambystoma* [12, 15–18]. Polymerase chain reactions (PCR) included 2 µl extracted DNA, 0.5 µM of each primer, and 1X GoTaq Master Mix (Promega) to a total volume of 20 µl. Conditions for PCR were as follows: initial denaturation step of 2 minutes at

**Table 2. Primers used for PCR amplification of the mitochondrial D-loop.** Primers THR (forward) and 651 (reverse) were used for initial amplification of entire ~1300 bp. Internal primers 007 (forward), DL3 (reverse), and DL1 (reverse) were used for Sanger sequencing (Shaffer & McKnight, 1996).

Primer Name	Sequence 5' → 3'
THR (forward)	AACATCGATCTTGTAAGTC
007 (forward)	GCACCCAAAGCAAATTCCTTG
DL3 (reverse)	TTCGATCCAATTGATGAATG
DL1 (reverse)	AATATTGATAATTCAGCTCCG
651 (reverse)	GTAAGATTAGGACCAAATCT

<https://doi.org/10.1371/journal.pone.0260178.t002>

95°C, followed by 35 cycles of 15s at 95°C, 15s at 53°C, and 90s at 72°C. This program ended with a final extension of 10 min. at 72°C. Amplified PCR product was cleaned prior to cycle sequencing by exonuclease I/shrimp alkaline phosphatase (New England Biolabs) and used for bi-directional Sanger sequencing on an ABI 3730 automated sequencer (MCLAB). Sequence chromatograms were imported and visualized using SEQUENCHER 5.2 (Gene Codes Corporation). Sequences were aligned using ClustalW Multiple Alignment option [19] (Thompson et al. 2003) as implemented in Bioedit [20].

### Phylogenetic reconstructions for mitochondrial haplotypes

Phylogenetic reconstructions were estimated for all unique mitochondrial D-loop haplotypes and an additional 53 *A. barbouri*, 83 unisexual ambystomatids, and 27 *A. texanum* sequences obtained from Genbank. Both maximum likelihood (ML) and Bayesian optimality criteria were used for phylogenetic analyses. Maximum-likelihood analyses were performed using the software RAxML [21] on the CIPRES Science Gateway [22] under the GTR+G model. Nodal support was estimated using rapid bootstrapping (1000 replicates). Bayesian phylogenetic reconstructions were performed using MrBayes 3.2.1 [23] also on the CIPRES Science Gateway. The best model of substitution was selected by Modeltest [24], implemented in MEGA X [25] using Bayesian information criterion (BIC). The Markov chain Monte Carlo (MCMC) algorithm ran for 10,000,000 generations, sampling every 1,000 generations. Two independent runs were performed and the resulting trees were combined after the deletion of a burnin (25%). Convergence of the MCMC chain was assessed by estimating the effective sample size for parameters (>200) and by visualizing parameter trace plots using the software Tracer 1.7.1 [26, 27]. A majority-rule consensus tree was generated and nodal support was estimated by posterior probabilities.

### GBS library preparation and sequencing

A total of 227 individuals were included for GBS sequencing including 201 individuals from 12 populations of *A. barbouri* in Tennessee, 17 individuals from five populations of *A. barbouri* in Kentucky, five individuals from two populations of *A. texanum*, and two sampled outgroups *A. talpoideum* (N = 1) and *A. mabeei* (N = 1; Table 1). Genotyping-by-Sequencing libraries were prepared using the restriction enzyme ApeKI following the protocol of Elshire et al. [28]. Genomic DNA was quantified using Quant-iT Picogreen dsDNA Assay Kit (Thermo Fisher Scientific), and all samples were standardized to between 5–6.5 ng/uL (50–65 ng total genomic DNA). Extracted DNA was digested with the restriction enzyme ApeKI. Adaptors containing PCR binding sites and individual barcodes were ligated onto digested DNA. Barcoded DNA was pooled and PCR amplified using primers that bind to the ligated adaptors (see Elshire et al. [28] for primer sequences). The resulting PCR products were cleaned using the Qiagen PCR purification kit and then cleaned again using the AxyPrep Mag PCR Clean-up kit

(Axygen, Big Flats, New York, USA). The distribution of the fragment size in the PCR product was determined using an Agilent 2100 BioAnalyzer (Agilent Technologies Inc., Santa Clara, CA, USA). Barcoded libraries were sequenced using the Illumina NextSeq (Illumina Inc., San Diego, CA, USA) with a 75 bp single end read chemistry.

### SNP discovery and filtering

The *Stacks* program *process\_radtags* was used to filter and demultiplex raw reads based on bar-coded sequences [29]. We used the seven-step de novo clustering pipeline *ipyrad* v. 3.5 [30] to generate and filter SNP datasets used in downstream analyses. Quality filtering of raw sequence reads converted bases with Phred scores <33 to Ns, reads with more than 5 Ns were removed. Reads were clustered using a sequence similarity threshold of 90% both within and between sampled individuals, with a minimum read depth of six. Individuals with fewer than 500,000 reads were excluded from downstream analyses. Loci with observed heterozygosity ( $H_o$ ) greater than 0.5 were removed to filter out possible paralogs. The final SNP dataset was then filtered to remove loci deviating from Hardy-Weinberg equilibrium ( $p < 0.05$ ), loci genotyped in less than 60% of individuals, and SNPs with a minor allele frequency (MAF) less than 0.01. Only one SNP was retained per sequenced tag.

### Population genetic variation and effective population size

Estimates of within population genetic variation for 12 Tennessee *A. barbouri* populations were estimated using the R-package *DiveRsity* v. 1.9.9 [31]. Summary statistics were estimated separately for each population and included the proportion of SNPs that were polymorphic within each population ( $P$ ), allelic richness ( $A_R$ ), and observed and expected heterozygosity ( $H_o$  and  $H_e$ ). Estimates of  $P$ ,  $A_R$ ,  $H_o$ , and  $H_e$  were estimated for the entire SNP dataset and for the subset of SNPs that were polymorphic within each population.

Effective population size ( $N_E$ ) was estimated for Tennessee *A. barbouri* populations using both the linkage disequilibrium (LD) method and the heterozygote-excess method [32] as implemented in the program *N\_EEstimator* v2 [33]. Populations with sample sizes of less than 20 individuals were excluded from analyses as these datasets did not provide enough signal for reliable  $N_E$  estimation [34]. The minor allele frequency (MAF) parameter was set at 0.01 and 95% confidence intervals were estimated using the parametric chi-squared method.

**Population structure.** Population pairwise  $F_{ST}$  values [35] and  $G''_{ST}$  [36] were estimated for the 12 Tennessee *A. barbouri* populations using the R package *diveRsity* [31]. Hierarchical partitioning of genetic variation across Tennessee populations of *A. barbouri* was examined using Analysis of Molecular Variances (AMOVA) as implemented in *Arlequin* 3.5 [37, 38]. Results from phylogenetic reconstructions based on mitochondrial D-loop haplotypes were used to generate hypotheses regarding higher-level structuring of populations. Significance of variance components was determined using 1000 permutations.

The optimal number of genetic clusters ( $K$ ) based on genomic SNPs was estimated using both a multivariate approach and a Bayesian-based assignment method. Discriminant Analysis of Principal Component (DAPC) was performed using the 'adegenet' package in *R* (v. 2.1.4) [39, 40]. The optimal number of principal components (PCs) was determined by cross-validation using the 'xvalDapc' function (1000 replicates) and the number of PCs with the highest assignment success was retained for DAPC analyses. The *find.clusters* function was first used to identify the optimal value of  $K$  based on a Bayesian Information Criterion (BIC) process. The optimal  $K$  was then used to perform a DAPC analysis to describe the relationship between the genetic clusters. Individual membership probabilities were assessed using the function *compoplot*. DAPC analysis was performed on three hierarchical datasets; these included 1) 12

populations of *A. barbouri* in Tennessee, 2) all sampled Tennessee and Kentucky *A. barbouri* populations, and 3) all *A. barbouri* populations and four *A. texanum* individuals from AAFB.

Bayesian assignment tests were performed on Tennessee *A. barbouri* populations using the program STRUCTURE 2.3.4 [41]. Kentucky *A. barbouri* populations were not included in assignment analyses due to small sample sizes at these sites. Values of  $K$  (number of populations) ranged from 1 to 12 populations with 20 replicate runs per value of  $K$ ; MCMC simulations were performed for a burn-in of 250,000 iterations and an additional 1,000,000 iterations were retained for the final analysis. Results were summarized using the software package CLUMPAK [42]. The optimal number of groups ( $K$ ) was determined using the Delta- $K$  method [43] as implemented in STRUCTURE HARVESTER v0.6.94 application [44]. The Distruct (1.1) application in CLUMPAK was used to produce the final barplots.

### Multispecies coalescent using BPP

A multispecies coalescent (MSC) model was used to examine taxonomic boundaries, estimate species trees, and estimate population divergence times ( $t$  = years before present) as implemented in the software BPP v. 4.4 [45]. Input datasets were formatted as full-length sequence alignments from our GBS sequencing and included all *A. barbouri* and *A. texanum* populations. Loci represented by fewer than 100 individuals were removed from the analysis. Joint species-delimitation and species-tree reconstructions were performed using the A11 model, where individuals were assigned to sampled populations. Run parameters were as follows: uniform rooted trees were used as the species model prior, the theta prior ( $\theta = 4N_E\mu$ ) was assigned a gamma distribution with  $\alpha = 3$  and  $\beta = 0.04$  and the prior for tau ( $\tau = t\mu$ ) was gamma distributed with  $\alpha = 3$  and  $\beta = 0.2$ . Three independent MCMC runs were implemented, each with a burn-in of 5,000, a sample frequency of 10 and a total of 50,000 iterations. Species delimitation and species-tree topologies with the highest posterior probabilities from the A11 runs were used as input for the A00 analyses. The A00 model estimates divergence time parameters and long-term  $N_E$ , assuming fixed species assignments and a fixed species tree. The priors for A00 were the same as priors used for the A11 model. Three independent runs were also performed for the A00 analysis, each with a burn-in of 5,000, a sample frequency of 10 and a total of 50,000 iterations. MCMC chains were pooled and absolute estimates of  $\tau$  were converted to years before present (YBP) in the BPPR using the genome wide mutation rate estimate for the vertebrate nuclear genome where  $\mu = 1.21 \times 10^{-9}$  [46, 47].

## Results

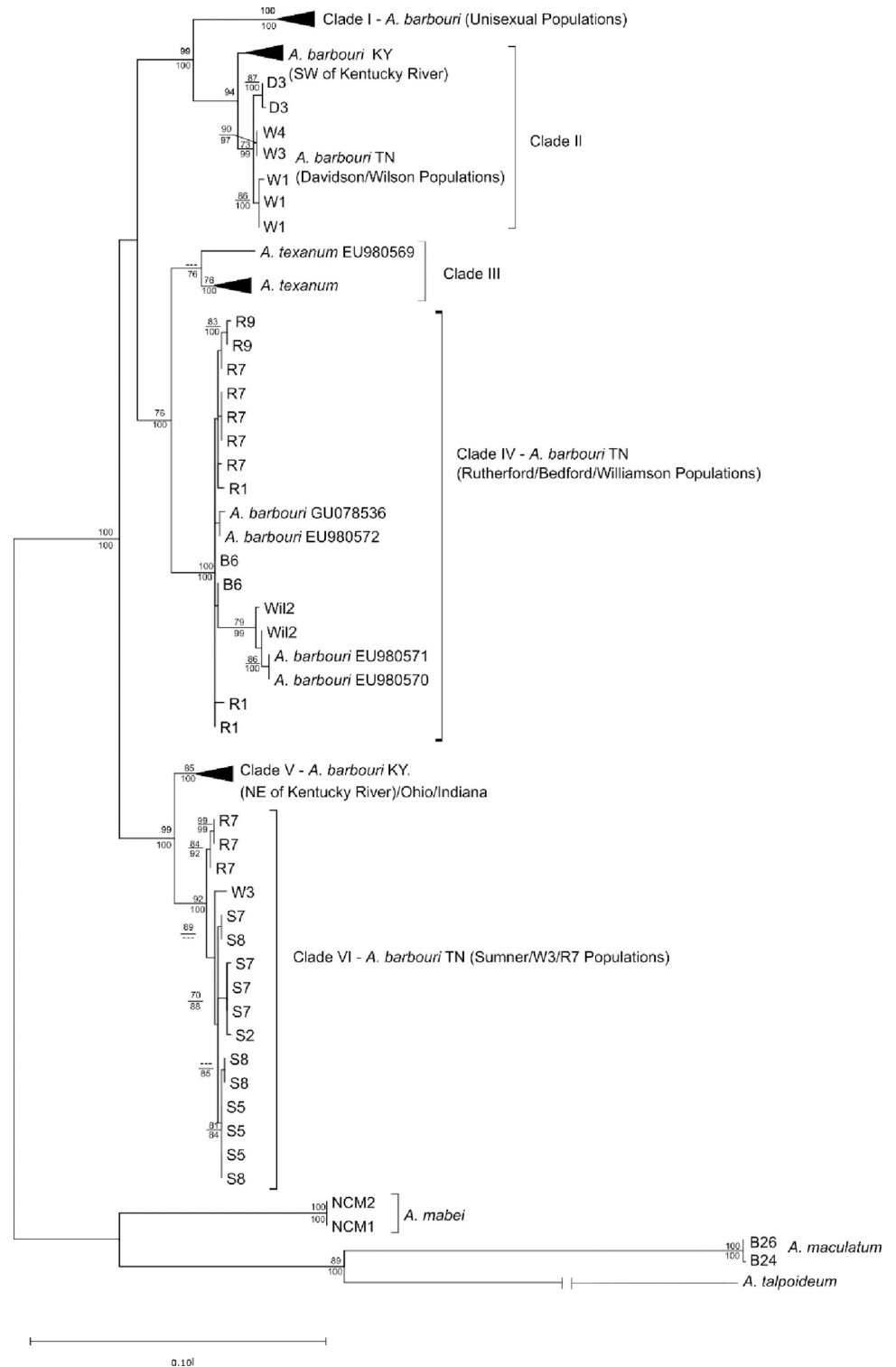
### Mitochondrial sequencing

Sanger sequencing of the mitochondrial D-loop resulted in an ~1100 bp sequence alignment in the 81 individuals selected for sequencing. A total of 60 unique haplotypes were identified from the newly sequenced individuals, 39 haplotypes were identified from *A. barbouri* populations sampled in Tennessee (See S1 Table for GenBank accessions), and 15 haplotypes were identified from populations sampled in Kentucky. An additional six haplotypes were sequenced from outgroups. The final alignment contained 126 variable sites (excluding outgroups); of these, 120 sites were parsimony informative. No haplotypes were shared across multiple populations.

### Phylogenetic reconstructions

Both Bayesian and ML phylogenetic reconstructions of mitochondrial D-loop haplotypes identified six major clades across *A. barbouri* and *A. texanum* haplotypes (Fig 3). The Clade I





**Fig 3. Phylogenetic tree of *A. barbouri* mitochondrial haplotypes.** Maximum likelihood phylogenetic reconstructions of unique mitochondrial D-loop haplotypes from *Ambystoma barbouri* and *A. texanum* under a GTRCAT model of evolution as performed by RaxML. Bootstrap support values above branches are shown for nodes with 70% support or greater. Values below nodes indicate posterior probabilities from Bayesian reconstructions under a T92+G model of sequence evolution as performed by MrBayes. Asterisks denote accession numbers for sequences downloaded from Genbank.

<https://doi.org/10.1371/journal.pone.0260178.g003>

included all haplotypes belonging to the unisexual lineage of *A. barbouri* with 100% posterior probability (pp) and 100% bootstrap support (bs) for Bayesian and ML analysis, respectively. In *Clade II*, *A. barbouri* from Wilson Co. (populations W1, W3, and W4) and Davidson Co. (population D3) formed a monophyletic clade that also includes Kentucky *A. barbouri* populations from southwest of the Kentucky River (pp = 100, bs = 95%). *Clade I* and *Clade II* were recovered as sister clades in both analyses with strong support (pp = 100, bs = 99%). *Clade III* included all *A. texanum* individuals (pp = 99, bs = 71) with the exception of a single *A. texanum* haplotype obtained from Genbank (ID EU980569, collected from Lawrence KS); this sequence is recovered as basal to *Clade III* in both ML and Bayesian analyses. *Clade IV* groups with *Clade III* and included Tennessee *A. barbouri* individuals from Rutherford Co. (populations R1, R7, and R9), Bedford Co. (B6), and Williamson Co.; pp = 100, bs = 71%). *Clade V* included all Kentucky individuals from populations northeast of the Kentucky River, Ohio, and Indiana (pp = 100, bs = 86%). Finally, *Clade VI* is sister to *Clade V* and included all Tennessee *A. barbouri* from Sumner Co. (populations S2, S5, S7, and S8), a single haplotype from Wilson population W3, and three haplotypes sequenced from Rutherford population R7 (pp = 100, bs = 99%). On a larger scale, all *A. barbouri* and *A. texanum* are part of the same well-supported monophyletic group in both analyses (pp = 100, bs = 100%), such that all *A. texanum* haplotypes are nested within *A. barbouri*.

### GBS sequencing

The average number of retained sequence reads per individual generated from sequencing of GBS libraries was  $5.86 \times 10^6 \pm 5.9 \times 10^6$ ; 220 individuals were retained after removal of low coverage individuals (defined as  $< 500,000$  total reads). A total of  $1.56 \times 10^6$  loci were recovered from the de novo assembly in ipyrad and 440,629 loci were retained after filtering. Additional filtering for SNPs (i.e. filtering for low representation, HWE, MAF, and max heterozygosity) recovered 1,169 SNPs for the dataset that included only Tennessee *A. barbouri*, 516 SNPs in the dataset that included all sampled *A. barbouri* populations in Tennessee and Kentucky, and 500 SNPs in the dataset that included all *A. barbouri* and *A. texanum* populations.

### Within Population Genetic Variation

Much of the genetic variation within our SNP dataset was in the form of fixed differences between populations (Table 3). The proportion of SNPs that were polymorphic within populations ranged from 0.045 (W3 and Wil2) to 0.267 (S7). When only polymorphic loci were considered, the Will2 population had the highest estimates of  $H_o$  and  $H_e$ ; however, this population had a low proportion of polymorphic SNPs overall. Rutherford County populations (R1 and R7) had the lowest estimates of  $H_o$  and  $H_e$ , when considering only polymorphic loci and for the entire SNP dataset.

### Effective population size

Estimates of effective population size ( $N_E$ ) were generated for five populations with adequate sampling ( $N \geq 20$ ), including S7, S8, R1, R7, and D3 (Table 4). Results from LD analyses ranged from  $N_E = 15$  (R1) to  $N_E = 108$  (108) (S8); the upper confidence interval for S8 was  $\infty$ , indicating low signal for this dataset. Results from the heterozygote excess method were inconclusive ( $\infty$ ) for two of the five analyzed populations. Estimates from the remaining three populations ranged from 24 (D3) to 406 (S7); however, CIs included  $\infty$  for two of these estimates (S7 and S8).

**Table 3. Standard measures of genetic diversity for 12 populations of *Ambystoma barbouri* in Tennessee based 586 SNP loci.** Summary statistics include the average number of individuals genotyped per locus ( $N$ ) and the proportion of SNPs that were polymorphic within each population ( $P$ ). Allelic richness ( $A_R$ ), observed heterozygosity ( $H_O$ ), and expected heterozygosity ( $H_e$ ) were calculated for all loci and again for only those loci that were polymorphic within each population.

Population	N	P	All Loci			Polymorphic Loci		
			$A_R$	$H_O$	$H_e$	$A_R$	$H_O$	$H_e$
Sumner 5	9.58	0.108	1.042	0.018	0.020	1.460	0.170	0.154
Sumner 7	22.78	0.267	1.036	0.027	0.025	1.204	0.102	0.092
Sumner 8	12.97	0.185	1.020	0.025	0.025	1.259	0.136	0.127
Davidson 3	22.02	0.192	1.041	0.018	0.016	1.243	0.094	0.083
Wilson 1	7.49	0.112	0.986	0.019	0.023	1.395	0.173	0.163
Wilson 3	7.07	0.045	1.032	0.017	0.016	1.514	0.217	0.184
Wilson 4	6.79	0.089	1.049	0.019	0.020	1.602	0.210	0.220
Williamson 2	4.11	0.045	1.000	0.012	0.026	1.690	0.257	0.229
Rutherford 1	16.71	0.092	1.026	0.009	0.008	1.303	0.096	0.091
Rutherford 7	23.57	0.140	1.032	0.011	0.011	1.240	0.076	0.079
Rutherford 9	7.74	0.056	1.022	0.009	0.009	1.470	0.160	0.159
Bedford 6	8.74	0.060	1.026	0.010	0.009	1.467	0.167	0.142
Average Total	12.46	0.116	1.026	0.016	0.017	1.404	0.155	0.144

<https://doi.org/10.1371/journal.pone.0260178.t003>

## Population structure

Pairwise  $F_{ST}$  estimates for *A. barbouri* populations within Tennessee were geographically structured in a hierarchical fashion (Table 5). Overall,  $F_{ST}$  values averaged  $0.356 \pm 0.229$  and ranged from no differentiation ( $F_{ST} = 0$ ) to highly differentiated ( $F_{ST} = 0.703$  for S5/R1). In general, populations north of the Cumberland River (S2, S5, S7, and S8) showed evidence of long-term isolation from populations south of the Cumberland River. Pairwise  $F_{ST}$  estimates comparing populations north and south of the Cumberland River averaged  $0.606 \pm 0.047$ , while  $F_{ST}$  estimates among Tennessee populations north of the Cumberland River averaged only  $0.078 \pm 0.041$ . Populations south of the Cumberland were further structured; populations in Wilson and Davidson Counties (D3, W1, W3, and W4) were diverged from populations further south in Williamson, Rutherford, and Bedford counties. Estimates of pairwise  $F_{ST}$  comparing populations within Wilson/Davidson averaged  $0.104 \pm 0.035$  and within Williamson, Rutherford and Bedford Counties averaged  $0.060 \pm 0.060$ , while between group pairwise  $F_{ST}$  values were much higher, averaging  $0.284 \pm 0.060$ . Pairwise  $G^{*ST}$  were very similar to  $F_{ST}$  estimates and demonstrated the same hierarchical pattern.

The three monophyletic clades of Tennessee *A. barbouri* recovered from mitochondrial phylogenetic reconstructions were further examined using genome-wide SNP genotypes in an AMOVA framework (Table 6). Most variance in SNP genotypes could be attributed to differences between the three clades (63.64%). Differences between populations within clades only

**Table 4. Estimates of effective population sizes ( $N_E$ ) for five populations of *Ambystoma barbouri* in Tennessee based on the Linkage Disequilibrium (LD) method and the heterozygosity excess method as performed by  $N_e$  Estimator.** The 95% confidence intervals were estimated by the parametric chi-squared method.

Population	N	Linkage Disequilibrium			Heterozygote Excess		
		$N_E$	CI Lower	CI Upper	$N_E$	CI Lower	CI Upper
Sumner 7	35	84.5	51.5	207.2	406.4	20.7	$\infty$
Sumner 8	21	107.7	28.8	$\infty$	192.7	18.8	$\infty$
Davidson 3	29	52.1	33.2	106.4	24.0	13.0	182.8
Rutherford 1	20	14.8	8.7	30.1	$\infty$	14.6	$\infty$
Rutherford 7	28	30.7	20.9	51.6	$\infty$	22.8	$\infty$

<https://doi.org/10.1371/journal.pone.0260178.t004>

**Table 5. Population pairwise  $F_{ST}$  estimates (above diagonal) and  $G''_{ST}$  (below diagonal) averaged across 584 SNP loci for 12 populations of *Ambystoma barbouri* in Tennessee.**

	Populations	1	2	3	4	5	6	7	8	9	10	11	12
1.	Sumner 5	–	0.11	0.017	0.619	0.631	0.621	0.547	0.637	0.675	0.652	0.672	0.673
2.	Sumner 7	0.105	–	0.107	0.585	0.598	0.586	0.52	0.593	0.632	0.61	0.627	0.630
3.	Sumner 8	0.019	0.109	–	0.581	0.589	0.579	0.519	0.588	0.627	0.604	0.623	0.626
4.	Davidson 3	0.621	0.592	0.602	–	0.161	0.054	0.093	0.201	0.29	0.253	0.296	0.300
5.	Wilson 1	0.634	0.588	0.58	0.167	–	0.089	0.138	0.247	0.349	0.239	0.332	0.357
6.	Wilson 3	0.626	0.567	0.572	0.06	0.09	–	0.081	0.244	0.332	0.265	0.328	0.342
7.	Wilson 4	0.553	0.515	0.521	0.104	0.126	0.077	–	0.171	0.269	0.192	0.262	0.276
8.	Williamson 2	0.623	0.540	0.552	0.185	0.223	0.244	0.153	–	0.017	0.057	0.034	0.000
9.	Rutherford 1	0.703	0.628	0.656	0.283	0.387	0.365	0.317	0.024	–	0.116	0.022	0.04
10.	Rutherford 7	0.679	0.620	0.643	0.252	0.273	0.285	0.223	0.058	0.113	–	0.106	0.148
11.	Rutherford 9	0.671	0.589	0.610	0.272	0.333	0.338	0.27	0.033	0.023	0.099	–	0.075
12.	Bedford 6	0.677	0.592	0.618	0.274	0.362	0.351	0.286	0.000	0.041	0.137	0.073	–

<https://doi.org/10.1371/journal.pone.0260178.t005>

accounted for 2.09% of the total molecular variance. Differences between individuals accounted for the remaining 34.27% of variance in SNP genotypes.

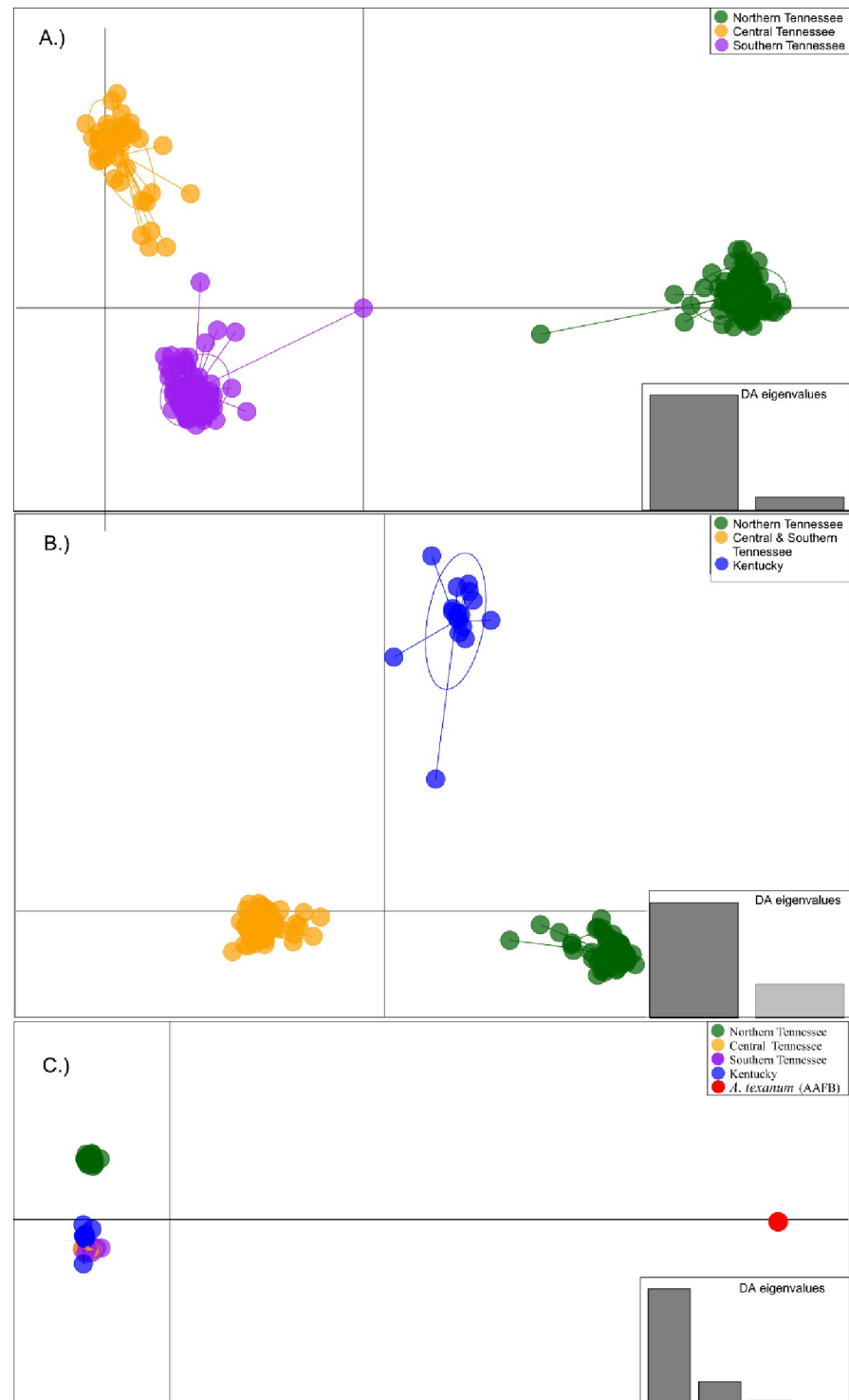
Results of BIC for DAPC analysis suggested that three genetic clusters ( $K = 3$ ) best represented genetic variation among populations sampled within Tennessee (Fig 4A, S1 Fig). Membership of each cluster was geographically partitioned and consisted of a northern cluster that included populations from Sumner County, a central cluster comprised of individuals from Wilson, and Davidson counties, and a southern cluster that represented Bedford, Rutherford, and Williamson County populations. Membership probabilities and DAPC plots indicated some degree of admixture between the central and southern clusters, but no admixture within the northern cluster (S1 Fig). A separate analysis that included Tennessee *A. barbouri* and the five sampled populations from Kentucky also resulted in three genetic clusters (Fig 4B). Kentucky populations of *A. barbouri* were isolated from populations in Tennessee. Central and southern populations were assigned to the same cluster and populations in Sumner county were assigned to the third cluster. When *A. texanum* individuals were included in a third analysis, results of the BIC analysis indicated the number of genetic clusters was five (Fig 4C). Tennessee populations were assigned to northern, central, and southern clusters and scatter plots showed Tennessee populations to be tightly grouped with the fourth cluster that included all Kentucky populations. All *A. texanum* individuals formed the fifth cluster that was well separated from all *A. barbouri* individuals.

Analysis of Bayesian assignment tests for Tennessee populations of *A. barbouri* were consistent with results from DAPC analysis that suggested the presence of three geographically partitioned genetic clusters within the state (Fig 5). The optimal value of  $K$ , as determined by the Delta  $K$  method, indicated four genetic clusters; however, increasing  $K$  from 2 to 3 did not

**Table 6. Results of hierarchical analyses of molecular variation (AMOVA) for the SNP dataset from 12 populations of *A. barbouri* in Tennessee.** Assignment to mitochondrial clades are as follows: Clade II (S5, S7, and S8), Clade IV (D3, W1, W3, and W4), and Clade III (Wil2, R1, R7, R9, B6). Asterisks indicate significance of  $\Phi$  statistics based on 1000 permutations in Arlequin.

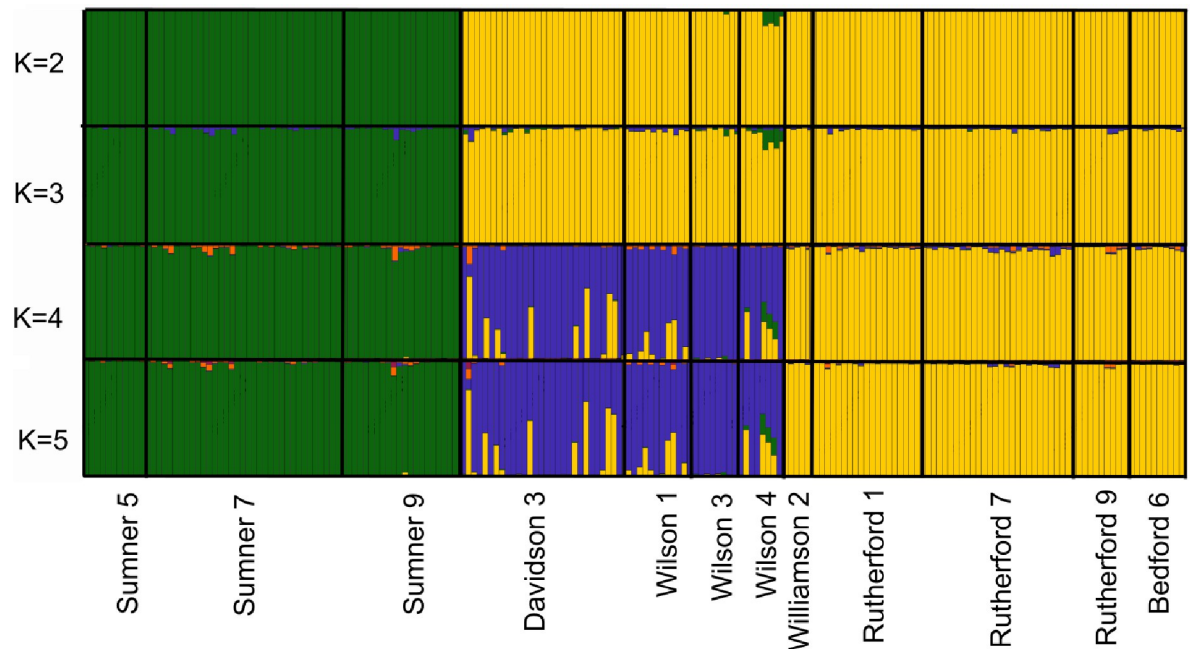
Source of variation	DF	Sum of squares	Variance components	Percentage of variation	$\Phi$ -Statistics
Among drainages	2	614.5	2.30	63.60	$\phi_{CT} = 0.636^*$
Among populations within drainages	9	31.24	0.08	2.09	$\phi_{SC} = 0.058^*$
Within populations	384	476.24	1.24	44.23	$\phi_{ST} = 0.657^*$
Total	395	1,121.97	3.62		

<https://doi.org/10.1371/journal.pone.0260178.t006>



**Fig 4. Scatter plot from discriminant analysis of principal components (DAPC).** Scatter plots of PCA 1 and PCA 2 from DAPC analysis of SNP genotypes for A) all Tennessee *A. barbouri* ( $N = 198$ ,  $K = 3$ ), B) all Tennessee *A. barbouri* and samples from five populations of *A. barbouri* in Kentucky ( $N = 215$ ,  $K = 3$ ) and C) all *A. barbouri* and *A. texanum* ( $N = 4$ ,  $K = 5$ ). The optimal number of clusters for each analysis was determined using the Bayesian Information Criteria (BIC). The final number of principal components used in each analysis was determined using the cross-validation method.

<https://doi.org/10.1371/journal.pone.0260178.g004>



**Fig 5. Bayesian assignment barplots.** Results of Bayesian Assignment tests based on SNPs generated from GBS sequencing. Barplots indicate individual assignment probabilities for samples from 12 *A. barbouri* populations in Tennessee ( $K = 2-5$ ).

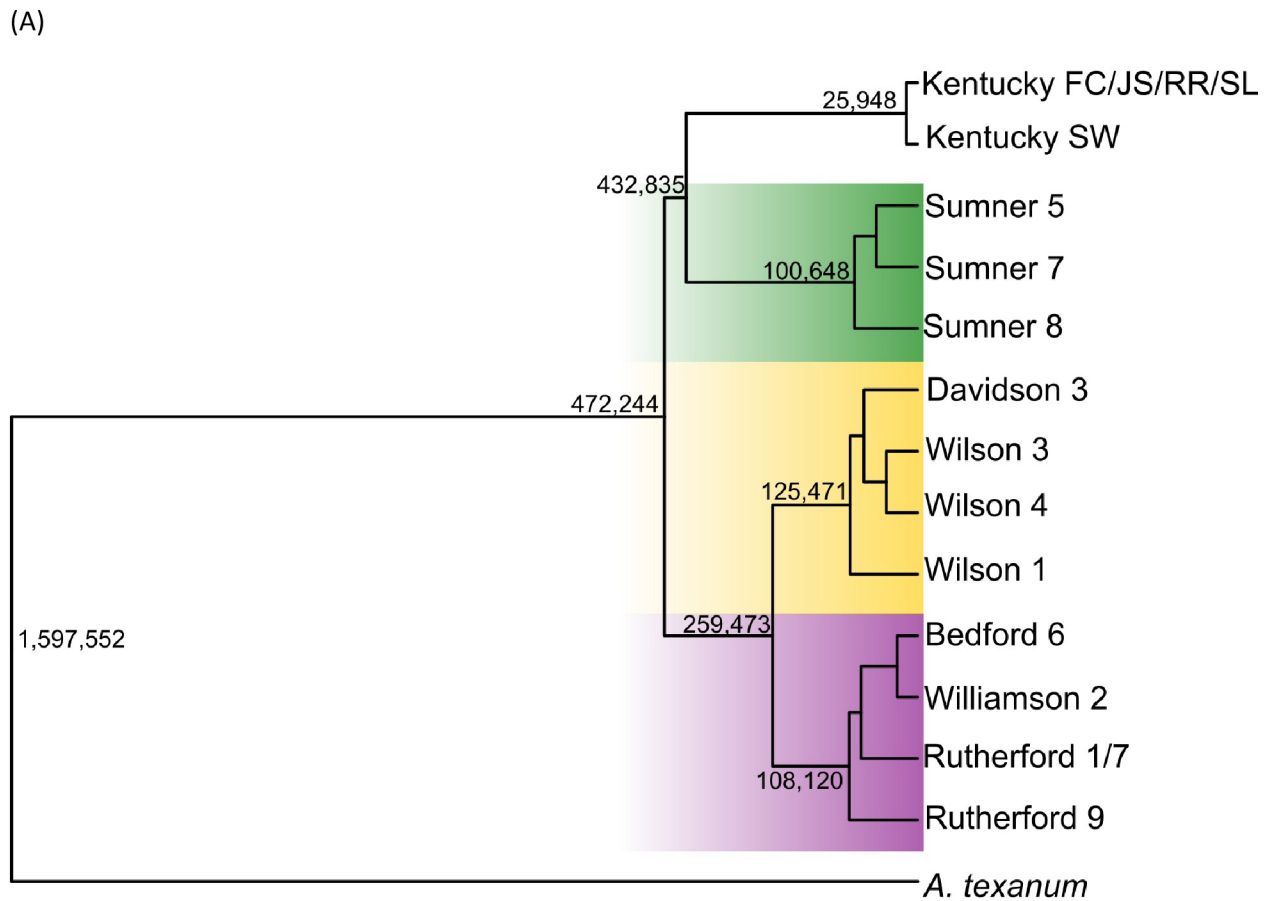
<https://doi.org/10.1371/journal.pone.0260178.g005>

change population assignments. Assignment plots for  $K = 2-3$  separated northern populations in Sumner County from central and southern populations. When  $K$  was increased to four, populations from Davidson and Wilson Counties were separated from southern populations in Bedford, Rutherford and Williamson Counties. Increasing the value of  $K$  to 5 increased the level of admixture but did not result in geographically meaningful partitions.

### Multispecies coalescent analyses

A total of 988 loci were retained for the MSC analyses performed by BPP. All three A11 runs recovered the same taxonomic groups and species tree topology with the highest posterior probability. Out of 18 sampled populations, 14 populations were genetically distinct. Within Tennessee, *A. texanum* and all *A. barbouri* populations, with the exception of R1-R7, were identified as distinct groups (Fig 6). Kentucky *A. barbouri* populations were split into two groups; populations FC, JC, RR, and SL were grouped together and the SW population was its own genetic group. The separation of individual populations in this analysis does not indicate that each population warrants recognition at the species-level as simulations have shown that BPP will split groups at the population-level when many loci are analyzed [48].

The A11 species tree reconstruction with the highest posterior identified the same Tennessee clades recovered in our mitochondrial gene trees with the exception of the placement of *A. texanum*. In the species tree, *A. texanum* was basal to all populations of *A. barbouri* in Tennessee and Kentucky. The estimated divergence time (A00 analysis) between *A. texanum* and *A. barbouri* was 1.6 million YBP. All Sumner *A. barbouri* formed a monophyletic group with Kentucky populations as a monophyletic sister group. Divergence time estimates indicated that Sumner populations and Kentucky populations shared a common ancestor < 500,000 YBP. All *A. barbouri* populations south of the Cumberland River were monophyletic (TMRCA 260,000 YBP) and were further split into two groups that corresponded to clades



(B)

Node	TMRCA (YBP)	CI (95%)	
<i>A. texanum</i> , <i>A. barbouri</i>	1,597,552	1,169,671	2,062,859
<i>A. barbouri</i> (all)	472,244	313,795	639,208
Kentucky (all), Sumner (all)	472,244	313,795	639,208
Kentucky (all)	24,982	13,432	41,552
Sumner (all)	100,648	45,292	162,312
Davidson, Wilson, Williamson, Bedford, Rutherford	259,743	175,719	350,108
Davidson, Wilson	125,471	75,781	195,323
Bedford, Williamson, Rutherford	108,120	62,980	164,210

**Fig 6. Multispecies coalescent analysis of SNP genotypes.** Species-tree (A) and divergence time estimates (B) from Bayesian multispecies coalescent method as implemented in the software BPP. Numbers at nodes also indicate divergence time estimates (t). Divergence time parameter  $\tau$  was converted to time in years before present (YBP) in the BPPR statistical package in R-studio. Shading in tree indicates ESU assignments as follows: Northern ESU (green), Central ESU (Yellow) and Southern ESU (Purple).

<https://doi.org/10.1371/journal.pone.0260178.g006>

recovered in the mitochondrial gene tree. Davidson and Wilson populations were recovered together in the first group (mitochondrial Clade II) and Rutherford, Bedford, and Williamson populations (mitochondrial Clade IV) formed the second group.

## Discussion

Patterns of genomic variation and taxonomic relationships identified here have important implications for developing management strategies aimed at the long-term conservation of *A. barbouri*. Despite the documented decline of *A. barbouri* populations in the Nashville Basin, results from GBS-derived SNP genotyping indicate that estimates of genetic variation in extant populations of Tennessee *A. barbouri* are similar to SNP-based datasets examined from other ambystomatids [34, 49]. Partitioning of genetic variation between Tennessee populations suggests that hydrogeography of the Nashville Basin has shaped patterns of gene flow. Both mtDNA and genomic SNP genotypes showed similar patterns with respect to population structure within the state of Tennessee that should be used to inform the designation of conservation units. Our results also reveal a complex history of Tennessee *A. barbouri* with more northern populations and with its sister species *A. texanum*. Despite marked differences in phylogenetic reconstructions based on mtDNA sequencing and nuclear genotypes, collectively these results indicate that Tennessee is genetically unique from northern *A. barbouri* and *A. texanum*.

### Genetic variation within populations

Loss of genetic variation resulting from rapid population decline threatens the long-term success of conservation efforts due to loss of adaptive potential and fixation of deleterious alleles [50]. For *A. barbouri* populations in the Nashville Basin, maintenance of adaptive genetic variation is particularly critical as these populations are faced with habitat alteration by urbanization in addition to climate change [51]. Evaluating the genetic health of populations is valuable for management and conservation planning; however, the interpretation of genetic variation estimates from SNP data is challenging due to the limited number of comparable studies utilizing reduced-representation methods in salamanders. Heterozygosity estimates obtained here are similar to estimates obtained in the handful of published SNP-based genetic studies. In a recent survey of SNPs from ddRAD sequencing, estimates of  $H_O$  ranged from 0.165 to 0.269 for populations of the Mole Salamander (*A. talpoideum*) and were slightly higher for populations of the Marbled Salamander (*A. opacum*; 0.207–0.298) [37]. Another SNP-based survey examined the effects of land use on genetic variation in the Northern Two-lined Salamander (*Eurycea bislineata*). Fusco et al. [52] reported nearly identical estimates of  $H_O$  from urban, suburban and rural salamander populations ( $H_O = 0.265, 0.278$  and  $0.275$ , respectively) and concluded that genetic variation had been maintained, despite habitat disturbance. However, estimates of heterozygosity may not be directly comparable across independent studies as filtering parameters in bioinformatic pipelines can influence population genetic summary statistics. Stringent filtering criteria may preferentially retain loci in conserved regions of the genome and downwardly bias estimates of genetic variation [53]. There have been previous population genetic studies in *A. texanum* and *A. barbouri* populations based on microsatellite markers. Micheletti and Storfer [54] estimated genetic variation at 11 microsatellite loci in 76 populations of *A. barbouri* distributed throughout Kentucky, Ohio, and Indiana; the average  $H_e$  from their study ranged from 0.67–0.81, respectively. Few conclusions can be made by comparing results from these prior microsatellites studies to our results based on SNP genotypes; microsatellites are known to have a much higher rate of mutation than nucleotide substitutions, increasing the expected amount of allelic diversity for a given population size [55].

### Effective population size

A reduction in estimates of contemporary effective population size ( $N_E$ ) may also signal a decline in the genetic health of at-risk populations. Our estimates of  $N_E$  were low ( $N_E$  from the



LD method averaged 58.0 across five populations), but were not unusual for LD-derived estimates of  $N_E$  in salamanders. Published estimates of  $N_E$  reported for amphibians have typically been under 100 [49, 56, 57]. Effective population size estimates reported here were similar to values reported for the congeneric endangered California Tiger Salamander (*A. californiense*;  $N_E$  values of 11–64) [58], Leora's Stream Salamander (*A. leorae*;  $N_E$  values of 17–45) [59] and for the Long-toed Salamander (*A. macrodactylum*;  $N_E = 23$ –2070 [60]). In *A. macrodactylum*, Savage et al. [61] estimated  $N_E$  for 47 breeding populations and more than half of these estimates were less than 50. Life history factors may provide some explanation for low  $N_E$  in this group as salamanders often exhibit high variance in reproductive success and larval survival within populations. Variance in reproductive success will increase relatedness among individuals in a population, inflating linkage disequilibrium across loci and reducing  $N_E$ .

Obtaining robust estimates of  $N_E$  and other demographic parameters in at-risk species can be challenging due to the sample sizes required for LD detection. Many threatened taxa are inherently rare and prohibitively difficult to sample in large numbers. For robust estimation of  $N_E$ , it is recommended that sample sizes be greater than 30 for most systems [62]. When the number of sampled individuals is much smaller than the effective size, LD-based  $N_E$  estimates can be downwardly biased and confidence intervals can be large (or infinite) due to inadequate signal in the dataset [63]. We limited our  $N_E$  analysis to populations with sample sizes  $> 20$  individuals, which left only five populations for  $N_E$  estimation. The infinite upper confidence interval for our  $N_E$  estimate in population Sumner 8 was likely due to an inadequate sample size for this population ( $N = 21$ ).

## Phylogenetic reconstructions and population structure

Mitochondrial gene-tree reconstructions suggest a complex biogeographic history for *A. barbouri* in Tennessee that involves populations of unisexual ambystomatids, *A. texanum*, and populations of *A. barbouri* from the northern core region (Kentucky, Indiana, and Ohio). Our results support findings by Bogart et al. [12] demonstrating a monophyletic relationship of unisexual ambystomatid mtDNA haplotypes nested within *A. barbouri*. The phylogenetic reconstruction shown here further details this relationship whereby unisexual ambystomatids likely shared a maternal common ancestor with *A. barbouri* populations in Kentucky and populations from Middle Tennessee. This pattern, coupled with phylogenies based on nuclear genes, points to a history of repeated hybridization events initiated by an *A. barbouri* maternal ancestor in the southern portion of its range.

With regards to their mitochondrial lineage, *A. texanum* is nested within present-day *A. barbouri*, rendering *A. barbouri* paraphyletic. This phylogenetic pattern contradicts the scenario proposed by Kraus and Petranka [9], that stream-dwelling *A. barbouri* descended from the more widespread pond-breeding *A. texanum*. Species-tree reconstructions based on genome-wide SNP data are incongruent with relationships recovered from the mitochondrial trees. This species-tree topology adheres to conventional taxonomic and geographical boundaries (Fig 6), where *A. texanum* is a sister-species to all *A. barbouri* populations. The TMRCA for *A. barbouri* and *A. texanum* based on SNP genotypes was estimated at  $\sim 1.6$  million YBP making this a relatively recent split between these ecologically distinct species [64]. Incongruence between mitochondrial gene trees and reconstruction based on nuclear data are not uncommon and are often attributed to ancestral lineage sorting, introgression, and/or sex-biased dispersal [65, 66]. Determining the cause of mitonuclear discordance would require additional sampling across the distributions of both *A. barbouri* and *A. texanum*; however, it is relevant to note that recent studies have demonstrated that hybridization is common across salamander lineages (including the genus *Ambystoma*) and may facilitate rapid diversification [67].

The hydrogeography of the Nashville Basin, together with cyclic glacial movements during the Pleistocene, may have shaped contemporary patterns of genetic variation in Tennessee populations of *A. barbouri*. Mitochondrial-based phylogenetic reconstructions and partitioning of genetic variation at SNP loci (including assignment tests, DAPC analyses, AMOVA, and MSC reconstructions) identify the same three genetically distinct groups in Tennessee; these include a northern cluster, a central cluster, and a southern cluster. The Cumberland River appears to have served as a major barrier to gene flow between the northern and the central/southern clusters. The central and southern clusters both occur south of the Cumberland River and are divided by smaller regional drainage patterns. Individuals in the northern and central clusters occupy the Old Hickory Lake and Lower Stones River watersheds, respectively. The central cluster is bordered by the Cumberland River to the North and by the East Fork of the Stones River to the east. The southern cluster includes populations from three watersheds including the West Fork Stones River Watershed, Upper Duck River Watershed, and Mill Creek Watershed; the southern cluster (with the exception of the Wil2 population) is bordered by the Stones River to the north and by the Duck River to the south.

Mitochondrial-based phylogenetic reconstructions revealed that Tennessee *A. barbouri* populations are not monophyletic with respect to northern populations, which may reflect a history of repeated range expansions from Tennessee into the northern end of *A. barbouri*'s present day distribution. The estimated timing of divergence between the three Tennessee clusters falls within the mid-late Pleistocene. Glacial advances during the Pleistocene did not extend into Kentucky and Tennessee and this region likely served as a refuge for both terrestrial and aquatic fauna [68]. As glaciers retreated, populations at the northern limits of these refugia would have advanced northward into newly habitable territory. These expansions would have left a genetic signature characterized by reduced genetic variation in the northern end of their distribution as a result of founder effects. Repeated expansions of *A. barbouri* northward from Tennessee may explain the polyphyletic relationships of contemporary Tennessee *A. barbouri* with respect to populations in the north. Specifically, mitochondrial *A. barbouri* populations in northeastern Kentucky, Ohio, and Indiana (Clade V) are nested within Tennessee *A. barbouri* as are mitochondrial haplotypes from ambystomatid unisexuals and *A. texanum*. Genetic evidence of repeated northward expansions has been reported for the Eastern Woodrat (*Neotoma floridana*) [69] and other southeastern fauna (reviewed in Hewitt 2000; [70, 71]).

## Management implications and conclusions

Results from our genomic survey have specific implications for the design of management and conservation strategies that may improve the long-term persistence of *A. barbouri* in Tennessee. First, patterns of genetic variation in mitochondrial and nuclear genomes support the assignment of three genetically distinct units for management that warrant the designation of evolutionary significant units (ESUs), where ESUs are defined as groups of populations that show phylogeographic differentiation for mtDNA haplotypes and divergence in allele frequencies in nuclear markers [72, 73]. These three units include a Northern ESU encompassing all Sumner County populations (S2, S5, S7, and S8), a Central ESU that includes populations from eastern Davidson and Wilson Counties (D3, W1, W3 and W4), and a Southern ESU that includes populations in Williamson, Rutherford, and Bedford counties (Wil2, B6, R1, R7, and R9). Patterns of differentiation between these three groups of populations suggest a long history of genetic isolation at both mitochondrial and nuclear markers, such that different groups are likely to possess unique combinations of adaptive genetic variation and have likely experienced independent evolutionary trajectories.

Evolutionarily and phylogenetically distinctive populations contribute disproportionately to genetic diversity and should be ranked highly in regards to conservation priority. Geographically peripheral populations are frequently observed to be genetically less-redundant than more central populations and are more likely to possess potentially adaptive genetic variation [74]. For *A. barbouri*, Tennessee populations are geographically disjunct from the majority of *A. barbouri* breeding populations [14] and occupy the southernmost edge of the species distribution. Results from this study suggest that Tennessee *A. barbouri* should be prioritized for conservation planning as these populations are both genetically diverse and evolutionarily distinct from populations in the northern part of their distribution. The distinctiveness of these populations is further evidenced by observed differences in reproductive life-history traits, including mean diameter of early stage ovum size and number of eggs per clutch [8]. Prioritizing peripheral populations with adaptive genetic variation and evolutionary potential is even more critical when considering environmental challenges that accompany climate change. Amphibians in general are very sensitive to climate change as their reproductive life histories are linked to temperature and precipitation [75]. However, there is evidence that populations at the warm-range edge of their distribution are more resilient [76]. Protection of *A. barbouri* breeding sites in Tennessee may be instrumental to ensuring the long-term viability of this species as a whole.

## Supporting information

**S1 Table. Genbank accessions for mitochondrial D-loop sequences used in phylogenetic reconstructions.**

(DOCX)

**S1 Fig.** Assignment probabilities from DAPC analysis for A.) sampled populations of *A. barbouri* in Tennessee B.) sampled populations of *A. barbouri* in Tennessee and Kentucky, and C.) all sampled populations of *A. barbouri* and two representative populations of *A. texanum*.

(DOCX)

## Acknowledgments

We thank Ryan Hanscom for assistance with sample collections. Joanna Bellan assisted with mitochondrial sequencing and data analysis. Andrea Drayer provided *A. barbouri* tissue samples from Kentucky populations. We also thank Kristin Womble for GIS assistance and for producing the map in Fig 2.

## Author Contributions

**Conceptualization:** N. Wade Hubbs, Carla R. Hurt, Brian Leckie, David Withers.

**Formal analysis:** Carla R. Hurt.

**Funding acquisition:** Carla R. Hurt.

**Investigation:** N. Wade Hubbs, Carla R. Hurt, John Niedzwiecki.

**Methodology:** Brian Leckie.

**Project administration:** Carla R. Hurt.

**Writing – original draft:** N. Wade Hubbs, Carla R. Hurt.

**Writing – review & editing:** Carla R. Hurt, John Niedzwiecki, David Withers.

## References

1. Geist J. Strategies for the conservation of endangered freshwater pearl mussels (*Margaritifera margaritifera* L.): a synthesis of conservation genetics and ecology. *Hydrobiologia*. 2010 May; 644(1):69–88.
2. Withers DI, Condict K, McCoy R. A guide to the rare animals of Tennessee. Division of Natural Areas, Tennessee Department of Environment and Conservation, Nashville, Tennessee. 2009.
3. Tennessee Wildlife Resources Agency Rules of Biodiversity. Rules and Regulations for in Need of Management, Threatened, and Endangered Species. Chapter 1660-01-32. 2018.
4. Anderson MA, Campbell JR, Carey AN, Dodge DR, Johnston RA, Mattison ER, et al., Population survey of the streamside salamander in the Nashville Basin of Tennessee. *Southeast Nat*. 2017; 13(1):101–107."
5. Wright BR, Farquharson KA, McLennan EA, Belov K, Hogg CJ, Grueber CE. A demonstration of conservation genomics for threatened species management. *Mol Ecol Resour*. 2020 Nov; 20(6):1526–41. <https://doi.org/10.1111/1755-0998.13211> PMID: 32562371
6. Zhang S, Li C, Li Y, Chen Q, Hu D, Cheng Z, et al., Genetic Differentiation of Reintroduced Père David's Deer (*Elaphurus davidianus*) Based on Population Genomics Analysis. *Front Gen*. 2021:1611. <https://doi.org/10.3389/fgene.2021.705337> PMID: 34557217
7. Devitt TJ, Wright AM, Cannatella DC, Hillis DM. Species delimitation in endangered groundwater salamanders: Implications for aquifer management and biodiversity conservation. *Proc Nat Acad Sci*. 2019 Feb 12; 116(7):2624–33. <https://doi.org/10.1073/pnas.1815014116> PMID: 30642970
8. Niemiller ML, Glorioso BM, Nicholas C, Phillips J, Rader J, Reed E, et al., Notes on the reproduction of the Streamside Salamander, *Ambystoma barbouri*, from Rutherford County, Tennessee. *Southeast Nat*. 2009; 8(1):37–44.
9. Kraus F, Petranka JW. A new sibling species of *Ambystoma* from the Ohio River drainage. *Copeia*. 1989 Feb 27:94–110.
10. Niedzwiecki JH. Evolutionary history and hybridization of two mole salamander sister species from different habitats. University of Kentucky; 2005.
11. Lampert KP, Schartl M. A little bit is better than nothing: the incomplete parthenogenesis of salamanders, frogs and fish. *BMC Biology*. 2010 Dec; 8(1):1–3. <https://doi.org/10.1186/1741-7007-8-78> PMID: 20687905
12. Bogart JP, Bi K, Fu J, Noble DW, Niedzwiecki J. Unisexual salamanders (genus *Ambystoma*) present a new reproductive mode for eukaryotes. *Genome*. 2007 Feb; 50(2):119–36. <https://doi.org/10.1139/g06-152> PMID: 17546077
13. Eastman JM, Niedzwiecki JH, Nadler BP, Storfer A. Duration and consistency of historical selection are correlated with adaptive trait evolution in the Streamside Salamander, *Ambystoma barbouri*. *Evolution*. 2009; 63(10):2636–2647. <https://doi.org/10.1111/j.1558-5646.2009.00741.x> PMID: 19500149
14. Niemiller ML, Glorioso BM, Nicholas C, Phillips J, Rader J, Reed E, et al., Status and distribution of the Streamside Salamander, *Ambystoma barbouri*, in middle Tennessee. *Am Midl Nat*. 2006; 156(2):394–399.
15. Shaffer HB, McKnight ML. The polytypic species revisited: genetic differentiation and molecular phylogenetics of the Tiger Salamander *Ambystoma tigrinum* (Amphibia: Caudata) complex. *Evolution*. 1996; 50(1):417–433. <https://doi.org/10.1111/j.1558-5646.1996.tb04503.x> PMID: 28568845
16. Charney ND, Ireland AT, Bettencourt BR. Mapping genotype distributions in the unisexual *Ambystoma* complex. *J Herpetol* 2014; 48(2):210–219.
17. Church SA, Kraus JM, Mitchell JC, Church DR, Taylor DR. Evidence for multiple Pleistocene refugia in the postglacial expansion of the Eastern Tiger Salamander, *Ambystoma tigrinum*. *Evolution* 2003; 57(2):372–383. [https://doi.org/10.1554/0014-3820\(2003\)057\[0372:EFMPRI\]2.0.CO;2](https://doi.org/10.1554/0014-3820(2003)057[0372:EFMPRI]2.0.CO;2) PMID: 12683533
18. Zamudio KR, Savage WK. Historical isolation, range expansion, and secondary contact of two highly divergent mitochondrial lineages in Spotted Salamanders (*Ambystoma maculatum*). *Evolution*. 2003; 57(7):1631–1652. <https://doi.org/10.1554/02-342> PMID: 12940367
19. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics*. 2003; (1):2–3.
20. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acid*. 1999; 41:95–98.
21. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014 May 1; 30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623

22. Miller MA, Pfeiffer W, Schwartz T. The CIPRES science gateway: a community resource for phylogenetic analyses. In Proceedings of the 2011 TeraGrid Conference: extreme digital discovery. 2011; Jul 18: 1–8.
23. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001; 17(8): 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754> PMID: 11524383
24. Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics (Oxford, England)* 1998; 14(9):817–818. <https://doi.org/10.1093/bioinformatics/14.9.817> PMID: 9918953
25. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018 Jun; 35(6):1547. <https://doi.org/10.1093/molbev/msy096> PMID: 29722887
26. Lanfear R, Hua X, Warren DL. Estimating the effective sample size of tree topologies from Bayesian phylogenetic analyses. *Gen Biol Evol*. 2016 Aug 1; 8(8):2319–32. <https://doi.org/10.1093/gbe/evw171> PMID: 27435794
27. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 2018 Jan; 4(1):vey016. <https://doi.org/10.1093/ve/vey016> PMID: 29942656
28. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One*. 2011; 6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379> PMID: 21573248
29. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. *G3-Genes Genom Genet*. 2011; 1(3):171–182. <https://doi.org/10.1534/g3.111.000240> PMID: 22384329
30. Eaton DAR. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*. 2014; 30(13):1844–1849. <https://doi.org/10.1093/bioinformatics/btu121> PMID: 24603985
31. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. diveRsity: An R package for the estimation of population genetics parameters and their associated errors. *Meth Ecol Evol*. 2013; 4(8): 782–788.
32. Zhdanova OL, Pudovkin AI. Nb\_HetEx: a program to estimate the effective number of breeders. *Jour Hered*. 2008 Nov 1; 99(6):694–5. <https://doi.org/10.1093/jhered/esn061> PMID: 18703539
33. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol Ecol Resour*. 2014; 14(1):209–214. <https://doi.org/10.1111/1755-0998.12157> PMID: 23992227
34. Nunziata SO, Lance SL, Scott DE, Lemmon EM, Weisrock DW. Genomic data detect corresponding signatures of population size change on an ecological time scale in two salamander species. *Mol Ecol*. 2017 26(4):1060–1074. <https://doi.org/10.1111/mec.13988> PMID: 28026889
35. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984 11(1):1358–70. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x> PMID: 28563791
36. Meirmans PG, Hedrick PW. Assessing population structure: FST and related measures. *Mol Ecol Res*. 2011 Jan; 11(1):5–18.
37. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res*. 2010; 10(3):564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x> PMID: 21565059
38. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; 131(2):479–491. <https://doi.org/10.1093/genetics/131.2.479> PMID: 1644282
39. Jombart T, Ahmed I. Adegnet 1.3–1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 2011; 27(21): 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521> PMID: 21926124
40. Jombart T, Collins C. A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0. London, Imperial College London, MRC Centre for Outbreak Analysis and Modelling. 2015.
41. Pritchard JK, Wen W, Falush D. Documentation for STRUCTURE software: Version 2. Department of Human Genetics, University of Chicago, Chicago. 2003
42. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour*. 2015 Sep; 15(5):1179–91. <https://doi.org/10.1111/1755-0998.12387> PMID: 25684545
43. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol*. 2005; 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> PMID: 15969739

44. Earl DA, VonHoldt BM. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 2012; 4(2):359–361.
45. Yang Z. The BPP program for species tree estimation and species delimitation. *Current Zoology.* 2015; 61(5):854–65.
46. Allio R, Donega S, Galtier N, Nabholz B. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol Biol Evol.* 2017; 34(11):2762–2772. <https://doi.org/10.1093/molbev/msx197> PMID: 28981721
47. Reis MD, Yang Z. Bayesian molecular clock dating using genome-scale datasets. In *Evolutionary genomics 2019* (pp. 309–330). Humana, New York, NY.
48. Leaché AD, Zhu T, Rannala B, Yang Z. The spectre of too many species. *Sys Biol.* 2019 Jan 1; 68(1):168–81. <https://doi.org/10.1093/sysbio/syy051> PMID: 29982825
49. McCartney-Melstad Evan, Vu Jannet K., and Bradley Shaffer H. "Genomic data recover previously undetectable fragmentation effects in an endangered amphibian." *Molecular Ecology* 27, no. 22 (2018): 4430–4443.
50. Kardos M, Armstrong E, Fitzpatrick S, Hauser S, Hedrick P, Miller J., et al., 2021. The crucial role of genome-wide genetic variation in conservation.
51. Schmidt C, Garroway CJ. The population genetics of urban and rural amphibians in North America. *Mol. Ecol.* 2021 <https://doi.org/10.1111/mec.16005> PMID: 34053153
52. Fusco N.A., Pehek E. and Munshi-South J. Urbanization reduces gene flow but not genetic diversity of stream salamander populations in the New York City metropolitan area. *Evol. App.* 2010; 14(1): 99–116.
53. Huang H, Knowles LL. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Sys Bio* 2016; 65(3): 357–365. <https://doi.org/10.1093/sysbio/syu046> PMID: 24996413
54. Micheletti SJ, Storfer A. An approach for identifying cryptic barriers to gene flow that limit species' geographic ranges. *Mol Ecol.* 2017; 26(2):490–504. <https://doi.org/10.1111/mec.13939> PMID: 27886423
55. Haasl RJ, Payseur BA. Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity* 2011; 106(1):158–171. <https://doi.org/10.1038/hdy.2010.21> PMID: 20332809
56. Jehle R, Arntzen JW. Microsatellite markers in amphibian conservation genetics. *Herpetol J.* 2002; 12:1–9.
57. Heredia-Bobadilla RL, Monroy-Vilchis O, Zarco-González MM, Martínez-Gómez D, Mendoza-Martínez GD, Sunny A. Genetic structure and diversity in an isolated population of an endemic mole salamander (*Ambystoma rivulare* Taylor, 1940) of central Mexico. *Genetica.* 2016 Dec; 144(6):689–98. <https://doi.org/10.1007/s10709-016-9935-9> PMID: 27796527
58. Wang IJ, Johnson JR, Johnson BB, Shaffer HB (2011) Effective population size is strongly correlated with breeding pond size in the endangered California Tiger Salamander, *Ambystoma californiense*. *Conserv Genet.* 2011; 12(4):911–920.
59. Sunny A, Monroy-Vilchis O, Fajardo V, Aguilera-Reyes U. Genetic diversity and structure of an endemic and critically endangered stream river salamander (Caudata: *Ambystoma leorae*) in Mexico. *Cons Gen.* 2014 Feb; 15(1):49–59.
60. Funk WC, Tallmon DA, Allendorf FW. Small effective population size in the Long-toed Salamander. *Mol Ecol.* 1999; 8(10):1633–1640. <https://doi.org/10.1046/j.1365-294x.1999.00748.x> PMID: 10583827
61. Savage WK, Fremier AK, Bradley Shaffer H. Landscape genetics of alpine Sierra Nevada salamanders reveal extreme population subdivision in space and time. *Mol Ecol.* 2010 19(16):3301–3314. <https://doi.org/10.1111/j.1365-294X.2010.04718.x> PMID: 20701683
62. Waples RS, Do CH. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary applications.* 2010 May; 3(3):244–62. <https://doi.org/10.1111/j.1752-4571.2009.00104.x> PMID: 25567922
63. Waples RS, Do CH. LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular ecology resources.* 2008 Jul; 8(4):753–6. <https://doi.org/10.1111/j.1755-0998.2007.02061.x> PMID: 21585883
64. Vences M. and Wake D.B., 2007. Speciation, species boundaries and phylogeography of amphibians. *Amphibian biology*, 7, pp.2613–2671.
65. Nichols R. Gene trees and species trees are not the same. *Trends Ecol Evol.* 2001 Jul 1; 16(7):358–64. [https://doi.org/10.1016/s0169-5347\(01\)02203-0](https://doi.org/10.1016/s0169-5347(01)02203-0) PMID: 11403868

66. Toews DP, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol*. 2012 Aug; 21(16):3907–30. <https://doi.org/10.1111/j.1365-294X.2012.05664.x> PMID: 22738314
67. Patton AH, Margres MJ, Epstein B, Eastman J, Harmon LJ, Storfer A. Hybridizing salamanders experience accelerated diversification. *Scientific reports*. 2020 Apr 16; 10(1):1–2.
68. Jacquemin SJ, Ebersole JA, Dickinson WC, Ciampaglio CN. Late Pleistocene fishes of the Tennessee River Basin: an analysis of a late Pleistocene freshwater fish fauna from Bell Cave (site ACb-2) in Colbert County, Alabama, USA. *PeerJ*. 2016; 2(4):e1648. <https://doi.org/10.7717/peerj.1648> PMID: 26855876
69. Hayes JP, Harrison RG. Variation in mitochondrial DNA and the biogeographic history of woodrats (*Neotoma*) of the eastern United States. *Syst Biol*. 1992; 41(3):331–
70. Hewitt G. The genetic legacy of the Quaternary ice ages. *Nature*. 2000; 405(6789):907–13. <https://doi.org/10.1038/35016000> PMID: 10879524
71. Hewitt GM. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol Jour Linnean Soc*. 1996; 58(3):247–76.
72. Moritz C. Defining ‘evolutionarily significant units’ for conservation. *Trends Ecol Evol*. 1994; 9(10):373–375. [https://doi.org/10.1016/0169-5347\(94\)90057-4](https://doi.org/10.1016/0169-5347(94)90057-4) PMID: 21236896
73. Moritz C. Conservation units and translocations: strategies for conserving evolutionary processes. *Hereditas* 1999; 130(3):217–228.
74. Volkmann L, Martyn I, Moulton V, Spillner A, Mooers AO. Prioritizing populations for conservation using phylogenetic networks. *PONE* 2014; 9(2):e88945. <https://doi.org/10.1371/journal.pone.0088945> PMID: 24586451
75. Corn PS. Climate change and amphibians. *Biodiv Cons*. 2005; 28(1).
76. Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibáñez C, et al., Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proc Nat Acad Sci*. 2019; 116(21):10418–10423. <https://doi.org/10.1073/pnas.1820663116> PMID: 31061126