A Genomic Perspective on the Conservation Status of the Endangered Nashville Crayfish (*Faxonius shoupi*)

Carla Hurt (churt@tntech.edu)
Tennessee Technological University https://orcid.org/0000-0001-5850-033X
Parker Hildreth
Tennessee Technological University: Tennessee Tech University
Carl Williams
Tennessee Wildlife Resources Agency

Research Article

Keywords: Faxonius shoupi, crayfish, Mitochondrial DNA, Genotyping by Sequencing, Endangered Species

Posted Date: April 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-393405/v1

Version of Record: A version of this preprint was published at Conservation Genetics on March 24th, 2022. See the published version at https://doi.org/10.1007/s10592-022-01438-6.

Abstract

The Nashville crayfish (*Faxonius shoupi*) was federally listed as an endangered species in 1986 due to its limited distribution in the Mill Creek watershed; this waterway lies in the rapidly developing Nashville basin and has experienced habitat degradation due to agricultural run-off, contamination, and urban development. Recovery efforts, including dam removal and restoration of riparian zones, have improved conditions in Mill Creek and *F. shoupi* has increased in numbers and recolonized extirpated stream segments. However, a history of demographic bottlenecks and restricted gene flow may have negatively impacted the long-term recovery of this species. A recently discovered population of *F. shoupi* in a disjunct segment of the Lower Tennessee River at the Pickwick Tailwater may provide an additional source of genetic variation. Uncertainty surrounding the origins of the Pickwick population and its taxonomic relationship to *F. shoupi* in Mill Creek raises questions about the conservation and management implications of this population. We used mitochondrial sequencing and SNP genotyping to assess genetic variation and connectivity of *F. shoupi* in the Mill Creek drainage and to investigate the taxonomy and demographic history of the newly discovered population at Pickwick. We found substantial genetic variation and evidence of connectivity for samples throughout Mill Creek for both mitochondrial and genome-wide SNPs. Our results also suggest a recently severed connection between crayfish in Pickwick and Mill Creek. Unique mitochondrial haplotypes and SNP variation in the Pickwick population highlight the need for prioritizing this population in future conservation and management plans for this species.

Introduction

The Nashville Crayfish (*Faxonius shoupi*) is the only federally protected crayfish in the state of Tennessee. Listing of this species in 1986 was enacted in response to the limited distribution of *F. shoupi* in Middle Tennessee and the loss of quality habitat and ongoing development in the Nashville Basin. The known distribution of *F. shoupi* from the time of its discovery in 1938-1939 (first identified as *Cambarus propiquus sanborni*, Fleming 1939) until 2019, has primarily been limited to Mill Creek (MC) and its tributaries (Davidson and Williamson Counties; Fig. 1). There have also been reported occurrences of *F. shoupi* in Big Creek (Elk River Drainage), the South Harpeth River, and in Richland Creek (Cumberland River Drainage) (Bouchard 1974). The greatest threats to the persistence of *F. shoupi* in MC has been poor water quality and siltation from agriculture, competition with the invasive crayfish, and rapid urbanization within Williamson and Davidson counties (USFWS 2018). Restoration projects have been implemented in MC in an effort to improve water quality, restore channel stability, enhance aquatic habitat, and restore riparian buffer function. Recent surveys have shown that *F. shoupi* can now be observed in high number, even in heavily developed areas; as a result, the USFWS proposed to remove this species from the federally endangered species list (USFWS 2018). However, disruption of gene flow due to habitat disturbance and impoundments as well as loss of genetic variation from recent bottlenecks can still inhibit the long-term adaptive potential and resiliency of this species.

Repeated sampling efforts of *F. shoupi* populations throughout the MC drainage show that populations were negatively impacted by urbanization and development (Bouchard 1984, Barrociere 1986, USFWS 2018). Specimen numbers from surveys in 1986 were less than a third of the numbers collected from the same sites in 1969 (Hobbs et al. 1969). Also, in the Sevenmile tributary to MC, *F. shoupi* appears to have been replaced by an undescribed species later identified as *F. durelli* (Bouchard 1972). More recent surveys in MC and its tributaries indicate that *F. shoupi* has responded favorably to ongoing habitat restoration efforts. Recolonization surveys have shown that *F. shoupi* is able to quickly recolonize habitat following local extirpations (Carpenter 2002). Despite recovery in numbers, these demographic bottlenecks can still have negative fitness consequences by depleting standing genetic variation, fixing deleterious alleles, and reducing adaptive potential to environmental change.

The recent discovery of a disjunct population of *F. shoupi* in the Lower Tennessee River at the Pickwick Tailwater (PW) in Hardin County, Tennessee suggests that this species may be more widely distributed and may occupy larger waterways than previously thought. A biological survey of the Lower Tennessee River in 2019 produced specimens morphologically identified as *F. shoupi* between TR mile 205.1 downstream to TR mile 203.7. Diagnostic characters for the PW population, including first form male gonopod, cephalothorax morphology and chelae morphology, match the description of *F. shoupi* specimens from MC populations. However, the PW population also exhibits a unique pigmentation pattern characterized by pale vermiculations on a contrasting darker pigmented region of the cephalothorax compared to the small pale speckles restricted to punctations observed on the MC specimens (Fig. 2). Mill Creek and PW sites differ dramatically in their habitat characteristics. Mill Creek is a second order stream with a maximum depth of less than two meters. In contrast, river width at the PW site is more than 300 meters and crayfish were collected from a depth of nearly four meters. Interestingly, *Faxonius shoupi* at PW co-occur with a number of widespread stream-dwelling crayfish that are not known to occupy the MC drainage; these include *Cambarus sp. nov.*, *F. forceps*, *F. mirus*, and *F. placidus*. The population of *F. shoupi* at PW appears to be well established, however the extent of the range distribution in the Tennessee River remains undetermined until comprehensive surveys efforts are completed.

The relationship of the *F. shoupi* PW to populations in MC has important implications for management strategies and conservation status of this species. The discovery of an additional, viable population of *F. shoupi* would suggest that this species is not as vulnerable as previously thought. However, the taxonomic placement and the origins of the PW population are unclear. Genetic evidence is needed to support morphological

assignment of the PW population to *F. shoupi*; traditional morphological traits used for crayfish taxonomy frequently contradict molecular evidence (Taylor and Knouft 2006). It is also unclear if the PW population of *F. shoupi* is the result of a recent, anthropogenic release or if this population represents an older native population that was only recently discovered. Genetic tools can be used to test hypotheses regarding colonization history. Anthropogenic mediated introductions, such as a bait-bucket release, would be established by a small number of propagules; this founder event would leave a detectable genetic signature. Specifically, population bottlenecks quickly remove rare alleles from a population resulting in an excess of heterozygosity relative to expectations for a population in mutation-drift equilibrium. Also, this bottleneck event would reduce genome-wide genetic variation relative to source populations of the species. Information on the taxonomic relationship and origins of the crayfish at PW is needed to effectively incorporate this population into management strategies for this species.

We used both mitochondrial sequencing and genotype-by-sequencing datasets to investigate population structure and demographic histories of *F. shoupi* in MC and to investigate the history and taxonomic status of the recently discovered population in the Lower Tennessee River downstream of Pickwick Dam. Our specific objectives were to assess levels of genetic variation within populations of *F. shoupi*, test for genetic signatures of recent population declines, estimate levels of gene-flow throughout the MC drainage, and reconstruct the relationship between the MC and PW sites. Phylogenetic reconstructions and a Bayesian multispecies coalescent model were applied to estimate population divergence times between the two disjunct populations. Results from this study are interpreted in the context of management considerations for maintaining genetic variation and the adaptive potential of this species.

Methods

Collections. A total of 76 tissue samples were obtained from crayfish morphologically identified as *F. shoupi* from three sites within MC, Tennessee and one site in the Tennessee River at Pickwick Tailwater (Table 1, Fig. 1). Collections at PW were conducted by scuba divers operating from an anchored dive boat at depths ranging from approximately 2-4 meters and a river width of 308 meters. Crayfish were collected by overturning loose slab rocks and quickly grabbing specimens with gloved hands. The specimens were placed in mesh holding bags, brought to the surface where they were processed for data before being released.

Sampling sites in the MC drainage were selected to span the currently known distribution of the species with portions of the upper, middle (type locale) and lower reaches serving as designated sites and to facilitate comparisons between MC and PW. Tissues from outgroup species were collected from the Green River in Buffalo River Drainage (*F. durelli*, type locality), Barrens Fork in Collins River Drainage *(F. placidus*, BF population) and Sinking Creek in the Lower Cumberland - Old Hickory Lake Drainage (*F. placidus*, type locality) (Table 1). Tissue samples from *F. shoupi* were obtained from leg sections separated at autonomizing joints (approximately 5 mm) and individuals were immediately released on site. Crayfish are known to undergo autotomy of appendages and regeneration initiates after ~6 days, depending on the intermolt period (Durand 1960). Complete specimens for outgroups were preserved in 70% ethanol and will be deposited as voucher specimens in the Tennessee Tech Ichthyology collection. Tissues were stored in RNAlater at -20°C for preservation of DNA prior to extraction.

Mitochondrial sequencing. Genomic DNA was extracted from leg tissues using the EZNA Tissues DNA Kit (Omega Bio-tek) except that DNA was eluted in water in the final step. The 5' end of the mitochondrial gene cytochrome oxidase I (COI) was amplified using universal primers LCO1490 (GGTCAACAAATCATAAAGATATTGG) and HCO2198 (TAAACTTCAGGGTGACCAAAAAATCA) (Folmer et al. 1994). Conditions for polymerase chain reactions (PCR) were as follows: initial denaturation step of 5 minutes at 95°C followed by 35 cycles of 15 s at 95°C, 15 s at 54°C, 45 s at 72°C. This program ended with a final extension of 10 minutes at 72°C. PCR products were cleaned prior to cycle sequencing reactions 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 version 5.2 (Gene Codes Corp.). Sequences were aligned using the ClustalW alignment algorithm (Thompson et al. 1994) as implemented in Bioedit (Hall 1999). Alignments were refined by eye and protein-coding genes were examined for stop codons to check for miscalled bases and pseudogenes.

Mitochondrial Haplotype Diversity. Estimates of genetic variation within sampled populations at the mitochondrial COI gene, including genotypic diversity (H_{GD}) and nucleotide diversity (π_M) were performed in Arlequin 3.5.2 (Excoffier et al. 2010). Mitochondrial haplotypes from all *F. shoupi* individuals and outgroups were trimmed to an equal length of 632 base pairs and used in a reconstruction of a minimum spanning haplotype network as performed by the package *pegas* (Pardis 2010) in the R-studio suite (Racine 2012). Structuring of genetic variation for mitochondrial COI haplotypes was investigated using Analysis of Molecular Variance (AMOVA) as performed by Arlequin (Excoffier et al. 1992). The hierarchical analysis investigated the proportion of molecular variance due to difference between drainages (i.e. MC populations and PW; F_{ST}), variation between populations within drainages (F_{SC}), and variation within sampled populations (F_{CT}).

Phylogenetic Analyses. Phylogenetic reconstructions of mitochondrial haplotypes were estimated using both Bayesian and maximum likelihood (ML) optimality criteria. Both analyses assumed the HKY-Gamma model of evolution (Hasegawa et al. 1985); this was selected as the best-fit model under Bayesian Information Criteria in as performed by MEGA X (Kumar et al. 2018). A Bayesian phylogenetic reconstruction was performed using MrBayes version 3.2.7a on the CIPRES Science Gateway (Miller et al., 2010). 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 (first 10% of trees). A majority-rule consensus tree was generated and nodal support was estimated by posterior probabilities. A ML analysis was performed using MEGA X and the tree with the highest log likelihood was retained. Nodal support was estimated by bootstrap analysis with 1000 replicates.

Genotyping by Sequencing. Genotyping by sequencing (GBS) was used to identify and genotype SNPs across all four *F. shoupi* sites, *F. placidus* LC and *F. placidus* BF (Table 2) following the protocol described by Elshire et al. (2011). Genotyping by sequencing library construction used the same DNA extractions that were used for mitochondrial Sanger sequencing. Extracted DNA was standardized to 5 ng/µl and quantified using the Quantifluor dsDNA System (Promega). Standardized DNA was digested with the restriction enzyme ApeKI (New England Biolabs) and barcoded adapters were ligated to cut restriction sites. Each individual was labelled with a unique barcode. Ligated DNA was pooled and PCR amplified using primers complementary to ligated adapters. The resulting PCR product was cleaned using the Qiaquick PCR purification kit and the distribution of PCR fragment lengths was measured using an Agilent 2100 BioAnalyzer (Agilent Technologies). The resulting library was cleaned using the AxyPrep Mag bead purification system (Agilent) and sequenced on an Illumina NexSeq 500 (Illumina Inc.) with the 75 bp, single-end read chemistry.

Genotyping and SNP filtering. Single Nucleotide Polymorphisms were identified and filtered using the *de novo map* STACKS pipeline (Catchen et al. 2013). In *ustacks*, the minimum number of raw reads needed to form a stack (allele) was set to *m*=3 and the number of mismatches between alleles within individuals *M*=2. A catalog of loci was constructed in *cstacks* where *n*=2 mismatches were allowed between sample tags. The *populations* function was used to estimate population summary statistics and to generate input datasets for downstream analyses. One source of bias in GBS datasets is mutations that occur at restriction enzyme cut-sites; the accumulation of substitutions in restriction sites can influence population genetic inferences by causing allele drop-out and biasing datasets to more conserved regions of the genome (Guatier et al. 2013; Silliman et al. 2021). We investigated the influence of filtering parameters for missing data on the retention of loci and variable sites using the filtering flags in the *populations* function. Estimates of contemporary effective population sizes and tests for bottlenecks used the p1r70 dataset, where SNPs were required to be present in 70% of individuals in a single population. This dataset was used to maximize the number of loci within populations instead of shared loci across populations. Estimates of within population genetic variation and population structure (Bayesian assignment tests, DAPC, and pairwise F_{ST}' and Φ_{ST} estimates) were performed using the p2r60 dataset, so that only SNPs present in at least two out of four of the *F. shoupi* populations were included in these analyses. Maximum heterozygosity was set at 0.60 to remove homeologous loci. Tests for Hardy-Weinberg equilibrium (HWE) were performed in Plink 2.0 (Chang 2015) and loci with *p*-values below 0.01 in any single populations were removed from further analyses.

Genetic Variation at SNP loci. Estimates of genetic variation including mean observed heterozygosity (H_o), expected heterozygosity (*He*), and nucleotide diversity at variable sites (π_N) were estimated in the program *populations* within the STACKS workflow. We estimated the average allelic richness (A_{RN}) and private allelic richness (P_{RN}) using a rarefaction approach to correct for unequal sample size across sites as implemented in the software ADZE (Szpiech et al. 2008). Due to the loss of shared loci across species we did not include *F. placidus* LC and *F. placidus* BF in rarefaction analyses.

Demographic Histories. Contemporary effective populations size (N_e) estimates for all populations of *F. shoupi* and *F. placidus* were performed using the LD method (Waples and Do 2008), as implemented in NeEstimator v2.01 (Do et al. 2014). The LD method measures the strength of association of alleles at independent loci to infer the degree of drift as a function of N_e. A threshold of 0.02 for the allelic frequency was chosen in order to reduce bias resulting from low-frequency variants. Confidence intervals (95%) were estimated using parametric bootstrap. Linkage disequilibrium based estimates of N_e were adjusted for physical linkage using the equation by Waples et al. (2016).

We used the estimated chromosome number from the congeneric *F. virilis* for N_e adjustments (*Chr* = 200, Mlinarec 2016).

$$\frac{\widehat{N_e}}{N_e} = 0.098 + 0.219 \times \ln(Chr)$$

Rapid declines in population size can also leave a signature on patterns of allelic variation within populations due to a loss of low frequency alleles. We tested for evidence of recent population bottlenecks using the program BOTTLENECK (Piry et al. 1999). A test for excess heterozygosity (compared with mutation-drift equilibrium) was performed using the 1-tailed Wilcoxon signed rank test (α = 0.05), which detects historical bottlenecks (0.25 – 2.5 times N_e generations; Cornuet and Luikart 1996). Simulation parameters were set for 1000 iterations and assumed an infinite allele model (IAM). We also tested for a mode-shift in allele frequency distributions to detect a genetic signature of more recent bottlenecks (Luikart et al. 1998). Recent bottlenecks result in a rapid loss of low frequency alleles, resulting in the allele frequency spectrum to a greater number of alleles at intermediate frequencies then expected under mutation-drift equilibrium. The allele frequency shift is detectable for several dozen generations.

Population Structure. Pairwise population differentiation at SNP loci was summarized using pairwise F_{ST}' and Φ_{ST} as implemented in the program *Populations* in the STACKS workflow. F_{ST}', an analogue of F_{ST}, is a standardized measure of differentiation that is independent of the amount of genetic variation that is present within populations (Meirmans 2006). The Φ_{ST} statistic is an AMOVA-based measure of population differentiation that incorporates information regarding the molecular distance between haplotypes (Excoffier 1992).

A Bayesian assignment test was implemented in the software STRUCTURE v. 2.3.4 (Pritchard et al. 2000) on all *F. placidus* and *F. shoupi* populations; *F. durelli* was not included due to small sample size. We performed ten independent runs for each value of K (1-6) assuming the admixture model. Markov chain Monte Carlo (MCMC) settings included two independent MCMC chains, each with a burnin of 50,000 iterations followed by 100,000 iterations. The optimal valued of *K* was examined by plotting the average natural log of the probability of the data for *K*=1-6 (Ln Pr(X|*K*); Pritchard and Wen 2003). We also investigated the optimal value of *K* using the Δ*K* method as implemented in Structure Harvester; however, due to small standard deviation across replicates for K=3, the ΔK method could not be performed (Evanno 2005; Earl 2012). Permuted membership assignment coefficients from the ten independent runs were estimated using Clumpp v. 1.1.2 (Jokobsson and Rosenberg, 2007). Structure bar plots were generated using STRUCTURE PLOT (Ramasamy et al. 2014).

Discriminant analysis of principal components (DAPC) was performed in R using the package 'adegenet' (Jombart et al. 2010). This method determines the number of populations or groups without prior knowledge of sample origins. Discriminant analysis of principal components differs from Bayesian clustering in that it does not assume that loci are unlinked and that populations are randomly mating. The optimal number of clusters was determined using BIC as performed by the *find.clusters* function.

Divergence time Estimation. Divergence times estimates for the split between the *F. shoupi* MC sites and PW were estimated using the Bayesian multispecies-coalescent-with-introgression (MSci) method implemented in the A00 model in BPP v.4.2 (Yang 2015). The input tree for the BPP analysis was based on results from the mtDNA gene tree. Two different models were used for divergence time estimates. The first model assumed no migration after initial population divergence between *F. shoupi* populations at MC and PW. The second model allowed bidirectional migration between the two drainages. Each analysis of 50,000 MCMC generations was ran twice from different starting seeds with a burn-in period of 5000 using algorithm 0 (default fine-tuning parameter, e = 2) and an estimated heredity (a = 4, b = 4); this gave consistent parameter estimates between replicate runs and ESS values >1000 for all parameters. We consider speciation probability values >0.95 as strong support for all speciation events. The theta prior (3, 0.04), tau prior (3,0.2) phi prior (1,1) finetune automatic update. The posterior probabilities of the model with migration (M1) and without migration (M2) were compared using the stepping-stones method as implemented in the bppr package (https://github.com/dosreislab/bppr, see also, and Rannala and Yang 2017). The stepping-stone method analysis used 16 independent MCMC chains for each model (50,000 generations per chain).

Results

Mitochondrial Haplotypes. Sequencing of mitochondrial COI identified 28 unique haplotypes across all sampled populations and outgroups (Table 2). *Faxonius shoupi* sampled from MC-M and PW had the greatest number of haplotypes and haplotypic diversity. The COI sequence alignment used in MSN analysis was trimmed to an equal length of 632 bp; this reduced the number of haplotypes from 28 to 21. The resulting MSN network showed four common mitochondrial haplotypes shared at high frequency across all sampled populations of *F. shoupi* (Fig. 3). All haplotypes within *F. shoupi* were connected by at most three mutational steps. Haplotypes from *F. placidus* LC, *F. placidus* BF, and *F. durelli* were separated from *F. shoupi* haplotypes by a minimum of 33 steps. Phylogenetic reconstructions based on Bayesian and ML criteria resulted in nearly identical topologies. Both criteria recovered a well-supported monophyletic clade that included all haplotypes from the four sampled *F. shoupi* populations (Fig. 4). Haplotype sequences were not structured by populations; the 18 haplotypes belonging to *F. shoupi* formed an unresolved polytomy in both Bayesian and ML reconstructions.

The results of the AMOVA for COI haplotypes (Table 3) showed that, while the majority of genetic variation was found within sampled populations (98.4%), some structuring of haplotype diversity could also be attributed to differences between the populations in MC and PW (3.16%). None of the variance in mitochondrial haplotype diversity could be attributed to differences between sampled sites within the MC watershed.

SNP Genotyping. We genotyped a total of 128 individuals from four populations of *F. shoupi* and two populations of *F. placidus*. (Table 2). Three individuals with low sequence coverage (less than 500,000 sequenced tags) were removed from further analysis. The average number of reads across retained individuals was 3,931,286 sequenced tags per individual. The *denovomap* STACKS pipeline generated 2,031,321 loci, with a mean coverage of 14.3 across individuals and the mean number of sites per locus was 82. The influence of filtering parameters, including the minimum number of populations (p) and the proportion of sequenced individuals (r), greatly influenced the number of genotyped SNPs. The number of retained loci, variant sites, and mean number of sampled individuals per locus for different parameters are listed in Supplement 1.

Within Population Genetic Variation. Estimates of within population genetic variation including Het_{Obs}, Het_{Exp}, and π_N at SNP loci showed that the *F. shoupi* at PW had the highest levels of genetic variation (Table 2). Within the MC drainage, the MC-U population was more variable than

downstream sites MC-M and MC-L. Heterozygosity was similar for *F. placidus* LC and for populations of F. *shoupi*. *Faxonius placidus* BF had the lowest estimates of genetic variation overall. Rarefaction estimates of the average A_{RN} and P_{RN} for PW and MC-U populations were similar. Downstream sites within MC had lower estimates of A_{RN} and P_{RN} when accounting for differences in sample size (Supplement 2).

Effective Population Size and Bottlenecks. Estimates of effective population sizes based on the LD method ranged from 19-41; *F. shoupi* MC-M had the highest estimate of N_e and MC-U had the lowest estimates. Estimates of N_e for *F. shoupi* PW and *F. placidus* LC were unreliable ($N_e = \infty$), likely due to a lack of signal in the dataset (Table 4). Tests for historical bottlenecks indicated that several populations had experienced recent declines in numbers. The mode-shift test showed a shift in the spectrum of allele frequencies in four of six populations. All MC populations of *F. shoupi* and the *F. placidus* LC population showed evidence of a recent bottleneck. Allele frequency spectrums for *F. shoupi* PW and *F. placidus* BF were normal. The Wilcoxin test for excess heterozygosity was not significant for any of the examined populations.

Population Differentiation SNPs. Estimates of pairwise population differentiation showed little evidence of genetic structuring among populations of *F. shoupi* (Table 5). Pairwise F_{ST}' and Φ_{ST} estimates between populations of *F. shoupi* were less than or equal to 0.001 and 0.01, respectively. Populations of *F. placidus* showed moderate levels of differentiation from *F. shoupi*. The average F_{ST}' and Φ_{ST} estimates for *F. placidus* LC/*F. shoupi* populations were 0.057 and 0.078, respectively; *F. placidus* BF/*F. shoupi* averaged 0.018 and 0.033, respectively.

Results from the Bayesian assignment analyses suggested that *K*=3 provided the best fit for the data as determined by (Ln Pr(X|*K*). We assessed membership assignments for a range of values of *K* (*K*=2-5). Permutated membership coefficients for all values of *K* are shown in Figure 5. At *K*=2, admixture was observed across all populations; however, there was no geographic partitioning of individuals. For *K*=3-5, the *F. placidus* LC population is assigned to a separate group from *F. shoupi* and *F. placidus* BF. The *F. placidus* BF population showed split membership at K=3, but clustered with *F. shoupi* for K=4-5. To detect further structuring of genetic variation within *F. shoupi*, we performed a separate analysis which only included the four sampled *F. shoupi* populations as suggested by Vaha et al. (2007); however, no further subdivision was found (results not shown). The BIC curve for the DAPC clustering analysis that included *F. shoupi* and *F. placidus* indicated that two (K=2) was the optimal number of explanatory clusters (Fig. 6). Cluster 1 included all populations of *F. shoupi* as well as the *F. placidus* BF population. Cluster 2 only included individuals form the *F. placidus* LC population.

The divergence time estimates for the split between *F. shoupi* at MC and PW were similar for models with migration (M1) and without migration (M2). The estimate for model M1 was 69,638 YBP (Fig. 7) and M2 was 65,971 YBP (95% C.I. 32,828 – 100,400). The marginal log likelihood values were greater for model M1 than for model M2 (Supplement 3) and the relative posterior probability for M1 was 1, indicating a greater likelihood for the model with no post-divergence gene flow.

Discussion

The Nashville crayfish has demonstrated remarkable genetic resiliency despite a history of habitat loss, impoundment, and contamination in the MC watershed. Our results suggest that MC populations have recently experienced population declines; however, substantial genetic variation is indicated by both mitochondrial haplotypes and genome-wide SNP datasets. Genetic connectivity among populations has been maintained and may reflect this species' capacity for dispersal and rapid colonization (Carpenter 2002). Our molecular investigation of the newly discovered PW population shows that this population shares a recent connection with *F. shoupi* in the MC watershed. Despite this similarity, the PW population also possesses unique genetic variation in both the mitochondrial and nuclear genome, improving the long-term adaptive outlook for this species. Results from this study provide the first glimpse of population-level genetic variation and structure in *F. shoupi* and give insights into the genetic consequences of the management history of this species.

Genetic Variation and Demographic History. Allele frequency distributions for genome-wide SNPs indicate that MC populations have experienced recent bottlenecks. The mode-shift test detected an upward shift in the distribution of allele frequencies, indicative of a recent population bottleneck, at all three MC sites. Despite evidence of recent population declines, estimates of genetic variation at mitochondrial haplotypes across MC populations were still high compared to estimates obtained from other at-risk North American crayfish species. Comparable surveys of mitochondrial gene diversity (also referred to as haplotype diversity) in the genus *Cambarus* averaged 0.360 and 0.297, for sampled populations of *C. jezerinaci* and *C. parvoculus*, respectively (Thoma and Fetzner 2008). Gene diversity estimates for our MC samples averaged 0.826; estimates obtained here are similar to estimates from native populations of the invasive *Procambarus clarkii* (H_{GD} averaged 0.844; Yi et al. 2018). Rapid recovery following bottleneck events can limit the loss of genetic variation (Nei 1975). Experimental evidence has shown that *F. shoupi* can recover quickly following population depletion; their ability to rapidly colonize and recover from loss of numbers may have mitigated the loss of genetic variation that can accompany demographic bottlenecks (Carpenter 2002). This study is among the first to use GBS derived SNPs to assess patterns of population-level genetic variation in crayfish, so it is not possible to place our estimates of heterozygosity in the context of similar studies. Furthermore, bioinformatic parameters can influence estimates of genetic variation, confounding cross-study comparisons. We did see consistent patterns of genetic variation across sampling sites within MC. Individuals sampled from the MC-U site demonstrated higher levels of genetic variation than downstream sites (MC-M and MC-L). This same pattern was observed after correcting for

unequal sample sizes (Supp. 2), although MC-U had the lowest estimates of private allelic richness of the three sampled sites. Genetic variation measures for both mitochondrial and nuclear markers were slightly higher for the *F. shoupi* population at PW than for any of the sampled *F. shoupi* populations in MC.

Despite little evidence of population differentiation, the *F. shoupi* PW population was more variable and harbored unique genetic variation not present in *F. shoupi* populations from MC sites. The *F. shoupi* PW population possessed the greatest number of unique mitochondrial haplotypes (5 unique haplotypes; Fig 4) and high estimates of genetic variation for genome-wide SNPs. It is possible that the Pickwick population has retained genetic variation because it has not experienced the levels of disturbance and siltation that has been documented along MC. Despite impoundments along the stretch of the Tennessee river, water quality and flow regimes near the Pickwick Tailwater have been maintained. Greater demographic stability is also supported by the mode-shift test that did not detect a shift in the allele frequency spectrum for PW.

Estimates of effective population sizes across *F. shoupi* populations at MC were markedly low; the average N_e across *F. shoupi* populations in MC was 23.7. It is difficult to establish a baseline for comparison of N_e estimates in crayfish as few empirical studies have investigated contemporary N_e from multi-locus molecular data. A recent study using microsatellite genotypes to estimate N_e across populations of *Cambarus pristinus* recovered much larger N_e values than our estimates for *F. shoupi*. Point estimates of N_e for populations of *C. pristinus* ranged from 137 to 1348 using the same LD method applied here (Grubb 2020). Several factors may reduce N_e relative to census numbers (N); demographic bottlenecks will reduce the N_e/N as will variation in reproductive success. Little is known about the reproductive life-history of *F. shoupi*. Many crayfish demonstrate r-selected life histories with females carrying between 200-400 eggs that hatch every spring (Bouchard 1976). Organisms with high fecundity typically experience high mortality in early life-stages so that very few parents successfully pass down genetic material to the next generation. This high variance in reproductive success reduces the N_e /N (termed the Hedgecock effect; Hedrick 2005). The relationship of the Ne/N ratio as a function of the variance in the number of progeny per parent (V_k) is estimated as follows (Wright 1938):

$$\frac{N_e}{N} \approx \frac{4}{V_k + 2}$$

In animals that exhibit this “sweepstake reproductive success”, Ne/N can be reduced to a small fraction (<<0.01; Hedgecock and Pudovkin 2011).

Population Structure. Molecular results indicate that *F. shoupi* populations throughout MC have maintained geneflow despite documented habitat disturbances and impoundments. Historically, drainage from MC into the Cumberland River was impeded by a series of three dams. As of 2018, all dams were removed, restoring connectivity for the entire length of MC downstream to the Cumberland River (Cumberland River Compact 2018). Both mitochondrial haplotypes and genome-wide SNPs suggest that genetic connectivity has been maintained along MC. Analysis of molecular variance indicated that variation in mitochondrial haplotypes could not be explained by differences in sampling sites within the MC watershed. Pairwise measures of population differentiation (F_{ST} and ϕ_{ST}) were less than 0.01 for all comparisons; for reference, values of F_{ST} less than 0.05 are generally interpreted as little structuring of genetic differentiation (Hartl and Clark 1997).

The *F. shoupi* PW population was genetically similar to populations in MC; however, we did detect some molecular signatures of recent divergence. The PW population possessed genetic variation for both mitochondrial haplotypes and nuclear SNPs that was not present in any of the sampled MC sites. Our Bayesian MSci analysis suggested that connectivity between these sites was severed less than 100,000 YBP. A recent connection between populations in the two drainages is also supported by morphological similarities in the gonopod structure of first form males, and the presence of secondary dentation of the dactyl. However, differences in color patterns and morphological characters have been observed between the two populations. Specifically, the saddle pigmentation is predominately mottled in the PW population and solid in the MC population, the proportional width of the eyes is greater in PW than in MC, finally in males the presence of setae within the palm is more pronounced in PW than in MC.

One explanation for the disjunct distribution of *F. shoupi* is that the MC and PW are relictual populations of what historically was a more widespread distribution throughout the Cumberland and Tennessee watersheds. Reports of *F. shoupi* in the Elk and Harpeth River Systems lend support to the hypothesis of a now extant, larger distribution of *F. shoupi* in Middle Tennessee (Barrociere 1986). Populations at Big Creek and the South Harper were presumably extirpated. The population at Richland Creek may have been replaced by the closely related *F. placidus*, which is the predominant, stream dwelling crayfish in this basin (Bouchard 1984). The disjunct distribution of *F. shoupi* is shared by other freshwater taxa; specifically, there are about 36 mussel species demonstrating disjunct distributions that are endemc to the Tennessee and Cumberland Drainages (Haag and Cicerello 2016). This shared assemblage, long recognized as unique and first termed by Ortmann (1918) as the Cumberlandian Fauna, may be the result of the two drainages previously being isolated from the Mississippi and Ohio Rivers but connected to each other (Galloway et al. 2011). The presence of *F. shoupi* in the lower Tennessee River may represent a remnant population of this former connection. Alternatively, the newly found PW population could also be the result of a recent, human-mediated introduction as a bait-bucket release, however our molecular results do not support this. Introduced populations are typically initiated by a small number of founders and, as a result, demonstrate patterns of genetic variation indicative of a population bottleneck (i.e. an excess of heterozygosity and shifts in the

distribution of allele frequencies; Fitzpatrick et al. 2012). The PW populations was the only *F. shoupi* population that did not have a mode-shift in its allele frequency distribution. Colonizers of introduced populations will pass along a small subset of genetic variation from their source, so that genetic variation in the introduced site will be a subset of what is observed in the native population range. The PW population did not reveal any evidence of reduced genetic variation for either mitochondrial haplotype or genome-wide SNP surveys. Furthermore, the PW population possessed unique genetic variation that could not be traced back to any of our MC samples.

The inclusion of samples from multiple *F. placidus* populations highlights the need for a thorough molecular-based evaluation of this group. *Faxonius placidus,* as currently described, is a widespread species associated with the Cumberland, Barren, Tennessee, and Lower Ohio rivers (Poly and Wetzel, 2003). Our phylogenetic reconstructions based on mitochondrial haplotypes and pairwise measures of differentiation based on SNPs reveal that *F. placidus* is paraphyletic with respect to *F. shoupi*; the *F. placidus* BF population is more closely related to a monophyletic *F. shoupi* than it is to the *F. placidus* population at the type locality (population LC). Integrative species delimitation methods are currently being used to investigate the relationship between populations of *F. placidus* throughout Tennessee (Hildreth et al. in prep).

Implications for management. Faxonius shoupi has served as the flagship species for the protection and restoration of MC. Prior to the discovery of the PW population, *F. shoupi* was widely considered to be the only known species whose distribution was restricted entirely to the Nashville basin (Barrociere 1986). Habitat in MC has been compromised by loss of riparian buffer zones, contamination from sewage or chemical spills, and siltation (Miller and Hartfield, 1985). Federal listing of *F. shoupi* has resulted in improvements to this watershed that have not only benefited *F. shoupi*, but have also restored habitat to a number of co-distributed freshwater fauna. Despite evidence of demographic bottlenecks*, F. shoupi* have maintained moderate levels of genetic variation and connectivity throughout MC; the ability of this species to rapidly colonize and expand in numbers may have mitigated some of the loss of genetic variation that can accompany sustained demographic declines.

The discovery of a second *F. shoupi* population in the Tennessee River improves the long-term outlook for this species. While it is not possible to conclusively determine the origins of the PW population, our results demonstrate that the PW population is genetically variable and should be prioritized for conservation planning along with *F. shoupi* in the MC watershed. Although this species is considered to be a recovery success, the distribution of *F. shoupi* is still restricted in its range and vulnerable to declines should quality habitat be lost. The disjunct population at PW should be included in management planning for *F. shoupi;* this site is at risk of degradation associated with urbanization and agriculture, the same factors that threaten MC populations (Bizwell and Mattingly 2010). Potential threats unique to the PW include the introduction of invasive species and pathogens from the Tombigbee waterway, pollution from upstream sources, and changes to flow regimes from impoundments. Further sampling along the mainstem of the Tennessee and Cumberland rivers is needed to determine the full extent of the range of *F. shoupi;* the discovery of *F. shoupi* at PW indicates that this species may inhabit deeper riverine habitat than previously thought. Clarification of the full distribution of this species will be useful when reviewing the conservation status of this species.

Declarations

Competing Interests

The authors have no competing interests to declare.

Funding

Funding for this project was provided by the USFWS and through the Undergraduate Research and Creative Activities Program (URECA) at Tennessee Tech University

Author Contributions

All authors contributed to study design, conception and sampling. CH and PH performed the molecular work. CH performed the computational analysis and the wrote the first draft of the manuscript. CW and PH further edited the manuscript.

Conflicts of Interest/Competing Interests

The authors have no competing interests to declare.

Availability of data and material

Raw sequence reads obtained from GBS sequencing are available at NIH Sequence Read Archive: Accession Number XXXX. Mitochondrial sequences from Sanger sequencing are a deposited on Genbank: Accession Numbers XXXXX.

Animal Research N/A

Ethics approval N/A

Consent to participate N/A

Consent for publication N/A

Acknowledgments

We thank Philip Hedrick and Gerry Dinkins for comments on earlier versions of the manuscript. We also thank Jeffrey Simmons for assistance with field collections. Edward Lisic and the URECA Team grant program provided partial funding for this work.

References

1. Barrociere LJ (1986) The ecological assessment and distribution status of the Nashville Crayfish, *Orconectes shoupi*. Thesis, Tennessee Technological University
2. Bizwell EA, Mattingly HT (2010) Aggressive Interactions of the Endangered Nashville Crayfish, *Orconectes shoupi*. Southeast Nat 9:359–372
3. Bouchard RW (1972) A contribution to the knowledge of Tennessee crayfish. Dissertation, University of Tennessee, Knoxville
4. Bouchard RW (1974) Crayfishes of the Nashville Basin, Tennessee, Alabama and Kentucky (Decapoda, Astacidae). The Association of Southeastern Biologists Bulletin 21:41
5. Bouchard RW (1976) Geography and ecology of crayfishes of the Cumberland Plateau and Cumberland Mountains, Kentucky, Virginia, Tennessee, Georgia, and Alabama. Part I. The genera of *Procambarus* and *Orconectes*. Freshwater Crayfish 2:563–584
6. Bouchard RW (1984) Distribution and status of the endangered crayfish Orconectes shoupi (Decapoda: Cambaridae). Report, USFWS
7. Carpenter J (2002) Recolonization of depleted areas by the Nashville Crayfish. Report, USFWS
8. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Molecular ecology 22:3124–3140
9. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4(1):s13742–015
10. Cornuet JM, Gordon L (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014
11. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources 14:209–214
12. Durand JB (1960) Limb regeneration and endocrine activity in the crayfish. The Biological Bulletin 118:250–261
13. Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources 4:359–361
14. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one 6:e19379
15. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491
16. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources 10:564–567
17. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology 14:2611–2620
18. Fitzpatrick BM, Fordyce JA, Niemiller ML, Reynolds RG (2012) What can DNA tell us about biological invasions? Biol Invasions 14:245–253
19. Fleming RS (1939) The larger crustacea of the Nashville region. Journal of the Tennessee Academy of Science 13:261–264
20. Folmer O, Hoeh WR, Black MB, Vrijenhoek RC (1994) Conserved primers for PCR amplification of mitochondrial DNA from different invertebrate phyla. Mol Mar Biol Biotech 3:294–299
21. Galloway WE, Whiteaker TL, Ganey-Curry P (2011) History of Cenozoic North American drainage basin evolution, sediment yields, and accumulation in the Gulf of Mexico basin. Geosphere 7:938–973
22. Gautier M et al (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Molecular ecology 22:3165–3178
23. Grubb B (2020) Phylogeography and population genetics of a headwater-stream adapted crayfish, *Cambarus pristinus* (Decapoda: Cambaridae), from the Cumberland Plateau in Tennessee. M.S. Thesis Austin Peay State University
24. Hartl DL, Clark AG (1997) Principles of population genetics, vol 116. Sinauer associates, Sunderland

25. Loiseau MT, Pudlo P, Kerdelhué C, Estoup A (2013) Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. Mol Ecol 22:3766–3779
26. Haag WR, Cicerello RR (2016) A distributional atlas of the freshwater mussels of Kentucky. Scientific and Technical Series 8, Kentucky State Nature Preserves Commission, Frankfort
27. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In Nucleic Acids Symposium Series 41:95–98
28. Hartl DL, Clark AG (1997) Principles of Population Genetics, 3nd edn. Sinauer Associates, Inc, Sunderland
29. Hasegawa M, Kishino H, Yano T (1985) Dating the human-ape split by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174
30. Hedrick P (2005) Large variance in reproductive success and the Ne/N ratio
31. Evolution, 59:1596–1599
32. Hedgecock D, Pudovkin AI (2011) Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. Bull Mar Sci 87:971–1002
33. Hobbs HH Jr (1969) On the distribution and phylogeny of the crayfish genus Cambarus. The distributional history of the biota of the southern Appalachians. Part I: Invertebrates. Research Division Monograph 1:93–178
34. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806
35. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genom 11:94
36. Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered 89:238–247
37. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549
38. Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60:2399–2402
39. Meirmans PG, Hedrick PW (2011) Assessing population structure: FST and related measures. Molecular ecology resources 11:5–18
40. Miller AC, Hartfield PD (1987) A Study of *Orconectes shoupi*, Mill Creek Basin, Tennessee, 1985. Mississippi Museum of Natural Science Jackson, Mississippi
41. Miller MA, Pfeiffer W, Schwartz T (2011) The CIPRES science gateway: a community resource for phylogenetic analyses. In Proceedings of the 2011 TeraGrid Conference: extreme digital discovery 1–8
42. Mlinarec J, Porupski I, Maguire I, Klobučar G (2016) Comparative karyotype investigations in the white-clawed crayfish *Austropotamobius pallipes* species complex and stone crayfish A. Torrentium (Decapoda: Astacidae). J Crustac Biol 36:87–93
43. Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 9:1–10
44. Ortmann AE (1918) The naiads (freshwater mussels) of the upper Tennessee drainage. With notes on synonymy and distribution. Proceedings of the American Philosophical Society 57:521–626
45. Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26:419–420
46. Piry S, Luikart G, Cornuet JM (1999) Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. J Hered 90:502–503
47. Poly WJ, Wetzel JE (2003) Distribution and taxonomy of three species of *Orconectes* (Decapoda: Cambaridae) in Illinois, USA. J Crustac Biol 23:380–390
48. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
49. Racine JS (2012) RStudio: a platform-independent IDE for R and Sweave. 167–172
50. Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG (2014) STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus 3:1–3
51. Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24:2498–2504
52. Taylor CA, Knouft JH (2006) Historical influences on genital morphology among sympatric species: gonopod evolution and reproductive isolation in the crayfish genus *Orconectes* (Cambaridae). Biol J Lin Soc 89:1–12
53. Thoma RF, Fetzner JW Jr (2008) Taxonomic status of *Cambarus* (*Jugicambarus*). In: *jezerinaci*, spiny scale crayfish (Powell River crayfish). Report, Virginia Department of Game & Inland Fisheries Richmond, Virginia
54. US Fish and Wildlife Service (1987) Nashville crayfish recovery plan. US Fish and Wildlife Service, Atlanta

55. US Fish and Wildlife Service (2018) Species Status Assessment Report for the Nashville Crayfish (*Orconectes shoupi*). US Fish and Wildlife Service, Atlanta
56. VÄHÄ JP, Erkinaro J, Niemela E, Primmer CR (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. Mol Ecol 16:2638–2654
57. Rannala B, Yang Z (2017) Efficient Bayesian species tree inference under the multispecies coalescent. Syst Biol 66:823–842
58. Reynolds J, Souty-Grosset C, Richardson A (2013) Ecological roles of crayfish in freshwater and terrestrial habitats. Freshwater Crayfish 19:197–218
59. Silliman K, Indorf J, Knowlton N, Browne B, Hurt C (2021) Base-substitution mutation rate across the nuclear genome of Alpheus snapping shrimp and the timing of isolation by the Isthmus of Panama. BMC Ecology and Evolution (in press)
60. Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24:2498–2504
61. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight *matrix* choice. Nucleic acids research 22:4673–4680
62. Waples RS, Do CH (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular ecology resources 8:753–756
63. Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular ecology resources 14:209–214
64. Waples RS (2016) Making sense of genetic estimates of effective population size. 4689–4691
65. Wetzel JE, Poly WJ, Fetzner JW Jr (2004) Morphological and genetic comparisons of golden crayfish, *Orconectes luteus*, and rusty crayfish, *O. rusticus*, with range corrections in Iowa and Minnesota. J Crustac Biol 24:603–617
66. Wright S (1938) Size of population and breeding structure in relation to evolution. Science 87:430–431
67. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst Biol 60:150–160
68. Yang WS, Noh JH, Jeon NJ, Kim YC, Ryu S, Seo J, Seok SI (2015) High-performance photovoltaic perovskite layers fabricated through intramolecular exchange. Science 348:1234–1237
69. Yi S, Li Y, Shi L, Zhang L, Li Q, Chen J (2018) Characterization of Population Genetic Structure of red swamp crayfish, *Procambarus clarkii*, in China. Scientific reports 8:1–11

Tables

Table 1 – Sample sites for populations of *F. shoupi*, *F. placidus*, and *F. durelli* included for mitochondrial sequencing and SNP genotyping. Populations denoted by an asterisk indicate the type locality for respective species. Mapped coordinates are shown in Figure 1. Sample size (*N*) refers to the number of individuals included for mitochondrial COI sequencing and GBS-based SNP genotyping, respectively.

Population	Drainage	Coordinates	
***F. shoupi* MC-U**	Mill Creek (Upper)	35.95308	-86.6686
F. shoupi* MC-M	Mill Creek (Middle)	36.05942	-86.6717
***F. shoupi*MC-L**	Mill Creek (Lower)	36.14497	-86.7129
***F. shoupi*PW**	Tennessee River (Pickwick Tailwater)	35.06425	-88.2594
***F. placidus* LC ***	Lower Cumberland-Old Hickory Lake	36.21910	-86.3093
***F. placidus* BF**	Collins River (Barrens Fork)	35.67462	-85.7767
***F. placidus* PW**	Tennessee River (Pickwick Tailwater)	35.06425	-88.2594
F. durelli*	Buffalo River	35.28885	-87.7625

Table 2. Within population genetic variation summary statistics for *Faxoniusshoupi* and two populations of *F. placidus*. N_a and N_b indicate the samples sizes for mitochondrial COI haplotypes and SNP genotypes, respectively. N_{Hap} indicates the number of unique COI haplotypes from each population. H_{GD} indicates the genotypic diversity for COI haplotypes. π_M and π_N are nucleotide diversity estimates for COI haplotypes and genomic SNPs, respectively. Het_{Obs} is the mean observed heterozygosity, Het_{Exp} is the mean expected heterozygosity under Hardy-Weinberg

equilibrium. Mean heterozygosity (Het_{obs} and Het_{exp}) and π_N were calculate for both variant sites and across all sites. F_{IS} is the mean value of the inbreeding coefficient across variant SNP loci. H_{GD} and π_M were not calculated for *F. placidus*populations due to limited sampling. A_{RN} and P_{RN} indicate the mean allelic richness and private allelic richness using rarefaction to adjust for sample size (g=25).

	MtDNA Haplotypes				Genomic SNPs								
Population	N_a	N_{Hap}	H_{GD}	π_M	N_b	Variant Sites Only			All Sites (Variant + Fixed)			F_{IS}	A_{RN}	P_{RN}
						Het_{Obs}	Het_{Exp}	π_N	Het_{Obs}	Het_{Exp}	π_N			
F. shoupi MC-U	25	8	0.813	0.010	24	0.073	0.071	0.073	0.00050	0.00049	0.00050	0.025	1.948	0.094
F. shoupi MC-M	21	12	0.852	0.011	25	0.046	0.047	0.049	0.00033	0.00033	0.00035	0.017	1.915	0.120
*F. shoupi*MC-L	26	9	0.812	0.004	16	0.048	0.047	0.048	0.00034	0.00033	0.00034	0.020	1.850	0.118
*F. shoupi*PW	27	11	0.909	0.011	27	0.065	0.077	0.079	0.00046	0.00055	0.00056	0.094	1.901	0.173
*F. placidus*LC	3	1	—	—	18	0.087	0.066	0.068	0.00057	0.00043	0.00045	-0.028	—	—
F. placidus BF	3	2	—	—	18	0.044	0.038	0.039	0.00028	0.00024	0.00025	-0.005	—	—

Table 3. Analysis of Molecular Variance (AMOVA) for populations of *Faxoniusshoupi* using mitochondrial COI haplotypes. Populations were grouped according to drainage so that group 1 included all populations in Mill Creek (MC-U, MC-M, and MC-L) and group 2 only included the PW population.

Source of Variation	d.f.	Sum of squares	Percentageof Variation	Statistics
Among Drainages	1	2.21	3.16%	F_{ST} = 0.031
Among Populations within Drainages	2	1.94	0.00%	F_{SC} = 0.001
Within Populations	95	93.50	98.4%	F_{CT} = 0.031

Table 4. Demographic parameter estimates for demographic parameters for *Faxoniusshoupi* and sampled populations of *F. placidus*. Contemporary effective population sizes (N_e) for populations of *F. shoupi* and *F. placidus* based on the linkage disequilibrium (LD) method as performed in NeEstimator (Waples et al. 2014). 95% confidence intervals (CI) are based on non-parametric bootstrapping. Results from the Wilcoxin test (1-tailed test for heterozygosity excess) and the mode-shift test as performed by Bottleneck (Piry et al. 1999).

Population	Ne (0.02)	95% CI	Wilcoxin Test (IAM)	Mode Shift Test
***F. shoupi* MC-U**	19	15 – 23	n.s.	shifted
***F. shoupi*MC-M**	41	29 – 70	n.s.	shifted
***F. shoupi*MC-L**	30	21 – 48	n.s.	shifted
***F. shoupi*PW**	∞	- ∞	n.s.	normal
***F. placidus* LC**	19	16 -27	n.s.	shifted
***F. placidus*BF**	∞	-∞	n.s.	normal

Table 5. Pairwise measures of population differentiation for all sampled *F. shoupi* and *F. placidus* populations based on genomic SNPs. Numbers below diagonal indicate average F_{ST}' values. Numbers above diagonal indicate $\Phi_{ST.}$

	F. shoupi				*F. placidus*	
	MC-U	MC-M	MC-L	PW	LC	BF
F. shoupi MC-U	—	0.006	0.006	0.005	0.077	0.043
F. shoupi MC-M	0.000	—	0.007	0.008	0.082	0.042
F.shoupi MC-L	0.000	0.001	—	0.009	0.092	0.029
*F. shoupi*PW	0.000	0.000	0.001	—	0.061	0.018
F. placidus A	0.056	0.051	0.076	0.044	—	0.046
F. placidus BF	0.023	0.022	0.020	0.006	0.026	—

Figures

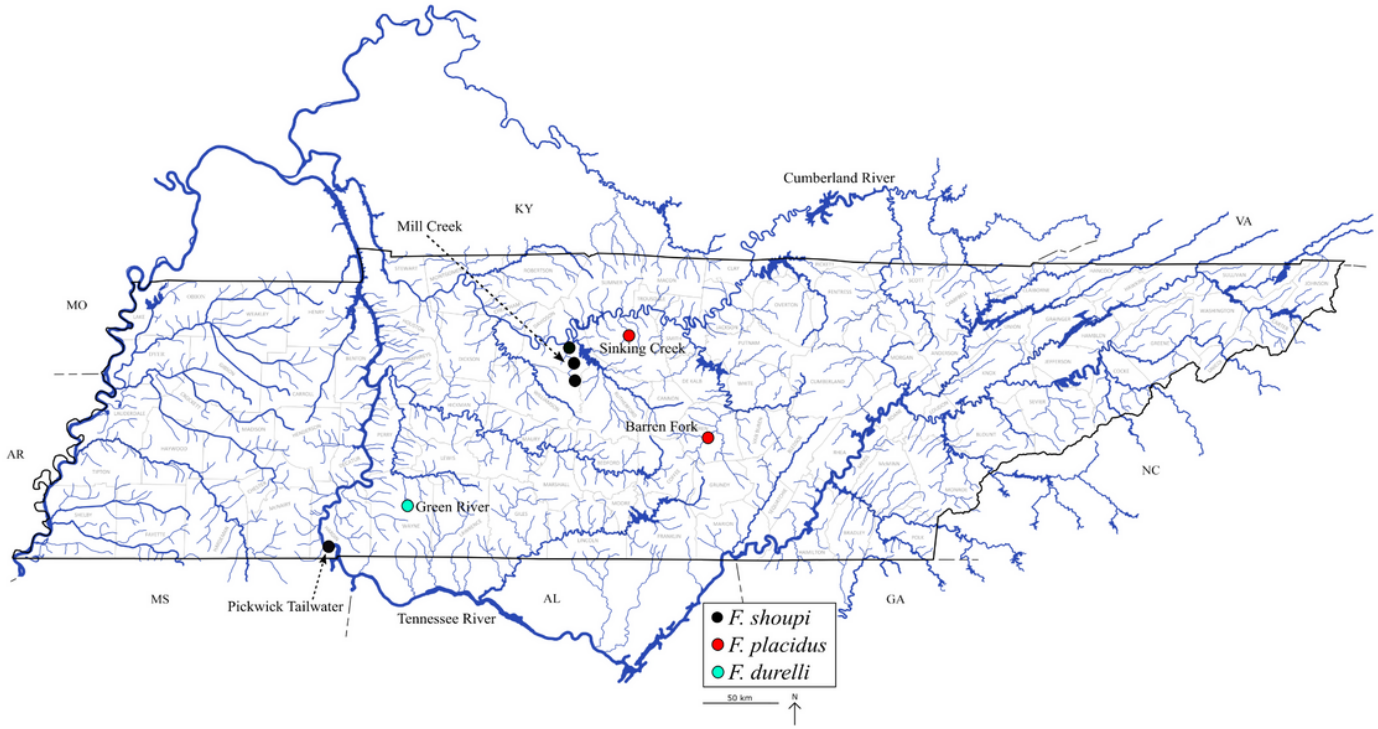

Figure 1

Map showing all sites sampled for F. shoupi, F. placidus, and F. durelli. Sites are color coded by species as indicated in the legend. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 2

Images of shoupi (first form males) from Sevenmile Creek (tributary to Mill Creek) and Pickwick Tailwater (Tennessee River).

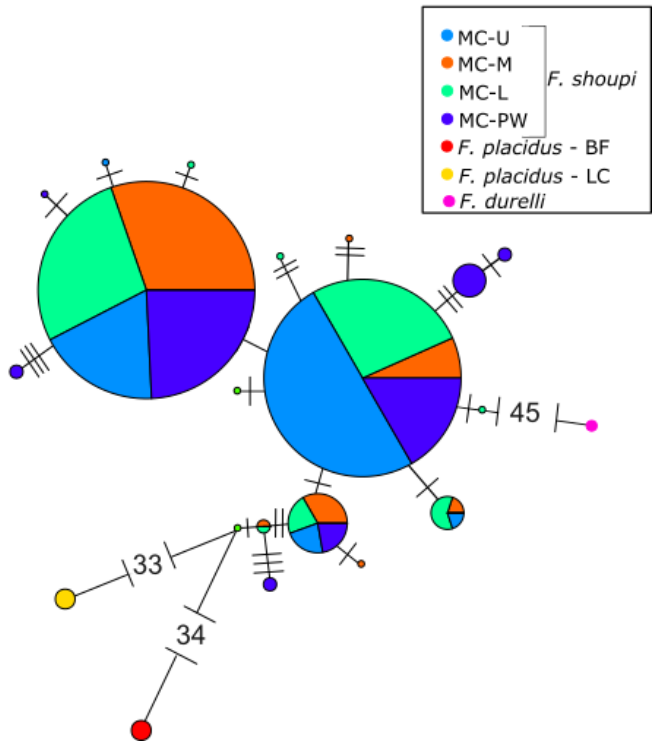

Figure 3

Minimum Spanning Haplotype Network illustrating the relationships between 20 mitochondrial COI haplotypes identified from F. shoupi, F. placidus, and F. durelli individuals. The size of the circles corresponds to the number of individuals represented by each haplotype. Colors correspond to sampled populations. Lines between circle indicate the number of mutational steps between each haplotype.

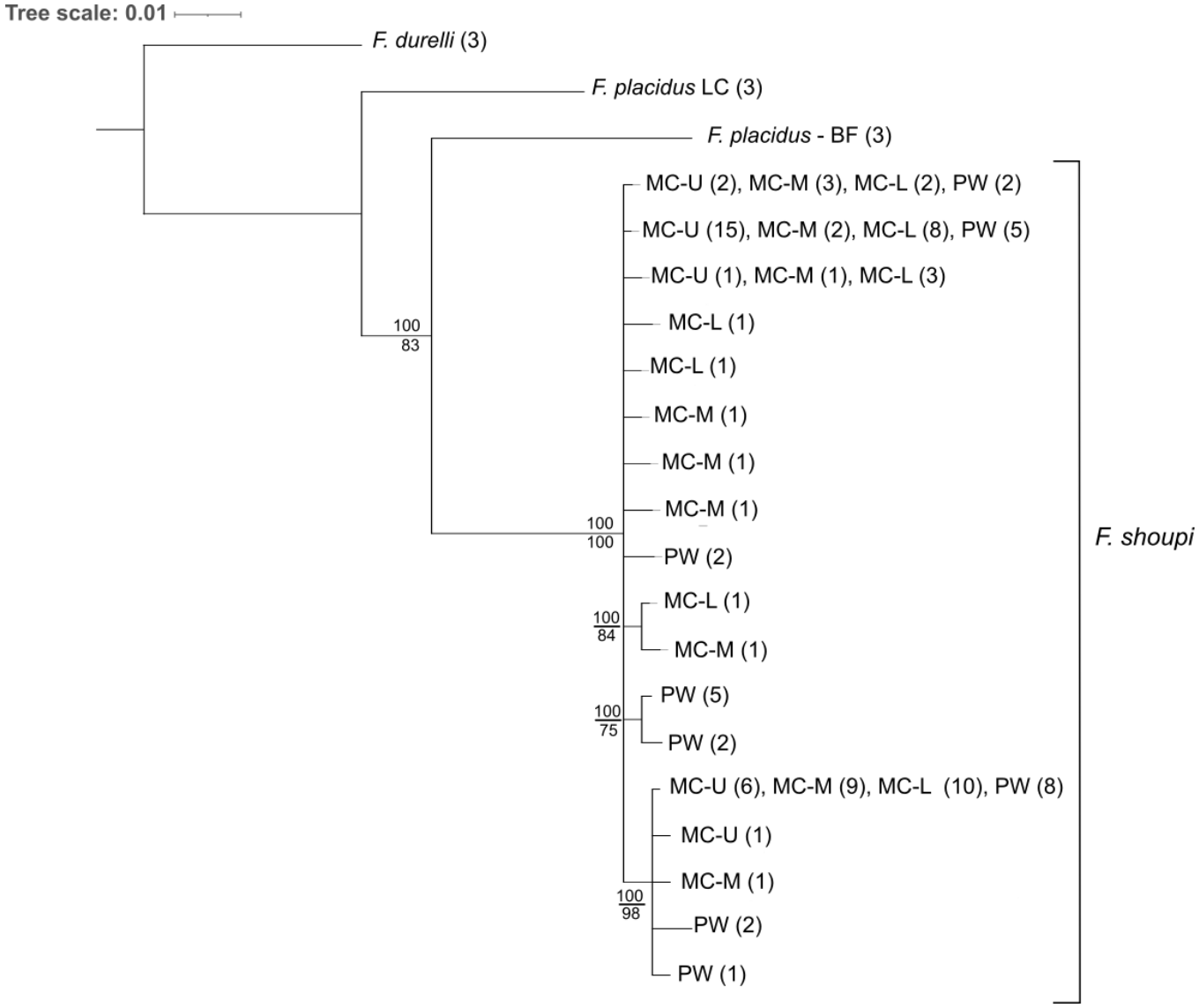

Figure 4

Bayesian phylogenetic reconstruction of 21 unique mitochondrial COI haplotypes representing four sampled populations of F. shoupi as well as representative sequences from F. placidus and F. durelli. Populations IDs are indicated at tips with haplotype copy number shown in parentheses. Numbers at nodes indicate posterior probability support (below line) and ML bootstrap support (above line).

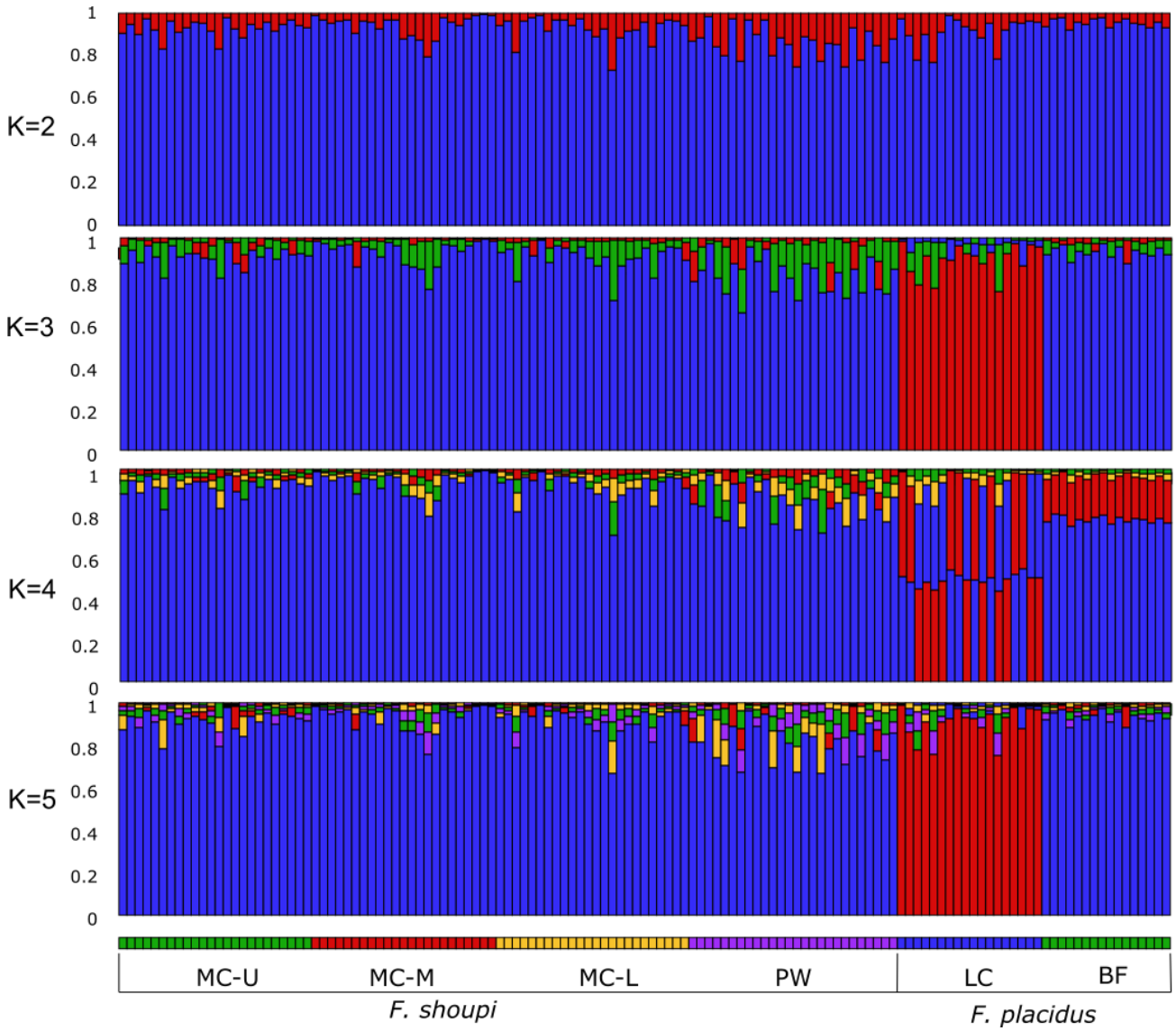

Figure 5

Results of Bayesian cluster analysis as performed by STRUCTURE. Bar plot indicates assignment probabilities for 131 individuals based on the SNP loci generated from GBS. Bar plots shown are for K=2 through K=5.

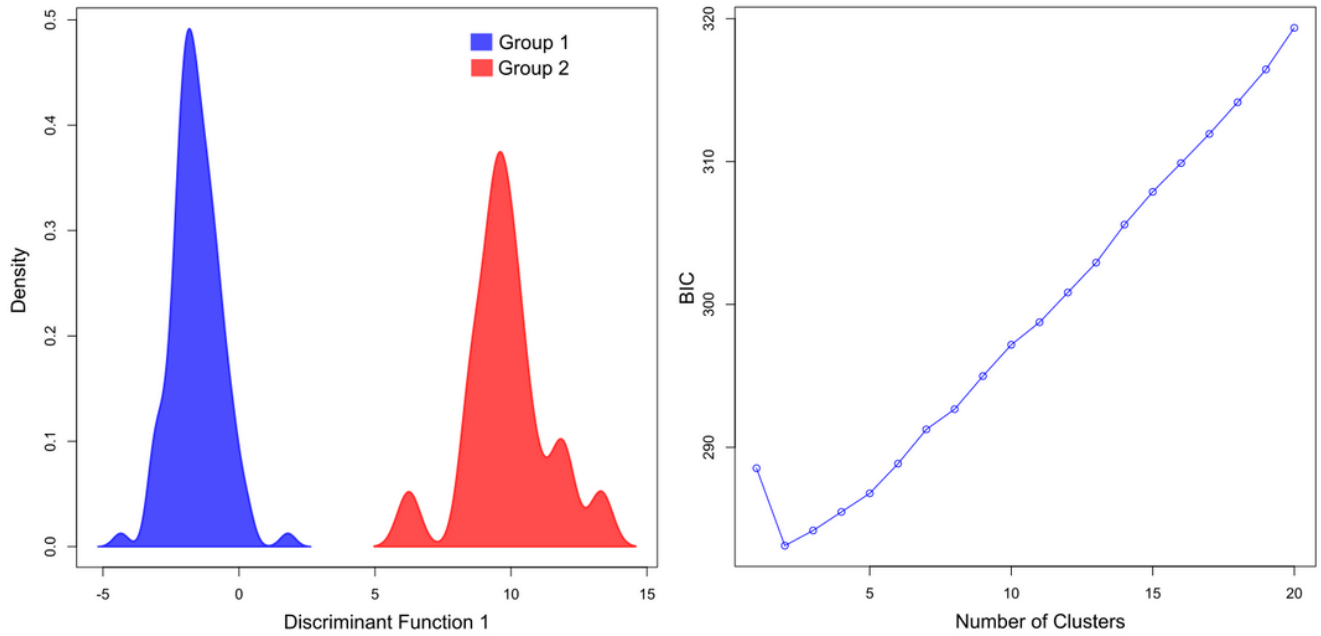

Figure 6

Results of discriminate analysis of principal components (DAPC) based on SNP genotypes. A) Density plot of discriminant function when all F. shoupi and F. placidus individuals are included in the analysis. Group 1 includes all individuals from F. shoupi populations and from the F. placidus BF population. Group 2 includes all individuals from the F. placidus LC population B) Bayesian Information Criterion value for K=1-20.

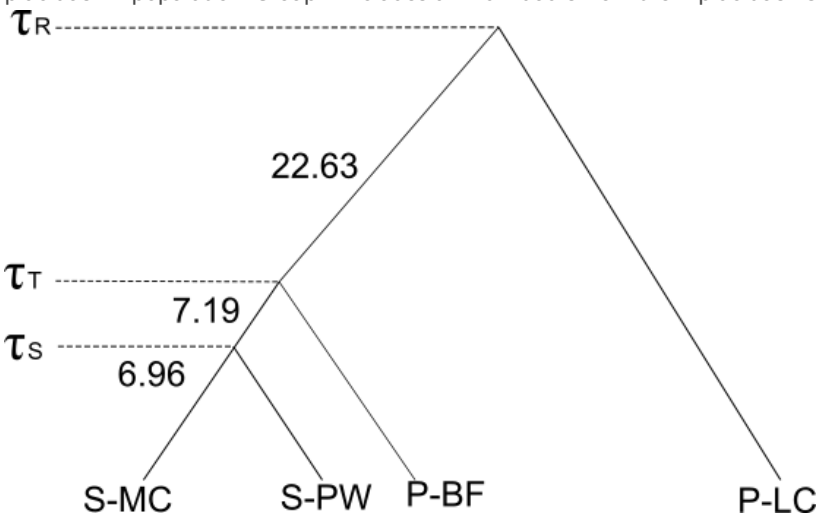

Figure 7

A Model M1 used for divergence time estimates using the Bayesian multispecies-coalescent-with-introgression (MSci) method as implemented in BPP (Flouri et al., 2018). Values on branches are posterior means of divergence time (YBP) estimates x 10-4

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supp1bioinformatics.docx
- Supp2rarefaction.docx
- Supp3BPPM2.docx