University of North Carolina at Asheville
Journal of Undergraduate Research
Asheville, North Carolina
May 2021

Using Genetic Data to Determine the Origins of a Potential Introduced Population of Mottled Sculpin (*Cottus bairdii*) Across the Eastern Continental Divide Near Asheville, NC

Isabel Johnson
Biology Department
University of North Carolina at Asheville
One University Heights
Asheville, North Carolina 28804 USA

Faculty Advisors: Dr. R. Graham Reynolds and Dr. David Gillette

Abstract

The Mottled Sculpin (*Cottus bairdii*) is a widely distributed benthic freshwater fish native to major rivers in the Mississippi River drainage of the US and Canada. In North Carolina, the species is found in Gulf-coast draining rivers on the western side of the Eastern Continental Divide. It is particularly widespread and abundant in the French Broad River Drainage. Nevertheless, several populations are known from the Atlantic-draining Broad River Basin. These populations all occur near the headwaters of the Broad River Drainage, just to the east of the Eastern Continental Divide. While Mottled Sculpin has been widely introduced elsewhere due to their use as bait, no consensus has been reached regarding the status of the Broad River sculpin in North Carolina, particularly whether they are native or introduced. Twenty samples from each of two populations were collected, including the presumed native Cane Creek population, a tributary of the Haw River, and the presumed introduced population from the Green River, a tributary of the Broad River. Two loci of mitochondrial DNA were sequenced, generating about 1,500 base pairs of sequence data combined. My hypothesis was that if the Broad River populations are introduced, they should share mitochondrial haplotypes with the French Broad population. Phylogenetic analysis revealed a significant number of shared haplotypes between the two populations, thus revealing that the Broad River population was introduced. In addition, the Haw Creek population revealed a divergence within its population, which is a curious observation on which warrants further investigation.

1. Introduction

Non-native species are a major threat to biodiversity, and non-native fish presence has been found to lead to long-term decreases in both *alpha* and *beta* diversity in freshwater ecosystems, therefore, it is important to understand whether a species has been introduced in an area so that it can be properly managed, and potential impacts on native species can be mitigated.[1] As populations require genetic diversity to maintain fitness across generations, it is important to have an understanding of the degree of genetic diversity within both native and introduced populations so that the health and success of the invasive population can be better understood.[2] Analyzing the genetic diversity of introduced populations can bring insights into their colonization and establishment, as well as potential evolutionary responses to novel environments.[3]

Previous studies comparing population genetics of native and introduced populations have found varying degrees of genetic variation between populations. If a large degree of differentiation is found between native and introduced populations, it may be an indication of bottleneck effects in the introduced population.[4] High levels of genetic variation in the introduced population may indicate a larger introduced population from the native range, or a larger degree of gene flow.[5] This indicates a high degree of crossover between populations, which if occurring, is important information for managers.[6] In addition, lower degrees of diversity in the invasive populations may indicate low

resiliency in the population, so having an understanding of this can give insight into how serious the invasive population's presence in the non-native habitat may be.[7]

Examining the genetic diversity of introduced fish species can give insight into the management of these species. Mottled sculpin are an abundant and widely distributed species in the southern Appalachian region, but there is uncertainty about the ability of this species to persist in coming decades, particularly due to its dependence on cold-water habitats, which may reduce its range in coming decades as global climate change warms water temperatures, forcing the species out of its habitat.[8]

Mottled sculpin are indigenous to the French Broad River (FBR) Basin in North Carolina (Fig. 1), where they are common inhabitants of cooler headwater streams.[9] The FBR Basin is located on the western slope of the Eastern Continental Divide, which runs through Western North Carolina just to the East of Asheville, and eventually drains to the Gulf. But, several populations of sculpin are known in headwater streams of the Broad River Basin, which is on the eastern slope of the Eastern Continental Divide and drains to the Savannah River and then to the Atlantic. A few fish species occur on both sides of the continental divide, including the Rosyside Dace (*Clinostomus funduloides*), and it appears that that species naturally occurs in the Broad River.[10] It is generally thought that sculpins were introduced in the Broad River, [11–13] likely through bait bucket translocations, which is where people who use them as bait then subsequently release them into new areas. Sculpin have been introduced elsewhere via this vector, including into the Colorado River.[14–15] But, some researchers believe that sculpin in the Broad River Basin are native, and the product of historical stream capture events.[13] This led me to hypothesize that if the populations in the upper Broad River Basin are introduced relatively recently,[12] then they should be genetically similar to Cane Creek (Swannanoa/FBR) drainage populations. This genetic similarity should likely extend to sharing of mitochondrial DNA (mtDNA) haplotypes or the interdigitation of haplotypes in a phylogeny. If they were not introduced and are instead the result of historical stream capture, then populations would be native to the Broad River, and thus should show significant genetic divergence from French Broad populations, extending to a lack of shared mtDNA haplotypes, or reciprocal monophyly in phylogenetic analysis. This study aims to compare the population genetics of new and introduced populations in these streams, by examining whether the populations are genetically divergent from one another, and the level of genetic diversity within the two populations.

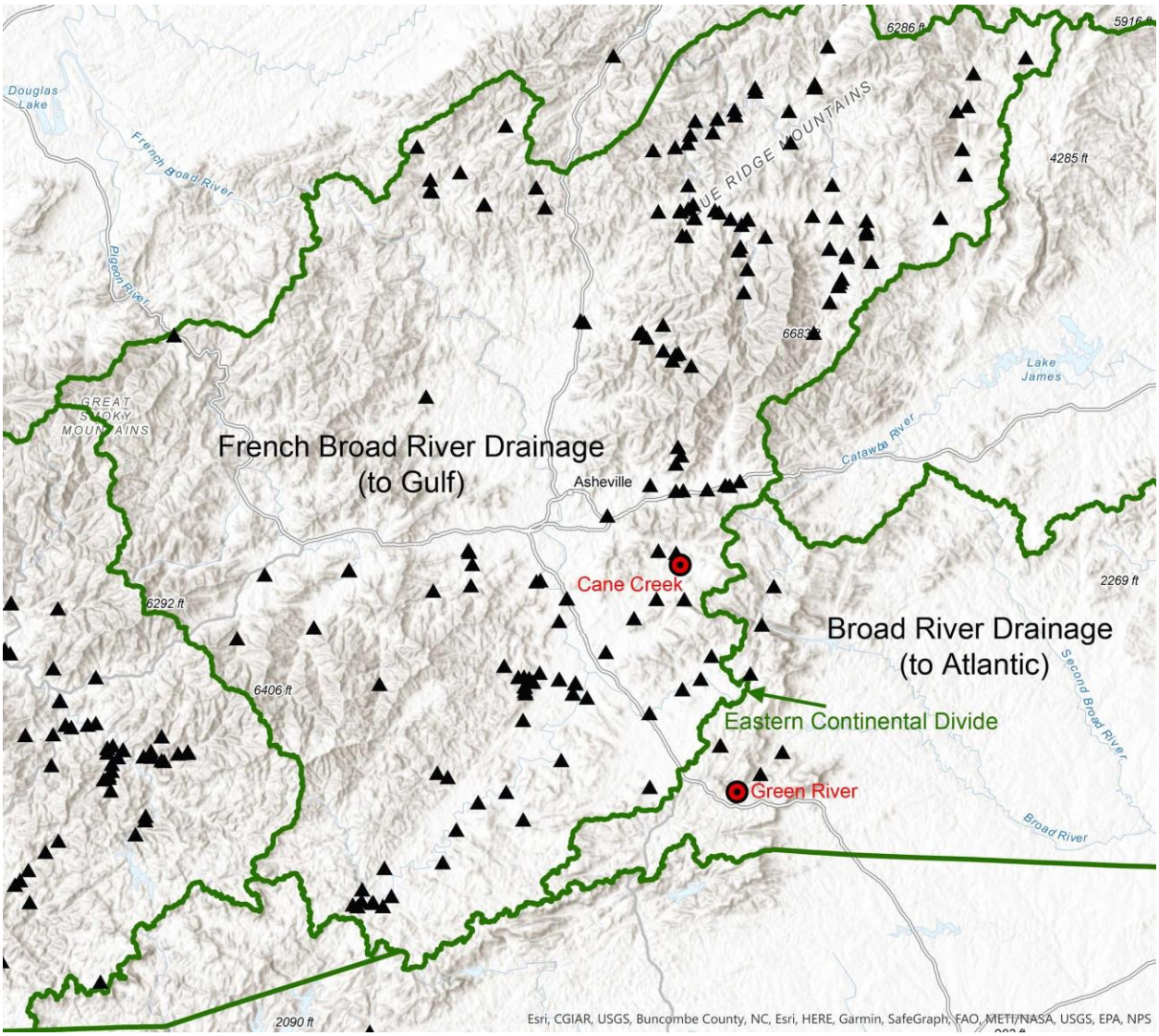

Figure 1. Map of study region.

Figure 1. This map of the Western North Carolina region shows all museum records of mottled sculpin with associated lat/long coordinates obtained from a search of the VertNet portal in March 2021 (black triangles). The green lines indicate separation among major drainage basins. The red dots are sampling sites, with Cane Creek to the north and Green River to the south. The Eastern Continental Divide, separating the Gulf-draining streams and rivers, is labeled. Map rendered in ArcGIS Pro 10.

Figure 2. Mottled sculpin collected from the Green River, March 2021. Photograph by Graham Reynolds.

2. Methods

2.1 Sample Collection

All available US museum records of mottled sculpin (Fig. 2) were downloaded as a .csv file from the VertNet Database (www.vertnet.org) using the search terms "Cottus bairdii North Carolina." These records were then sorted to filter out the ones without resolvable latitude and longitude coordinates. These records were then imported into ArcGIS Pro and mapped onto a topographic basemap to visualize the presumed historical range and the presumed introduced populations of the species (Fig. 1). River basins were overlaid onto the map using a river basin shapefile downloaded from the North Carolina Department of Environmental Quality GIS database (https://data-ncdenr.opendata.arcgis.com/datasets/297f08153f3f4567b779ac1dda7ea374_0).

Sampling sites, one from each side of the Eastern Continental Divide, were selected based on the following criteria: 1) ease of access, 2) known records of mottled sculpin nearby, and 3) and permitted fish collection at the site. The two sites selected that fit these criteria were Cane Creek and the Green River. Cane Creek (35.5178, -82.4022) is a tributary of the Swannanoa River, which drains to the French Broad River on the north end of the Biltmore Estate. This entire drainage basin is a Mississippi drainage, hence these waters end in the Gulf of Mexico. Green River (35.2640, -82.3244) is an upper tributary of the Broad River, with headwaters in Dupont State forest. The Green River passes through two impoundments, Lake Summit and Lake Adger, before joining the Broad River north of Sandy Springs, NC.

Twenty individual mottled sculpin were collected from each of the two sites (Table 1). Sculpin were collected using a back-pack style electrofisher (Fig. 3) to temporarily stun the fish, working upstream until the appropriate number of samples had been collected by dipnetting stunned fish. Captured fish were kept in a 5-gallon bucket until processing. For each fish, a tissue sample was collected by cutting off a small portion of the caudal fin and storing it in a 95% ethanol solution in individual cryotubes.

Table 1. Sampling locations and sample numbers for specimens used in this study.

Sample Site	Presumed Origin	Drainage	Lat/Long (decimal degrees)	# specimens
Green River Tributary	introduced	Savannah (Atlantic)	35.2640, -82.3244	20
Cane Creek	natural	French Broad (Gulf)	35.5178, -82.4022	20

Figure 3. Electrofishing in Cane Creek, Fairview, North Carolina. Photograph by Graham Reynolds.

2.2 Genetic Data Collection

In the lab, each fin sample was macerated using a razor blade, and the tissue was then digested using Proteinase K and a buffer solution. RNAse was added to the solution to digest any unneeded RNA. Digestions took place overnight, and the following day DNA was extracted from the digested tissue samples using a Wizard SV(R) DNA extraction kit (Promega, Madison, Wisconsin). Purified DNA was then eluted into water and stored at -20 degrees C.

I performed polymerase chain reaction (PCR) on each sample to amplify the mitochondrial regions cytochrome B (protein-coding) and the D-loop Control Region. Primer sequences and PCR conditions were obtained from Baumsteiger et al. (2012)[16] and primers were ordered from ThermoFisher. I used the following recipe for PCR reactions: 12.5 □l Promega GoTaq Green mastermix, 2.5 □l H20, 2.5 □l Forward primer, 2.5 □l Reverse primer, and 5 □l of template DNA. The cytochrome B locus was amplified using the primers L14724 (5-GTGACTTGAAAAACCACCGTT-3) and H15915 (5-CAACGATCTCCGGTTTACAAG-3) using the following conditions: 94°C for 5 min., 35 cycles at 94°C for 1 min., 48°C for 1 min., and 72°C for 1 min., and a final 72°C extension for 5 min.[16-18] The D-loop control region was amplified using the primers CR-A (5-CCTGAAGTAGGAACCAGATG-3) and CR-E (5-TTCCACCTCTAACTCCCAAAGCTAG-3) and the following conditions: 94°C for 5 min., 35 cycles at 94°C for 30 sec., 50°C for 1 min., and 72°C for 90 sec., and a final 72°C extension for 5 min.[16,19] An agarose gel was used to ensure that each amplification was successful. Samples were then mailed to NC State for cleanup followed by Sanger sequencing in both directions.

2.3 Genetic Data Analysis

I imported sequences as .abi files into Geneious R10 (www.geneious.com). I created contiguous sequences (contigs) from the forward and reverse primer reads for each individual using the de novo assembly function. I then aligned all contigs with each other and a representative sequence downloaded from GenBank (directly within Geneious). If sequences appeared reversed in the initial alignment, I created a reverse complement, then realigned the sequences. I checked each alignment, one alignment for D-loop and one alignment for CytB, by eye and removed spurious gaps, as well as trimmed the ends of the alignment with lots of missing data. I then created a maximum likelihood phylogeny for each alignment separately using the RaxML function in Geneious[20]. I used 1,000 rapid bootstrapping repetitions and extracted the best-scoring ML tree. I visualized trees by exporting them as .nexus files, then opened them in FigTree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

I then exported each alignment as a .fasta file, then imported the .fasta file into RStudio (running R v. 4.0.3) using the *read.dna* function from the package ape[21]. I analyzed each mtDNA locus separately. I attached population information for each sample by creating a vector of population labels and attaching it using the *rownames* function from base R. I then used the package pegas[22] to collapse sequences into haplotypes and then visualized a haplotype network in RStudio using the function *haploNet* from pegas. I colored each population using internal functions in haploNet. Finally, I calculated a genetic distance matrix for each alignment using the *dist.dna* function in ape and visualized the matrix by turning it into a heatmap using the function *heatmap* in R.

3. Results

3.1 Sampling

I was able to sample 20 individual mottled sculpin from each of the two locations, representing populations on either side of the Eastern Continental Divide (Fig. 1). Zero mortality was observed among either sculpin or other fish captured at each site during electrofishing. At both sites, mottled sculpin was the most abundant species captured. At the Green River site, no other fish species were observed or captured. At the Cane Creek site, four other species were captured: swannanoa darter (*Etheostoma swannanoa*), fantail darter (*E. flabellare*), Gilt darter (*Percina evides*), and central stoneroller (*Campostoma anomalum*).

3.2 Genetic diversity and divergence

I found 13 haplotypes at each locus. For the CytB locus, no haplotypes are shared among individuals, but haplotypes are also interdigitated among the Green River and Cane Creek sites (e.g., there isn't a pattern of divergence between the two sites; Fig. 4). For the D-loop locus, there is also no pattern of divergence, and two haplotypes are shared between the two sites (Fig. 5; Table 2).

The heatmap analysis (Fig. 6) and phylogenetic tree (Fig. 7) showed two interesting patterns. First, Green River and Cane Creek form a mixed clade, indicating no divergence among those two water bodies. Second, Cane Creek has a surprisingly divergent 2nd lineage present. Both of these patterns are also apparent for the D-loop locus (Figs. 8,9).

Table 2. Characteristics and summary statistics of the mitochondrial locus for each population.

locus	length	# haplotypes	Proportion of private haplotypes
D-loop	423	13	0.85
CytB	1,077	13	1.0

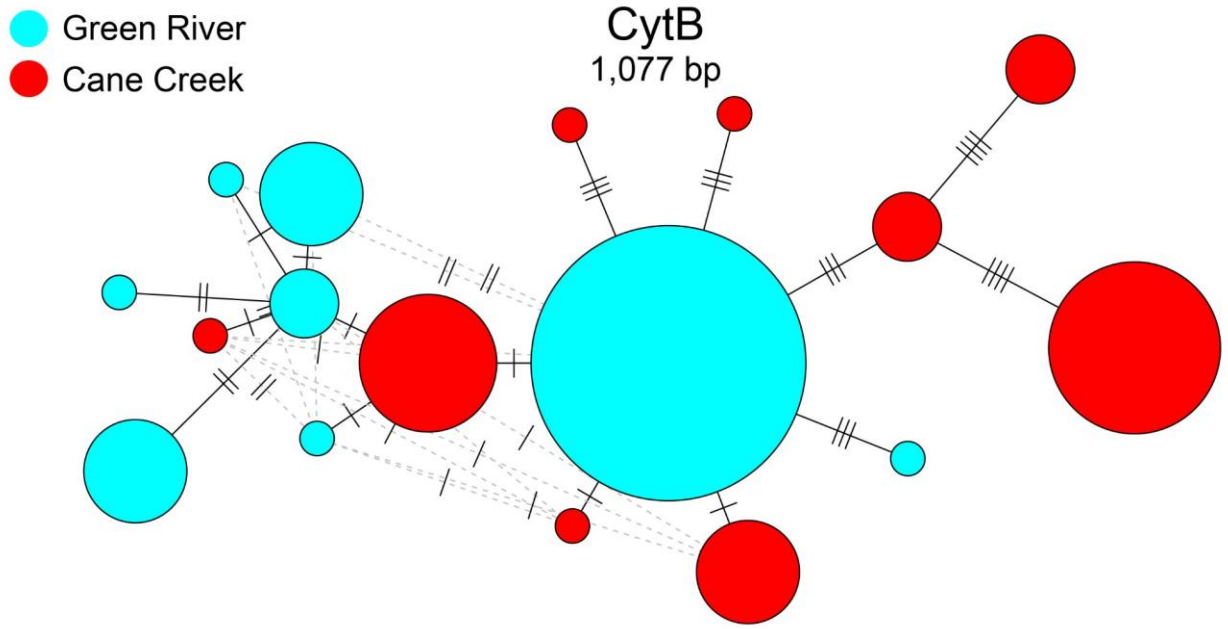

Figure 4. Haplotype network analysis of the cytochrome B locus. Haplotypes are represented as circles, and circles are scaled by the number of sequences that each haplotype has. Line connecting haplotypes represent evolutionary relationships, with tick marks representing individual base-pair substitutions separating haplotypes. Note that while no haplotypes are shared, there is no obvious pattern of divergence.

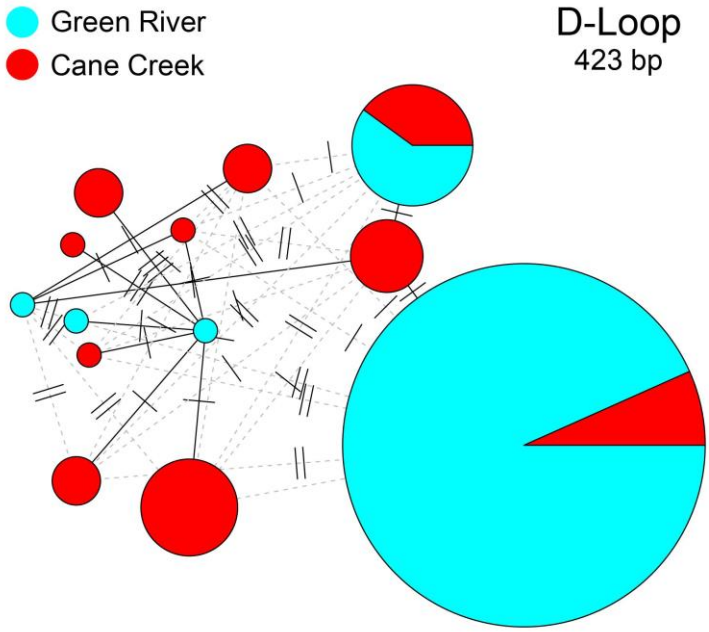

Figure 5. Haplotype network analysis of the D-loop locus. Haplotypes are represented as circles, and circles are scaled by the number of sequences that each haplotype has. Line connecting haplotypes represent evolutionary relationships, with tick marks representing individual base-pair substitutions separating haplotypes. Note that two haplotypes are shared, including the most common haplotype.

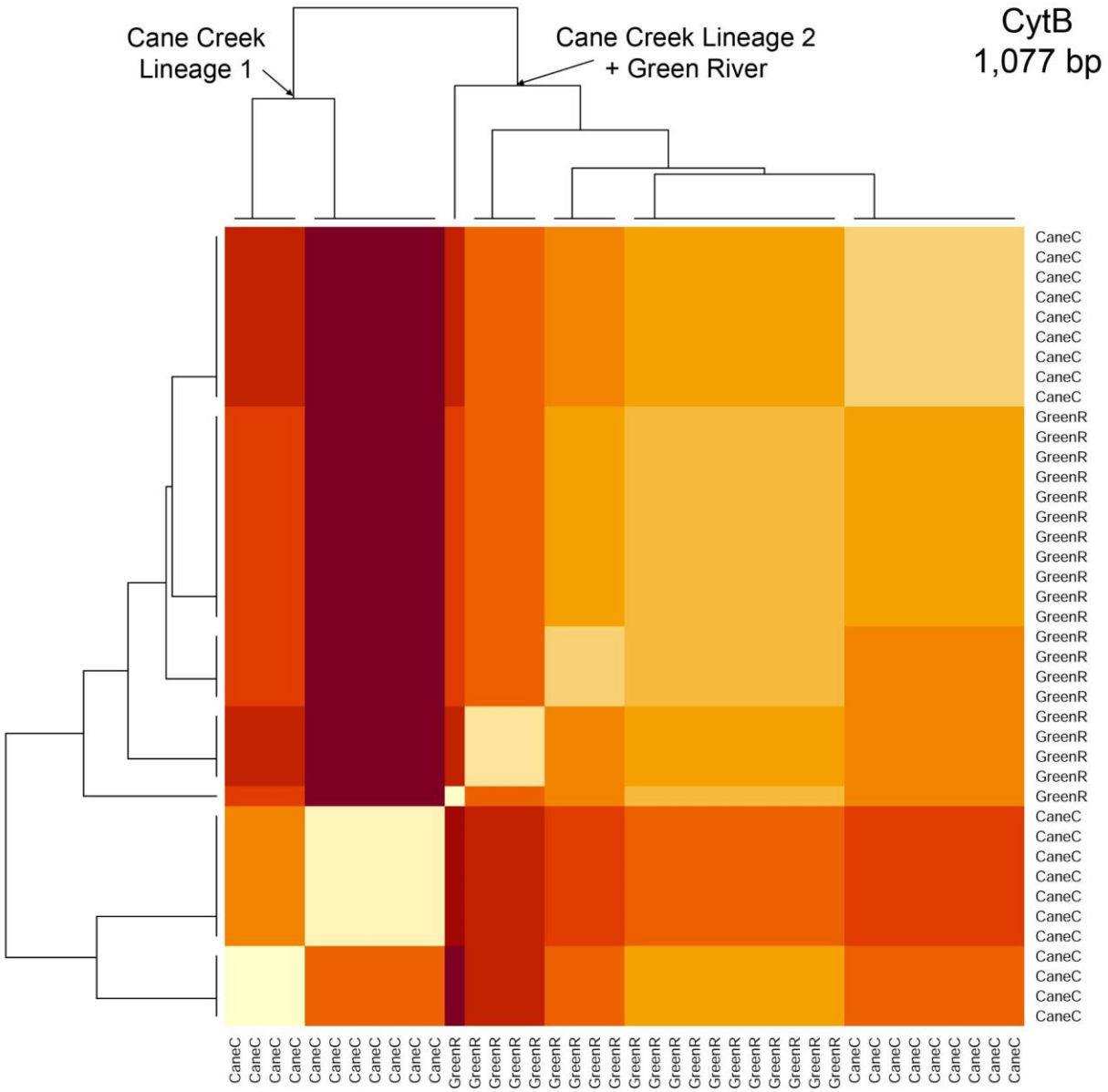

Figure 6. Genetic Distance heat map showing the relative phylogenetic closeness of samples to one another at the cytochrome B region with lighter shades representing more closely related sequences. A neighbor-joining phylogeny is shown outside of the matrix. Note that the matrix is reflected across the diagonal.

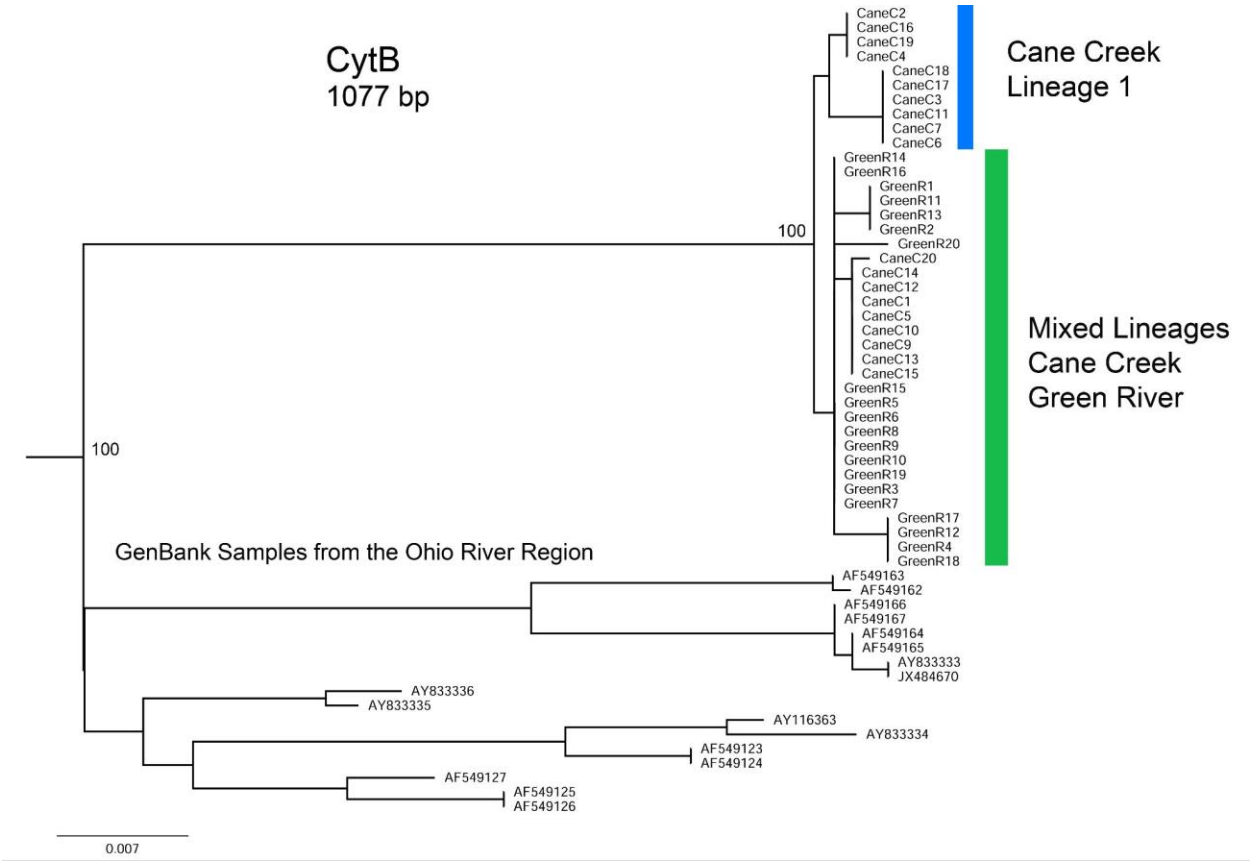

Figure 7. Maximum likelihood tree of the cytochrome B locus. Note the 2 distinct lineages present in Cane Creek, as well as the mixture of Cane Creek and Green River. Significant bootstrap values are shown on the tree (if no value is shown, the relationship was not significant. Outgroup samples from GenBank are also shown.

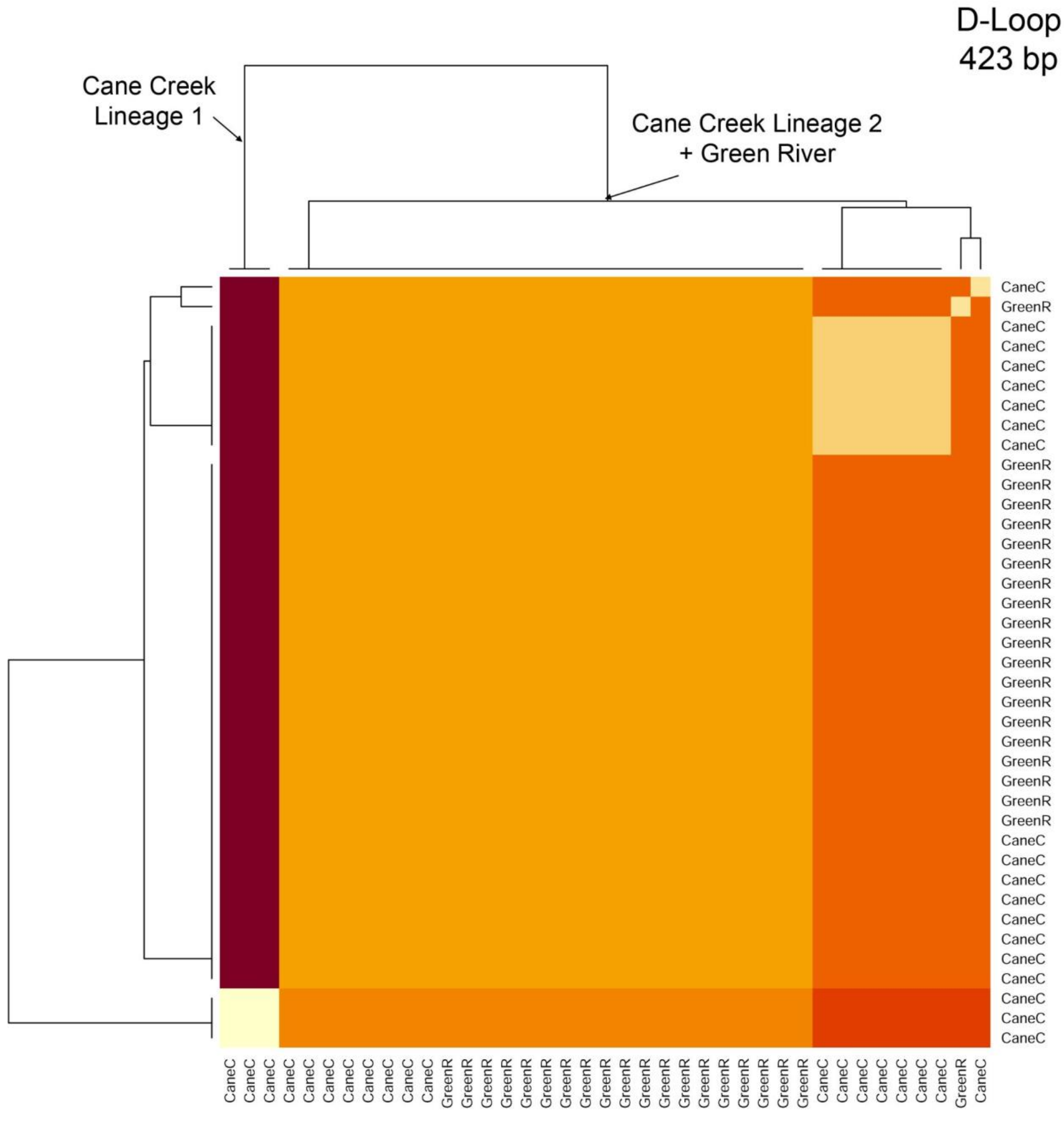

Figure 8. Genetic Distance heat map showing the relative phylogenetic closeness of samples to one another at the D-loop region, with lighter shades representing more closely related sequences. A neighbor-joining phylogeny is shown outside of the matrix. Note that the matrix is reflected across the diagonal.

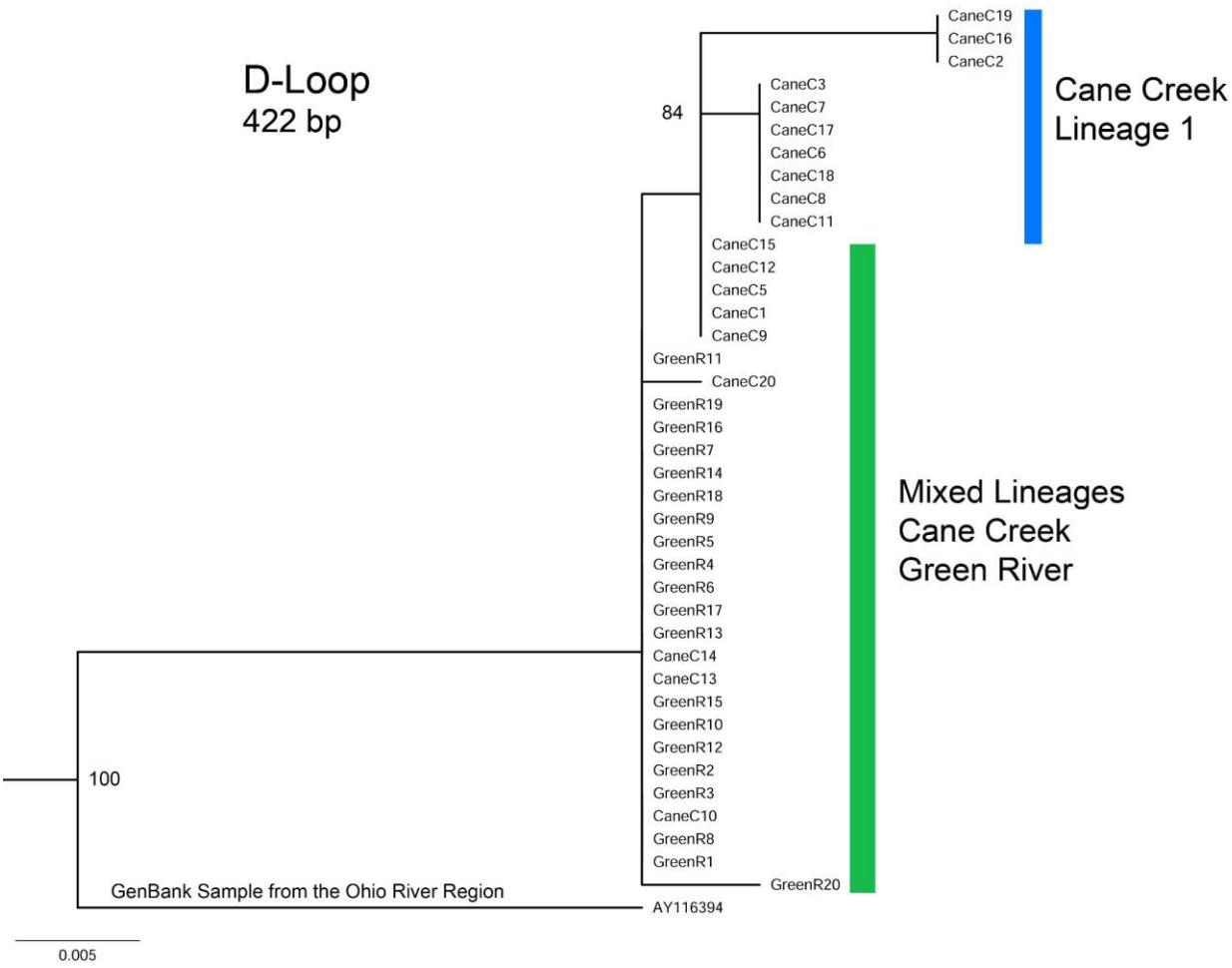

Figure 9. Maximum likelihood tree of the D-loop locus. Note the 2 distinct lineages present in Cane Creek, as well as the mixture of Cane Creek and Green River. Significant bootstrap values are shown on the tree (if no value is shown, the relationship was not significant. An outgroup sample from GenBank is also shown.

4. Discussion

This study sought to determine whether sculpin was native or introduced in the Broad River by examining mitochondrial haplotypes at two loci, using a total of 1,500 base pairs of sequence data. I was able to successfully sequence all samples at the Cytb locus, and all but one sample from Cane Creek at the D-loop locus. I found that mitochondrial haplotypes are interdigitated between Cane Creek and Green River and that no divergence was observed between the two populations. Further, some haplotypes are shared at both sites. These data strongly suggest that the Green River population is indeed introduced, which solves a long-standing question regarding their status[12]. These results indicate that aquatic resources managers could create management plans to attempt to inhibit further mottled sculpin population spread, and eradicate the population currently present. Admittedly, this would prove challenging, as the most effective method of invasive freshwater fish removal is electrofishing, which is a highly labor-intensive and time-consuming process[23]. Other methods of chemical removal like chemical treatments may also be effective, but often have unintended consequences on non-target species[23]. Further study will need to be done in order to determine the extent to which mottled sculpin introduction in affecting stream ecosystem dynamics in the Broad River watershed. However, based on previous research, mottled sculpin is most likely outcompeting or interfering with native species in some capacity. Indeed, I observed that at the Green River site, mottled sculpin were abundant and the only fish species that we observed.

I found an unexpected result in the Cane Creek population, which showed two very distinct mitochondrial lineages within a single 50 m stretch of stream. The reason for this result is unknown, and thus would be an interesting area of further study. A likely scenario is that the Cane Creek population of sculpin shows a mixture of divergent mitochondrial lineages. This could be owing to 1) historical stream capture, whereby a divergent population of sculpin entered the Cane Creek Drainage via a connection made between streams; 2) a result of introduction of sculpin from elsewhere in the FBR Drainage into Cane Creek; or 3) evidence of incomplete lineage sorting and mitochondrial genetic diversity present in FBR Drainage populations. Given my current data showing that sculpin get moved around by people in Western North Carolina, I suspect that sculpin introduction from elsewhere in the French Broad River Drainage is highly plausible.

Nevertheless, it is important to note that we do not yet have a comprehensive picture of sculpin phylogeography in the French Broad River Basin. Hence, a broader study including many sampling sites across the region could begin to characterize the extent of genetic diversity and geographic partitioning of diversity so that the biogeography and evolution of mottled sculpin within the basin can be further understood.

5. Acknowledgements

I thank Dr. Gillete for assistance in identifying appropriate field study sites, as well as for help in the field.

6. References

1. Moi DA, Alves DC, Figuerido BRS, Antiquira PAP, Mello FT, Jeppeson E, Romero GQ, Mormul RP, Bonecker CC. 2021. Non-native fishes homogenize native fish communities and reduce ecosystem multifunctionality in tropical lakes over 16 years. Sci Total Environ. 769.
2. Zhou Y, Wu J, Wang Z, Li G, Zhou L, Gui J. 2021. Microsatellite polymorphism and genetic differentiation of different populations screened from genome survey sequencing in red-tail catfish (*Hemibagrus wyckioides*). Aquac Rep. 19.
3. Almerao MP, Delaunay C, Coignet A, Peiro DF, Pinet F, South-Grosset C. 2018. Genetic diversity of the invasive crayfish *Procambarus clarkii* in France. Limnol. 69: 135-141.
4. Li H, Liang X, Zou S, Liu Y, Clerq PD, Slipinski A, Pang H. 2017. New EST-SSR models reveal strong genetic differentiation in native and introduced populations of mealybug destroyer *Cryptolaemus montrouzieri*. Biol Control. 109:21-26.
5. Kwong RM, Broadhurst LM, Keener BR, Coetzee JA, Knerr N, Martin GD. 2017. Genetic analysis of native and introduced populations of the aquatic weed *Sagittaria platyphylla*- implications for biological control in Australia and South Africa. Biol Control. 112: 10-19.
6. Agdamar S, Tarkan AS. 2019. High genetic diversity in an invasive freshwater fish species *Crassius gibelio*, suggests establishment success at the frontier between native and invasive ranges. Zool Anz. 283: 192-200.
7. Agdamar S, Tarkan AS, Keskin E, Top N, Dogac E, Baysal O, Emiroglu O. 2015. The role of environmental factors and genetic diversity on colonization success of non-native fish, *Lepomis gibbosus* from the western part of Turkey. Biochem Syst Ecol. 58: 195-203.
8. Rashleigh B, Grossman GD. 2005. An individual-based simulation model for mottled sculpin (*Cottus bairdi*) in a southern Appalachian stream. Ecol. Model. 187(2-3): 247-258.
9. Arndt RG, Foltz JW. 2009. Freshwater Fishes of South Carolina (No. 22). Univ of South Carolina Press.
10. Teai H. 2019. Using Morphology and Mitochondrial Cytochrome b Sequences in Determining Origins and Genetic Similarities among Populations of Rosyside Dace (Clinostomus funduloides Girard) in Western North Carolina. University of North Carolina Asheville Journal of Undergraduate Research.
11. Menhinick EF. 1991. The freshwater fishes of North Carolina. North Carolina Wildlife Resources Commission. 227 pp.
12. Tracy BH, Rohde FC, Hogue GM. 2020. An annotated atlas of the freshwater fishes of North Carolina. SFC Proc. 1(60).

13. Fuller P, Neilson M. c2012. Nonindigenous Aquatic Species Database: *Cottus bairdii* Girard, 1850 [internet]. Gainesville(FL): U.S. Geological Survey. [updated 2012 Mar 5; cited 2021 Apr 11] Available from: https://nas.er.usgs.gov/queries/factsheet.aspx?SpeciesID=502
14. Miller RR. 1952. Bait fishes of the lower Colorado River, from Lake Mead, Nevada, to Yuma, Arizona, with a key for identification. Calif Fish Game 38:7-42.
15. Miller RR, Lowe CH. 1967. Fishes of Arizona. C.H. Lowe, ed. The vertebrates of Arizona, part 2. 133-151.
16. Baumsteiger J, Kinziger AP, Aguilar A. 2012. Life history and biogeographic diversification of an endemic western North American freshwater fish clade using a comparative species tree approach. Mol Phylogenet Evol 65: 940-952.
17. Kinziger AP, Wood RM, Douglas ME. 2003. Molecular systematics of the polytypic species *Cottus hypselurus* (Teleostei: Cottidae). Copeia 2003(3): 624–627.
18. Schmidt TR, Gold JR. 1993. Complete sequence of the mitochondrial cytochrome b gene in the cherryfin shiner, *Lythrurus roseipinnis* (Teleostei: Cyprinidae). Copeia 1993: 880–893.
19. Lee WJ, Conroy J, Howell WH, Kocher T. 1995. Structure and evolution of teleost mitochondrial control regions. J Mol Evol. 41: 54–66.
20. Stamatakis A. 2006. RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. J Bioinform 22(21):2688-2690.
21. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. J Bioinform 35: 526-528.
22. Paradis E. 2010. pegas: an R package for population genetics with an integrated–modular approach. J Bioinform 26:419-420.
23. Rytwinski T, Taylor JJ, Donaldson LA, Britton JR, Browne DR, Gresswell RE, Lintermans M, Prior KA, Pellatt MG, Vis C, Cooke SJ. 2019. The effectiveness of non-native fish removal techniques in freshwater ecosystems: a systematic review. Environment Rev 27(1):71-94.