Final Report for the Hydro Research Foundation

Freshwater mussel landscape genomics:

Physical not genetic isolation in river networks

By Matthew Robert Fuller

Duke University Nicholas School of the Environment,
Environment Hall Box 90328, Durham, NC 27708, USA

Correspondence: matt.fuller@duke.edu

Keywords: Next-Generation Sequencing (NGS), Restriction-site Associated DNA sequencing (RADseq), dam removal, *Elliptio complanata*, genetic population structure, gene flow

Short title: Freshwater mussel landscape genomics

Abstract

The landscape genomics of the freshwater mussel *Elliptio complanata* was evaluated across the Neuse River Basin in North Carolina, USA. Genetic evaluations were made at 25 sites throughout the drainage network using next generation restriction-site associated DNA sequencing to identify single nucleotide polymorphisms as genetic markers for genotyping individuals. Population genetic structure for the species was minimal across the landscape despite heavy physical fragmentation of the system by dams. The genetic variation observed among individuals and sites could not be explained by landscape variables such as river distance or intervening dams between sample locations. Furthermore, individual morphology and demographics were not related to the genetic clusters. These results suggest *E. complanata* in the Neuse River Basin may not have experienced genetic divergence as a result of the physical isolation and habitat fragmentation experienced over the last 400 years. Instead, rare jumps past barriers by migratory American Eel hosts, short disturbances allowing barrier permeability, or the genetic memory of old individuals remaining in the reproductive population have maintained population genetic structure evenly across the landscape. Even though these populations have avoided the negative impacts of genetic erosion while in isolation, there are potential demographic, environmental, and ecological impacts of habitat fragmentation that should be considered when managing dams. These considerations are especially important when management is considering their removal and more than a single species for conservation.

Introduction

Historically, in North America, aquatic organisms were able to disperse throughout a river network and only rarely encountered natural barriers (e.g., waterfalls) to movement. This high level of connectivity allowed populations from distant parts of river networks to exchange individuals and interbreed. In addition to supplementing the population demographically (adding individuals via immigrants), these long distance migrants also increased the rate of gene flow between populations. A higher exchange rate of genes between distant populations maintains greater genetic diversity within those populations and avoids the potential risks of genetic erosion (Honnay & Jacquemyn 2007). This is important because populations with greater genetic diversity are predicted to be more fit and capable of adapting to or surviving major disturbances such as climate change.

A major anthropogenic disturbance to river networks has been habitat alteration. In river networks, these habitat alterations take the form of habitat loss, degradation, and fragmentation (Fuller *et al.* 2015). The causes of these habitat alterations are tightly linked to human use of water. Habitat loss in river networks often is the result of river habitat being converted to reservoirs or water being withdrawn from the channel causing the river to run dry. Habitat degradation is easily observed at point source pollution sites (e.g. – downstream of waste water treatment outfalls), but is also known from non-point source pollution in landscapes heavily impacted by agriculture (the result of slow leaching of fertilizer into groundwater that eventually provides streams with a majority of their base flow). Anthropogenic habitat fragmentation is seen in river networks primarily in the form of dams and culverts (physical barriers), but also exists in chemical and biological forms (Fuller *et al.* 2015). Despite these forms of habitat alteration being separated into different categories, the occurrence of one form often leads to or coincides with

another form. For example, a toxic chemical outflow into a river might degrade a segment of the river habitat, but also make it impassible for many aquatic species. As a result, not only has the habitat degraded chemically, but the habitat is lost to organisms that would otherwise have lived there, and it has created a chemical barrier to movement between the upstream river tributaries and the downstream river segments.

Dams are a primary example of an aquatic habitat alteration mechanism that can result in the loss, degradation, and fragmentation of riverine habitat. Small mill dams were first built across the eastern US in the 1600s at low densities (Walter & Merritts 2008), but now an estimated 2,000,000 small dams (reservoir capacity less than 62,000m^3) are scattered throughout the river networks of the US (Graf *et al.* 1993). Larger dams are estimated around 87,000 (US Army Corps of Engineers 2013). As these large and small dams age, they are often decommissioned, abandoned, or become functionally obsolete. Consequently, the cost of maintaining the dams surpasses the cost of repairing or removing them (Poff & Hart 2002). By 2020, the United States Federal Emergency Management Agency predicts that approximately 85% of the dams in the US will have reached their anticipated operational life span, which means a wave of dam removals may be necessary soon (Stanley & Doyle 2002).

Besides the aging infrastructure motivation, dam removal is also promoted (and even marketed via mitigation banking) for restoring river habitat and connectivity within river networks (Poff & Hart 2002; FEMA 2013). Currently, the dispersal ability of aquatic species can be severely restricted by dams in both the up- and downstream direction (Weigel *et al.* 2013). Migratory salmon are a notable species impaired by dams blocking their upstream movement to spawning grounds (Kiffney *et al.* 2009). Additionally, non-migratory fish also are unable to move past

dams and become functionally separated from adjacent populations (Morita & Yamamoto 2002). For other common aquatic animals, such as freshwater mussels, the relationship with dams is ambiguous.

A majority of freshwater mussels require a fish as a host during the larval stage of their life cycle (Strayer *et al.* 2004). Movement of the fish host during larval attachment is the only major mode of dispersal to new branches of a river network for mussels (Schwalb *et al.* 2012). Most mussels have host-parasite relationships with several fish species in their native habitat, but some mussels can only develop on a single species of fish (Schwalb *et al.* 2011). Given the dependence of mussels on their fish hosts for long-distance dispersal and life cycle completion, any impact of dams on fishes also impacts mussels. So, when dams fragment fish populations and restrict their dispersal, freshwater mussel populations are similarly (and possibly more adversely) affected (Watters 1992). However, currently available research has been unable to consistently document the same local impacts of fragmentation on mussel populations. Recent studies have noted faster growth rates and more abundant and diverse mussel communities downstream of small mill dams (Singer & Gangloff 2011; Gangloff *et al.* 2012). However, dams are regularly cited as negatively affecting mussel populations locally by changing water temperatures, sedimentation rates, eliminating fish hosts, and/or restricting their dispersal through river networks (Strayer *et al.* 2004). It is apparent that all dams are not having the same impact on mussels and therefore the diversity of impacts of dams on mussels needs clarification.

Removing dams has the ability to reconnect fragmented river segments and the mussel/fish populations/communities that were previously isolated. Given the unclear relationship between dams and mussels, it is important to determine what variation among dams (e.g. – dam size,

location within the river network, type of water release, etc.) results in the varying impact on mussels. To avoid the site-specific nuances of a single dam's impact on a freshwater mussel species, we can look at species whose distributions span entire river networks to see observe the cumulative impacts of dams in the system. With a better understanding of the dam-mussel relationship at the basin scale, it may be possible to use the cumulative effects of barriers across space to inform the local habitat shifts caused by dams.

This research has two primary objectives. The first is to establish a landscape genomic approach for identifying areas of a river network that are restricting gene flow between populations of a freshwater mussel species (*Elliptio complanata*). The second objective is to try and correlate these areas of restricted gene flow with hypothesized barriers to movement in the river network. For the first objective, 25 empirical sample sites within the Neuse River Basin in North Carolina, USA were sampled and 32 *E. complanata* from each site were sequenced using next generation sequencing technology to map population genetic structure across the river network. Using the measured population genetic structure from the empirical genetic data, correlations were used to try and establish links between both current and historic barriers to movement and the estimated rates of gene flow between sampled populations. These relationships between gene flow rate and barrier location are able to inform whether the current genetic patterns observed in *E. complanata* are derived from historic barriers (natural geographic barriers) or more recent barriers (dams and culverts) and could inform how these barriers are managed in the future.

Methods

Field sampling

Empirical sample sites were located within the Neuse River Basin in North Carolina, USA (Figure 1). The Neuse River Basin is approximately 440km long (down the mainstem) and covers an area of approximately 14,600km^2. Geologically, it is split into two main ecoregions (the coastal plain encompassing the eastern coastal reaches of the river network and the piedmont that encompasses the western headwaters of the basin) by the Fall Line which runs north-south along the eastern half of the United States. The mean annual flow of the Neuse River is approximately 9 million m^3/day (Billingsley *et al.* 1957). Similar statistics for each sample site are noted in Table 1. At each site, 40 individual *Elliptio complanata* were collected by hand and returned to the lab for tissue sampling.

Genomic sample preparation and sequencing

Foot tissue was clipped from each mussel and then DNA extracted using Wizard® SV Genomic DNA Purification System (Promega catalogue# A2360). DNA quantitation was measured using a Qubit 2.0 Fluorometer to select 32 individuals (from the collected 40) that contained the most consistent DNA concentrations (excluded very high and very low DNA concentrations as potential outliers). These individuals were used to build double-digest restriction-site associated DNA sequencing (ddRADseq) libraries for next generation sequencing (Peterson *et al.* 2012). The restriction enzymes used for the ddRADseq libraries were "*MspI*" and "*SbfI*". Three pilot study sites were built into three unique libraries (32 individuals from each population in three separate libraries) and then sequenced on three lanes of a patterned Illumina flow cell (32 individuals from one site/population per lane). For the remaining 22 sites, libraries were built by

mixing individuals from different sites into a single library (16 individuals from two different sites in each library of 32 individuals)and then sequencing two optically balanced libraries on the same lane (64 individuals total on a lane made up of 16 individuals from 4 separate sites). These measures to mix individuals across libraries and flow cell lanes were established to help avoid systematic lane and library-build patterns in the genetic sequences. All libraries were run at the Duke University Duke Center for Genomic and Computational Biology facility on an Illumina Hi-seq 2500 using single-end 50bp sequencing standards.

Genomic sequence data processing

Raw sequences were demultiplexed and processed (SNP discovery) using the bioinformatics pipeline Stacks (Catchen *et al.* 2013). During the initial steps of Stacks (STACKS:process_radtags) sequences specific to each individual were demultiplexed and sequences of low quality (raw Phred score < 10 on a Phred 33 scale) were removed. Additionally, the sequences tagged by the Illumina Hi-seq as low quality sequences were also removed. Stacks of sequences were then aligned and built de novo because no reference genome exists for *E. complanata*. During this de novo stack assembly (STACKS:denovo_map.pl) the minimum number of identical sequences to generate a stack (flag "-m" for denovo_map.pl) was set at three while the number of nucleotide mismatches permitted between loci for a single individual (flag "-M") was set to three (Catchen *et al.* 2013). Mismatch allowance for the entire catalogue (flag "-n") was set to two (Catchen *et al.* 2013). Generating the final set of loci and SNPs for analysis was done using the "populations" program in Stacks (Catchen *et al.* 2013). Filtering parameters in STACKS:populations were set such that a minimum of 75% of all individuals were sequenced at each loci (flag "-r"), the number of reads at each locus must be at least 10 (flag "-m"), and the minimum minor allele frequency for a given nucleotide site needed

to be greater than 0.05 (flag "--mim_maf") (Catchen *et al.* 2013). The final haplotypes for each individual were then evaluated to identify population genetic structure in the system.

Population genetic analysis

To evaluate population genetic structure, three different analyses were used to check for consistency in the results. First, the population divergence statistic F_{ST} was calculated for all site-by-site pairwise interactions. This distance matrix for all sites was then used to build a neighbor-joining tree to identify which sites were most and least genetically related to each other. Second, discriminant analysis of principle components (DAPC) was used to ordinate individuals and identify population clusters related to landscape variables (Jombart *et al.* 2010). Lastly, analysis of population structure was conducted using the statistical software STRUCTURE v2.3.4 (Pritchard *et al.* 2000; Pritchard 2010) to estimate the number of genetic clusters in the data set. Post-processing of the Structure files was done using Structure Harvester (Earl & VonHoldt 2012) and CLUMPP version 1.1.2 (Jakobsson & Rosenberg 2007). All analyses, unless otherwise specified, were conducted in the statistical software R (R Core Team 2016) using the packages "adegenet" (Jombart *et al.* 2008, 2010; Jombart & Ahmed 2011), "ape" (Paradis *et al.* 2004; Popescu *et al.* 2012), "ecodist" (Goslee & Urban 2007), "hierfstat" (Goudet & Jombart 2015), and "pegas" (Paradis 2010).

Results

For the 800 sequenced *E. complanata*, 360 loci were identified for use to compare *E. complanata* population genetic structure among the 25 sites. Each loci was sequenced in at least 75% of all individuals and in at least 50% of the sampled sites. The total number of alleles per site ranged from 388 to 461 at the 360 biallelic-loci (Figure 2) indicating a relatively even number of alleles represented across all sites.

Genetic relationships evaluated using the F_{ST} statistic (Meirmans & Hedrick 2011) suggest genetic isolation exists between some sites in this study system (Figure 3). Many pairwise F_{ST} values were greater than 0.1, which is often used as a threshold between minimal and significant genetic differentiation and population structure between sampled populations. To investigate these site-by-site relationships, a neighbor-joining tree was built using the pairwise F_{ST} distance matrix among all sites (Figure 4). The neighbor-joining tree did not group sites residing within the same major tributaries of the Neuse River Basin. A good example of this is how the Little River ("LittleRvr" suffix in labels of tree) sites were scattered among all major branches of the tree when they should probably cluster very closely on the same primary branch (Figure 4).

DAPC ordination was used to identify individuals and sites that were genetically similar. For individuals, two distinct clusters were identified using the first two discriminant functions from the DAPC (Figure 5). These two groups separated primarily along the first discriminant function. Sites, however, did not separate into any distinct clusters/groups (Figure 5). Instead, most sites contained individuals in both genetic clusters that separated along discriminant function 1. Attempts to group individuals based on their size (an approximation for age), ecoregion

(Piedmont or Coastal Plain), landscape elevation above sea level, distance to the mainstem Neuse River outlet, or upstream watershed area were unsuccessful at generating any patterns associated with the individual genetic clusters. Similar grouping evaluations were made using potential laboratory processes that could have systematically resulted in the groups. These method evaluations were also unsuccessful in producing a relationship that might explain the genetic separation demonstrated in the DAPC ordination (results not shown). In case the genetic groupings were the result of morphologically cryptic species, the two distinct genetic clusters were also evaluated separately (individuals in each group were evaluated using DAPC) to see if new genetic patterns existed among the sites. Evaluation of each genetic cluster of individuals separately also did not provide any insight to the mechanism driving the genetic diversity among the individuals and sites (data not shown).

Similar to the DAPC results, using STRUCTURE to assign individuals to genetic groups also found genetic clusters unrelated to the geography of the system (Figure 6). Structure defined two genetic clusters as the most likely grouping for the individuals with three or four genetic groups as the second and third most likely number of genetic groupings. This also agrees with what the neighbor-joining tree and F_{ST} population genetic data, which did not group sites based on their location in the river network (Figure 4).

Discussion

The landscape genetics of *E. complanata* in the Neuse River Basin suggest that little to no genetic isolation currently exists across the river network that is associated with landscape variation. The genetic variation that does exist within the system is greatest within rather than

between sites. There are a few plausible explanations for this pattern of genetic variation that include “genetic memory” dampening genetic drift, and species-specific life history traits or demographics.

Genetic drift occurs when alleles in a population become fixed or drop out of a population gene pool as a result of random genetic mixing and mutation during reproduction (Slatkin 1987). This fixation of alleles occurs more rapidly in small, isolated populations and is a primary target for measuring the impacts of habitat fragmentation. The shift in allele frequencies across the river network was the target response variable for understanding how freshwater mussels in our study system were responding to the heavy fragmentation by dams and was expected to be a useful tool for characterizing the impact of dams on dispersal through the river network. The lack of genetic variation across the river network that was related to any landscape variable was, therefore, unexpected. However, there are a few potential reasons for this genetic consistency across the basin.

One potential reason for the lack of genetic variation across the study basin is the maintenance of old genetic alleles in the reproductive population. The study species, *E. complanata*, can live for relatively long periods of time (50-70 years). This generation period means that at a minimum, there could be some individuals that are only five or six generations beyond their ancestors that lived prior to a majority of the mill dams that first fragmented the river network 400 years ago. As a result, old gene frequencies may persist in the genetically-effective population and wash out any impact of physical isolation that would otherwise manifest in the population genetics through drift. These old individuals continue to reproduce until death and therefore can maintain the old genetic “memory” in the system even while in isolation.

A second potential reason for a lack of population genetic structure for *E. complanata* within the Neuse River basin may be due to their demographics. *E. complanata* is one of the most common and abundant mussels found in coastal watersheds on the eastern half of North America (Haag 2012). Very abundant species generally have large reproductive populations and the rates of allele fixation or loss from genetic drift occur fastest in small, isolated populations (Waller 2015). When the demographics of large breeding populations is combine with the genetic memory of old individuals in the breeding population, the effects of genetic drift can be very slow to manifest. This means even though dams and other barriers may be isolating populations physically from exchanging genetic material, the population genetic structure of the river network may not reflect that physical isolation yet because genetic drift is occurring too slowly to be detected yet.

E. complanata are considered fish host generalists (though the American Eel has been hypothesized as a primary/important host in some areas (Lellis *et al.* 2013; Reese *et al.* 2014)) and can therefore complete its life cycle using a variety of fish hosts. These species are both migratory (e.g., American Eel) as well as local residents (e.g., sculpins) (Lellis *et al.* 2013). This is important because the diversity of hosts allows *E. complanata* to have successful recruitment both in isolation with resident fishes, but also across large spatial extents when attaching to a migrating fish. Furthermore, American Eel are able to climb wet vertical surfaces when less than 100mm long (Arai 2016), which makes them capable of overcoming many barriers that would otherwise restrict movement of other species. Since *E. complanata* can metamorphose on these small, Elver-stage eels, it is possible that *E. complanata* is occasionally jumping barriers by attaching to these hosts. If dispersal past dams can be achieved only three or four times per 100 generations, then the effects of genetic drift in isolation can be masked/reversed (Slatkin 1987).

The dams in the Neuse River Basin are generally small mill dams with only a few large facilities (Hoenke *et al.* 2014). Many of these mills dams may be passable by organisms without the ability to climb vertical walls during major floods, or, even temporarily passible during their repair or the construction of a new dam adjacent to the old failing dam. Mill dams were also sometimes built with gates to flush the reservoir when it began to fill in with sediment. This flushing may provide passage for mature adults to retreat downstream past barriers occasionally. Any of these dispersal events might provide opportunity for individuals to pass what are generally assumed to be permanent and complete barriers.

With no genetic variation associated with distance, it is assumed that the mussels, though isolated physically in space, have not developed distinct genetic variants in isolation as a result of mutation or genetic drift. Therefore, the removal of dams in these systems could re-establish gene flow connectivity between the isolation populations without any risk of negative genetic recombination issues. Furthermore, the timeframe of isolation for *E. complanata* in this system suggests that species with similar movement and life history patterns may be somewhat resistant to the effects of habitat fragmentation from a genetic erosion perspective. The "genetic memory" of populations with long-lived individuals in a system helps to buffer the local and short-term impacts of physical isolation in the landscape. If these physical barriers are properly managed (built, but then removed after short, specified time-periods), the benefits of the barrier could be reaped (Doyle *et al.* 2003) and the impact to species could be minimized by removal before reaching a period that would push the species to a state of genetic vulnerability.

While it is interesting that the sampled populations show no genetic erosion from their likely physical isolation in the Neuse River Basin, it is important to note that genetic erosion is only

one of many potential negative impacts of being physically isolated in space. Therefore, the genetic impacts of habitat fragmentation are only one of many stresses that are relieved when defragmenting a system through barrier removal. As dam infrastructure ages across North America, decisions to remove or repair dams will require information on which barriers should be removed first to restore functioning metapopulations quickly (Fuller *et al.* 2015). Genetic as well as demographic, environmental, and ecological considerations beyond a single species should be weighed in this decision process.

Acknowledgements

Primary funding for this research came from the Hydro Research Foundation and the US Fish and Wildlife Service with small grants from the Society for Freshwater Science, The Garden Club of America, and Duke University Nannerl Keohane Award also contributing to the project.

References

Arai T (2016) *Biology and Ecology of Anguillid Eels* (T Arai, Ed,). CRC Press.

Billingsley GA, Fish RE, Schipf RG (1957) *Water resources of the Neuse River Basin, North Carolina*.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Doyle MW, Harbor JM, Stanley EH (2003) Toward policies and decision-making for dam removal. *Environmental Management*, **31**, 453–465.

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

FEMA (2013) *Dam Safety in the United States: A Progress Report on the National Dam Safety Program Fiscal Years 2008 through 2011*.

Fuller MR, Doyle MW, Strayer DL (2015) Causes and consequences of habitat fragmentation in river networks. *Annals of the New York Academy of Sciences*, **1355**, 31–51.

Gangloff MM, McCormick M, Siefferman L (2012) *Effects of small dams on freshwater bivalve assemblages in North Carolina piedmont and coastal plain streams*. North Carolina Wildlife Resources Commission.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Goudet J, Jombart T (2015) hierfstat: Estimation and Tests of Hierarchical F-Statistics.

Graf WL, Board WST, Council NR (1993) Landscapes, commodities, and ecosystems: The relationship between policy and science for American rivers. In: *Sustaining Our Water Resources*, pp. 11–42. The National Academies Press, Washington, DC.

Haag WR (2012) *North American freshwater mussels natural history, ecology, and conservation*. Cambridge University Press, Cambridge England ; New York.

Hoenke KM, Kumar M, Batt L (2014) A GIS based approach for prioritizing dams for potential removal. *Ecological Engineering*, **64**, 27–36.

Honnay O, Jacquemyn H (2007) Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conservation Biology*, **21**, 823–831.

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.

Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Jombart T, Ahmed I, Calboli F *et al.* (2008) Package " adegenet ." *Bioinformatics Application Note*, **24**, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new

method for the analysis of genetically structured populations. *BMC Genetics*, **11**.

Kiffney PM, Pess GR, Anderson JH *et al.* (2009) Changes in Fish Communities Following Recolonization of the Cedar River, Wa, USA by Pacific Salmon after 103 Years of Local Extirpation. *River Research and Applications*, **25**, 438–452.

Lellis WA, St. John White B, Cole JC *et al.* (2013) Newly documented host fishes for the Eastern Elliptio Mussel (Elliptio complanata). *Journal of Fish and Wildlife Management*.

Meirmans PG, Hedrick PW (2011) Assessing population structure: FST and related measures. *Molecular Ecology Resources*, **11**, 5–18.

Morita K, Yamamoto S (2002) Effects of habitat fragmentation by damming on the persistence of stream-dwelling charr populations. *Conservation Biology*, **16**, 1318–1323.

Paradis E (2010) Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**, 419–420.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**.

Poff NL, Hart DD (2002) How Dams Vary and Why It Matters for the Emerging Science of Dam Removal. *Bioscience*, **52**, 659–668.

Popescu AA, Huber KT, Paradis E (2012) Ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics*, **28**, 1536–1537.

Pritchard JK (2010) Documentation for structure software : Version 2 . 3. *In Practice*, **6**, 321–326.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Core Team (2016) R: A Language and Environment for Statistical Computing.

Reese SP, Huffner M, Feindt J (2014) Mussel population and distribution on Buffalo Creek , an American Eel stocked tributary to the West Branch Susquehanna River. *Journal of the Pennsylvania Academy of Science*, **88**, 63–66.

Schwalb AN, Morris TJ, Mandrak NE, Cottenie K (2012) Distribution of unionid freshwater mussels depends on the distribution of host fishes on a regional scale. *Diversity and Distributions*, n/a-n/a.

Schwalb AN, Poos MS, Ackerman JD (2011) Movement of logperch-the obligate host fish for endangered snuffbox mussels: implications for mussel dispersal. *Aquatic Sciences*, **73**, 223–231.

Singer EE, Gangloff MM (2011) Effects of a small dam on freshwater mussel growth in an Alabama (U.S.A.) stream. *Freshwater Biology*, **56**, 1904–1915.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

375 Stanley EH, Doyle MW (2002) A geomorphic perspective on nutrient retention following dam
376 removal. *Bioscience*, **52**, 693–701.

377 Strayer DL, Downing JA, Haag WR *et al.* (2004) Changing Perspectives on Pearly Mussels,
378 North America's Most Imperiled Animals. *Bioscience*, **54**, 429–439.

379 US Army Corps of Engineers (2013) National Inventory of Dams.

380 Waller DM (2015) Genetic rescue: a safe or risky bet? *Molecular Ecology*, **24**, 2595–2597.

381 Walter RC, Merritts DJ (2008) Natural streams and the legacy of water-powered mills. *Science*,
382 **319**, 299–304.

383 Watters GT (1992) Unionids, Fishes, and the Species-Area Curve. *Journal of Biogeography*, **19**,
384 481–490.

385 Weigel DE, Connolly PJ, Martens KD, Powell MS (2013) Colonization of Steelhead in a Natal
386 Stream after Barrier Removal. *Transactions of the American Fisheries Society*, **142**, 920–
387 930.

Tables

Table 1. Empirical site information for all 25 study sites.

Site Number	River Name	Major Neuse River Tributary	Lat	Long	NC County	Ecoregion	Watershed Area (km2)	Elevation (m)	Sample Date
1	South Flat River	Falls Lake Tributary	36.256763	-78.944047	Person	Piedmont	140.88	144.2	18-Nov-15
2	Deep Creek	Falls Lake Tributary	36.24044	-78.88891	Person	Piedmont	83.06	120.845	18-Nov-15
3	Moccasin Creek	Contentnea Creek	35.896366	-78.310517	Franklin	Piedmont	20.66	78.9668	16-Nov-15
4	Smith Creek	Falls Lake Tributary	36.08842	-78.60244	Granville	Piedmont	16.28	95.8034	29-Nov-15
5	Moccasin Creek	Contentnea Creek	35.813648	-78.256806	Johnston	Piedmont	72.26	61.8966	30-Oct-15
6	Brier Creek	Crabtree Creek	35.859187	-78.812136	Wake	Piedmont	32.14	87.1567	09-Dec-15
7	Little River	Little River	35.91381	-78.386818	Wake	Piedmont	67.52	80.7519	16-Nov-15
8	Little River	Little River	35.88179	-78.375459	Wake	Piedmont	95.47	78.7962	16-Nov-15
9	Little River	Little River	35.821778	-78.351681	Wake	Piedmont	145.59	67.9511	30-Oct-15
10	Little River	Little River	35.666638	-78.258479	Johnston	Piedmont	270.97	50.8198	16-Nov-15
11	Little River	Little River	35.753408	-78.297865	Johnston	Piedmont	194	60.1784	30-Oct-15
12	Richland Creek	Crabtree Creek	35.834102	-78.720203	Wake	Piedmont	16.62	81.352	07-Dec-15
13	Big Arm Creek	Marks Creek	35.70401	-78.417023	Johnston	Piedmont	9.64	53.1969	29-Nov-15
14	Marks Creek	Marks Creek	35.751893	-78.429294	Wake	Piedmont	33.53	56.0474	29-Nov-15
15	Great Swamp	Contentnea Creek	35.608883	-77.952268	Wilson	Coastal Plain	98.83	23.5176	11-Dec-15
16	Contentnea Creek	Contentnea Creek	35.687801	-77.94714	Wilson	Coastal Plain	609.38	22.5189	16-Nov-15
17	Swift Creek	Swift Creek PDMT	35.718939	-78.751931	Wake	Piedmont	54.41	90.8891	07-Dec-15
18	Indian Well Swamp	Swift Creek CP	35.459254	-77.283604	Pitt	Coastal Plain	38.51	8.81818	11-Dec-15
19	Swift Creek	Swift Creek PDMT	35.68827	-78.68157	Wake	Piedmont	97.85	76.4903	25-Nov-15
20	Neuse River	Neuse River	35.64742	-78.40546	Johnston	Piedmont	2986.68	40.2397	16-Dec-15
21	Core Creek	Neuse River	35.253172	-77.28708	Craven	Coastal Plain	144.27	2.19813	12-Dec-15
22	Whitley's Creek	Neuse River	35.21487	-77.68293	Lenoir	Coastal Plain	9.52	12.0585	16-Dec-15
23	Eno River	Falls Lake Tributary	36.04681	-79.01093	Orange	Piedmont	294.23	120.383	06-Nov-14
24	Neuse River	Neuse River	35.78443	-78.53688	Wake	Piedmont	2275.12	47.8991	11-Nov-14
25	Little Creek	Contentnea Creek	35.67696	-78.15254	Wilson	Piedmont	8.89	46.3913	20-Nov-14

Figures (with captions below)

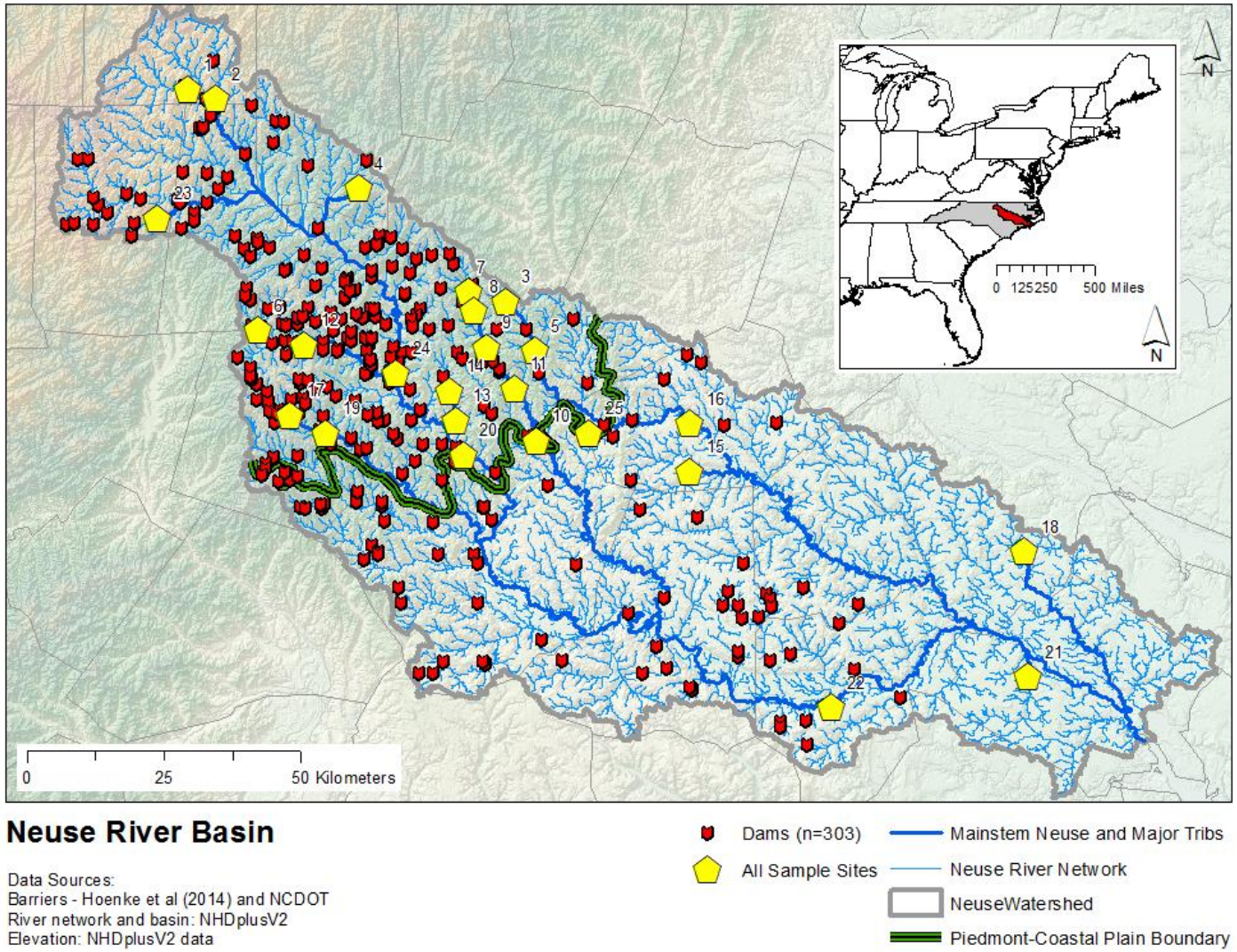
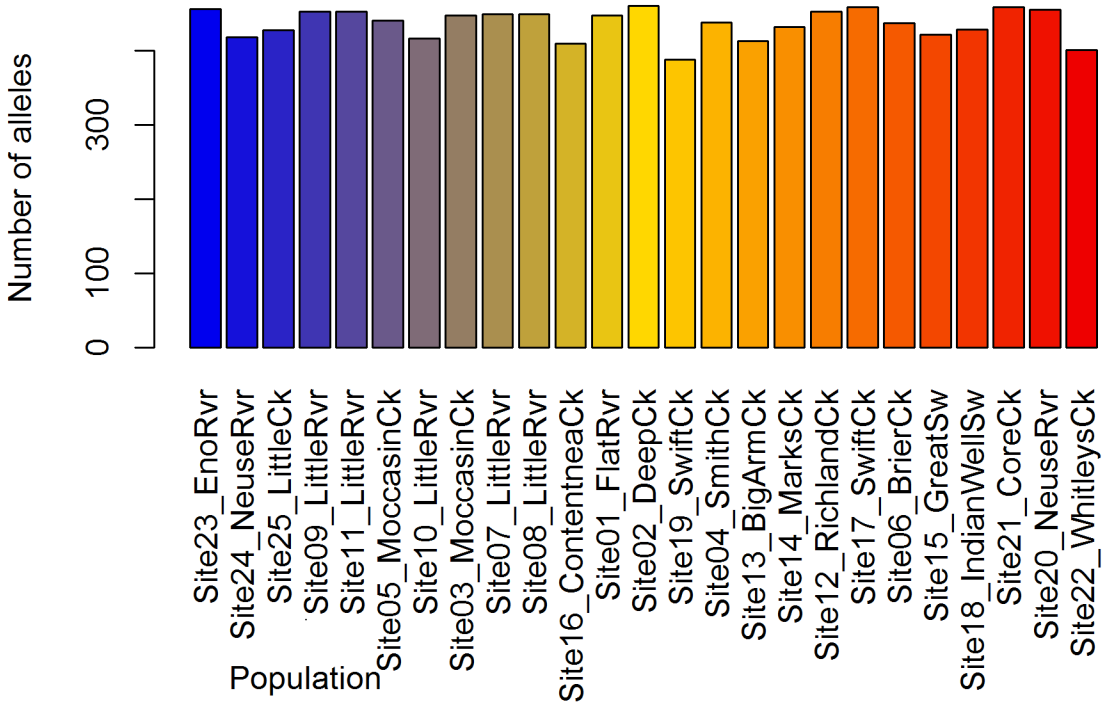


Figure 1. Site map for empirical sampling. Piedmont ecoregion is to the north and west of the boundary line (toward the headwaters), while the coastal plain extends to the south and east of the boundary.



394
395 **Figure 2. Allele counts at each study site.**

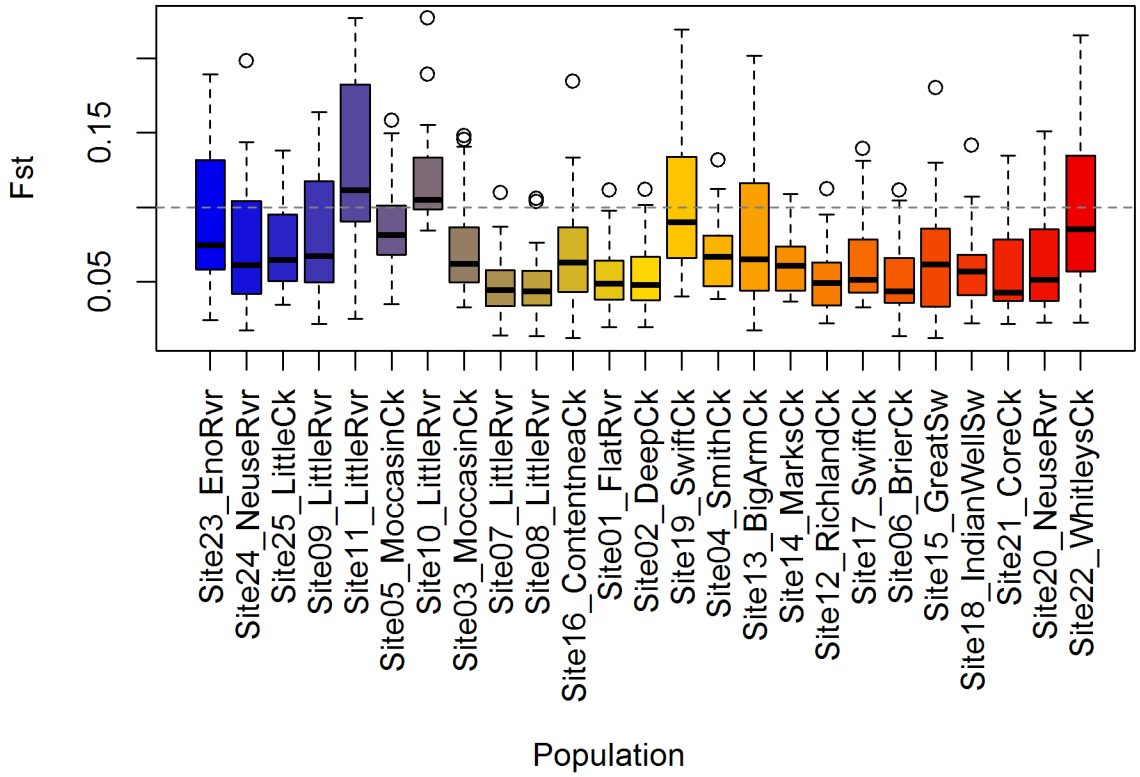


Figure 3. Population genetic structure represented by Wright's F_{ST} statistic. F_{ST} values greater than 0.1 represent moderate levels of genetic isolation.

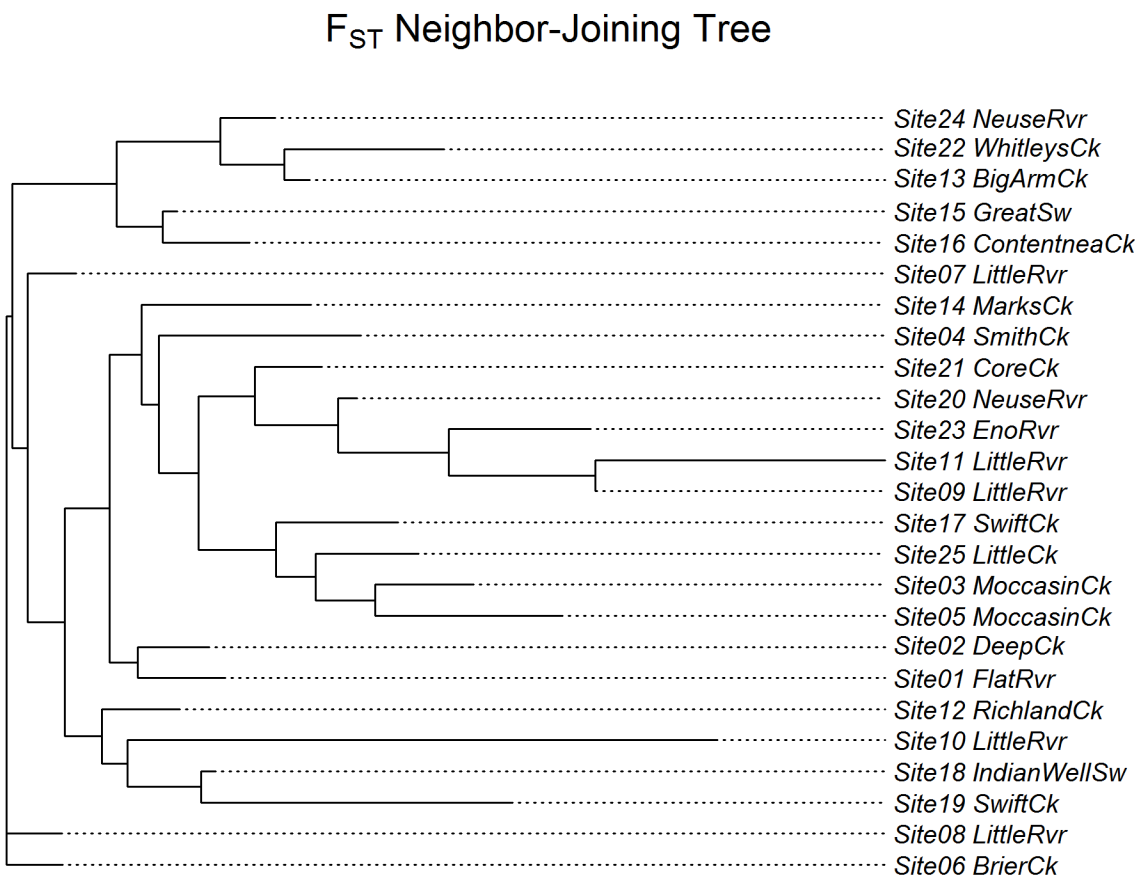


Figure 4. Neighbor-joining tree using F_{ST} to define splits and branch lengths among all the sample sites.

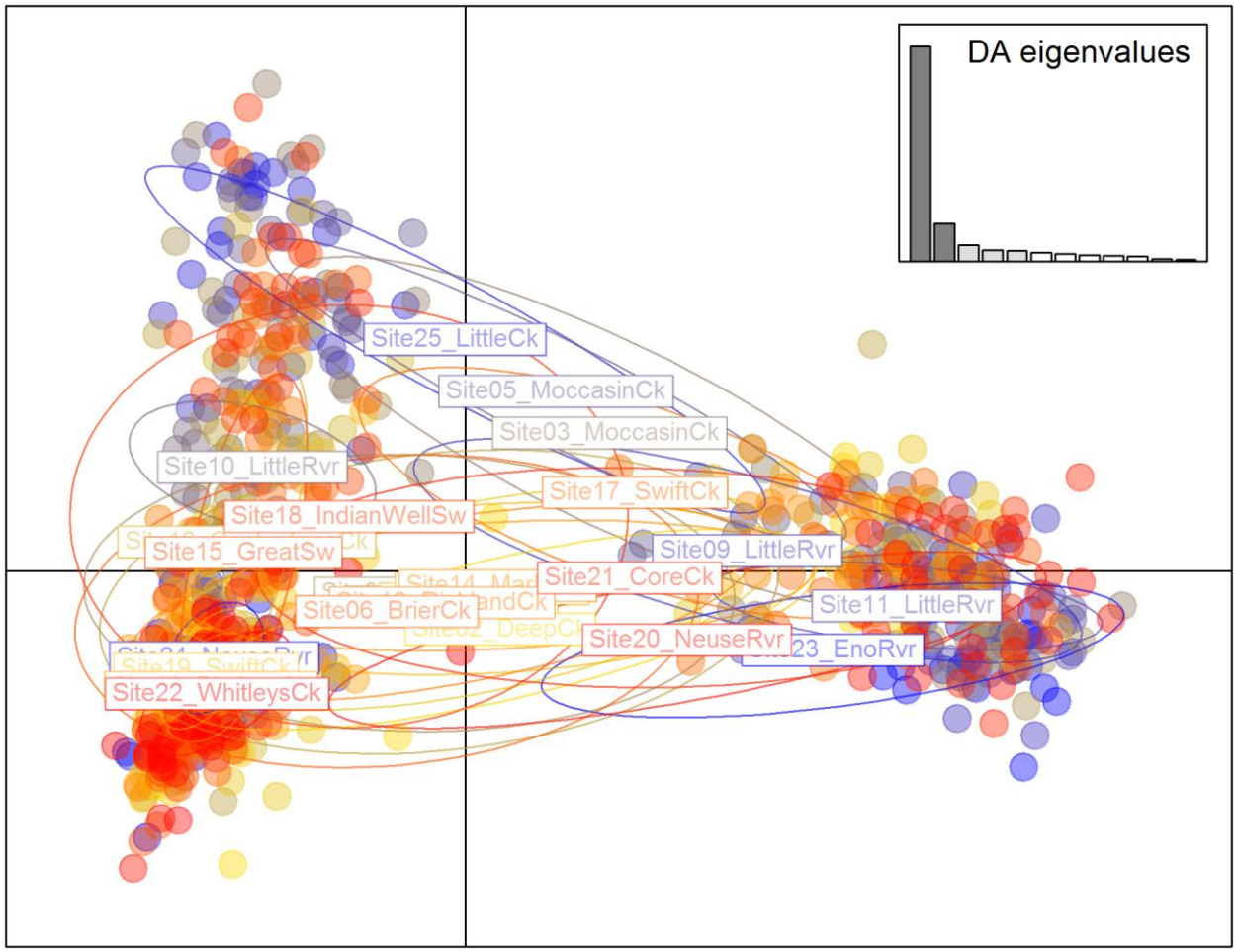


Figure 5. DAPC ordination with individuals represented by points and sites represented by box labels placed at the center of mass of individuals for a site. Ellipses help identify the spread of individuals across the ordination space. Colors are the same for sites and their individuals.

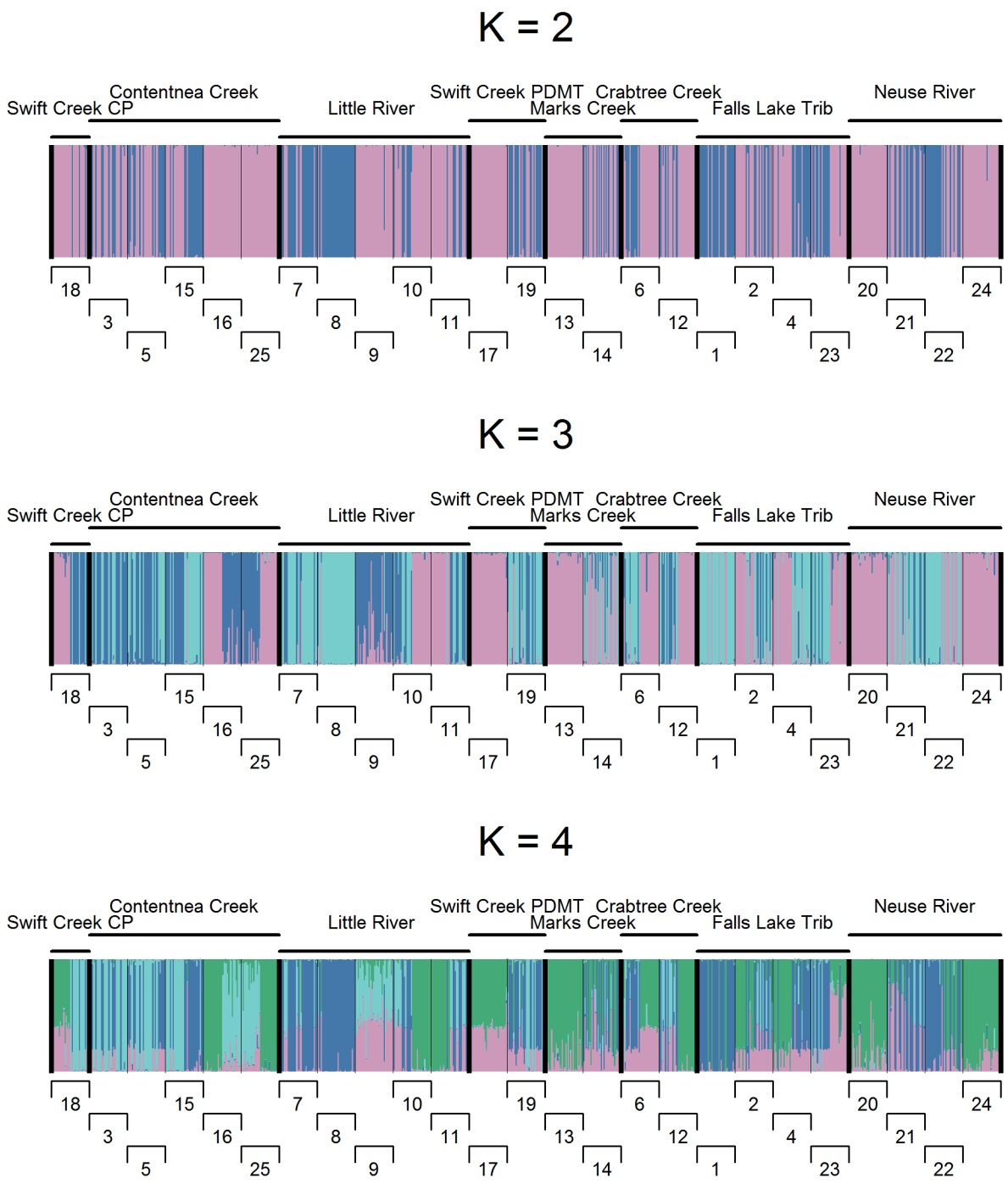


Figure 6. Structure plots for the three most likely numbers of genetic groups for *E. complanata* selected using the Evanno et al (YEAR) method. Each bar (vertical colored line) in the plots represents an individual mussel in the study. The different colors of those lines represent the different genetic groups. Lines with more than one color had genetics that made it probable that they could have belonged to more than one genetic group. Sites and their individuals are labeled/bracketted below each plot while above each plot are labels/brackets identifying the groups of sites within each major tributary.