University of North Carolina Asheville
Journal of Undergraduate Research
Asheville, North Carolina
May 2017

# Mitochondrial *Cytochrome-b* Gene Comparison Suggests Mixed Origins for Newly Discovered Populations of Rosyside Dace (*Clinostomus funduloides Girard*) from the Upper French Broad Basin

Andrew Simons
Biology
The University of North Carolina Asheville
One University Heights
Asheville, North Carolina 28804 USA

Faculty Advisors: Dr. Jennifer Ward, Dr. David Gillette

## Abstract

The rosyside dace, *Clinostomus funduloides,* is a minnow commonly found in rocky pools of fast-running headwaters and streams throughout its range in the eastern United States. Though *C. funduloides* has not been officially catalogued in the Upper French Broad (UFB) basin in North Carolina, populations were recently discovered in two UFB sites adjacent to the Catawba and Broad basins, where this species is abundant. To determine the likely origin of the UFB populations, the 1140 bp mitochondrial cytochrome-b gene (*CYT-B*) was first amplified and sequenced. Then, a NCBI GenBank BLAST was performed to identify sequences within this database sharing the highest identity percentage similarity to UFB sequences. Haplotypes represented by UFB *CYT-B* sequences and all *C. funduloides CYT-B* sequences available from the Genbank database were then subjected to Tamura-Nei analysis to establish phylogenetic relationships. Divergence time estimates were calculated using nucleotide percent divergence values returned the accepted molecular clock estimate for Cyprinidae *CYT-B*. Results indicate that both UFB populations consist of haplotypes sharing >98% similarity to *C. funduloides* from the adjacent Catawba basin, including a haplotype sharing 100% identity similarity. All UFB and Catawba *CYT-B* haplotypes form a clade which is separated by 2.64 MY from the clade formed by all other represented haplotypes. It can be concluded that some individuals in the UFB populations represent recent introduction from an extant Catawba population, while other UFB specimens represent Pleistocene-divergent haplotypes that may also be endemic to the Catawba, introduced from other river basins bordering the UFB, or may represent an endemic UFB population. This represents the first attempt to comprehensively study the newly discovered UFB populations. Future research will serve to confirm origin of other haplotypes present in the UFB, further resolve phylogenetic relationships of regional populations, and will serve to inform any necessary habitat conservation policy.

## 1. Introduction

The rosyside dace (*Clinostomus funduloides Girard*) is a minnow (Cyprinidae) common to rocky pools within fast-running headwaters and streams throughout its range in the eastern United States [1,2]. Though the Upper French Broad drainage basin in the North Carolina Appalachians sits squarely within this range [2], and meets multiple criteria for suitable habitat [3,4], *C. funduloides* has not been officially catalogued within the Upper French Broad (UFB) system. Recently, however, populations of *C. funduloides* have been discovered in UFB headwaters immediately west of the Eastern Continental Divide (D. Gillette, University of North Carolina at Asheville, and W. Thomas Russ of the North Carolina Wildlife Resources Commission; personal communication). Given the close proximity of these sites both to the Catawba and Broad headwaters, where *C. funduloides* is abundant, bait-bucket transport or other means of accidental introduction are suspected [5].

Species introduction to non-endemic areas is common, but it is also possible that ostensibly introduced species are in fact endemic [5]. With respect to the southern Appalachian river basins, ancestral *C. funduloides* populations were likely separated and isolated by the formation of discrete lakes and river systems upon Miocene continental uplift [6], then further divided and united by multiple glaciation and recession events during the Pleistocene [1]. In a closely related species, *Richardsonius egregious (Lahontan redside)*, comparative molecular dating of the mitochondrial *cytochrome-b* gene (*CYT-B)* suggested initial vicariance of a Pliocene ancestral population giving rise to two distinct clades, each in turn yielding multiple divergent haplotypes through the Pleistocene as these populations were further divided by changing drainage patterns [7]. Though the exact nature and timeline of the southern Appalachians' formation may be debated, the similar geologic history of the region is likely reflected by divergence patterns within its aquatic fauna [1], including *C. funduloides* populations isolated by the formation of mountain-bound river basins.

Comparison of *CYT-B* is therefore particularly useful in determining relationships both among related species [8-10] and among conspecific populations [5,7,11]. Maternally inherited through cytoplasm, the highly conserved *CYT-B* region reflects a population's evolutionary history. The slow rate of accumulated point mutations in *CYT-B* distinguishes among populations, and comparative analysis of these mutations provides estimates of the likely divergence timeline [1,7,10]. These comparative data therefore establish phylogenetic relationships between species and populations [7,10], trace their origin [7], and often shape and support conservation policy [12,13].

Thus, this research used *CYT-B* to determine likely origin of newly discovered UFB populations of *C. funduloides* in Buncombe County and Translylvania County, North Carolina. Given the species' absence from official, published wildlife records of the UFB basin to date, these populations are likely the result of recent human activity, namely the accidental introduction of individuals from Catawba or Broad headwaters. This work represents the first comprehensive study of these UFB *C. funduloides* populations, and results will not only answer questions raised upon their discovery, but will also serve to inform state and federal bureaus regarding any eradication or conservation policy necessary.

## 2. Methods

Four specimens of *C. funduloides* were collected from each of two sites: Flat Creek in Buncombe County, North Carolina (Lat. 35.6384, Lon. -82.3112; Pop. 1), and from Crab Creek in Transylvania County, North Carolina (Lat. 35.2345, Lon. -82.6188; Pop. 2). These were euthanized by immersion in Tricaine-S tricaine methanosulfonate solution from Western Chemical Inc. ™, and stored at -20 ºC. Specimens were thawed at room temperature, and DNA was extracted from muscle tissue via Qiagen DNeasy Blood & Tissue Kit™, using spin-column protocol. DNA concentration was measured by Nanodrop™ spectrophotometer, and *CYT-B* was amplified using previously published primers in the following 50 µL PCR [11]: 10 µL Promega 2.5 mM dNTP Mix, 5 µL New England BioLabs 10x Standard Taq reaction buffer, 5 µL Fisher 25 mM $MgCl_2$, 1 µL each of forward and reverse primer (10 mM) from Eurofins Genomics, 0.5 µL Fisher *Taq* polymerase, 25.5 µL PCR water, and 2 µL template DNA. Samples were cycled in a BioRad T100™ thermal cycler with the following program: initial denaturation at 94 ºC for 3 minutes; 30 cycles consisting of denaturation at 94 ºC for 30 seconds, annealing at 52 ºC for 30 seconds, and extension at 72 ºC for 1 minute; followed by a final extension at 72 ºC for 10 minutes. PCR products were subjected to electrophoresis on ethidium bromide-stained 0.7% agarose gel at 90 volts, then visualized under ultraviolet light, using Carestream MI™ imaging software and apparatus.

Samples exhibiting PCR products of ~1140 nucleotides were purified according to QIAquick PCR Purification Kit™ protocol adjusted for 40 µL PCR product. Purified PCR product concentration was then measured via Nanodrop™ spectrophotometer, and samples were prepared for both forward and reverse sequencing by the following: First, an aliquot of PCR product yielding ~70 ng DNA was dispensed into a 1.5 mL microcentrifuge tube. To this volume, 1 µL of either forward or reverse 10 mM primer and a volume of PCR water were added, for a total of 12 µL. Samples were then submitted to North Carolina State University Genomic Sciences Laboratory for Sanger sequencing.

Raw forward and reverse sequence files were subjected to NCBI GenBank BLAST to confirm identity as *C. funduloides CYT-B,* and sequences were manually spliced, aligned, and verified with respect to any potential mutations noted. There were no insertions nor deletions in final sequences. A second NCBI GenBank nucleotide BLAST was performed on final sequences to determine the GenBank subject sequence with the highest identity percentage relative to UFB queries, and the accession number of the subject sequence was cross referenced with published literature to determine the source of the specimen.

Finally, all haplotypes represented by UFB specimens, as well as those represented by all other *C. funduloides* sequences currently registered in GenBank, were subjected to Jukes-Cantor and Tamura-Nei analyses, using Geneious R10™ software. Both models used a neighbor-joining tree-build method with bootstrap resampling of 100,000 replicates, and node representation acceptance threshold of 50%. Following determination of phylogenetic topology, divergence time estimates were calculated from nucleotide divergence percentages returned by Tamura-Nei analysis, using currently accepted molecular clock estimates for Cyprinidae *CYT-B* [1,10].

## 3. Results

PCR amplification of *CYT-B* yielded product for Pop. 1 samples RD1, RD2, RD3, RD4, and for Pop. 2 samples RD11, RD12, and RD13 (Fig.1). Full 1140 bp *CYT-B* sequences were obtained for RD1, RD2, RD11, RD12, and RD13, while partial sequences were obtained for RD3 (795 bp) and RD4 (794 bp). NCBI GenBank BLAST confirmed identification as *C. funduloides CYT-B* for all Pop. 1 and Pop. 2 sequences, with four haplotypes represented: RD2, RD4, and RD11 all shared 100% nucleotide identity similarity with NCBI GenBank accession ID JX442999.1, a specimen collected from Mill Creek in the Catawba basin, McDowell County, North Carolina [14]. RD1 shared >99% identity similarity with JX442999.1, and had a transition at nucleotide 51 (Fig 2). RD3 also shared >99% identity similarity with JX442999.1, shared the RD1 transition at nucleotide 51, had a transversion at nucleotide 354, and two additional transitions at nucleotides 651 and 669 (Fig 3). RD12 and RD13 shared 98% identity similarity with JX442999.1, shared the RD1 transition at nucleotide 51, and had 17 additional transitions at nucleotides 30, 231, 351, 405, 408, 450, 456, 468, 618, 630, 720, 741, 807, 909, 915, 943, and 996 (Fig 4). Except for nucleotide 943 at the codon first position in RD12 and RD13, all other mutations were in the codon's third position. Still, no mutation altered amino acid sequences in any haplotypes represented, respective to JX442999.1.

Jukes-Cantor and Tamura-Nei analyses each returned phylogenetic trees with identical topology and highly similar bootstrap values. As expected, however, given the differing rates of nucleotide substitution assumed in the Tamura-Nei model, relative percent divergence values differed slightly from those of the Jukes-Cantor model, and were afforded preference accordingly. For all Tamura-Nei model nodes, bootstrap values were above 69, suggesting a sufficiently high degree of confidence in the consensus tree topology (Fig. 6). As such, the phylogeny suggests two distinct regional clades of *C. funduloides*. One includes haplotypes of Pop. 1, Pop. 2, and GenBank JX442999.1, all from western North Carolina (NC). The second includes haplotypes from Big Richland Creek in Humphreys County, Tennessee (TN), and from an undisclosed site in the Southeast region (SE) near Knoxville, TN [15] (Fig. 6).

The NC clade shows an estimated divergence of 2.08 MY from its common ancestor with the TN/WV/SE clade (Table 1). Within the NC clade, the three haplotypes represented in Pop. 1 by RD1, RD3, and by RD2, RD4 (sharing 100% identity with GenBank JX442999.1) are the most closely related, at a maximum divergence range of 290,000 years. As a sub-clade, these are divergent by at least 420,000 years from the ancestor in common with the Pop. 2 haplotype represented by RD12 and RD13 (Table 1, Fig. 6). Of the TN/SE clade, the TN haplotypes show a similar divergence range to that of the Pop. 1 haplotypes, at a maximum range of 230,000 years (Table 1, Fig. 6). The WV haplotype, likely representing a different regional clade, shows a divergence of 2.33 MY from the ancestor in common with that giving rise to the TN and NC clades (Fig. 6).

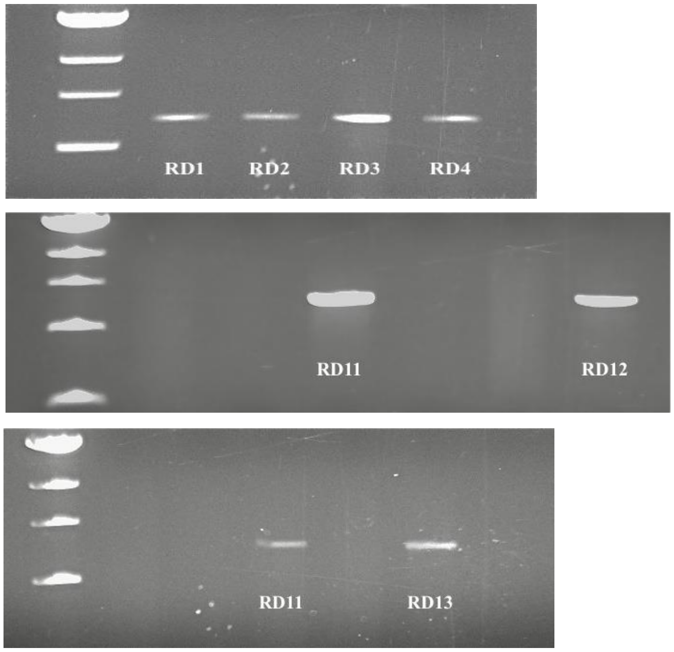

Figure 1. Visualization of ethidium bromide-stained 0.7% electrophoresis gel of *CYT-B* PCR amplification product, confirming presence of ~1136 base-pair fragment amplified from DNA extraction of Pop. 1 samples RD1, RD2, RD3, and RD4 (top), and from Pop. 2 samples RD11, RD12, and RD13 (center, bottom). The 1 kb standard at the left of each gel indicates migration distance of fragment lengths 3 kb, 2 kb, 1.5 kb, 1kb, and 0.5 kb, respectively, from top to bottom of ladder portions depicted.

RD2, RD4, RD11

```
   1 ATGGCAAGCC TACGAAAAAC GCACCCACTA ATAAAAATCG CTAATGATGC ACTAGTTGAC
  61 CTACCAACGC CATCCAATAT TTCCGTAATA TGAAACTTCG GATCCCTCCT AGGATTATGT
 121 TTAATTACCC AAATCCTGAC AGGGCTATTC TTAGCCATAC ACTACACCTC TGATATCTCA
 181 ACTGCATTCT CATCTGTAAC ACATATCTGC CGAGATGTTA ACTATGGTTG GCTCATTCGA
 241 AACATACATG CCAACGGCGC CTCATTCTTC TTCATCTGTA TTTATATACA TATTGCCCGG
 301 GGCCTATACT ATGGGTCTTA CCTTTATAAG GAGACCTGAA ACATTGGCGT AGTTCTACTT
 361 CTTCTGGTAA TAATAACAGC CTTCGTTGGC TACGTTCTGC CATGGGGGCA AATATCCTTT
 421 TGAGGTGCCA CCGTAATTAC AAATCTACTG TCAGCAGTTC CTTATATGGG CGACACCCTC
 481 GTCCAATGAA TCTGAGGGGG CTTCTCAGTA GATAACGCAA CGCTAACGCG ATTCTTCGCC
 541 TTCCATTTCC TCTTTCCATT CGTCATCGCC GGTGCAACCA TTTTACACCT ACTCTTCTTA
 601 CACGAAACGG GATCAAATAA CCCCGCTGGA CTAAACTCTG ACGCCGATAA AATTTCTTTC
 661 CACCCGTATT TCTCGTATAA AGATCTTCTT GGCTTTGTAC TAATACTATT GGCTCTTACG
 721 TCGCTAACTT TATTTTCCCC AACTCTCCTC GGTGACCCAG AGAATTTTAC CCCAGCAAAC
 781 CCGCTGGTCA CCCCACCACA TATTCAGCCT GAATGATACT TCCTATTTGC CTACGCCATC
 841 CTACGATCCA TTCCAAACAA GCTAGGAGGA GTCCTAGCCC TGTTGTTTAG CATTTTAGTG
 901 CTCATGGTTG TACCAATTTT ACACACCTCA AAACAACGAG GATTAACTTT CCGCCCAGTT
 961 ACCCAATTTC TATTCTGAAC CCTAGTAGCA GATATGATTA TCCTGACATG AATTGGAGGC
1021 ATACCCGTAG AACACCCATA CATTATCATC GGCCAAGTCG CGTCGGTCCT ATACTTTGCA
1081 CTATTCCTCG TTCTTGCCCC GCTTGCAGGG TGAGTGGAAA ATAAAGCATT AAAATGAGCT
```

Figure 2. Full *CYT-B* sequence representing haplotype shared by specimens RD2, RD4, and RD11, and sharing 100% nucleotide identity similarity with NCBI GenBank accession JX442999.1.

RD1

1 ATGGCAAGCC TACGAAAAAC GCACCCACTA ATAAAAATCG CTAATGATGC **G**CTAGTTGAC
61 CTACCAACGC CATCCAATAT TTCCGTAATA TGAAACTTCG GATCCCTCCT AGGATTATGT
121 TTAATTACCC AAATCCTGAC AGGGCTATTC TTAGCCATAC ACTACACCTC TGATATCTCA
181 ACTGCATTCT CATCTGTAAC ACATATCTGC CGAGATGTTA ACTATGGTTG GCTCATTCGA
241 AACATACATG CCAACGGCGC CTCATTCTTC TTCATCTGTA TTTATATACA TATTGCCCGG
301 GGCCTATACT ATGGGTCTTA CCTTTATAAG GAGACCTGAA ACATTGGCGT AGTTCTACTT
361 CTTCTGGTAA TAATAACAGC CTTCGTTGGC TACGTTCTGC CATGGGGGCA AATATCCTTT
421 TGAGGTGCCA CCGTAATTAC AAATCTACTG TCAGCAGTTC CTTATATGGG CGACACCCTC
481 GTCCAATGAA TCTGAGGGGG CTTCTCAGTA GATAACGCAA CGCTAACGCG ATTCTTCGCC
541 TTCCATTTCC TCTTTCCATT CGTCATCGCC GGTGCAACCA TTTTACACCT ACTCTTCTTA
601 CACGAAACGG GATCAAATAA CCCCGCTGGA CTAAACTCTG ACGCCGATAA AATTTCTTTC
661 CACCCGTATT TCTCGTATAA AGATCTTCTT GGCTTTGTAC TAATACTATT GGCTCTTACG
721 TCGCTAACTT TATTTTCCCC AACTCTCCTC GGTGACCCAG AGAATTTTAC CCCAGCAAAC
781 CCGCTGGTCA CCCCACCACA TATTCAGCCT GAATGATACT TCCTATTTGC CTACGCCATC
841 CTACGATCCA TTCCAAACAA GCTAGGAGGA GTCCTAGCCC TGTTGTTTAG CATTTTAGTG
901 CTCATGGTTG TACCAATTTT ACACACCTCA AAACAACGAG GATTAACTTT CCGCCCAGTT
961 ACCCAATTTC TATTCTGAAC CCTAGTAGCA GATATGATTA TCCTGACATG AATTGGAGGC
1021 ATACCCGTAG AACACCCATA CATTATCATC GGCCAAGTCG CGTCGGTCCT ATACTTTGCA
1081 CTATTCCTCG TTCTTGCCCC GCTTGCAGGG TGAGTGGAAA ATAAAGCATT AAAATGAGCT

Figure 3. Full *CYT-B* sequence representing haplotype of RD1, sharing >99% nucleotide identity similarity with NCBI GenBank accession JX442999.1, with transition of the codon's third position at nucleotide 51 (bold, underlined).

RD3

1 ATGGCAAGCC TACGAAAAAC GCACCCACTA ATAAAAATCG CTAATGATGC **G**CTAGTTGAC
61 CTACCAACGC CATCCAATAT TTCCGTAATA TGAAACTTCG GATCCCTCCT AGGATTATGT
121 TTAATTACCC AAATCCTGAC AGGGCTATTC TTAGCCATAC ACTACACCTC TGATATCTCA
181 ACTGCATTCT CATCTGTAAC ACATATCTGC CGAGATGTTA ACTATGGTTG GCTCATTCGA
241 AACATACATG CCAACGGCGC CTCATTCTTC TTCATCTGTA TTTATATACA TATTGCCCGG
301 GGCCTATACT ATGGGTCTTA CCTTTATAAG GAGACCTGAA ACATTGGCGT AGT***G***CTACTT
361 CTTCTGGTAA TAATAACAGC CTTCGTTGGC TACGTTCTGC CATGGGGGCA AATATCCTTT
421 TGAGGTGCCA CCGTAATTAC AAATCTACTG TCAGCAGTTC CTTATATGGG CGACACCCTC
481 GTCCAATGAA TCTGAGGGGG CTTCTCAGTA GATAACGCAA CGCTAACGCG ATTCTTCGCC
541 TTCCATTTCC TCTTTCCATT CGTCATCGCC GGTGCAACCA TTTTACACCT ACTCTTCTTA
601 CACGAAACGG GATCAAATAA CCCCGCTGGA CTAAACTCTG ACGCCGATAA **G**ATTTCTTTC
661 CACCCGTA**C**T TCTCGTATAA AGATCTTCTT GGCTTTGTAC TAATACTATT GGCTCTTACG
721 TCGCTAACTT TATTTTCCCC AACTCTCCTC GGTGACCCAG AGAATTTTAC CCCAGCAAAC
781 CCGCTGGTCA CCCCA----- ---------- ---------- ---------- ----------

Figure 4. Partial *CYT-B* sequence representing haplotype of RD3, sharing >99% nucleotide identity similarity with NCBI GenBank accession JX442999.1, with transitions of the codon's third positon at nucleotides 51, 651 and 669 (bold, underlined), and transversion of the codon's third position at nucleotide 354 (bold, underlined, italicized).

RD12, RD13

1 ATGGCAAGCC TACGAAAAAC GCACCCACT**G** ATAAAAATCG CTAATGATGC **G**CTAGTTGAC

61 CTACCAACGC CATCCAATAT TTCCGTAATA TGAAACTTCG GATCCCTCCT AGGATTATGT

121 TTAATTACCC AAATCCTGAC AGGGCTATTC TTAGCCATAC ACTACACCTC TGATATCTCA

181 ACTGCATTCT CATCTGTAAC ACATATCTGC CGAGATGTTA ACTATGGTTG **A**CTCATTCGA

241 AACATACATG CCAACGGCGC CTCATTCTTC TTCATCTGTA TTTATATACA TATTGCCCGG

301 GGCCTATACT ATGGGTCTTA CCTTTATAAG GAGACCTGAA ACATTGGCGT **G**GTTCTACTT

361 CTTCTGGTAA TAATAACAGC CTTCGTTGGC TACGTTCTGC CATG**A**GG**A**CA AATATCCTTT

421 TGAGGTGCCA CCGTAATTAC AAATCTACT**A** TCAGC**G**GTTC CTTATAT**A**GG CGACACCCTC

481 GTCCAATGAA TCTGAGGGGG CTTCTCAGTA GATAACGCAA CGCTAACGCG ATTCTTCGCC

541 TTCCATTTCC TCTTTCCATT CGTCATCGCC GGTGCAACCA TTTTACACCT ACTCTTCTTA

601 CACGAAACGG GATCAAA**C**AA CCCCGCTGG**G** CTAAACTCTG ACGCCGATAA AATTTCTTTC

661 CACCCGTATT TCTCGTATAA AGATCTTCTT GGCTTTGTAC TAATACTATT GGCTCTTAC**A**

721 TCGCTAACTT TATTTTCCCC **G**ACTCTCCTC GGTGACCCAG AGAATTTTAC CCCAGCAAAC

781 CCGCTGGTCA CCCCACCACA TATTCA**A**CCT GAATGATACT TCCTATTTGC CTACGCCATC

841 CTACGATCCA TTCCAAACAA GCTAGGAGGA GTCCTAGCCC TGTTGTTTAG CATTTTAGTG

901 CTCATGGT**C**G TACC**G**ATTTT ACACACCTCA AAACAACGAG GA***C***TAACTTT CCGCCCAGTT

961 ACCCAATTTC TATTCTGAAC CCTAGTAGCA GATAT**A**ATTA TCCTGACATG AATTGGAGGC

1021 ATACCCGTAG AACACCCATA CATTATCATC GGCCAAGTCG CGTCGGTCCT ATACTTTGCA

1081 CTATTCCTCG TTCTTGCCCC GCTTGCAGGG TGAGTGGAAA ATAAAGCATT AAAATGAGCT

Figure 5. Full *CYT-B* sequence representing haplotype of RD12 and RD13, sharing 98% nucleotide identity similarity with NCBI GenBank accession JX442999.1, with transitions of the codon’s third position at nucleotides 30, 51, 231, 351, 405, 408, 450, 456, 468, 618, 630, 720, 741, 807, 909, 915, and 996 (bold, underlined), and at the codon’s first position at nucleotide 943 (bold, underlined, italicized).

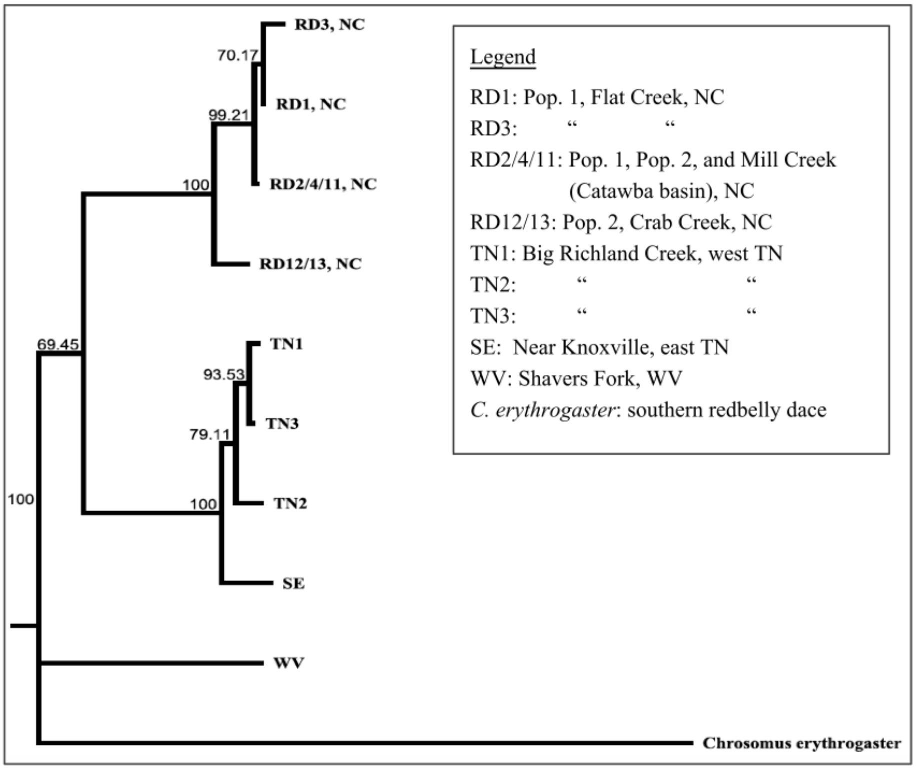

Figure 6. Tamura-Nei consensus tree of haplotypes represented by both UFB populations, GenBank accession JX442999.1 from Mill Creek in the Catawba basin, and all other *C. funduloides CYT-B* sequences available in the

GenBank database, showing bootstrap values at branch nodes, and indicating relative percent nucleotide divergence via branch length.

Table 1. divergence estimates of haplotypes within and among clades determined by Tamura-Nei phylogenetic analysis.

| **Subject haplotype/clade** | **Divergence (MY)** | **from ancestor in common with** |
|---|---|---|
| Pop.1 / Catawba | 0.027–0.29 | –Same Clade– |
| TN, Big Richland Creek | 0.037–0.23 | –Same Clade– |
| Pop. 1 / Catawba | 0.72 | RD12/13, Pop. 2 |
| TN/SE Clade | 1.83 | NC Clade |
| NC Clade | 2.08 | TN/SE Clade |
| WV | 2.33 | NC and TN Clades |
| *Chrosomus erythrogaster* | 6.81 | *C. funduloides* |

## 4. Discussion

These data do support the hypothesis of recent introduction to the UFB, given the 100% nucleotide identity similarity of RD2, RD4, and RD11 to GenBank accession JX442999.1, collected from Mill Creek in the Catawba basin just opposite the Eastern Continental Divide (ECD) from the Pop.1 site. Given the estimated time of these river basins' geologic separation [1,6], the molecular clock estimates for Cyprinidae *CYT-B* [1,10], and the divergence time estimates of a closely related species undergoing similar vicariance events during the same periods [7], it is unlikely that an endemic Catawba basin population would share an identical haplotype with any endemic UFB population separated at least since the Pleistocene. Though back mutation is possible, likelihood of reversion only in the same nucleotide(s) is low. Since *C. funduloides* has not been officially catalogued from the UFB before now, and since the Catawba basin population from Mill Creek in McDowell County, NC is separated by merely ~5 km across the ECD, this further suggests introduction into the Pop.1 UFB site from this nearby Catawba population with respect to the haplotype represented by RD2 and RD4. As RD11 in the Pop. 2 UFB site also shares the same GenBank JX442999.1 haplotype, and given the ~60 km distance separating this site from the Mill Creek site across the ECD, this also suggests recent introduction.

Regarding RD1, RD3, and RD12/13 haplotypes, however, these data suggest two possibilities: that one or more endemic UFB populations did diverge from neighboring river basin populations and have escaped notice until now, or that multiple introduction events from other sources have yielded these multiple haplotypes represented in the UFB. Given the relatedness of RD1 and RD3 haplotypes to that of RD2/4/11 and Genbank JX442999.1, following a similar pattern to that of TN haplotypes from Big Richland Creek in Humphreys County (Table 1, Fig. 6), it is possible that these are also endemic to the Catawba basin at or near Mill Creek, and represent a diverse population at this site. In this case, it is possible that all Pop. 1 UFB haplotypes were introduced from the adjacent Catawba basin. Whether the Catawba diversity follows separation and reunification events through repeated Pleistocene glaciation, or whether this population itself reflects multiple introductions from discrete drainages along the Catawba or from adjacent river basins, additional genetic data from these regions will likely resolve the relationships.

To this end, additional data will undoubtedly serve to further resolve the relationship of the Pop. 2 UFB haplotype represented by RD12 and RD13. If both UFB populations were introduced from the same source population, we can expect identical haplotype representation, provided that a sufficient number of introduced individuals avoids any significant bottleneck effect. Given the small sample size from each of these UFB sites (n=4), it is possible that additional collection will confirm identical haplotype profiles at both locations. If, however, different haplotype profiles reflect either significant bottleneck effect or introduction from multiple sites, a systematic survey and collection from all river basins bordering the UFB will serve to determine origin of RD12/13, as well as RD1 and RD3, and will further resolve phylogenetic relationships of *C. funduloides* in the southern Appalachian region.

While it is possible that an endemic UFB population diverged from populations currently bordering the UFB, it is apparent that both this study and the GenBank database currently lack comprehensive comparative data to support this. The phylogeny here (Fig. 6) may reflect an eastern and western clade separated by the ECD, as similarly described by Houston *et al*. [7], but it cannot be demonstrated without sufficient sampling from all regional drainage basins. Given the profile of point mutations that the RD12/13 haplotype shares with the TN/WV/SE clade (Fig. 6), and given its estimated divergence time from haplotypes RD1, RD3 and that represented by GenBank JX442999.1 (Table 1), it may also represent an easternmost population along a cline approaching the ECD. If this is not the case, likely introduction from the nearby Broad, Savannah, or Tuckasegee basins may be demonstrated via additional genetic data.

Even with sufficient data, however, it is possible that a most parsimonious phylogeny may display apparent gaps and incongruities [7], and that successive translocations and introductions by human activity may complicate determining native haplotypes within a site or region [5,7]. Given the *C. funduloides* behavior of spawning over other species' nests, hybridization is also possible. Including genetic data from other loci may therefore be appropriate in further study. Demonstrating the existence of an endemic UFB population may therefore pose a challenge, but the results of future research may indicate whether any conservation measures are warranted. Given the high sensitivity of *C. funduloides* to pollution and siltation, and its preference for adequate riparian vegetative cover [13], it is possible that endemic UFB populations declined and disappeared in a similar pattern as observed in other regions [4]. If so, any refugia deserve further attention, especially considering the extensive agricultural development of land bordering the UFB Pop. 2 site. Determining a comprehensive phylogeny of all regional populations, regardless of challenge, is therefore imperative to drafting any potential conservation policy.

With current results here, it cannot be maintained with certainty that a unique population of *C. funduloides* exists in the UFB. By systematically sampling populations within a given radius of the UFB population, future research may resolve regional phylogeny and eliminate the possibility that all population haplotypes were recently introduced from bordering river basins. Additionally, future sample registry will also provide comparative data to supplement any lacking from GenBank or other publicly accessible databases. While the origin of a potential UFB population is yet unclear, it is perhaps more unclear, if any population is indeed endemic, how it escaped official notice until recently. Given habitat loss and destruction due to widespread logging and development through the last century, it is possible that these activities all but eliminated an endemic UFB population [3,1]. In any case, the existence of *C. funduloides* in the UFB basin is a curiosity that demands future attention.

## 5. Acknowledgements

In addition to Dr. Jennifer Ward and Dr. David Gillette, the author wishes to extend thanks to Dr. Rebecca Hale, Dr. Matthew Greene, Dr. Graham Reynolds, Dr. David Clarke, W. Thomas Russ, the North Carolina Wildlife Resources Commission, and to Professor Landon Ward, Alexander Lappie, Torey Todd, Biology 443, and Henry Gates for their contributions to this research.

## 6. References

1. Berendzen PB, Simons AM, Wood RM, Dowling TE, Secor CL. Recovering cryptic diversity and ancient drainage patterns in eastern North America: Historical biogeography of the *Notropis rubellus* species group (Teleostei: Cypriniformes). Mol Phylogen Evol. 2008;46:721-737.
2. NatureServe, Freshwater Biodiversity Unit, IUCN: *Clinostomus funduloides* [Internet]. c2016. The IUCN Red List of Threatened Species, Version 2016-3 [updated 2013; cited 2017 Mar 20]. Available from: http://maps.iucnredlist.org/map.html?id=184083
3. Meitzen KM. Stream flow changes across North Carolina (USA) 1955-2012 with implications for environmental flow management. Geomorphology. 2016;252:171-184.
4. Rice DL, Phinney GJ. Distribution and status of the rosyside dace, *Clinostomus funduloides* Girard, (Cyprinidae) in southern Ohio. Ohio J Sci. 1985;85(4):159-164.
5. Scott CH, Cashner M, Grossman GD, Wares JP. An awkward introduction: phylogeography of *Notropis lutipinnis* in its 'native' range and the Little Tennessee River. Ecol of Freshwater Fish. 2009;18:538-539.

6. Gallen SF, Wegmann KW, Bohnenstiehl DR. Miocene rejuvenation of topographic relief in the southern Appalachians. GSA Today. 2013;23(2):4-10.

7. Houston DD, Shiozawa DK, Riddle BR. The roles of Neogene geology and late Pleistocene lake levels in shaping the genetic structure of the Lahontan redside shiner *Richardsonius egregius* (Teleostei: Cyprinidae). Biol J Linn Soc. 2011;104:163-176.

8. Ghorashi SA, Fatemi SM, Amini F, Houshmand M, Salehi-Tabar R, Hazaie K. Phylogenetic analysis of anemone fishes of the Persian Gulf using mtDNA sequences. Afr J Biotech. 2008;7(12):2074-2080.

9. Nwafili SA, Gao T. Genetic diversity in the mtDNA control region and population structure of *Chrysichthys nigrodigitatus* from selected Nigerian rivers: Implications for conservation and aquaculture. Arch Pol Fish. 2016;24:85-97.

10. Houston DD, Shiozawa DK, Riddle BR. Phylogenetic relationships of the western North American cyprinid genus, Richardsonius, with an overview of phylogeographic structure. Mol Phyl Evol. 2010;55:259-273.

11. Yang J, Hsu K, Liu Z, Kuo P, Tang W, Zhou Z, Liu D, Bao B, Lin H. The population history of *Garra orientalis* (Teleostei: Cyprinidae) using mitochondrial DNA and microsatellite data with approximate Bayesian computation. Evol Biol. 2016;16(73):[15 p.].

12. Alves MJ, Coelho H, Collares-Pereira MJ, Coelho MM. Mitochondrial DNA variation in the highly endangered cyprinid fish *Anaecypris hispanica*: importance for conservation. Heredity. 2001;87:463-473.

13. Kubach K, Scott M. Rosyside dace: *Clinostomus funduloides*. SC SWAP, Supplemental Volume: Species of conservation concern. 2015;[3 p.]. Available from: http://www.dnr.sc.gov/swap/supplemental/freshwaterfish/rosysidedace2015.pdf

14. Schönhuth S, Shiozawa DK, Dowling TE, Mayden RL. Molecular systematics of western North American cyprinids ( Cypriniformes: Cyprinidae). Zootaxa. 2012;3586:281-303.

15. Hulsey CD, Hollingsworth PR. Do constructional constraints influence cyprinid (Cyprinidae: Leuciscinae) craniofacial coevolution?. Biol J Lin Soc. 2011;103:136-146.

16. Utz RY, Hilderbrand RH, Raesly RL. Regional differences in patterns of fish species loss with changing land use. Biol Conserv. 2010;143:688-699.