Assessing the prevalence of common patterns and unique events in the formation of biotas: a study of fish taxa of the North American central highlands

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

DOMINIK HALAS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ANDREW SIMONS, ADVISOR
AUGUST 2011

Copyright Dominik Halas 2011

Acknowledgements

None of this work would have been possible without the help of a great many people. First of all, I must thank my advisor Andrew Simons, for his guidance as this project changed shape again and again over the years, and most of all for his incredible patience during all the times when it seemed that no progress was being made. I also thank my committee members, Sharon Jansa, George Weiblen, and Scott Lanyon for their advice and discussion. Keith Barker and Ken Kozak also helped considerably with many of the analyses that would otherwise have been opaque to me.

I thank Jacob Egge, Brett Nagle, Peter Hundt, Brett Young, and Michael Ghedotti for their help with collecting specimens in the field, and Brandon Demuth and Belinda Befort for help in the lab. Peter Berendzen, Rick Mayden, Rob Wood, Bernie Kuhajda, Thomas Near, and Jacob Egge all loaned or donated specimens which contributed greatly to the quality of the analyses. Funding from the Dayton Wilkie Fellowships of the Bell Museum at the University of Minnesota, and from a National Science Foundation Doctoral Dissertation Improvement Grant made much of the analysis possible.

I must thank the various friends I have made during my six years in Minnesota, all of whom contributed to the study in various ways, even if only by helping me to relax when I wasn't working on the project. Alexi Powell, Brian Barber, Wendy Clement, Ben Lowe, Bailey McKay, Hernan Vazquez, and Jake Musser were not only friends, but also helped at one point or another by explaining some analytical method to me, or suggesting a new direction for me to investigate. I must particularly thank Dustin Haines, Will Ratcliff, Becky Stark, Alexi Powell and Belinda Befort for being wonderful friends to me in good times and bad.

Finally, I must thank my parents, for their constant support and unending patience, and I hope that they will find that the three years I've taken to finish a Master's degree, and six years to finish a Ph. D. have been worthwhile.

Abstract

The Central Highlands fish fauna is a complex assemblage of hundreds of species, many showing a high degree of endemicity. This diversity has traditionally been explained by invoking relatively simple vicariance and dispersal scenarios. In my study, I assembled all existing phylogenies of fishes from the Central Highlands to conduct a Phylogenetic Analysis for the Comparison of Trees, which revealed that vicariance and dispersal events have both played a role in the formation of the Central Highlands ichthyofaunas. Evidence of co-ordinated dispersal and area reticulation was also discovered. The fauna of some regions, in particular the Tennessee River basin, is shown to have accumulated due to multiple vicariance and dispersal events.

One mitochondrial and seven nuclear loci were sequenced for the members of the *Etheostoma zonale* species group. A high degree of geographical structure was found, with twelve reciprocally monophyletic mitochondrial clades in *E. zonale*, and nine in *E. lynceum*. Species tree analysis demonstrated the monophyly of both species. One clade found in the Upper Tennessee River basin was found to have undergone introgression with a second clade.

One mitochondrial and three nuclear loci were sequenced for the members of the *Luxilus zonatus* species group. Four mitochondrial introgression events were found in the group, involving all three members of the *L. zonatus* group as well as two other species of *Luxilus*. The geographical extent of the introgression events varies, as does their time of occurrence.

A Hierarchical Approximate Bayesian Computation was performed on a set of sister clades of three taxa found in the Ozark and Ouachita Mountains: two clades of *Etheostoma zonale*, two clades of *E. blennioides*, and *Luxilus cardinalis* and *L. pilsbryi*. All taxon pairs share the same geographical split; the analysis showed that two of these divergences did not happen at the same time as the third, and are thus pseudocongruent.

The results of this study show that the Central Highlands ichthyofauna has a more complex history than has been believed, and suggest methods of reconstructing such complex histories which should prove useful in biogeographical studies of similarly complex systems.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

One of the simpler questions that can be asked, but harder questions to be answered, about a group of species found together in the same area, is the question of how those species came to be found in that place together. The ways in which species assemblages form, and the mechanisms leading to their formation, are an ongoing focus of research in a number of biological disciplines, including historical biogeography. Until recently, the systematic study of such assemblages has presupposed that the histories of species assemblages are simple, made up of a single pattern of two types of events: vicariance, in which species found over a broad area divide in two due to the formation of barriers to the movement of individual organisms, and dispersal, in which individual species colonize new areas and then diverge (e.g. Ronquist 1997). Vicariant events are expected to form common patterns of species-area relationships among different taxa, because the free movement of multiple unrelated populations is impeded in the same way at the same time, whereas dispersal is not expected to form any patterns, because only the organism doing the dispersing is directly affected. Another common assumption has been that each area of the globe has a single history of vicariant events which have generated its diversity (Humphries 2000). More recently, various phenomena, which under these viewpoints were dismissed as noise, have come to be seen as important in the evolution of biota. One such phenomenon is co-ordinated dispersal (Erwin 1979, 1981), in which multiple taxa disperse into the same area at the same time due to the disappearance of a previously-existing barrier to their dispersal. Even though the mechanism of diversification involves dispersal from one region to another, rather than formation of a barrier resulting in a loss in connectivity among populations, co-ordinated dispersal is still expected to produce common patterns, because multiple species in an ancestral area will be able to disperse into a new area at more or less the same time upon the disappearance of the barrier between the ancestral and new areas. A second phenomenon whose importance has only recently been recognized is area reticulation, the idea that the same area may have been involved in multiple vicariance events over time (Brooks 1990). One specific type of reticulation, pseudocongruence (Cunningham and Collins 1994, Donoghue and Moore 2003), the existence of identical species-area relationships in different taxa due to events occurring at different times, has been viewed more as a limitation on biogeographical methods than an interesting phenomenon in its own right (Soltis et al. 2006). The finite extent of earth's surface, and cyclical patterns of climatic events such as changes in temperature or precipitation, may well result in similar changes in biota occurring at different periods; to determine the significance of these effects to diversification requires treating pseudocongruence as a starting point for further research. The extent to which the histories of biotas are reticulated is still unclear, although several studies have shed some light on this question (Folinsbee and Brooks 2007, Halas et al. 2005, Hembree 2006, Dominguez-Dominguez et al. 2006)

The ichthyofauna of the Central Highlands of the United States provides an excellent model system to study these questions. It is a highly diverse fauna, with several hundred species present (Page and Burr 2011), giving multiple lineages from which data can be obtained. The region in which these lineages are found has been extensively sampled, making the existence of still-undiscovered species that could confound the reconstruction of phylogenies or biogeographical patterns unlikely. The geology of the region is well known (Thornbury 1965, Melhorn and Kempton 1991), so that inferred patterns of diversification can be correlated to historical events. Finally, fish are limited in their available means of dispersal, suggesting that co-ordinated events, if they do occur, are likely to leave still-visible evidence today. Freshwater fishes even have an obvious avenue for co-ordinated dispersal: stream capture events, in which the headwaters of one stream are captured by a more rapidly-eroding neighboring stream. The fishes inhabiting the captured stream all cross a barrier to dispersal (the previously existing drainage divide between the streams) at the same time; if they survive and eventually differentiate, the stream capture event will be seen in retrospect to be a co-ordinated dispersal event.

Many biogeographical studies of Central Highlands fish taxa exist, but very few have compared patterns among taxa. The most notable such study, that of Mayden (1988a) predicted that the biogeographical patterns of Central Highlands fish taxa would follow various pre-Pleistocene drainage patterns whose existence is independently known from geological data; this is known as the pre-Pleistocene vicariance hypothesis. This hypothesis was contrasted to the Central Highlands-Pleistocene dispersal hypothesis, which states that the Pleistocene glaciations, by improving habitat conditions in much of unglaciated southeastern North America, allowed multiple species to disperse from their center of origin in the Tennessee and Cumberland River basins to other areas of the Central Highlands. Mayden's study, however, used only seven clades, and an implicit assumption of the method used, component compatibility analysis (Zandee and Roos 1987), is that there is a single pattern of vicariance which is more or less displayed by each component clade; it is therefore problematic to use such a study as evidence that there is, in fact, one overall biogeographical pattern for Central Highlands fishes. More recent studies of more than one taxon (Turner et al. 1996, Strange and Burr 1997, Berendzen 2005, Soltis et al. 2006) have compared individual phylogenies to Mayden's hypothesis, but have not examined multiple phylogenies together.

To examine these questions concerning the complexity of historical biogeographical systems, I have compiled a set of molecular phylogenies of fish taxa from the Central Highlands, and analyzed them using a method of comparative biogeographical analysis known as Phylogenetic Analysis for the Comparison of Trees (Wojcicki and Brooks 2005). This method allows for the recovery of reticulate histories for an area, so that if the species assemblage of an area has multiple origins, each of these separate origins can, in theory, be obtained from the analysis. To further this analysis, I have also compiled multilocus datasets for two groups of species found in the Central Highlands: the *Etheostoma zonale* species group, which includes one species found throughout the Central Highlands, and one adjacent to them in the Mississippi embayment, and the *Luxilus zonatus* species group, three species found in the Ozark Highlands. Furthermore, I have compiled a set of three sister pairs of divergent taxa found in the same geographical area to explicitly test for pseudocongruence.

New advances in phylogenetic methods, allowing for the increasing use of multilocus data, allow questions concerning phenomena such as pseudocongruence and area reticulation to be answered better than ever before. Species tree methods such as *BEAST (Heled and Drummond 2010) allow for the estimation of species trees from multilocus data, reducing the error inherent in gene tree / species tree incongruence (Knowles 2009). Multilocus estimates of divergence times and numbers of distinct divergence events are possible with Hierarchical Approximate Bayesian Computation, providing a test for the presence of pseudocongruence in a biogeographical data set.

Extended Bayesian skyline plots (Heled and Drummond 2008) allow one to reconstruct the history of population expansion and contraction in taxa, potentially aiding in distinguishing past dispersal events, and particularly co-ordinated dispersal events, from vicariance events.

Using these methods, in this study I reconstruct the general patterns of diversification present in the Central Highlands fish fauna, present the biogeographical history of the *Etheostoma zonale* and *Luxilus zonatus* species groups, and obtain evidence for the existence of pseudocongruence in diversification events among fish taxa in the Ozark mountains.

CHAPTER 2

Common and unique events in the diversification of the fishes of the North American Central Highlands: A comparative biogeographical test of the pre-Pleistocene vicariance hypothesis.

Introduction

One of the main goals of historical biogeography is to recover evidences of the processes and events that are responsible for the biodiversity we see today. Such questions are particularly compelling in areas with an especially high level of diversity. One such area is the Central Highlands of North America, the Appalachian, Ozark and Ouachita Mountains, which has a greater diversity of freshwater fish species than any other area of comparable size outside of the tropics. Of the 831 known species of freshwater fish native to North America north of Mexico, approximately one third are found in the Central Highlands region (Page and Burr 2011). Because of their location, the Central Highlands have been extensively sampled for over a century, so that the ichthyofauna of the region is now well known. Molecular phylogenies are increasingly available for many of these taxa, making possible a comparative biogeographical study to attempt to determine the nature of the events which generated this multitude of fish species.

The North American Central Highlands consist of three separate highland areas: to the west of the Mississippi River, the Ozark Mountains and the Ouachita Mountains, which are separated by floodplain of the Arkansas River, and to the East of the Mississippi, the Eastern Highlands, which include the Appalachian Mountains and other highland areas to their west. The Ozark and Ouachita Mountains together form the Interior Highlands; the Interior Highlands and Eastern Highlands comprise the Central Highlands. The Central Highlands entirely escaped the Pleistocene glaciations, a factor which must have contributed to the high species diversity there today. Within the Central Highlands, most of the streams and rivers are high in gradient. The Arkansas and Mississippi floodplains, which divide the Central Highlands into three, are broad, low-gradient rivers which at present form barriers to the dispersal of highland fishes from one region to another. To the north of the Central Highlands is a formerly glaciated lowland area known as the Central Lowlands (Figure 2.1). Before the Pleistocene much of the Central Lowlands is believed to have been physiogeographically similar to the Central Highlands (Thornbury 1965). The Pleistocene glaciations, which covered almost all of the Central Lowlands, must have caused either the extinction or southward dispersal of much of the original ichthyofauna of this region; many of the fish species found in the Central Lowlands today have been derived largely through dispersal from the Central Highlands.

Nearly the entirety of the Central Highlands is drained by the Mississippi River and its tributaries; the major exception is the Mobile River, which drains the southernmost part of the Eastern Highlands directly into the Gulf of Mexico. The major river systems of the Central Highlands and Central Lowlands are shown in Figure 2.2. The Ouachita Mountains are drained primarily by the Ouachita River. The Ozarks are drained by the Arkansas, White, and Missouri Rivers. That part of the Eastern Highlands within the Mississippi River basin is drained by the Tennessee, Cumberland and Ohio Rivers. A large part of the upper Ohio River basin, along with the Wabash and upper Mississippi River basins, forms that part of the Central Lowlands which drains into the Mississippi. Many of the rivers draining the Central Highlands have their confluence in lowland areas to the north or south of the Highlands; as a result, dispersal by many species between major river systems within each highland area is limited.

Historically, there have been two major hypotheses to account for the diversification of the Central Highlands ichthyofauna. The first, known as the Central Highlands-Pleistocene dispersal hypothesis (Near et al. 2001), was formulated during heyday of theories of centers of origin. This hypothesis states that most Central Highlands species evolved in the Eastern Highlands, which today have the most species within the region, and dispersed westward across the Central Lowlands to the Interior Highlands during the Pleistocene, when climatic conditions and the geological effects of the glaciation created suitable habitat conditions connecting the Central Highlands. This hypothesis predicts that the earliest divergences within Central Highlands taxa should occur among taxa found in the Eastern Highlands, that Central Lowlands taxa should group with those in the Interior Highlands, and that divergences among Interior Highlands taxa should be the most recent.

The second hypothesis, known as the pre-Pleistocene vicariance hypothesis, was first stated by Pflieger (1971) and developed by Mayden (1985, 1987a, b) and Wiley and Mayden (1985). This hypothesis states that much of the ichthyofaunal diversity of the Central Highlands was already in place prior to the Pleistocene, and was shaped by the drainage patterns that existed during the Tertiary, prior to their massive rearrangement as a result of the Pleistocene glaciations. The major differences in drainage pattern between the Pleistocene and today are shown in Figure 2.3. There are a number of specific predictions made by the pre-Pleistocene vicariance hypothesis with regard to these distributions. For instance, the Ohio River was, prior to the Pleistocene, two separate river systems: the Old Ohio River, which drained into the Mississippi in the same area that the Ohio does now, but which included only the lower half of the modern Ohio's drainage basin; and the Teays River, which drained the upper half of the modern Ohio River basin into the upper Mississippi River by way of a channel through northern Indiana and Illinois. Taking this drainage pattern into account, the pre-Pleistocene vicariance hypothesis predicts that species found in the Upper Ohio River will be more closely related to species in the upper Mississippi River basin and the Ozark Highlands than to species in the lower Ohio River basin. Another example is the Appalachian River, the hypothesized former connection of the upper Tennessee River to the Mobile River, rather than the lower Tennessee River. This pattern predicts that a close relationship should exist between species found in the Upper Tennessee and those in the Mobile Basin.

Mayden (1985, 1987a, b) and Wiley and Mayden (1985) accumulated examples of taxa which could shed light on these competing hypotheses, but the first study to actually incorporate a comparative biogeographical analysis of multiple Central Highlands taxa was that of Mayden (1988a), who performed a component compatibility analysis (Zandee and Roos 1987) of seven clades of Central Highlands fish species. Mayden's analysis found support for the pre-Pleistocene vicariance hypothesis over the Pleistocene dispersal hypothesis. Other studies which have examined more than one Central Highlands taxon (Turner et al. 1996, Strange and Burr 1997, Berendzen 2005, Soltis et al. 2006) have not combined the phylogenies of each taxon in a single analysis. Thus, Mayden's 1988 study is still the most recent hypothesis of biogeographical relationships in the Central Highlands ichthyofauna based upon comparative analysis of multiple clades. However, component compatibility analysis can only reconstruct a single set of relationships among biogeographical areas, and so may not be an entirely appropriate method for the analysis of regions with complex histories. Co-ordinated dispersal, in which multiple taxa disperse into the same area at the same time due to barriers disappearing (Erwin 1979, 1981), and area reticulation, the involvement of the same area in multiple vicariance events over time (Brooks 1990) cannot be reconstructed with component compatibility analysis, yet there is reason to believe both may have been important in the diversification of the Central Highlands ichthyofauna. Changes in drainage patterns not only separate populations which once mingled freely, they also allow taxa to enter river systems they were previously barred from dispersing to. Any such event may allow multiple species to enter a new drainage system simultaneously, resulting in co-ordinated dispersal. If the drainage pattern of a region has changed more than once, as is hypothesized to be the case with the Cumberland, Duck, and Tennessee River systems (Starnes and Etnier 1986), then area reticulation will occur. A historical biogeographical method which can reconstruct such events is Phylogenetic Analysis for the Comparison of Trees, or PACT (Wojcicki and Brooks 2005). Where the taxa in an area have more than one historical origin, this taxon will be repeated in the final cladogram, once for each independent historical origin of taxa in that area. Another phenomenon which can confound historical biogeographical analyses, and which may well have occurred in the Central Highlands, is pseudocongruence (Cunningham and Collins 1994, Donoghue and Moore 2003), the existence of identical species-area relationships in different taxa due to events occurring at different times. Since PACT as originally formulated made use of the branching pattern alone of species-area cladograms, it is susceptible to being misled by pseudocongruence. To attempt to account for this, divergence times at nodes within cladograms were used to determine whether or not matching patterns of species-area relationships in different taxa actually represent the same event (Lim 2008).

In the current study, a PACT analysis is performed on twenty-one clades of Central Highlands fishes for which mitochondrial cytochrome *b* sequence data is available, to produce a species-area cladogram which shows the full pattern of diversification of these taxa. The pre-Pleistocene vicariance hypothesis is tested, along with the new hypothesis that the Central Highlands ichthyofauna shows reticulations and evidence of significant dispersal, as a result of a more complex series of historical events than is captured in the pre-Pleistocene vicariance hypothesis.

Methods

The literature was searched for phylogenies of Central Highlands fish taxa which included allopatrically-distributed taxa within the Central Highlands, so that each phylogeny would contribute some biogeographical information to the analysis, and which included mitochondrial cytochrome *b* sequence data, so that divergence percentages at the nodes in each phylogeny could be used to correlate divergences across phylogenies in the PACT analysis. Phylogenies showing divergences between different species, and those between allopatric monophyletic populations within a single species, were both considered, and both allopatric species and allopatric populations were treated equivalently in the subsequent analysis. Where sister species were not reciprocally monophyletic, they were combined to form a single terminal taxon. All clades used in the study are listed in Table 2.1, with the sources from which their phylogenies were obtained listed. The cladogram used for *Etheostoma blennioides* grafted the phylogeny of the clades found in the Ozarks as presented by Berendzen (2005) to the phylogeny of Piller and Bart (2009), which did not include representatives of each of the Ozark populations of this species. The study area was divided into areas for the PACT analysis on the basis of the distributions of allopatric taxa in the phylogenies examined. The study area was divided into a total of twenty-six areas (Figure 2.4); the boundary between each area is based on the boundary of at least one pair of allopatric taxa in one of the phylogenies examined. The boundaries were drawn so that taxa which were completely allopatric would not be shown as occurring in the same area. Some clades included taxa found outside the study area, to the east (Atlantic drainages) south (Gulf Coast drainages, and the Mississippi River Embayment) north (Great Lakes drainages) or west (the Great Plains). These taxa were retained for the PACT analysis, but no correspondences were found between any of them, and, for simplicity, they are excluded from all figures shown here. Ranges of all taxa were obtained from the publications listed in Table 2.1, from the Atlas of North American Freshwater Fishes (Lee et al. 1980) and various state ichthyofaunas (Boschung and Mayden 2004, Burr and Warren 1986, Etnier and Starnes 1993, Jenkins and Burkhead 1994, Robison and Buchanan 1988) with additional information for *Nothonotus* from Etnier and Williams (1989).

For each phylogeny, the cytochrome *b* sequences used to estimate the published phylogeny were downloaded from Genbank (www.ncbi.nlm.nih.gov/genbank) or, where available, the alignments used in the published studies were downloaded from Treebase (www.treebase.org) or another website as specified in the paper. As these sequences were all from a protein-coding gene with the same length throughout all taxa examined, alignment was performed trivially in the program MacClade 4.06 (Maddison and Maddison 2003) ; in the rare cases where the sequences did not align trivially (due to missing data at the beginning of the sequence) the program ClustalX version 2.0 (Larkin et al. 2007) was used to create a new alignment. The aligned data set for each taxon was then analyzed in MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) using the same partitioning strategies, models of sequence evolution, and number of mcmc and burnin generations as in the original published analyses, using the implementation of MrBayes on the online CIPRES Portal (now the CIPRES Science Gateway). The original analyses in these papers all used model testing to determine the most appropriate model of molecular evolution to use. In the case of those taxa for which a Bayesian analysis was not performed in the original paper, Modeltest 3.7 was used to find an appropriate model of sequence evolution for an unpartitioned analysis of each data set, using the Akaike Information Criterion (AIC) to select the best model. Models of evolution, mcmc generations, and burnin generations for each analysis are given in Table 2.2.

Rather than using the divergence times for each node as estimated in the MrBayes analysis, the topology from the output cladograms from MrBayes was retained, but branch lengths were re-estimated under a maximum likelihood model. This was done in order to avoid the known problem of MrBayes whereby divergence times can be drastically inflated, while the topology of the tree and the lengths of branches relative to each other remain unchanged (Brown et al. 2010) and to do so more quickly than would be possible by repeatedly running problematic MrBayes analyses with newly-adjusted branch length priors until each cladogram was the correct length. For each Bayesian analysis, the topology with the highest posterior probability was converted to a maximum likelihood cladogram in PAUP* version 4.0b (Swofford 2002). Maximum likelihood scores were calculated using the same model of evolution used to reconstruct the cladogram in MrBayes, if that model was not partitioned, or by using the unpartitioned model specified by Modeltest 3.7 under the AIC, if the original model was partitioned. Each cladogram was then tested for adherence to a molecular clock using a likelihood ratio test (Huelsenbeck and Crandall 1997). For each cladogram, likelihood values were obtained in PAUP* first without enforcing a molecular clock, and then with the clock enforced, and a likelihood ratio test performed. If there was no significant difference in likelihood scores for clock-enforced and unenforced trees, that cladogram was set aside for the next stage of the analysis. If there was a significant difference in scores, the tree was made ultrametric with the non-parametric rate smoothing method (Sanderson 1997) as implemented in the program TreeEdit version 1.0a10 (Rambaut and Charleston 2001), using the default “weight rate difference across root” option. This method was used for the sake of expediency; while more rigorous rate-smoothing methods exist, given that precise sequence divergence percentages were not required for the analysis, their use was not considered necessary in this case.

Percentages of sequence divergence between taxa, which were determined by obtaining the node depths for each divergence from the completed maximum likelihood cladograms, rather than absolute estimates of clade age, were used to compare divergences across different clades chronologically because absolute age estimates are not available for many phylogenies. Several of the percid phylogenies used were calibrated based on fossil evidence; however, all such studies used centrarchid fossils (Near and Benard 2004). It has recently been shown that centrarchids are only distantly related to percids (Smith and Craig 2007), and it is questionable whether divergence dates based on distantly related external calibration points are reliable. Rather than attempting to calibrate divergence dates from those studies that did provide them with divergence percentages from those that did not, sequence divergence percentages were used for all phylogenies in order to treat each contributing phylogeny consistently in the PACT analysis. This does assume, however, that there is a molecular clock among all taxa used in the analysis, or at least that differences in rate among taxa are trivial compared to the differences in divergence times.

For phylogenies which were originally generated with multiple loci or with a mix of molecular and morphological data, the topology used in the PACT analysis was that from the published cladogram generated with the most inclusive data set which still included all relevant taxa, rather than the topology as reconstructed in the maximum likelihood analyses described above, which used only cytochrome *b* data. Some of the papers from which phylogenies were obtained used species tree methods to reconstruct phylogenies; where these were available, the topology recovered using a species tree method was preferred to that recovered with a maximum likelihood, maximum parsimony, or Bayesian analysis of a concatenated data set. In all cases where the topology as published did not perfectly match the topology recovered in the maximum likelihood analyses performed above, only those divergence percentages for bipartitions which matched across both topologies were used in the PACT analysis. That is, if a bipartition in the redone maximum likelihood analysis included all the same taxa on each side of the partition as a bipartition in the original published topology, the percent divergence at the base of the clade containing this bipartition was used in the PACT analysis. Otherwise, nodes in the inclusive topology for which no divergence percentage was available in the redone maximum likelihood analysis were allowed to float freely between the first younger and older nodes for which divergences were available. In figures 2.5, 2.6 and 2.10, solid lines joining at a node indicate that the divergence percentage of that node is known, while dashed lines emerging up and down from a node indicate that the divergence percentage of that node is unknown.

All phylogenies were assembled by hand using the PACT algorithm (Wojcicki and Brooks 2005). As each cladogram was added to those already included in the analysis, the divergences in each cladogram were compared to those of each cladogram already included the analysis individually; that is, all possible pairwise comparisons between pairs of cladograms in the analysis were made. This served to reduce the possibility that the order of adding cladograms to the analysis would influence the final species-area tree. Although a computerized implementation of the algorithm is not yet available, PACT has several features which nonetheless make its use in a study of this kind desirable. The PACT algorithm allows one to construct a species-area cladogram which incorporates all information available in each input cladogram; that is, it does not discard any known speciation events prior to conducting the analysis. It also makes no assumptions about the relative weights of dispersal and vicariant events. Unlike methods such as Primary Brooks Parsimony Analysis (Brooks 1985) or Component Compatibility Analysis (Zandee and Roos 1987), PACT does not impose a single branching pattern on the areas included in the study. That is, if the taxa found in a certain area have more than one historical origin, that area will be duplicated in the species-area cladogram resulting from the analysis, once for each independent origin of a taxon in that area. The analysis as performed in this study was further restricted by using the divergence times of nodes (Lim 2008) to determine whether divergences in different taxa involving the same areas would be combined in the final species-area cladogram. If two splits which would otherwise be combined in a PACT analysis (under the hypothesis that they represent the same biogeographical event) had a more than two-fold difference in the percent sequence divergences at the nodes in question, these were not combined in the analysis.

After the PACT was complete, divergence events were counted to determine the number of divergences due to common events, and due to unique events. A common event was defined as any divergence in one clade which matched temporally and spatially to another divergence in a different clade; that is, the there was at least partial overlap in the ranges of the descendant taxa in both clades on either side of the divergence, and the percent sequence divergence of one event was no more than twice as much as that of the other. Where more than two taxa showed such an overlap, their divergences were combined into the same common divergence event only if the percent sequence divergences of the taxon showing the highest percent sequence divergence was no more than twice that of the taxon showing the lowest percent sequence divergence. Each divergence in a cladogram which did not find a match in any other cladogram was counted as a unique event.

Results

Figure 2.5 shows the biogeographical correspondence between ten distinct clades: the *Erimystax x-punctatus* group, *Hybopsis amblops*, the *Hypentelium nigricans* group, the *Nothonotus vulneratus* group, the *Nothonotus tippecanoe* group, *Etheostoma blennioides*, *Etheostoma caeruleum*, the *Percina antesella* group, the *Percina squamata* group, and *Percina evides*. Together, the ranges of these clades span the entire study area. There are ten distinct groups of divergence events among these clades which show a biogeographical and temporal correlation; these will be referred to as Common Divergence Events. Figure 2.6 is a simplification of 2.5, showing only those branches of each clade involved in the Common Divergence Events, and removing those branches that represent unique divergence events.

The earliest divergence found in these clades is Common Divergence Event 3, which separates taxa in the *Nothonotus vulneratus*, *N. tippecanoe*, *Erimystax x-punctatus*, and *Etheostoma blennioides* groups found in the Tennessee River drainage (areas V, W, and X) from all other areas to the north and west (Figure 2.7). The second-oldest divergence, Common Divergence Event 4, involves two different clades: the *Percina antesella* and *Hypentelium nigricans* groups. In both these clades, taxa found in the Mobile River drainage (area Z) are split from taxa found in all areas to the north and west (Figure 2.7). *Hypentelium nigricans* and *Percina evides* are involved in Common Divergence Event 5, which appears to be a trans-Mississippian split, dividing taxa found in the Mississippi River basin west of the Mississippi (areas A to H) from those found to its east (areas L to Y) (Figure 2.7).

Common Divergence Event 6 involves the *Hybopsis amblops* and *Erimystax x-punctatus* groups. In both clades, taxa found in the Arkansas River (area E) are split from taxa found to the north and east (Figure 2.8). Common Divergence Event 7 includes the *Erimystax x-punctatus*, *Hybopsis amblops*, *Etheostoma blennioides*, *Nothonotus vulneratus*, and *Percina squamata* clades. This event divides taxa found in the Wabash, Kentucky, and Upper Ohio Rivers (areas L to O) from taxa found both to the south and the west of these areas (Figure 2.8). Two of the taxa found in areas L to O also occur in area P, the Green River. Three of these taxa, *E. x-punctatus*, *E. blennioides*, and *N. vulneratus*, also took part in Common Divergence Event 3; the fourth participant in Common Divergence Event 3, the *N. tippecanoe* clade, is not split by Common Divergence Event 7. Its representative here, *N. tippecanoe* itself, either was not affected by this split, or only dispersed from one side of the dispersal barrier to the other after the split occurred. Of the five clades involved in Common Divergence Event 7, only one, *E. blennioides*, shows a split between Upper Ohio River taxa and taxa found both to the west (in the Ozarks) and to the south (in the Cumberland River); one clade, *E. x-punctatus*, has taxa only to the west, but not to the south, of the split, and the remaining three have taxa only to the south, but not to the west. It is possible that this split actually represents two distinct divergence events that are very close together chronologically.

Common Divergence Event 8 is a split between taxa found, on the one hand, in the Upper Mississippi River and the Missouri and Meramec drainages in the northern half of the Ozarks (areas A to D) and on the other hand in the Arkansas, White, and Black River drainages in the southern half of the Ozarks (areas E to H) (Figure 2.9). Three taxa are involved in this split: *Hypentelium nigricans*, *Percina evides*, and *Etheostoma blennioides*. The most recent four divergence events involve two taxa: *Etheostoma caeruleum* and *E.blennioides*. Common Divergence Event 9 is a split between areas in the northern Ozarks (areas B to G) and areas to the east of the Mississippi (areas M to X) (Figure 2.9). Common Divergence Event 10 is a split between the Upper Mississippi River and Missouri River drainages on one side (areas A to C) and the Meramec and Black Rivers on the other (areas D and G) (Figure 2.9). These two splits are actually a trichotomy in *E. caeruleum*; resolved in the way given here, these divergence events match the order found in *E. blennioides*. Common Divergence Event 11 divides taxa found in the White and Black Rivers (areas F and G) from those found in the Arkansas and Little Red Rivers (areas E and H) to their south and west (Figure 2.9). The most recent divergence, Common Divergence Event 12, divides taxa in the Osage River (area B) from those in the Gasconade River (area C) (Figure 2.9).

The ten clades which are involved in these Common Divergence Events include a total of 47 divergence events. Of these, 26 are included in the Common Divergence Events; the remaining 21 are geographically unique divergences (either dispersals, or vicariances for which there is no evidence in other taxa.) There are also several cases of failure to diverge: the case of *Nothonotus tippecanoe* already mentioned; *Hypentelium nigricans*, which did not take part in Common Divergence Event 11, despite apparently already being found in the relevant areas; and *Erimystax x-punctatus*, *Hypentelium nigricans* and *Percina evides*, which did not take part in Common Divergence Event 10 even though all three were apparently found in the relevant areas when that event took place.

The biogeographical correspondence between the *Luxilus zonatus*, *Erimystax dissimilis*, *Noturus albater*, *Etheostoma variatum*, and *Etheostoma sagitta* groups is shown in Figure 2.10. All five clades are restricted to the Ozark and Ouachita Mountains west of the Mississippi River (areas E to I) and, east of the Mississippi, to the Ohio River basin (areas L to O) and the upper Cumberland River (area U) with the exception of *Erimystax dissimilis*, which is also found in the remainder of the Cumberland River, the Green River, and the Tennessee River (areas P to X).

The *Etheostoma variatum* and *E. sagitta* clades are each other's sister group; each clade also has representatives west and east of the Mississippi River. The divergence between the two groups may thus represent a duplication event. Both groups, along with the *Erimystax dissimilis* group, take part in Common Divergence Event 1, an apparent vicariance between taxa in drainages east of the Mississippi on the one hand, and taxa in the Ozarks on the other (Figure 2.7) In the case of the *E. variatum and sagitta* groups, there is a large geographical area between the taxa found to either side of the split; presumably both clades had representatives in the Wabash and Kaskaskia Rivers (areas K and L) which are now extirpated or extinct. Common Divergence Event 2 is a split within the Ozark Mountains, between the Missouri, Meramec, and Black Rivers to the North (areas B, C, D, and G) and the White and Little Red Rivers to the south (areas F and H) (Figure 2.7). Three taxa participate in this split: the *Luxilus zonatus* clade, the *Noturus albater* clade, and the *Etheostoma variatum* clade. The *Etheostoma sagitta* clade has a representative to the north of this split, but not to the south, while the *Erimystax dissimilis* clade has a representative to the south, but not the north; both may be cases of either extinction or a failure to disperse. The *Etheostoma variatum* clade includes an unresolved trichotomy which spans Common Divergences 1 and 2; there is not sufficient evidence from the species-area cladogram to determine which divergence event is actually the earlier of the two.

Of the 14 divergence events found in these five clades, six are accounted for by Common Divergence Events, one is a duplication, and seven represent geographically unique divergences.

Of the phylogenies examined, several did not contain any divergences which appeared be common with divergences in any other clades. These include *Chrosomus,* the *Notropis rubellus* group, and the *Erimystax insignis* group, all cyprinids; all clades except for the *Noturus albater* group in *Noturus (Rabida)*, ictalurids; and the percids *Etheostoma (Catonotus)*, the *E. spectabile* group, and all *Nothonotus* clades either than the *N. vulneratus* and *N. tippecanoe* clades. The divergences in these clades represent another 54 unique divergence events; thus, among all the clades examined in this study, there is a total of 115 divergence events, 32 (28 percent) of which are Common Divergence Events.

Discussion

The present analysis shows many similarities to that of Mayden (1988a) (Figure 2.11), despite the differing methods used. Both analyses recover a split between the Mobile River (area Z) and the Mississippi River drainage (all other areas). Both show splits between the Ohio, Cumberland, and Tennessee River basins (areas M to Y) and the Interior Highlands, upper Mississippi, and Illinois Rivers (areas A to L). Both also show splits in the Ozarks between the rivers draining northward into the Missouri River (areas A to D), and those draining southward into the Arkansas River (areas E to H). However, there are also notable differences. Mayden's analysis grouped the Green River with the Cumberland and Tennessee, in a sister group to the remainder of the Ohio basin and all areas west of the Mississippi. The current analysis, however, most frequently groups the Green River (area P) with other tributaries of the Ohio River. Mayden's analysis found only equivocal evidence for a split between the Tennessee River and the Cumberland River; such a split is found in the current study (Common Divergence Event 4). Mayden's study placed the Ouachita Highlands (areas I and J) in a sister relationship with the Arkansas River drainage (area E); in the present study those areas and the White and Black Rivers (areas F and G) form a polytomy. Some of these differences can no doubt be attributed to the different selection of taxa between the two studies: Mayden's study used seven clades with phylogenies based on morphological characteristics; only three of those clades are among the fifteen used here.

A major difference between the results of this analysis and that of Mayden (1988a) is that the difference in methods used allowed for the recovery of reticulate histories of many of the areas involved. The Duck and Tennessee Rivers (areas V, X and Y) are a prime example. They are involved in Common Divergence Event 3, which splits them from river basins to the North, including the Cumberland and Ohio Rivers, in Common Divergence Event 5, which splits off areas to the west of the Mississippi from the Ohio, Cumberland and Tennessee Rivers grouped together; in Common Divergence Event 7, which splits the Ohio River from the Cumberland and Tennessee Rivers, and in Common Divergence Event 9, which splits areas in the Ozark Mountains from the Ohio, Cumberland, and Tennessee Rivers. There is clearly more than a single vicariance event involved; furthermore, there must have been dispersal between vicariance events, to explain why clades affected by more recent vicariances were not affected by earlier ones. The two taxa affected by Common Divergence Event 5, *Hypentelium nigricans* and *Percina evides*, have undifferentiated populations spanning the Ohio, Cumberland, and Tennessee River basins. Since Common Divergence Event 3 split the area occupied by these populations into two, their lack of differentiation needs an explanation. It is quite possible that they were not throughout most of this area at the time of Common Divergence Event 3, and dispersed throughout this range at a later date, or, they were present at that time, but Common Divergence Event 3 affected only certain taxa, perhaps because of differences in environmental tolerances which allowed certain species to continue moving from one side of the barrier to another, while others were no longer able to. Common Divergence Event 3 may represent the hypothesized prior configuration of the Duck and Tennessee drainages, which drained southward into the Gulf of Mexico, while the Cumberland River drained west and north into the Mississippi, before the Tennessee River changed course and began draining northward to a confluence with the Cumberland (Starnes and Etnier 1986); this would place Common Divergence Event 3 during the Late Neogene Period. If percent sequence divergence is converted to absolute time using a rate of 2 % per million years, as usd for darters by Near and Benard (2004), Common Divergence Event 3 is placed between eight and six million years ago, during the late Miocene. Two of the taxa which take part in Common Divergence Event 7 dividing the Ohio River basin from those to the south, have representatives to the south of this split which straddle the split in Common Divergence Event 3. These may both be taxa that did not enter the Tennessee and Duck River basins until after Common Divergence Event 3 occurred. The population of *Hybopsis amblops* found in areas Q, R, S, T, V, and W, may have spread from the Cumberland River (areas Q to T) to the Duck (V) and Tennessee (W) after these rivers were connected at their mouths in the Pleistocene, while *Percina squamata*, found in areas S, T, X, and Y, in the upper portions only of the Cumberland and Tennessee basins, may have spread into the Tennessee through a stream capture in the Cumberland Gap region, or its current range may be a remnant of a larger range that previously covered the lower Tennessee and Cumberland basins (Starnes and Etnier 1986). In contrast to these two taxa, the clade of *Etheostoma blennioides* found in the Cumberland River, and *Nothonotus sanguifluus* + *N. microlepidus*, which also took part in Common Divergence Event 7, are restricted to the Cumberland, as they are from lineages old enough to have been present during Common Divergence Event 3.

Reticulation is also seen in the placement of the Black River (area G). Common Divergence Event 2 groups it with the Missouri River drainages to the North (areas B to D), separating it from the upper White River and Arkansas drainages to the south (areas E, F, and H) while the more recent Common Divergence Event 8 groups it with these southern drainages, separating it from those to the North. The shift of the Black River from one grouping to another may be connected with the shift of the Mississippi River from the west to the east side of Crowley's Ridge (Robison 1986), an elevated area parallel to the Mississippi River in northeastern Arkansas and southeastern Missouri. Prior to the Pleistocene, the Mississippi flowed to the west of this ridge, and each of the tributaries of the Black River flowed independently into Mississippi. Habitat for highland fishes in the Mississippi was more likely to have been suitable upstream, towards the Missouri, than downstream towards the mouth of the White and Arkansas Rivers, which would favor a connection between the Black River and Missouri River drainages. After the Mississippi shifted to the east of Crowley's Ridge, the proximity between the Black and Missouri Rivers was lost, and the reduced flow in the former Mississippi River basin (now the lower part of the Black River) may have allowed fishes from the White River to more easily spread to the Black, and vice versa.

The present analysis and that of Mayden (1988a) differ prominently in their support for the existence of the Teays River system. The Teays River is believed to have connected the upper Ohio River drainage directly to the upper Mississippi River, while the Green and Cumberland Rivers drained separately into the lower Mississippi directly via the Old Ohio River (Figure 2.3). In Mayden's analysis, the Green River forms a clade with the Cumberland and Tennessee Rivers, while the upper Ohio River forms a clade with the upper Mississippi, Ozark, and Ouachita drainages, supporting the existence of the Teays. In the present analysis, however, the Green River (area P) groups most often with the upper Ohio drainages (areas M, N and O). This may be due to the difference in taxon sampling between the two studies. Notably, only one of the taxa in this study (*Etheostoma blennioides*) found in the Green River has a population endemic to that system. The other taxa found in the Green are spread throughout other river systems as well, and so their presence in the Green as well as the upper Ohio may be due to dispersal after the connection of the upper and lower Ohio systems. One taxon, *Hypentelium nigricans*, does show evidence for the existence of the Teays system, in that a population in the New River (area N) is sister to populations west of the Mississippi (Berendzen et al. 2003). No other taxa, however, repeat this pattern in the current study.

In addition to the clades which do show biogeographical overlap, a number of clades with cytochrome *b* sequence data were examined, but found to have no biogeographical overlap with other taxa. Most of these taxa have divergences which, for the most part, are older than those within the clades which do show overlap. Given that the older a common divergence event, the more likely evidence for it is to have been obscured by subsequent vicariance and dispersal events, it should not come as a surprise that clades with old divergences show little biogeographical congruence with other clades.

Among the clades which were involved in Common Divergence Events, 32 of their divergences took place at a Common Divergence Event, while 29 were unique events; this means that 47% of all divergences were common among taxa. However, if one includes the 54 divergences among clades for which no common divergences were found, this number drops to 25%. If it is true that many of the clades with no common divergences represent old taxa that may be the only surviving representatives of the vicariance events which resulted in their diversification, the percentage of divergences due to common events would increase; even so, it is clear that unique dispersal events are responsible for a significant proportion of the diversity in the Central Highlands.

The timing of the divergences seen depends, of course, on the scaling factor used to convert amounts of sequence divergence to time. Using a figure of two percent sequence divergence between lineages per million years, as obtained for darter lineages by Near and Benard (2004), the beginning of the Pleistocene 2.5 million years ago becomes a sequence divergence of 0.025. This places seven of the twelve Common Divergence Events prior to the beginning of the Pleistocene. Using a smaller figure for percent sequence divergence per million years, as has been estimated for some taxa (Berendzen, Gamble et al. 2008) would increase the number of divergence events prior to the Pleistocene; a figure of one percent, for instance, places all except Common Divergence Event 12 in the Pliocene or earlier. Thus, while the high proportion of unique divergence events indicates that the pre-Pleistocene vicariance hypothesis is insufficient to account for the diversity of the Central Highlands ichthyofaunas, the timing of the common divergence events does support the pre-Pleistocene vicariance hypothesis as an explanation for many of the common events seen.

The results of this analysis depend on the degree to which the assumptions are justified. A key assumption in this regard is that divergences across clades would be considered to represent the same event if there was anywhere up to a 100 percent difference in the percent sequence divergence at the respective nodes. Obviously, it would be too restrictive to require perfect agreement in percent divergence before combining taxa in the species-area cladogram, as there are numerous sources of error involved in determining the percent divergence at a node. Beyond the error in determining the true percent divergence at nodes *within* taxa, there is also the fact that the rate of evolution, and thus the amount of time represented by a particular percent divergence, must certainly vary to at least some extent *across* taxa. (This is necessarily so in this particular data set, given that some of the taxa used failed a molecular clock test for uniformity of rate *within* each taxon.) Finally, there is the fact that, even if true percent divergences of cytochrome *b* at each node within each phylogeny could be known perfectly, and the chronological calibrations of these divergences could be known as well, allowing for the perfect determination of true divergence times for cytochrome *b* within each taxon, there would still be variation in the divergence time due to coalescent effects. Taking all of these factors into account, the percentage difference between sequence divergences chosen may be considered reasonable. Of course, there is no empirical justification for such a particular choice. This study was limited, however, by the requirement to reconstruct all the phylogenies under consideration in a reasonable amount of time, and the lack of multilocus data for most taxa. As multi-locus phylogenetic studies become the norm, date estimates obtained through the use of coalescent-based species tree methods such as *BEAST (Heled and

Drummond 2010) should allow one to decide whether or not to combine divergence events with more justification from the data – for instance, one could decide that two divergence events may be combined if the 95% highest posterior density interval for each divergence time overlaps. To achieve greater precision, methods such as Hierarchical Approximate Bayesian Computation (Hickerson et al. 2006) could be used to test for apparent simultaneous divergence across different taxa. A separate issue is rate variation across taxa. The use of fossil calibrations during species or gene tree reconstructions would of course help to specify absolute date ranges for divergence times (and determine if there is evidence for rate variation across taxa) but it is problematic for the ichthyofauna of the Central Highlands, since of the two most speciose families in the Central Highlands, the Percidae and Cyprinidae, the Percidae are very poorly represented in the fossil record, as are the Central Highlands clades of the Cyprinidae (Cavender 1986). Another potential advance would be to use methods of ancestral area reconstruction such as Lagrange (Ree and Smith 2008) to determine ancestral ranges at nodes prior to combining them using PACT; however, any method that reconstructs the ancestral range of a taxon using information only from that taxon may give a result that differs from that obtained by conducting a PACT analysis first, and then reconstructing ancestral ranges based on the information it gives on concordance of divergence events.

Conclusion

The ichthyofaunal diversity of the Central Highlands of North America is clearly complex in origin. It cannot be explained as the result of the piecewise division of one widespread range throughout the region for each basal taxon due to a single series of vicariance events. The history of the Central Highlands ichthyofauna is, rather, a reticulate one. The Tennessee and Duck River basins, which surely not coincidentally are the focal point for the diversity within the Central Highlands, feature in no less than four independent divergence events, but these are not the only reticulations found in the region. The unique divergence events seen in the analysis, which account for more than half of all divergence events, contribute greatly to the reticulation seen, as each dispersal into an area represents another event responsible for a portion of the species diversity of that area. Common divergence events, however, are still responsible for a significant portion of all divergences. While many of these events are no doubt vicariances, some may be co-ordinated dispersal events – due to the loss of barriers to dispersal allowing multiple taxa to colonize new areas. The fact that different common divergence events include the same areas also indicates that much dispersal must have been taking place between divergences. Thus, while the pre-Pleistocene divergence hypothesis is supported by many of the divergence events found, the full story of fish diversification in the Central Highlands includes many dispersal events as well. Some specific predictions made by the pre-Pleistocene vicariance hypothesis are poorly supported by the present study, however, such as the former existence of the Teays River; and much of the branching pattern seen in Mayden's (1988a) analysis is suggested to be an artifact of the method used.

As sequence data accumulates, and techniques for the reconstruction of species trees and divergence times improve, our ability to determine whether apparent correlations in species-area relationships across clades are actually the result of shared divergence events is expected to increase. Improved methods to reconstruct changes in population sizes in the past will also help us to reconstruct what happened to each daughter population as populations split into two. This promises to increase our knowledge of specific events in past biogeographical systems, and also provide valuable data to other fields, such as community ecology and the study of adaptation, where knowledge of the order and timing of species' arrivals in an area can be invaluable. However, stochastic variation in rates of evolution and in the coalescent process itself will prevent much of the history of biotas from being known. In this regard, it would be invaluable to model how random variation in rates of evolution and other phenomena which affect phylogenetic and biogeographical reconstruction influence the results which are drawn. This would help to better understand how much of a study such as the current one can be considered reliable information about the past, and how much is determined entirely by the method used, and not the actual history.

Table 2.1 Clades of fishes used in analysis.

Clade	Number of taxa	Source
Cyprinidae		
Erimystax dissimilis group	3	Simons 2004
Erimystax x-punctatus group	5	Simons 2004
Hybopsis amblops	6	Berendzen, Gamble et al. 2008
Luxilus zonatus group	3	Dowling and Naylor 1997
Catostomidae		
Hypentelium nigricans group	5	Berendzen et al. 2003
Ictaluridae		
Noturus albater group	2	Egge and Simons 2009
Percidae		
Nothonotus vulneratus group	4	Keck and Near 2008
Nothonotus tippecanoe group	2	Keck and Near 2008
Etheostoma blennioides	16	Piller and Bart 2009; Berendzen 2005
Etheostoma caeruleum	11	Ray et al. 2006
Etheostoma sagitta group	3	Switzer 2004
Etheostoma variatum group	7	Switzer 2004
Percina antesella group	3	Near 2002
Percina evides	3	Near et al. 2001
Percina squamata group	2	Near 2002
Unused clades		
Chrosomus	6	Strange and Mayden 2009
Erimystax insignis group	2	Simons 2004
Notropis rubellus group	9	Berendzen, Simons et al. 2008
Noturus (Rabida) clades other than *N. albater* group	18	Egge and Simons 2009
Etheostoma (Catonotus)	7	Hollingsworth and Near 2009
Etheostoma spectabile group	11	Bossu and Near 2009
Nothonotus clades other than *N. vulneratus* and *N. tippecanoe* groups	12	Keck and Near 2008

Table 2.2 Models of evolution and run specifications for MrBayes analyses and molecular clock likelihood ratio tests. Models which were not taken from the original papers, but rather newly determined for the present analysis, are indicated with an asterisk.

Clade	Model of evolution for MrBayes	MCMC generations	Burnin	Model of evolution for molecular clock LRT	Molecular clock rejected?
Cyprinidae					
Erimystax	GTR+SSR	2×10^6	2.5×10^5	GTR+SSR	Yes
Hybopsis amblops	1st codon: SYM+I+G 2nd codon: GTR+I 3rd codon: GTR+I+G	2×10^6	1×10^5	TrN+I+G	No
Luxilus zonatus group	GTR+I+G*	2×10^6	5×10^5	GTR+I+G	No
Catostomidae					
Hypentelium nigricans group	GTR+G	2×10^6	1×10^5	TrN+G	Yes
Ictaluridae					
Noturus (Rabida)	GTR+I+G	5×10^6	1×10^6	GTR+I+G	Yes
Percidae					
Nothonotus	1st codon: SYM+I+G 2nd codon: GTR+I 3rd codon: GTR+I+G	6×10^6	1×10^6	GTR+I+G	Yes
Etheostoma blennioides	1st codon: SYM+I+G 2nd codon: GTR 3rd codon: GTR+I+G	5×10^6	5×10^5	GTR+I+G	No
Etheostoma caeruleum	1st codon: HKY+I+G 2nd codon: GTR+I 3rd codon: GTR+I+G	6×10^6	1.5×10^5	GTR+I+G	No

Etheostoma sagitta group and *E. variatum* group	1st codon: GTR+I+G 2nd codon: GTR+I+G 3rd codon: GTR+I+G	6 x 10^6	5 x 10^5	GTR+I+G	No
Percina antesella group and *P. squamata* group	GTR+I+G	5 x 10^6	1 x 10^6	GTR+I+G	No
Percina evides	HKY+G*	2 x 10^6	5 x 10^5	HKY+G	No
Unused clades					
Chrosomus	1st codon: SYM+I+G 2nd codon: GTR+G 3rd codon: GTR+I+G	5 x 10^6	2 x 10^5	GTR+I+G	Yes
Notropis rubellus group	GTR+I+G	2 x 10^6	1 x 10^5	GTR+I+G	No
Etheostoma (Catonotus)	1st codon: SYM+I+G 2nd codon: HKY+I 3rd codon: GTR+I+G	6 x 10^6	1 x 10^6	TIM+I+G	No
Etheostoma spectabile group	1st codon: SYM+I+G 2nd codon: GTR+I+G 3rd codon: GTR+I+G	1.5 x 10^7	1 x 10^6	GTR+I+G	No

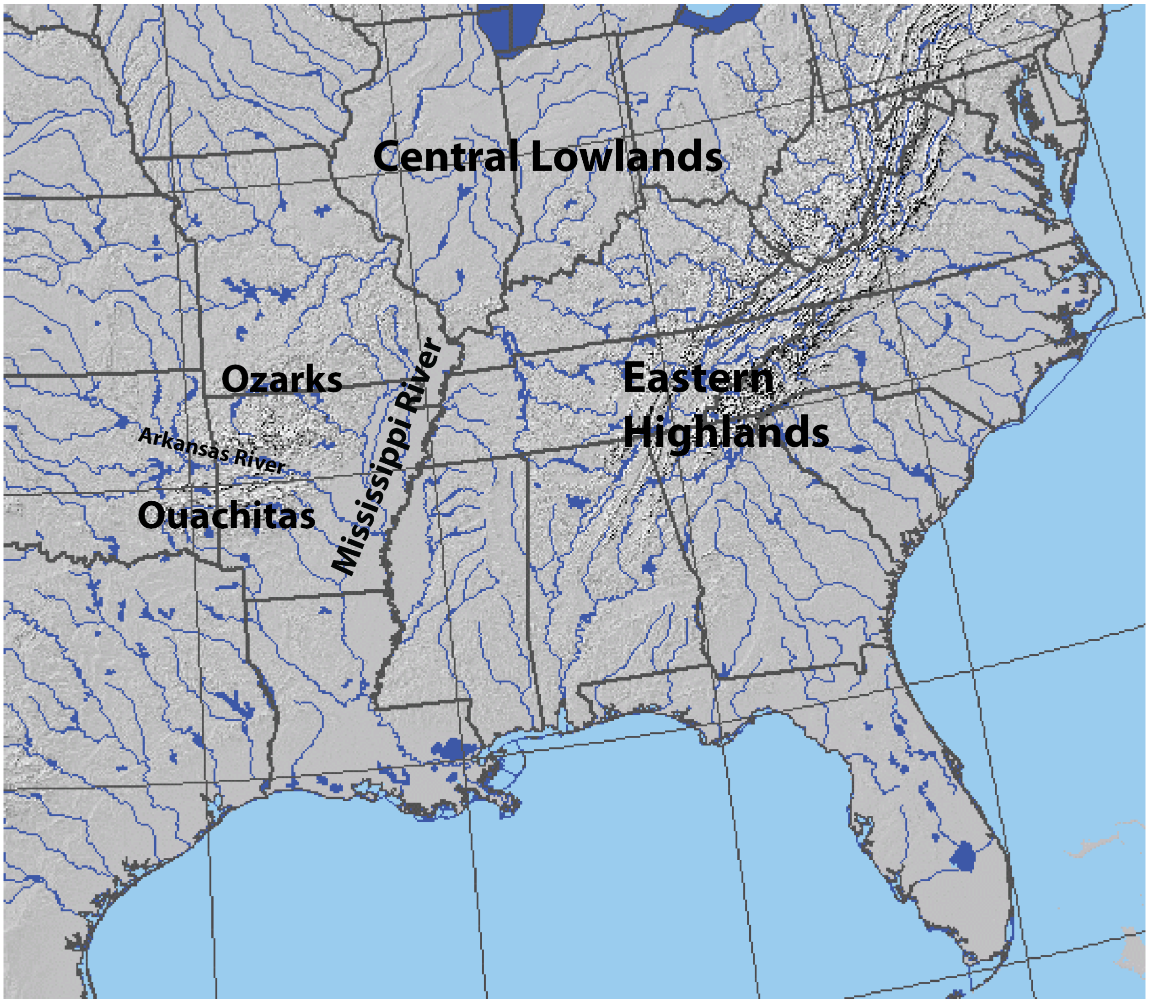


Figure 2.1 Map of southeastern North America, showing the Central Highlands, Central Lowlands, and the Arkansas and Mississippi Rivers.

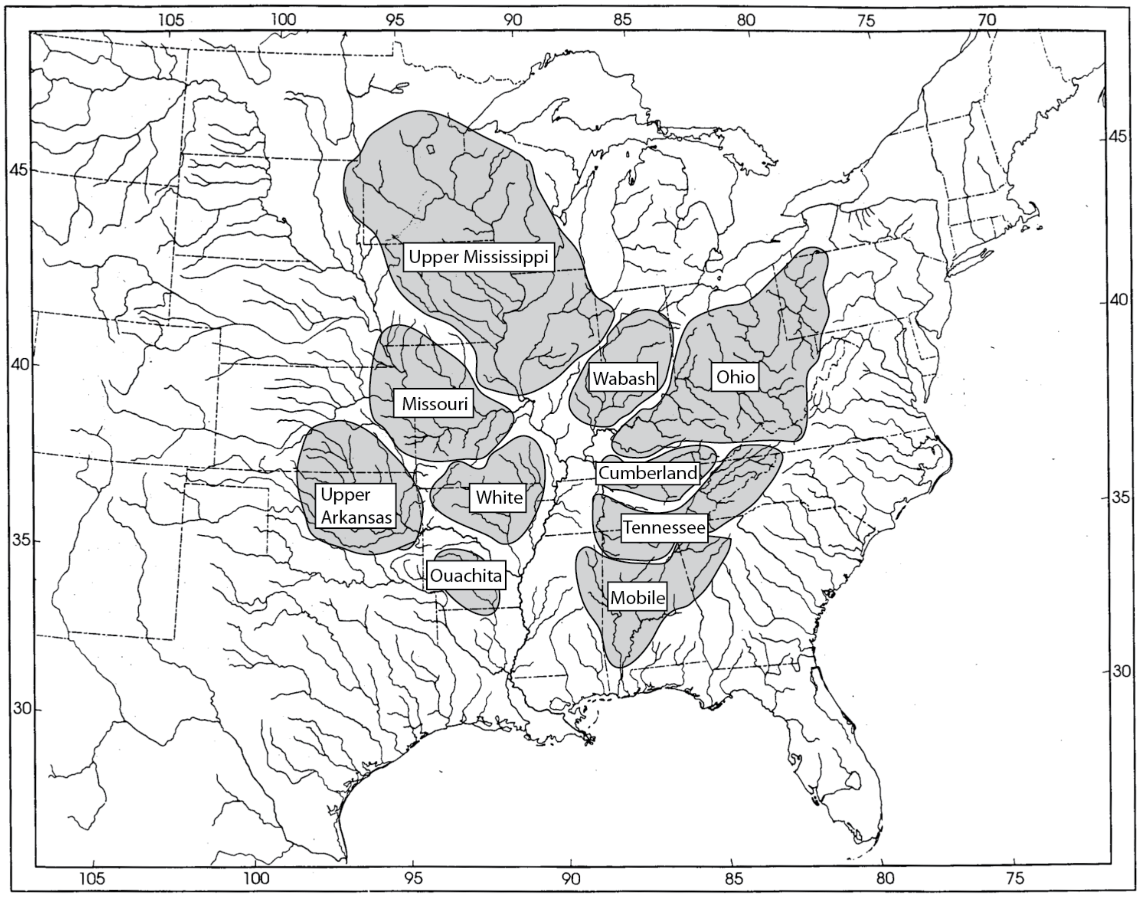


Figure 2.2 Major river systems of the Central Highlands.

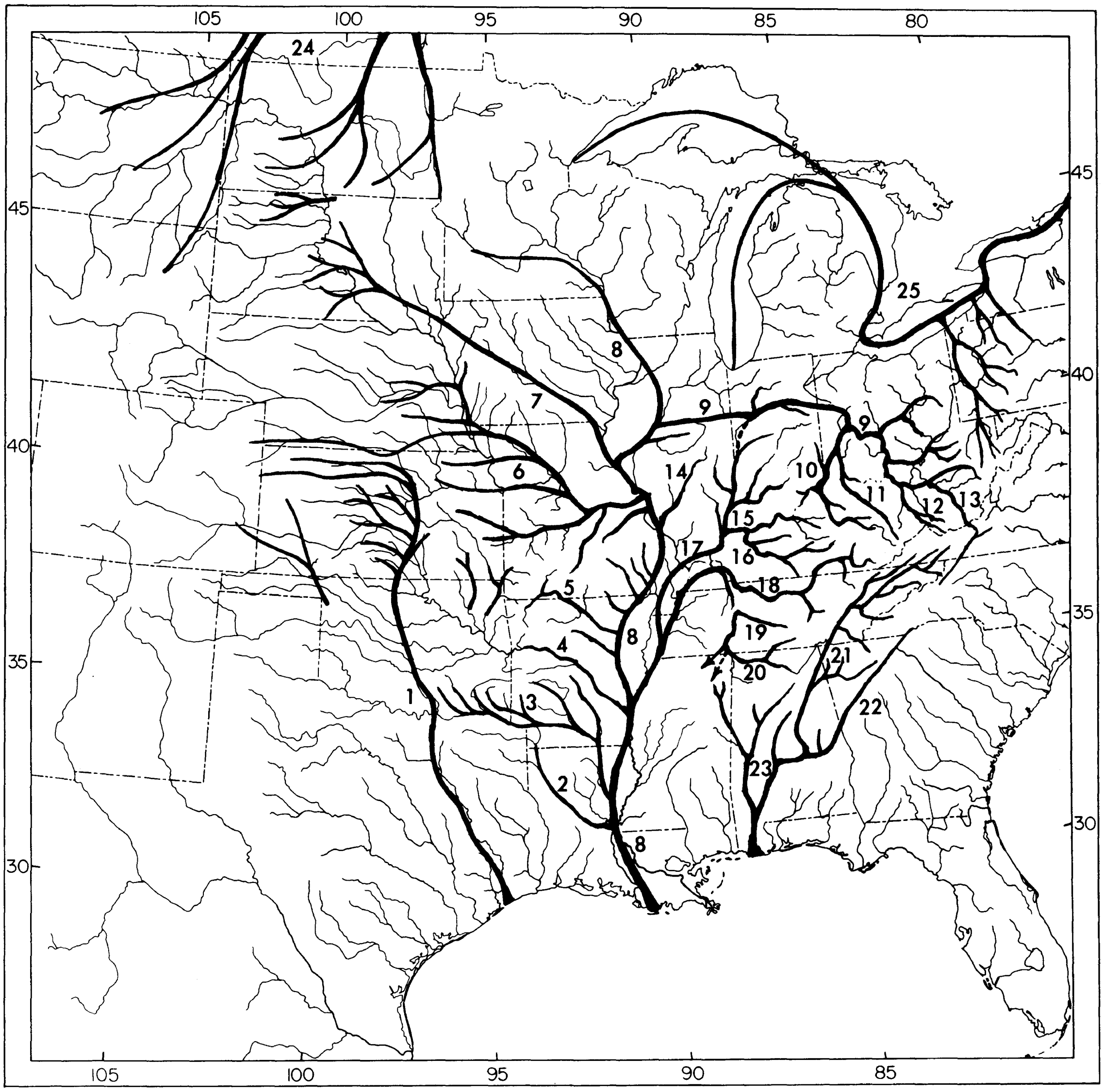


Figure 2.3 Preglacial drainages of southeastern North America superimposed on present-day drainages. From Mayden (1988a). 1 – Plains Stream 2 – Old Red River 3 – Old Ouachita River 4 – Old Arkansas River 5 – White River 6 – Old Grand-Missouri River 7 – Ancestral Iowa River 8 – Old Mississippi River 9 – Teays River 10 – Old Kentucky River 11 – Old Licking River 12 – Old Big Sandy River 13 – Kanawha River 14 – Kaskaskia River 15 – Wabash River 16 – Green River 17 – Old Ohio River 18 – Old Cumberland River 19 – Old Duck River 20 – Old Tennessee River 21 – Appalachian River 22 – Old Tallapoosa River 23 – Mobile Basin 24 – Hudson Bay Drainage 25 – St. Lawrence River.

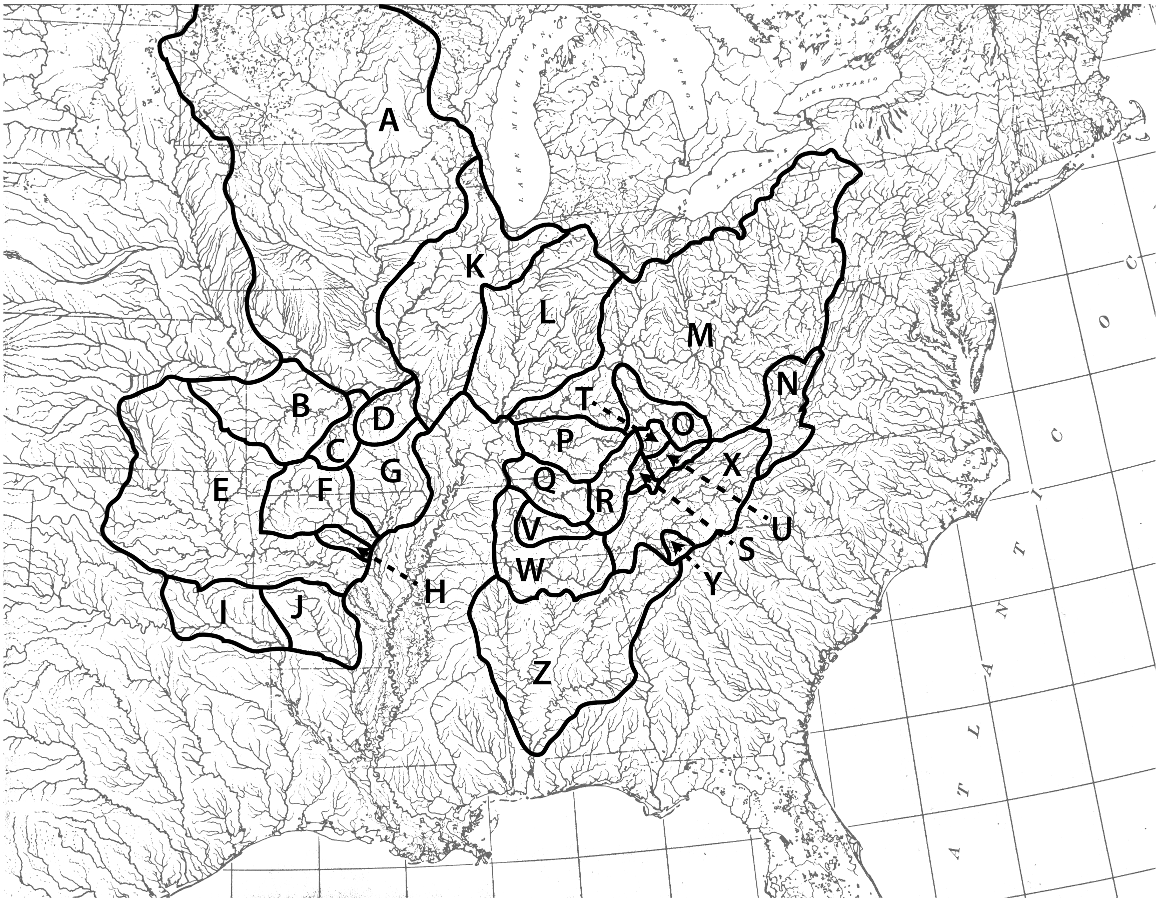


Figure 2.4 North American Central Highlands divided into areas for PACT analysis: a) upper Mississippi River b) Osage River c) Gasconade River d) Meramec River e) Arkansas River f) White River g) Black and St. Francis Rivers h) Little Red River i) Kiamichi and Little Rivers j) Ouachita River k) Illinois and Kaskaskia Rivers l) Wabash River m) upper Ohio River n) New River o) Kentucky River p) Green River q) lower Cumberland River r) middle Cumberland River s) Big South Fork t) Rockcastle River u) upper Cumberland River v) Duck River w) lower Tennessee River x) upper Tennessee River y) Hiwassee River z) Mobile River

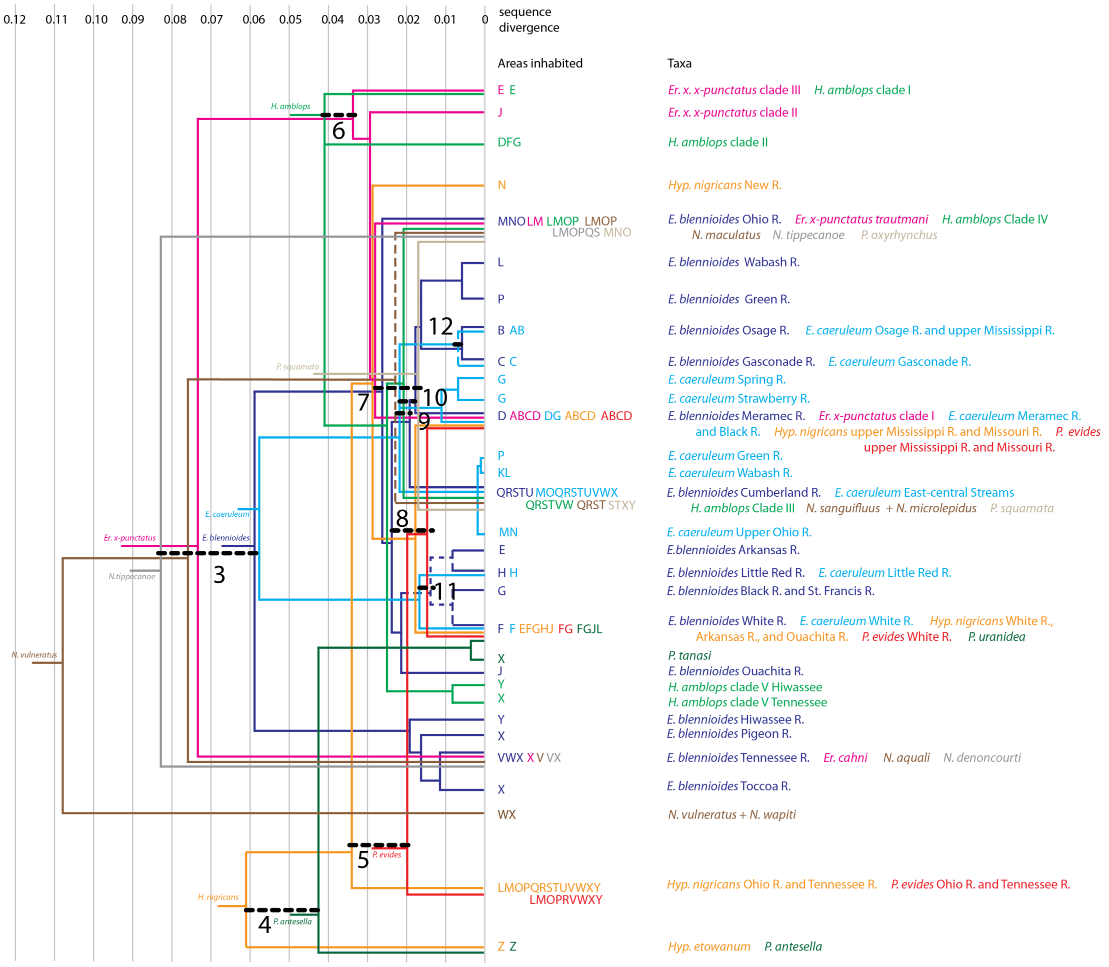

0.12
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
sequence divergence
Areas inhabited
Taxa
E E
Er. x. x-punctatus clade III
H. amblops clade I
J
Er. x. x-punctatus clade II
DFG
H. amblops clade II
N
Hyp. nigricans New R.
MNO LM LMOP LMOP LMOPQS MNO
E. blennioides Ohio R.
Er. x-punctatus trautmani
H. amblops Clade IV
N. maculatus
N. tippecanoe
P. oxyrhynchus
L
E. blennioides Wabash R.
P
E. blennioides Green R.
B AB
E. blennioides Osage R.
E. caeruleum Osage R. and upper Mississippi R.
C C
E. blennioides Gasconade R.
E. caeruleum Gasconade R.
G
E. caeruleum Spring R.
G
E. caeruleum Strawberry R.
D ABCD DG ABCD ABCD
E. blennioides Meramec R.
Er. x-punctatus clade I
E. caeruleum Meramec R. and Black R.
Hyp. nigricans upper Mississippi R. and Missouri R.
P. evides upper Mississippi R. and Missouri R.
P
E. caeruleum Green R.
KL
E. caeruleum Wabash R.
QRSTU MOQRSTUVWX QRSTVW QRST STXY
E. blennioides Cumberland R.
E. caeruleum East-central Streams
H. amblops Clade III
N. sanguifluus + N. microlepidus
P. squamata
MN
E. caeruleum Upper Ohio R.
E
E.blennioides Arkansas R.
H H
E. blennioides Little Red R.
E. caeruleum Little Red R.
G
E. blennioides Black R. and St. Francis R.
F F EFGHJ FG FGJL
E. blennioides White R.
E. caeruleum White R.
Hyp. nigricans White R., Arkansas R., and Ouachita R.
P. evides White R.
P. uranidea
X
P. tanasi
J
E. blennioides Ouachita R.
Y
H. amblops clade V Hiwassee
X
H. amblops clade V Tennessee
Y
E. blennioides Hiwassee R.
X
E. blennioides Pigeon R.
VWX X V VX
E. blennioides Tennessee R.
Er. cahni
N. aquali
N. denoncourti
X
E. blennioides Toccoa R.
WX
N. vulneratus + N. wapiti
LMOPQRSTUVWXY LMOPRVWXY
Hyp. nigricans Ohio R. and Tennessee R.
P. evides Ohio R. and Tennessee R.
Z Z
Hyp. etowanum
P. antesella
H. amblops
6
12
P. squamata
7
10
9
E. caeruleum
8
Er. x-punctatus
E. blennioides
3
N.tippecanoe
11
N. vulneratus
5
P. evides
H. nigricans
4
P. antesella

Figure 2.5 PACT analysis of ten Central Highlands clades. Each clade is shown in a different color as follows: *Erimystax x-punctatus* group, pink; *Hybopsis amblops*, light green; *Hypentelium nigricans* group, orange; *Nothonotus tippecanoe* group, gray; *Nothonotus vulneratus* group, brown; *Percina antesella* group, dark green; *Percina squamata* group, tan; *Percina evides*, red; *Etheostoma blennioides*, violet; and *Etheostoma caeruleum*, light blue. Colored dashed lines in cladograms represent nodes for which a cytochrome *b* percent sequence divergence is unavailable. Letters under "Areas Inhabited" correspond to areas in Figure 2.4, and represent the range of each terminal taxon. Numbers in figure indicate Common Divergence Events; black dashed lines join divergence events in different clades hypothesized to represent a Common Divergence Event (see text for details). The name of each terminal taxon is given under "Taxa".

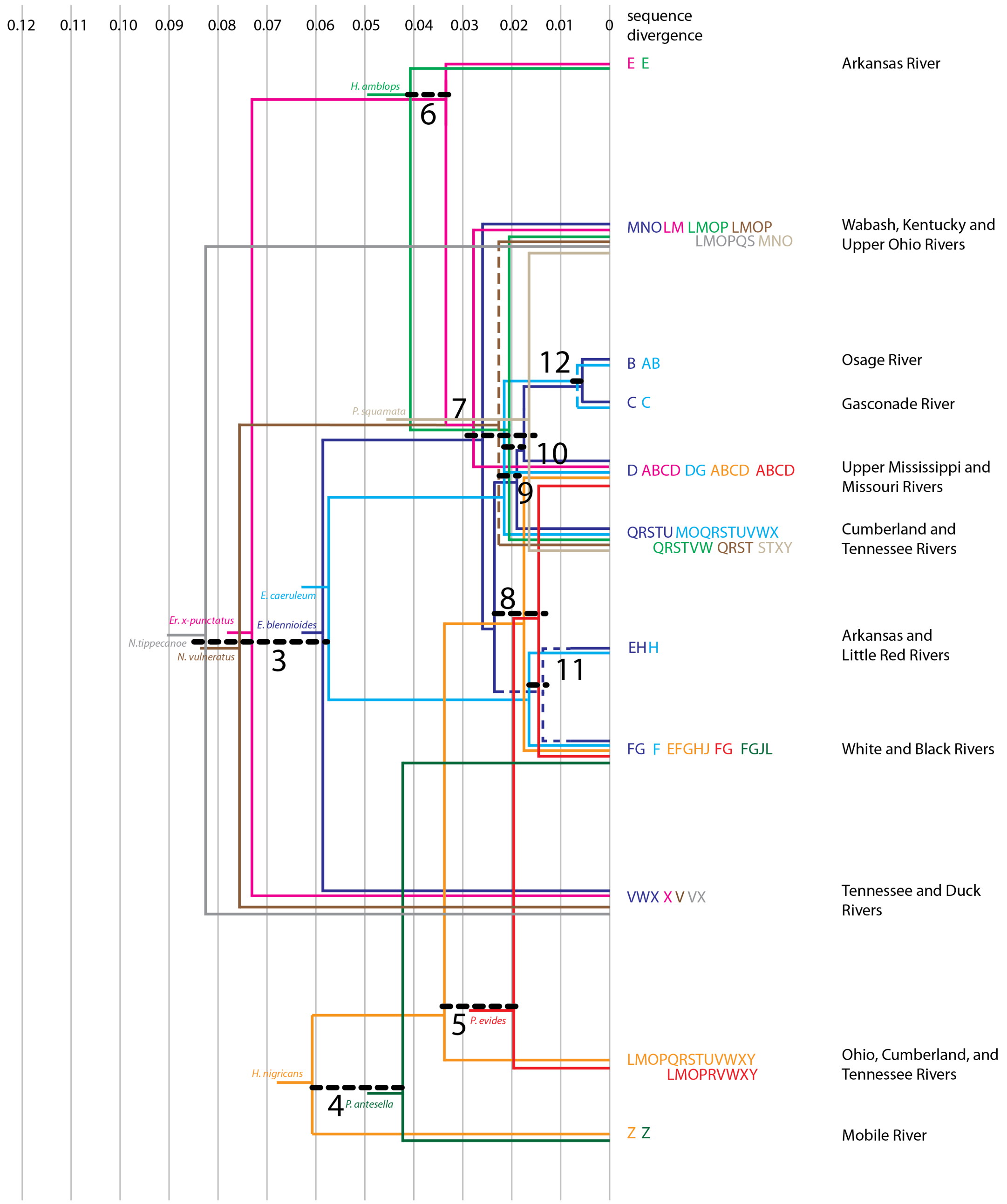


Figure 2.6 Simplified version of Figure 2.5, showing only those branches of each clade involved in Common Divergence Events. Names of rivers at right are the major drainages inhabited by each set of congruent taxa. Other details are as for Figure 2.5.

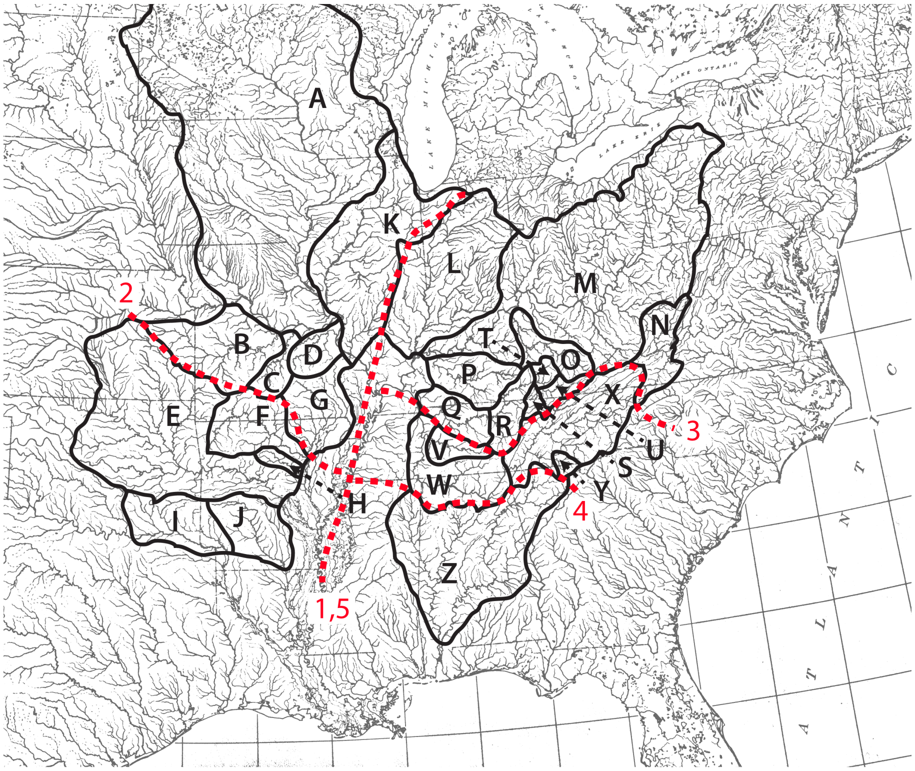


Figure 2.7 Map showing location of splits in Common Divergence Events 1, 2, 3, 4, and 5.

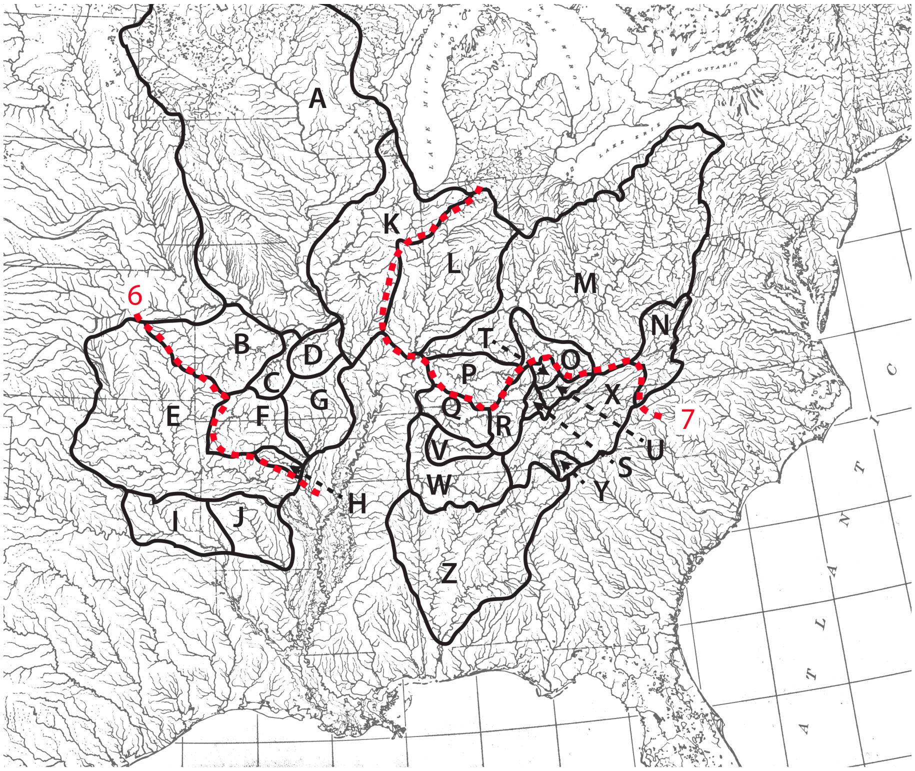


Figure 2.8 Map showing location of splits in Common Divergence Events 6 and 7.

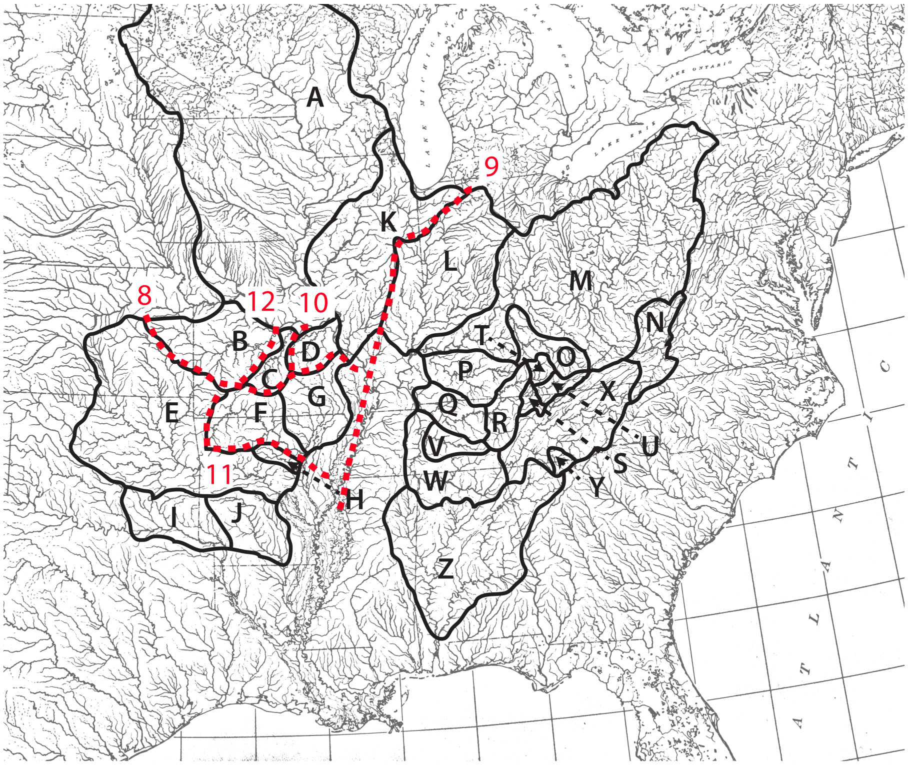


Figure 2.9 Map showing location of splits in Common Divergence Events 8 to 12.

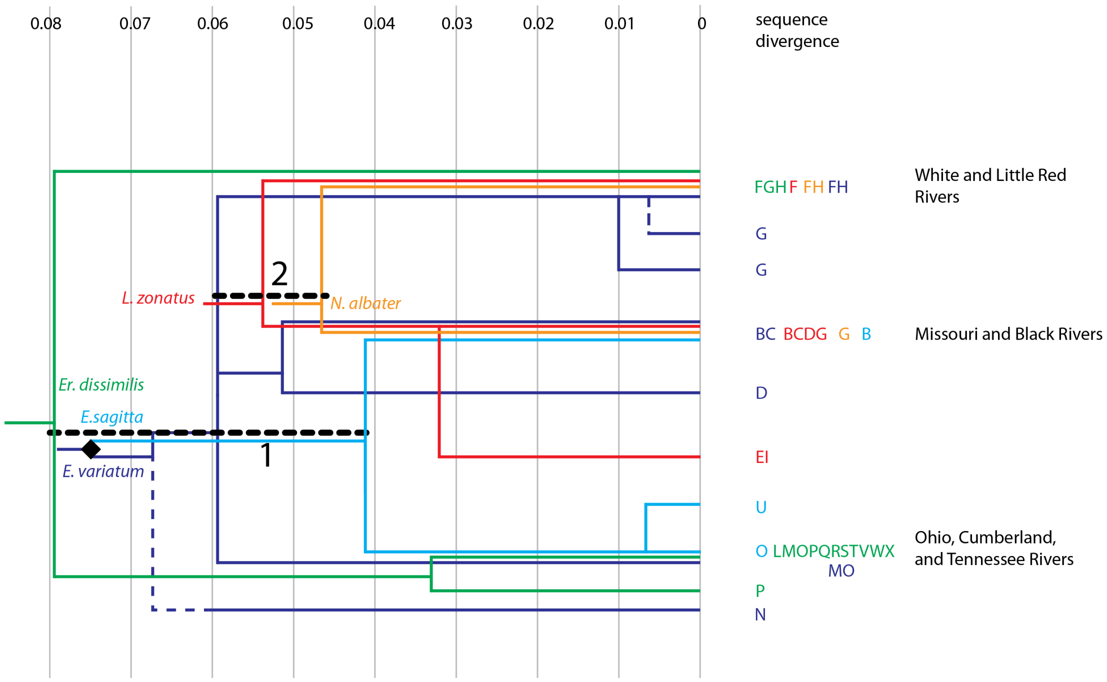


Figure 2.10 PACT analysis of five Central Highlands clades. Each clade is shown in a different color as follows: *Luxilus zonatus* group, red; *Erimystax dissimilis* group, green; *Noturus albater* group, orange; *Etheostoma sagitta* group, light blue; and *Etheostoma variatum* group, violet. All other details are as for Figure 2.5.

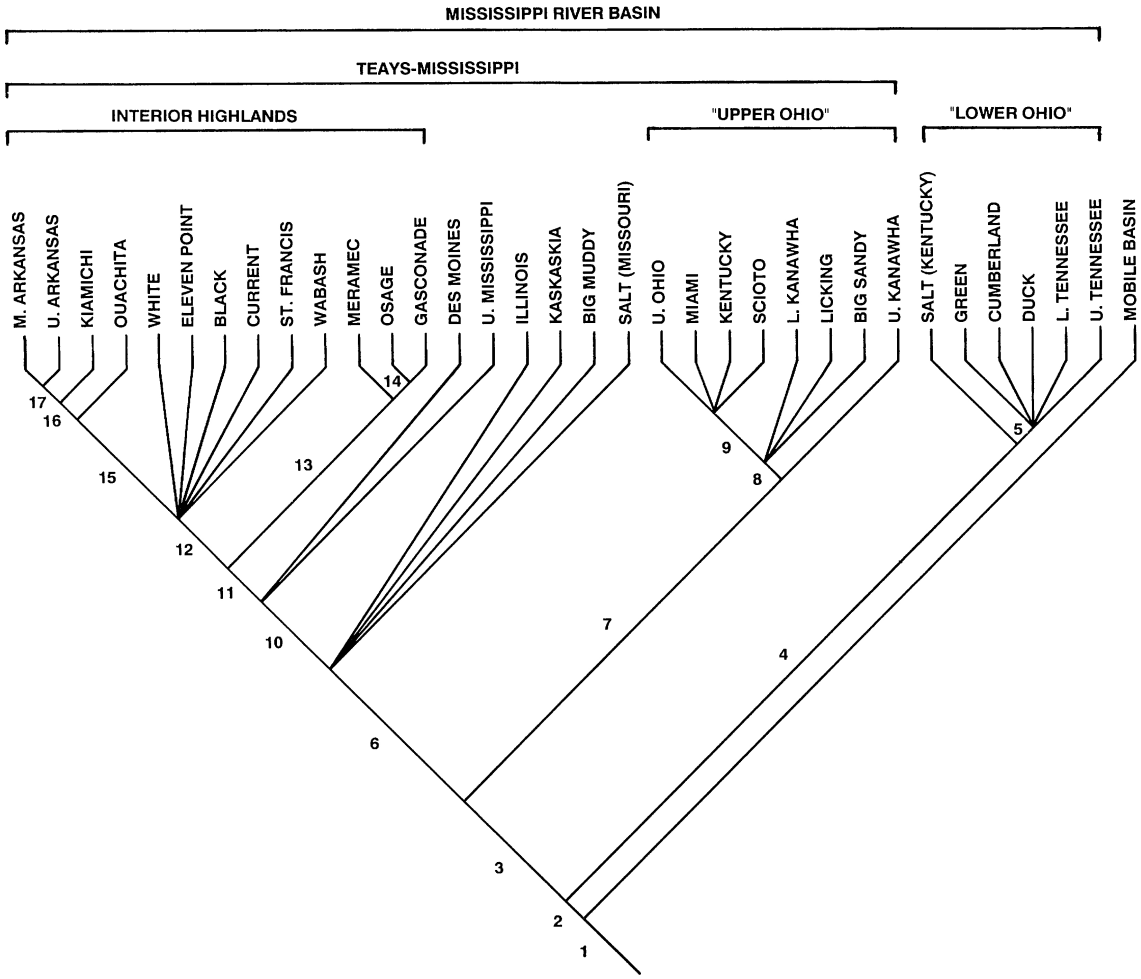


Figure 2.11 Mayden's (1988a) area cladogram of Central Highlands drainages, based upon a comparative biogeographical analysis of seven clades of fishes.

CHAPTER 3

Geographical differentiation and incipient speciation in the *Etheostoma zonale* species group

Introduction

The identity of many wide-ranging taxa which have previously been considered single, variable species is now being reconsidered, as molecular methods uncover far greater genetic diversity within such taxa than has been expected on morphological grounds alone, and as the increasing adoption of the phylogenetic species concept (Eldredge and Cracraft 1980) prompts the belief that each lineage within such variable taxa should in fact be considered a separate species. However, where the different phylogenetic lineages in such variable taxa have diverged recently, distinguishing species boundaries can become difficult. For one, not all lineages may evolve at the same rate, so that lineages that are clearly separate may be sister to lineages with very few synapomorphies to distinguish them. More significantly, because the ability to interbreed is a symplesiomorphic character (Rosen 1979), where lineages differentiate largely due to allopatry, loss of that allopatry may lead to interbreeding, making it difficult to reconstruct the history of the lineages involved, or even to determine how many lineages may have existed at one time. Within the Central Highlands of North America, a number of wide-ranging, variable species have been found to actually consist of groups of species (e.g. *Notropis rubellus* (Wood et al. 2002), *Etheostoma spectabile* (Ceas and Page 1997), and *Etheostoma punctulatum* (Mayden 2010); in this study, I show that the darter *Etheostoma zonale* is most likely another such group of species.

The *Etheostoma zonale* species group contains two currently-recognized species: the Banded Darter, *Etheostoma zonale*, found throughout the Interior and Ouachita Highlands and the Central Lowlands of North America, in the Mississippi River drainage, and small areas of the Great Lakes drainage; and the Brighteye Darter *E. lynceum*, found in the Mississippi Embayment region of the coastal plain, in the lower Mississippi River drainage and several Gulf Coast drainages. The widespread distribution of the species and its known morphological variation (Tsai and Raney 1974) have led to predictions that it would be a valuable source of information about the general pattern of diversification within the Central Highlands fish fauna (Mayden

1988a); however, until now, a molecular study of the group throughout its range has not been attempted.

Etheostoma zonale was described as *Poecilichthys zonalis* from the Holston River in Virginia by Cope in 1868. Jordan (1880) described *Nanostoma vinctipes* from a tributary of the Illinois River; this was soon synonymized with *E. zonale*. Jordan and Gilbert (1886) described *E. zonale arcansanum* from the Ouachita River; this was later considered a separate species before being dropped altogether as invalid. Hay (1881) described *Nanostoma elegans* from the Chickasawhay River in Mississippi; the species was later moved to *Etheostoma*,and the name was changed to *E. lynceum* by Hay in Jordan 1885 because it was preoccupied in *Etheostoma*. This species was later considered to be a subspecies of *E. zonale*. The first systematic study of variation in *E. zonale* was conducted by Tsai and Raney in 1974. They divided *E. zonale* into a nominate subspecies and *E. z. lynceum*, and divided these subspecies into seven and three morphological races, respectively, based upon various characters including the degree of squamation on the cheeks, breast and belly, lateral line count, and numbers of infraorbital pores and branchiostegal rays. Starnes and Etnier (1986) re-elevated *E. lynceum* to species status based on a lateral line count non-overlapping with neighboring populations of *E. zonale*, and features of the pattern of coloration, giving us the two species that are recognized in the group today.

I investigated the phylogenetic relationships of the *E. zonale* species group using mitochondrial cytochrome *b* and seven nuclear loci, six of which were newly developed for this study. I obtained specimens from the ranges of each of the morphological races described by Tsai and Raney (1974). I obtained gene trees for each locus; I also derived a species tree using *BEAST (Heled and Drummond 2010). Unlike gene trees, which show only the history of divergence of a single gene, species trees show the history of divergence of species or populations. Due to phenomena such as incomplete lineage sorting, the pattern of divergence of gene trees may not match that of the species which transmit them. The analysis of multilocus sequence data using species tree methods is thus essential in determining the actual divergence history of species (Knowles 2009).

I discovered a significant case of introgression between two divergent clades in *Etheostoma zonale*. Such introgression events are well-known in various darter taxa

(Piller et al. 2008, Ray et al. 2008, Keck and Near 2010); such events can cause difficulties in determining divergence times accurately.

I hypothesized that the *E. zonale* species group, due to its apparent morphological diversity, would show substantial geographical variation, and a temporal sequence of divergence that would correlate with that seen in other taxa from the Central Highlands.

Methods

Individuals of *Etheostoma zonale* were collected from 61 localities across the range of the species, and individuals of *E. lynceum* from 23 localities. Specimens were collected with seines and backpack shockers, and transported to the laboratory in liquid nitrogen or ethanol. Other species of *Etheostoma* were used as outgroup taxa. Specimens of *Etheostoma baileyi*, *E. barrenense*, *E. blennioides*, *E. rupestre*, *E. simoterum*, *E. thalassinum*, and *E. variatum* were used as outgroups for all nuclear genes sequenced; mitochondrial cytochrome *b* sequences of these species along with *E. atripinne*, *E. blennius*, *E. inscriptum*, *E. oophylax*, *E. podostemone*, *E. rafinesquei*, *E. raneyi*, and *E. swannanoa* were downloaded from Genbank.

Genomic DNA was extracted using QIAamp™ tissue extraction kits (Qiagen Inc.) according to the manufacturer's instructions. The complete mitochondrial cytochrome *b* gene was amplified using the polymerase chain reaction (PCR). PCR was performed in a total volume of 25 µL, containing 5 µL 5x Green GoTaq Flexi Buffer, 1.5 µL 25 mM $MgCl_2$, 0.5 µL 10 mM dNTPs, 0.03 nmol each of the forward and reverse primers, 0.125 µL GoTaq Flexi DNA polymerase, and approximately 0.5 µg of DNA, under the following thermocycler settings: initial denaturation at 95.0 ºC for 1 minute; 30 cycles of 95.0 ºC for 30 seconds, 53.0 ºC for 1 minute, and 72.0 ºC for 2 minutes; and a final extension at 72.0 ºC for 10 minutes, with the reaction terminating at 4 ºC. Cytochrome *b* was sequenced using the primers HA and LA of Schmidt et al. (1998). New internal primers were designed to allow complete sequencing of both strands: zon1f [5′ GTGCCACTGTYATYACYAACC 3′] and zon1r [5′ CCYGTYTCATGRAGGAARAG 3′]. Amplified DNA was purified using 4 units of exonuclease 1 and 0.4 units of shrimp alkaline phosphatase per reaction. Automated sequencing was performed using Big Dye terminator cycle sequencing at the DNA Sequencing and Analysis Facility of the Biomedical Genomics Center at the University of Minnesota. Sequences were checked for accuracy and assembled using Sequencher 4.7 (Gene Codes Corporation).

Due to initial difficulty with sequencing the entire cytochrome *b* gene, an alternative set of primers was used to sequence a 739 bp fragment of the cytochrome *b* gene for some samples. The primers zonalef [5′ GCTACTGCAACCAACCCYAACAC 3′] and zonaler[5′ GGAGTTAGGGGTGGGAGTTAAAA 3′] were used to amplify cytochrome *b* with the same reactants as given above, except that 3.0 µL of 25 mM $MgCl_2$ was used per reaction, under the following thermocycler settings: initial denaturation at 94.0 ºC for 5 minutes; 32 cycles of 94 ºC for 30 seconds, 54.4 ºC for 45 seconds, and 72.0 ºC for 1 minute; and a final extension at 72.0 ºC for 5 minutes, with the reaction terminating at 4.0 ºC. These amplifications were sequenced with a pair of internal sequencing primers: IF1 [5′ TACAAGAACCTYTAATGGCAAGC 3′] and IR1 [5′ TTAATGTGSGGAGGCGTCACTA 3′]; otherwise, purification and sequencing was as stated above. This resulted in the sequence for the entire gene being available, but only 739 bp with data for both strands. Some individuals which initially only had the 739 bp fragment sequenced later had cytochrome *b* amplified again, using the primers HA and LA, and were sequenced in only one direction using the primer HA; when combined with the 739 bp fragment this produced a full 1140 bp sequence, with both strands sequenced except for 24 bases at the 5′ end of the gene which were sequenced in only one direction.

526 basepairs of the 1st intron of the nuclear s7 gene were amplified using PCR under the same reaction conditions as described above for cytochrome *b* using the HA and LA primers, except that the annealing temperature used was 60 ºC. The primers used were S7RPEX1F and S7RPEX2R of Chow and Hazama (1998). Amplified DNA was purified and sequenced as described for cytochrome *b* above.

Six new nuclear intron loci were also developed for use in this study. The *Danio rerio* (zebrafish) genome build Zv9 was searched at the National Center for Biotechnology Information website for introns of a suitable length for sequencing in one segment, and with flanking exons of a suitable length for primer development. Where such introns were found, the flanking exons were were BLAST searched against the entire zebrafish genome, in order to weed out gene duplications. Exons without duplications elsewhere in the genome were then BLAST-like Alignment Tool (BLAT) searched at http://genome.ucsc.edu against the *Tetraodon*, fugu, medaka, and stickleback genomes to find conserved sequences. Where sequences of a suitable length for primer design were found in at least one of the other genomes examined, the exon sequences were used to design primers in Primer3 (http://frodo.wi.mit.edu/primer3/). All primer pairs developed were tested using the same PCR conditions as were followed for cytochrome *b* amplification with the HA and LA primers, except that the annealing temperature used was always 2 ºC less than the lower of the two melting temperatures of the primers. PCR products from loci which amplified successfully were purified and sequenced as described above for cytochrome *b*. Six loci sequenced successfully and showed variation within *Etheostoma*; their names and chromosomal locations are given in Table 3.1, and the primers designed for each locus, along with the annealing temperatures used during amplification, are given in Table 3.2. The nedd4l locus included a length of approximately 20 to 30 thymines in a row, which made sequencing difficult. Internal primers were designed to permit clear sequencing of a portion of the intron. The forward strand of pabpc4 did not sequence well with the primer used for amplification, and so a more specific sequencing primer was designed. The annealing time for the slco4a1 PCR reaction was increased to two minutes. Apart from this change, and the changes in annealing temperature, amplification conditions for all six nuclear loci were identical to those for cytochrome *b*.

Nuclear genes heterozygous at two or more sites were phased using one of three methods: sequences which were length heterozygotes were phased by eye in Sequencher (Sousa-Santos et al. 2005; Flot et al. 2006). Sequences which were not length heterozygotes were phased using the program PHASE (Stephens et al. 2001, Stephens and Scheet 2005) as implemented in the program DnaSP v5 (Librado and Rozas 2009). Homozygous sequences, sequences with only one heterozygous site, and sequences already phased due to a length heterozygosity were included as known sequences. For each run of the algorithm, the final portion of the run was set to be ten times as long as previous portions, and the algorithm was run five times for each locus; otherwise, default settings were used. If the phase of a sequence was reconstructed with a certainty threshold greater than 90 percent for all sites, this reconstruction was used in all subsequent analyses. Sequences which were not reconstructed by PHASE with this level of confidence were phased by PCR amplification using allele-specific primers (Petersson et al. 2003). Pairs of primers differing only at the last base at the 3′ end were designed, with a length sufficient for the melting temperature of the primer to match the melting temperature, within one or two degrees Celsius, of the regular amplification primer to be used at the other end of the sequence being phased. After allele-specific sequences were obtained, PHASE was re-run, using the sequences obtained using allele-specific sequencing as known haplotypes. If sequences which were previously reconstructed by PHASE with high confidence now showed a lower degree of confidence, allele-specific primers were also designed for these sequences. This cycle was repeated until all remaining unphased sequences were reconstructed by PHASE with high confidence.

All sequences for all genes will be deposited in Genbank.

Nuclear genes were tested for recombination using the phi test (Bruen et al. 2006) as implemented in Splitstree (Huson and Bryant 2006), as this test is suitable for detecting recombination both within populations of a single species, and across species. DNAsp v5 was used to determine the number of polymorphic sites at each locus. For each locus, the program jModelTest (Posada 2008, Guindon and Gascuel 2003) was used to select a suitable model of evolution. The Akaike Information Criterion was used to select models, and to select between an unpartitioned model, and one partitioned by codon position, for cytochrome *b*. Gaps in nuclear loci were coded for analysis using the program SeqState (Müller 2005), which codes gaps as binary presence/absence characters, or, where two or more gaps with different start and end points overlap, as multistate characters. Gaps with ambiguous alignments were not coded. Gaps with more than two states were excluded, to avoid having an additional model of evolution for a partition with very few characters; this applied to a total of 43 mostly parsimony-uninformative gap characters. Some taxa had very long gaps, which spanned several other gaps in other taxa; to allow those smaller gaps to be coded as binary presence or absence characters, these long, overlapping gaps were coded as missing data in the taxa that displayed them. Maximum likelihood analyses were performed in Garli 2.0 (Zwickl 2006). For nuclear genes, all analyses were partitioned, with one partition for the sequence data, and another for the gap data. For all loci, the Mkv model (Lewis 2001) was used for the gap data. One hundred bootstrap replicates were run for each locus. To reduce the amount of time for bootstrapping in Garli, for cytochrome *b*, the bootstrap analysis was run on a dataset with all duplicate haplotypes removed, except for those required to obtain a bootstrap value for clades with no haplotype variation. The Akaike Information Criterion (AIC) as implemented in Mrmodeltest 2.3 (Nylander 2004) was used to determine suitable models for analysis of each gene in MrBayes; the AIC was also used to compare a model partitioned by codon to an unpartitioned model for the cytochrome *b* dataset. Cytochrome *b* was analyzed in MrBayes 3.1.2 (Huelsenbeck and Ronquist 2003; Ronquist and Huelsenbeck 2005) using a model partitioned by codon; nuclear genes were run with one partition for sequence data and a second for gap data. All runs were 10,000,000 generations long, with a burnin of 1,000,000 generations.

In several cases, apparent incongruence was detected between mitochondrial and nuclear gene trees. To determine if this incongruence reflected a genuine difference in the data sets, the approximately unbiased test (Shimodaira 2002) was performed. In all cases, a clade that was monophyletic according to one gene was paraphyletic or polyphyletic according to another. An unconstrained analysis was performed in MrBayes 3.1.2 on the locus in which the clade in question was para- or polyphyletic, and then a second analysis was performed in which the clade was constrained to be monophyletic. In both cases the model and run conditions used were the same as for the standard MrBayes analyses mentioned above. The five percent of trees in the posterior distribution with the highest posterior probability for each analysis were loaded into PAUP* 4.0, (Swofford 2002) and site likelihoods calculated, using the model of sequence evolution for each locus determined using jModelTest. The site likelihood files were then combined and used as the input for the program Consel v0.1j (Shimodaira and Hasegawa 2001), which was run under default settings to conduct the AU test. If any of the trees from the constrained analysis were not significantly different than the best tree from the unconstrained analysis, the hypothesis that the incongruence among data sets is genuine was rejected for that locus.

A species tree analysis was conducted using *BEAST (Heled and Drummond 2010). The terminals for the analysis were the seven outgroup taxa for which nuclear gene sequences were obtained, the 21 allopatric clades for *Etheostoma zonale* and *E. lynceum* obtained from the cytochrome *b* gene tree, and the South Fork Holston clade of *E. zonale*, which appeared as a separate clade in nuclear gene trees, but was intermingled with the Upper Tennessee River clade in the cytochrome *b* data. Models for each locus were selected using MrModelTest. Cytochrome *b* was run unpartitioned for the analysis; the cytochrome *b* data set was also reduced to ten individuals per clade. In clades with more than ten individuals, the individuals for the *BEAST analysis were chosen so that as much of the phylogenetic structure within each clade as possible would be represented. All loci were run with empirical base frequencies. Molecular clock likelihood ratio tests for each locus were performed in PAUP* to determine whether each locus should follow the strict clock model, or the relaxed clock uncorrelated exponential model. Default priors and operators were used. Two analyses were run: one with cytochrome *b* and all seven nuclear loci; and a second analysis with nuclear loci only. For the cytochrome *b* + nuclear analysis, the cytochrome *b* sequence data for the South Fork Holston race was represented as missing. Missing data were also used to represent outgroup taxa at loci for which a sequence was not obtainable. Several runs were performed for each analysis to assess convergence between runs, and then combined to obtain the final species tree. The analysis including cytochrome *b* had difficulty converging on the same posterior values across runs; the results of the best run were used to select more informative priors for a new set of runs. The species trees were calibrated using a sequence divergence of 2.0% per million years (Near and Benard 2004).

To test for population growth in populations of *Etheostoma zonale* and *E. lynceum*, the R2 Test (Ramos-Onsins and Rozas 2002) was used. The R2 test was performed on the cytochrome *b* sequence data for each allopatric clade with a sufficient number of sequences (at least ten) for the test. The test was performed in DnaSP v5; to determine the significance of the test statistic for each test, 10,000 coalescent simulations were performed in DnaSP.

Extended Bayesian Skyline Plots (Heled and Drummond 2008) were obtained for the Upper Tennessee and South Fork Holston clades of *Etheostoma zonale* using the program BEAST. Due to the incongruence between mitochondrial and nuclear gene trees with regard to these two races, and the geographical distribution of individuals attributable to the South Fork Holston race, it was hypothesized that the Upper Tennessee Race is in the process of replacing the South Fork Holston race; accordingly, the Upper Tennessee Race should show evidence of population growth in an Extended Bayesian Skyline Plot, while the South Fork Holston Race should not. An EBSP was obtained for the Upper Tennessee clade using all loci, while the EBSP for the South Fork Holston clade was obtained using nuclear loci only. In both analyses, substitution models, clock models, and trees were unlinked for all partitions. Site models for each locus were selected using MrModelTest; where MrModelTest chose the F81 model, HKY was used in its place. An initial run of the Upper Tennessee clade data using the models selected showed poor convergence for cytochrome *b* and s7, the three loci for which a GTR model was selected. The analysis was rerun using the HKY model for these three loci, and much better convergence was observed. Empirical base frequencies were used for all loci. A molecular clock likelihood ratio test was performed for each locus to determine whether or not to use the strict clock model in BEAST. A rate of 2.0 ± 0.2 % sequence divergence per million years (Near and Benard 2004) was set as the prior for mitochondrial divergence rate. For the South Fork Holston race, which did not have cytochrome *b* data, an equivalent prior was placed on nedd4l, converted by using the relative rate of nedd4l to cytochrome *b* in the *BEAST species tree analysis. The prior for all other clock rates was uniform on the range [0,0.2]. The weights for the "demographic.populationMeanDist", "demographic.indicators", and "demographic.scaleActive" operators were set to 160, 320, and 60, respectively, for the Upper Tennessee clade, and 140, 280, and 60 for the South Fork Holston clade. For each locus in both analyses, the weight for the "Substitution rates and heights" operator was increased to 15, and the weight for the kappa operator was increased to 2. The mean of the distribution of population sizes was set to 1. Each analysis was run for 50,000,000 generations.

Results

Cytochrome *b* was successfully sequenced for 324 ingroup individuals. Of this, 130 individuals had only a 739 bp fragment sequenced, while 194 had the full 1140 bp locus sequenced. Several individuals were sequenced first to obtain a 739 bp fragment, and then again, starting from a new PCR, to obtain the full-length locus; these individuals served to test the PCR and sequencing error rate for cytochrome *b*. No errors were detected in over 10,000 bp of duplicate sequence, giving an error rate of less than 0.01%.

Between 71 and 122 ingroup individuals were sequenced for each of the seven nuclear loci, giving twice as many alleles. In addition, sequences of each intron were obtained from between four and seven of the outgroup taxa. Loci varied substantially in length; the range of lengths is given in Table 3.3. All loci also included gaps of various lengths, some of which were parsimony-informative. The number of polymorphic and parsimony-informative sites at each locus, along with the number of sequences obtained at each locus, the length of each aligned locus, and the number of gaps coded for each locus are given in Table 3.4.

All nuclear loci were tested for recombination; nedd4l, s7 and eif3c gave significant results. Examination of the allele network produced in SplitsTree for eif3c suggested that *E. variatum* was the source of the signal of recombination; removing this taxon led to a non-significant result for the test of recombination. Nedd4l and s7 also tested positive for recombination, with no obvious single source of the signal; the output of the recombination test in DnaSP was used to determine where to trim the aligned sequences until there was no longer a signal of recombination. The aligned nedd4l sequences went from 453 bp to 346 bp after trimming, while s7 went from 526 bp to 394 bp. Nedd4l varied greatly in length among the outgroup taxa. Only four of the outgroup taxa contained the sequenced fragment of the intron; the sequence which was obtained for the remaining three species either overlapped only slightly with the ingroup fragment (*E. rupestre*) or not at all (*E. simoterum* and *E. barrenense*). Individuals of *E. lynceum* from the Homochitto, Amite, Tangipahoa, Pearl and Pascagoula populations had a repeating AT element of indeterminate length, which prevented sequencing the entire length of the locus. As a result, sequences of slco4a1 for these populations were not confirmed by sequencing in both directions. All individuals sequenced from the Homochitto population had a heterozygous site to either side of this repeating element, preventing allele-specific phasing. The program PHASE was unable to determine the phase of these individuals, and so they were excluded from analyses. For stx5a, *E. simoterum* and *E. barrenense* had a similar repeating AT element. The individual of *E. simoterum* sequenced was heterozygous at three sites, spanning the repeating element, preventing its use in subsequent analyses. Except for these cases, all loci were successfully phased using one of the three methods described above. Allele-specific sequencing resulted in duplicated PCR reactions for over 1000 bp of sequence in all nuclear loci, this served as a test of the PCR and sequencing error rate for these loci. No errors were detected, giving an error rate of less than 0.1%.

Models of evolution selected for likelihood and Bayesian analyses of each locus are given in Table 3.5. The AIC selected a model partitioned by codon position for cytochrome *b*. The results of the likelihood analysis are shown in Figure 3.2. Cytochrome *b* shows a high degree of geographically-ordered variability: there are nine clades in *E. lynceum* and 12 in *E. zonale* with monophyly well-supported, whose members are geographically separate from those in all other clades. No geographic locality includes individuals belonging to two highly divergent, well-supported distinct clades. All of these clades are also strongly supported in the Bayesian analysis, except for the Ouachita and Upper Arkansas clade. The distributions of the clades are mapped in figure 3.3.

The basal split in the ingroup is between five southern clades of *E. lynceum*, and four northern clades of *E. lynceum* along with all of *E. zonale*, rendering *E. lynceum* paraphyletic. Within the southern *E. lynceum* clade, found in southern Mississippi and small parts of Louisiana and Alabama, there are five well-supported clades, with a well-supported branching order, except for the most recent divergence between the Tangipahoa and Pearl clades. The order of divergence does not proceed geographically from east to west or vice versa, but rather goes from either end of the distribution towards the middle. The Pearl River clade itself shows some geographical structuring, with individuals from the Strong and Upper Little Rivers together forming one subclade, those from McGee Creek a second subclade, and those from the Wolf River, which drains separately into the Gulf of Mexico from the Pearl River, forming a third subclade.

Within the clade consisting of northern *E. lynceum* and *E. zonale*, there are seven clades with no support for the branching order among them. The first of these clades is the northern *E. lynceum* clade, found in central and northern Mississippi and western Tennessee and Kentucky, which is divided into four smaller clades. The first split within this clade is between the Yazoo river clade, and three other clades, one found in the Hatchie and Wolf River systems, one in the Obion River system, and one in the Big Black River system. The Yazoo clade itself shows geographical structure, with one monophyletic subclade found in the Yalobusha and Yocona Rivers, and a second found in the Little Tallahatchie River.

The remaining clades all belong to *Etheostoma zonale*. One of these clades is a well supported grouping of two clades, one found in the Green River in Kentucky, and one found in the Upper Tennessee River. The Upper Tennessee River clade includes all populations in the Tennessee and its tributaries down to the confluence of the Elk River with the Tennessee in Alabama, with the exception of populations in the Little Tennessee River and Hiwassee River (upstream of its confluence with the Ocoee). The Upper Tennessee clade shows some geographical structuring; for instance, individuals from the French Broad River system form a monophyletic subclade within the clade, as do individuals from the Sequatchie River system, and individuals from the North Fork and main stem of the Holston River. The most widely distributed haplotype, however, is spread from Richland Creek in the westernmost part of the clade's range, to the Middle Fork of the Holston, in the northeasternmost part of the range.

A clade with weak bootstrap support and no Bayesian support is formed by the Cumberland, Duck, Little Tennessee, and Hiwassee River clades. The Little Tennessee and Hiwassee clades form a strongly supported clade, as do the Duck and Cumberland River clades. The Duck River clade includes individuals from within the Duck River basin, as well as from Indian Creek, which drains directly into the Tennessee River upstream of its confluence with the Duck. The Cumberland River clade shows geographical structure; individuals from the Rockcastle River in Kentucky, in the upstream portion of the Cumberland Basin, form a clade, as do the remaining individuals in the Cumberland clade. Individuals from the Big South Fork of the Cumberland form a subclade within this second clade.

A fourth clade is formed by two clades found in the Upper Ohio and Kentucky River systems; each of these two clades is strongly supported. A fifth clade is formed by individuals from the Upper Mississippi River, which has the largest range of any of the cytochrome *b* clades. Individuals in this clade are found from the northern Ozark mountains in Missouri, east to Illinois and north to Minnesota. Most individuals from throughout this range also share a single haplotype. Individuals from the Meramec River and Gasconade River systems in Missouri, however, each form subclades.

The final clade is formed by three clades in the White, Little Red, and Ouachita and Upper Arkansas river systems, in the Ouachita and southern Ozark mountains. The basal split in this group is between the White River clade, which is spread throughout the upper White, Black, and St. Francis river systems in the southern Ozarks of Arkansas and Missouri, and the remaining two clades. One of these is the Little Red River clade, found only in the Little Red River drainage in central Arkansas; the final clade, the Ouachita and Upper Arkansas clade, is more broadly distributed in Arkansas, Oklahoma, Kansas and Missouri. Within this clade, individuals from the Saline River in central Arkansas form a subclade sister to all remaining individuals. Individuals from the Ouachita River drainage also form a subclade, but the remaining individuals from the Upper Arkansas basin are paraphyletic with respect to this clade.

The seven nuclear gene trees are shown in Figures 3.4-3.10. Each nuclear locus showed some geographical structuring, but not as much as seen in cytochrome *b*; in none of the nuclear gene trees do all individuals belong to one allopatric clade or another. A key feature of four of the seven nuclear genes is the presence of a clade not evident in the cytochrome *b* data: in the Upper Tennessee River, individuals from the Clinch, Middle Fork Holston, and Elk Rivers in the northern part of the drainage form a separate clade or clades that are well-separated from the remaining individuals in the Upper Tennessee River clade. These individuals do not even form a separate subclade in the cytochrome *b* gene tree. Because the distribution of this clade is congruent with that of

Tsai and Raney's (1974) South Fork Holston race, it is referred to here as the South Fork Holston clade.

In the eif3c gene tree (Figure 3.4), the South Fork Holston individuals form two separate branches, one of which includes some individuals of *E. lynceum* as a sub-branch. Individuals belonging to the Upper Ohio and Kentucky cytochrome *b* populations form a well-supported clade here, as do individuals belonging to the Upper Mississippi, White, Little Red, and Ouachita and Upper Arkansas populations. All other individuals show very little geographical structure, however, and the most common haplotype is shared by *E. lynceum* and individuals in the Tennessee and Cumberland river drainages.

The ncl1 gene tree (Figure 3.5) shows greater geographical structure. The South Fork Holston clade actually does form a well-supported clade, but it is nested within a well-supported Upper Tennessee and Hiwassee clade. There is some geographical structure within *E. lynceum*; individuals belonging to the two northernmost populations, the Obion and Hatchie, form two branches within the tree, and individuals belonging to the seven southern populations also form two branches. The Duck, Cumberland, and Little Tennessee populations form a well-supported clade; within this clade, the Little Tennessee population forms a well-supported clade, while the two other populations are paraphyletic. The Upper Ohio and Kentucky populations again forms a well-supported clade.

The nedd4l gene tree (Figure 3.6) features a well-supported South Fork Holston clade; however, some individuals from the Clinch River grouped with individuals from the Upper Tennessee River instead. Most of *E. lynceum* forms a well-supported clade, with the exclusion of individuals from the Obion, and some Yazoo individuals. The Duck and Cumberland populations together form a clade. Little Tennessee individuals form a clade, but one that is not supported.

In the pabpc4 gene tree (Figure 3.7), the South Fork Holston clade is well-supported, as is a clade consisting of the five southern populations of *E. lynceum*. The Upper Ohio and Kentucky populations form a clade which renders what would otherwise be an Upper Tennessee, Little Tennessee and Hiwassee clade paraphyletic.

The s7 gene tree (Figure 3.8) groups *E. lynceum* together, not as a monophyletic clade, but rather as a series of paraphyletic steps at the base of the tree. It is possible that the tree is misrooted, and *E. lynceum* is actually monophyletic at this locus. The South Fork Holston clade does not appear; its members are mixed with individuals from the Upper Tennessee and Hiwassee populations in a clade which only receives marginal support. The Duck and Cumberland populations are each paraphyletic basal to this clade. The Little Tennessee population does form a clade that receives strong support.

In the slco4a1 gene tree (Figure 3.9), the Upper Tennessee, Little Tennessee, Hiwassee, and South Fork Holston populations all form a clade that receives strong support. *E. lynceum* forms two clades, neither of which are supported: one including the six southernmost populations (sequences for the Homochitto population are not available) and one including the four northern populations. The Duck and Cumberland populations together form a well-supported clade.

The stx5a gene tree (Figure 3.10) includes a well-supported South Fork Holston clade, and a well-supported monophyletic *E. lynceum*. It also includes well-supported Cumberland, Duck, and Upper Ohio and Kentucky clades, and a marginally-supported Ouachita and Upper Arkansas clade.

An approximately unbiased test was conducted for each of the four loci (eif3c, nedd4l, pabpc4, and stx5a) which showed a South Fork Holston clade or grouping distinct from individuals in the remainder of the Upper Tennessee river cytochrome *b* clade. The test compared likelihoods of unconstrained trees for each locus to likelihoods of trees constrained to make all members of the Upper Tennessee cytochrome *b* clade monophyletic. The AU test did not reject monophyly of the Upper Tennessee populations for nedd4l (p=0.358) or stx5a (p=0.309), but did reject monophyly for eif3c (p=0.002) and pabpc4 (p=0.002). An AU test was also conducted to determine whether monophyly of *E. lynceum* in the cytochrome *b* gene tree could be rejected; the test was unable to reject monophyly (p=0.738).

Cytochrome *b* and nuclear trees for all loci indicating the identity of each individual included in the analysis are given as Figures 3-S1 to 3-S8 in the supplement.

The *BEAST species tree generated from cytochrome *b* and the seven nuclear loci is shown in Figure 3.11. The tree was generated from 100,000,000 post-burnin generations spread across four separate runs, achieving convergence of parameters across runs, and satisfactory effective sample size (ESS) values for all key parameters in the combined runs. The branching pattern shows several clear differences from the cytochrome *b* tree. *Etheostoma lynceum* is monophyletic with strong support; it is divided into the same two major clades as appear in the cytochrome *b* tree, but the branching pattern differs within each. Within *E. zonale* there are further differences. *Etheostoma zonale* is monophyletic with strong support, and there is more support for the branching order among its major clades than there is in the cytochrome *b* tree. It is divided into two major clades: one including northern and western populations, and one including southeastern populations in the Tennessee and Cumberland River basins. In the northern and eastern clade, whose monophyly is strongly supported, the Green River clade groups with the Kentucky and Upper Ohio clades with strong support, rather than with the Upper Tennessee clade as in the cytochrome *b* tree. This may be an indication of mitochondrial introgression from Tennessee River populations as the source of cytochrome *b* in the Green River clade; such an event must have happened sufficiently far in the past for the Green River clade's cytochrome *b* to become reciprocally monophyletic with that of the Upper Tennessee clade. A second clade within the northern and western clade is formed by the Upper Mississippi clade grouping with strong support with the White, Little Red, and Ouachita and Upper Arkansas clades. The remaining clades form the southeastern clade, which is however poorly supported. Within this grouping, the Upper Tennessee clade is most closely related to the Hiwassee clade; the South Fork Holston clade is sister to these two clades and the Little Tennessee clades, though with weak support, while the Duck and Cumberland form a clade with strong support that is sister to the Upper Tennessee, Little Tennessee, South Fork Holston and Hiwassee clades with weak support.

The species tree generated from nuclear loci only (Figure 3.12) was obtained from 150,000,000 post-burnin generations across five runs. It is similar to the first species tree, but with some changes in branching pattern and generally lower levels of support. *Etheostoma lynceum* is still monophyletic with strong support; within the species, the Big Black clade has switched from being among the northern clades to the southern ones. The branching order of the five southernmost clades also differs. The monophyly of *E. zonale* is not strongly supported, but the northern and western clade still shows high levels of support both for its monophyly and the monophyly of most of its subclades. In the southeastern clade, the South Fork Holston clade has changed position, to be sister to the five remaining clades within this grouping; however, support for the branching order within the southeastern clade is weak to nonexistent, with the exception of the strongly supported grouping of the Duck and Cumberland clades and the Upper Tennessee and Hiwassee clades. A relative rate of evolution cytochrome *b* to eif3c of 0.134 was obtained from the cytochrome *b* + nuclear *BEAST analysis; this was used to make the divergence percentages in each species tree comparable. In both trees, a rate of sequence divergence of 2 % per million years, as used in darters by Near and Benard (2004) places the divergence of the *E. zonale* species group from its nearest relatives in the Miocene, and the divergence between *E. zonale* and *E. lynceum* in the late Pliocene. The divergence between the two main clades of *E. zonale* occurred in the late Pliocene or early Pleistocene, and all subsequent diversification is Pleistocene in date.

The results of population growth tests for each of the cytochrome *b* clades using the R2 test are given in Table 3.6. The Upper Ohio, Upper Mississippi, and Little Red clades all show a significant signal of recent population growth, as does the Ouachita subclade of the Ouachita and Upper Arkansas clade.

Extended Bayesian skyline plots were obtained for the Upper Tennessee and South Fork Holston clades; both analyses were run for 50,000,000 generations and achieved convergence of parameters and satisfactory ESS values. The plots are shown in Figure 3.13. The plot for the Upper Tennessee race shows an increase in population size within the past 40,000 years (assuming a generation time of one year). The 95 % highest posterior density (HPD) around the posterior for the number of population size changes ranged from one to three, with a mode of one population size change. The plot for the South Fork Holston population shows no increase in population size; the 95 % HPD lies between zero and two changes, with a mode of zero.

Discussion

Molecular data demonstrates that the *E. zonale* species group is highly differentiated geographically; *E. zonale* itself contains twelve allopatric cytochrome *b* clades, while *E. lynceum*, with a much smaller range than *E. zonale*, contains nine. Some of these clades correspond to those defined on morphological grounds by Tsai and Raney (1974) while others do not. Tsai and Raney's Upper Mississippi River race corresponds to the Upper Mississippi clade; their Ohio River race corresponds to the clade containing the Kentucky and Upper Ohio clades, and their Hiwassee River race corresponds to the clade containing the Hiwassee and Little Tennessee clades. Their Arkansas, Black, Tennessee, and South Fork Holston races do not correspond, however, to monophyletic groups of cytochrome *b* clades. In *E. lynceum*, the Pascagoula River race corresponds to the Pascagoula clade, but the other two races defined by Tsai and Raney do not correspond to the cytochrome *b* phylogeny. None of the nuclear data define monophyletic, allopatric groups nearly to the same extent as cytochrome *b*; however, taking the loci together, a species tree is produced with a better-supported backbone than seen in any of the individual gene trees.

When compared with the PACT cladogram depicted in Chapter 2, the species tree for *E. zonale* shows a number of clear similarities, which are shown in Figure 3.14. The southeastern clade of *E. zonale* does not appear to share many similarities with the taxa included in the PACT analysis; none of the taxa in the PACT analysis show the degree of differentiation within the Tennessee and Cumberland River drainages that is shown by *E. zonale*. The split between the Cumberland and Tennessee drainages in *E. zonale* appears to be recent compared to similar splits in other taxa. Within the northern and western clade, however, the split between populations in the Ohio, Kentucky, and Green river basins and those west of the Mississippi appears to be a part of Common Divergence Event 7; as *Erimystax x-punctatus* and *Etheostoma blennioides* show the same pattern at the same time. The split between the Upper Mississippi clade and the three clades to its south matches the pattern of Common Divergence Event 8, also seen in *Percina evides*, *Hypentelium nigricans*, and *E. blennioides*. The split between the White River clade and the Ouachita and Upper Arkansas and Little Red clades matches the pattern of g Event 11, also seen in *E. blennioides* and *E. caeruleum*. Finally, the split between the Ouachita and Upper Arkansas and Little Red clades in *E. zonale* matches the corresponding split in *E. blennioides*, which creates an additional Common Divergence Event. Thus, except in the Tennessee and Cumberland River drainages, *E. zonale* shows a pattern of divergence which matches that of other taxa, particularly its relative *E. blennioides*, suggesting that vicariance was in large part the cause of its diversification in these areas. The population growth analysis using the R2 test complicates this assessment slightly. Positive evidence for population growth is found in the Upper Ohio and Upper Mississippi clades; this should not be surprising, as much of the range of both these clades was glaciated until the end of the Pleistocene. Population growth is also evident in the unglaciated Little Red and Ouachita river drainages, however. The vicariant splitting of a previously-widespread population should not result in a signal of population growth. Population growth would be expected if a clade originates by dispersal across a barrier; the presence of a common biogeographical pattern then suggests that dispersal due to the temporary loss of a barrier may be the cause of divergence in such a situation. The nature of the Little Red and Arkansas river divergence is further examined in Chapter 5.

Basing his analysis on Tsai and Raney's (1974) delimitation of races in *E. zonale* and *E. lynceum*, Mayden (1988a) made a number of predictions concerning the biogeography of the species. Mayden predicted that *E. zonale* evolved in Central Highland drainages including the Tennessee River, which is not contradicted by the present study. His prediction that the Ohio River race is monophyletic and more closely related to races west of the Mississippi River than to the Tennessee River race is also borne out. Mayden's prediction that the Green River race is more closely related to the Tennessee, Duck, and Cumberland River populations than to the Ohio River race is not supported, except by the sister relationship of the Green and Upper Tennessee River populations in the cytochrome *b* gene tree. Also refuted is Mayden's prediction that the Upper Mississippi race would turn out to be non-monophyletic, with the Missouri River populations more closely related to populations in the southern Ozarks than to populations in the Upper Mississippi.

The most unexpected result of the nuclear gene analyses is the evidence for the existence of Tsai and Raney's (1974) South Fork Holston race, and its hybridization with the Upper Tennessee clade. Four nuclear genes show that individuals from range of Tsai and Raney's South Fork Holston race form a lineage separate from the Upper Tennessee clade; in two genes this is supported by the rejection, according to the AU test, of monophyly for the two clades together. There is no evidence of the South Fork Holston clade in the mitochondrial data, however, suggesting introgression; the status of several of the nuclear loci indicates that full-level hybridization may be ongoing. The ancestral mitochondrial lineages of this race are now apparently gone; it is also unclear whether the three loci which do not distinguish the Upper Tennessee and South Fork Holston races represent the ancestral state prior to differentiation of the clades, or are the result of more recent hybridization. Hybridization appears to have made the least progress in the South Fork of the Holston River itself, where no individuals belong to the Upper Tennessee race at those loci that do distinguish the races. In the Clinch River, some individuals bear alleles at these loci belonging to the South Fork Holston clade, while others bear alleles of the Upper Tennessee clade. Some individuals bear one allele from each clade, confirming hybridization rather than simple coexistence in this population. Individuals from the Sequatchie and Toccoa Rivers display an intriguing sign of hybridization. While the gene tree analysis for nedd4l places all individuals in these two populations with the Upper Tennessee clade, this is based upon a fragment of nedd4l reduced in length to remove the signal of recombination from the data set. This analysis used the left-hand portion of the full nedd4l sequence; a separate maximum likelihood analysis on the right-hand portion that was removed (not shown) grouped four alleles from the Sequatchie River and two from the Toccoa River with the South Fork Holston clade. This is clear evidence of recombination, and thus hybridization, between the two clades in these river systems. Tsai and Raney did not report any morphological similarity of specimens from these rivers to the South Fork Holston race; this matter is one that bears further investigation.

The South Fork Holston clade is unusual in that it is disjunct in its distribution. The Middle Fork Holston and Elk River populations, part of the South Fork Holston River itself, are contiguous, but the Clinch River population is separated from these two populations by the North Fork and the main stem of the Holston River, which appear to contain only individuals belonging to the Upper Tennessee clade. The Sequatchie and Toccoa River populations are also disjunct from each other, and a great distance away from the Clinch and South Fork Holston populations. This disjunct distribution is most easily explained if the South Fork Holston clade was once distributed throughout the Upper Tennessee River drainage, but has been displaced by most of its former range by the Upper Tennessee clade dispersing from downstream (Figure 3.16). If this is the case, one would expect the Upper Tennessee clade to show a historical increase in population size, with the remaining representatives of the South Fork Holston clade showing constant population size or a decrease in population. Extended Bayesian skyline plots (Figure 3.13) confirm this hypothesis. The South Fork Holston clade shows a modal number of past population size changes of zero, while the 95 % highest posterior density of the number of population size changes for the Upper Tennessee clade excludes zero, and the plot for this population shows an increase in population size starting about 40,000 years ago. Why one taxon should replace another closely-related taxon like this is unclear; the Upper Tennessee race may have some non-evident selective advantage over the South Fork Holston race. It is also unclear if the Upper Tennessee Race invaded the Upper Tennessee upon gaining this advantage, or if it was previously held back by either a lack of suitable habitat or lack of a drainage connection from entering the Upper Tennessee from its previous downstream range. A number of other fish taxa show differentiation between the main stem and the headwaters of the Upper Tennessee River basin (e. g. *Nothonotus camurus* and *N. chlorobranchius* (Keck and Near 2008)) and the possibility that a similar phenomenon is responsible should be considered.

The species tree analyses performed in this study rely on the assumption that incongruence among loci is due to incomplete lineage sorting. For the South Fork Holston clade, this was clearly not the case. Since the cytochrome *b* sequences belonging to South Fork Holston individuals are introgressed Upper Tennessee clade haplotypes, these had to be discarded and replaced with missing data for the species tree analysis. This does not seem to have had a terrible effect on the analysis, but it is important to remember that this introgression was only detected because the Upper Tennessee and South Fork Holston clades have diverged sufficiently. Introgression events between sister taxa may not be detectable by examination of individual gene trees, and it seems apparent that any such events would confound estimates of divergence times for those taxa. This needs to be taken into account when considering divergence times of closely-related taxa in groups in which introgression is known to be prevalent; ideally methods of species tree analysis will be developed that explicitly account for introgression.

Species delimitation

Given the pattern of differentiation seen in the various nuclear gene trees, the question of how many species actually belong to the *E. zonale* species group naturally arises. Given that the putative species are all allopatrically distributed, and, given the evidence of hybridization seen in this study, not reproductively isolated, this question is, unfortunately, intertwined with the question of species concepts. The phylogenetic species concept defines a species as "a diagnosable cluster of individuals within which there is a parental pattern of ancestry and descent, beyond which there is not, and which exhibits a pattern of phylogenetic ancestry and descent among units of like kind" (Eldredge and Cracraft 1980). Under this concept, *E. zonale* may be divisible into six species, and *E. lynceum* into two; the general lineage species concept of de Queiroz (1998) also supports such a conclusion.

In many ways, the South Fork Holston clade of *E. zonale* has the greatest claim to being a distinct species. It shows six unique, unreversed synapomorphies at the ncl1, nedd4l, and stx5a loci: four transversions and two transitions. It is also morphologically distinct from the other races of *E. zonale*, having eight infraorbital pores rather than seven; however, this is apparently a plesiomorphic trait (Tsai & Raney 1974). Complicating the recognition of this race as a species is its hybridization with the Upper Tennessee clade. The hybridization has progressed the least in the Middle Fork Holston and Elk Rivers; and it is these populations only which show the full suite of synapomorphies in all individuals. The degree of hybridization seen in the remaining populations with alleles belonging to the South Fork Holston clade suggests that it may represent a species that is in the late stages of extinction through hybridization.

The Upper Ohio and Kentucky clades together form another putative species. They appear as a monophyletic clade at four loci, and share six unique unreversed synapomorphies at three loci; four transitions and two transversions. However, no unique morphological characters are known for the race. More problematic is the status of the Green River clade. The species tree analysis places it with the Kentucky and Upper Ohio clades with strong support, even though no locus includes all individuals of these races in a monophyletic group. Indeed, most loci show the Green River race to be polyphyletic. Stx5a places the Green River clade with the Upper Mississippi and White River clades, an intriguing result, as Tsai and Raney considered the Green River population to be an intergrade between the Ohio River race and the Black River race (which forms a part of the White River clade.) Much of the Green River was impounded during the Pleistocene, forming a glacial lake (Thornbury 1965), which may have made it unsuitable habitat for most highland fish species (Mayden 1988a). It is possible, then, that the Green River population was recolonized after this period by populations from both the Upper Ohio and Kentucky River clades on one hand, and by populations from the White River clade on the other, crossing the Mississippi when it was of a sufficiently high gradient to form suitable habitat. Tsai and Raney's assessment of the Green River population as an intergrade thus appears to be correct, complicating matters if *E. zonale* is to be subdivided into new species.

The Duck and Cumberland clades could also be defined as species, or, more conservatively, as one species together. The Duck and Cumberland clades together show five unique unreversed transitions and two unique unreversed transversions at three loci. The Cumberland clade shows two unique unreversed transversions and one unique unreversed transition at stx5a, while the Duck clade shows two unreversed transitions at this locus, which are, however, repeated in outgroup taxa. Tsai and Raney grouped the fish from these drainages along with those from the Tennessee River in a single Tennessee race, and found little morphological difference between the Cumberland and Tennessee drainages. The distinctiveness of these clades molecularly, with very little morphological difference from the Upper Tennessee clade, suggests that they may have been evolving neutrally for a long time, in an area where habitat conditions have changed little for millions of years.

The Little Tennessee clade is supported by two transversions at the ncl1 locus; Tsai and Raney grouped this population with the Hiwassee population in their Hiwassee race, but gave no unique derived morphological characters that distinguish this race. The Hiwassee clade is not at all distinct at any of the nuclear loci; indeed, it usually shares a common haplotype with members of the Upper Tennessee clade. These two clades together never form a monophyletic group exclusive of other clades, and they do not appear to have any morphological synapomorphies. These two clades may, therefore, represent the plesiomorphic state in *E. zonale*.

The populations of *E. zonale* west of the Mississippi form a well-supported clade in the species tree analysis, but there are no synapomorphies for the group as a whole. The Upper Mississippi clade shows one unique transversion, but it is also notable for having several apparent morphological and developmental synapomorphies: Adult male breeding coloration is darker green in the Upper Mississippi clade than in other races, and the Upper Mississippi clade has ontogenetically faster yolk depletion (Simon and Wallus 2006). The Ouachita and Upper Arkansas clade is supported by one transition in stx5a; no unique synapomorphies link the White River clade to this clade. Should these four clades be separated from *E. zonale*, the valid name for the species would be *E. vinctipes*; if the Upper Mississippi clade were to be separated from the remaining clades, it would take the name *E. vinctipes*, while the remaining clades would take the name *E. arcansanum*.

While the numbers of characters supporting species status for each of these putative species may seem low, it is worth noting that, at the same loci, the monophyly of *E. lynceum* is supported by only three unique synapomorphies, two transitions and a transversion in stx5a. Starnes and Etnier (1986) removed *E. lynceum* from the synonymy of *E. zonale* on the basis of a difference in lateral line scale counts and coloration pattern, a rather greater degree of morphological differentiation than seen among most races of *E. zonale*. However, the evidence of the nuclear loci suggests that closer examination of morphological characteristics of the putative species described above may reveal an equivalent degree of morphological differentiation. Examination of male breeding coloration would be particularly worthwhile, as this is a transient character that may not have been noted in the past, and could serve as an isolating mechanism among lineages. Even if few diagnostic characters have evolved in each clade, it is evident from the gene trees that many of these clades are now evolving as distinct lineages, and so could be considered to be distinct species under the lineage species concept. If so, most of these species appear to be at an early stage in their diversification, particularly as undifferentiated lineages possessing plesiomorphic characters still exist (the Upper Tennessee clade.)

The status of the populations of *E. lynceum* also poses an interesting problem. Southern populations of *E. lynceum* are supported as a monophyletic grouping by one unique unreversed transition in nedd4l, and by a long insertion with a repeating AT element in slco4a1. Ncl1 and pabpc4 also support a division of the species into northern and southern groupings, though with no unreversed synapomorphies to support these groupings. However, the geographical division between the northern and southern groupings differs for each clade (Figure 3.17). This may indicate the presence of once-separate northern and southern species which are now undergoing hybridization. The northern clades of *E. lynceum* in the cytochrome *b* tree are more closely related to *E. zonale* than to southern *E. lynceum*; however, monophyly of *E. lynceum* is not rejected for cytochrome *b*. If this paraphyly is real, it may indicate introgression of a population of *E. zonale* into *E. lynceum*, although the source of such a population would appear to be extinct. Further examination of the populations of *E. lynceum* at additional loci is required before the phylogeographic history of these populations, and how many species are represented, can be assessed.

Conclusion

The *E. zonale* species group shows a high degree of geographical differentiation; twenty-one allopatric clades are identifiable in the cytochrome *b* gene tree, and several of these could be further subdivided into even smaller allopatric clades. There is no support for reconstructing a branching order for most of the main clades, however, which cripples the utility of this data in comparative biogeographic analyses.

Nuclear loci show substantially less biogeographical differentiation, and none of the loci by itself is useful in a biogeographical analysis. Taken together in a species tree analysis, however, these loci do recover strong support for part of the backbone of the phylogeny. Compared with that of other taxa, there is considerable correspondence, suggesting a vicariant history; however, evidence of population growth in some of these populations suggests that co-ordinated dispersal may also have played a role.

The South Fork Holston race of *E. zonale* is shown to have undergone extensive introgression with the Upper Tennessee race, which appears to be in the process of replacing it. The disjunct distribution of the South Fork Holston race suggests that such a process is ongoing, as do extended Bayesian skyline plots for each of the two taxa.

Based on the evidence of the various nuclear loci, there may be as many as nine species within the *E. zonale* species group, rather than the currently recognized two. If this higher number does reflect reality, it seems clear that most of the species in the group are in the early stages of differentiation.

Sampling of additional loci may serve to provide support for those relationships within the *E. zonale* species group which are still poorly supported; it may also help to clarify the number of species which exists in the group. Sampling of additional loci and additional populations is particularly necessary for the Green River clade of *E. zonale*, and in the central portion of the range of *E. lynceum*.

Table 3.1 Names and chromosomal locations of nuclear introns developed for use in this study.

Locus name	Abbreviation	Intron Number	Chromosome Number
eukaryotic translation initiation factor 3, subunit c	eif3c	13	12
nicalin	ncl1	8	22
neural precursor cell expressed, developmentally down-regulated 4-like	nedd4l	24	21
poly(A) binding protein, cytoplasmic 4 (inducible form)	pabpc4	8	17
solute carrier organic anion transporter family, member 4A1	slco4a1	11	23
syntaxin 5A	stx5a	10	14

Table 3.2 Primers for nuclear loci, with annealing temperatures for each primer pair.

Locus	Primer name	Primer sequence	Annealing temperature (ºC)
eif3c 13th intron	eif3cf13	5′ CGCAGRATCATGCACACMTA 3′	57.0
	eif3cr13	5′ GTARATGTGGCASAGGATGG 3′	
ncl1 8th intron	ncl1f8	5′ CCAGTCTGCTSCAGGACAAY 3′	50.9
	ncl1r8	5′ SGCCAGGTTGATYTTCTTRT 3′	
nedd4l 24th intron	nedd4l24f	5′ AGATGATGCTGGGMAAACAGA 3′	53.0
	nedd4l24r	5′ TGTCCAAAGTTGTCCTCRTCA 3′	
nedd4l 24th intron (internal primers)	nedd4l24if	5′ TTGAGTAGTCTTAATCCCTGACCT 3′	
	nedd4l24ir	5′ TGCAAAATTCCTCCTTAGCTG 3′	
pabpc4 8th intron	pabpc4f8	5′ GARGAGCGCAARGCYCAYCT 3′	53.0
	pabpc4r8	5′ TSASCTGRTTKGGTGCATAG 3′	
pabpc4 8th intron (sequencing primer)	pabpc4fs8	5′ GAGGAGCGCAAGGCCCACCT 3′	
slco4a1 11th intron	slco4a1f11	5′ GCCTTCCTCRCSTTCCTCTT 3′	54.3
	slco4a1r11	5′ ATCACRGAGCCGAAYGCTAT 3′	
stx5a 10th intron	stx5a10f	5′ GCAGAACATYGARAGCACMA 3′	52.7
	stx5a10r	5′ GGAGGAGACKGACTGGAAGT 3′	

Table 3.3 Variation in length of intron loci. All figures are lengths in bp.

Locus	Ingroup			Outgroup	
	Minimum	Mode	Maximum	Minimum	Maximum
eif3c	405	417	418	414	418
ncl1	562	585	587	526	609
nedd4l	432	453	456	451	453
pabpc4	399	402	410	407	409
s7	524	526	531	524	531
slco4a1	792	900	1140	892	905
stx5a	692	705	715	573*	717

*the shortest sequence, from *E. barrenense*, was not fully sequenced due to a repeating AT element in the middle of the intron. 573 bp is the length of sequence actually obtained.

Table 3.4 Number of polymorphic and parsimony informative sites per locus for the *E. zonale* species group, and for all sequences including outgroups. Numbers are also given for the 1st, 2nd, and 3rd positions of cytochrome *b*. Numbers of gap characters encoded for nuclear loci are also indicated.

	With outgroups			Ingroup only				
Locus	Poly-morphic	Parsimony informative	Number of sequences	Poly-morphic	Parsimony informative	Number of sequences	Alignment length	Gap Characters
cyt *b*	281	241	339	184	161	324	1140	NA
cyt *b* 1140 bp only*	428	375	209	268	240	194	1140	NA
1st pos	59	39	209	31	24	194	—	—
2nd pos	15	2	209	4	0	194	—	—
3rd pos	354	334	209	233	216	194	—	—
eif3c	70	58	162	49	35	156	417	16
ncl1	83	67	164	61	45	150	585	15
nedd4l	82	73	252	57	44	244	346	6
pabpc4	83	67	164	47	30	150	402	18
s7	128	100	202	82	55	188	394	12
slco4a1	176	142	156	101	72	142	892	33
stx5a	114	86	252	85	55	240	705	19

Sites with gaps are not included in the counts. *“Cyt *b* 1140 bp only” refers to the cytochrome *b* alignment from which all sequences 739 bp long have been removed, leaving only full 1140 bp sequences. Counts for codon positions in cytochrome *b* are given only for the full-length sequences. Alignment length is the number of characters in the matrix for likelihood analysis for each locus, after gaps found in more than half of all taxa have been removed.

Table 3.5 Models of evolution for Garli, MrBayes, and *BEAST analyses, and molecular clock test results for *BEAST analyses for each locus.

Locus	Model for Garli	Model for Bayesian analyses	Clock rejected?
cyt *b* 1st position	TrNef+I+G	SYM+I+G	—
cyt b 2nd position	TPM2uf	GTR	—
cyt *b* 3rd position	GTR+G	GTR+I+G	—
cyt *b* unpartitioned	GTR+I+G	GTR+I+G	N
eif3c	HKY+G	HKY+G	N
ncl1	TrN+G	HKY+G	N
nedd4l	HKY+G	HKY+G	N
pabpc4	HKY+G	HKY+G	N
s7	TrN+I+G	HKY+I+G	Y
slco4a1	GTR+I+G	GTR+I+G	Y
stx5a	HKY+G	HKY+G	Y

Table 3.6 Results of the R2 test for population growth for all cytochrome *b* clades in the *E. zonale* species group. Significant results (indicating population growth) are shown in bold. "NA" indicates clades with no variability.

	Homochitto	Pascagoula	Amite	Tangipahoa	Pearl	Yazoo	Obion
R2	0.2182	NA	0.2	0.1668	0.1348	0.2018	0.2778
p	0.87827	NA	0.33019	0.42095	0.11145	0.88426	0.95366
n	11	10	10	10	11	17	10

	Hatchie	Big Black	Green	Upper Tennessee	South Fork Holston	Hiwassee	Duck
R2	0.1436	0.3	0.2086	0.0652	0.1191	0.1849	0.1493
p	0.06435	0.67602	0.52005	0.063	0.12994	0.40642	0.35807
n	10	10	10	44	16	10	12

	Cumberland	Kentucky	Upper Ohio	Upper Mississippi	White	Little Red	Ouachita and Upper Arkansas	Ouachita only
R2	0.1993	0.1844	**0.0745**	**0.0597**	0.0776	**0.1342**	0.0837	**0.1172**
p	0.9003	0.37222	**<0.00001**	**0.00203**	0.056	**0.03003**	0.119	**0.01148**
n	23	10	16	32	26	10	29	11

"Upper Tennessee" includes all individuals from the Upper Tennessee cytochrome *b* clade, including those which belong to the South Fork Holston nuclear gene clade. "South Fork Holston" includes those individuals from the Upper Tennessee cytochrome *b* clade belonging to the South Fork Holston nuclear gene clade. "Ouachita only" is the subset of individuals from the Ouachita and Upper Arkansas clade found in the Ouachita River drainage. The Little Tennessee clade is not included because it is too poorly sampled (n=3) for the R2 test.

Appendix 3.1 Specimens Examined. The specimens used in this study are listed with the number of specimens from each site in parentheses and collection locality. All specimens are uncatalogued specimens in the Bell Museum of Natural History frozen tissue collection, unless otherwise indicated. Institutional abbreviations are as follows: MMNS = Mississippi Museum of Natural Science PLU= Pacific Lutheran University Natural History Collection SLU = Saint Louis University Ichthyological Collection UAIC = University of Alabama Ichthyological Collection YFTC = Yale Fish Tissue Collection.

Etheostoma lynceum

Homochitto clade

HomochittoA (5) Buffalo River, Wilkinson Co., MS, PLU uncat.
HomochittoB (5) Homochitto River, Jefferson Co., MS.
HomochittoC (1) McCall Creek, Lincoln Co., MS, MMNS 614.

Pascagoula clade

PascagoulaA (5) Bowie Creek, Covington Co., MS.
PascagoulaB (5) Chunky River, Clarke Co., MS.

Amite clade

Amite (2) East Fork Amite River, Amite Co., MS, PLU uncat.
Amite (8) East Fork Amite River, Amite Co., MS.

Tangipahoa clade

TangipahoaA (9) Bala Chitto Creek, Pike Co., MS.
TangipahoaB (1) Beaver Creek, Tangipahoa Co., LA, MMNS 320.

Pearl clade

PearlA (3) McGee Creek, Walthall Co., MS.
PearlB (2) Strong River, Simpson Co., MS, MMNS 622.
PearlC (3) Wolf River, Harrison Co., MS.
PearlD (3) Upper Little River, Marion Co., MS.

Big Black clade

Big BlackA (1) Big Black River, Montgomery Co., MS, MMNS 894.
Big BlackB (3) Bayou Pierre, Copiah Co., MS, UAIC uncat.
Big BlackC (3) Little Bayou Pierre, Claiborne Co., MS, UAIC uncat.
Big BlackD (1) Long Creek, Attala Co., MS, PLU uncat.
Big BlackE (2) Little Zilpha Creek, Attala Co., MS, PLU uncat.

Yazoo clade

YazooA (6) Puskus Creek, LaFayette Co., MS.
YazooA (1) Puskus Creek, LaFayette Co., MS, PLU uncat.
YazooB (4) Pumkin Creek, LaFayette Co., MS.
YazooC (6) unnamed creek, Yalobusha River drainage, Yalobusha Co., MS.

Hatchie clade

HatchieA (5) Spring Creek, Hardeman Co., TN, PLU uncat.
HatchieB (5) Wolf River, Fayette Co., TN.

Obion clade

Obion (4) Clear Creek, Henry Co., TN.

Obion (4) Clear Creek, Henry Co., TN, PLU uncat.

Etheostoma zonale

South Fork Holston clade

South Fork HolstonA (3) Clinch River, Claiborne Co., TN, UAIC uncat.
South Fork HolstonB (2) Clinch River, Scott Co., VA, UAIC uncat.
South Fork HolstonC (1) Clinch River, Hancock Co., TN, SLU uncat.
South Fork HolstonD (5) Elk River, Carter Co., TN, UAIC uncat.
South Fork HolstonE (5) Middle Fork Holston River, Washington Co., VA, UAIC uncat.

Upper Tennessee clade

Upper TennesseeA (2) French Broad River, Cooke Co., TN, YFTC 7119-7120.
Upper TennesseeB (2) Holston River, Hawkins Co., TN, UAIC uncat.
Upper TennesseeC (2) Indian River, Claiborne Co., TN, SLU uncat.
Upper TennesseeD (4) Little River, Blount Co., TN.
Upper TennesseeE (3) North Fork Holston River, Washington Co., VA, UAIC uncat.
Upper TennesseeF (4) Nolichucky River, Washington Co., TN.
Upper TennesseeG (5) Richland Creek, Giles Co., TN.
Upper TennesseeH (3) Sequatchie River, Marlon Co., TN.
Upper TennesseeI (3) Toccoa River, Fannin Co., GA, YFTC 6978-6980.

Hiwassee clade

Hiwassee (10) Valley River, Cherokee Co., NC.

Little Tennessee clade

Little Tennessee (3) Citico Creek, Monroe Co., TN.

Duck clade

DuckA (7) Buffalo River, Perry/Wayne Cos., TN.
DuckB (3) Duck River, Bedford Co., TN.
DuckC (2) Indian Creek, Hardin Co., TN, YFTC 7611-7612.

Cumberland clade

CumberlandA (2) Big South Fork, Scott Co., TN, YFTC 7086-7087.
CumberlandB (4) East Fork Stones River, Rutherford Co., TN, SLU uncat.
CumberlandC (10) Rockcastle River, Rockcastle/Laurel Cos., KY
CumberlandD (1) Smith Fork, Smith Co., TN, YFTC 5789.
CumberlandE (6) Turnbull Creek, Cheatham Co., TN.

Green clade

GreenA (5) Drakes Creek, Warren Co., KY.
GreenB (5) Green River, Green Co., KY.

Kentucky clade

KentuckyA (5) South Fork Kentucky River, Owsley Co., KY.
KentuckyB (5) Red Bird River, Clay Co., KY.

Upper Ohio clade

Upper OhioA (2) Big Darby Creek, Pickaway Co., OH.
Upper OhioB (2) Broken Straw Creek, Warren Co., PA, YFTC 11272-11273.
Upper OhioC (1) French Creek, Crawford Co., PA, YFTC 11284.
Upper OhioD (5) Licking River, Bath/Rowan Cos., KY.

Upper OhioE (2) Walhonding River, Coshocton Co., OH.

Upper OhioF (4) Dry Fork and Whitewater River, Hamilton Co., OH.

Upper Mississippi clade

Upper MississippiA (3) Big Piney River, Texas Co., MO.

Upper MississippiB (1) Ferson-Otter Creek, Kane Co., IL, YFTC 1363.

Upper MississippiC (1) Hickory Creek, Bureau Co., IL, YFTC 1661.

Upper MississippiD (3) Huzzah Creek, Crawford Co., MO.

Upper MississippiE (6) Otter Creek, La Salle Co., IL.

Upper MississippiF (1) West Fork Mazon River, Grundy Co., IL, YFTC 990.

Upper MississippiG (5) Pomme de Terre River, Polk Co., MO.

Upper MississippiH (1) County Ditch No. 1, Whiteside Co., IL, YFTC 1655.

Upper MississippiI (5) Root River, Olmsted Co., MN.

Upper MississippiJ (3) Upper Iowa River, Fillmore Co., MN.

Upper MississippiK (1) Black River, Jackson Co., WI, YFTC 1178.

Upper MississippiL (2) Yellow Medicine River, Yellow Medicine Co., MN.

White clade

WhiteA (2) Crooked Creek, Marion Co., AR.

WhiteB (2) Eleven Point River, Randolph Co., AR.

WhiteC (5) James River, Stone Co., MO.

WhiteD (5) Kings River, Carroll Co., AR.

WhiteE (4) St. Francis River, Madison Co., MO.

WhiteF (3) Spring River, Fulton Co., AR.

WhiteG (5) Strawberry River, Sharp Co., AR.

Little Red clade

Little RedA (5) Archey Fork Little Red River, Van Buren Co., AR.

Little RedB (5) Middle Fork Little Red River, Van Buren Co., AR.

Ouachita and Upper Arkansas clade

Ouachita and Upper ArkansasA (7) Caddo River, Montgomery Co., AR.
Ouachita and Upper ArkansasB (3) Caddo River, Pike Co., AR.
Ouachita and Upper ArkansasC (5) Elk River, McDonald Co., MO.
Ouachita and Upper ArkansasD (3) South Fork Fourche La Fave River, Perry Co., AR, YFTC 1685-1687.
Ouachita and Upper ArkansasE (2) Mulberry River, Johnson Co., AR, SLU 840.03.
Ouachita and Upper ArkansasF (1) Ouachita River, Montgomery Co, AR.
Ouachita and Upper ArkansasG (3) Saline River, Grant Co., AR.
Ouachita and Upper ArkansasH (5) Spring River, Jasper Co., MO.

Etheostoma baileyi (1) Red Bird River, Clay Co., KY.

Etheostoma barrenense (1) Drakes Creek, Warren Co., KY.

Etheostoma blennioides (1) Valley River, Cherokee Co., NC.

Etheostoma rupestre (1) Cahaba River, Bibb CO., AL.

Etheostoma simoterum (1) Otter Creek, Wayne Co., KY.

Etheostoma thalassinum (1) Mountain Creek, Greenville Co., SC.

Etheostoma variatum (1) Walhonding River, Coshocton Co., OH.

Cytochrome *b* outgroup sequences:

Etheostoma atripinne (AF288444), *E. baileyi* (AF288423), *E. barrenense* (AF288424), *E. blennioides* (AF288426), *E. blennius* (AF288427), *E. inscriptum* (AF288435), *E.oophylax* (AF412524), *E. podostemone* (AF045346), *E. rafinesquei* (AF288439), *E. raneyi* (AF288441), *E. rupestre* (AF288442), *E. simoterum* (AF288445), *E. swannanoa* (AF288446), *E. thalassinum* (AF288448), *E. variatum* (AF289266).

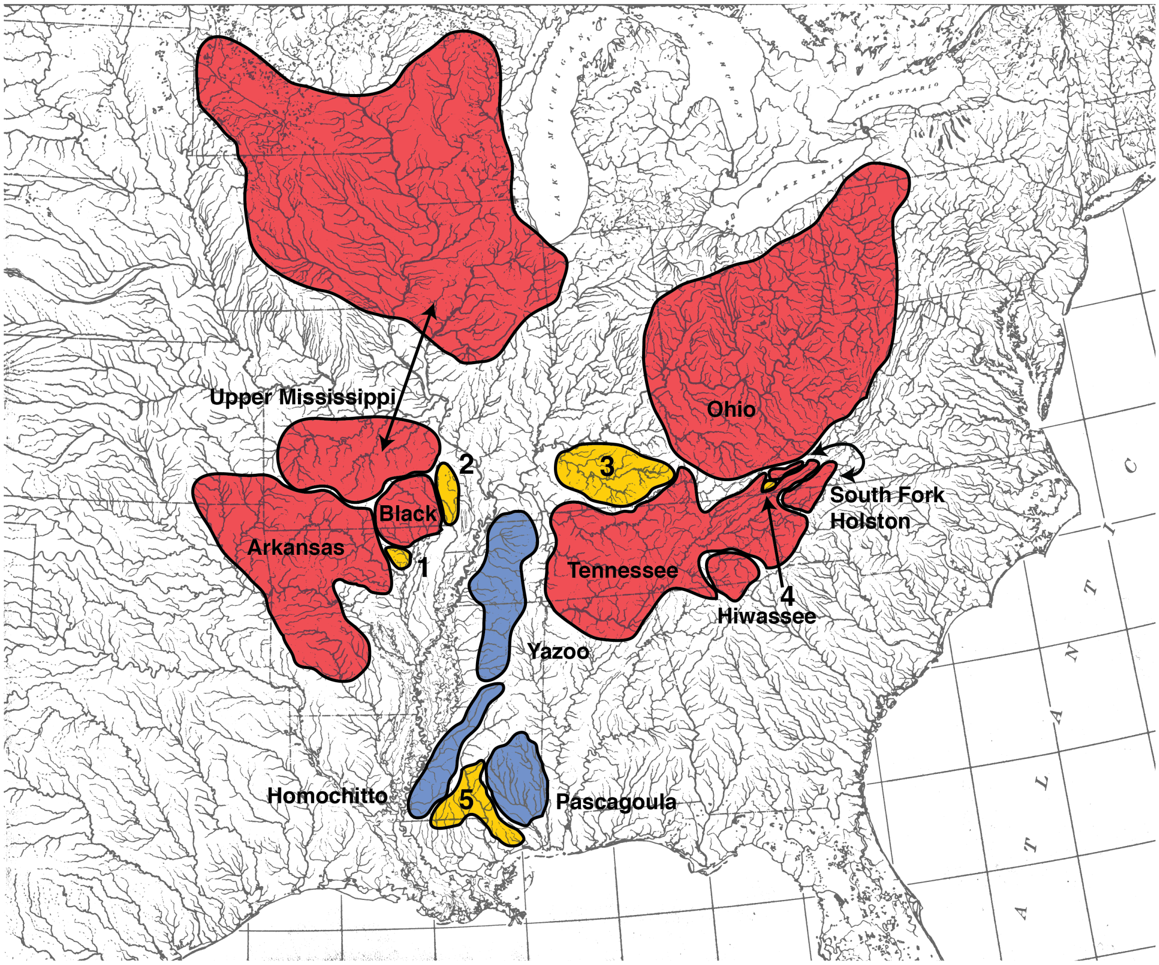


Figure 3. 1 Morphological races of *Etheostoma zonale* and *E. lynceum* according to Tsai and Raney (1974). Red indicates races of *E. zonale*, blue, races of *E. lynceum*, and yellow, intergrades between races, numbered as follows: 1, 2: Black River Race × Arkansas River Race, 3: Black River Race × Ohio River Race, 4: South Fork Holston River Race × Tennessee River Race, 5: Homochitto River Race × Pascagoula River Race.

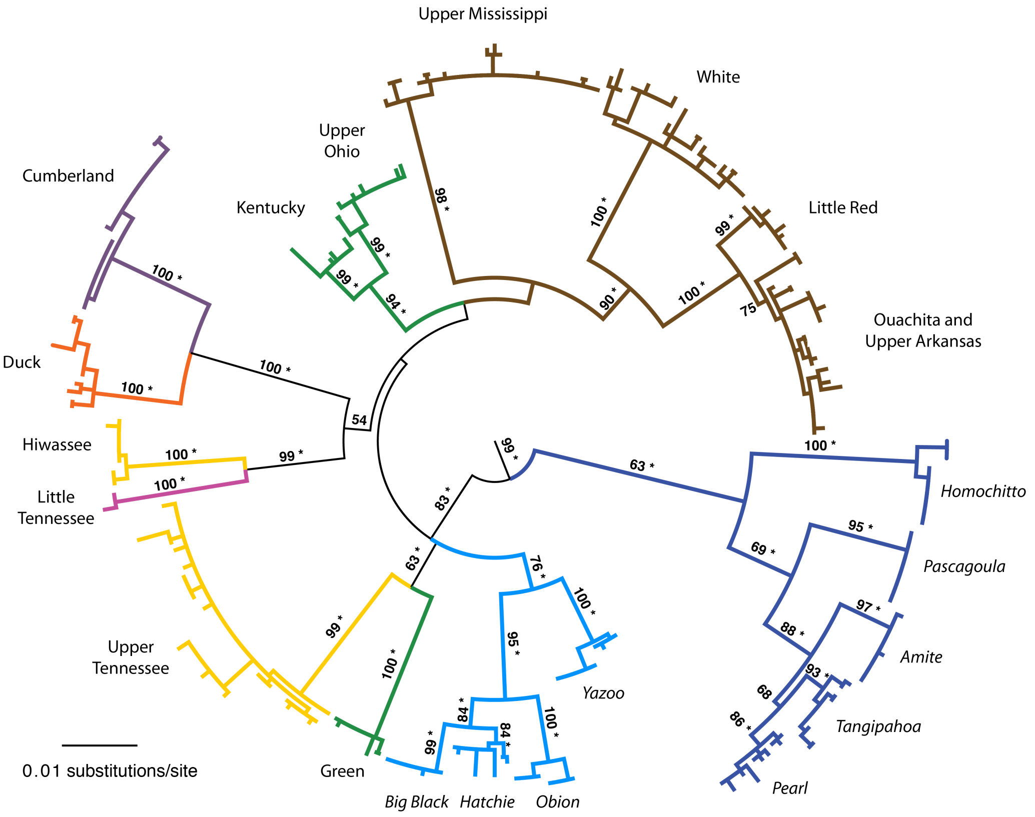


Figure 3.2 Cytochrome *b* maximum likelihood gene tree for the *Etheostoma zonale* species group. Clades are named after the drainage system they are found in. Colors represent putative species according to the multi-locus analysis. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Populations of *E. zonale* are indicated in plain text, and those of *E. lynceum* in italics. Support values of branches within each named clade are not shown. Outgroups used in analysis not shown. Tree ln likelihood: -9295.778696.

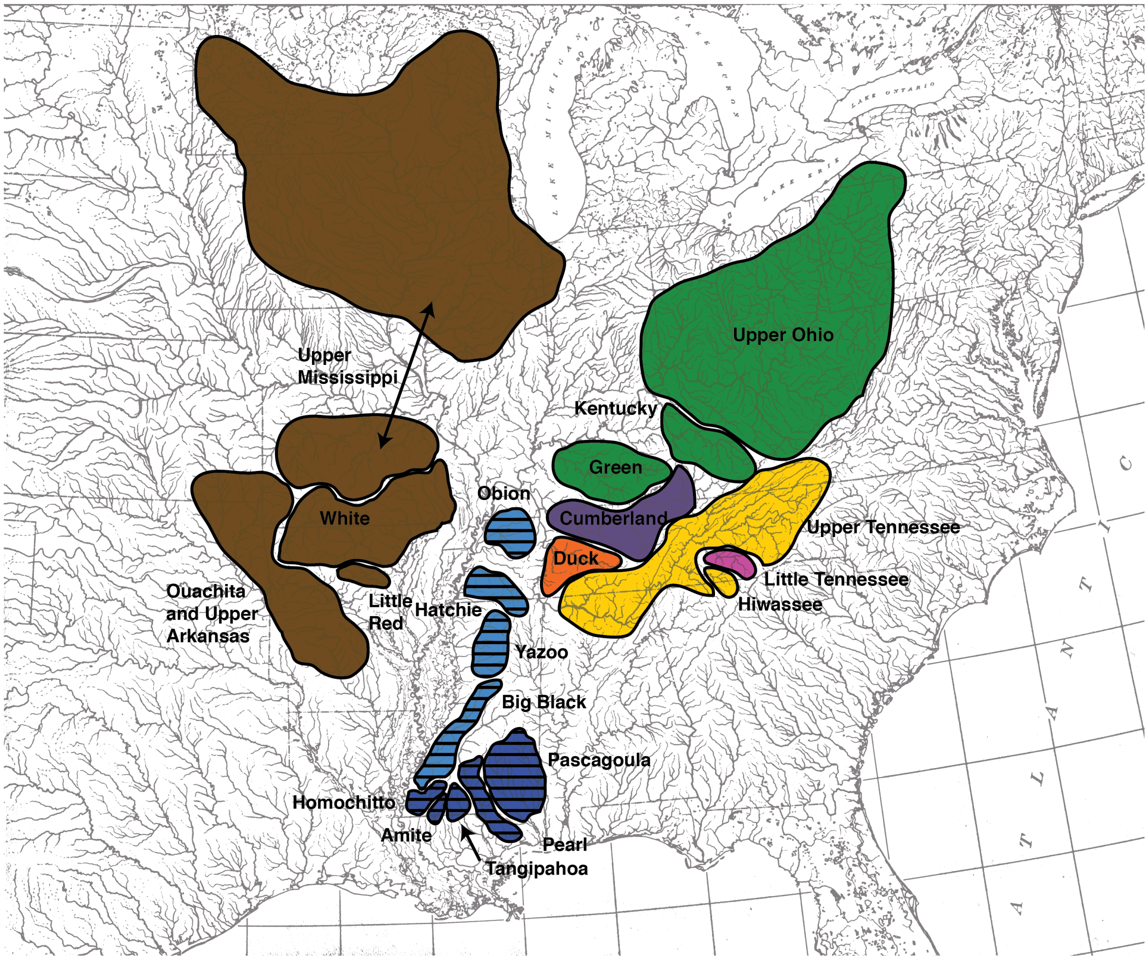


Figure 3.3 Distribution of each named cytochrome *b* clade in the *Etheostoma zonale* species group. Colors correspond to colors of branches in figure 3.2. Areas with solid color are clades of *E. zonale*; those with hatching are clades of *E. lynceum*.

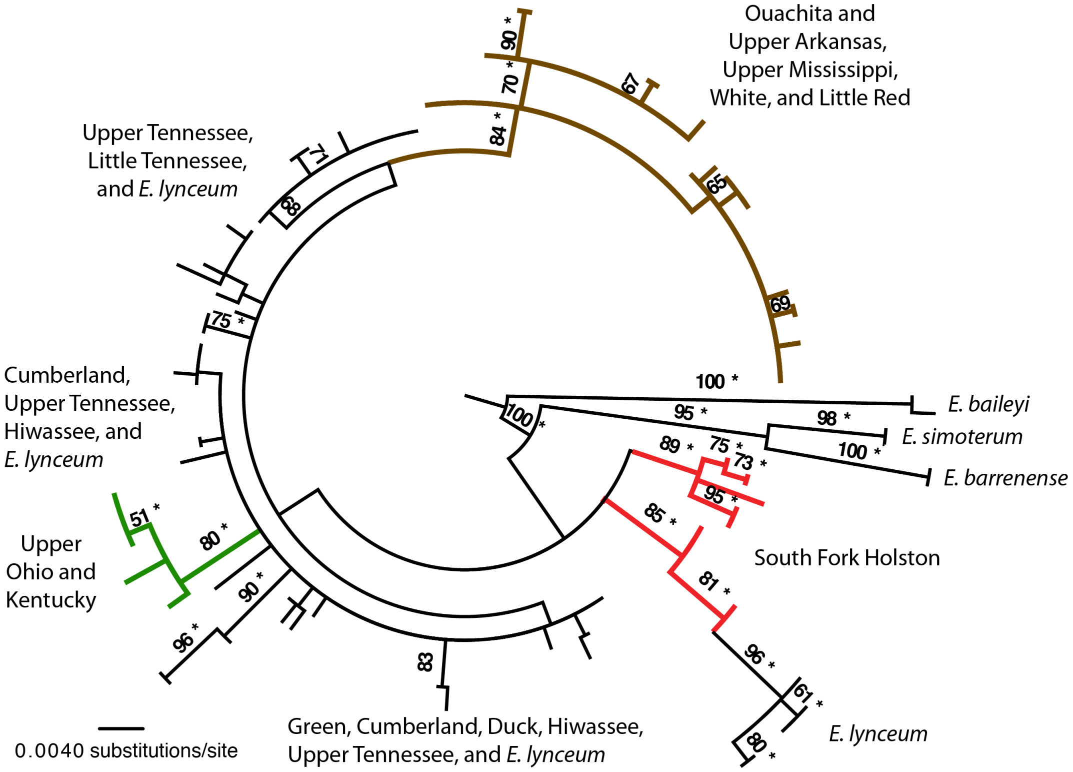


Figure 3.4 Eif3c maximum likelihood gene tree for the *Etheostoma zonale* species group. Colors correspond to map areas in figure 3.3. Uncolored branches do not show a clear geographical pattern. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -1482.645397.

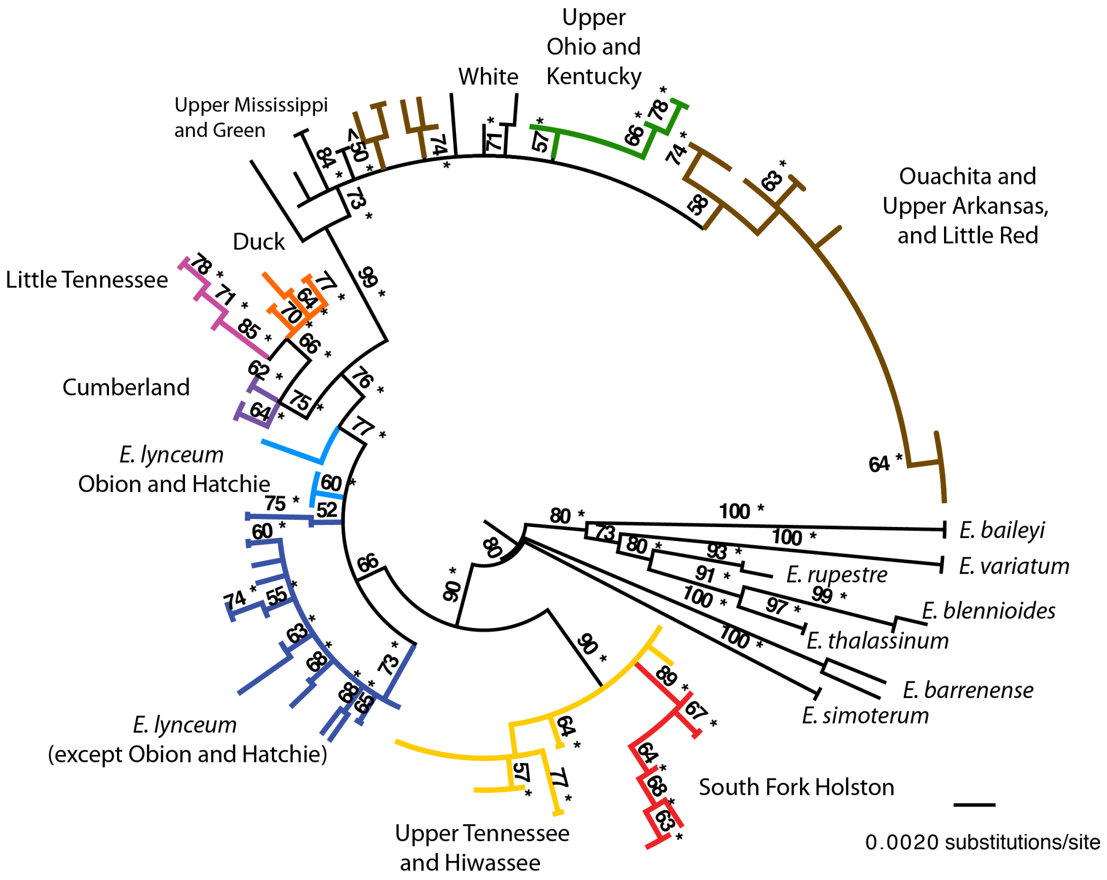


Figure 3.5 Ncl1 maximum likelihood gene tree for the *Etheostoma zonale* species group. Colors correspond to map areas in figure 3.3. Uncolored branches do not show a clear geographical pattern. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -1874.883824.

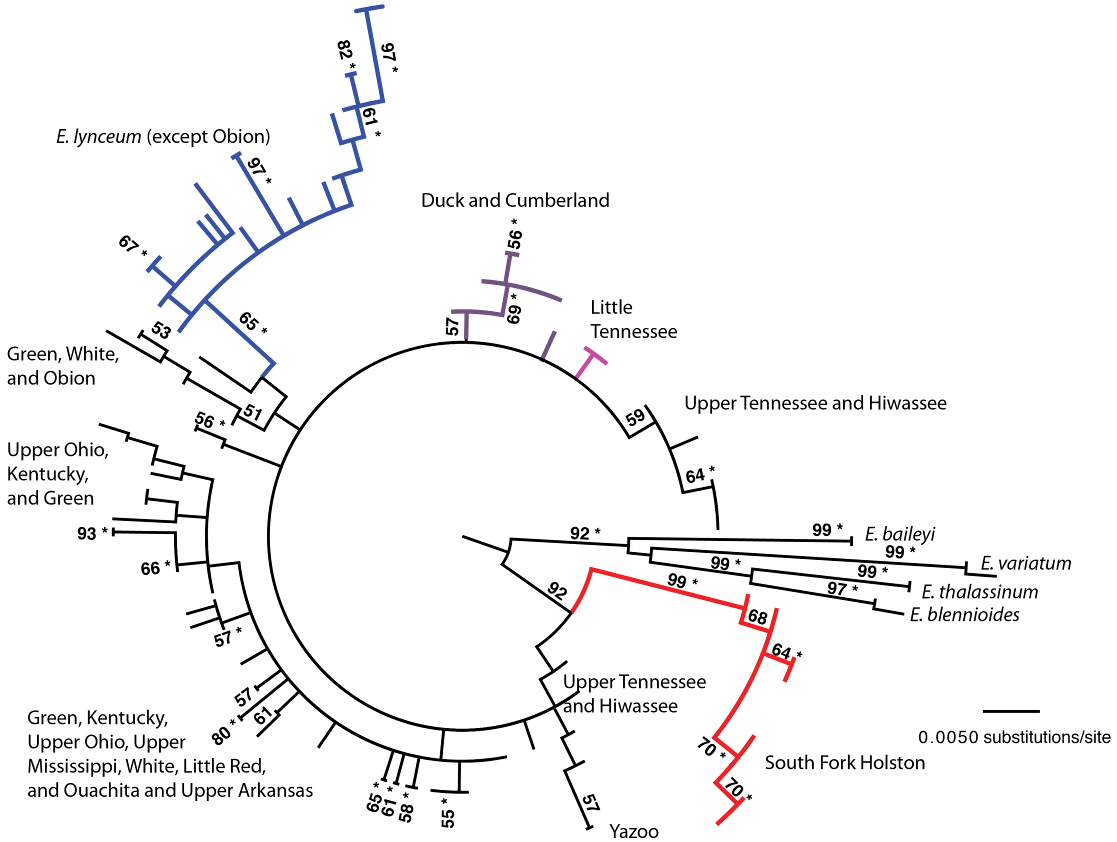


Figure 3.6 Nedd4l maximum likelihood gene tree for the *Etheostoma zonale* species group. Colors correspond to map areas in figure 3.3. Uncolored branches do not show a clear geographical pattern. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -1364.569092.

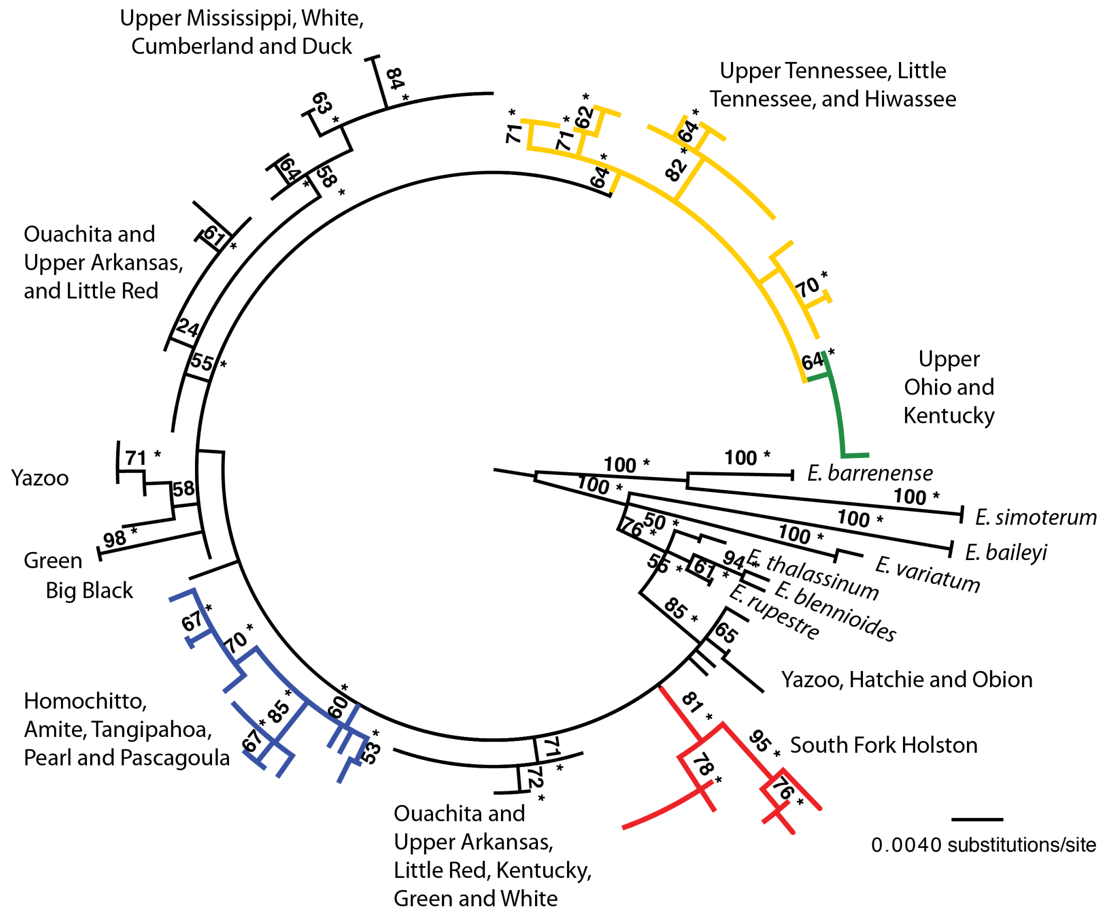


Figure 3.7 Pabpc4 maximum likelihood gene tree for the *Etheostoma zonale* species group. Colors correspond to map areas in figure 3.3. Uncolored branches do not show a clear geographical pattern. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -1386.421292.

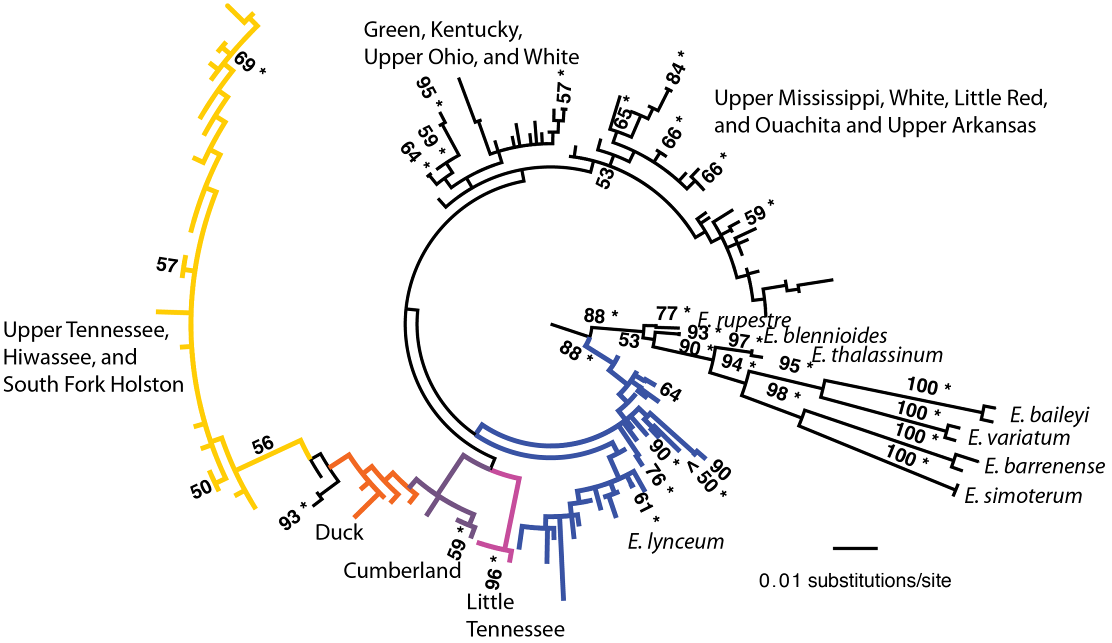


Figure 3.8 S7 maximum likelihood gene tree for the *Etheostoma zonale* species group. Colors correspond to map areas in figure 3.3. Uncolored branches do not show a clear geographical pattern. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -2371.420745.

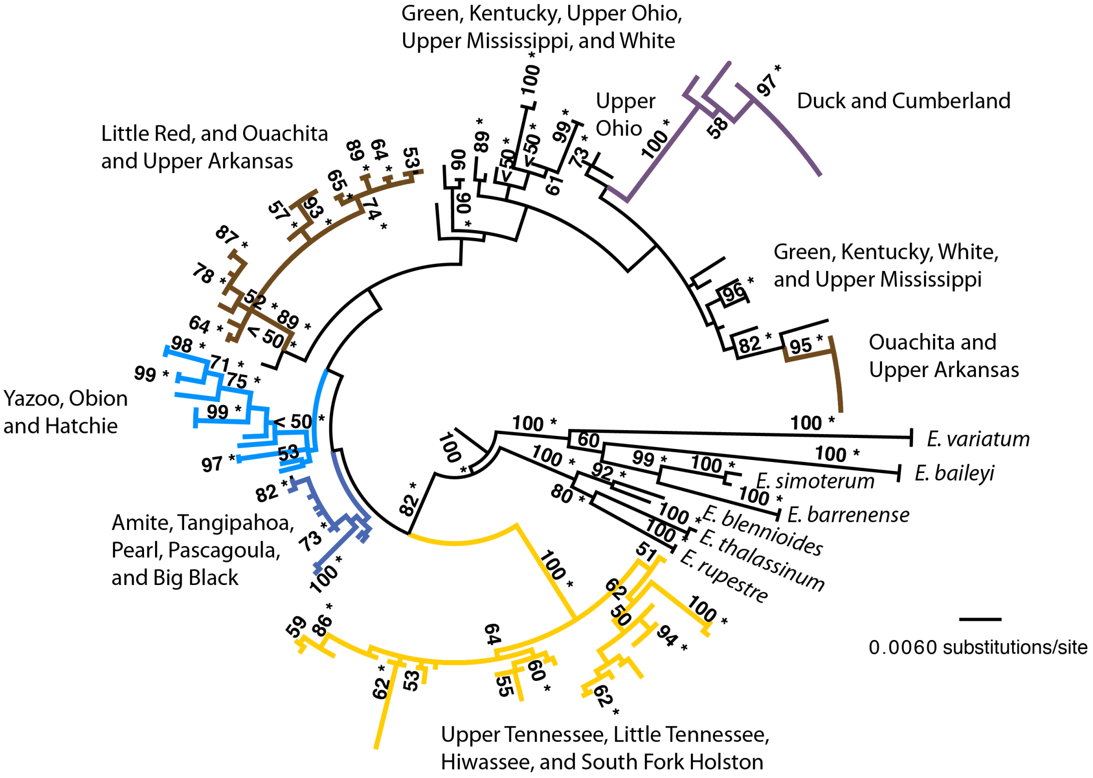


Figure 3.9 Slco4a1 maximum likelihood gene tree for the *Etheostoma zonale* species group. Colors correspond to map areas in figure 3.3. Uncolored branches do not show a clear geographical pattern. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -4516.146314.

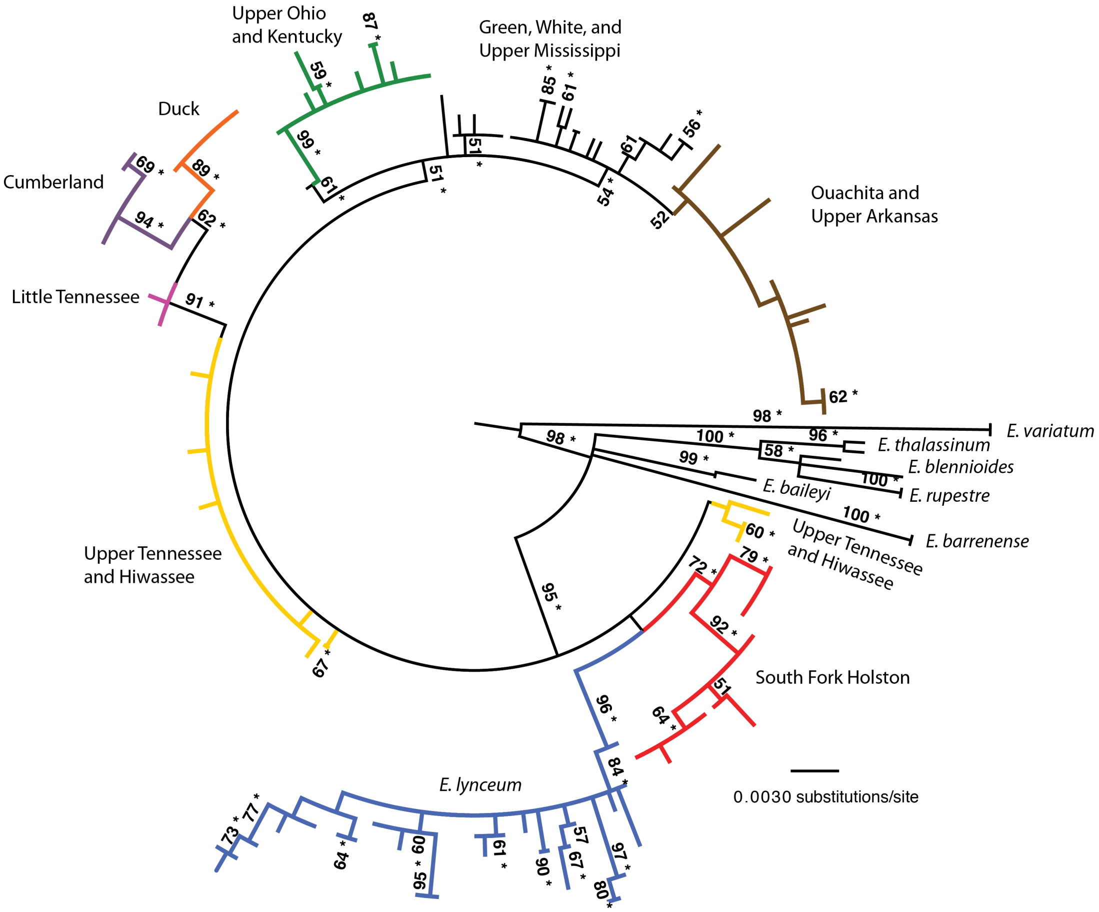


Figure 3.10 Stx5a maximum likelihood gene tree for the *Etheostoma zonale* species group. Colors correspond to map areas in figure 3.3. Uncolored branches do not show a clear geographical pattern. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -2321.594447.

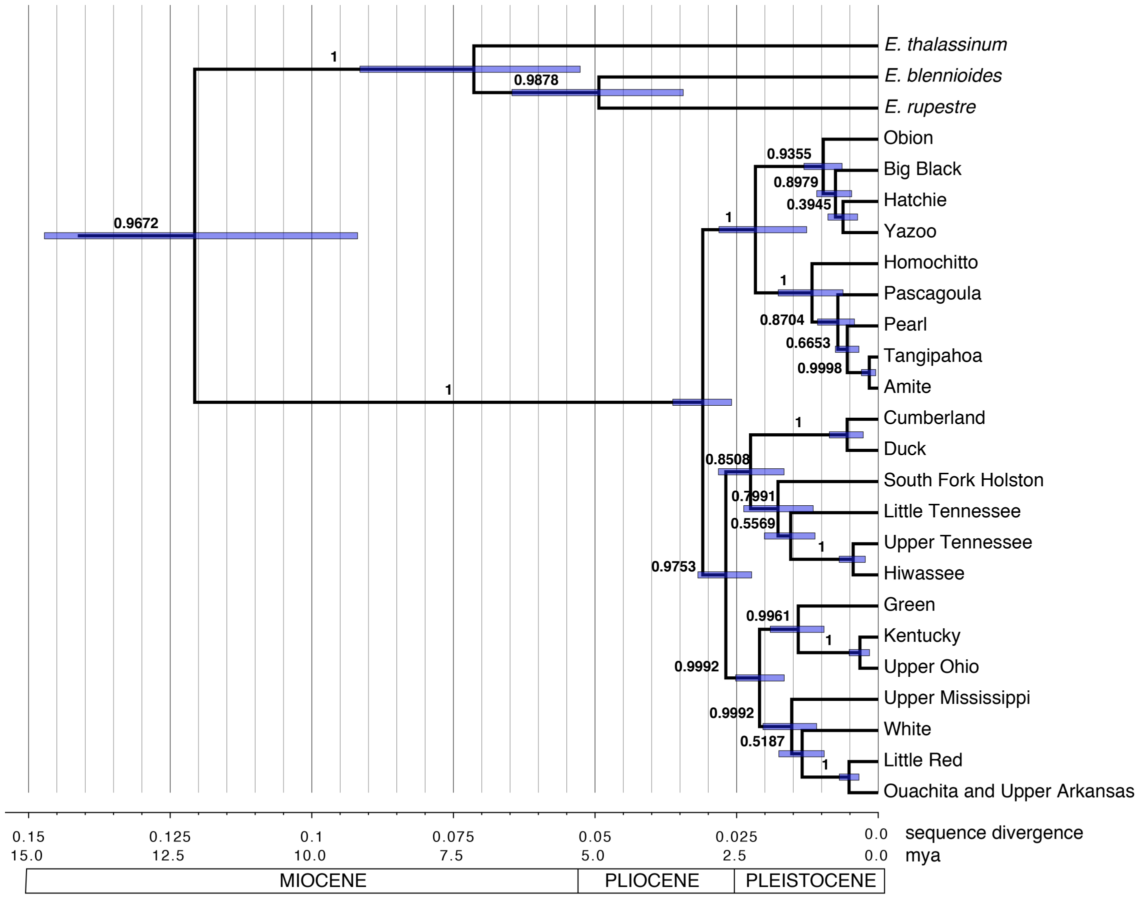


Figure 3.11 Species tree for *Etheostoma zonale* species group generated using *BEAST. Analysis performed with cytochrome *b* and seven nuclear loci. Numbers above branches are posterior probabilities for nodes. Confidence intervals at nodes are 95% highest posterior densities. Scale is in percent sequence divergence and millions of years before the present, converted using a rate of 0.01% per lineage per million years. Additional outgroup taxa have been removed from the tree.

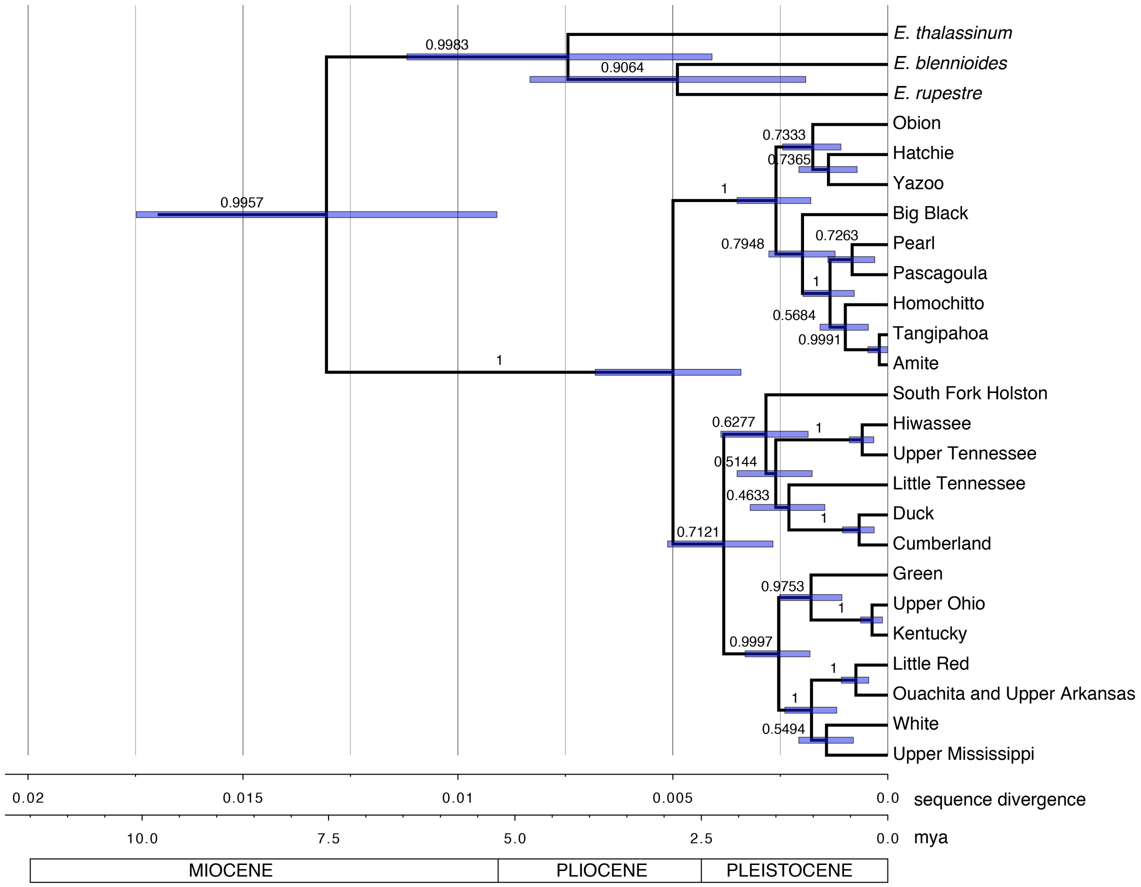
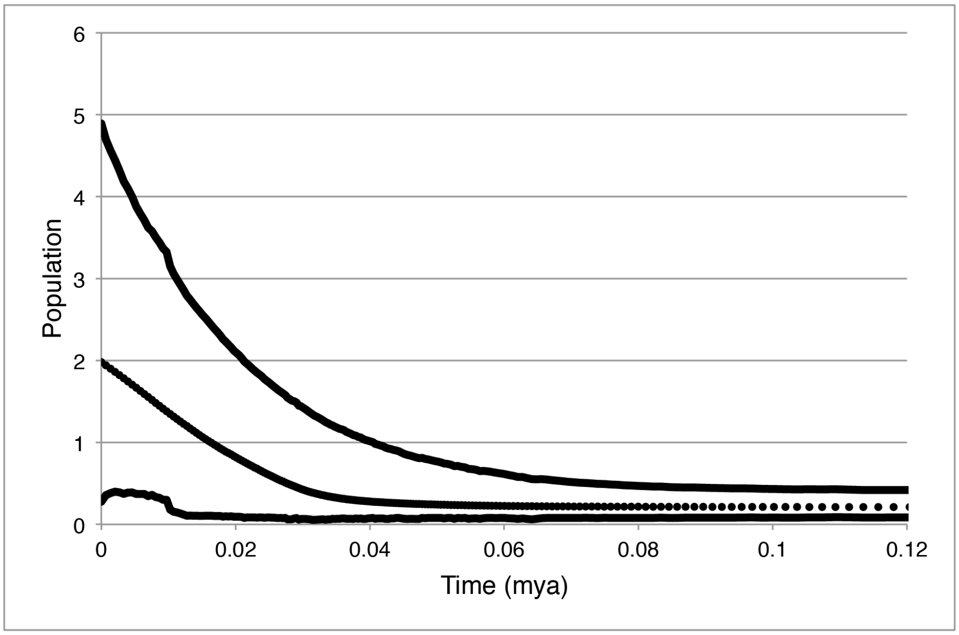


Figure 3.12 Species tree for *Etheostoma zonale* species group generated using *BEAST. Analysis performed with seven nuclear loci only. Numbers above branches are posterior probabilities for nodes. Confidence intervals at nodes are 95% highest posterior densities. Scale is in percent sequence divergence and millions of years before the present, converted using a rate of 0.01% per lineage per million years. Additional outgroup taxa have been removed from the tree.



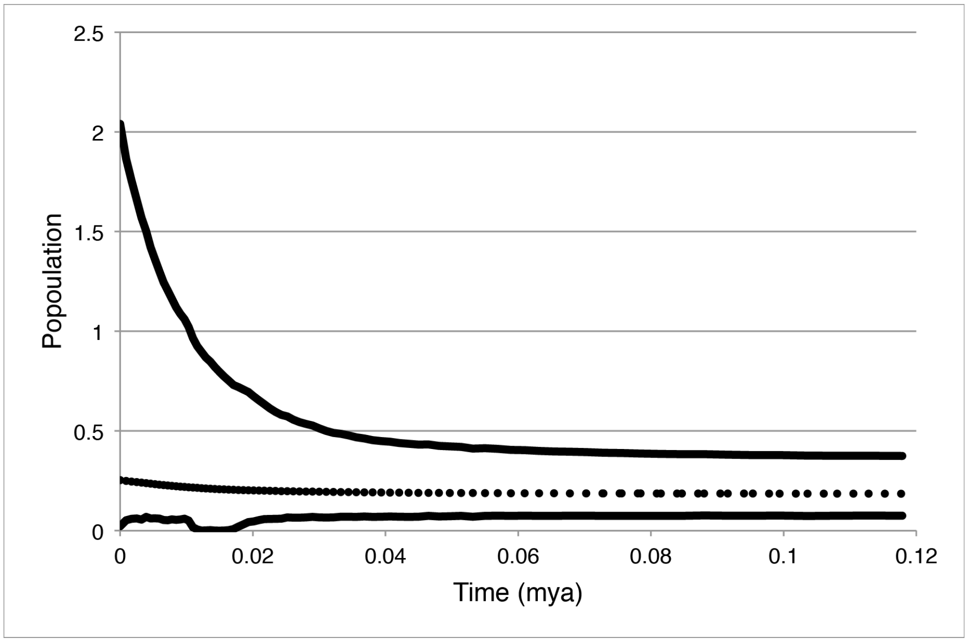


Figure 3.13 Extended Bayesian skyline plots for the Upper Tennessee (left) and South Fork Holston (right) clades. Solid lines are the 95% highest posterior densities for population size; the middle dotted line is the median population size. Population size is relative to generation time; for a generation time of one year, the scale is in millions of individuals. The time scale assumes a generation time of one year.

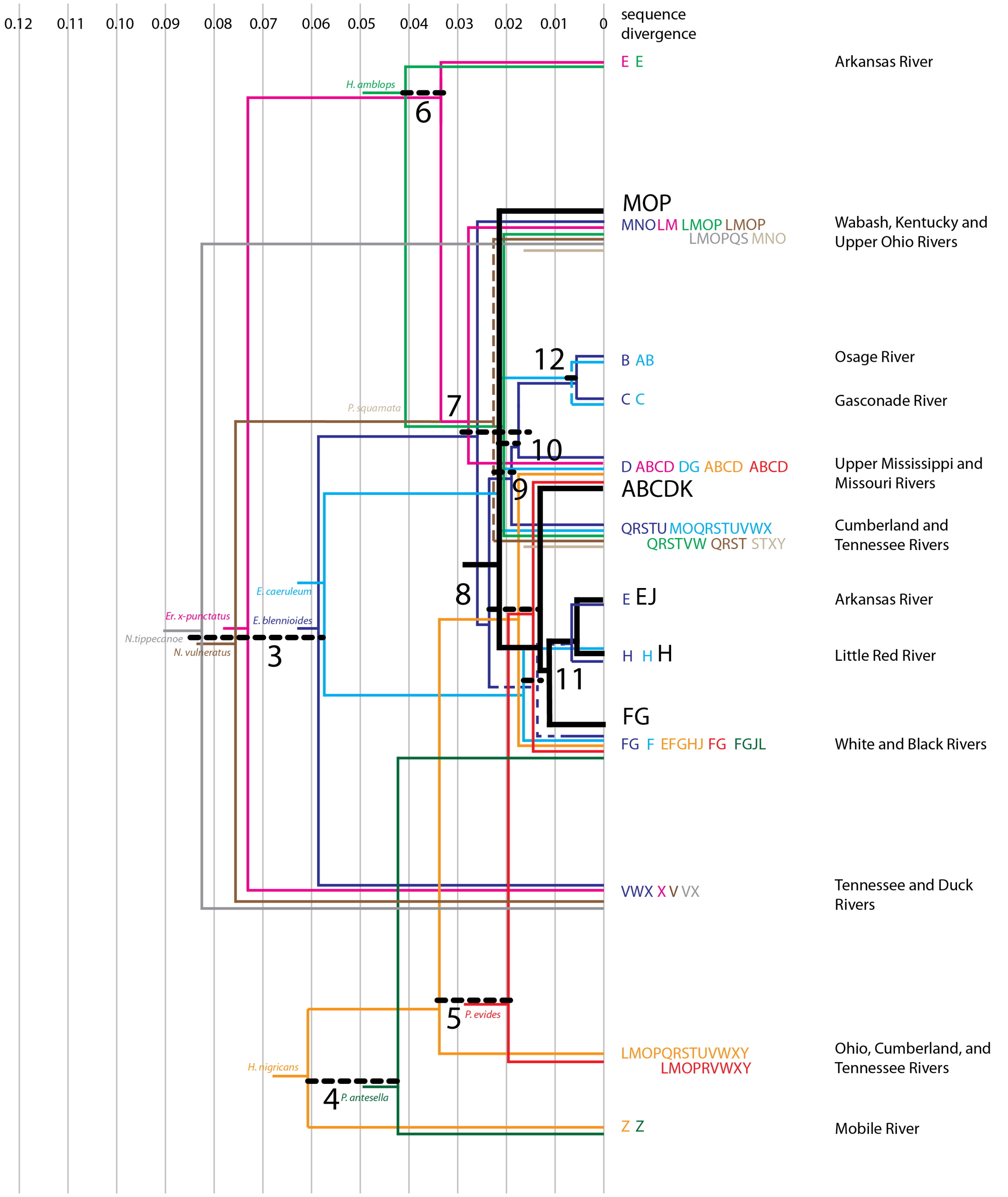


Figure 3.14 Correspondence of part of the *E. zonale* species tree, consisting of the northern and western clade, with the PACT analysis as presented in Chapter 2. Heavy solid black lines represent the newly-added *E. zonale* lineages; each clade included is indicated as follows by the heavy black letters giving its range: MOP – Upper Ohio, Kentucky, and Green River clades; ABCDK – Upper Mississippi clade; EJ – Ouachita and Upper Arkansas clade; H – Little Red clade; FG – White clade.

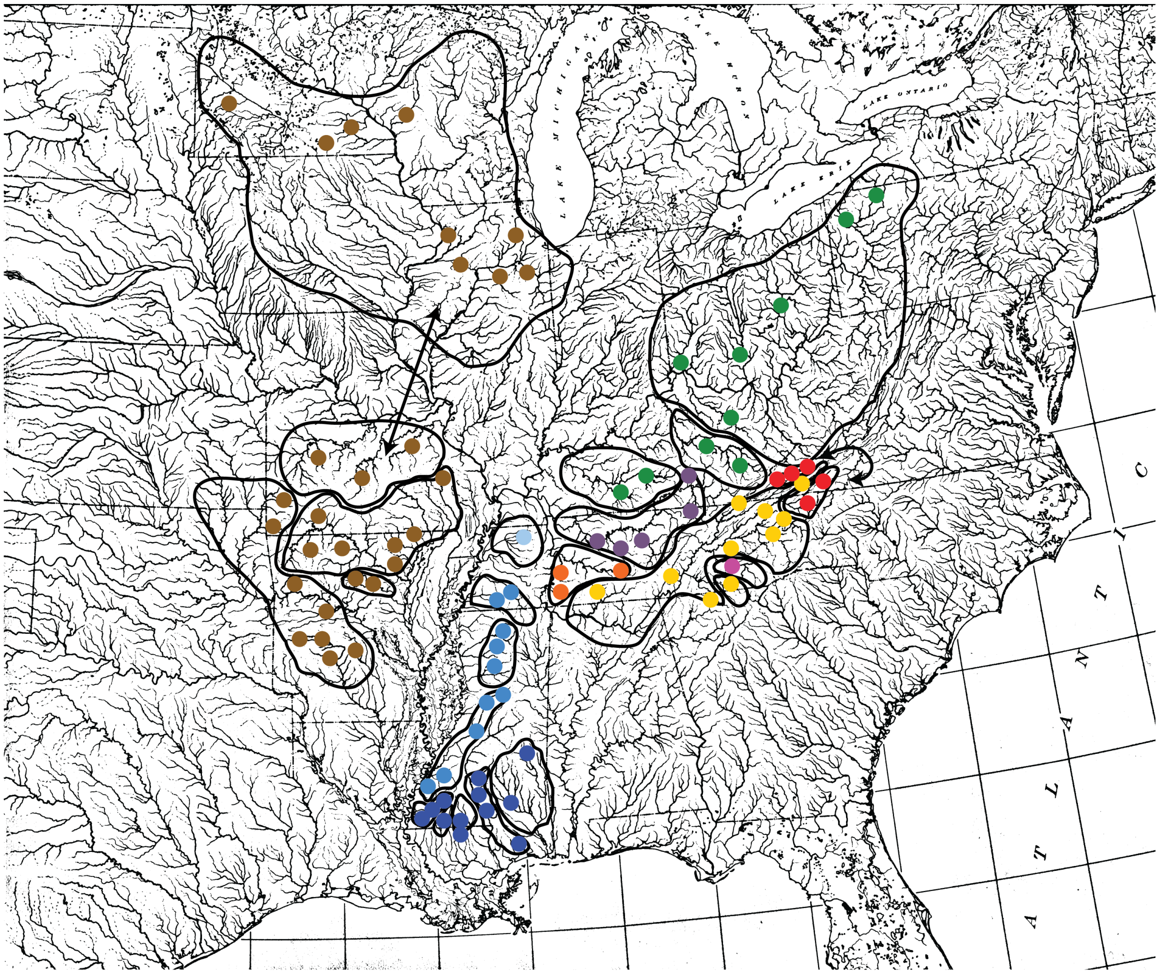


Figure 3.15 Sampling localities, and putative species delimitation of the *Etheostoma zonale* species group based upon species tree and invididual nuclear gene analyses. Black lines encircle the terminals used in the species tree analysis. Each dot represents a sampling locality, and each color represents a different species. Yellow: *E. zonale* sensu stricto; red: *Etheostoma* sp. South Fork Holston; light purple: *Etheostoma* sp. Little Tennessee; dark purple: *Etheostoma* sp. Cumberland; orange: *Etheostoma* sp. Duck; green: *Etheostoma* sp. Ohio; brown: *Etheostoma vinctipes*; dark blue: *Etheostoma lynceum* sensu stricto; light blue: *Etheostoma* sp. Obion; mid blue: intermediates between *E. lynceum* sensu stricto and *Etheostoma* sp. Obion.

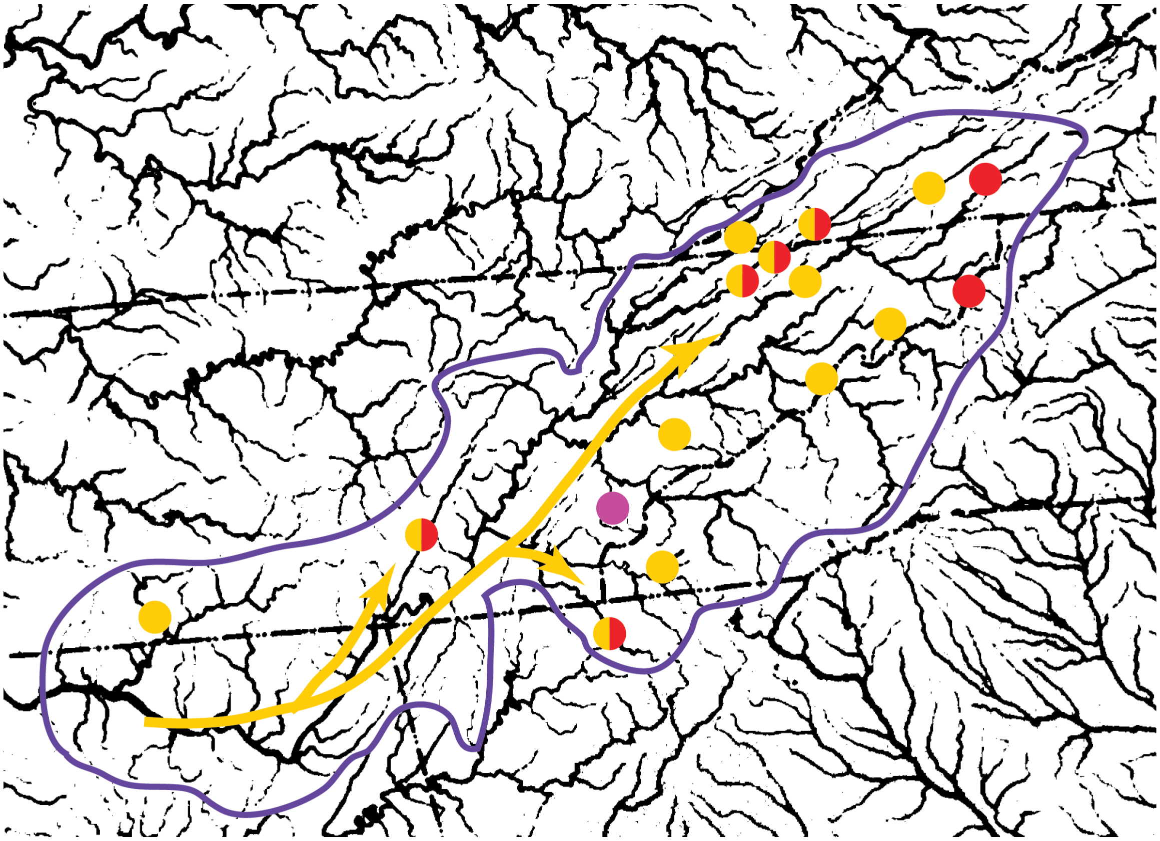


Figure 3.16 Distribution of nuclear gene clades in the upper Tennessee River basin (outlined). Yellow represents members of the Upper Tennessee and Hiwassee clades; red represents members of the South Fork Holston clade, and purple represents the Little Tennessee clade. The arrow indicates the putative direction of dispersal of the Upper Tennessee clade. The three split dots at the top of the map represent the Clinch River populations, which contain alleles from the South Fork Holston clade and the Upper Tennessee clade; the two lower split dots represent the Sequatchie and Toccoa River populations, which include individuals with alleles at the nedd4l locus that are recombinant between the South Fork Holston and Upper Tennessee clades.

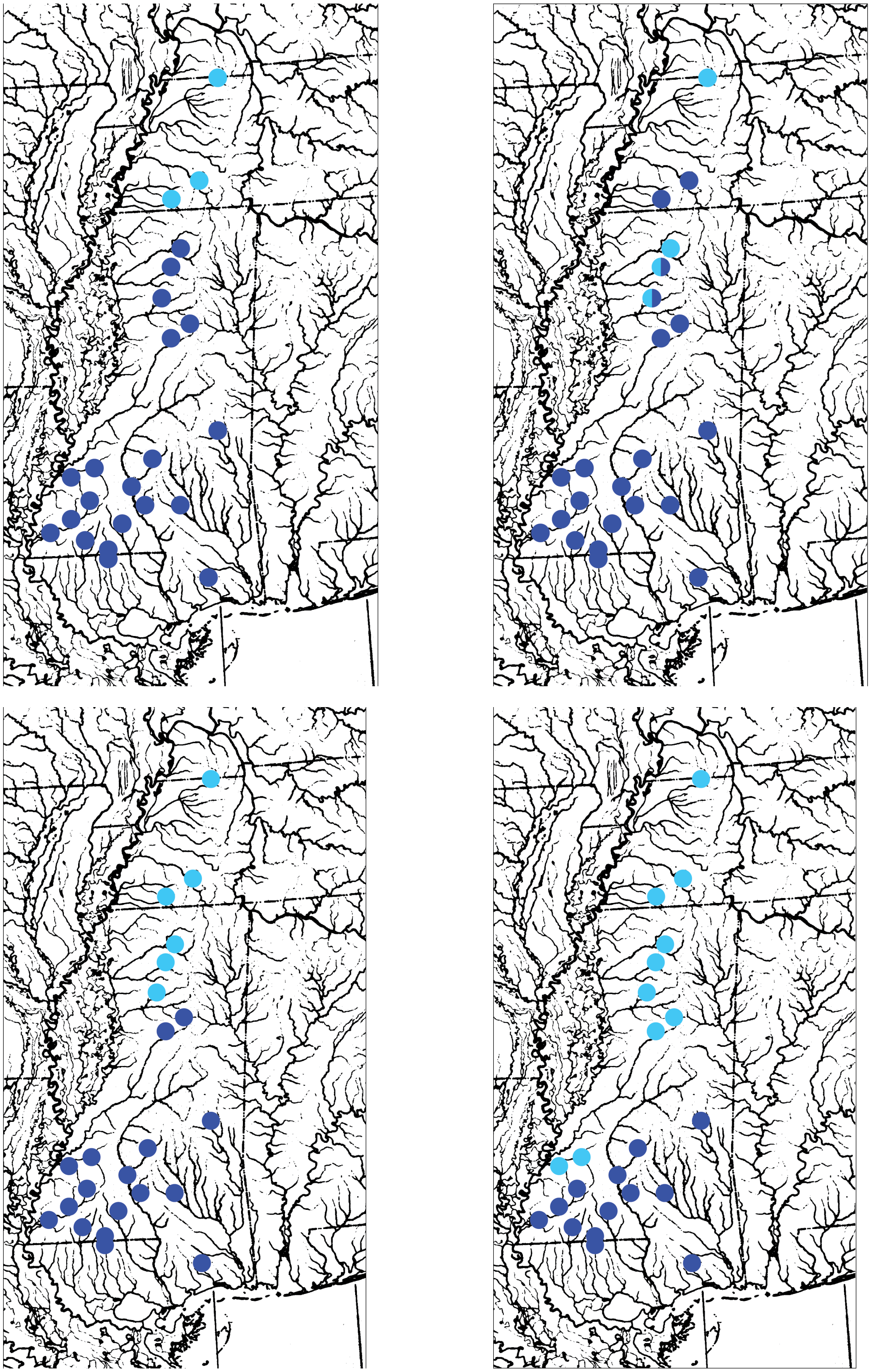


Figure 3.17 Geographical splits seen in various loci for *Etheostoma lynceum*. Top left: split between northern and southern allele groupings in ncl1; top right: nedd4l; bottom left: slco4a1; and bottom right: pabpc4 and cyt *b*.

CHAPTER 4

An embarrassment of introgression: systematics of the *Luxilus zonatus* species group

Introduction

Hybridization occurs in all kinds of sexually-reproducing organisms, in some to a far greater extent than others; these latter cases are the ones that produce many of our thornier species problems. When hybridization is not an evolutionary dead-end, but leads to the permanent transfer of genetic material from one species or population to another, this is known as introgression. Introgression is one of the main forms of horizontal transfer of genes among eukaryotes. Identifying introgression is important not only to study it for its own sake, but because it makes reconstructing the phylogenetic relationships of organisms much more difficult than it might otherwise be.

The *Luxilus zonatus* species group was selected for the current study for a number of reasons: its members are allopatrically distributed throughout the Ozarks, giving information about divergence events within this region; it had not yet been thoroughly sampled, meaning that a new analysis could considerably improve on our existing estimates of its phylogenetic relationships and divergence times; and because one of the apparent divergences within the group was geographically congruent with divergences in two *Etheostoma* species, allowing it to serve for the test of pseudocongruence presented in Chapter 5. In the course of collecting data for this analysis, several mitochondrial introgression events were discovered throughout the clade, which are now described.

The *Luxilus zonatus* species group contains three species, each allopatrically distributed in one or more of the main Ozark drainages. The first member of the group to be described was *Luxilus zonatus* by Agassiz (in Putnam 1863), as *Alburnus zonatus*. *Luxilus pilsbryi* was described by Fowler in 1904 as *Notropis pilsbryi*. Mayden (1988b) described *L. cardinalis*, distinguishing it from *L. pilsbryi*, and created a hypothesis of the relationships of the *L. zonatus* species group based upon morphological and allozymatic characters, which placed *L. zonatus* sister to the other two species. At various times, the *Luxilus zonatus* species group has been considered to be one, two, or three species, and has been placed in various genera; Mayden (1988b) gives an account of the systematics of the group up to the publication of his paper. Dowling and Naylor (1997) using a single cytochrome *b* sequence for each member of the *L. zonatus* species group, along with sequences for each of the other members of *Luxilus* in a maximum parsimony analysis, recovered a monophyletic *L. zonatus* species group sister to *L. cerasinus* in a monophyletic *Luxilus.* Mayden et al. (2006) in a broadly-sampled study of the genus *Notropis* and related genera including *Luxilus*, recovered a monophyletic *Luxilus zonatus* species group, but its relationships with other taxa were unresolved; no evidence for the monophyly of *Luxilus* was found.

Hybridization is well known in the Cyprinidae (Stauffer et al. 1979, Cooper 1980, Poly 1997, Eisenhour and Piller 1997) including within the genus *Luxilus* (Meagher and Dowling 1991); this study adds four new cases to the record not only of hybridization, but of introgression, within the Cyprinidae.

The distribution and sequence of divergence of the members of the *Luxilus zonatus* species group appears to correlate with that of other taxa (Chapter 2); I hypothesized that a multi-locus analysis of the group would continue to support these apparent vicariance events, while providing a better estimate of the actual divergence times of each member of the group.

Methods

Individuals of *Luxilus cardinalis* were collected from three localities across the range of the species, individuals of *L. pilsbryi* from five localities, and individuals of *L. zonatus* from nine localities. Specimens were collected with seines and backpack shockers, and transported to the laboratory in liquid nitrogen or ethanol. Other species of *Luxilus* and *Notropis* were used as outgroup taxa. Specimens of *Luxilus chrysocephalus* from three localities, *Notropis rubellus* from two localities, as well as single individuals of *L. cornutus*, *N. atherinoides, N. boops, N. greenei, N. leuciodus, N. nubilus, N. ozarcanus*, *N. photogenis*, *N. stilbius*, *N. stramineus*, *N. telescopus*, and *N. volucellus* were used as outgroups for all nuclear genes sequenced; mitochondrial cytochrome *b* sequences of these species along with *Agosia chrysogaster*, *Campostoma anomalum*, *Cyprinella spiloptera*, *Erimystax x-punctatus*, *Hybopsis amblops*, *Luxilus albeolus*, *L. cerasinus*, *L. coccogenis*, *L. zonistius*, *Lythrurus ardens*, *Notropis calientis*, *N. chrosomus*, *N. dorsalis*, *N. edwardraneyi*, *N. heterolepis*, *N. hudsonius*, *N. longirostris*, *N. sabinae*, *N. shumardi*, *N. texanus*, *Opsopoeodus emiliae*, *Phenacobius uranoscopus*, *Pimephales notatus*, *Pteronotropis hubbsi*, and *P. signipinnis* were downloaded from Genbank. Existing cytochrome *b* sequences of the ingroup taxa were also downloaded from Genbank, while new cytochrome *b* sequences were obtained for the specimens of *Luxilus chrysocephalus* and *L. cornutus* used in this study.

Genomic DNA was extracted using QIAamp™ tissue extraction kits (Qiagen Inc.) according to the manufacturer's instructions. The complete mitochondrial cytochrome *b* gene was amplified using the polymerase chain reaction (PCR). PCR was performed in a total volume of 25 µL, containing 5 µL 5x Green GoTaq Flexi Buffer, 1.5 µL 25 mM $MgCl_2$, 0.5 µL 10 mM dNTPs, 0.03 nmol each of the forward and reverse primers, 0.125 µL GoTaq Flexi DNA polymerase, and approximately 1.0 µg of DNA, under the following thermocycler settings: initial denaturation at 95.0 ºC for 1 minute; 30 cycles of 95.0 ºC for 30 seconds, 53.0 ºC for 1 minute, and 72.0 ºC for 2 minutes; and a final extension at 72.0 ºC for 10 minutes, with the reaction terminating at 4 ºC. Cytochrome *b* was sequenced using the primers HA and LA of Schmidt et al. (1998). New internal primers were designed to allow complete sequencing of both strands:

LuxF1 [5′ ATTTGAGGRGGCTTTTCAGT 3′] and LuxR1 [5′ TAGGAAGTGGAAGGCGAAGA 3′]. For a subset of nine individuals, PCR was performed twice, and for each individual the light strand was sequenced from one reaction, and the heavy strand from the other, to test for the effect of PCR error on sequencing. Amplified DNA was purified using 4 units of exonuclease 1 and 0.4 units of shrimp alkaline phosphatase per reaction. Automated sequencing was performed using Big Dye terminator cycle sequencing at the DNA Sequencing and Analysis Facility of the Biomedical Genomics Center at the University of Minnesota. Sequences were checked for accuracy and assembled using Sequencher 4.7 (Gene Codes Corporation).

Three new nuclear intron loci were also developed for use in this study. The *Danio rerio* (zebrafish) genome build Zv9 was searched at the National Center for Biotechnology Information website for introns of a suitable length for sequencing in one segment, and with flanking exons of a suitable length for primer development. Where such introns were found, the flanking exons were were BLAST searched against the entire zebrafish genome, in order to weed out gene duplications. Exons without duplications elsewhere in the genome were then BLAT searched at http://genome.ucsc.edu against the *Tetraodon*, fugu, medaka, and stickleback genomes to find conserved sequences. Where sequences of a suitable length for primer design were found in at least one of the other genomes examined, the exon sequences were used to design primers in Primer3 (http://frodo.wi.mit.edu/primer3/). All primer pairs developed were tested using the same PCR conditions as were followed for cytochrome *b* amplification, except that the annealing temperature used was always 2 ºC less than the lower of the two melting temperatures of the primers. PCR products from loci which amplified successfully were purified and sequenced as described above for cytochrome *b*. Three loci sequenced successfully and showed variation within *Luxilus*: the 8th intron of nicalin (ncl1), found on the 22nd chromosome in *Danio*; the 3rd intron of ribosomal protein SA (rpsa), found on the 6th chromosome; and the 4th intron of ribosomal protein S3 (rps3), found on the 18th chromosome. The primers designed for each locus, along with the annealing temperatures used during amplification, are given in Table 4.1. Apart from the change in annealing temperature, amplification conditions for all three nuclear loci were identical to those for cytochrome *b*.

Nuclear genes heterozygous at two or more sites were phased using one of three methods: sequences which were length heterozygotes were phased by eye in Sequencher (Flot et al. 2006). Sequences which were not length heterozygotes were phased using the program PHASE, with homozygous sequences, sequences with only one heterozygous site, and sequences already phased due to a length heterozygosity included as known sequences. Five separate runs of PHASE were performed for each locus, with a final pass ten times the length of previous ones. If the phase of a sequence was reconstructed with greater than 90 percent certainty at all sites, this phasing was used all analyses. Sequences which were not reconstructed by PHASE with this level of accuracy were phased using allele-specific primers (Petersson et al. 2003). Pairs of primers differing only at the last base at the 3′ end were designed, with a length sufficient for the melting temperature of the primer to match the melting temperature, within one or two degrees Celsius, of the regular amplification primer to be used at the other end of the sequence being phased. After allele-specific sequences were obtained, PHASE was re-run, using the sequences obtained using allele-specific sequencing as known haplotypes. If sequences which were previously reconstructed by PHASE with high confidence now showed a lower degree of confidence, allele-specific primers were also designed for these sequences. This cycle was repeated until all remaining unphased sequences were reconstructed by PHASE with high confidence.

All newly-obtained sequences for all genes will be deposited in Genbank.

Nuclear genes were tested for recombination using the phi test (Bruen et al. 2006) as implemented in Splitstree (Huson and Bryant 2006), as this test is suitable for detecting recombination both within populations of a single species, and across species. DNAsp v5 was used to determine the number of polymorphic sites at each locus. For each locus, the program jModelTest (Posada 2008, Guindon and Gascuel 2003) was used to select a suitable model of evolution. The Akaike Information Criterion was used to select models, and to select between an unpartitioned model, and one partitioned by codon position, for cytochrome *b*. Gaps in nuclear loci were coded for analysis using the program SeqState (Müller 2005). Gaps with ambiguous alignments were not coded. In addition, gaps with more than two states were excluded, to avoid having an additional model of evolution for a partition with very few characters. Some taxa had very long gaps, which spanned several other gaps in other taxa; to allow those smaller gaps to be coded, these long gaps were coded as missing data in the taxa that displayed them. Maximum likelihood analyses were performed in Garli 2.0 (Zwickl 2006). For nuclear genes, all analyses were partitioned, with one partition for the sequence data, and another for the gap data. For all loci, the Mkv model (Lewis 2001) was used for the gap data. One hundred bootstrap replicates were run for each locus. Mrmodeltest 2.3 (Nylander 2004) was used to determine suitable models for analysis of each gene in MrBayes. Cytochrome *b* was analyzed in MrBayes 3.1.2 (Huelsenbeck and Ronquist 2003; Ronquist and Huelsenbeck 2005) using a model partitioned by codon; nuclear genes were run with one partition for sequence data and a second for gap data. All runs were 10,000,000 generations long, with a burnin of 1,000,000 generations.

To test for population growth in each species, the R2 Test (Ramos-Onsins and Rozas 2002) was used. The R2 test was performed in DnaSP v5 on the cytochrome *b* and nuclear sequences for each species. To determine the significance of the test statistic for each test, 10,000 coalescent simulations were performed.

A species tree analysis was conducted using *BEAST (Heled and Drummond 2010). The terminals for the analysis were the three species in the *Luxilus zonatus* species group, and various subsets of outgroup taxa: one analysis was run with only those six outgroup taxa for which nuclear gene sequence data was available for all three loci, and a second was run with the eleven outgroup taxa for which sequence data was available for at least two of the three nuclear loci. For this second analysis, the nuclear gene sequence of taxa for which a sequence was not available was represented as missing data. Cytochrome *b* sequences which appeared to have originated due to mitochondrial introgression were excluded from the analyses, as were three ncl1 sequences in *L. cardinalis* that appeared to be the result of hybridization. Models for each locus were selected using MrModelTest. Cytochrome *b* was run unpartitioned for the analysis. All loci were run with empirical base frequencies. Molecular clock likelihood ratio tests for each locus were performed in PAUP* to determine whether each locus should follow the strict clock model, or the relaxed clock uncorrelated exponential model. Default priors and operators were used. Several runs were performed for each analysis to assess convergence between runs, and then combined to obtain the final species tree. For each analysis, a priors-only data set was also created and run in *BEAST, to determine the influence of the priors on the posterior values obtained for each parameter. The species trees were calibrated using a sequence divergence of 2.0% per million years (Near and Benard 2004).

Results

Cytochrome *b* was successfully sequenced for 66 ingroup individuals. Nine individuals were sequenced from two separate PCRs, one for sequencing the forward strand, and one for the reverse strand. No discrepancies were found between the strands for each of these sequences, giving a PCR and sequencing error rate of less than 0.01 percent.

Forty-one ingroup individuals were sequenced at each of the three nuclear intron loci, giving 82 sequences for each locus altogether. In addition, sequences of each intron locus were sequenced for three individuals of *Luxilus chrysocephalus* and one of *L. cornutus*; intron sequences were also obtained for between six and eleven outgroup members of the genus *Notropis*. Rpsa showed little length variation; rps3 had little variability in length in the ingroup, but much more in the outgroup, and ncl1 showed considerable length variability in both the ingroup and outgroup. Ranges and modes of intron length are given in Table 4.2. All loci also included gaps of various lengths, some of which were parsimony-informative. The number of polymorphic and parsimony-informative sites at each locus, along with the number of sequences obtained at each locus, the length of each aligned locus, and the number of gaps coded for each locus are given in Table 3.3.

The three nuclear loci were tested for recombination; ncl1 gave a significant result. The recombination test in DnaSP was used to determine where to trim the aligned sequences until there was no longer a signal of recombination. The aligned ncl1 sequences went from 405 bp to 323 bp after trimming. All loci sequenced were successfully phased using one of the three techniques described in the Methods. Allele-specific sequencing resulted in duplicated PCR reactions for over 1000 bp of sequence in all nuclear loci, this served as a test of the PCR and sequencing error rate for these loci. No errors were detected, giving an error rate of less than 0.1%.

Models of evolution selected for likelihood and Bayesian analyses of each locus are given in Table 4.4. The AIC selected a model partitioned by codon position for cytochrome *b*. The results of the likelihood analysis are shown in Figure 4.2. There are three well-supported clades on the tree, each corresponding to one of the three ingroup species; however, none of the three ingroup species is monophyletic, due to apparent mitochondrial introgression. A *Luxilus cardinalis* clade is sister to a *Luxilus pilsbryi* clade with strong support; however, the *L. cardinalis* clade includes three individuals of *L. pilsbryi* from the James River in Missouri. The *L. pilsbryi* clade includes all five individuals of *L. zonatus* sampled from the Strawberry River in the Black River drainage in Arkansas. In neither case do the introgressed individuals form a clade. Sister to these two clades is an *L. zonatus* clade with no introgressed individuals. This clade only includes individuals of *L. zonatus* from the Black and St. Francis rivers. There is no notable geographic structuring within any of these three clades. The *L. zonatus*, *L. cardinalis* and *L. pilsbryi* clades together form a strongly-supported clade. They are sister to a clade consisting of *Notropis greenei*, *Opsopoeodus emiliae*, and *Pimephales notatus*, but with no support.

Sister to the clades discussed so far is a clade consisting of *L. chrysocephalus*, *L. cornutus*, and *L. albeolus*; this clade is supported in the Bayesian analysis but not the likelihood bootstrap. This clade also includes the remaining individuals of *L. zonatus*. One branch of this clade consists of individuals of *L. chrysocephalus* along with *L. zonatus* from the Missouri and Meramec River drainages. The *L. zonatus* individuals form a strongly-supported clade. The net between-group distance between *L. chrysocephalus* and *L. zonatus* in this clade is 0.0216, calculated using the formula $\delta = \delta_{xy} - (\delta_x + \delta_y)/2$ (Nei and Li 1979), where δ_x and δ_y are the mean distances within the *L. zonatus* clade and the subclade of four *L. chrysocephalus* individuals sister to it, and δ_{xy} is the average distance between the two clades. Using a percent sequence divergence of 2.0 % per million years (Near and Benard 2004) this indicates a time for introgression of the *L. chrysocephalus* mtDNA into *L. zonatus* of slightly over one million years ago. There is geographical structuring within this clade: the Gasconade and Osage River populations of *L. zonatus* form a strongly-supported clade that renders the Meramec River population paraphyletic.

The other branch of the *L. chrysocephalus* + *L. cornutus* + *L. albeolus* clade includes *L. cornutus* and *L. albeolus* in a well-supported sister relationship. The *L. cornutus* clade, however, also includes individuals of *L. zonatus* from the St. Francis and Black Rivers, and individuals of *L. chrysocephalus* from the St. Francis and Gasconade Rivers. The remaining species of *Luxilus* appeared among the outgroup taxa, *L.*

cerasinus being the furthest removed from the *L. zonatus* species group. The sampling localities for all ingroup species, and their assignments to the five haplogroups described above, are shown in Figure 4.1.

The three nuclear gene trees are shown in Figures 4.3-4.5. *L. zonatus* was recovered as monophyletic at two loci and *L. cardinalis* at one; none of the loci recovered a monophyletic *L. pilsbryi.* One locus recovered a monophyletic *L. zonatus* species group.

The rpsa gene tree (Figure 4.3) recovers a monophyletic *L. zonatus* with weak bootstrap and no Bayesian support; this is sister to *L. cornutus* with strong support. A clade consisting of *L. pilsbryi*, *L. cardinalis*, and *Notropis telescopus*, while unsupported, is sister to *L. chrysocephalus*, *N. rubellus*, and *N. stilbius* with Bayesian but not bootstrap support. *L. cardinalis* is monophyletic with bootstrap but not Bayesian support; *L. pilsbryi* is paraphyletic with respect to *L. cardinalis* and *N. telescopus*. The long branch leading to *N. telescopus* suggests that its placement here may be in error, perhaps due to long-branch repulsion (Siddall 1998).

The rps3 gene tree (Figure 4.4) shows a monophyletic *L. zonatus* species group, with Bayesian and weak bootstrap support. *L. zonatus* is monophyletic, also with Bayesian and weak bootstrap support. There is some geographical structure within the species; a weakly-supported clade includes all of the Missouri and Meramec River populations, along with one St. Francis River individual. Neither *L. pilsbryi* nor *L. cardinalis* are monophyletic.

Ncl1 (Figure 4.5) shows less structure than the other two introns. *L. chrysocephalus*, *L. cornutus*, *N. greenei*, *N. boops*, and *N. telescopus* all fall within the clade including the *L. zonatus* species group. All three ingroup species are polyphyletic, and there is no apparent geographical structuring. Three *L. cardinalis* alleles fall far out in the outgroup; these may represent a hybridization event, and were excluded from most subsequent analyses.

The *BEAST species tree, generated from all four loci, is shown in Figure 4.6. The tree was generated from 130,000,000 post-burnin generations spread across two separate runs, achieving convergence of parameters across runs, and satisfactory effective sample size (ESS) values for all key parameters in the combined runs. The species tree recovers a well-supported sister-group relationship of *L. pilsbryi* and *L. cardinalis*; *L. zonatus* is recovered as the weakly-supported sister to these species. The two successive outgroups are *L. cornutus* and *L. chrysocephalus*; the grouping of the five *Luxilus* species is strongly supported. The 95 % highest posterior density intervals for the divergence times are quite broad; using a percent sequence divergence of 2.0 % per million years (Near and Benard 2004) the *L. zonatus* species group may have originated in the Miocene or Pliocene, *L. zonatus* diverged from its sister species in the late Miocene, Pliocene, or early Pleistocene; and *L. cardinalis* and *L. pilsbryi* diverged in the Pliocene or Pleistocene. A second species tree analysis using only those outgroup taxa for which sequence data was available at all four loci produced similar results (not shown).

The results of population growth tests for each of the clades at all loci using the R2 test are given in Table 4.5. For cytochrome *b*, only non-introgressed individuals were used; the exception is the clade of *L. zonatus* found in the Missouri and Meramec Rivers, for which a separate test was performed. *L. cardinalis* shows a signal of growth in cytochrome *b*, and *L. pilsbryi* shows a strong signal of growth in rpsa.

Discussion

There are two features which are immediately evident in the gene trees for the *L. zonatus* species group: the lack of any significant geographic structuring within species, and the multiple mitochondrial introgression events which have occurred, which are almost the only source of any geographical structuring that exists. The lack of geographical structuring is largely consistent with what is known about the morphological variation in the species group. Mayden (1988b) reported variation in adult size and tuberculation among *L. cardinalis* from the northern and southern parts of its range. All specimens of *L. cardinalis* included in this study are from the northern part of the range; it would be worthwhile to obtain additional specimens from the southern part of the range to determine if there is any notable genetic variation among them. No variation was found by Mayden in *L. pilsbryi*, consistent with the lack of geographical structuring seen in any of its loci; or in *L. zonatus*. The limited geographic structuring seen in *L. zonatus* is worth further investigation.

Of the various introgression events observed in the *L. zonatus* species group, two admit of a simple explanation: the introgression of *L. cardinalis* mtDNA into James River *L. pilsbryi*, and the introgression of *L. pilsbryi* mtDNA into Strawberry River *L. zonatus*. Both of these introgressions are limited geographically to single sampling localities in this study. Both localities are adjacent to the range of the source species of the mtDNA, thus, a minor stream capture event, introducing members of one species into the range of the other, may account for both introgressions. More complex, however, are the introgressions of *L. chrysocephalus* and *L. cornutus* mtDNA into *L. zonatus*. The introgression of *L. chrysocephalus* mtDNA into *L. zonatus* in the Missouri and Meramec River drainages appears to have taken place about one million years ago. However, all known *L. chrysocephalus* populations within this region are themselves introgressed, bearing only *L. cornutus* mtDNA (Duvernell and Aspinwall 1995). At the time that the introgression took place, *L. chrysocephalus* in the Missouri River basin may have still had their ancestral mtDNA. Alternatively, this mtDNA lineage may have entered Missouri basin *L. zonatus* from the Black or St. Francis Rivers, where some *L. chrysocephalus* retain their ancestral mtDNA, through a stream capture event. Duvernell and Aspinwall hypothesized that *L. cornutus* mtDNA may have a selective advantage in F_1 hybrids; the results of this study raise the intriguing possibility that *L. chrysocephalus* mtDNA, but not *L. cornutus* mtDNA, has a selective advantage in F_1 hybrids between *L. zonatus* and *L. chrysocephalus*. The presence of *L. cornutus* mtDNA in some, but not all, individuals of both *L. chrysocephalus* and *L. zonatus* in the Black and St. Francis River drainages only confuses the question, however. *L. zonatus* in the Black and St. Francis River drainages with *L. cornutus* mtDNA are more closely related to *L. chrysocephalus* with *L. cornutus* mtDNA from the Gasconade River drainage than they are to *L. cornutus* itself. This suggests that *L. chrysocephalus* may have acted as a conduit for the introgression of *L. cornutus* mtDNA into *L. zonatus*, with no contact between *L. cornutus* and *L. zonatus* ever having been necessary. A similar pattern of mitochondrial introgression through a conduit species has been observed in darters of the genus *Nothonotus* in the Cumberland and Tennessee Rivers (Keck and Near 2010). The present-day distributions of the three species support this hypothesis: *L. chrysocephalus* and *L. zonatus* overlap broadly in range, but *L. cornutus* is almost entirely allopatric to *L. zonatus* (Figure 4.7). *Luxilus chrysocephalus* and *L. cornutus* do overlap in range, but the closest that the current hybrid zone comes to the Ozarks is northeastern Illinois. *Luxilus cornutus* mtDNA may have spread into *L. chrysocephalus* in the Ozarks through geographically intermediate populations, or *L. cornutus* may have had a southward range extension in the past, into the Ozark drainages. This possibility would suggest that *L. cornutus* also overlapped in range with *L. zonatus* in the past, complicating the explanation of the origin of *L. cornutus* mtDNA in *L. zonatus*. More thorough sampling of *L. zonatus*, and *L. chrysocephalus* in the Missouri, Meramec, Black and St. Francis River drainages, along with ancestral range reconstructions based upon paleoclimate data, may be able to elucidate the true history of those populations of *L. zonatus* with *L. chrysocephalus* mtDNA.

In contrast to the mtDNA results, there is little to no evidence of any introgression in the nuclear loci. In this, the *L. zonatus* species group is consistent with many other fish taxa where introgression has occurred (Scribner et al. 2001). The one apparent exception is the three alleles of *L. cardinalis* which for ncl1 group with various species well in the outgroup; there area insufficient data for the outgroup taxa in the present analysis to determine the source of these alleles.

The species tree analysis for the *L. zonatus* species group suffers from broad posterior densities for divergence times. Thus, few definite conclusions can be drawn about the history of diversification within the group. However, it appears likely that the origin of the group predates the Pleistocene. *Luxilus zonatus* also appears to have diverged prior to the Pleistocene, while the *L. cardinalis – L. pilsbryi* divergence is most likely to have taken place within the Pleistocene. The topology of the relationships within the group is unchanged from previous analyses. Compared to the divergence times derived from Dowling and Naylor's (1997) single cytochrome *b* sequences, however, the divergence times seen in the species tree are more recent. The more recent divergence times are sufficient to bring the divergence between *L. zonatus* and *L. pilsbryi* + *L. cardinalis*, part of Common Divergence Event 2 in Chapter 2 (Figure 2.10) into overlap with Common Divergence Event 8 (Figure 2.5), which involves the same areas, although the divergences in Event 8 place the Black River basin on a different side of the split than seen in the *L. zonatus* species group. This perhaps demonstrates, more than anything else, the importance of obtaining well-supported divergence time estimates for comparative phylogeographic analyses. Furthermore, any conclusions about diversification times in this group must be tempered by the lack of a calibration based upon closely-related species; a sequence divergence of one percent per million years, rather than two percent, would, for instance, place most divergences in the group within the Pleistocene. The sister relationship of the *L. zonatus* species group with *L. chrysocephalus* and *L. cornutus* seen in the species tree is suggestive, as a recent study of the phylogeny of *Luxilus* and *Notropis* has not recovered a monophyletic *Luxilus* (Mayden et al. 2008). The limited outgroup sampling in the current study prevents any definite conclusions from being drawn, but broader multilocus sampling is clearly warranted. A flaw in the study is that current species tree methods do not allow for the possibility of introgression. *BEAST and other species tree methods such as BEST (Liu 2008) assume that all gene tree/species tree discordance is due to incomplete lineage sorting. This is obviously not the case in the *Luxilus zonatus* species group, and so all introgressed individuals had to be removed from the cytochrome *b* data set prior to the analysis. This may not be significant for recently introgressed populations such as the Strawberry River population of *L. zonatus* and the James River population of *L. pilsbryi*, but there is clearly a significant loss of data in the removal of the Missouri River populations of *L. zonatus.* Given the extent to which hybridization and introgression occurs throughout all kingdoms of life (Arnold et al. 2008), the development of methods of species tree analysis that take into account introgression is a priority; however, the problem may prove too complex to be computationally tractable in all but the simplest cases.

This study demonstrates the value of geographically broad sampling in phylogenetic studies. Prior to the current study, only one gene sequence for each of the three *Luxilus zonatus* group species was available on Genbank, the three cytochrome *b* sequences published in Dowling and Naylor (1997). Had Dowling and Naylor sequenced an individual of *L. zonatus* from the Strawberry, Meramec, St. Francis or Missouri River basins, they would not have recovered a monophyletic *L. zonatus* species group. Indeed, if *L. zonatus* is equally abundant throughout its range, the majority of individuals belonging to the species do not have the ancestral mitochondrial DNA of the species. One might thus say that Dowling and Naylor were lucky to sequence an individual from the minority of the population that still bears its ancestral mitochondrial DNA.

Conclusion

The previously hypothesized relationships of the *L. zonatus* species group are supported by the current study: *Luxilus zonatus* is sister to the species pair of *L. pilsbryi* and *L. cardinalis*. The instances of mitochondrial introgression presented in this study are previously unknown, but are perhaps not overly surprising, given both the broadcast spawning habit of the members of *Luxilus*, and the previously known introgression in *L. chrysocephalus* and *L. cornutus* from the same region (Duvernell and Aspinwall 1995). The zone of introgression in southeastern Missouri between *L. chrysocephalus* and *L. cornutus* is now shown to include *L. zonatus* as well, while the introgression zone between *L. cornutus* and *L. chrysocephalus* in the Missouri River drainage also involves *L. zonatus*, but in an unusual way: all *L. chrysocephalus* present in this drainage have *L. cornutus*-type mtDNA, while all *L. zonatus* present have *L. chrysocephalus* type mtDNA.

The divergence times obtained for the *L. zonatus* species group have broad posterior density intervals, and do not contradict the previous analysis based on cytochrome *b* data alone. The *L. zonatus* species group appears to have diverged during the Pliocene or the early Pleistocene; subsequently, various introgression events occurred in the group during the Pleistocene, involving either transfer of mtDNA from one member of the group to another, or from *L. chrysocephalus* and *L. cornutus* to *L. zonatus*.

Table 4.1 Primers for nuclear loci, with annealing temperatures for each primer pair.

Locus	Primer name	Primer sequence	Annealing temperature (ºC)
ncl1 8th intron	ncl1f8	5′ CCAGTCTGCTSCAGGACAAY 3′	50.9
	ncl1r8	5′ SGCCAGGTTGATYTTCTTRT 3′	
rpsa 3rd intron	rps3f4	5′ CTACAAGCTGCTSGGAGGMC 3′	50.6
	rps3r4	5′ TAGTTSACKGGGTCTCCRCT 3′	
rps3 4th intron	rpsa3f	5′ ATTGTTGCCATYGARAAYCC 3′	55.6
	rpsa3r	5′ GCWGCCTGRATCTGATTGGT 3′	

Table 4.2 Variation in length of intron loci. All figures are lengths in bp.

Locus	Ingroup			Outgroup	
	Minimum length	Modal length	Maximum length	Minimum length	Maximum length
ncl1	396	405	453	335	1081
rps3	393	395	395	392	769
rpsa	331	331	331	315	342

Table 4.3 Number of polymorphic and parsimony informative sites per locus for the *L. zonatus* species group, and for all sequences including outgroups. Numbers are also given for the 1st, 2nd, and 3rd positions of cytochrome *b*. Numbers of gap characters encoded for nuclear loci are also indicated.

	With outgroups			Ingroup only				
Locus	Polymorphic	Parsimony informative	Number of sequences	Polymorphic	Parsimony informative	Number of sequences	Alignment length	Gap characters
cyt *b*	475	415	115	213	202	66	1140	—
1st pos	87	58	115	17	16	66	—	—
2nd pos	29	6	115	1	1	66	—	—
3rd pos	359	351	115	195	185	66	—	—
ncl1	39	23	106	40	29	82	323	9
rps3	54	42	102	18	12	82	395	8
rpsa	66	47	112	32	21	82	331	8

Sites with gaps are not included in the counts. Ingroup counts are higher for ncl1 than counts with the outgroup included because several outgroup taxa contain large gaps which reduced the total number of included sites.

Table 4.4 Models of evolution for Garli, MrBayes, and *BEAST analyses, and molecular clock test results for *BEAST analyses for each locus.

Locus	Model for Garli	Model for Bayesian analyses	Clock rejected?
cyt *b* 1st position	TIM2+I+G	SYM+I+G	—
cyt b 2nd position	TIM2+I+G	F81+G	—
cyt *b* 3rd position	TIM2+G	GTR+I+G	—
cyt *b* unpartitioned	TIM1+I+G	GTR+I+G	Y
ncl1	JC+I+G	HKY+I+G	Y
rps3	TIM3+G	HKY+G	N
rpsa	TVMef+G	GTR+G	N

Table 4.5 Results of the R2 test for population growth for *Luxilus* at all loci. Significant results (indicating population growth) are shown in bold.

		cyt *b*	ncl1	rps3	rpsa
L. cardinalis	R2	**0.1108**	0.1001	0.1903	0.0957
	p	**0.04505**	0.06907	0.88878	0.08551
	n	10	17	17	17
L. pilsbryi	R2	0.1073	0.1519	0.1253	**0.0728**
	p	0.146	0.746	0.35942	**<0.00001**
	n	15	22	22	22
L. zonatus	R2	NA	0.1076	0.0924	0.0734
	p	NA	0.47	0.29032	0.10223
	n	NA	40	40	40
L. zonatus Black River	R2	0.1166	NA	NA	NA
population	p	0.2	NA	NA	NA
	n	15	NA	NA	NA
L. zonatus	R2	0.1044	NA	NA	NA
Missouri and Meramec River	p	0.088	NA	NA	NA
population	n	15	NA	NA	NA

In *L. zonatus*, the Black River population includes those individuals with unintrogressed mtDNA; the Missouri River population includes those individuals with *L. chrysocephalus*-type mtDNA.

Appendix 4.1 Specimens Examined. The specimens used in this study are listed with the number of specimens from each site in parentheses and collection locality. All specimens are uncatalogued specimens in the Bell Museum of Natural History frozen tissue collection.

Luxilus cardinalis

ArkansasA (6) Elk River, McDonald Co., MO.
ArkansasB (3) Shoal Creek, Newton Co., MO.
ArkansasC (1) Spring River, Jasper Co., MO.
U66601

Luxilus pilsbryi

Little Red (3) Archey Fork of Little Red River, Van Buren Co., AR.
WhiteA (3) Crooked Creek, Marion Co., AR.
WhiteB (5) James River, Stone Co., AR.
WhiteC (3) Kings River, Carroll Co., AR.
WhiteD (3) Richland Creek, Washington Co., AR.
U66602

Luxilus zonatus

BlackA (5) Black River, Reynolds Co., MO.
BlackB (3) Current River, Ripley Co., MO.
BlackC (3) Eleven Point River, Oregon Co., MO.
BlackD (3) Spring River, Sharp Co., AR.
BlackE (5) Strawberrry River, Sharp Co., AR.
Gasconade (5) Big Piney River, Texas Co., MO.
Meramec (5) Huzzah Creek, Crawford Co., MO.
Osage (5) Pomme de Terre River, Polk Co., MO.

St. Francis (5) St. Francis River, Madison Co., MO.
U66600

Luxilus cornutus (1) St. Croix River, Pine Co., MN.

Luxilus chrysocephalus

Gasconade (2) Big Piney River, Texas Co. MO.
St. FrancisA (1) St. Francis River, Madison Co., MO.
St. FrancisB (1) Village Creek, Cross Co., AR.
Black (1) Strawberry River, Sharp Co., AR.

Notropis atherinoides (1)

Notropis boops (1) North Rolling Fork, Boyle Co., KY.

Notropis greenei (1) Current River, Ripley Co., MO.

Notropis leuciodus (1) South Toe River, Yancey Co., NC.

Notropis nubilus (1) Big River, Washington Co., MO.

Notropis ozarcanus (1) Spring River, Fulton Co., AR.

Notropis photogenis (1) North Fork New River, Watauga Co., NC.

Notropis rubellus

Wabash (1) Tippecanoe River, Fulton Co., IN.
New (1) North Fork New River, Ashe Co., NC.

Notropis stilbius (1) Conasauga River, Polk Co., TN.

Notropis stramineus (1) Root River, Olmsted Co., MN.

Notropis telescopus (1) Current River, Ripley Co., MO.

Notropis volucellus (1) Mississippi River, Ramsey Co., MN.

Cytochrome *b* outgroup sequences:

Agosia chrysogaster (DQ324093), *Campostoma anomalum* (AF452079), *Cyprinella spiloptera* (U66605), *Erimystax x-punctatus* (AF117172), *Hybopsis amblops* (AF117152), *Luxilus albeolus* (U66598), *L. cerasinus* (U66599), *L. chrysocephalus* (U66595, U66596), *L. coccogenis* (U66603), *L. cornutus* (U66597), *L. zonistius* (U66604), *Lythrurus ardens* (U17268), *Notropis atherinoides* (AY096008), *N. boops* (AF352261), *N. calientis* (AF469139), *N. chrosomus* (AF342262), *N. dorsalis* (AF117162), *N. edwardraneyi* (AF352263), *N. longirostris* (AF117178), *N. nubilus* (AF352265), *N. photogenis* (AF352280), *N. rubellus* (AF117194), *N. sabinae* (AF117188), *N. shumardi* (AF117200), *N. stilbius* (AF352285), *N. telescopus* (AF352289), *Opsopoeodus emiliae* (U17270), *Phenacobius uranops* (AY486056), *Pimephales notatus* (U66606), as well as *Notropis greenei*, *N. heterolepis*, *N. hudsonius*, *N. leuciodus*, *N. ozarcanus*, *N. stramineus*, *N. texanus*, *N. volucellus, Pteronotropis hubbsi*, and *P. signipinnis* (all from Mayden et al. (2006), not on Genbank)

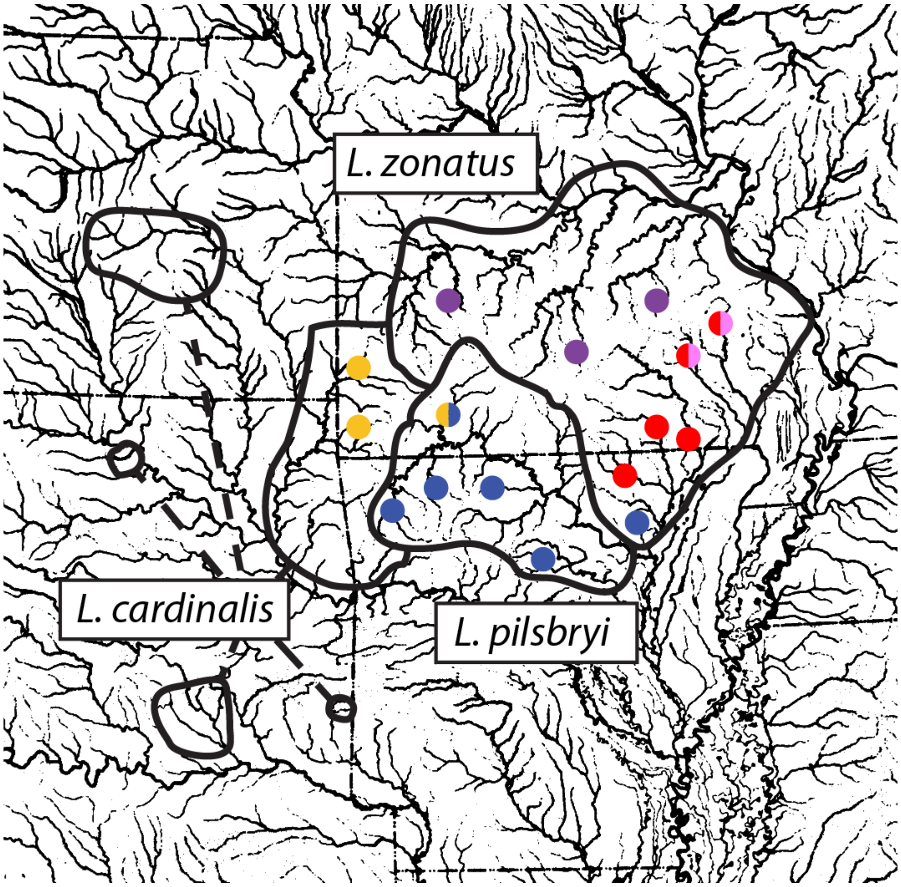


Figure 4.1 Distribution of the *Luxilus zonatus* species group, with sampling sites with cytochrome *b* haplogroup indicated. Species ranges are encircled by solid lines; *L. cardinalis* is found in the five disjunct areas joined by the dashed lines. Sampling sites are represented by dots; the color of each dot represents the cytochrome *b* haplogroup found at that site. Red = *L. zonatus* haplogroup, blue = *L. pilsbryi* haplogroup, yellow = *L. cardinalis* haplogroup, purple = *L. chrysocephalus* haplogroup, pink = *L. cornutus* haplogroup

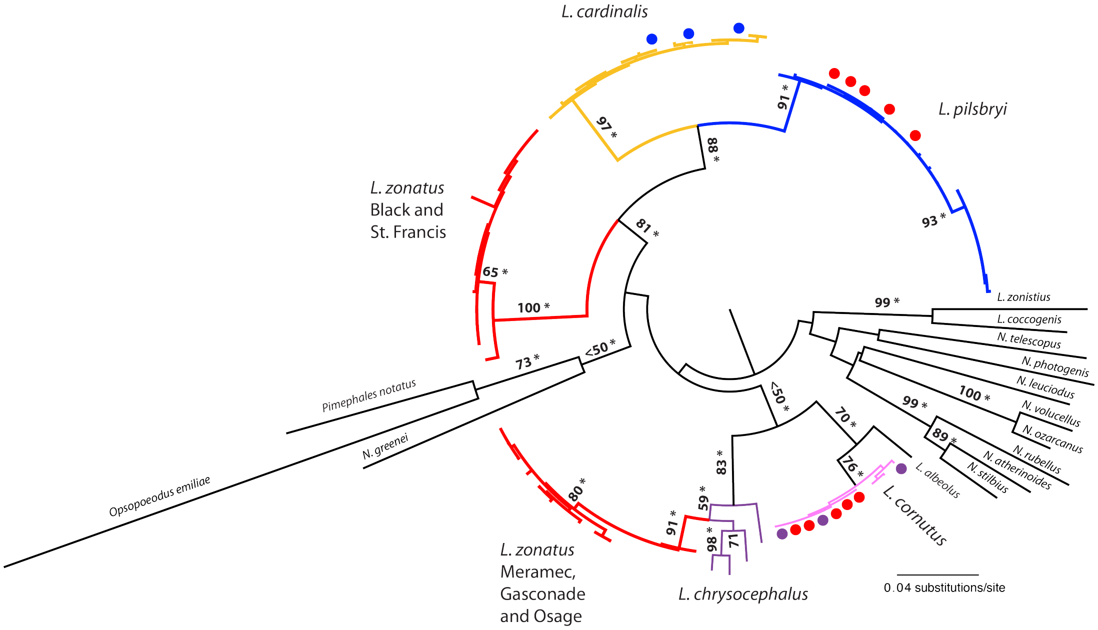


Figure 4.2 Cytochrome *b* maximum likelihood gene tree for the *Luxilus zonatus* species group. Yellow represents *L. cardinalis*, blue represents *L. pilsbryi*, red represents *L. zonatus*, purple represents *L. chrysocephalus*, and pink represents *L. cornutus*. Individuals which are apparently mitochondrially introgressed are indicated as dots of their respective species' colors. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Support values for some very short branches within species are not shown. Additional outgroups used in analysis not shown. Tree ln likelihood: -14504.781.

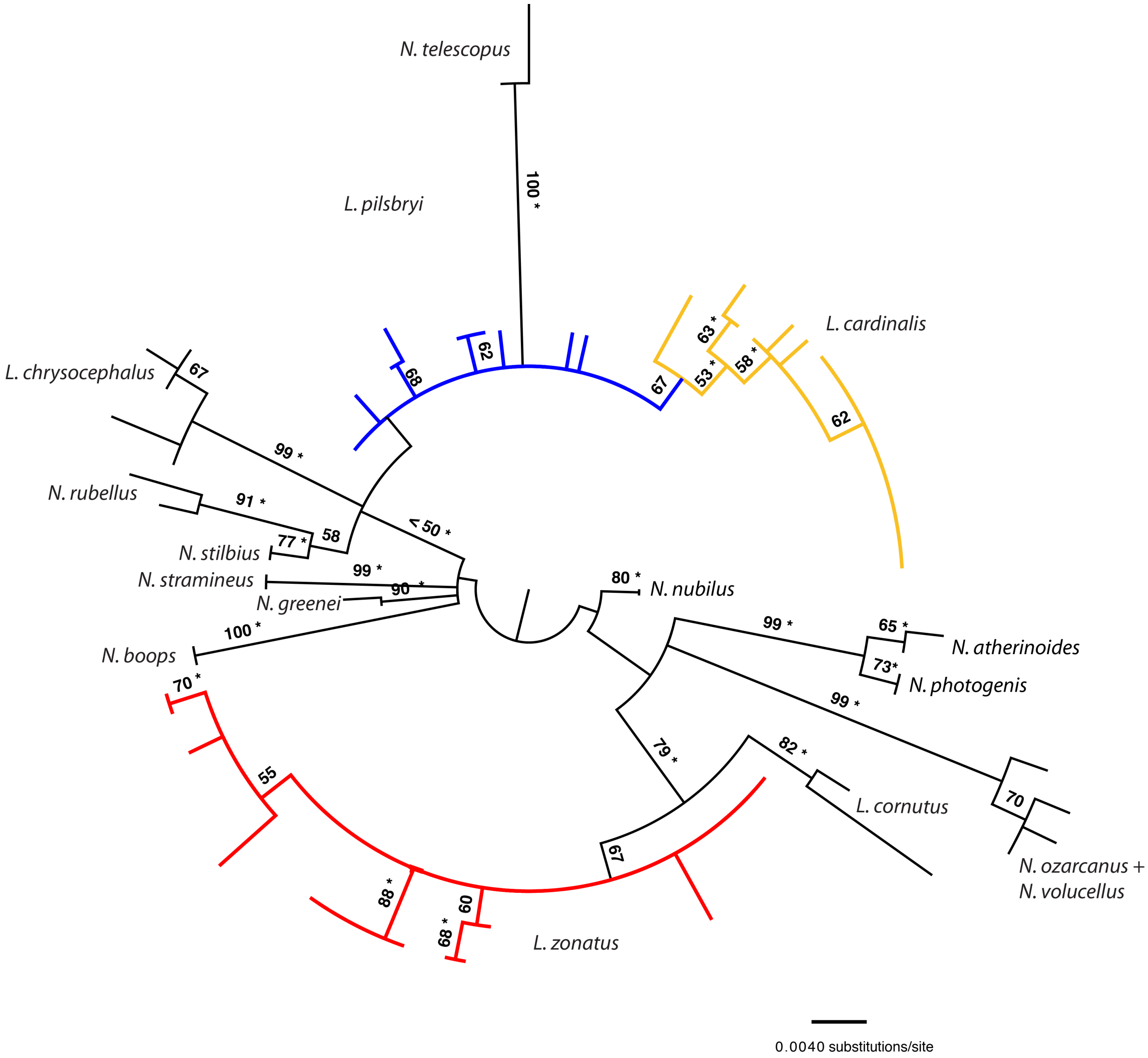


Figure 4.3 Rpsa maximum likelihood gene tree for the *Luxilus zonatus* species group. Yellow represents *L. cardinalis*, blue represents *L. pilsbryi*, and red represents *L. zonatus*. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -1180.587578.

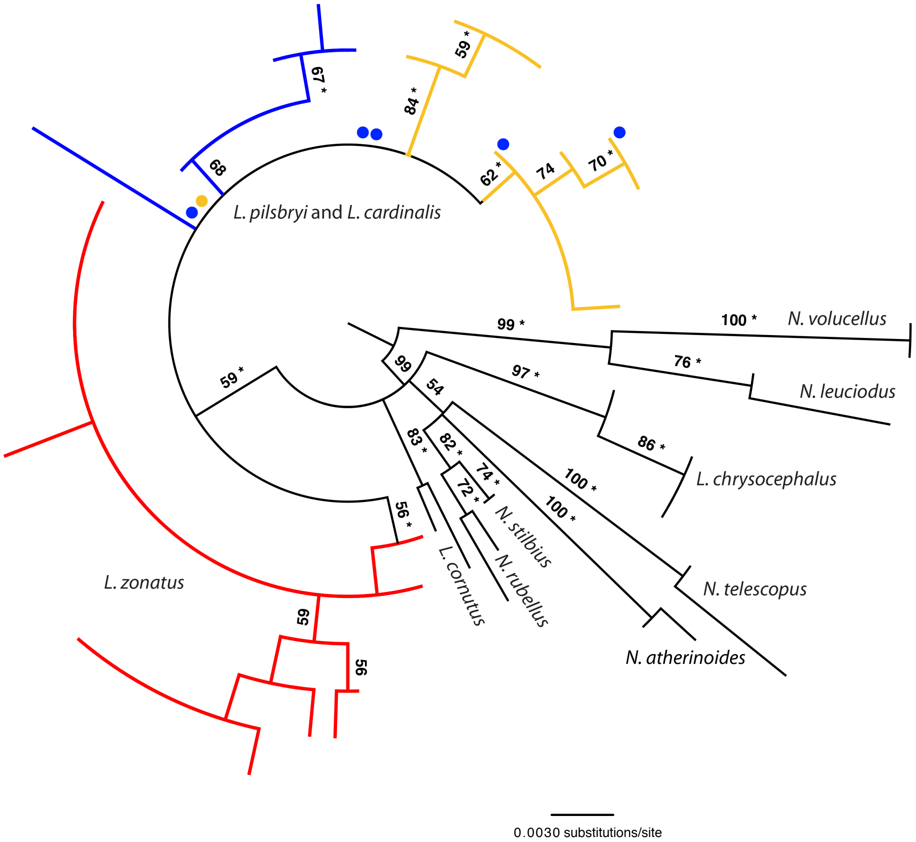


Figure 4.4 Rps3 maximum likelihood gene tree for the *Luxilus zonatus* species group. Yellow represents *L. cardinalis*, blue represents *L. pilsbryi*, and red represents *L. zonatus*. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -1053.445861.

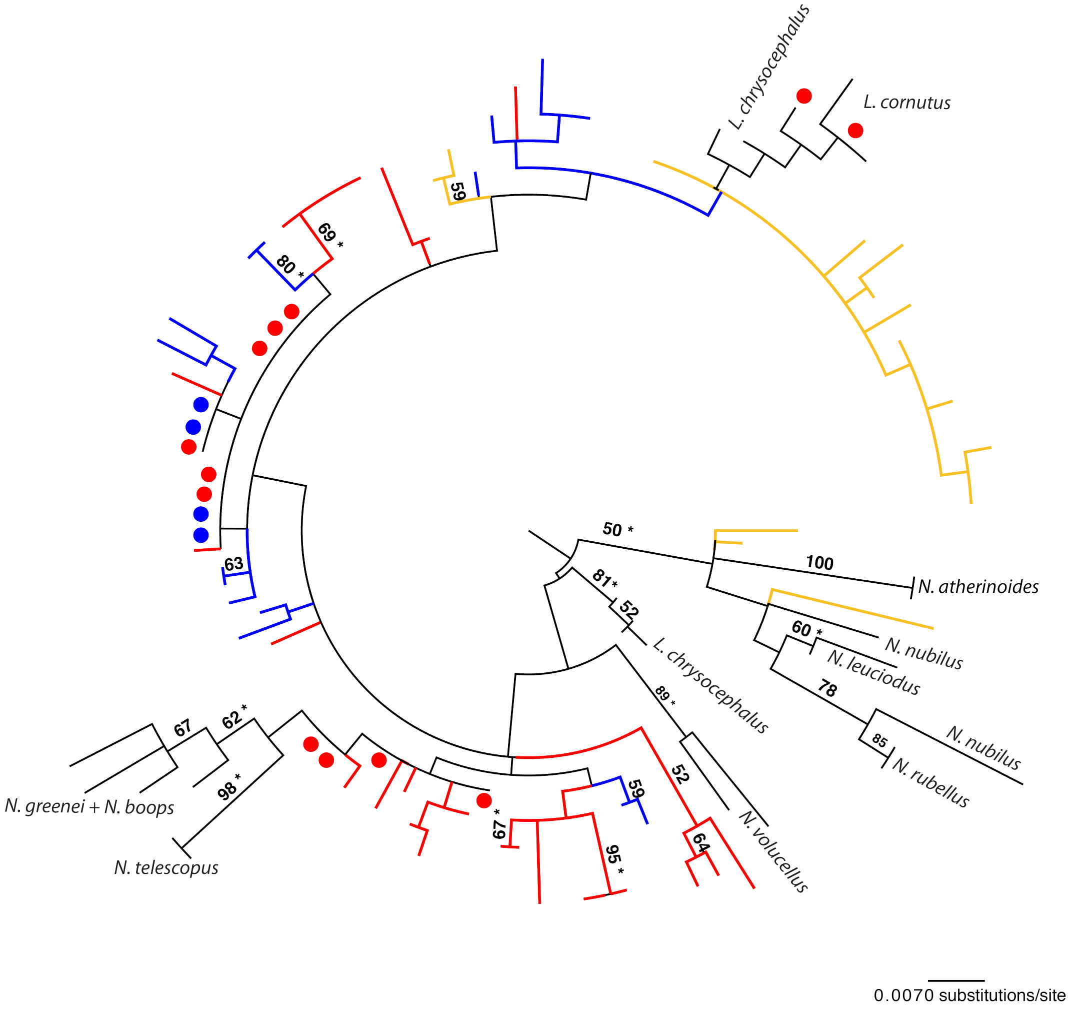


Figure 4.5 Ncl1 maximum likelihood gene tree for the *Luxilus zonatus* species group. Yellow represents *L. cardinalis*, blue represents *L. pilsbryi*, and red represents *L. zonatus*. Bootstrap percentages over 50% are shown; asterisks indicate nodes with a greater than 95% posterior probability in Bayesian analyses. Tree ln likelihood: -1515.619205.

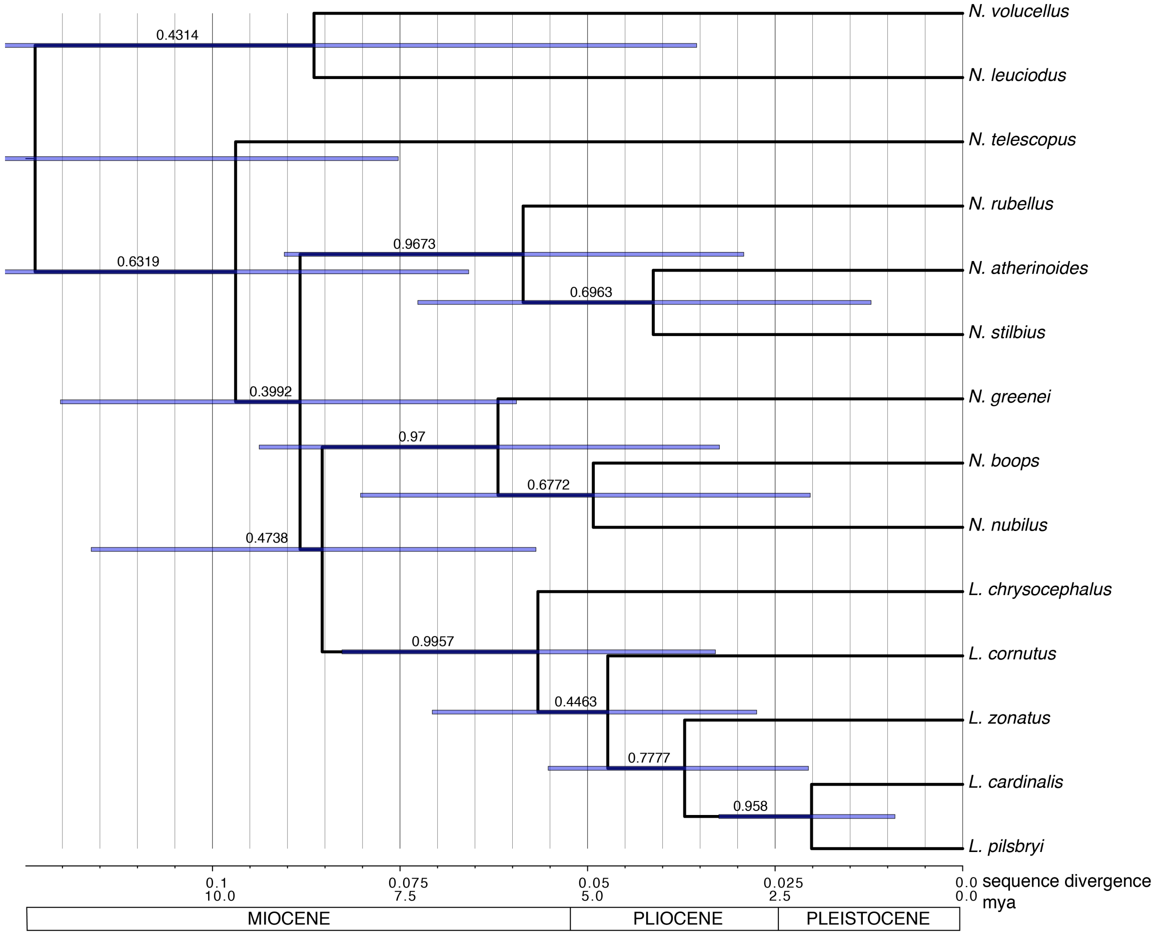


Figure 4.6 Species tree for *Luxilus zonatus* species group generated using *BEAST. Numbers above branches are posterior probabilities for nodes. Confidence intervals at nodes are 95% highest posterior densities. Scale is in percent sequence divergence and millions of years before the present, converted using a rate of 0.01% per lineage per million years.

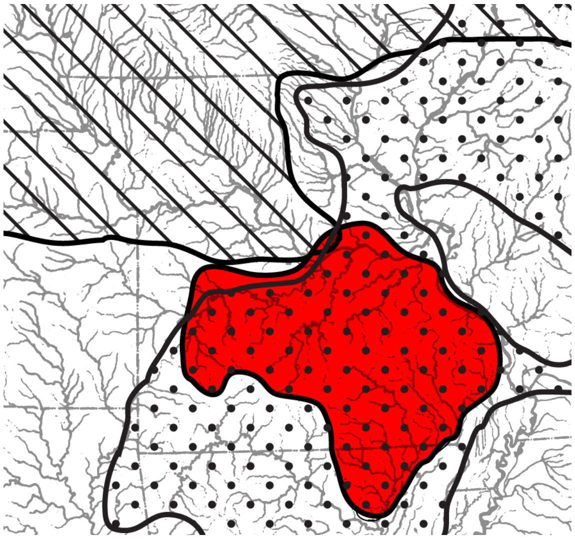


Figure 4.7 Ranges of *Luxilus zonatus* (red), *L. chrysocephalus* (dots) and *L. cornutus* (diagonal lines.)

CHAPTER 5

Detecting pseudocongruence: a test of three geographically congruent divergent taxon pairs in the Ozark highlands

Introduction

Historical biogeographers seek to explain the distributions of species across the globe by recourse to the past geological and biological events which have shaped their distribution. To do this, common patterns among species are sought, which may be explained through the action of common mechanisms which ideally can be identified by consulting the geological record. The discovery of common patterns among species should thus be a source of joy to the historical biogeographer; this is no longer always the case, however, now that the phenomenon known as pseudocongruence (Cunningham and Collins 1994, Donoghue and Moore 2003) has been identified. Pseudocongruence occurs when divergence events occurring at different times produce identical patterns of species-area relationships. It is potentially quite common in areas where a divergence can recur multiple times, with episodes of dispersal in between each divergence event producing new fodder for the divergence event to act upon. Identifying pseudocongruent events is thus necessary in a historical biogeographical analysis if one is to avoid oversimplifying the histories of the taxa being examined. Since pseudocongruence involves identical spatial relationships but different temporal relationships, methods which allow one to precisely determine the temporal components of phylogenies should be useful in detecting its presence in the biogeographical history of a region. One such method is *BEAST, a method of Bayesian species tree analysis which also produces estimates of divergence times. A second method is Hierarchical Approximate Bayesian Computation (HABC) (Hickerson et al. 2006, Huang et al. 2011). HABC uses hyperparameters which characterize multiple phylogenetic datasets simultaneously by describing various aspects of the differences in their divergence times, such as the number of different divergence times across all phylogenies. The most recent version of HABC implemented in the program msBayes allows for the use of multilocus data, allowing for a direct comparison with the results of a *BEAST analysis.

For this study, divergence events known from the fish fauna of the Central Highlands were searched to find a group of various pairs of sister taxa belonging to different clades, all showing the same geographical divergence (see Chapters 2 and 3). Sister taxa without any subsequent mitochondrial divergence of either sister pair were used, because these should retain the most evidence about the timing of their divergence of any kind of taxon pair.

Three taxon pairs were identified in the Ozark Mountains showing similar divergence patterns: two sister populations of *Etheostoma blennioides*, one found in the Arkansas River drainage, and the other in the Little Red River drainage; two sister populations of *Etheostoma zonale*, one found in the Upper Arkansas and Ouachita river drainages, and the other in the Little Red River drainage, and the sister species pair *Luxilus cardinalis*, found in the Arkansas and Ouachita river drainages, and *Luxilus pilsbryi*, found in the White and Little Red river drainages. The distributions of each of these taxon pairs are shown in Figure 5.1. Though the ranges of the three taxon pairs are not identical, they all show a split between the Arkansas and Little Red river drainages.

Divergence times for each of the three taxon pairs were determined using *BEAST. Hierarchical Approximate Bayesian Computation (Hickerson et al. 2006, Huang et al. 2011) was also used to test for pseudocongruence in these taxa. The HABC method creates simulated datasets by randomly drawing parameters from a prior distribution, and comparing the hyperparameters of these datasets to those of the observed dataset; those which are closest are kept. The method thus determines the likeliest hyperparameters, and thus the likeliest number of different divergences and divergence times, to have produced the observed data set. Unlike *BEAST, msBayes does not just determine divergence times, but it also directly determines how many separate divergence events are most likely to be represented among the set of divergences input to the program. It is thus perhaps the most powerful tool available to detect pseudocongruence.

Based on the divergence times seen in the prior phylogenetic analyses for each species, I hypothesized that *E. zonale* and *E. blennioides* would show evidence of divergence due to the same event, while *Luxilus cardinalis* and *L. pilsbryi* would show evidence of having diverged at an earlier time. I also hypothesized that the divergences would be the result of dispersal across a barrier which opened, closed, and then opened again, resulting in indications of population growth in at least one population in each of the divergent pairs subsequent to their divergence.

Methods

Sequences for *Etheostoma zonale*, *Luxilus cardinalis* and *Luxilus pilsbryi* used in this study were the same as those in Chapters 3 and 4. Sequences for *Etheostoma blennioides* were obtained from 6 localities in the Arkansas and Little Red River basins. Cytochrome *b* sequences from individuals of *Luxilus pilsbryi* that had introgressed with *L. cardinalis* were not used.

Etheostoma blennioides genomic DNA was extracted using QIAamp™ tissue extraction kits (Qiagen Inc.) according to the manufacturer's instructions. The complete mitochondrial cytochrome *b* gene was amplified using the polymerase chain reaction (PCR). PCR was performed in a total volume of 25 µL, containing 5 µL 5x Green GoTaq Flexi Buffer, 1.5 µL 25 mM $MgCl_2$, 0.5 µL 10 mM dNTPs, 0.03 nmol each of the forward and reverse primers, 0.125 µL GoTaq Flexi DNA polymerase, and approximately 0.5 µg of DNA, under the following thermocycler settings: initial denaturation at 95.0 ºC for 1 minute; 30 cycles of 95.0 ºC for 30 seconds, 53.0 ºC for 1 minute, and 72.0 ºC for 2 minutes; and a final extension at 72.0 ºC for 10 minutes, with the reaction terminating at 4 ºC. Cytochrome *b* was sequenced using the primers HA and LA of Schmidt et al. (1998). New internal primers were designed to allow complete sequencing of both strands: blen1f [5′ GTGCCACAGTCATCACCAATC 3′] and blen1r [5′ CCTGTTTCATGGAGAAAGAG 3′]. Amplified DNA was purified using 4 units of exonuclease 1 and 0.4 units of shrimp alkaline phosphatase per reaction. Automated sequencing was performed using Big Dye terminator cycle sequencing at the DNA Sequencing and Analysis Facility of the Biomedical Genomics Center at the University of Minnesota. Sequences were checked for accuracy and assembled using Sequencher 4.7 (Gene Codes Corporation). Cytochrome *b* sequences for individuals included in the analyses in Berendzen (2005) were obtained from the author.

Individuals of *Etheostoma blennioides* were also sequenced at four nuclear loci: ncl1, nedd4l, slco4a1, and stx5a. Each locus was amplified using the reaction conditions described for *E. zonale* in Chapter 3. The nedd4l locus included a length of approximately 20 to 30 thymines in a row, which made sequencing difficult. Internal primers were designed to permit clear sequencing of a portion of the intron:

nedd4l24ifblennioides [5′ TTGAGTWGTCTTAATCCCTGACCT 3′] and nedd4l24irblennioides [5′ TGCAAAAGTCATCCTTNGCTGT 3′].

Nuclear genes heterozygous at two or more sites were phased using one of three methods: sequences which were length heterozygotes were phased by eye in Sequencher (Sousa-Santos et al. 2005; Flot et al. 2006). Sequences which were not length heterozygotes were phased using the program PHASE (Stephens et al. 2001, Stephens and Scheet 2005) as implemented in the program DnaSP v5 (Librado and Rozas 2009). Homozygous sequences, sequences with only one heterozygous site, and sequences already phased due to a length heterozygosity were included as known sequences. For each run of the algorithm, the final portion of the run was set to be ten times as long as previous portions, and the algorithm was run five times for each locus; otherwise, default settings were used. If the phase of a sequence was reconstructed with a certainty threshold greater than 90 percent for all sites, this reconstruction was used in all subsequent analyses. Sequences which were not reconstructed by PHASE with this level of confidence were phased by PCR amplification using allele-specific primers (Petersson et al. 2003). Pairs of primers differing only at the last base at the 3′ end were designed, with a length sufficient for the melting temperature of the primer to match the melting temperature, within one or two degrees Celsius, of the regular amplification primer to be used at the other end of the sequence being phased. After allele-specific sequences were obtained, PHASE was re-run, using the sequences obtained using allele-specific sequencing as known haplotypes. If sequences which were previously reconstructed by PHASE with high confidence now showed a lower degree of confidence, allele-specific primers were also designed for these sequences. This cycle was repeated until all remaining unphased sequences were reconstructed by PHASE with high confidence.

All sequences for all genes will be deposited in Genbank.

Nuclear genes were tested for recombination using the phi test (Bruen et al. 2006) as implemented in Splitstree (Huson and Bryant 2006), as this test is suitable for detecting recombination both within populations of a single species, and across species.

A species tree analysis was conducted for *Etheostoma zonale and E. blennioides* using *BEAST (Heled and Drummond 2010). The terminals for the analysis were the

Upper Arkansas and Little Red river populations of *E. blennioides*, the Ouachita – Upper Arkansas and Little Red river populations of *E. zonale*, the Hiwassee River populations of *E. blennioides* and *E. zonale*, included to reduce the stem length of the *E. blennioides* and *E. zonale* branches, and *E. baileyi*, *E. barrenense*, *E. rupestre*, *E. simoterum*, *E. thalassinum*, and *E. variatum*. The loci used were cytochrome *b*, along with the four nuclear intron loci for which sequences were available for both *E. blennioides* and *E. zonale*: ncl1, nedd4l, slco4a1, and stx5a. Models for each locus were selected using MrModelTest (Nylander 2004). Cytochrome *b* was run unpartitioned for the analysis. All loci were run with empirical base frequencies. Molecular clock likelihood ratio tests for each locus were performed in PAUP* (Swofford 2002) to determine whether each locus should follow the strict clock model, or the relaxed clock uncorrelated exponential model. Default priors and operators were used. Missing data was used to represent outgroup taxa at loci for which a sequence was not obtainable. Several runs were performed for each analysis to assess convergence between runs, and then combined to obtain the final species tree. For *Luxilus*, the existing species tree analysis (Chapter 4) was compared to this analysis; the analyses could not be combined into one because of the amount of divergence between cyprinids (*Luxilus*) and percids (*Etheostoma*).

Hierarchical Approximate Bayesian Computation (Hickerson et al. 2006, Huang et al. 2011) as implemented in the program msBayes was used to test for simultaneity of the divergences in *Luxilus*, *Etheostoma blennioides*, and *E. zonale*. Because msBayes discards data at base positions including gaps, for *E. zonale*, only 1140 bp cytochrome *b* sequences were used. The mutation rate scalar for mitochondrial sequences was set at ten times the rate for nuclear sequences, based on the results of the species tree analyses in *BEAST. An initial run of 100,000 draws from the hyper-prior was performed to assess the suitability of the priors. Based on this run, the prior for the upper limit of tau (divergence time) was reset to 0.7. Using this new setting, 2,000,000 random draws from the hyper-prior were simulated. Based on the results of this second run, a third run of 2,000,000 draws from the hyper-prior was performed, with the number of tau classes (the number of distinct divergence events) set to two. Divergence times in tau were converted to time in years using a cytochrome *b* mutation rate of 2.0 ± 0.2 % sequence divergence per million years (Near and Benard 2004); rates for nuclear loci were scaled to this using the posterior relative rates obtained from the *BEAST analyses.

To test for population growth in each of the two populations of each species, the R2 Test (Ramos-Onsins and Rozas 2002) was used. The R2 test was performed on the cytochrome *b* sequence data and nuclear locus data for all six clades. The test was performed in DnaSP v5; to determine the significance of the test statistic for each test, 10,000 coalescent simulations were performed in DnaSP.

Extended Bayesian Skyline Plots (Heled and Drummond 2008) were obtained for each of the two clades of all species, for a total of six separate plots. In all analyses, substitution models, clock models, and trees were unlinked for all partitions. In *Etheostoma blennioides*, there was no variation among the ncl1, nedd4l, and stx5a sequences for the Arkansas population, nor among the nedd4l sequences for the Little Red population, so these loci could not be used for the analyses. All loci for all species were run using the HKY model and a strict clock. Empirical base frequencies were used for all loci. A rate of 2.0 ± 0.2 % sequence divergence per million years (Near and Benard 2004) was set as the prior for mitochondrial divergence rate. The prior for all other clock rates was uniform on the range [0,0.2]. The weights for the "demographic.populationMeanDist", "demographic.indicators", and "demographic.scaleActive" operators were changed, based on the recommendation in the Extended Bayesian Skyline Plot tutorial, set to 160, 320, and 60, respectively, for *Etheostoma zonale*, 40, 80, and 16 for the Arkansas population of *E. blennioides*, 80, 160, and 32 for *E. blennioides*, and 80, 160, and 30 for *Luxilus*. For each locus in all analyses, the weight for the "Substitution rates and heights" operator was increased to 15, and the weight for the kappa operator was increased to 2. The mean of the distribution of population sizes was set to 1. Each analysis was run for 50,000,000 generations.

Results

New cytochrome *b* sequences were obtained for seven individuals of *E. blennioides*. Sequences for the four nuclear introns were obtained for seven individuals from the Arkansas River population, and nine individuals from the Little Red River population. The numbers of polymorphic sites at each of these loci, along with the same numbers for the *E. zonale* and *Luxilus* populations are shown in Table 5.1. No evidence of recombination was found in any of the nuclear loci used in the *BEAST and msbayes analyses.

Models of evolution selected for the *BEAST analysis are shown in Table 5.2. The *BEAST species tree is shown in Figure 5.2. The tree was generated from 100,000,000 post-burnin generations spread across two separate runs, achieving convergence of parameters across runs, and satisfactory effective sample size (ESS) values for all key parameters in the combined runs. The 95 % highest posterior density of divergence times for *E. blennioides* overlaps somewhat with that for *E. zonale*. Using a percent sequence divergence of 2.0 % per million years (Berendzen 2005), the mean divergence date of the Little Red and Arkansas populations of *E. blennioides* is slightly less than 1,000,000 years ago, while that of the corresponding *E. zonale* populations is approximately 500,000 years ago. *Luxilus cardinalis* and *L. pilsbryi* have a mean divergence date of approximately 2,000,000 years ago. The 95 % HPD of divergence times for *Luxilus cardinalis* and *L. pilsbryi* (Figure 5.3) overlaps with that of *E. blennioides*, but not with that of *E. zonale*.

The first run of the msBayes analysis resulted in two separate divergences having the highest posterior probability. Divergence times in the second run were converted from units of tau to years using an average mutation rate per site per locus across all three divergent pairs of 3.33 x 10^{-}9 per year; this is based on the mean posterior rate values obtained for each locus and taxon in the *BEAST analyses, with the rate for cytochrome *b* for all taxa set to 2.0 x 10^{-}9 per year. The plots of posterior density for each divergence are shown in Figure 5.4. In the second run, the more recent divergence was very recent indeed: the modal divergence time was only 9,200 years ago, while the median divergence time is 151,000 years ago. This divergence corresponds to the *E.*

zonale and *E. blennioides* divergences. The older divergence, corresponding to the *Luxilus* divergence, has a modal divergence time of 1,170,000 years ago, and a median divergence time of 1,320,000 years ago. The more recent divergence forms a rather steep peak in the posterior distribution, while the older peak is very broad.

Tables 5.3 to 5.5 give the results of the R2 test for all divergent pairs at all loci sequenced. Neither *E. blennioides* population showed evidence of population growth at any locus. The Ouachita and Upper Arkansas population of *E. zonale* showed evidence of population growth at the eif3c locus; *L. cardinalis* showed evidence of population growth at cytochrome *b*, and *L. pilsbryi* showed evidence at population growth at the rpsa locus.

The extended Bayesian skyline plots for each population are shown in Figure 5.5. Table 5.6 gives the 95 % highest posterior densities for the number of population size changes in each population. A constant population size was rejected for both *L. cardinalis* and *L. pilsbryi*; the plots for both species show an increase in population size starting 500,000 to 200,000 years ago. For the four remaining populations, constant population size is not rejected, but for all populations except the Little Red River population of *E. blennioides*, the mode of the number of population size changes is one. The timing of the increase in population size varies from 300,000 years ago for the Arkansas population of *E. zonale*, to only 10,000 years ago for the Little Red River population of *E. zonale*.

Discussion

Given that it confounds events with distinct cause pseudocongruence has the potential to be one of the greatest threats to the accurate reconstruction of biogeographical history. The development of methods to detect its presence would therefore be of great help in the practice of biogeography. Using the three taxon pairs that diverge across the Arkansas–Little Red River drainage divide, I attempted to determine whether the use of multilocus methods of determining divergence times could fill this role.

The results of the *BEAST and msBayes analyses both support the existence of two separate, pseudocongruent events in the divergences of the three pairs of taxa examined. The analyses agree on which taxa took part in which divergences, but they do not agree on the timing of the divergences. *BEAST sets the more recent divergence, between the Arkansas and Little Red river populations of *E. blennioides* and *E. zonale*, between 500,000 to less than 1,000,000 years ago, and the divergence between the two *Luxilus* species at 1,000,000 to less than 3,500,000 years ago. msBayes, however, makes the *E. zonale* and *E. blennioides* divergences much more recent: the modal divergence time is less than 10,000 years ago, and the median divergence time is still only 151,000 years ago. The *Luxilus* divergence time does overlap between the two analyses: the modal divergence time in msBayes is 1,170,000 years ago, and the median is 1,320,000 years ago, overlapping with the most recent portion of the 95 % HPD interval for divergence time in the *BEAST analysis. The spread of the posterior for divergence time of *Luxilus* is quite broad in the msBayes analysis, as it is in the *BEAST analysis; it is therefore heartening that msBayes was still able to reject a single divergence event for all three taxa. The broad span in divergence time estimates for *Luxilus* may be due to the limited data available (only four loci) or to the quality of the data used; the ncl1 locus, with its poorly resolved phylogeny, is an obvious culprit. The difference in divergence time estimates between *BEAST and msBayes, on the other hand, may be an artifact of the way each estimates divergence times; the strong leftward peak of the younger divergence time estimate in the msBayes plot may indicate that msBayes is less likely than *BEAST to conclude that lineages have actually diverged when there is little sequence divergence between them. Given that the purpose of msBayes is solely to determine the number and timing of various divergence events, while that of *BEAST is primarily to reconstruct the topology of species trees, with divergence times secondary, the younger divergence times obtained by msBayes may command more weight.

The extended Bayesian skyline plots do not aid directly in determining divergence times, but they do give additional information about what was happening to each population after they diverged, and thus might shed some light on the nature of those divergences. The plots both show an increase in population size for the two species of *Luxilus* over the past several hundred thousand years. This suggests that their ranges might have been much smaller when they originally diverged. The increase in population size appears to start several hundred thousand years more recently than the estimated time of divergence for the two species, so the population growth does not appear to be an indication that the species originated through dispersal across a barrier. Constant population size cannot be rejected for the remaining populations, but an increase in size is still most likely for all of them except the Little Red River population of *E. blennioides*, which shows a constant, small population size throughout its existence. The increase in the size of the Arkansas population of *E. zonale* starts 250,000 years ago; this is well past the divergence time estimated by *BEAST, but earlier than the median time estimated by msBayes. Were an increase in the number of loci used in the analyses to confirm that the population expansion and divergence occurred at the same time, this would be good evidence that the Arkansas population of *E. zonale* originated through dispersal from the Little Red River population. The other two populations which show evidence of growth, the Arkansas River population of *E. blennioides* and the Little Red River population of *E. zonale*, show it beginning much more recently, though still within the upper limit for the dispersal time found in the msBayes analysis. A co-ordinated dispersal event in which *E. zonale* and *E. blennioides* were able to enter the Arkansas River drainage through a stream capture event is not ruled out by these data. It is notable that the extended Bayesian skyline plots detected a much stronger signal of population growth than did the R2 tests of the various loci; this may be due to the amplification of signal in the multilocus extended Bayesian skyline plot method over the single locus R2 test.

A flaw in this study is its dependence on a single calibration of sequence divergence to time, based on the centrarchid study of Near and Benard (2004). Clearly, determining whether two divergence events that show a different degree of sequence divergence represent the same event is entirely dependent on being able to accurately calibrate the sequence divergence. It would be useful to be able to easily calculate the minimum change in the percent sequence divergence calibrations for different pairs of taxa which would result in msBayes recovering their divergences as a single event.

Another flaw is that there is no independent method of testing for the existence of pseudocongruence. It would be just as incorrect to assume that any taxa which msBayes says did not diverge at the same time are pseudocongruent as it would be to say that any taxa which show the same geographical pattern diverged due to the same event. In order to resolve this problem, if it is indeed possible, it will be necessary to determine at what point the differences in rate of evolution between two different taxa are simply too great if we are to assume that they diverged at the same time, and thus can reliably conclude that they are in fact pseudocongruent. Doing so might be possible in taxa which have excellent fossil records. Although this is not the case in *Luxilus* and *Etheostoma*, it is notable, however, that the two taxon pairs whose divergences are recovered as the same event show more similarity in range than the third taxon pair (Figure 5.1). The member of the third taxon pair that is found in the Little Red River, *Luxilus pilsbryi*, is also found in the White River drainage, while the other two taxon pairs' representatives in the Little Red River are confined to it. It is also notable that both *E. blennioides* and *E. zonale* include populations in the White River that are less closely related to the Arkansas and Little Red river populations of their respective species. Thus, the biogeographical pattern of all three taxon pairs is not identical, though the split between the Arkansas and Little Red rivers is seen in all three taxa. It may be the case that close examination of the ranges of potentially pseudocongruent taxa in this way may be sufficient to reveal which taxa are actually pseudocongruent; this seems a potentially fruitful avenue for further research. Ideally, such comparisons may allow for an independent test of which taxa are actually pseudocongruent, thus allowing one to determine whether or not the taxa determined by msBayes to have diverged at different times actually did diverge at different times, or if msBayes itself is in error.

Conclusion

Pseudocongruence is a phenomenon which makes accurate reconstruction of historical biogeographical events difficult. The methods used in this analysis, the use of *BEAST to estimate divergence times, and of msBayes to measure divergence times as well as directly determine whether a set of divergences have or have not occurred due the same event, show promise in detecting pseudocongruence. Using these methods, I determined that *Etheostoma blennioides*, *E. zonale*, and the *Luxilus pilsbryi – L. cardinalis* species pair, all of which show divergent populations on either side of the Arkansas – Little Red river drainage divide, actually represent two distinct divergence events, one for the two *Etheostoma* species, and one for *Luxilus*. Extended Bayesian skyline plots suggested that co-ordinated dispersal may have been the method of diversification at least for the *Etheostoma* species. These results suggest that, where sufficient multilocus genetic data are available, pseudocongruence can actually be detected using the appropriate methods. This should then allow for more accurate biogeographical reconstructions of past divergence events, as researchers will not be conflating data from separate events in the past.

Appendix 5.1 Specimens Examined. The specimens used in this study are listed with the number of specimens from each site in parentheses and collection locality. All specimens are uncatalogued specimens in the Bell Museum of Natural History frozen tissue collection, unless otherwise indicated. Institutional abbreviations are as follows: MMNS = Mississippi Museum of Natural Science PLU= Pacific Lutheran University Natural History Collection SLU = Saint Louis University Ichthyological Collection UAIC = University of Alabama Ichthyological Collection YFTC = Yale Fish Tissue Collection.

Etheostoma blennioides

Little Red River clade

Little RedA (2) Middle Fork Little Red River, Searcy Co., AR, SLU uncat.
Little RedB (2) Archey's Fork Little Red River, Van Buren Co., AR, SLU uncat.
Little RedC (3) Middle Fork Little Red River, Van Buren Co., AR.
Little RedD (2) Little Red River, Van Buren Co., AR.

Arkansas River clade

ArkansasA (2) Shoal Creek, Newton Co., MO.
ArkansasB (5) Shoal Creek, Cherokee Co., KS, UAIC 10072.13.

Hiwassee River clade

HiwasseeA (1) Valley River, Cherokee Co, NC.

Etheostoma zonale

Little Red River clade

Little RedA (5) Archey Fork Little Red River, Van Buren Co., AR.
Little RedB (5) Middle Fork Little Red River, Van Buren Co., AR.

Ouachita and Upper Arkansas clade

Ouachita and Upper ArkansasA (7) Caddo River, Montgomery Co., AR.
Ouachita and Upper ArkansasB (3) Caddo River, Pike Co., AR.
Ouachita and Upper ArkansasC (5) Elk River, McDonald Co., MO.
Ouachita and Upper ArkansasD (3) South Fork Fourche La Fave River, Perry Co., AR, YFTC 1685-1687.
Ouachita and Upper ArkansasE (2) Mulberry River, Johnson Co., AR, SLU 840.03.
Ouachita and Upper ArkansasF (1) Ouachita River, Montgomery Co, AR.
Ouachita and Upper ArkansasG (3) Saline River, Grant Co., AR.
Ouachita and Upper ArkansasH (5) Spring River, Jasper Co., MO.

Hiwassee River clade

Hiwassee (10) Valley River, Cherokee Co., NC.

Etheostoma baileyi (1) Red Bird River, Clay Co., KY.
AF288423

Etheostoma barrenense (1) Drakes Creek, Warren Co., KY.
AF288424

Etheostoma blennioides (1) Valley River, Cherokee Co., NC.

Etheostoma rupestre (1) Cahaba River, Bibb CO., AL.
AF288442

Etheostoma simoterum (1) Otter Creek, Wayne Co., KY.
AF288445

Etheostoma thalassinum (1) Mountain Creek, Greenville Co., SC.
AF288448

Etheostoma variatum (1) Walhonding River, Coshocton Co., OH.
AF289266

Table 5.1 Loci used for each taxon in the msBayes analysis, with sample size for each of the two populations, and the sequence length of each locus.

Taxon	Locus	Sample Size: Arkansas population	Sample Size: Little Red population	Sequence Length	Polymorphic Sites
E. blennioides	cyt b	7	9	1102	20
	ncl1	14	18	582	4
	nedd4l	14	18	453	1
	slco4a1	14	18	882	15
	stx5a	14	18	695	11
E. zonale	cyt b	14	8	1140	33
	eif3c	18	16	408	4
	ncl1	18	16	572	9
	nedd4l	20	16	453	11
	pabpc4	18	16	399	7
	s7	18	12	526	17
	slco4a1	18	16	882	27
	stx5a	20	16	705	9
Luxilus	cyt b	11	15	1135	88
	ncl1	17	22	403	24
	rps3	20	22	391	10
	rpsa	20	22	331	15

Table 5.2 Models of evolution and molecular clock test results for each locus in the *BEAST analysis.

Locus	Model	Clock Rejected?
cyt b	GTR+I+G	N
ncl1	HKY+G	N
nedd4l	HKY+G	N
slco4a1	GTR+G	Y
stx5a	HKY+G	N

Table 5.3 Results of the R2 test for population growth for both populations of *E. blennioides* at all loci. "NA" indicates loci with no variability.

		cyt *b*	ncl1	nedd4l	slco4a1	stx5a
Arkansas	R2	0.2464	NA	NA	0.1275	NA
	p-value	0.61827			0.14907	
	n	7			14	
Little Red	R2	0.2292	0.1259	NA	0.1735	0.1765
	p-value	0.77353	0.15644		0.83	0.859
	n	9	18		18	18

Table 5.4 Results of the R2 test for population growth for both populations of *E. zonale* at all loci. Significant results (indicating population growth) are shown in bold.

		cyt *b*	eif3c	ncl1	nedd4l	pabpc4	s7	slco4a1	stx5a
Arkansas	R2	0.1462	**0.1046**	0.1071	0.1521	0.1516	0.1247	0.161	0.1089
	p	0.524	**0.0068**	0.0981	0.6131	0.5843	0.35	0.774	0.1087
	n	14	18	18	20	18	18	18	20
Little Red	R2	0.1768	0.1344	0.1428	0.1187	0.15	0.1455	0.1766	0.2421
	p	0.10829	0.1792	0.3582	0.1749	0.4213	0.1550	0.831	0.6667
	n	8	16	16	16	16	12	16	16

Table 5.5 Results of the R2 test for population growth for *Luxilus* at all loci. Significant results (indicating population growth) are shown in bold.

		cyt *b*	ncl1	rps3	rpsa
L. cardinalis	R2	**0.1108**	0.1001	0.1903	0.0957
	p	**0.04505**	0.06907	0.88878	0.08551
	n	10	17	17	17
L. pilsbryi	R2	0.1073	0.1519	0.1253	**0.0728**
	p	0.146	0.746	0.35942	**<0.00001**
	n	15	22	22	22

Table 5.6 Modes and 95% highest posterior densities of number of population size changes in extended Bayesian skyline plots.

	Mode	Lower limit of HPD	Upper limit of HPD
E. blennioides Arkansas	1	0	2
E. blennioides Little Red	0	0	1
E. zonale Arkansas	1	0	3
E. zonale Little Red	1	0	2
L. cardinalis	1	1	3
L. pilsbryi	1	1	3

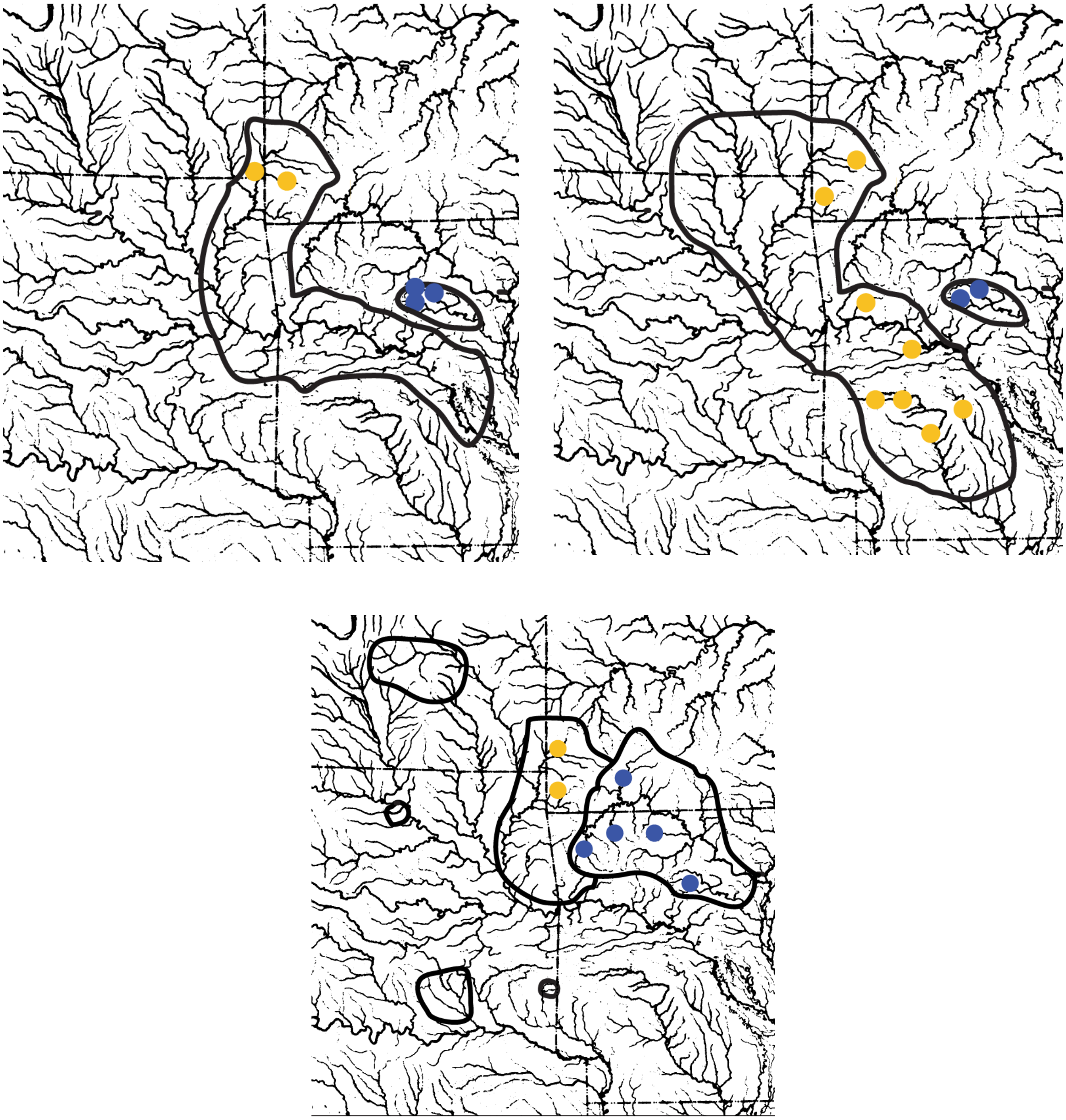


Figure 5.1 Distributions of the sister populations of *Etheostoma blennioides* (top left), *E. zonale* (top right) and *Luxilus* (bottom). Arkansas River populations shown in yellow; Little Red river populations shown in blue. Dots indicate sampling localities.

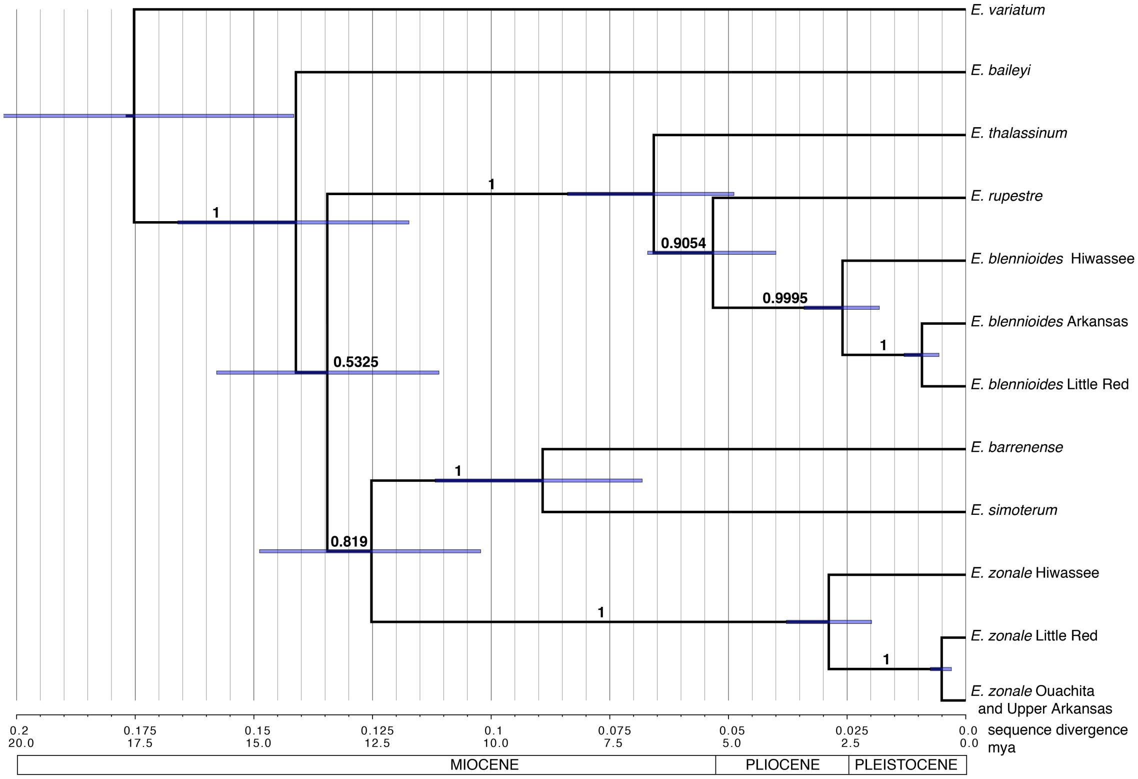


Figure 5.2 Species tree for *Etheostoma* species generated using *BEAST. Numbers above branches are posterior probabilities for nodes. Confidence intervals at nodes are 95% highest posterior densities. Scale is in percent sequence divergence and millions of years before the present, converted using a rate of 0.01% per lineage per million years.

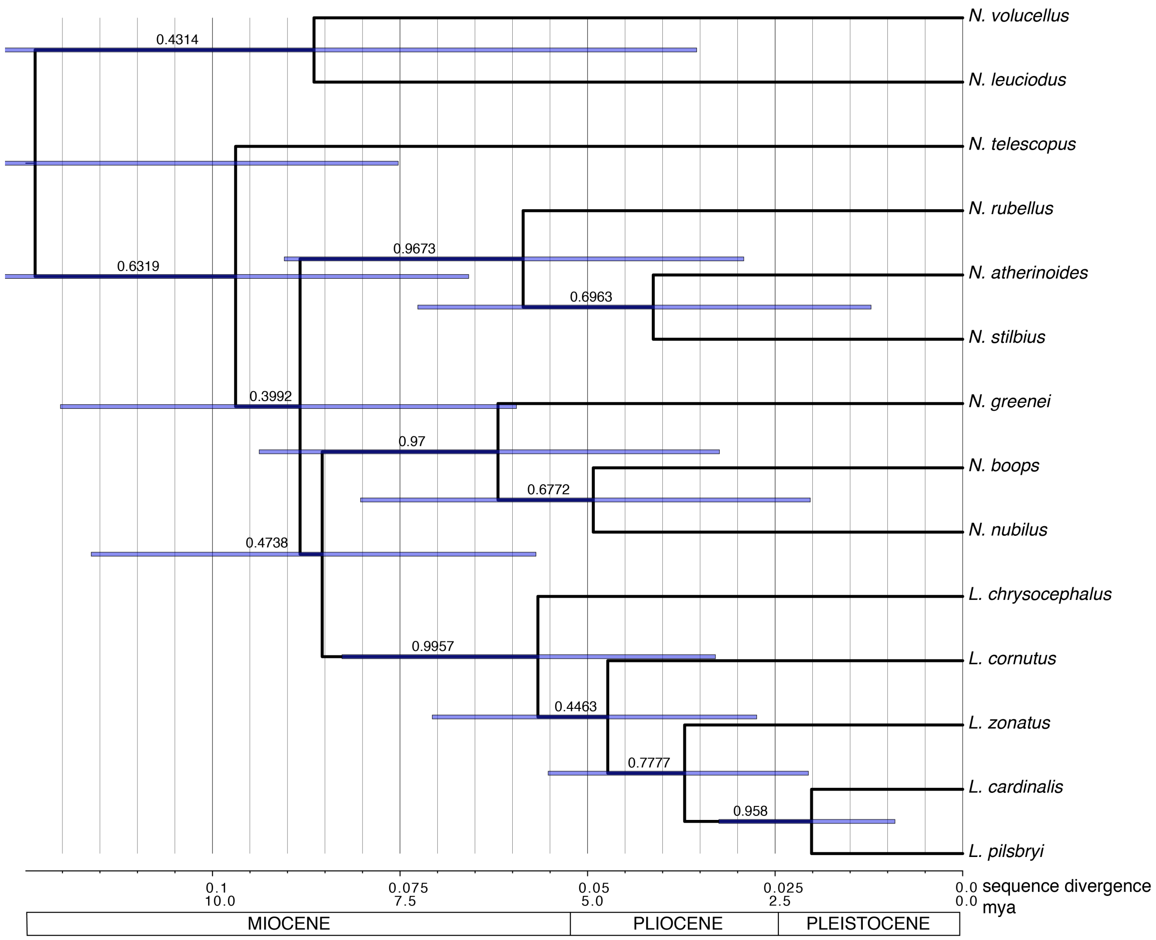


Figure 5.3. Species tree for *Luxilus* species generated using *BEAST. Numbers above branches are posterior probabilities for nodes. Confidence intervals at nodes are 95% highest posterior densities. Scale is in percent sequence divergence and millions of years before the present, converted using a rate of 0.01% per lineage per million years.

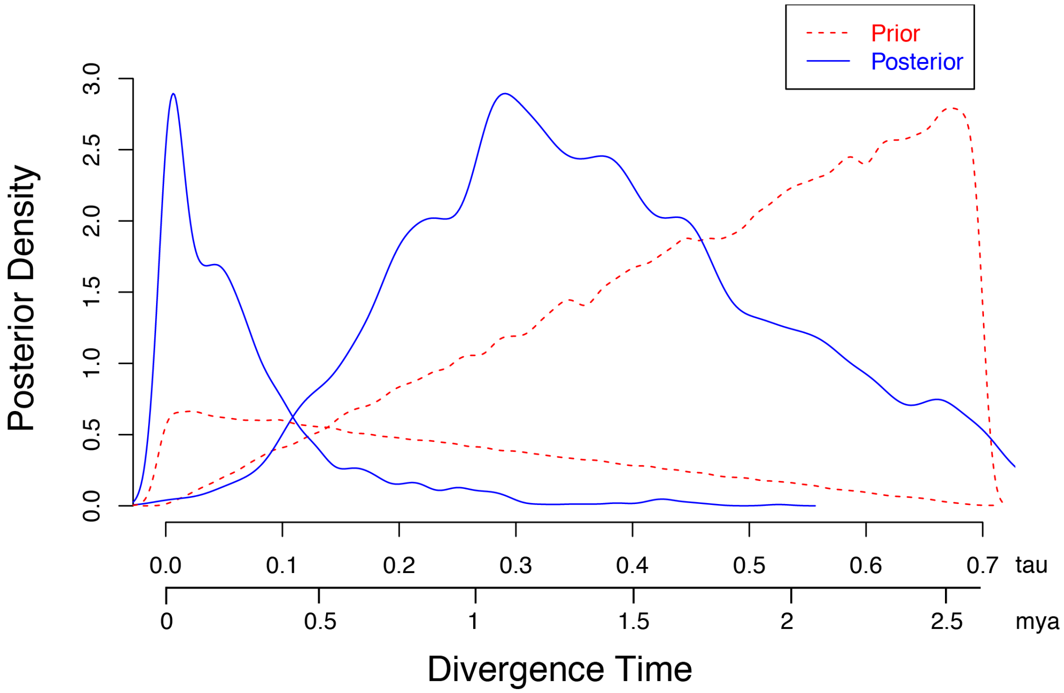


Figure 5.4 Estimates of divergence times from msBayes. The left-hand blue peak is the posterior distribution of the divergence times of *Etheostoma blennioides* and *E. zonale*, while the right-hand blue peak is the posterior distribution of the divergence times of *Luxilus*; the two red peaks are the prior distributions for the divergence times. The time scale assumes a generation time of one year.

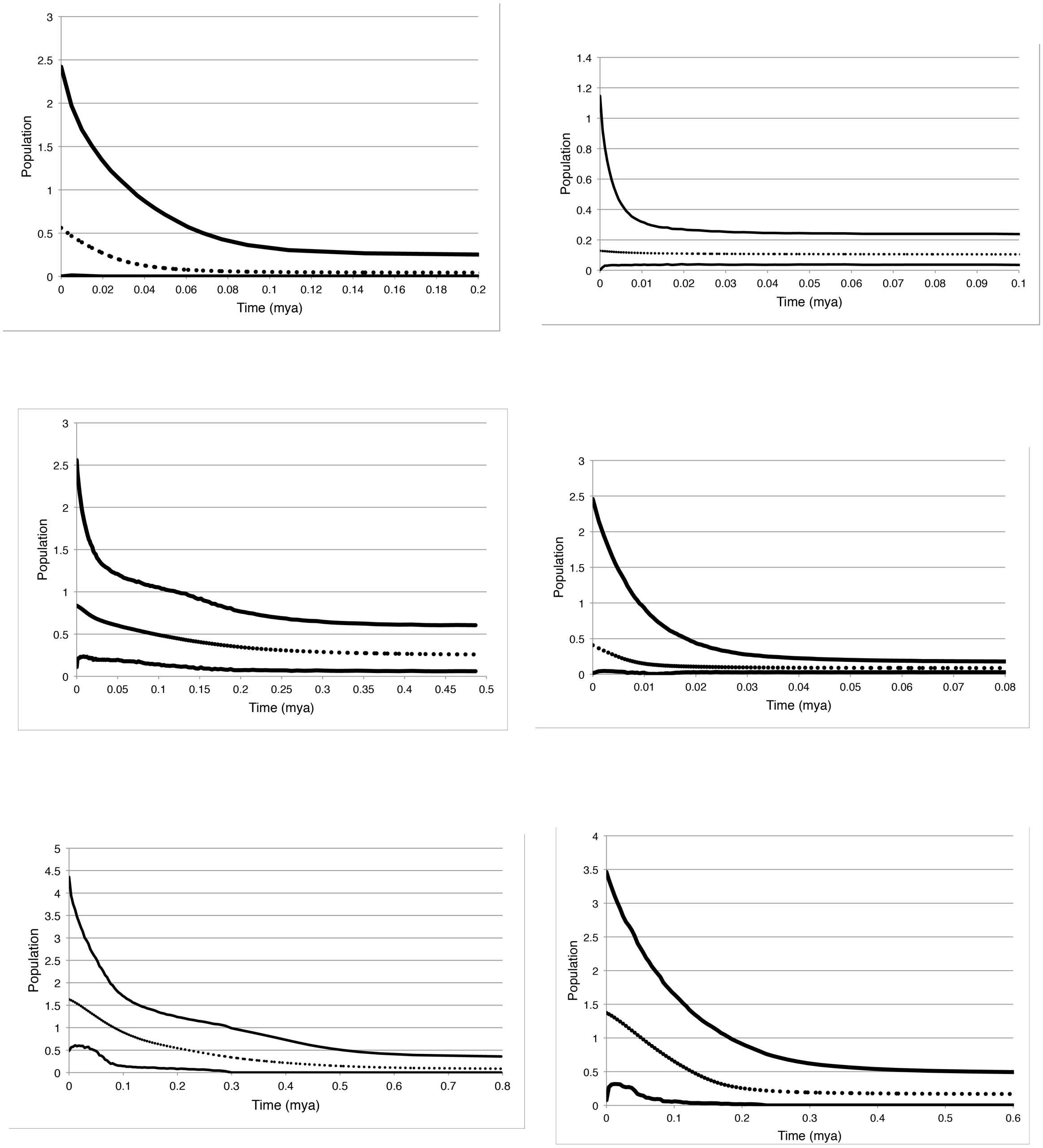


Figure 5.5 Extended Bayesian skyline plots for *Etheostoma blennioides* (top), *E. zonale* (middle) and *Luxilus* (bottom). Plots on the left are for Arkansas River populations; those on the right are for Little Red River populations. Solid lines are the 95% highest posterior densities for population size; the middle dotted line is the median population size. Population size is relative to generation time; for a generation time of one year, the scale is in millions of individuals. The time scale assumes a generation time of one year. Note that the scales of the horizontal and vertical axes differ between plots.

CHAPTER 6

Conclusion

The Central Highlands fish fauna is the result of a complex history of many kinds of divergence events. Vicariance, unique dispersal, and co-ordinated dispersal events have all intertwined to generate the rich biodiversity seen today. Hybridization and introgression have created a more literal intertwining, in the exchange of genetic material between previously-distinct Central Highlands taxa suggested by phylogenetic analyses. These events have generated additional biodiversity while making the reconstruction of the origins of that biodiversity a more difficult affair.

A comparative biogeographical analysis of the Central Highlands fish fauna using the PACT algorithm (Wojcicki and Brooks 2005) revealed that both common divergence events and unique dispersal events have contributed significantly to the diversity of the fauna. While many of the common divergence events are no doubt vicariance events, there is also area reticulation, with several areas taking part in multiple common divergence events. This reticulation suggests that co-ordinated dispersal may well be involved as well.

The *Etheostoma zonale* species group is shown to be highly geographically structured. A species tree analysis using *BEAST (Heled and Drummond 2010) recovers several divergence events which correspond to common divergence events from the PACT analysis. Other divergences show no such correspondence. Evidence from population growth analyses suggests that co-ordinated dispersal events may be involved in some of the divergences. One of the races of the *E. zonale* group is found to have introgressed with a second race in the upper Tennessee River drainage.

The previously-known relationships among the members of the *Luxilus zonatus* species group are confirmed by a new multi-locus analysis. Previously unknown, however, is the existence of at least four separate mitochondrial introgression events, involving all three members of the *L. zonatus* group and two other species of *Luxilus*. The introgression events appear to differ in both their time of occurrence and geographical extent.

Pseudocongruence (Donoghue and Moore 2003) was found to exist in at least one set of divergent fish clades in the Central Highlands. The evidence from these taxa suggests that divergence time analysis using Hierarchical Approximate Bayesian Computation (Hickerson et al. 2006, Huang et al. 2011) can detect pseudocongruence among divergences between taxa that are geographically congruent. These results of this analysis parallel that of the PACT analysis, which also found two distinct divergence events between these taxa.

Although the Central Highlands fish fauna is a historically complex system, the evidence of the present study suggests that improved prospects for the collection of genetic data at many independent loci, in conjunction with new methods of data analysis that can incorporate and effectively use such multilocus data, will allow its history to become increasingly better known over time. This in turn suggests that other biogeographical systems which are not so tractable as the Central Highlands fish fauna will also have further secrets to reveal, once the right tools are found and used in the right way. However, the increasing complexity being discovered in systems such as the Central Highlands poses a problem for many phylogenetic methods, which are not designed to deal with the level of complexity that exists in natural systems. Species tree reconstruction methods, for example, cannot yet incorporate introgression, which this study found to be commonplace. It is quite possible that increasingly complex methods of analysis will only uncover further complexity in the data; this in turn suggests that for many biogeographical systems, many of the details of their history can never actually be known.

Bibliography

Arnold, M. L., Sapir, Y., and Martin, N. H. 2008. Genetic exchange and the origin of adaptations: prokaryotes to primates. *Philosophical Transactions of the Royal Society B* **363**: 2813-2820.

Berendzen, P. B. Z. 2005. Phylogeography of Central Highlands fishes: discovering cryptic diversity and ancient drainage patterns in North America. Ph. D. dissertation, University of Minnesota.

Berendzen, P. B., Gamble, T., and Simons, A. M. 2008. Phylogeography of the bigeye chub *Hybopsis amblops* (Teleostei: Cypriniformes): early Pleistocene diversification and post-glacial range expansion. *Journal of Fish Biology* **73**: 2021-2039.

Berendzen, P. B., Simons, A. M., Wood, R. M., Dowling, T. E., and Secor, C. L. 2008. Recovering cryptic diversity and ancient drainage patterns in eastern North America: Historical biogeography of the *Notropis rubellus* species group (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution* **46**: 721-737.

Berendzen, P. B., Simons, A. M., and Wood, R. M. 2003. Phylogeography of the northern hogsucker, *Hypentelium nigricans* (Teleostei: Cypriniformes): genetic evidence for the existence of the ancient Teays River. *Journal of Biogeography* **30**: 1139-1152.

Boschung, H. T., Jr., and Mayden, R. L. 2004. Fishes of Alabama. Smithsonian Books, Washington, D. C.

Bossu, C. M., and Near, T. J. 2009. Gene trees reveal repeated instances of mitochondrial DNA introgression in orangethroat darters (Percidae: *Etheostoma*). *Systematic Biology* **58**: 114-129.

Brooks, D. R. 1985. Historical ecology: a new approach to studying the evolution of ecological associations. *Annals of the Missouri Botanical Garden* **72**: 660-680.

Brooks, D.R. 1990. Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. *Systematic Zoology* **39**: 14-30.

Brown, J. M., Hedtke, S. M., Lemmon, A.R., and Moriarty Lemmon, E. 2010. When trees grow too long: investigating the causes of highly inaccurate Bayesian branch-length estimates. *Systematic Biology* **59**: 145-161.

Bruen, T. C., Philippe, H., and Bryant, D. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665-2681.

Burr, B. M., and Warren, M. L., Jr. 1986. A distributional atlas of Kentucky fishes. Kentucky Nature Preserves Commission Scientific and Technical Series 4. Frankfort, KY.

Cavender, T. M. 1986. Review of the fossil history of North American freshwater fishes. In: C. H. Hocutt and E. O. Wiley (eds.) The zoogeography of North American freshwater fishes. John Wiley & Sons, New York, NY. pp. 699-724.

Ceas P.A., and Page L.M. 1997. Systematic studies of the *Etheostoma spectabile* complex (Percidae; subgenus *Oligocephalus*), with descriptions of four new species. *Copeia* **1997**: 496–522.

Chow, S., and Hazama, K. 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology* **7**: 1247-1263.

Cooper, J. E. 1980. Egg, Larval, and Juvenile Development of Longnose Dace, *Rhinichthys cataractae*, and River Chub, *Nocomis micropogon*, with notes on their hybridization. *Copeia* **1980**: 469-478.

Cope, E. C. 1868. On the distribution of fresh-water fishes in the Allegheny region of southwestern Virginia. *Journal of the Academy of Natural Sciences of Philadelphia* **2nd Series 6**: 207-247.

Cunningham, C. W., and T. M. Collins.1994. Developing model systems for molecular biogeography: vicariance and interchange in marine invertebrates. In: Schierwater, B., B. Streit, P. Wagner, and R. DeSalle (eds.) *Molecular ecology and evolution: approaches and applications*. Basel, Switzerland: Birkhauser Verlag. pp. 405-433.

Dominguez-Dominguez, O., I. Diadrio, and G. Perez-Ponce de Leon. 2006. Historical biogeography of some river basins in central Mexico evidenced by their goodeine freshwater fishes: a preliminary hypothesis using secondary Brooks parsimony analysis. *Journal of Biogeography* **33**: 1437-1447.

Donoghue, M. J., and B. R. Moore. 2003. Toward an integrative historical biogeography. *Integrative and Comparative Biology* **43**: 261–270.

Dowling, T. E., and Naylor, G. J. P. 1997. Evolutionary relationships of minnows in the genus *Luxilus* (Teleostei:Cyprinidae) as determined from cytochrome *b* sequences. *Copeia* **1997**: 758-765.

Duvernell, D. D., and Aspinwall, N. 1995. Introgression of *Luxilus cornutus* mtDNA into allopatric populations of *Luxilus chrysocephalus* (Teleostei: Cyprinidae) in Missouri and Arkansas. *Molecular Ecology* **4**: 173-181.

Egge, J. J. D., and Simons, A. M. 2009. Molecules, morphology, missing data and the phylogenetic position of a recently extinct madtom catfish (Actinopterygii: Ictaluridae). *Zoological Journal of the Linnean Society* **155** 60-75.

Eisenhour, D. J., and Piller, K. R. 1997. Two new intergeneric hybrids involving *Semotilus atromaculatus* and the genus *Phoxinus* with analysis of additional *Semotilus atromaculatus–Phoxinus* hybrids. *Copeia* **1997**: 204-209.

Eldredge, N., and Cracraft, J. 1980. Phylogenetic patterns and the evolutionary process. Columbia University Press, New York.

Erwin, T.L. 1979. Thoughts on the evolutionary history of ground beetles: hypotheses generated from comparative faunal analyses of lowland forest sites in temperate and tropical regions. In: Erwin, T. L., G. E. Ball, and D. R. Whitehead (eds.) *Carabid beetles – their evolution, natural history, and classification*. The Hague: W. Junk. pp. 539–592.

Erwin, T.L.1981. Taxon pulses, vicariance, and dispersal: an evolutionary synthesis illustrated by carabid beetles. In: Nelson, G., and D. E. Rosen (eds.) *Vicariance biogeography – a critique*. New York: Columbia University Press. pp. 159–196.

Etnier, D. A., and Starnes, W. C. 1986. *Etheostoma lynceum* removed from the synonymy of *E. zonale* (Pisces: Percidae). *Copeia* **1986**: 832-836.

Etnier, D. A., and Starnes, W. C. 1993. The fishes of Tennessee. University of Tennessee Press, Knoxville, TN.

Etnier, D. A., and Williams, J. D. 1989. *Etheostoma (Nothonotus) wapiti* (Osteichthyes: Percidae), a new darter from the southern bend of the Tennessee River system in

Alabama and Tennessee. *Proceedings of the Biological Society of Washington* **102**: 987-1000.

Flot, J.-F., Tillier, A., Samadi, S., and Tillier, S. 2006. Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes* **6**: 627-630.

Folinsbee, K. E., and D. R. Brooks. 2007. Miocene hominoid biogeography: pulses of dispersal and differentiation. *Journal of Biogeography* **34**: 383-397.

Fowler, H. W. 1904. Notes on fishes from Arkansas, Indian Territory and Texas. *Proceedings of the Academy of Natural Sciences* **56**: 242-249.

Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696-704.

Halas, D. D. Zamparo, and D. R. Brooks. 2005. A historical biogeographical protocol for studying biotic diversification by taxon pulses. *Journal of Biogeography* **32**: 249-260.

Hay, O. P. 1881. On a collection of fishes from eastern Mississippi. *Proceedings of the U. S. National Museum* **3**: 488-515.

Heled, J. and Drummond, A. J. 2008. Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* **8**: 289.

Heled, J., and Drummond, A. J. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**: 570-580.

Hembree, D. I. 2006. Amphisbaenian paleobiogeography: evidence of vicariance and geodispersal patterns. *Palaeogeography Palaeoclimatology Palaeoecology* **235**: 340-354.

Hickerson, M. J., Stahl, E. A., and Lessios, H. A. 2006. Test for simultaneous divergence using approximate Bayesian computation. *Evolution* **60**: 2435-2453.

Hollingsworth, P. R., Jr., and Near, T .J. 2009. Temporal patterns of diversification and microendemism in Eastern Highland endemic barcheek darters (Percidae: Etheostomatinae). *Evolution* **63**: 228-243.

Huang, W., Takebayashi, N., Qi, Y., and Hickerson, M. J. 2011. MTML-msBayes: Approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. *BMC Bioinformatics* **12**: 1.

Huelsenbeck, J. P., and Crandall, K. A. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* **28**: 437–466.

Huelsenbeck, J. P., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754-755.

Humphries, C. J. 2000. Form, space and time: which comes first? *Journal of Biogeography* **27**: 11-15.

Huson, D. H. and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**:254-267

Jenkins, R. E., and Burkhead, N. M. 1994. Freshwater fishes of Virginia. American Fisheries Society, Bethesda, MD.

Jordan, D. S. 1880. Descriptions of new species of North American fishes. *Proceedings of the U. S. National Museum* **2**: 235-241.

Jordan, D. S. 1885. A catalogue of the fishes known to inhabit the waters of North America, north of the Tropic of Cancer, with notes on the species discovered in 1883 and 1884. *Report of the U. S. Commissioner of Fish and Fisheries* **1885**: App. E, XXIV: 789-973.

Jordan, D. S., and Gilbert, C. H. 1886. List of the fishes collected in Arkansas, Indian Territory, and Texas, in September 1884, with notes and descriptions. *Proceedings of the U. S. National Museum* **9**: 1-25.

Keck, B. P, and Near, T. J. 2008. Assessing phylogenetic resolution among mitochondrial, nuclear, and morphological datasets in *Nothonotus* darters (Teleostei: Percidae). *Molecular Phylogenetics and Evolution* **46**: 708-720.

Keck, B. P., and Near, T. J. 2010. Geographic and temporal aspects of mitochondrial replacement in *Nothonotus* darters (Teleostei: Percidae: Etheostomatinae). *Evolution* **64**: 1410-1428.

Knowles, L. L. 2009. Estimating species trees: methods of phylogenetic analysis when there is incongruence among genes. *Systematic Biology* **58**: 463-467.

Larkin,M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. .2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947-2948.

Lee, D. S., Gilbert, C. R., Hocutt, C. H., Jenkins, R. E., McAllister, D. E., and Stauffer, J., Jr. (eds.) 1980. Atlas of North American freshwater fishes. North Carolina State Museum of Natural History, Raleigh, NC.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**: 913-925.

Librado, P., and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.

Lim, B. K. 2008. Historical biogeography of New World emballonurid bats (tribe Diclidurini): taxon pulse diversification. *Journal of Biogeography* **35**: 1385-1401.

Liu, L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* **24**: 2542-2543.

Maddison, D. R., and Maddison, W. P. 2003. MacClade 4: Analysis of phylogeny and character evolution. Version 4.06. Sinauer Associates, Sunderland, MA.

Mayden, R. L. 1985. Biogeography of Ouachita highland fishes. *The Southwestern Naturalist* **30**: 195-211.

Mayden, R. L. 1987a. Historical ecology and North American highland fishes: a research program in community ecology. In: W. J. Matthews and D. C. Heins (eds.) Community and evolutionary ecology of North American stream fishes. University of Oklahoma Press, Norman, OK. pp. 210-222.

Mayden, R. L. 1987b. Pleistocene glaciation and historical biogeography of North American central-highland fishes. In: W. C. Johnson (ed.) Quaternary environments of Kansas. Kansas Geological Survey, Guidebook Series 5. Lawrence, KS. pp. 141-151.

Mayden, R. L. 1988a. Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. *Systematic Zoology* **37**: 329-355.

Mayden, R. L. 1988b. Systematics of the *Notropis zonatus* species group, with description of a new species from the Interior Highlands of North America. *Copeia* **1988**: 153-173.

Mayden, R. L. 2010. Systematics of the *Etheostoma punctulatum* species group (Teleostei: Percidae), with descriptions of two new species. *Copeia* **2010**: 716-734.

Mayden, R. L., Simons, A. M., Wood, R. M., Harris, P. M., Kuhajda, B. R. 2006. Molecular systematics and classification of North American notropin shiners and minnows (Cypriniformes: Cyprinidae). In: De Lourdes Lozano-Vilano, M., and Contreras-Balderas, A. J. (eds.) Studies of North American desert fishes in honor of E. P. (Phil) Pister, conservationist. Universidad Autonoma de Nuevo Leon, Melhorn, W. N., and J. P. Kempton (eds.) 1991. Geology and hydrology of the Teays-Mahomet bedrock valley system. Geological Society of America Special Paper 258.

Meagher, S., and Dowling, T. E., Hybridization between the cyprinid fishes *Luxilus albeolus*, *L. cornutus*, and *L. cerasinus* with comments on the proposed hybrid origin of *L. albeolus*. *Copeia* **1991**: 979-991.

Müller, K. 2005. SeqState – primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics* **4**: 65-69.

Near, T. J. 2002. Phylogenetic relationships of *Percina* (Percidae: Etheostomatinae). *Copeia* **2002**: 1-14.

Near, T. J., and Benard, M. F. 2004. Rapid allopatric speciation in logperch darters (Percidae: *Percina*). *Evolution* **58**: 2798-2808.

Near, T. J., Page, L. M., and Mayden, R. L. 2001. Intraspecific phylogeography of *Percina evides* (Percidae: Etheostomatinae): an additional test of the Central Highlands pre-Pleistocene vicariance hypothesis. *Molecular Ecology* **10**: 2235-2240.

Nei, M., and Li, W.-H.1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences USA* **76:** 5269-5273.

Nylander, J. A. A. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Page, L. M., and Burr, B. M. 2011. Peterson Field Guide to Freshwater Fishes of North America North of Mexico: Second Edition. Houghton Mifflin Company, Boston, MA.

Petersson, M., Bylund, M., and Alderborn, A. 2003. Molecular haplotype determination using allele-specific PCR and pyrosequencing technology. *Genomics* **82**: 390-396.

Pflieger, W. L. 1971. A distributional study of Missouri fishes. *University of Kansas Publications: Museum of Natural History* **20**: 225-570.

Piller, K. R., and Bart, H. L., Jr. 2009. Incomplete sampling, outgroups, and phylogenetic inaccuracy: A case study of the Greenside Darter complex (Percidae: *Etheostoma blennioides*) *Molecular Phylogenetics and Evolution* **53**: 340-344.

Piller, K. R., Bart, H. L., and Hurley, D. L. 2008. Phylogeography of the Greenside Darter complex, *Etheostoma blennioides* (Teleostomi: Percidae): a wide-ranging polytypic taxon. *Molecular Phylogenetics and Evolution* **46**: 974-985.

Poly, W. J. 1997. Characteristics of an intergeneric cyprinid hybrid, *Campostoma anomalum* x *Luxilus* sp. indet. (Pisces: Cyprinidae), from the Portage River, Ohio. *Ohio Journal of Science* **97**: 40-43.

Posada, D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**: 1243-1256.

Putnam, F. W. 1863. List of the fishes sent by the museum to different institutions, in exchange for other specimens, with annotations. *Bulletin of the Museum of Comparative Zoology* **1**: 1-16.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In: Howard, D. J., and Berlocher, S. H. (eds.) Endless forms: species and speciation. New York: Oxford University Press.

Rambaut, A., and Charleston, M. 2001. TreeEdit Phylogenetic Tree Editor v1.0 alpha 8. http://tree.bio.ed.ac.uk/software/treeedit.

Ramos-Onsins, S. E., and Rozas, J. 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**: 2092-2100.

Ray, J. M., Lang, N. J., Wood, R. M., and Mayden, R. L. 2008. History repeated: recent and historical mitochondrial introgression between the Current Darter, *Etheostoma* *uniporum*, and Rainbow Darter, *Etheostoma caeruleum*, (Teleostei: Percidae). *Journal of Fish Biology* **72**:418–434.

Ray, J. M., Wood, R. M., and Simons, A. M. 2006. Phylogeography and post-glacial colonization patterns of the rainbow darter, *Etheostoma caeruleum* (Teleostei: Percidae). *Journal of Biogeography* **33**: 1550-1558.

Ree, R. H., and Smith, S. A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* **57**: 4-14.

Robison, H. W. 1986. Zoogeographic implications of the Mississippi River basin. In: C. H. Hocutt and E. O. Wiley (eds.) The zoogeography of North American freshwater fishes. John Wiley & Sons, New York, NY. pp. 267-285.

Robison, H. W., and Buchanan, T. M. 1988. Fishes of Arkansas. University of Arkansas Press, Fayetteville, AR.

Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* **46**: 195-203.

Ronquist, F., and Huelsenbeck, J. P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.

Rosen, D. E. 1979. Fishes from the uplands and intermontane basins of Guatemala: Revisionary studies and comparative geography. *Bulletin of the American Museum of Natural History* **162**: 267-376.

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**: 1218-1231.

Schmidt, T. R., Bielawski, J. P., and Gold, J. R. 1998. Molecular phylogenetics and evolution of the cytochrome *b* gene in the cyprinid genus *Lythrurus* (Actinopterygii: Cypriniformes). *Copeia* **1998**: 14-22.

Scribner, K. T., Page, K. S., and Bartron, M. L. 2001. Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. *Reviews of Fish Biology and Fisheries* **10**:293–323.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51**: 492-508.

Shimodaira, H., and Hasegawa, M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**: 1246-1247.

Siddall, M. E. 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris zone. *Cladistics* **14**: 209-220.

Simon, T. P, and Wallus, R., 2006. Reproductive biology and early life history of fishes in the Ohio River drainage. CRC Press: Boca Raton, FL.

Simons, A. M. 2004. Phylogenetic relationships in the genus *Erimystax* (Actinopterygii: Cyprinidae) based on the cytochrome *b* gene. *Copeia* **2004**: 351-356.

Smith, W. M., and Craig, M. T. 2007. Casting the percomorph net widely: the importance of broad taxonomic sampling in the search for the placement of serranid and percid fishes. *Copeia* **2007**: 35-55.

Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* **15**: 4261-4293.

Sousa-Santos, C., Robalo, J. I., Collares-Pereira, M.-J., and Almada, V. C. 2005. Heterozygous indels as useful tools in the reconstruction of DNA sequences and in the assessment of ploidy level and genomic constitution of hybrid organisms. *DNA Sequence* **16**: 462-467.

Starnes, W. C., and Etnier, D. A. 1986. Drainage evolution and fish biogeography of the Tennessee and Cumberland Rivers drainage realm. In: C. H. Hocutt and E. O. Wiley (eds.) The zoogeography of North American freshwater fishes. John Wiley & Sons, New York, NY. pp. 325-361.

Stauffer, J. R., Jr., Denoncourt, R. F., Hocutt, C. H., Jenkins, R. E. 1979. A description of the cyprinid fish hybrid, *Notropis chrysocephalus* × *Notropis photogenis*, from the Greenbrier River, West Virginia. *Natural History Miscellanea* **204**: 1-6.

Stephens, M., and Scheet, P. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics* **76**: 449-462.

Stephens, M., Smith, N. J., and Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**: 978-989.

Strange, R. M., and B. M. Burr. 1997. Intraspecific phylogeography of North American highland fishes: a test of the Pleistocene vicariance hypothesis. *Evolution* **51**: 885-897.

Strange, R. M., and Mayden, R. L. 2009. Phylogenetic relationships and a revised taxonomy for North American cyprinids currently assigned to *Phoxinus* (Actinopterygii: Cyprinidae). *Copeia* **2009** 494-501.

Switzer, J. F. 2004. Molecular systematics and phylogeography of the *Etheostoma variatum* species group (Actinopterygii: Percidae). Ph. D. dissertation, Saint Louis University.

Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Thornbury, W. D. 1965. Regional geomorphology of the United States. New York: John Wiley and Sons.

Tsai, C.-F., and Raney, E. C. 1974. Systematics of the Banded Darter, *Etheostoma zonale* (Pisces: Percidae). *Copeia* **1974**: 1-24.

Turner, T. F., J. C. Trexler, D. N. Kuhn, and H. W. Robison. 1996. Life-history variation and comparative phylogeography of darters (Pisces: Percidae) from the North American Central Highlands. *Evolution* **50**: 2023-2036.

Wiley, E. O., and Mayden, R. L. 1985. Species and speciation in phylogenetic systematics, with examples from the North American fish fauna. *Annals of the Missouri Botanical Garden* **72**: 596-635.

Wojcicki, M., and Brooks, D.R. 2005. PACT: an efficient and powerful algorithm for generating area cladograms. *Journal of Biogeography* **32**: 755–774.

Wood, R.M., Mayden, R.L., Matson, R.H., Kuhajda, B.R., and Layman, S.R. 2002. Systematics and biogeography of the *Notropis rubellus* species group (Teleostei: Cyprinidae). *Bulletin: Alabama Museum of Natural History* **22**: 37-80.

Zandee, M., and M. C. Roos. 1987. Component compatibility in historical biogeography. *Cladistics* **3**: 305-332.

Zwickl, J. D. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph. D. dissertation, The University of Texas at Austin.