Gene Trees Reveal Repeated Instances of Mitochondrial DNA Introgression in Orangemouth Darters (Percidae: Etheostoma)

CHRISTEN M. BOSSU1,2 AND THOMAS J. NEAR3,*

1Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37917, USA;
2Present address: Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA;
3Department of Ecology and Evolutionary Biology and Peabody Museum of Natural History, Yale University, New Haven, CT 06520, USA;
*Correspondence to be sent to: Department of Ecology and Evolutionary Biology and Peabody Museum of Natural History, Yale University, New Haven, CT 06520, USA; E-mail: thomas.near@yale.edu.

Abstract.—Phylogenies of closely related animal species are often inferred using mitochondrial DNA (mtDNA) gene sequences. The accuracy of mtDNA gene trees is compromised through hybridization that leads to introgression of mitochondrial genomes. Using DNA sequences from 6 single-copy nuclear genes and 2 regions of the mitochondrial genome, we investigated the temporal and geographic signature of mitochondrial and nuclear introgression in the Etheostoma spectabile clade. Phylogenetic analyses of the nuclear genes result in the monophyly of the E. spectabile clade; however, with respect to sampled specimens of 5 species (Etheostoma fragi, Etheostoma uniporum, Etheostoma pulchellum, Etheostoma barri, and E. spectabile), the mitochondrial phylogeny is inconsistent with E. spectabile clade monophyly. Etheostoma uniporum and E. fragi are both fixed for heterospecific mitochondrial genomes. Limited nuclear introgression is restricted to E. uniporum. Our analyses show that the pattern of introgression is consistently asymmetric, with movement of heterospecific mitochondrial haplotypes and nuclear alleles into E. spectabile clade species; introgressive hybridization spans broad temporal scales; and introgression is restricted to species and populations in the Ozarks. The introgressed mitochondrial genome observed in E. fragi has an obscure phylogenetic placement among darters, an ancient age, and is possibly a mitochondrial fossil from an Etheostoma species that has subsequently gone extinct. These results indicate that introgression, both ancient and more contemporaneous, characterizes the history of diversification in the E. spectabile species clade and may be relatively common among clades comprising the species-rich North American freshwater fauna. [Gene phylogeny; hybridization; introgression; mtDNA; nuclear; molecular clock; reproductive isolation; species tree.]

A pattern emerging from several recent phylogenetic investigations among closely related species is that sampled DNA sequences from genes with independent evolutionary histories often do not converge onto a similar inferred species tree (Shaw 2002; Bachtrog et al. 2006). One perspective that is gaining wider acceptance is one championed by Maddison (1997, p. 523), where “all of the gene trees are part of the species tree, which can be visualized like a fuzzy statistical distribution, a cloud of gene histories.” Maddison (1997) suggested that phylogenetic analysis of multiple loci should be undertaken in an explicit coalescent framework, because it is expected that some gene trees will deviate from the underlying species tree due to stochastic variation in the coalescent process. However, it is important to consider that comparing and highlighting incongruence between gene trees can be biologically meaningful, ensuring that processes other than incomplete lineage sorting, such as hybridization, are considered and not lost in the “cloud of gene histories.”

Typically, phylogenetic relationships of animal species with recent diversification are investigated with gene trees inferred from mitochondrial DNA (mtDNA) gene sequences (Avise 1986). Mitochondrial genes are thought to have a better chance of tracking the species tree due to a higher mutation rate relative to nuclear genes, and ancestral alleles shared between incipient species will sort faster due to a smaller effective population size as a consequence of uniparental inheritance (Pamilo and Nei 1988; Moore 1995). However, it is important to consider that the mitochondrial genome is genetically a single locus and not necessarily a representative of the multitude of evolutionary histories of the unlinked genes in the nuclear genome. There are several mechanisms that will lead to incongruence between gene trees and the species tree, most notably incomplete lineage sorting of ancestral polymorphisms and introgression resulting from interspecific hybridization (Maddison 1997; Funk and Omland 2003). This prospect is particularly troubling for phylogenies of closely related species based only on cytoplasmic genome sequence data, because there is a relatively high frequency of heterospecific mitochondrial and chloroplast lineages that have crossed species boundaries in many plant and animal clades (Chan and Levin 2005).

Despite the possibility of being misled by using a single locus, the fact remains that mitochondrial gene sequences have an appreciably higher probability of resolving relationships among closely related animal species than DNA sequences from a single nuclear gene and perhaps substantial sets of nuclear genes (Moore 1995; Hudson and Coyne 2002). Given the need for phylogenetic trees among closely related species for comparative studies, detecting mtDNA introgression is a very important aspect of studying the phylogenetic history of these clades (Funk and Omland 2003; Grant et al. 2005). A notable pattern in the form of biased introgression of maternally inherited genomes can aid phylogeneticists in this endeavor. There are several mechanisms that independently, and in combination, could bias introgression of the cytoplasmic genome: sexual selection and asymmetric reproductive barriers (Chan and Levin 2005), demographic effects (Rieseberg et al. 1996b), differences in the magnitude of selection on particular
genes (Funk and Omland 2003), and cytonuclear compatibilities (Rieseberg et al. 1996a). This biased cytoplasmic introgression manifests itself without introgression of alleles from the nuclear genome. Examples include Hawaiian crickets (Shaw 2002), the Drosophila yakuba species group (Bachtrog et al. 2006), Arctic and Brook char (Doiron et al. 2002), mormyrid fishes (Sullivan et al. 2004), European cyprinid fishes (Freyhof et al. 2005), and lamprologin cichlid fishes (Schelly et al. 2006). In all these cases, phylogenetic analysis of mtDNA sequences would reflect the heterospecific origin of the mitochondrial genomes, resulting in an incorrect conclusion with regard to the inferred species phylogeny. The method that has shown the most promise to detect mitochondrial introgression involves a direct assessment of incongruence between mitochondrial and nuclear gene trees, as well as differences between molecular phylogenies and morphological characters (Shaw 2002; Bachtroge et al. 2006). Because of the uniparental inheritance of the mitochondrial genome, this straightforward phylogenetic approach can also identify the directionality of introgression, and given the phylogenetic pattern coupled with molecular branch lengths, it may also provide information on the relative timing of introgressive hybridization events.

Our focus on mtDNA introgression is the result of ongoing investigations in the phylogenetic relationships of darters, a clade of approximately 225 freshwater teleost fish species endemic to eastern North America. In particular, published studies have found substantial incongruence between the mtDNA gene tree and a single nuclear gene tree that is consistent with extensive mtDNA introgression in several species of the Etheostoma spectabile species clade (Lang and Mayden 2007; Ray et al. 2008), collectively called the orangethroat darters. This clade contains 8 recognized species (Distler 1968; Ceas and Page 1997; Ceas and Burr 2002; Nelson et al. 2004). 2 subspecies that we regard here as species, and at least 4 undescribed species. Breeding males of the E. spectabile clade are brilliantly colored and superficially resemble the rainbow darter, Etheostoma caeruleum, differing only in breeding pigmentation pattern, pectoral fin ray counts, and infraorbital canal characteristics (Trautman 1930; Page 1983). Reflecting the overall similarity in color pattern, E. caeruleum and species of the E. spectabile species clade have traditionally been classified in the darter subclade Oligocephalus (Page 1981, 1983; Bailey and Etner 1988). Species in the E. spectabile clade and E. caeruleum co-occur throughout their respective distributions in the Eastern and Interior Highlands of eastern North America (Fig. 1), and putative F1 hybrid specimens have been documented (Distler 1968). In addition, natural hybridization between Etheostoma pulchellum with Etheostoma radiosum and Etheostoma whipplei has been documented (Linder 1955, 1958; Branson and Campbell 1969; Echelle et al. 1974; Mendelson 2003a).

Phylogenetic analyses of mitochondrial and nuclear genes indicate that Oligocephalus is not monophyletic and the E. spectabile species clade is distantly related to E. caeruleum (Lang and Mayden 2007). The gene trees presented in Lang and Mayden (2007) also reveal that although the E. spectabile species clade was monophyletic for the sampled nuclear gene, the species complex was not monophyletic in the mitochondrial gene tree. In particular, mtDNA haplotypes sampled from Etheostoma uniporum were closely related to those sampled from E. caeruleum, and mtDNA haplotypes sampled from Etheostoma fragi were not closely related to any of the sampled darter species. The similarity of mtDNA haplotypes observed in E. uniporum and E. caeruleum was explored in Ray et al. (2008), where they discovered phylogenetic patterns consistent with hybridization, followed by introgression of E. caeruleum mtDNA haplotypes into E. uniporum. The discrepancy between the mitochondrial and the nuclear gene trees observed within the E. spectabile species clade appears consistent with multiple instances of heterospecific mtDNA introgression.

Our research goals are to investigate the extent of introgression of mitochondrial genomes and nuclear alleles in the E. spectabile clade through comparisons of gene trees inferred from mtDNA and multiple nuclear genes. We are interested in determining if extensive mtDNA introgression has occurred without nuclear introgression, as seen in other animal clades (Powell 1983; Spolsky and Uzzell 1984; Bachtroge et al. 2006). Incongruence between inferred mitochondrial and nuclear gene trees would suggest that mtDNA introgression has been a feature of the diversification of the E. spectabile species clade, whereas limited incongruence among the phylogenies inferred from the nuclear genes would indicate a lack of introgression in the nuclear genome. We intend to focus on both the temporal and the geographic contexts of introgression, asking 2 main questions. First, what is the relative timing of introgression events as determined through comparison of gene trees? Second, given that species of the E. spectabile complex co-occur with E. caeruleum in 2 areas with rich aquatic biodiversity (Fig. 1), are hybrids and genomes with introgressed alleles found throughout their overlapping geographic distributions, or is introgression constrained to particular geographic regions? Our research strategy is unique among published studies involving teleost fishes in that it uses 6 single-copy nuclear genes to investigate mtDNA introgression and estimate phylogenetic relationships within a clade of closely related species.

**Materials and Methods**

**Specimen Sampling, DNA Sequencing, and Alignment**

Etheostoma spectabile clade species were collected throughout their geographic distributions using a seine net (Fig. 1; Supplementary Table A1, http://www.oxfordjournals.org/our_journals/sysbio/). On collection, specimens were anesthetized using MS-222, and a tissue biopsy was taken from the right pectoral fin and cataloged in the Yale Fish Tissue Collection. The
FIGURE 1. Geographic distribution of *Etheostoma spectabile* clade species. The geographic distributions of species in the Eastern Highlands are colored grey and include *Etheostoma bison*, *Etheostoma tecumsehi*, *Etheostoma lawrencei*, *Etheostoma kantuckeense*, Mamequit darter, Sheltowee darter, and Ihiyo darter. Region of ambiguous species designation between *E. spectabile* and *E. pulchellum* is indicated. Approximate sampling localities are shown with solid circles. Inset map shows the geographic distribution of *Etheostoma caeruleum*.

Specimens were fixed in formalin and later transferred to 70% ethanol and deposited in the fish collection at the Yale Peabody Museum of Natural History (YPM) or the University of Tennessee Research Collection of Fishes. Dr. Patrick Ceas (St. Olaf College) provided additional specimens of *E. spectabile* and *E. pulchellum*. The darter species *E. caeruleum*, *E. radiosum*, and *E. whipplei* were sampled to ensure inclusion of potential parental species involved in hybridization with *E. spectabile* clade species. To explore the phylogenetic relationships of the *E. spectabile* clade with respect to the entire darter clade, representative species from all the major darter lineages were sampled. Sequences of the mtDNA gene cytochrome *b* (*cyt*-*b*) collected from specimens sampled across the geographic distribution of *E. caeruleum* were downloaded from GenBank (Ray et al. 2006) to investigate the geographic context of introgression. The non-darter percids, *Sander vitreus* and *Perca flavescens*, were included as outgroups in all phylogenetic analyses.

Nucleic acids were isolated with standard phenol-chloroform extraction and ethanol precipitation methods or with a Qiagen DNAnase Tissue Extraction Kit (QIAGEN, Valencia, CA) following the manufacturer’s protocol. The mitochondrial genes *cyt*-*b* and NADH subunit 2 (*ND2*) were amplified using polymerase chain reaction (PCR), with primers and cycling conditions reported in Near et al. (2000) and Kocher et al. (1995). Nuclear genes were amplified using PCR, with primers and cycling conditions reported in previous studies; *S7* ribosomal protein intron 1 (Chow and Hazama 1998), *RAG1* exon 3 (Lopez et al. 2004), *mixed lineage leukemia* (*MLL*) (Dettaê¨ı and Lecointre 2005), Tmo-4C4 (Streeiman and Karl 1997), and *myoglobin* (*Mb*) (Grove et al. 2004). Primers for the *KELCH* locus were provided by Leos Kral (*KELCH*-F GCAAGAAACCAGGC-TAACA and *KELCH*-R GCATGAAATGCGACTGT). After initial sequencing, specific primers were designed for PCR and sequencing of *MLL* (*MLL*-F2 TTGCCCTCAGCCCTTCTTCTACGGG and *MLL*-R2 CTGAGSGACAGTTGGGCTC) and *Mb* (MVO-F2b AGGCTCTCCAAAGTGTGCG and MVO-R2 GCATGGTCCCTTGCCCT). All 6 nuclear genes were sequenced for a subset of our specimens that included 46 individuals sampled from 27 darter species. Individuals from multiple populations of *E. fragi*, *E. uniporum*, *E. caeruleum*, and putative hybrid specimens were sampled. Thermal cycling for PCR amplification of the nuclear genes included an initial denaturation step of 94°C for 3 min, followed by 30 cycles of denaturing at 94°C (30 s); primer annealing at 57.4°C (30 s) for *RAG1* exon 3
and MLL, and 55°C for Mb, Tmo-4C4, and KELCH; and extension at 72°C (1.5 min), followed by a final incubation period of 72°C for 5 min.

Amplified DNA products from PCR were cleaned prior to direct DNA sequencing using enzymatic purification with exonuclease 1 and shrimp alkaline phosphatase, incubated at 37°C for 15 min, followed by 80°C for 15 min to inactivate the enzymes, or using a Qiagen Qiaquick PCR Purification Kit (QIAGEN). Cleaned PCR products were used as templates for Big Dye (Applied Biosystems, Foster City, CA) cycle sequencing, and sequencing reactions were read using an ABI 3100 automated sequencer at the Molecular Systematics and Conservation Genetics Laboratory (Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT). Primers were designed for sequencing the RAG1 locus (RAG1F2 GATCTTTTCAGCCCTGCAYNCCC and RAG1R2 GCTCAAAAGGCTTGTACTGRC). Contiguous sequences were built using Sequencer (GeneCodes, Ann Arbor, MI). The cyt b and ND2 sequences were aligned by eye using the text editor in PAUP* v. 4.01 (Swofford 2002). All nuclear genes were aligned in ClustalX 1.8 (Thompson et al. 1997), followed by minor adjustments using the data editor in MacClade 4.0 (Maddison and Maddison 2000). Some individuals contained heterozygous genotypes for the sampled nuclear loci, and any heterozygous base pair positions were coded using standard degeneracy codes.

Phylogenetic Analyses and Species Tree Estimation

Partitioned Bayesian analyses were used to generate phylogenetic trees from the mitochondrial and nuclear gene alignments (Larget and Simon 1999). Phylogenies were estimated using several combinations of the sampled mitochondrial and nuclear genes: cyt b for all sampled darter specimens including the supplemental E. caeruleum sequences sampled from GenBank (Ray et al. 2006), the combined cyt b and ND2 mitochondrial gene alignment for E. spectabile complex species and other major darter lineages, each nuclear gene separately, and all nuclear genes combined into a single data matrix. Character partitions used in the partitioned Bayesian analyses were designated based on intron and exon regions and codon positions in coding regions for each gene. The partitioning scheme used in this study is similar to that used in other studies of darter phylogeny that have used both mitochondrial and nuclear gene sequences (Page et al. 2003; Lang and Mayden 2007; Keck and Near 2008; Piller et al. 2008). The models of molecular evolution for each character partition used in the Bayesian analyses were selected using the Akaike Information Criterion as executed in the computer program Modeltest v. 3.06 (Posada and Crandall 1998). Partitioned Bayesian analyses were run in MrBayes v. 3.1.2 (Ronquist and Huelsenbeck 2003) for 20 million generations, and parameter convergence was verified with Tracer v1.4 (Rambaut and Drummond 2003). Four chains were run simultaneously in each analysis. The burn-in period for the Bayesian analyses was $1 \times 10^6$ generations, and the parameters and trees generated before the burn-in period were discarded.

Given the potential for introgression in the E. spectabile clade and resulting incongruence between different gene phylogenies, it is important to explore different strategies to estimate the species tree from the individual gene trees. There are differing opinions regarding the merits of different methods to estimate a species tree, given data sampled from multiple genes. These methods include construction of a consensus tree from the individual gene trees (Jennings and Edwards 2005) and estimation of the species tree from a concatenated multiple-gene data set (Nylander et al. 2004; Rokas and Carroll 2005; Kubatko and Degnan 2007). Two methods have been developed to incorporate assumptions of coalescent models in estimating the optimal species tree: one minimizes the number of coalescent events across multiple gene trees (Maddison 1997; Maddison and Knowles 2006; Carstens and Knowles 2007), and the second uses a Bayesian strategy to estimate the species tree (Edwards et al. 2007; Liu and Pearl 2007). These methods assume that the incongruence among gene trees is due to incomplete lineage sorting; thus, our species tree analyses focused on the nuclear gene trees to remove any incongruence caused by mitochondrial introgression. We performed a partitioned Bayesian analysis that combined the data from the 6 sampled nuclear genes as discussed above. Mesquite v. 2.01 (Maddison and Maddison 2007) was used to estimate the species tree, given assumptions of lineage sorting and minimization of deep coalescences (Maddison 1997). A gene tree was estimated from each of the 6 nuclear loci using maximum likelihood trees estimated using the computer program RAxML (Stamatakis 2006), and these trees were used in Mesquite along with an association matrix that related each of the 46 sampled gene sequences to the 27 sampled species. A heuristic tree search within the Mesquite coalescent module using deep coalescences with multiple loci criterion and subtree pruning and regrafting branch swapping was used to find optimal species trees that minimized deep coalescence events. When multiple optimal trees were returned, a 50% majority rule consensus tree was generated using the computer program Phyutility (Smith and Dunn 2008).

A Bayesian method implemented in the computer program BEST v. 2.0 was also used to estimate the species tree (Liu and Pearl 2007). We incorporated multiple alleles of the 6 nuclear genes from the E. spectabile clade species Etheostoma burri, E. fragi, E. pulchellum, E. spectabile, and E. uniporum. Multiple alleles of the nuclear genes were also sampled from E. caeruleum. The optimal substitution models presented in Table 1 were for each locus. The prior distribution of theta ($\theta$) was an inverse gamma distribution with $\alpha = 3$ and $\beta = 0.03$ (Liu et al. 2008). Chains were run for $1 \times 10^8$ generations, and the frequency of sampling was every 1000 generations. Convergence of the chains was assessed by examining the log-likelihood values using the computer program Tracer v. 1.4 (Rambaut and Drummond...
Table 1. Selected optimal molecular evolutionary models for character partitions identified for each gene

<table>
<thead>
<tr>
<th>Gene</th>
<th>Length (bp)</th>
<th>Model of first codon</th>
<th>Model of second codon</th>
<th>Model of third codon</th>
<th>Model of intron</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytb</td>
<td>1140</td>
<td>TrN+I+G</td>
<td>GTR+I+G</td>
<td>GTR+I+G</td>
<td>NA</td>
</tr>
<tr>
<td>ND2</td>
<td>1047</td>
<td>TrN+I+G</td>
<td>GTR+I+G</td>
<td>GTR+I+G</td>
<td>NA</td>
</tr>
<tr>
<td>S7 intron 1</td>
<td>543</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>TVM+I</td>
</tr>
<tr>
<td>Tmo4C4</td>
<td>473</td>
<td>F81+I</td>
<td>F81</td>
<td>SYM+I</td>
<td>NA</td>
</tr>
<tr>
<td>Mb</td>
<td>409</td>
<td>TrNef</td>
<td>TrN</td>
<td>TrN+I</td>
<td>TrN</td>
</tr>
<tr>
<td>KELCH</td>
<td>754</td>
<td>K81uf+I</td>
<td>K81uf</td>
<td>TrN+I</td>
<td>TrN</td>
</tr>
<tr>
<td>RAG1 exon 3</td>
<td>1377</td>
<td>TVM+I</td>
<td>TVM+I</td>
<td>TVM+I</td>
<td>TVM+I</td>
</tr>
</tbody>
</table>

Abbreviation: NA = the gene does not contain the specified region.

The significance of constraining the gene trees that did not support the monophyly of the *E. spectabile* clade and *E. uniporum* specimens was explored using Bayes factors that quantified the support of alternative hypotheses given the observed data (Suchard et al. 2005). A Bayes factor is defined as the ratio of marginal likelihoods, or the likelihood of data under a particular model after parameter estimation, from 2 competing hypotheses. The Bayes factor can be interpreted as the success of each hypothesis at predicting the data (Kass and Raftery 1995). In this case, the null hypotheses were the optimal gene trees resulting from the unconstrained Bayesian analyses and were compared against the alternative hypotheses or the phylogenies that resulted from the constrained searches to enforce monophyly of the *E. spectabile* species clade and the sampled *E. uniporum* specimens. Each sampled gene that did not result in optimal gene trees that support the monophyly of the *E. spectabile* clade and *E. uniporum* was analyzed in MrBayes with the appropriate monophyly constraint. The Markov chain Monte Carlo (MCMC) parameters were identical to the unconstrained analyses discussed above. The marginal log-likelihoods were retrieved via the “sump” function in MrBayes and interpretations were based on the guidelines presented by Kass and Raftery (1995).

Relative Timing of Introggression

The relative timing among inferred mtDNA introgression events was estimated using an uncorrelated lognormal (UCLN) model of rate variation executed in the computer program BEAST v. 1.4.7 (Drummond et al. 2006; Drummond and Rambaut 2007). A uniform prior with a mean of 100 arbitrary time units with upper and lower bounds at 99 and 101 time units was set for the most recent common ancestor (MRCA) of the entire darter clade on the mtDNA gene tree. BEAST was run for $4 \times 10^7$ generations with a Yule speciation prior and the optimal partitioned molecular evolution models used in the MrBayes analyses. We used the computer program Tracer v. 1.4 to determine convergence and measure the effective sample size of estimated parameters and relative ages. The ages of the MRCA for nodes in the mtDNA gene tree characterized by introgression were tracked and the distribution of relative age estimates that included the 95% highest posterior density (HPD) was reported.

RESULTS

Alignment and Optimal Molecular Evolutionary Models

The initial alignment of the cyt b mtDNA contained sequences from 273 individuals that included all *E. spectabile* clade species and the *E. caeruleum* sequences reported in Ray et al. (2006). The alignment consisted of 1140 base pairs with no insertions or deletions. Replicated haplotypes were removed, leaving 170 individuals in the cyt b alignment. All new DNA sequences generated for this study were submitted to GenBank (FJ381003–FJ381428), and all data matrices and resulting trees were submitted to TreeBASE (http://www.treebase.org; S2262). Following the protocols outlined in Brandley et al. (2005), the optimal molecular evolutionary models for each gene were chosen by comparison of Bayes factors for each partitioning scheme. These models as well as the length of the alignments for all sampled genes are shown in Table 1.

Phylogenetic Analyses of mtDNA Sequences

The *E. spectabile* clade was not monophyletic in the phylogenies resulting from the Bayesian analyses of the cyt b and combined cyt b and ND2 mtDNA alignments, and the standard deviation of split frequencies was less than 0.05, indicating convergence with regard to the tree topology (Figs. 2 and 3). A core clade containing specimens of all *E. spectabile* clade species, except *E. uniporum* and *E. fragi*, was resolved and supported with significant posterior probabilities in the Bayesian inferred mtDNA phylogenies ($>95\%$ posterior probability). However, sampled specimens of 5 *E. spectabile* clade species (*E. fragi, E. uniporum, E. pulchellum, E. burri, and E. spectabile*) were distributed among 7 different
clades in the mtDNA phylogenies (Fig. 2). Most notably, all the haplotypes sampled from *E. uniporum* and *E. fragi* were distantly related to the *E. spectabile* clade. As reported in Ray et al. (2008), all the *E. uniporum* mtDNA haplotypes were phylogenetically nested within those sampled from *E. caeruleum*. Most of the sampled *E. uniporum* haplotypes form a distinct clade of specimens sampled throughout the limited geographic distribution of the species (Fig. 1), but the mtDNA haplotypes of several *E. uniporum* specimens were more closely related to *E. caeruleum* sampled from the same river drainage (Fig. 2). Average uncorrected pairwise distances of cyt* b* and ND2 between *E. uniporum* haplotypes and those from the core *E. spectabile* species clade in the Bayesian phylogeny ranged between 17% and 19%, and were as high as 7% between specimens sampled from the widespread *E. uniporum* clade and *E. caeruleum*. Our new mtDNA phylogenies also showed that mtDNA haplotypes of *E. fragi* did not appear closely related to any sampled darter lineage (Figs. 2 and 3). Average uncorrected pairwise distances at cyt* b* and ND2 between *E. fragi* and specimens in the core *E. spectabile* clade were 20% and the minimum pairwise distance between *E. fragi* and *E. caeruleum* mtDNA gene sequences was 15%.

Sampled mtDNA haplotypes from *E. spectabile*, *E. pulchellum*, and *E. burri* were not monophyletic in the phylogenetic analyses (Figs. 2 and 3). Haplotypes sampled from these species either were distributed in the *E. spectabile* species clade or in the clade of *E. caeruleum* haplotypes or were closely related to *E. whipplei* (Figs. 2 and 3). These populations of *E. spectabile* and *E. burri* are polymorphic with regard to their mitochondrial genomes because they include individuals with *E. spectabile*-like mitochondrial haplotypes and those that exhibit *E. caeruleum*-like mitochondrial haplotypes. In 2 instances, all sampled individuals from a population were fixed for haplotypes that were similar to other either *E. caeruleum* or *E. whipplei*. Three specimens of *E. spectabile* sampled from Doe Run (St. Francis River System, St. Francis County, MO) had mtDNA haplotypes typical of *E. caeruleum* found in the same river system. The 3 *E. pulchellum* specimens sampled from Mill Branch (Arkansas River Drainage, White County, AR) contained haplotypes that were closely related to those found in specimens of *E. whipplei* collected at the same locality (Figs. 2 and 3). There were 2 examples where individuals of *E. spectabile* and *E. burri* sampled from the same location possessed 2 divergent mtDNA haplotypes that were either phylogenetically nested in the core *E. spectabile* species clade or most closely related to *E. caeruleum* haplotypes sampled from the same river drainages.
**Phylogenetic Analyses of Nuclear Gene DNA Sequences**

In stark contrast to the phylogenies resulting from the mtDNA genes, Bayesian phylogenetic analyses for each of the 6 sampled nuclear genes provided substantial support for monophyly, or monophyletic to the exclusion of a single specimen sampled from the *E. spectabile* species clade (Figs. 4 and 5). The standard deviation of split frequencies was less than 0.05, indicating convergence with regard to the tree topology. The alleles sampled at the 6 nuclear gene loci for *E. uniporum* and *E. fragi* were monophyletic and related to a clade comprising the remaining *E. spectabile* species or were each nested in a clade containing all other *E. spectabile* clade species. Bayesian analyses for 3 of the 6 nuclear genes (*KELCH, MLL*, and *Mb*) resulted in phylogenies that contained a clade of all specimens sampled from the *E. spectabile* species complex supported with significant Bayesian posterior probabilities (Fig. 4). Bayesian phylogenies inferred from the *S7* ribosomal protein intron 1 and *RAG1* exon 3 supported near monophyly of the *E. spectabile* clade because in each of these 2 gene trees, a single *E. uniporum* specimen had an allele that was more closely related to the alleles sampled from *E. caeruleum* (Fig. 5a,b). Bayesian analysis of the *Tmo*-4C4 locus resulted in a phylogeny that lacked resolution with regard to interpreting the monophyly of the *E. spectabile* clade (Fig. 5c).

The *E. spectabile* species clade was monophyletic and supported with a significant posterior probability in a Bayesian phylogenetic analysis of the concatenated nuclear genes, and the standard deviation of split frequencies was less than 0.05, indicating convergence.
with regard to the tree topology (Fig. 6). *E. fragi* and *E. uniporum* were most closely related to one another, and all other sampled *E. spectabile* species were resolved as a clade with significant Bayesian posterior support. Only 2 interspecific nodes in the phylogeny were supported with a Bayesian posterior probability less than 1.00, and the phylogeny presented well-resolved relationships among the major darter lineages (Fig. 6).

The *E. spectabile* species clade was also monophyletic in the species analysis that minimized deep coalescences (Fig. 7a). A total of 6 optimal species trees, differing only in outgroup relationships and in the relationship

![Figure 4](image-url)  
**FIGURE 4.** Phylogenies resulting from partitioned Bayesian analyses of the nuclear genes (a) *KELCH*, (b) *MLL*, and (c) *Mb*. The gray box highlights the *Etheostoma spectabile* clade species. Asterisks identify nodes supported by Bayesian posterior probabilities $\geq 95\%$.

![Figure 5](image-url)  
**FIGURE 5.** Partitioned Bayesian analyses of the nuclear genes (a) ribosomal protein S7 intron 1 (S7), (b) RAG1 exon 3 (RAG1), and (c) Tmo-4C4 (Tmo). The gray box highlights the *Etheostoma spectabile* clade species. Asterisks identify nodes supported by Bayesian posterior probabilities $\geq 95\%$. 

---
between *Etheostoma vitreum* and the clade containing the species *E. microperca*, *E. founticola*, and *E. proeliare* were combined in a majority rule consensus tree. The *E. spectabile* species clade was found in each of the 6 species trees, as was the sister species relationship of *E. uniporum* and *E. fragi*.

Similar to the concatenated data Bayesian analysis and the parsimony minimizing deep coalescence species tree analysis, the BEST species tree analysis resulted in a monophyletic *E. spectabile* species clade that was supported with a significant Bayesian posterior probability (Fig. 7b). *E. caeruleum* was not closely related to the *E. spectabile* species clade, but was more closely related to *E. luteovinctum*, *E. radiosum*, and *E. whipplei*.

**Phylogenetic Congruence of Gene Trees**

The Bayesian phylogenies inferred from the mitochondrial and nuclear genes differ substantially with regard to the relationships of the *E. spectabile* species clade (Figs. 2–6). Comparing the optimal unconstrained mtDNA gene trees against the constrained mtDNA gene trees that reflect the monophyly of the *E. spectabile* species clade resulted in very high Bayes factors, providing very strong support for the optimal mtDNA gene trees and rejecting monophyly of the *E. spectabile* clade. The Bayes factor calculated for the contrast between the optimal S7 gene tree and the *E. spectabile* species clade constrained S7 gene tree was moderate and interpreted as strong support for the optimal S7 gene tree. Comparing the optimal gene trees for RAG1 and Tmo-4C4 with the constrained gene trees reflecting monophyly of the *E. spectabile* clade resulted in Bayes factors that are interpreted as positive support for the optimal unconstrained phylogenetic hypotheses, even the lowest Bayes factor for RAG1 provides significant support for introgression in the *E. spectabile* species complex.

**Relative Timing of mtDNA Introgression**

Estimation of relative divergence times in the mtDNA cyt b gene tree using the UCLN model in BEAST revealed that introgressed mtDNA lineages have differing ages of origination (Fig. 8). The MRCA of the entire
darter clade was calibrated with an age of 100 relative time units, and the UCLN mean age estimate for this node was 93.8 (95% HPD: 86.0–100.5). The estimated relative age for the mtDNA haplotype lineage observed in _E. fragi_ was 68.4 (95% HPD: 59.9–76.7) relative time units. The widespread _E. uniporum_ clade that is phylogenetically nested within sampled _E. caeruleum_ haplotypes was much younger, with an estimated relative age of the stem lineage at 11.8 (95% HPD: 9.5–14.2) time units and a crown node age of 9.3 (95% HPD: 6.7–11.9) time units (Fig. 8).

The _E. uniporum_ mtDNA haplotypes that were phylogenetically nested within _E. caeruleum_ haplotypes sampled from the same river drainages ranged in estimated relative age from 9.9 (95% HPD: 7.9–11.9) to 1.0 (95% HPD: 0.4–1.7) relative time units. The relative age of the haplotype lineage observed in _E. pulchellum_ specimens that are closely related to _E. whipplei_ was estimated as

<table>
<thead>
<tr>
<th>Gene</th>
<th>Constrained group</th>
<th>(H_0): unconstrained</th>
<th>(H_\Lambda): constrained monophyly</th>
<th>Bayes factor (= 2\log B_{10} = 2(H_0 - H_\Lambda))</th>
<th>Bayes factor interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All taxa mtDNA</td>
<td><em>Etheostoma spectabile</em> group monophyly</td>
<td>-15 982.70</td>
<td>-17 043.39</td>
<td>2121.38</td>
<td>Very strong</td>
</tr>
<tr>
<td>Subsample mtDNA</td>
<td><em>Etheostoma spectabile</em> group monophyly</td>
<td>-24 001.02</td>
<td>-25 053.67</td>
<td>2105.30</td>
<td>Very strong</td>
</tr>
<tr>
<td>Subsample S7</td>
<td><em>Etheostoma spectabile</em> group monophyly</td>
<td>-3241.29</td>
<td>-3300.85</td>
<td>119.12</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td><em>Etheostoma uniporum</em> group monophyly</td>
<td>-3241.29</td>
<td>3274.07</td>
<td>65.56</td>
<td>Strong</td>
</tr>
<tr>
<td>Subsample RAG1 exon 3</td>
<td><em>Etheostoma spectabile</em> group monophyly</td>
<td>-4280.20</td>
<td>-4284.55</td>
<td>8.70</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td><em>Etheostoma uniporum</em> group monophyly</td>
<td>-4280.20</td>
<td>-4285.01</td>
<td>9.62</td>
<td>Positive</td>
</tr>
<tr>
<td>Subsample Tmo-4C4</td>
<td><em>Etheostoma spectabile</em> group monophyly</td>
<td>-1722.98</td>
<td>-1728.76</td>
<td>11.56</td>
<td>Positive</td>
</tr>
</tbody>
</table>
FIGURE 8. Histograms of the posterior density of relative ages for the MRCA of each identified mtDNA introgression event, as estimated using BEAST, with the same taxon sampling as the mitochondrial gene tree presented in Figure 2. Circles identify the node age presented in the histogram, and the color of the marked node (gray or black) matches that of the histogram when multiple nodes are dated in a single graph. Thick branches in the phylogenies mark heterospecific haplotype lineages found in Etheostoma spectabile clade species. The mean age is reported as well as the 95% upper and lower HPD. The first histogram estimates the relative age of the MRCA of the entire darter clade. 2.5 (95% HPD: 1.4–3.6) time units, and the estimated relative ages of mtDNA haplotypes more closely related to *E. caeruleum*, but found in specimens of *E. burri* and *E. spectabile*, ranged between 0.5 (95% HPD: 0.1–1.0) and 1.8 (95% HPD: 1.1–2.5) relative time units (Fig. 8).

**DISCUSSION**

Through comparisons of phylogenies inferred from mitochondrial and nuclear genes, we have been able to reveal a complex pattern of repeated mitochondrial introgression, with little introgression of nuclear alleles in the *E. spectabile* species clade. Our results were intimated in previous studies that had a more limited sampling of specimens and nuclear genes. Lang and Mayden (2007) presented a phylogeny based on mtDNA sequences of the ND2 gene that showed extensive nonmonophyly of the *E. spectabile* clade, but the group was monophyletic in the phylogeny inferred from the nuclear S7 ribosomal protein intron 1. In the mtDNA phylogeny, *E. uniporum* was more closely related to *E. caeruleum*, and *E. fragi* was nested in an inclusive and monophyletic *Etheostoma* but not closely related to the *E. spectabile* clade or *E. caeruleum*. A more intensive sampling revealed that all *E. uniporum* mtDNA haplotypes were phylogenetically nested in *E. caeruleum* and never closely related to other species in the *E. spectabile* clade (Ray et al. 2008).

Because phylogenetic incongruence among genes can result from both introgression and ancestral polymorphism (Funk and Omland 2003), much effort has been aimed at distinguishing between these 2 processes (Sang and Zhong 2000; Holder et al. 2001; Nielsen and Wakeley 2001; Buckley et al. 2006). Although our analyses did not explicitly test for ancestral polymorphism through
coalescent-based simulations (Knowles and Maddison 2002; Knowles 2004), the fact that mtDNA haplotypes sampled from *E. uniporum* and *E. fragi* were more closely related to distantly related *Etheostoma* lineages argues against lineage sorting events being responsible for this phylogenetic pattern. Supporting this conclusion is the striking and well-supported monophyly of the *E. spectabile* species clade resulting from our phylogenetic analyses of nuclear genes, both the single-gene and the combined-gene analyses (Figs. 4–6) and the species trees inferred from the 6 nuclear genes using 2 different methods (Fig. 7).

There are multiple instances of introgressive hybridization that have occurred at different temporal periods in the evolutionary history of the *E. spectabile* species clade. The *E. uniporum* and *E. fragi* mitochondrial genomes have been completely replaced with those from other *Etheostoma* species and lineages. In addition, the nuclear gene trees suggest rare and limited introgression of *E. caeruleum* alleles into the nuclear genome of *E. uniporum*. The introgression of *E. caeruleum* mtDNA across *E. uniporum* species boundaries has occurred at a minimum during 2 distinct temporal periods in the history of the species. The 3 oldest mean UCLN estimates for the ages of the introgressed *E. caeruleum* haplotypes in *E. uniporum* are similar and range between 11.9 and 7.8 relative time units, and all 3 age estimates exhibit overlapping 95% HPD estimates (Fig. 8). These oldest introgressed mtDNA lineages include the geographically widespread *E. uniporum* clade as well as the haplotypes (*E. uniporum* M, O, and P) that are closely related to *E. caeruleum* haplotypes sampled from the Current River (Figs. 2 and 7). One sampled *E. uniporum* haplotype (T) was well nested in a clade of *E. caeruleum* sampled from the Strawberry, Spring, Gasconade, and Eleven Point rivers (Fig. 2). The mean UCLN age estimate for this haplotype was 1.0 (95% HPD: 1.7–0.4) relative time units, and its 95% HPD did not overlap with the lineage age estimate of any other heterospecific mtDNA haplotype observed in *E. uniporum*. This indicates a separate and more recent secondary introgression of *E. caeruleum* mtDNA in *E. uniporum* (Fig. 8).

Repeated hybridization between *E. caeruleum* and *E. uniporum* has resulted in the recurrent introgression of *E. caeruleum* mitochondrial genomes into *E. uniporum* and has also resulted in limited introgression of alleles in the nuclear genome. The S7 ribosomal intron 1 gene tree shows that the alleles sampled from *E. uniporum* C are more closely related to those sampled from *E. caeruleum* and *E. lutovenustum* (Fig. 5a), and alleles sampled from *E. uniporum* M for the RAG1 exon 3 locus are more closely related to those sampled from *E. caeruleum* (Fig. 5b). Alleles sampled from these specimens for KELCH, MLL, and Mb were all more closely related to other sampled *E. uniporum* specimens and nested in a monophyletic *E. spectabile* species clade (Fig. 4). Evidence of 2 separate instances of nuclear introgression in *E. uniporum* could indicate high introgression pressure on *E. uniporum* from *E. caeruleum*. And although conclusions about the timing of nuclear introgression are difficult, given this cursory examination of the nuclear genome, our analyses highlight the importance of using multiple independent nuclear markers. The origin of the introgressed mitochondrial genome found in *E. fragi* presents a problem that is much less obvious than the *E. caeruleum* introgressed mitochondrial genomes in *E. uniporum*. Our mitochondrial gene tree (Fig. 3) and the one presented in Lang and Mayden (2007) indicate that the mitochondrial genome in *E. fragi* is not closely related to mtDNA haplotypes observed in any extant *Etheostoma* species or lineage. The mean UCLN estimated age for the node that relates the introgressed mtDNA haplotypes in *E. fragi* to other darters in the mtDNA gene trees (Figs. 2 and 3) is quite old at 68.4 (95% HPD: 59.9–76.7) relative time units (Fig. 8). In contrast to the divergent phylogenetic placement inferred from the mtDNA haplotypes, gene trees estimated from each of the 6 nuclear genes show that *E. fragi* is more closely related to other species in the *E. spectabile* species clade (Figs. 4 and 5). Phylogenies resulting from analysis of the concatenated nuclear genes and the species trees estimated from minimizing deep coalescence and the Bayesian method implemented in BEST also result in monophyly of the *E. spectabile* clade and support the hypothesis that the geographically adjacent *E. fragi* and *E. uniporum* are sister species (Figs. 1, 6, and 7). A preliminary phylogenetic analysis of cyt b sampled from 222 of 225 darter species results in a phylogenetic nesting of the introgressed *E. fragi* mitochondrial genome in *Etheostoma* but is not closely related to any extant darter species. Given the obscure phylogenetic relationships and the ancient relative age, we hypothesize that the mitochondrial genome found in *E. fragi* represents an mtDNA fossil, meaning that the mtDNA haplotypes found in *E. fragi* have originated from an *Etheostoma* species that has subsequently gone extinct. A similar mechanism of recent extinction was proposed to explain orphan allozyme alleles present in the unisexual hybrid teleost fish species *Menidia clarki hubbsi* (Echelle et al. 1989), and mitochondrial lineages of extinct species have been reported in tetraploid *Hyla* tree frog species (Holloway et al. 2006).

The patterns of mtDNA introgression observed in *E. uniporum* and *E. fragi* raises questions regarding the occurrence and frequency of mtDNA introgression in other *E. spectabile* clade species. The mtDNA gene trees reveal that individuals of *E. spectabile*, *E. burri*, and *E. pulchellum* have heterospecific mtDNA haplotypes (Figs. 2 and 3) but do not have introgressed nuclear alleles (Fig. 6). The introgressed mtDNA haplotypes observed in *E. spectabile* and *E. burri* are phylogenetically nested in the *E. caeruleum* clade and are very similar to haplotypes of *E. caeruleum* sampled from the same river drainages. The heterospecific mtDNA haplotypes observed in *E. pulchellum* are phylogenetically closely related to the haplotypes sampled from *E. whipplei*, a syntopic congener (Figs. 2 and 3). We are confident that the sampled *E. whipplei* mtDNA haplotypes are not introgressed from another *Etheostoma* species because *E. whipplei* and *E. radiosum* are resolved as sister species.
in both the mitochondrial and the nuclear gene phylogenies (Figs. 2, 3, 6, and 7). This phylogenetic relationship is consistent with previous molecular phylogenetic analyses (Lang and Mayden 2007) and traditional taxonomic arrangements based on external morphology and male nuptial coloration (Hubbs and Black 1941; Moore and Rigney 1952). The UCLN relative age estimates for the introgressed mtDNA haplotypes in *E. spectabile*, *E. burri*, and *E. pulchellum* are very young (Fig. 8), indicating a fairly contemporaneous set of introgression events when compared with the patterns of mtDNA introgression observed in *E. uniporum* and *E. fragi*.

The patterns of mitochondrial introgression observed in the *E. spectabile* species clade, both contemporaneous and ancient, are consistently asymmetric. Heterospecific mtDNA haplotypes are only found in *E. spectabile* clade species, whereas *E. spectabile* clade mtDNA haplotypes are never introgressed in other species such as *E. caeruleum*. Asymmetric mitochondrial introgression is not uncommon and is observed in a diverse array of animal lineages (Good et al. 2003; McGuire et al. 2007; Takami et al. 2007). The asymmetric pattern of mitochondrial introgression observed in the *E. spectabile* clade could have resulted from asymmetric behavioral reproductive isolation, exemplified if heterospecific spawning between *E. spectabile* males and *E. caeruleum*/*E. whipplei* females occurs more frequently than the reciprocal heterospecific spawning event. This is the pattern observed in a laboratory behavioral study that demonstrated a significantly higher frequency of interspecific spawning between *E. radiosum* females and *E. pulchellum* males than between *E. pulchellum* females and *E. radiosum* males (Mendelson 2003a) that was facilitated through sneak fertilization of *E. pulchellum* males. Putative F1 hybrids between these 2 species have been reported (Branson and Campbell 1969); however, it is not known if *E. pulchellum* populations sympatric with *E. radiosum* contain introgressed *E. radiosum* mtDNA haplotypes.

Another mechanism that could result in the asymmetric introgression observed in the *E. spectabile* species clade is postzygotic isolation asymmetry, where reciprocal hybrid crosses yield different degrees of offspring viability or fitness (Tiffin et al. 2001; Bolnick and Near 2005; Turelli and Moyle 2007; Bolnick et al. 2008). One explanation for isolation asymmetry is the presence of genetic incompatibilities involving unparentally inherited genetic factors such as mitochondrial genomes (Turelli and Moyle 2007). A study of the freshwater teleost clade Centrarchidae demonstrated that species with accelerated rates of mitochondrial nucleotide substitution tend to be the worse maternal parent in experimental interspecific crosses, supporting the hypothesis that increased mitonuclear incompatibilities can result in isolation asymmetry (Bolnick et al. 2008). Interesting patterns have also emerged through experimental laboratory crosses performed on darter species that exhibit a wide range of divergence times (Hubbs and Strawn 1957; Hubbs 1958, 1959, 1967, 1971; Mendelson 2003b). Experimental hybrids from 2 sets of interspecific crosses, *E. caeruleum* females × *E. zonale* males, and *E. luteovinctum* females × *E. binotatum* males, exhibit high viability that is similar to control conspecific crosses, but in reciprocal heterospecific crosses, hybrid offspring of *E. zonale* or *E. binotatum* females underwent abnormal development patterns and had significantly reduced viability (Mendelson et al. 2006, 2007). There are no experimental data for reciprocal crosses between *E. spectabile* clade species and *E. caeruleum*; however, experiments should investigate if patterns of isolation asymmetry are concordant with expectations, given the asymmetry of mitochondrial introgression observed in our gene trees and previous studies by Lang and Mayden (2007) and Ray et al. (2008).

Selection on specific combinations of cytonuclear genotypes may result in the biased fixation of 1 parental mtDNA haplotype that would exhibit an asymmetric phylogenetic pattern (Seehausen 2004; Doiron et al. 2002). For example, wild brook charr populations, *Salvelinus fontinalis*, that inhabit colder, high-altitude lakes as compared with populations in lower-elevation warmer habitats, are fixed with mitochondrial haplotypes of arctic charr, *Salvelinus alpinus*, suggesting a selective advantage to introgressed brook charr populations (Doiron et al. 2002). Whether *E. caeruleum* mtDNA haplotypes confer a selective advantage onto *E. spectabile* clade species and a selective sweep is responsible for the complete replacement of the mitochondrial genome observed in *E. uniporum* requires further investigation.

The phylogenetic perspective on mtDNA introgression in the *E. spectabile* species clade provided by our analysis of mitochondrial and nuclear gene trees reveals a very interesting biogeographic pattern. Introgression of heterospecific mtDNA into *E. spectabile* clade species occurs only in those species distributed in the Ozarks and is not observed in species distributed in the Eastern Highlands. Two of the *E. spectabile* clade species (*E. spectabile* and *E. pulchellum*) are widely distributed in the Great Plains, Edwards Plateau, northern Ozarks, upper Mississippi River Basin, and northern tributaries of the Ohio River Drainage that were glaciated during the Pleistocene (Fig. 1). All remaining species have more restricted geographic distributions in the Ozark Mountains and Eastern Highlands, which are areas characterized by a high degree of diversity and endemism of freshwater teleost fish species and were relatively immune to the effects of Pleistocene glaciation (Wiley and Mayden 1985; Near et al. 2001; Near and Keck 2005; Soltis et al. 2006). Investigations of reported *E. spectabile* × *E. caeruleum* hybrids from northern tributaries of the Ohio River Drainage using allozyme analyses did not find evidence of hybridization between these 2 species (Martin and Richmond 1973; McLeod et al. 1980). The propensity for introgression of mtDNA in Ozark species of the *E. spectabile* clade is illustrated by a pattern of introgression from several sources; *E. uniporum* is fixed for mtDNA haplotypes that originated from *E. caeruleum*, whereas *E. fragi* is fixed for mtDNA haplotypes that originated from an *Ethostoma*
lineage that is presumably extinct. Even the more contemporaneous introgression of mtDNA haplotypes from *E. whipplei* and *E. caeruleum* into *E. spectabile*, and *E. burri* is found only in Ozark populations and never in species from the Eastern Highlands. One potential explanation for this pattern lies in the historical biogeography of the 2 lineages. Haplotype and morphological diversity of *E. caeruleum* is significantly higher in the Ozarks (Knapp 1964; Ray et al. 2006) and may be associated with the long-term stability of this region (Pflieger 1971). The phylogenetically basal position of *E. spectabile* species distributed in the Ozarks compared with that distributed in the Eastern Highlands suggests an Ozark origin for the clade, followed by colonization of the Eastern Highlands and previously glaciated areas. Thus, the Ozark region has been an area where *E. spectabile* species and *E. caeruleum* have co-occurred for a longer period of time, providing greater opportunities for hybridization and the observed introgression.

Introgressive hybridization events, both ancient and contemporaneous, have likely played a role in the diversification of the *E. spectabile* clade. Our nuclear analyses support the monophyly of the *E. spectabile* clade, a fact that would have been obscured if our inferences were based solely on the mtDNA phylogeny. This study highlights the necessity of nuclear gene phylogenies when investigating the evolutionary history of closely related species with a history of hybridization. Not only does our comparison of mitochondrial and nuclear gene trees uncover the presence of mitochondrial introgression in nearly one-third of the species in the *E. spectabile* clade, but it also refines the pattern, illustrating that introgressive hybridization spans temporal scales, yet remains restricted to species and populations in the Ozarks. Comparisons of demographic parameters, the evolution of pre- and postzygotic isolating mechanisms, differences in habitat utilization, and different histories of paleogeographic disturbance between the Ozarks and the Eastern Highlands are required to explore potential mechanisms resulting in the patterns of asymmetric mitochondrial introgression that characterize diversification of species in the *E. spectabile* clade.

**Supplementary Material**

Supplementary material can be found at http://www.oxfordjournals.org/our_journals/sysbio/.

**Funding**

The National Science Foundation (DEB-0716155) and the YPM supported this research.

**Acknowledgments**

This study was completed as a partial fulfillment for C.M.B.'s master of science degree, awarded by the Department of Ecology and Evolutionary Biology at the University of Tennessee. Our work on orange-throat darters was greatly facilitated by advice, discussions, and assistance in the field from P. A. Ceas, B. H. Bauer, R. C. Carlson, G. R. Dinkins, D. A. Etnier, R. C. Harrington, K. L. Heckman, P. R. Hollingsworth Jr., C. D. Hulsey, B. P. Keck, and W. C. Starnes. Gregory Watkins-Cowell provided collection assistance at the Peabody Museum of Natural History, and Jennifer Parris assisted with the University of Tennessee Research Collection of Fishes. B. H. Bauer, C. D. Hulsey, J. A. McGuire, and D. C. Tank provided comments on earlier versions of our manuscript.

**References**


First submitted 5 May 2008; reviews returned 8 July 2008; final acceptance 30 December 2008

Associate Editor: Mark S. Hafner