Molecular Phylogenetics and Evolution 53 (2009) 340–344

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Incomplete sampling, outgroups, and phylogenetic inaccuracy: A case study of the Greenside Darter complex (Percidae: *Etheostoma blennioides*)

Kyle R. Piller [a,*], Henry L. Bart Jr. [b]

[a] Southeastern Louisiana University, Dept. of Biological Sciences, Hammond, LA 70402, USA
[b] Tulane University Museum of Natural History, Belle Chasse, LA 70037, USA

ARTICLE INFO

Article history:
Received 29 January 2009
Revised 14 April 2009
Accepted 19 April 2009
Available online 3 May 2009

1. Introduction

The freshwater fish fauna of eastern North America is remarkable for its high species diversity (Mayden, 1992; Etnier and Starnes, 1993; Boschung and Mayden, 2004). One of the most diverse and well studied groups of freshwater fishes in this region are the darters (Teleostei: Percidae), which continue to be of high interest to evolutionary biologists and ecologists alike (Page, 1983; Bart and Page, 1992; Near, 2002).

The greenside darter, *Etheostoma blennioides*, is one of the most widely distributed species of Etheostomatine darters (Lee et al., 1980). Morphological variation within the complex has attracted much taxonomic interest and has resulted in the description of many species now considered synonymous with *E. blennioides* (Miller, 1968). A comprehensive analysis of morphological variation within the complex was conducted by Miller (1968), who recognized four subspecies (*blennioides, newmanii, gutselli*, and *pholidotum*), several morphological races, and three zones of morphological intergradation. Variation within the complex has also attracted the interest of molecular systematists, resulting in two similar phylogeographic studies recently published in this journal (Piller et al., 2008; Haponski and Stepien, 2008).

Piller et al. (2008) examined molecular variation in the mitochondrial cytochrome b gene, using data from 44 populations. Variation in the nuclear encoded S7 intron-1 was also examined for a subset of the specimens used in this study ($N = 24$). Multiple genetically distinctive clades were recovered, suggesting that there is more taxonomic diversity within the *E. blennioides* complex than is currently recognized (Piller et al., 2008). Haponski and Stepien (2008) examined variation in the same genetic markers for a larger number of individuals ($N = 345$), but fewer populations ($N = 19$). Conclusions of the two studies differed significantly. Here, we compare and contrast methodologies, analytical approaches, and conclusions in the two studies to examine why the taxonomic and systematic conclusions differed. Not only are these issues relevant to Greenside Darter taxonomy and systematics, but they are also relevant to phylogenetics in general. We argue that the main reason for the differences recovered in these studies is population sampling and the selective use of data from Piller et al. (2008) by Haponski and Stepien (2008).

2. Methods

The sequences analyzed in this study were derived from Genbank, based on data originally generated by Piller et al. (2008) and Haponski and Stepien (2008). Both studies used cytochrome b and S7 Intron 1. However, the differences in conclusions mainly involved cytochrome b data. Piller et al. (2008) assessed relationships using both maximum parsimony (MP) and Bayesian inference (BI) methodologies, whereas Haponski and Stepien (2008) used MP and maximum likelihood (ML) algorithms, but presented trees based only on ML analysis results. As stated earlier, the purpose of this study was to compare and contrast the results of two similar molecular phylogenetic studies. Therefore, to allow for direct comparisons among data sets, we generated phylogenetic hypotheses using only Bayesian inference (BI) methods.

A partitioned mixed-model Bayesian analysis (BI) was conducted using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) to assess relationships among populations of Greenside Darters. ModelTest 3.06 (Posada and Crandall, 1998) was used infer the best model of DNA sequence evolution using Akaike Information Criterion (AIC) (Posada and Buckley, 2004). Each of the three codon positions of cytochrome b was treated as a separate data partition and models of evolution separately were chosen for each partition using Modeltest.

Posterior probabilities were estimated using the Metropolis-coupled Markov chain Monte Carlo (Huelsenbeck et al., 2001). Bayesian analyses were run for five million generations using four chains, and trees were sampled every 100 generations. Burn-in was determined by examining a plot of maximum likelihood scores

* Corresponding author. Fax: +1 985 549 3851.
E-mail address: Kyle.Piller@selu.edu (K.R. Piller).

1055-7903/$ - see front matter
doi:10.1016/j.ympev.2009.04.010

against generations to determine the point at which likelihood values stabilized. The remaining (non-discarded) trees were used to calculate posterior probabilities on a 50% majority rule consensus tree. Branch support was tested using Bayesian posterior probabilities (BPP) (Holder and Lewis, 2003).

Three data sets were analyzed and compared in this study. First, we used the same cyt b data set Haponski and Stepien's (2008) analyzed using ML methods, but analyzed the data under a Bayesian framework. The second set of data was derived from Piller et al. (2008). Finally, we conducted a combined analysis of cytochrome b data from Piller et al. (2008) and Haponski and Stepien (2008). The results from each of these analyses were compared to each other. Different outgroups were used by Piller et al. (2008) and Haponski and Stepien (2008). Piller et al. (2008) rooted their trees with species from the *E. variatum* group, whereas Haponski and Stepien (2008) used multiple outgroup species including *E. variatum*, *E. bellum, E. camurum, E. rafinesquei, E. rupestre, E. blennius*, and *E. b. gutselli*. In order to make direct comparisons between the two studies and because of uncertainties regarding the appropriate choice of outgroup for *E. blennioides* (Bailey and Etnier, 1988; Porterfield, 1998; Porter et al., 2002), all three datasets were rooted with either *E. variatum*, or members of the *E. variatum* species group (*sensu* Hubbs and Black, 1940; McKeown et al., 1984).

3. Results

3.1. Phylogenetic analysis

Re-analysis of Haponski and Stepien's (2008) cytochrome b data under a partitioned mixed-model (pMM) Bayesian framework rooted with *E. variatum*, resulted in the same phylogenetic tree with similar nodal support values (Fig. 1A) as originally recovered by Haponski and Stepien (2008) using ML and rooted with other Etheostomatine taxa. The pMM Bayesian trees recovered by Piller et al. (2008) and the combined studies, also rooted with the *E. variatum* group, are depicted in Figs. 1B and 2. However, the results differed significantly (data not shown) when the combined dataset was rooted with Haponski and Stepien's (2008) outgroup taxa, including the lack of a monophyletic *E. blennioides* complex (*sensu* Miller, 1968).

Overall, there were many similarities in clades recovered in the three trees (Figs. 1A and B and 2). For example, Piller et al. (2008) and Haponski and Stepien (2008) both recovered distinctive *E. b. blennioides*, *E. b. pholidotum* (Great Lakes, Wabash, and Osage) and *E. b. newmanii* (Arkansas-White, Ouachita, and Cumberland) clades. However, there were discrepancies related to differences of the included populations. Piller et al. (2008) recovered a Tennessee River clade of Greenside Darters that was sister to *E. blennius*, whereas Haponski and Stepien (2008) recovered *E. b. gutselli* as sister to *E. blennius*. The combined dataset, inclusive of a wide-array of population sampling, produced a tree that was most similar to Piller et al. (2008).

4. Discussion

Phylogenetic inference is sensitive to a variety of factors including outgroup choice and taxon sampling (Swofford et al., 1996). It is clear that both of these issues have impacted phylogenetic relationships within the Greenside Darter complex based on our comparisons of Piller et al. (2008) and Haponski and Stepien (2008). Each of these issues is discussed below.

4.1. Outgroup choice

The choice as to which species or group of species should be used to polarize character states has been a point of discussion for many years (Maddison et al., 1984). Previous studies have argued that outgroups chosen should be closely related to the group under study, thereby reducing the potential of obtaining spurious topologies due to homoplasy (saturation effects common in distantly related groups) and long-branch attractions artifacts (Felsenstein, 1978; Wheeler, 1990). Furthermore, incorrectly rooting can result in misleading relationships that could affect subsequent taxonomic interpretations.

Haponski and Stepien (2008) included *E. gutselli* (or *E. b. gutselli sensu* Miller, 1968) as one of their outgroup taxa. However, *Etheostoma b. gutselli* was recognized as a member of the Greenside Darter complex by Miller (1968). One of Haponski and Stepien's (2008) objectives was to test the taxonomic validity of Miller's (1968) Greenside Darter subspecies. Piller et al.'s (2008) cyt b analysis, published prior to Haponski and Stepien's (2008) study, recovered *E. b. gutselli* (10 individuals) as a member of the Tennessee River basin clade of Greenside Darters. The results from both Piller et al. (2008) and Miller (1968) suggest a close relationship among *E. b. gutselli* and other members of the Greenside Darter complex. Therefore, excluding it from the ingroup precludes this test and makes the ingroup non-monophyletic, a violation of practice in phylogenetic systematics.

Piller et al. (2008) based choice of outgroups on recent molecular phylogenetic studies involving darters of subgenus *Etheostoma* and related subgenera (Porterfield, 1998; Porter et al., 2002). Porterfield (1998) inferred relationships among snubnose darters (Subgenus *Ulocentra*) and several outgroups from subgenus *Etheostoma* using the cytochrome b gene and recovered *E. rupestre* as sister to *E. blennioides*. Porter et al. (2002) used data from the mitochondrial control region and recovered *E. blennius* as the sister taxon to *E. blennioides*. Both studies noted a close relationship of the *E. variatum* group to the *E. blennioides* complex.

Haponski and Stepien (2008) included *E. b. gutselli* (or *E. gutselli*) as one of their outgroup taxa. *Etheostoma b. gutselli* has always been a member of the *E. blennioides* complex (i.e., part of a monophyletic *E. blennioides* species complex or group). Unlike some of Haponski and Stepien's (2008) other outgroup taxa, *E. b. gutselli* is closely related to other members of the *E. blennioides* species complex (Piller et al., 2008). Piller et al.'s (2008) cyt b analysis recovered *E. b. gutselli* (10 individuals) as a member of the Tennessee River basin clade of Greenside Darters, therefore, excluding it from the ingroup precludes this test and makes the ingroup non-monophyletic, a violation of practice in phylogenetic systematics.

4.2. Populations/taxon sampling

Studies have shown that the density of sampling of major lineages within the ingroup can strongly influence tree topology in phylogenetic reconstruction (Hillis, 1998; Zwickl and Hillis, 2002; Pollock et al., 2002). In this study, results from the combined analysis (Fig. 2) clearly show that population/taxon sampling has had a dramatic impact on the relationships within the *E. blennioides* complex. The lack of inclusion of particular populations or lineages, including Tennessee River basin populations, has lead to fallacious evolutionary relationships and subsequently, inappropriately derived taxonomic conclusions in regards to *E. b. gutselli* (Haponski and Stepien, 2008).

Haponski and Stepien (2008) claimed that they provide the first substantive evidence for recognizing *E. b. gutselli* as a distinct species, due to its close relationship with *E. blennius*, rather than to other populations/species within the *E. blennioides* complex. Although we agree that *E. b. gutselli* should be recognized as a distinct species, we feel that Haponski and Stepien's (2008) case for this is scientifically flawed because of their selective use of available data from Piller et al. (2008). Haponski and Stepien (2008) recovered a sister relationship between *E. blennius*, a species that

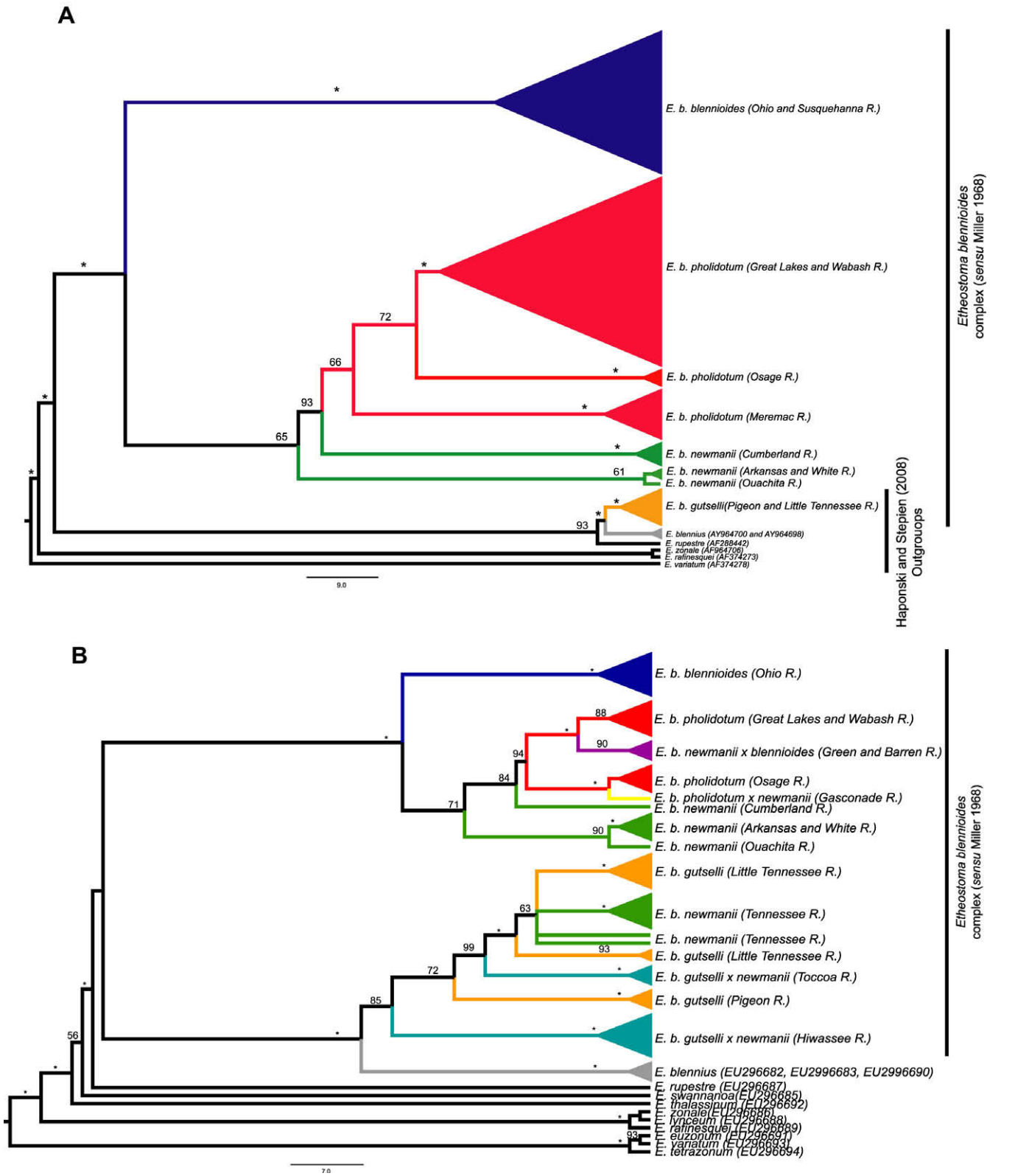

Fig. 1. Phylogram of the *Etheostoma blennioides* (Percidae) complex based on partitioned mixed-model Bayesian analyses of the cytrochrome b gene using data from (A) Haponski and Stepien (2008), and (B) Piller et al. (2008). Both studies were rooted with *Etheostoma variatum* or the *E. variatum* group. Asterisks represent posterior probabilities greater than 95.

is not considered to be a member of the *E. blennioides* species complex (Miller, 1968; Porterfield, 1998; Porter et al., 2002) and *E. b. gutselli* populations from the Pigeon and Little Tennessee Rivers (Fig. 1B). In Piller et al.'s (2008) cyt b tree, *E. blennius* is sister to a large clade of Greenside Darters from different parts of the Tennessee River basin, including *E. b. gutselli* from the Little Tennessee and Pigeon rivers, *E. b. newmanii* from the lower Tennessee River system, and populations that Miller (1968) presumed to be mor-

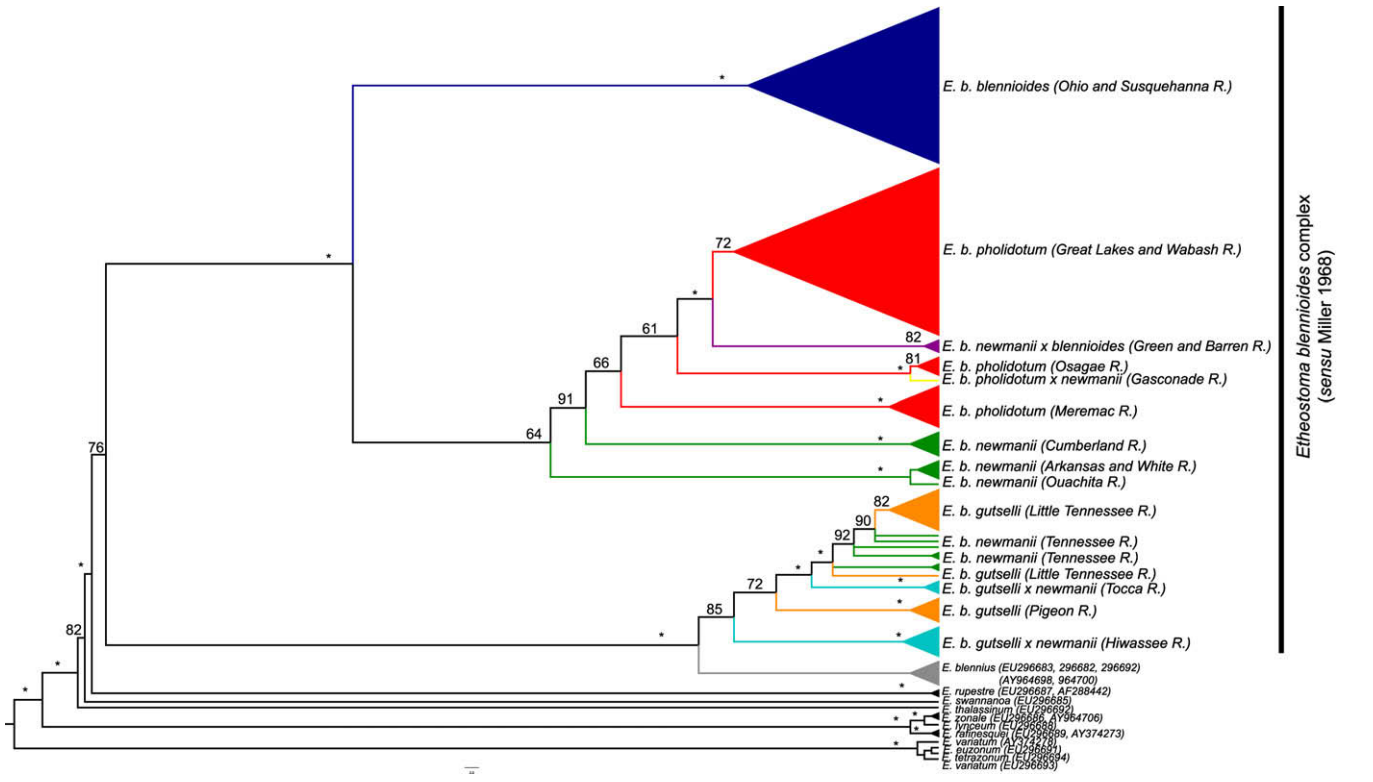

Fig. 2. Phylogram of the *Etheostoma blennioides* (Percidae) complex based on partitioned mixed-model Bayesian analyses of the cytrochrome b gene using combined data from Haponski and Stepien (2008) and Piller et al. (2008). The phylogram was rooted with the *E. variatum* group. Asterisks represent posterior probabilities greater than 95.

phological intergrades between *E. b. newmanii* and *E. b. gutselli* from the Hiwassee River system (Fig. 1A). Some of these populations were not included in Haponski and Stepien's (2008) study. Piller et al. (2008) concluded that the sister relationship of *E. blennius* to Tennessee River basin populations of the *E. blennioides* species complex reflected ancestral mitochondrial introgression. *Etheostoma b. gutselli* and *E. blennius* were recovered as distinct clades in Piller et al.'s (2008) nuclear S7 intron tree and Haponski and Stepien (2008) ignored this conclusion.

More seriously, Haponski and Stepien (2008) selectively used "unambiguous" cyt b sequences from Piller et al. (2008) to make their case that *E. b. gutselli* is sister to *E. blennius*. They state (p. 74) "samples identified morphologically by Piller et al. (2008) as *E. b. newmanii* from the Tennessee River clustered outside the Greenside Darter clade within the *E. b. gutselli* clade with 100% support." However, they failed to show these samples in their trees (cyt b and S7); stating only that more work is needed on the Tennessee River *E. b. newmanii*. Exclusion of these samples weakens Haponski and Stepien's (2008) case for the validity of *E. b. gutselli* and its sister relationship to *E. blennius*.

When all of the available cyt b data are considered together (Fig. 2), *E. b. gutselli* is not recovered as a monophyletic group sister to *E. blennius*. There is clear evidence of introgression of *E. b. gutselli* with *E. b. newmanii*, as Piller et al. (2008) argued. Examination of the combined data strongly suggests that Haponski and Stepien's (2008) result is based on their selective use of available cyt b and S7 intron data in building their trees.

The broader taxon sampling depicted in the combined analysis indicates that the evolutionary history of the *E. blennioides* species complex is decidedly more complex than the history presented by Haponski and Stepien (2008), with other evidence of gene flow and speciation not depicted in their trees. Their statement that *E. b. newmanii* and *E. b. pholidotum* are "invalid taxa", due to the fact that the taxa were recovered as polyphyletic in their trees, is taxonomically inaccurate. These taxa could be inaccurately delimited geographically or taxonomically. In fact, the combined analysis suggests that both of these taxa are more diverse (i.e., contain more species) than currently recognized. A proper taxonomic revision of the complex will likely result in the recognition of new taxa, in addition to the nominal *pholidotum* and *newmanii* forms, which will remain valid taxonomic names.

In summary, we show that the evolutionary history of the *E. blennioides* species complex presented by Haponski and Stepien (2008) is inaccurate and incomplete because of issues with their choice of outgroups, and (more seriously) their selective use of data from Piller et al. (2008). Our reanalysis of all of the available cyt b data clearly shows that the evolutionary histories of *E. blennius*, *E. b. gutselli* and Tennessee River basin populations of *E. b. newmanii* are intertwined. We believe that this result reflects past and present mitochondrial introgression, as argued by Piller et al. (2008) based on independent nuclear gene data. All other samples representing *E. b. blennioides*, *E. b. pholidotum*, and populations of *E. b. newmanii* north and west of the Tennessee River, form a separate major clade in all analyses, and it is clear from our combined analysis that subspecific names that have been assigned to these populations based on morphology, are not in agreement with the history inferred from genetic data. However, it is also clear that the taxonomic and nomenclatural issues within the *E. blennioides* complex cannot be resolved with limited amounts of molecular and morphological data, as Haponski and Stepien (2008) attempted to do with *E. b. gutselli*.

Acknowledgments

The authors would like to thank Brian Crother and Caleb McMahan, who graciously reviewed an earlier draft of this manuscript.

References

Bailey, R.M., Etnier, D.A., 1988. Comments on the subgenera of darters (Percidae) with descriptions of two new species from southcentral United States. Misc. Publ. Univ. Mich. Mus. Zool. 175, 1–48.

Bart Jr., H.L., Page, L.M., 1992. The influence of size and phylogeny on life history variation in North American percids. In: Mayden, R.L. (Ed.), Systematics, Historical ecology, and North American Freshwater Fishes. Stanford University Press, Stanford, CA., pp. 553–572.

Boschung Jr., H.T., Mayden, R.L., 2004. Fishes of Alabama. Smithsonian Press, Washington, DC.

Etnier, D.A., Starnes, W.C., 1993. The Fishes of Tennessee. Univ. Tenn. Press, Knoxville, TN.

Felsenstein, J., 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Haponski, A.E., Stepien, C.A., 2008. Molecular, morphological, and biogeographic resolution of cryptic taxa in the Greenside Darter *Etheostoma blennioides* complex. Mol. Phylo. Evol. 49, 69–83.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47, 3–8.

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. 4, 275–284.

Hubbs, C.L., Black, J.D., 1940. Percid fishes related to *Poecilichthys variatus*, with descriptions of three new forms. Occas. Pap. Mus. Zool. Univ. Mich. 416, 1–30.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Lee, D.S., Gilbert, C.R., Hocutt, C.H., Jenkins, R.E., McAllister, D.E., Stauffer Jr., J.R., 1980. Atlas of North American Freshwater Fishes. North Carolina State Museum of Natural History, Raleigh, NC. p. 854.

Maddison, W.P., Donoghue, M.J., Maddison, D.R., 1984. Outgroup analysis and parsimony. Syst. Zool. 33, 83–103.

Mayden, R.L., 1992. Systematics, Historical ecology and North American Freshwater Fishes. Stanford University Press, Stanford, CA.

McKeown, P.E., Hocutt, C.E., Morgan, R.P., Howard, J.T., 1984. An electrophoretic analysis of the *Etheostoma variatum* complex (Percidae: Etheostomatini), with associated zoogeographic considerations. Environ. Biol. Fishes 11, 95.

Miller, R.V., 1968. A systematic study of the greenside darter, *Etheostoma blennioides* Rafinesque (Pisces: Percidae). Copeia 1968, 1–40.

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

Page, L.M., 1983. Handbook of Darters. T.F.H. Publications, Neptune City, NJ.

Piller, K.R., Bart Jr., H.L., Hurley, D.L., 2008. Phylogeography of the greenside darter complex, *Etheostoma blennioides* (Teleostomi: Percidae): a wide-ranging polytypic taxon. Mol. Phylo. Evol. 46, 974–985.

Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51, 664–671.

Porter, B.A., Cavender, T.M., Fuerst, P.A., 2002. Molecular phylogeny of the snubnose darters, subgenus *Ulocentra* (genus *Etheostoma*, family Percidae). Mol. Phylo. Evol. 22, 364–374.

Porterfield, J.C. 1998. Phylogenetic systematic of snubnose darters (Percidae: *Etheostoma*), with discussion of reproductive behavior, sexual selection, and the evolution of male breeding color. Unpubl. PhD diss., Univ. of Illinois, Urbana Champaign, IL.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

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

Swofford, D.L., Olsen, J.L., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, MA., pp. 407–514.

Wheeler, W.C., 1990. Nucleic acid sequence phylogeny and random outgroups. Cladistics 6, 363–367.

Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51, 588–598.