ORIGINAL ARTICLE

AFLPs support deep relationships among darters (Percidae: Etheostominae) consistent with morphological hypotheses

TA Smith¹, TC Mendelson¹ and LM Page²

¹Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, MD, USA and ²Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

Recent attention has focused on the efficacy of amplified fragment length polymorphisms (AFLPs) for resolving deep evolutionary relationships. Here we show that AFLPs provide resolution of deep relationships within the family Percidae that are more consistent with previous morphological hypotheses than are relationships proposed by previous molecular analyses. Despite *in silico* predictions, we were able to resolve relatively ancient divergences, estimated at >25 MA. We show that the most distantly related species

share the fewest fragments, but suggest that large data sets and extensive taxon sampling are sufficient to overcome this obstacle of the AFLP technique for deep divergences. We compare genetic distances estimated from mitochondrial DNA with those from AFLPs and contrast traditional PAUP* Nei–Li AFLP genetic distances with a recently proposed method utilizing the Dice equation with constraining nucleotides. *Heredity* (2011) 107, 579–588; doi:10.1038/hdy.2011.50; published online 29 June 2011

Keywords: AFLP; darters; percidae; phylogeny; AFLP molecular clock; deep divergences

Introduction

Mitochondrial DNA (mtDNA) and nuclear introns are the most widely used genetic markers for phylogenetic analysis. However, discordance between topologies from nuclear and mitochondrial loci are common because of genes mutating at varying rates and being subjected to different stochastic processes (Maddison, 1997; Mendelson and Simons, 2006). Utilizing several loci and/or morphological characters have led to more robust topologies with greater statistical support, but, in many analyses, relationships utilizing multiple markers still remain unresolved or have low statistical support (Keck and Near, 2008; Ayache and Near, 2009). Examples where hybridization, introgression and saturation have confounded phylogenetic analysis are common in the literature (Mendelson and Simons, 2006; Keck and Near, 2010). Furthermore, for non-model organisms without a sequenced genome, primers for introns or other markers may not be available. Therefore, researchers may seek an alternative to using nucleotide sequences of a small number of loci for determining species relationships.

The amplified fragment length polymorphism (AFLP) technique (Vos *et al.*, 1995) is becoming a more common method of phylogenetic analysis, as it samples a large number of loci throughout the genome, including mitochondrial and nuclear DNA. AFLPs use a total digest of genomic DNA by two restriction endonu-

cleases, typically a rare cutter (six or more bases) and a common cutter (four bases) or two rare cutters. Theoretically, nucleotide polymorphisms that accrue over time between diverging species will yield different fragment lengths in the genomic digest. These different sized fragments are amplified by PCR, separated by gel electrophoresis and visualized by fluorescence. The pattern of fragments (bands on the gel) is used to determine patterns of speciation and to reconstruct evolutionary histories. Because AFLPs utilize the entire genome, they are able to detect rare and recent substitutions as well as retained alleles for hundreds or thousands of loci throughout the genome.

Although AFLPs underscore the importance of using a genome-wide set of genetic markers that includes both nuclear and mtDNA, their use is not without concerns. One main challenge of the AFLP technique in phylogenetics is that comigrating (same-size) fragments are assumed to be homologous, when in fact they may be homoplasious (Althoff *et al.*, 2007; Garcia-Pereira *et al.*, 2010). Homoplasy is not unique to AFLPs and presents a concern for all types of phylogenetic reconstruction. The likelihood of homoplasy because of saturation increases over time and can lead to hypotheses of close relationships in species that are distantly related. Using AFLPs may mitigate this problem; because different loci are subject to different stochastic processes and mutate at varying rates, utilizing the entire genome may capture phylogenetic signal that outweighs the noise of stochastic saturation.

Nonetheless, the probability of comigrating fragments being homologous decreases with species divergence, leading some to question the usefulness of AFLPs for deeper relationships (Althoff *et al.*, 2007; Garcia-Pereira *et al.*, 2010). A recent *in silico* study suggested that the utility of AFLPs breaks down at divergences >0.05

Correspondence: TA Smith, Department of Biological Sciences, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA.

E-mail: tsmith6@umbc.edu

Received 3 January 2011; revised 7 April 2011; accepted 24 May 2011; published online 29 June 2011

substitutions per position (from the most internal node to the tip of the tree), with all comigrating bands being non-homologous at a divergence (distance) of 0.4 substitutions per position (Garcia-Pereira *et al.*, 2010). In another *in silico* study of genomic data from *Drosophila*, Althoff *et al.* (2007) reported that only 14% (2/14) of comigrating fragments from the *D. simulans* and *D. yakuba* genomes (estimated to have diverged 6 MA) were homologous. Graves (2009), however, questioned the protocol of this study, citing a robust AFLP phylogeny of pinnipeds (Dasmahapatra *et al.*, 2009) in which authors eliminated bands that appeared to be non-homologous across all species (by only scoring those bands that amplified consistently with minimal size variation) and excluded smaller fragments. Using these methods, authors recovered moderate to strong phylogenetic signal at a divergence of ~15 MA (Higdon *et al.*, 2007; Dasmahapatra *et al.*, 2009).

In this study, we investigate the use of AFLP analysis to resolve deeper divergences in distantly related species of darters (Teleostei, subfamily Etheostominae). Darters are members of the family Percidae usually placed in four genera: *Ammocrypta*, *Crystallaria*, *Etheostoma* and *Percina*. They are one of the most speciose groups of freshwater fishes in North America with over 200 species. Based on analyses of the mitochondrial and nuclear DNA sequences calibrated to the absolute age of centrarchid fossils, darters are currently thought to have diverged from a most recent common ancestor (MRCA) around 25 million years ago (95% highest posterior density: 20.5–33.2 MA) (Hollingsworth and Near, 2009). Previous research using mtDNA and nuclear introns has not been able to fully resolve the darter phylogeny, and different genetic markers have yielded phylogenetic and morphological discordance; for example, the phylogenetic position of *Ammocrypta*, and the relationship of *Etheostoma zonale* to snubnose darters (Wood and Mayden, 1997; Song *et al.*, 1998; Sloss *et al.*, 2004; Keck and Near, 2008; Mendelson and Wong, 2010). Many sympatric darter species have been found to share haplotypes because of hybridization and mitochondrial replacement, confounding phylogenetic analyses (Ray *et al.*, 2008; Bossu and Near, 2009; Keck and Near, 2010).

Here we present an AFLP analysis of 69 darter species and show that most hypothesized relationships are consistent with previous morphological evidence. We propose that extensive taxon sampling and large numbers of AFLP fragments are sufficient to overcome the increased proportion of non-homologous comigrating bands in distantly related taxa. In addition, we propose an AFLP molecular clock based on previous studies that used sequence data calibrated to the absolute age of fossils of centrarchid fishes (Near and Benard, 2004; Near and Keck, 2005; Hollingsworth and Near, 2009). Finally, we show that the AFLP distance measures described by Mougél *et al.* (2002) may provide a more precise estimate of genetic distance than the more traditionally used Nei–Li equation.

Materials and methods

Taxonomic sampling and DNA extraction

Sixty-nine darter species representing the 4 genera and 25 subgenera recognized in the most recent classification

(Page, 2000) were used in this study (Table 1). Specimen identity and locality for each species in this study is presented in Supplementary Table 1. Two *Sander* (*S. canadense* and *S. vitreum*) and two *Perca* (*P. flavescens* and *P. fluviatilis*) species were used as outgroups to root the phylogeny. *Sander* and *Perca* are members of the subfamily Percinae (Song *et al.*, 1998). Genomic DNA from a single individual representing each species was extracted from tissue specimens using the QIAGEN DNeasy Tissue Kit (QIAGEN Inc., Valencia, CA, USA) and quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA).

AFLP protocol

The AFLP procedure was similar to that described by Vos *et al.* (1995). Genomic DNA (250 ng) was digested with *Eco*R1 and *Pst*I restriction endonucleases. Double-stranded adaptors were ligated onto the sticky ends of the genomic digest using a T4 DNA ligase (New England Biolabs, Ipswich, MA, USA). The product was diluted 1:10 and then subjected to two rounds of increasingly selective PCR amplification. In the first round, 2 µl of the diluted genomic enzyme digest and 1 µl of each primer were combined with 0.1 µl Taq DNA polymerase and 5.9 µl mastermix consisting of 0.05 µl 100 mM dNTPs (Fermentas, Glen Burnie, MD, USA), 0.30 µl 50 mM MgCl₂ (Invitrogen, Carlsbad, CA, USA), 4.55 µl dH₂O and 1 µl 10 × PCR reaction buffer (Invitrogen). A single selective nucleotide (adenine) was added to the end of the primer that was complementary to the adaptors ligated to the genomic digest (*Eco*R1: 5'-GACTGCGTACCAATTC-A; *Pst*I: 5'-GACTGCGTACATGCAG-A). A non-hot start Taq DNA polymerase (Go-Taq, Promega Corp., Madison, WI, USA) was used in the first selective amplification. In the second amplification, 2 µl of a 1:40 dilution of the first amplification was used as a template with 1 µl of each primer, 5.9 µl mastermix and 0.1 µl hot-start Taq DNA polymerase (Invitrogen, Fermentas or Qiagen). Primers in the second amplification had two additional selective nucleotides (A + GG, AG, AA, GA or AT), and the *Eco*R1 primer was fluorescently labeled (FAM, Invitrogen) for automated fragment analysis. A total of 20 primer pairs were used: *Eco*R1 + AGG with *Pst*I AAA, AAG, AGG, AGA: *Eco*R1 + AAG with *Pst*I AAA, AAG, AGG, AGA: *Eco*R1 + AAA with *Pst*I AAA, AAG, AGG, AGA: *Eco*R1 + AGA with *Pst*I AAA, AAG, AGG, AGA: *Eco*R1 + AAT with *Pst*I AAA, AAG, AGG, AGA.

One µl of the second round of amplified genomic digest was added to 9 µl Hi-Di Formamide (Applied Biosystems, Carlsbad, CA, USA), and was sent to Yale's DNA Analysis Facility for automated fragment analysis using a GeneScan LIZ 500 size standard (Applied Biosystems) on an ABIPrism 3730 DNA analyzer (Applied Biosystems). Fragments were analyzed using GeneMapper v.3.7 (Perkin-Elmer Corp., Waltham, MA, USA).

Phylogenetic analysis

Individuals were scored for the presence or absence of same sized fragments, or bands, using 1 to indicate the presence and 0 to show the absence of a band. GeneMapper panels were created for each primer pair for fragments between 75.0 and 450.0 bp using bin widths 0.4 bp wide. The threshold intensity for scoring

Table 1 Classification of Etheostominae from Page (2000), with recent changes in names of species

Genus <i>Ammocrypta</i>
<i>A. beanii</i>
<i>A. vivax</i>
Genus <i>Crystallaria</i>
<i>C. asprella</i>
Genus <i>Etheostoma</i>
Subgenus <i>Allohistium</i>
<i>E. cinereum</i>
Subgenus <i>Belophlox</i>
<i>E. fricksium</i>
<i>E. okaloosae</i>
Subgenus <i>Boleosoma</i>
<i>E. olmstedii</i>
Subgenus <i>Boleichthys</i>
<i>E. exile</i>
<i>E. gracile</i>
<i>E. fusiforme</i>
Least Darter group
<i>E. microperca</i>
<i>E. proeliare</i>
Subgenus <i>Catonotus</i>
Spottail Darter group
<i>E. crossopterygion</i>
<i>E. nigripinnatus</i>
<i>E. oophylax</i>
<i>E. squamiceps</i>
Fantail Darter group
<i>E. flabellare</i>
<i>E. kemicottii</i>
<i>E. percnurum</i>
Barcheek Darter group
<i>E. barbouri</i>
<i>E. smithi</i>
Subgenus <i>Doration</i>
<i>E. stigmaeum</i>
Subgenus <i>Etheostoma</i>
Greenside Darter group
<i>E. blennioides</i>
<i>E. blennioides</i>
<i>E. histrio</i>
Snubnose Darter group
<i>E. baileyi</i>
<i>E. coosae</i>
<i>E. orientale</i>
<i>E. zonale</i>
<i>E. zonistium</i>
Subgenus <i>Fuscatelum</i>
<i>E. parvipinnatus</i>
Subgenus <i>Loa</i>
<i>E. vitreum</i>
Subgenus <i>Litocara</i>
<i>E. sagitta</i>
Subgenus <i>Nothonotus</i>
<i>E. bellum</i>
<i>E. rufilineatum</i>
<i>E. tippecanoe</i>
Greenbreast Darter group
<i>E. jordani</i>
Spotted Darter group
<i>E. aquali</i>
Subgenus <i>Oligocephalus</i>
<i>E. caeruleum</i>
<i>E. hopkinsi</i>
<i>E. radiosum</i>
<i>E. swaini</i>
Orangethroat Darter group
<i>E. bison</i>
<i>E. lawrencei</i>
<i>E. luteovinctum</i>

Table 1 (Continued)

Subgenus <i>Ozarka</i>
<i>E. boschungii</i>
<i>E. tuscumbia</i>
Subgenus <i>Poecilichthys</i>
<i>E. osburni</i>
<i>E. variatum</i>
Subgenus <i>Vaillantia</i>
<i>E. chlorosoma</i>
Subgenus <i>Villora</i>
<i>E. edwini</i>
Genus <i>Percina</i>
Subgenus <i>Alvordius</i>
<i>P. macrocephala</i>
<i>P. maculata</i>
<i>P. peltata</i>
<i>P. roanoka</i>
Subgenus <i>Cottogaster</i>
<i>P. copelandi</i>
Subgenus <i>Ericosma</i>
<i>P. evides</i>
<i>P. palmaris</i>
Subgenus <i>Hadropterus</i>
<i>P. nigrofasciata</i>
<i>P. sciera</i>
<i>P. sipsi</i>
<i>P. smithvanizi</i>
Subgenus <i>Hypohomus</i>
<i>P. aurantiaca</i>
Subgenus <i>Imostoma</i>
<i>P. shumardi</i>
<i>P. vigil</i>
Subgenus <i>Odontopholis</i>
<i>P. stictogaster</i>
Subgenus <i>Percina</i>
<i>P. austroperca</i>
<i>P. caprodes</i>
Subgenus <i>Swainia</i>
<i>P. squamata</i>

a peak as 'present' varied among primer pairs between 400 and 3000. One challenge of working with AFLPs is that different primer pairs vary in their overall intensity, including the intensity of background noise, such that a single threshold cannot be applied across all primer pairs. Thresholds were therefore chosen to maximize phylogenetic signal and to minimize background noise. The threshold yielding the greatest phylogenetic signal (average bootstrap statistical support across nodes) for each primer pair was identified and applied individually by re-estimating the tree in PAUP* with the addition of each new primer pair. Automated peak scoring was modified manually by the removal of shoulders (stutter peaks) by a single individual who was unfamiliar with the taxon groups or labels to avoid systematic bias.

The resulting presence-absence matrix composed of 8056 characters was used to estimate phylogenetic relationships using distance and Bayesian methods. The matrix was translated into a distance matrix using the Dice distance equation with a Jukes-Cantor correction and 18 constraining nucleotides as described by Mougél *et al.* (2002) using DistAFLP (available from <http://pbil.univ-lyon1.fr/ADE-4/microb>). The Dice distance equation with constraining nucleotides is as follows:

$$d_{Dxy} = 1 - (n_{xy} / [n_{xy} + (A_{xy}/2)])^{1/r}$$

where n_{xy} is the number of fragments shared between individuals, Δ_{xy} is the number of fragments found only in individual x or y and r is the number of constraining nucleotides. This equation assumes that shared alleles between two species were present in a common ancestor halfway between them. The Jukes–Cantor correction, $\hat{f} = -(3/4) \ln[1 - (4/3)d_{D_{xy}}]$, accounts for unobserved substitutions because it assumes equal rates of substitutions between all base pairs. The number of shared fragments between individuals for each pairwise comparison was determined manually in Microsoft Excel. We calculated the number of constraining nucleotides based on the number of nucleotides involved in restriction and selective PCR: $r = E + P + N$, where E is the number of nucleotides at the restriction site of *EcoRI*, P is the number of nucleotides at the restriction site of *PstI* and N is the number of selective nucleotides ($r = 6 + 6 + 6$). These nucleotides constrain the results in that homologous fragments can be lost by a mutation at any one of the nucleotides involved in restriction or selective PCR; the greater the number of constraining nucleotides, the less likely a shared absence will be homologous. The distance matrix produced by DistAFLP with 1000 bootstrap replicates was analyzed with the Neighbor and Consense executables in PHYLIP Phylogeny Inference Package (Felsenstein, 1993) for neighbor-joining phylogenetic analysis.

The Nei–Li distance equation in PAUP* (Swofford, 2000) was also used to calculate Nei–Li genetic distances that traditionally have been used in AFLP phylogenetic analyses. PAUP* Nei–Li distances are based on a six base pair restriction site ($r = 6$) using the following equation from Nei and Li (1979):

$$\delta = -(3/2) \ln[(4\hat{S}^{1/2r} - 1)/3]$$

where $\hat{S}$ is an estimation of the proportion of ancestral restriction sites that remain unchanged in both individuals ($\hat{S} = 2n_{xy}/[x_x + n_y]$). Bootstrap analysis was performed in PAUP* using 1000 replicates and a neighbor-joining algorithm.

Finally, we conducted a Bayesian phylogenetic analysis in MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) using the restriction site binary model designed for restriction fragment data (for example, AFLPs) with the option 'noabsencesites' to compensate for bias introduced by the AFLP technique (characters that are absent in all taxa cannot be observed). Although Bayesian methods for AFLP markers have been developed, they are not computationally practical at this time. The analysis ran for 5 960 000 generations (convergence diagnostic < 0.01). Burn-in was determined to be 25% of the generations in all analyses using the software Tracer (available at <http://tree.bio.ed.ac.uk/software/tracer>) to determine the stationarity of all parameters. This large burn-in value was consistent with the observation that very divergent sequences take longer to converge.

AFLP molecular clock

We inferred the age of three nodes in our phylogeny based on estimates in the literature calibrated to the absolute age of fossils of centrarchid fishes (Near and Benard, 2004; Near and Keck, 2005). *Etheostoma tippecanoe* is estimated to have diverged from other species of the

subgenus *Nothonotus* $\sim 13.67 \pm 2.86$ MA (Near and Keck, 2005), *E. jordani* is estimated to have diverged from other *Nothonotus* $\sim 11.85 \pm 2.35$ MA (Near and Keck, 2005) and *Percina caprodes* and *P. austroperca* are estimated to have diverged $\sim 0.78 \pm 0.38$ MA (Near and Benard, 2004). To determine the AFLP genetic distances of *E. jordani* and *E. tippecanoe* from other species of *Nothonotus*, we used the average genetic distance of the two species from each of the remaining members of the *Nothonotus* group identified by our phylogeny. The three node age estimates were plotted as a function of the AFLP Dice genetic distance with a Jukes–Cantor correction and 18 constraining nucleotides ('DiceJC') and against PAUP* Nei–Li genetic distances in Excel to test for linearity (*Etheostoma barbouri* and *E. smithi* are estimated to have diverged around 16.3 MA (Hollingsworth and Near, 2009); however, this node age was excluded from the analysis because of the incongruence between mitochondrial and AFLP genetic distances).

The slope of the linear regression of AFLP genetic distance on the three node age estimates was used to determine the rate of AFLP divergence (as in Kropf *et al.*, 2009). The AFLP mutation rate was calculated from the divergence rate using the equation $D = 2\mu t' + 4N_{anc}\mu$, where D is the divergence rate, μ is the mutation rate, N_{anc} is the ancestral population size and t' is time expressed in generations after divergence (Kropf *et al.*, 2009). If we assume the ancestral population size was very small for each speciation event, $4N_{anc}\mu$ becomes negligible compared with the size of D , and the equation reduces to $D = 2\mu t'$. Divergence times for each species pair were then extrapolated using the equation of the linear regression of AFLP distance on node age.

AFLP and mitochondrial distance analysis

Cytochrome *b* (cyt *b*) sequences were obtained for 66 species from Genbank (<http://www.ncbi.nlm.nih.gov/genbank>), and jModelTest (Posada, 2008) was used to determine the parameters for the best fitting model of evolution. These parameters were implemented in PAUP* to estimate maximum likelihood (ML) genetic distances based on cyt *b*. Microsoft XLSTAT was used to perform a Mantel test to determine the significance of the correlation between ML cyt *b* and AFLP genetic distances.

Performance of AFLP at deeper divergences

Relative node ages of all internal nodes were determined using the DiceJC tree in the software Figtree (University of Edinburgh, v1.3.1). To estimate the performance of AFLPs across a range of divergence, relative node ages were compared with PAUP* bootstrap values and Bayesian posterior probabilities using Microsoft Excel. Microsoft Excel was also used to determine the correlation and statistical significance for each comparison.

Results

AFLP phylogenetic analysis

A total of 8056 characters were used in this analysis, of which 5942 (74%) were parsimony informative. The number of shared fragments between species ranged from a maximum of 235 to a minimum of 10 (average = 42 ± 0.360). A neighbor-joining cladogram

Figure 1 Neighbor-joining tree depicting hypothesized relationships among darters (Etheostominae), based on 8056 AFLP characters. *Perca* and *Sander* used as outgroups. Bootstrap support >50% (1000 replicates) for DiceJC genetic distances (left) and PAUP* Nei–Li genetic distances (right) are shown to the left of each node.

based on DiceJC genetic distance with bootstrap support >50% for both DiceJC and PAUP* Nei–Li genetic distances is shown in Figure 1. The Bayesian tree with posterior probability >0.5 is shown in Figure 2. *Percina*, *Ammocrypta* and *Etheostoma* are monophyletic: *Percina* 100% DiceJC bootstrap support (0.99 Bayesian posterior probability), *Ammocrypta* 100% (1.0), *Etheostoma* 52% or 64% (0.84) excluding *Etheostoma cinereum*. *E. cinereum*, the only member of the subgenus *Allohistium*, was recovered as sister to all other *Etheostoma* species with moderate bootstrap support (52%). However, Bayesian analysis (Figure 2) yielded a polytomy between *Percina*, *E. cinereum* and the remaining species of the genus *Etheostoma*.

One difference was observed between the topologies of the DiceJC and PAUP* Nei–Li distance-based trees. The tree based on DiceJC distances recovered *E. radiosum* as sister to a group consisting of *E. bison*, *E. lawrencei*,

E. caeruleum, *E. swaini*, *E. luteovinctum* and *E. exile* with strong bootstrap support (Figure 1). As in the Bayesian tree (Figure 2), the tree based on PAUP* Nei–Li distances recovered *E. radiosum* as sister to *E. bison* and *E. lawrencei*, with moderate bootstrap support (50%).

AFLP molecular clock

We indirectly calibrated the AFLP molecular clock using three estimated node ages from mtDNA and nuclear intron analyses that were calibrated to the absolute age of centrarchid fossils (Near and Benard, 2004; Near and Keck, 2005). We found a strong positive linear correlation with respect to time for both DiceJC and PAUP* Nei–Li distances ($r=0.997$; 0.998) (Figure 3). However, because of the limited calibrations to centrarchid fossils thus far, and because darter fossils are rare, we point out that the relationship between time and genetic distance may not

Figure 2 Bayesian Tree showing relationships among darters (Etheostomatinae). Posterior probabilities >0.50 shown and are based on 5960000 generations with a 25% burn-in.

be linear, particularly at node ages <1–2 million years. Ho *et al.* (2005) show that estimates of mutation rates are higher for recent divergences (<1–2 MA), which could lead to an overestimation of recent divergence times. Calibrations with additional fossils are needed to assess the utility of our molecular clock estimates.

We estimated an AFLP divergence rate of $\sim 0.0036 \pm 0.0024$ nucleotide substitutions per site per million years using DiceJC genetic distances, or 0.0111 ± 0.0068 using PAUP* Nei–Li distances. The estimated AFLP mutation rate, calculated from the divergence rate and assuming darters have one generation per year (Page, 1983), was found to be between 6×10^{-10} and 3×10^{-9} or between 2×10^{-9} and 9×10^{-9} nucleotide substitutions per site per generation using the DiceJC and the Nei–Li divergence rates, respectively. This mutation rate is an approximation because of the uncertainty in generation time (for example, although rare, some species do not mature until 2 years

of age, but then reproduce annually) and ancestral population size. A maximum DiceJC distance of 0.208 nucleotide substitutions per site was found in our study between *P. fluviatilis* (an outgroup) and *P. caprodes*, which would translate to ~ 48.9 MA using our calibrated clock. The maximum PAUP* Nei–Li distance for this species pair was 0.638, which also translates to a MRCA ~ 48.9 MA. The maximum distance within the darter ingroup was 0.177/0.540 between *E. hopkinsi* and *C. asprella*, which translates to a node age of ~ 40 MA (40.3 MA DiceJC and 40.1 MA PAUP* Nei–Li). A strong positive correlation was found between DiceJC and PAUP* Nei–Li node ages for each interspecific pairwise species comparison ($r = 0.9999$, $P < 0.0001$). We used our calibrated clock to estimate divergence times for all pairwise comparisons in our phylogeny (data not shown), and these node ages were compared with previously published divergence estimates based on mitochondrial and nuclear sequences.

Figure 3 The estimated age of nodes of the MRCA as a function of the AFLP pairwise genetic distances for three species-pairs of darters. Node ages and genetic distances were from *Etheostoma tippecanoe* from other *Nothonotus* darters (*E. jordani*, *E. bellum*, *E. rufileatum*, *E. aquali*), *E. jordani* from other *Nothonotus* darters (*E. bellum*, *E. rufileatum*, *E. aquali*) and *Percina caprodes* and *P. austroperca* (Near and Benard, 2004; Near and Keck, 2005). A full color version of this figure is available at the *Heredity* journal online.

Figure 4 Pairwise interspecific AFLP genetic distances as a function of ML cyt b genetic distances for 66 *Percidae* species. AFLP genetic distances were calculated with PAUP* Nei-Li and the Dice equation with a Jukes-Cantor correction using 18 constraining nucleotides (Mougel *et al.*, 2002). Correlation, PAUP* Nei-Li $r = 0.717$, DiceJC $r = 0.718$, Mantel Test, $P < 0.0001$. A full color version of this figure is available at the *Heredity* journal online.

AFLP and mitochondrial distance analysis

Supplementary Figure 1 shows the relationship between the number of shared AFLP bands and genetic distance, based on two methods of calculating genetic distance from restriction markers (DiceJC and PAUP* Nei-Li). DiceJC and PAUP* Nei-Li distances are strongly correlated with each other ($r = 1.00$, $P < 0.001$), and both increase logarithmically as the number of shared bands decrease; however, the means of the two genetic distances are significantly different (DiceJC: 0.131 ± 0.0004 , PAUP* Nei-Li: 0.398 ± 0.001 ; $t = 312$, $P < 0.0001$). PAUP* Nei-Li distances also exhibit a broader range (0.082–0.638) and much more scatter about the regression line than the DiceJC estimates (0.027–0.208) (residual sum of squares: PAUP* Nei-Li = 1.77, DiceJC = 0.183).

The pairwise comparison of DiceJC and PAUP* Nei-Li genetic distances with ML cyt b genetic distances resulted in a strong positive correlation (DiceJC: $r = 0.718$; PAUP*

Figure 5 Relationship between relative node age (as estimated in Figure) and statistical support >50% (0.50). Bootstrap $r = -0.526$, $P < 0.0001$. Posterior probability $r = -0.212$, $P = 0.123$. Large blue square indicates two data points at the same location. A full color version of this figure is available at the *Heredity* journal online.

Nei-Li: $r = 0.717$; Mantel Test: $P < 0.0001$) (Figure 4). Neither distance appears to be saturating.

Performance of AFLPs at deeper divergences

A statistically significant negative correlation was found between bootstrap statistical support and relative node age as estimated in FigTree (bootstrap: $r = -0.526$, $P = 3.76 \times 10^{-5}$) (Figure 5), indicating that older nodes had overall less bootstrap support than more recent nodes. That correlation remained statistically significant after the removal of two extreme data points ($r = -0.400$, $P = 0.003$). Posterior probabilities were not significantly correlated with relative node ages ($r = -0.212$, $P = 0.123$), except when two extreme data points were removed ($r = -0.374$, $P = 0.006$).

Discussion

In this study we used AFLPs to infer relationships among mostly distantly related darter species representing four genera and 25 subgenera. These appear to be the deepest divergences analyzed using AFLPs to date (Graves, 2009; Dasmahapatra *et al.*, 2009). We obtained a relatively robust tree with species of accepted genera and long-recognized subgenera clustering with statistical support. Relationships hypothesized from AFLPs agree with morphological hypotheses (for example, Bailey and Gosline, 1955; Page, 2000) to a greater extent than previous molecular analyses of nuclear and mitochondrial markers.

Phylogenetic relationships

The morphology-based genera first proposed in Bailey *et al.* (1954) and Bailey and Gosline (1955) have been almost universally accepted by researchers investigating evolutionary relationships among darters. One exception was Hubbs (1959), who continued to recognize *Percina* only for the morphologically distinctive logperches and *Hadropterus* for other species assigned by Bailey to *Percina*. In addition, several authors have recognized *Crystallaria* as a genus rather than as a subgenus of *Ammocrypta* (Simons, 1991; Page, 2000; Wood and Raley, 2000) and, most recently, *Nothonotus* as been recognized as a genus rather than as a subgenus of *Etheostoma* (Near

and Keck, 2005). The AFLP data strongly support the monophyly of *Percina* with the logperches deeply embedded within the genus. Moreover, the 46 species of *Percina* share a morphological synapomorphy, the presence of scutes on the breasts of males (Bailey *et al.*, 1954; Page, 1974).

The monophyly of *Ammocrypta* (with *Crystallaria* as a subgenus) or of *Ammocrypta* + *Crystallaria* has been nearly universally accepted since first recognized by Bailey *et al.* (1954), except by Simons (1992) who hypothesized *Ammocrypta* to be nested within *Etheostoma*. Our results strongly support the sister-group relationship of *Ammocrypta* and *Crystallaria*, and hypothesize this group to be sister to all other darters and not embedded within *Etheostoma*. The group containing the eight species of *Ammocrypta* and *Crystallaria* share a derived, exceptionally slender and translucent body. Recognition of *Crystallaria* as a genus is arbitrary but generally accepted because of the morphological distinctiveness of *Crystallaria* from species of *Ammocrypta* (Simons, 1991).

Although *Etheostoma* has never been diagnosed satisfactorily as a genus with morphological synapomorphies, it has been nearly universally accepted as monophyletic. The exception was Near and Keck (2005), who treated *Nothonotus* as a genus based on mtDNA sequence data in Song *et al.* (1998) and Sloss *et al.* (2004). However, Song *et al.* (1998) concluded that its phylogenetic position 'cannot be resolved with the cyt *b* dataset, as evidenced by the lack of bootstrap and decay support.' Sloss *et al.* (2004) found the position of *Nothonotus* to vary greatly depending on the type of analysis. Although these studies suggest that *Nothonotus* may be as divergent from other darters as the recognized genera are from one another, they neither identify the closest relative to *Nothonotus* nor indicate that *Etheostoma* with *Nothonotus* included creates a paraphyletic genus.

Our data show *Etheostoma* including *Nothonotus* to form a monophyletic group. The species of *Etheostoma* most distantly related to all other *Etheostoma* in our analysis is *E. cinereum*, a species that previously has been suggested to be the most divergent in the genus (Page and Whitt, 1973). Sloss *et al.* (2004) found *E. cinereum* to be the sister to the remaining *Etheostoma*, embedded in *Etheostoma*, or sister to *Ammocrypta* + *Crystallaria*, depending on the analysis. Our results indicate that recognizing *Nothonotus* as a genus and not *Allohistium* for *E. cinereum* creates a paraphyletic *Etheostoma*.

In addition to supporting the monophyly of long-established genera, AFLPs hypothesize a sister relationship between *Percina* and *Etheostoma*. Although this has been suggested previously in some analyses (for example, in a ML tree of mtDNA data by Sloss *et al.*, 2004), most workers have treated *Percina* as sister to the remaining genera, leaving *Ammocrypta* as the sister to *Etheostoma* (Kuehne and Barbour, 1983; Page, 1983). Other interesting relationships, some previously not hypothesized, among species and subgenera are suggested by the AFLP data, but await discussion pending the analysis of data on more species.

AFLP molecular clock

Based on our AFLP molecular clock, we estimate the MRCA of the subfamily Etheostominae ('darters') to have existed ~40 MA. Our estimated age of the MRCA

of *Percina* and the *Etheostoma* subgenus, *Catonotus*, of ~28 MA are similar to Hollingsworth and Near's (2009) estimated age of the MRCA of ~25 MA (95% highest posterior density: 20.5–33.2 MA). Likewise, Hollingsworth and Near (2009) estimated the barcheek darter group to have diverged from the fantail darter group (subgenus *Catonotus*) ~17 MA (95% highest posterior density: ~13–22). Our clock estimates the age of this node to be ~16 MA.

A notable exception to previous estimates is that *E. barbouri* and *E. smithi* are much more closely related (age of node ~7.0 MA) in our analysis than Hollingsworth and Near's (2009) estimate based on a multigene phylogeny using mitochondrial and nuclear loci (16.3 MA, 95% highest posterior density: 12.4–20.5). *E. barbouri* and *E. smithi* have a smaller genetic distance based on AFLP (nuclear) DNA than that based on mitochondrial cyt *b* nucleotide sequence (Figure 4). We interpret this cytonuclear discordance as evidence that one or the other of this species pair experienced mitochondrial replacement at some time in the past, most likely *E. barbouri*. Previous phylogenetic analyses showed that two mtDNA loci (cyt *b*, ND2) place *E. barbouri* outside the barcheek darter group whereas, AFLP, S7 nuclear intron and morphological data all recover *E. barbouri* as a barcheek darter (Page *et al.*, 2003; Mendelson and Simons, 2006). Interpretations of node age for *E. barbouri* and *E. smithi* are likely to be more reliable if based on nuclear rather than mtDNA distances.

AFLP and mitochondrial distance analysis

Restriction-site distance methods can be used with AFLP fragments because AFLP bands show a similar statistical behavior to restriction site data (Mougel *et al.*, 2002). Genetic distances estimated by two restriction-site distance metrics, DiceJC and PAUP* Nei–Li, are strongly correlated ($r=1.00$); however, PAUP* Nei–Li distances yield a greater amount of scatter around the regression line of distance on shared bands as compared with DiceJC distances (Supplementary Figure 1). This scatter suggests that PAUP* generated Nei–Li distances may not as accurately reflect genetic relationships, and perhaps suggests that, as shown by Mougel *et al.* (2002), constraining nucleotides are important in calculations of genomic divergence. Importantly, however, neither the topology (with the exception of one species, *E. radiosum*) nor the statistical support of nodes in the neighboring tree based on PAUP* Nei–Li distances were noticeably different from the DiceJC tree.

Several studies have suggested that the homology of comigrating bands should decrease with increasing genetic divergence. Garcia-Pereira *et al.* (2010) estimate that the utility of AFLPs breaks down at divergences >0.05 substitutions per position using PAUP* Nei–Li distances, with all comigrating bands being non-homologous at a divergence of 0.4 substitutions per position. Althoff *et al.* (2007) also predict that the number of comigrating bands containing homologous fragments decreases with increasing genetic distance. If this were the case in our study, we would expect to see a greater amount of scatter about the regression line of AFLP distances on ML cyt *b* distances at larger distances. As cyt *b* distances are derived from truly homologous loci, a

greater amount of scatter at larger distances would indicate a greater proportion of homoplasious AFLP bands. As in Dasmahapatra *et al.*, 2009, we show a strong positive relationship between PAUP* Nei–Li distances and ML cyt *b* distances between 0.082 and 0.638 substitutions per position, with a seemingly negligible increase in scatter about the regression line at larger distances, especially for DiceJC distances (Figure 4). This pattern suggests the phylogenetic signal-to-noise ratio is approximately constant through this level of divergence. Furthermore, because we are recovering moderate to strong phylogenetic signal at these deeper divergences, we suggest that a sufficient number of comigrating bands are indeed homologous at this level of divergence, and that perhaps large datasets and increased taxon sampling are sufficient to overcome noise created by homoplasious bands.

Performance of AFLP at deeper divergences

Our analysis of statistical support across increasing relative node ages indicates a significant effect of node age on bootstrap support but not on Bayesian posterior probability, unless two extreme data points are removed. This pattern suggests that statistical support for the deepest divergences relies on a few highly informative characters. Because the bootstrap method samples with replacement, a proportion of bootstrap replicates may fail to sample the few important characters that define older nodes. However, because both distance and Bayesian analyses yield moderate to strong phylogenetic signal for deeper relationships, and because posterior probabilities, which sample every character, do not show a significant correlation between node age and statistical support (Figure 5), we suggest that AFLPs are suitable for evaluating relationships among deeply divergent taxa.

We further suggest that extensive taxon sampling and large datasets (that is, a large number of AFLP fragments) may be necessary to overcome noise created by homoplasious fragments. As divergence time increases, species share fewer alleles (fragments), and the shared derived characters that once defined internal relationships are reduced in frequency. Sampling only a small fraction of the extant species or using a small number of characters could fail to detect these rare alleles shared between distantly related taxa. The greater the fraction of extant species (or individuals) sampled for a given taxonomic group, and the larger the AFLP dataset, the more likely two taxa would share common alleles (fragments) to link deeper relationships. Capturing these ‘intermediate’ species should therefore allow us to better detect internal relationships and reconstruct evolutionary histories. Most importantly, however, our study indicates that AFLPs can be used to resolve deeper divergences and determine species relationships at multiple levels of divergence.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We would like to thank P Ceas, D Eisenhour, B Freeman, P Harris, B Kuhajda, R Mayden, M Retzer, F Rohde, D Smith, M Thomas, S Walsh and R Wood for their

generous tissue contributions, and J Shultz, M Burns and K Omland for helpful comments on earlier drafts of the manuscript. We also are very grateful to N Lang for providing genomic DNA for our outgroups. This work was funded in part by a GAANN fellowship to TAS, NSF DEB-0718987 to TCM, and supported by CBI NIH T32GM066706 to TAS.

References

- Althoff DM, Gitzendanner MA, Segraves KA (2007). The utility of amplified fragment length polymorphisms in phylogenetics: a comparison of homology within and between genomes. *Syst Biol* **56**: 477–484.
- Ayache NC, Near TJ (2009). The utility of morphological data in resolving the phylogenetic relationships of darters as exemplified with *Etheostoma* (Teleostei: Percidae). *Bull Peabody Mus Nat Hist* **50**: 327–346.
- Bailey RM, Gosline WA (1955). Variation and systematic significance of vertebral counts in the American fishes of the family Percidae. *Misc Publ Mus Zool Univ Mich* **93**: 5–44.
- Bailey RM, Winn HE, Smith CL (1954). Fishes from the Escambia river, Alabama and Florida, with ecologic and taxonomic notes. *Proc Acad Nat Sci Phila* **106**: 109–164.
- Bossu CM, Near TJ (2009). Gene trees reveal repeated instances of mitochondrial DNA introgression in orangethroat darters (Percidae: *Etheostoma*). *Syst Biol* **58**: 114–129.
- Dasmahapatra KK, Hoffman JI, Amos W (2009). Pinniped phylogenetic relationships inferred using AFLP markers. *Heredity* **103**: 168–177.
- Felsenstein J (1993). *PHYLIP Phylogeny Inference Package* Version 3.5c. Distributed by the author. Department of Genetics, University of Washington: Seattle.
- Garcia-Pereira MJ, Caballero A, Quesada H (2010). Evaluating the relationship between evolutionary divergence and phylogenetic accuracy in AFLP data sets. *Mol Biol Evol* **27**: 988–1000.
- Graves J (2009). Phylogenetic analyses using AFLPs: Deeper AFLPs. *Heredity* **103**: 99.
- Higdon JW, Bininda-Emonds ORP, Beck RMD, Ferguson SH (2007). Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evol Biol* **7**: 216.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* **22**: 1561–1568.
- Hollingsworth PR, Near TJ (2009). Temporal patterns of diversification and microendemism in Eastern Highland endemic barcheck darters (Percidae: Etheostomatinae). *Evolution* **63**: 228–243.
- Hubbs C (1959). Laboratory hybrid combinations among Etheostomatine fishes. *Tex J Sci* **XI**: 49–56.
- Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754–755.
- Keck BP, Near TJ (2010). Geographic and temporal aspects of mitochondrial replacement in *Nothonotus* darters (Teleostei: Percidae: Etheostomatinae). *Evolution* **64**: 1410–1428.
- Keck BP, Near TJ (2008). Assessing phylogenetic resolution among mitochondrial, nuclear, and morphological datasets in *Nothonotus* darters (Teleostei: Percidae). *Mol Phylogenet Evol* **46**: 708–720.
- Kropf M, Comes HP, Kadereit JW (2009). An AFLP clock for the absolute dating of shallow-time evolutionary history based on the intraspecific divergence of southwestern European alpine plant species. *Mol Ecol* **18**: 697–708.
- Kuehne RA, Barbour RW (1983). *The American Darters*. University Press of Kentucky: Lexington, Kentucky.
- Maddison WP (1997). Gene trees in species trees. *Sys Bio* **46**: 523–536.

- Mendelson TC, Simons JN (2006). AFLPs resolve cytonuclear discordance and increase resolution among barcheek darters (Percidae: Etheostoma: Catonotus). *Mol Phylogenet Evol* **41**: 445–453.
- Mendelson TC, Wong MK (2010). AFLP phylogeny of the snubnose darters and allies (Percidae: Etheostoma) provides resolution across multiple levels of divergence. *Mol Phylogenet Evol* **57**: 1253–1259.
- Mougel C, Thioulouse J, Perriere G, Nesme X (2002). A mathematical method for determining genome divergence and species delineation using AFLP. *Int J Syst Evol Microbiol* **52**: 573–586.
- Near TJ, Benard MF (2004). Rapid allopatric speciation in logperch darters. *Evolution* **58**: 2798–2808.
- Near TJ, Keck BP (2005). Dispersal, vicariance, and timing of diversification in Nothonotus darters. *Mol Ecol* **14**: 3485–3496.
- Nei M, Li WH (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS USA* **76**: 5269–5273.
- Page LM (1974). The subgenera of Percina (Percidae: Etheostomatini). *Copeia* **1**: 66–86.
- Page LM (1983). *Handbook of Darters*. TFH Publications: Neptune City: NJ.
- Page LM (2000). Etheostomatinae. In: Craig JF (ed.). *Percid Fishes: Systematics, Ecology and Exploitation*. Blackwell Science: Oxford. pp 225–253.
- Page LM, Hardman M, Near TJ (2003). Phylogenetic relationships of barcheek darters (Percidae: *Etheostoma*, Subgenus *Catonotus*) with descriptions of two new species. *Copeia* **3**: 512–530.
- Page LM, Whitt GS (1973). Lactate dehydrogenase isozymes, malate dehydrogenase isozymes and tetrazolium oxidase mobilities of darters (Etheostomatini). *Comp Biochem Physiol* **44B**: 611–623.
- Posada D (2008). jModelTest: Phylogenetic model averaging. *Mol Biol Evol* **25**: 1253–1256.
- Ray JM, Lang NJ, Wood RM, Mayden RL (2008). History repeated: recent and historical mitochondrial introgression between the current darter *Etheostoma uniporum* and rainbow darter *Etheostoma caeruleum* (Teleostei: Percidae). *J Fish Biol* **72**: 418–434.
- Ronquist F, Huelsenbeck JP (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Simons AM (1991). Phylogenetic relationships of the crystal darter, *Crystallaria asprella* (Teleostei: Percidae). *Copeia* **4**: 927–936.
- Simons AM (1992). Phylogenetic relationships of the Boleosoma species group (Percidae: Etheostoma). In Mayden RL (ed.). *Systematics, Historical Ecology, and North American Freshwater Fishes*. Stanford University Press: Stanford, CA.
- Sloss BL, Billington N, Burr B (2004). A molecular phylogeny of the Percidae (Teleostei, Periformes) based on mitochondrial DNA sequence. *Mol Phylogenet Evol* **32**: 545–562.
- Song CB, Near TJ, Page LM (1998). Phylogenetic Relations among percid Fishes as inferred from mitochondrial cytochrome *b* DNA sequence data. *Mol Phylogenet Evol* **10**: 343–353.
- Swofford DL (2000). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.
- Wood RM, Mayden RL (1997). Phylogenetic relationships among selected darter subgenera (Teleostei: Percidae) as inferred from analysis of allozymes. *Copeia* **2**: 265–274.
- Wood RM, Raley ME (2000). Cytochrome *b* sequence variation in the crystal darter *Crystallaria asprella* (Actinopterygii: Percidae). *Copeia* **1**: 20–26.
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J *et al.* (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**: 4407–4414.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)