1-1-2017

# Phylogeography and species delimitation in convict cichlids (Cichlidae: Amatitlania): Implications for taxonomy and Plio-Pleistocene evolutionary history in Central America

Justin C. Bagley
*Brigham Young University*

Wilfredo A. Matamoros
*Universidad de Ciencias y Artes de Chiapas*

Caleb D. Mcmahan
*Louisiana State University*

Michael Tobler
*Kansas State University*

Prosanta Chakrabarty
*Louisiana State University*

*See next page for additional authors*



## Recommended Citation

Bagley, J., Matamoros, W., Mcmahan, C., Tobler, M., Chakrabarty, P., & Johnson, J. (2017). Phylogeography and species delimitation in convict cichlids (Cichlidae: Amatitlania): Implications for taxonomy and Plio-Pleistocene evolutionary history in Central America. *Biological Journal of the Linnean Society, 120* (1), 155-170. https://doi.org/10.1111/bij.12845

The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

*Biological Journal of the Linnean Society,* 2016, ••, ••–••. With 4 figures.

# Phylogeography and species delimitation in convict cichlids (Cichlidae: *Amatitlania*): implications for taxonomy and Plio–Pleistocene evolutionary history in Central America

JUSTIN C. BAGLEY[1]*, WILFREDO A. MATAMOROS[2], CALEB D. MCMAHAN[3,4], MICHAEL TOBLER[5], PROSANTA CHAKRABARTY[3] and JERALD B. JOHNSON[1,6]

[1]*Evolutionary Ecology Laboratories, Department of Biology, Brigham Young University, Provo, UT, 84602, USA*
[2]*Facultad de Ciencias Biológicas, Museo de Zoología, Colección de Ictiología, Universidad de Ciencias y Artes de Chiapas, CP 29039, Tuxtla Gutierrez, Chiapas, México*
[3]*Museum of Natural Science (Ichthyology), Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, 70803, USA*
[4]*The Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, IL, 60605, USA*
[5]*Division of Biology, Kansas State University, Manhattan, KS, 66506, USA*
[6]*Monte L. Bean Life Science Museum, Brigham Young University, Provo, UT, 84602, USA*

*Received 24 January 2016; revised 2 June 2016; accepted for publication 6 June 2016*

We investigate phylogeographic patterns and delimit species boundaries within *Amatitlania*, a genus of Central American cichlid fishes. Phylogenetic analyses of mitochondrial DNA sequences from 318 individuals spanning the geographical ranges of all three currently recognized *Amatitlania* species strongly supported one major clade, with a relatively diverged subclade corresponding to *A. kanna* samples from eastern Costa Rica and Panama. Gene trees and networks revealed marked incongruences between phylogeographic structure and morpho-species taxonomy as a result of species-level polyphyly. Bayes factor comparisons of species delimitation models accounting for incomplete lineage sorting under the multispecies coalescent decisively supported the recognition of two distinct species within *Amatitlania* corresponding to *Amatitlania nigrofasciata* and *A. kanna* lineages. The only clearly genetically and morphologically diagnosable species was *A. kanna*. These results strongly suggest that incomplete lineage sorting provides the best explanation for the polyphyly of *A. kanna*, whereas the polyphyly of *A. siquia* is likely a result of an imperfect taxonomy. Additional insights from coalescent-dating, network, and historical demographic analyses suggested that the two species of *Amatitlania* diversified only since the early Pleistocene, and that *A. nigrofasciata* experienced population expansions from approximately 200 000 years ago in the mid-late Pleistocene onward. We discuss implications of our results for the taxonomy and evolutionary history of *Amatitlania* and, more broadly, of Central American freshwater fishes. © 2016 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2016, **00**, 000–000.

KEYWORDS: *BEAST – Bayes factor delimitation – freshwater fishes – historical demography – mitochondrial DNA – model comparison – species trees – taxonomy.

## INTRODUCTION

Phylogeography has proven remarkably useful for understanding processes influencing the historical diversification of biotic lineages at and below the species level, as well as delimiting morphologically

*Corresponding author. Current addresses: Departamento de Zoologia, Universidade de Brasília, 70910-900 Brasília, DF, Brazil; Departamento de Zoologia e Botânica, IBiLCE, Universidade Estadual Paulista, 15054-000 São José do Rio Preto, SP, Brazil. E-mail: jcbagley@unb.br

'cryptic' species (Avise, 2000; Pons *et al.*, 2006; Bickford *et al.*, 2007). By linking genealogical divergences within species with data on geographical barriers and Earth history events, phylogeography permits the inference of how historical processes have shaped intraspecific genetic variation (Bermingham & Martin, 1998; Avise, 2000; Arbogast & Kenagy, 2001; Zink, 2002; Bagley & Johnson, 2014a). Phylogeography also provides critical information for systematics and conservation. Congeneric phylogeographic studies infer species-level phylogenies (Perdices *et al.*, 2002; Bagley *et al.*, 2011; Unmack *et al.*, 2012) and, although morphology-based taxonomy is often subjective and prone to overestimate or underestimate the number of distinct species or genetic lineages (e.g. 'splitters' vs. 'lumpers'; Dayrat, 2005), phylogeographic sampling can provide a more accurate picture of biodiversity in a clade (Moritz & Faith, 1998; Bickford *et al.*, 2007). That said, interpreting the number of species in a sample directly from gene trees is also subjective and potentially destabilizes taxonomy (O'Meara, 2010; Fujita *et al.*, 2012). Fortunately, a growing number of coalescent-based species delimitation methods (Pons *et al.*, 2006; Yang & Rannala, 2010; Grummer, Bryson & Reeder, 2014) hold promise for objectively defining species boundaries using genetic data. Applying these delimitation approaches using phylogeographic sampling can improve taxonomy, helping to avoid inaccurate biodiversity estimates and misallocation of conservation resources (Sites & Marshall, 2003; Agapow *et al.*, 2004).

The Neotropics present a 'preferred target' for biodiversity research because they encompass species-rich areas of the world (Rull, 2011), including multiple biodiversity 'hotspots' (Myers *et al.*, 2000). Within Neotropical North and South America, the freshwater fish assemblage comprises the most diverse group of vertebrates and, with approximately 7000 described and undescribed species, makes up almost half of global freshwater fish species richness (Reis, Kullander & Ferraris, 2003; Albert & Reis, 2011) and 11.2% of vertebrate species richness (IUCN, 2016). Despite their exceptional diversification, our knowledge of the processes responsible for the genetic diversity, species limits, and intraspecific diversification of Neotropical freshwater fishes remains limited in groups for which phylogeographic perspectives have yet to be developed.

The present study focuses on phylogeography and genetic-based species delimitation in *Amatitlania*, a wide-ranging genus of Neotropical 'convict cichlids' (family Cichlidae) endemic to fresh waters of the Central American (CA) Isthmus (Kullander, 2003; Schmitter-Soto, 2007a; Schmitter-Soto, 2007b). *Amatitlania* presently includes three species (Schmitter-Soto, 2007a; McMahan *et al.*, 2014). Günther (1867) originally described *Amatitlania nigrofasciata*, the type species of the genus, as '*Heros nigrofasciatus*' based on material from Lago Amatitlán, Guatemala. Subsequently, the species status of *A. nigrofasciata* has remained largely unquestioned, although its generic placement changed greatly, from *Heros* to *Cichlasoma* (Jordan & Evermann, 1898; and subgenus '*Archocentrus*' of Miller, 1966), to *Archocentrus* (Allgayer, 1994), to *Cryptoheros* (Allgayer, 2001). As '*Archocentrus nigrofasciatus*', the convict cichlid was traditionally considered to range from Guatemala to north-western Panama (Bussing, 1976; Bussing, 1998). Recently, however, in a morphological revision of *Archocentrus*, Schmitter-Soto (2007a) erected the genus *Amatitlania* for convict cichlids and re-described *A. nigrofasciata* as ranging from the Río Suchiate, Guatemala to the Río Sucio, El Salvador, on the Pacific versant, and from the Río Patuca, Honduras to Río Jutiapa, Guatemala, on the Atlantic. Schmitter-Soto (2007a) also described three new species from *A. nigrofasciata* material: *Amatitlania coatepeque*, endemic to Lago Coatepeque, El Salvador; *Amatitlania kanna*, ranging from the Cañaveral, Cricamola and Sixaola rivers of Costa Rica to the Atlantic drainages of north-western Panama, including the Changuinola, San San, and Róbalo rivers; and *Amatitlania siquia*, ranging from Atlantic and Pacific Costa Rica and Nicaragua through Río Yeguaré on the Pacific versant of Honduras. McMahan *et al.* (2014) undertook the first molecular and morphological analysis of *Amatitlania* to test the taxonomic distinctiveness of *A. coatepeque*. McMahan *et al.* (2014) resolved *A. coatepeque* as phylogenetically nested within *A. nigrofasciata*, with sufficient variation in diagnostic morphological characters rendering it invalid as a species; thus, they synonymized *A. coatepeque* within *A. nigrofasciata*.

The geological and paleoclimatic records of CA attest to a complex regional history over late Pliocene–recent (e.g. tectonism, cyclic cooling/drying during Pleistocene glaciations), making CA ideal for understanding the effects of Earth history on genetic diversity and historical demography (Bagley & Johnson, 2014a; Bagley & Johnson, 2014b). Given their wide distribution across geographical barriers in CA, *Amatitlania* provide excellent opportunities for studying species past demographic responses to historical processes. Here, we infer the phylogeographic history of *Amatitlania* and critically test the current taxonomic hypothesis of species limits in the genus using evolutionary genetic analyses of mitochondrial DNA (mtDNA) sequences. The objectives of the present study were three-fold: (1) to infer general phylogeographic and phylogenetic relationships in

the genus from comprehensive geographical and taxonomic sampling; (2) to delimit species and infer the species tree and timing of lineage diversification within *Amatitlania* using a recently developed Bayesian species delimitation approach (Grummer *et al.*, 2014); and (3) to test for genetic signals of historical demographic fluctuations in *Amatitlania* species in response to the dynamic Pliocene–Pleistocene history of CA. By comparing our genetic results to available morphological data, we identify important implications of our findings for taxonomy, as well as for understanding the evolutionary history of *Amatitlania* and the broader CA freshwater fish assemblage.

## MATERIAL AND METHODS

### SAMPLING

We sampled all nominal species and the entire geographical range of *Amatitlania* throughout CA during expeditions conducted between 1998 and 2012. We identified samples to species based on current taxonomy, including morphological and geographical distributions in Schmitter-Soto (2007a) and McMahan *et al.* (2014). Voucher specimens were deposited at the Louisiana State University Museum of Natural Science nd the Monte L. Bean Life Science Museum Fish Collection. We included mtDNA sequences from five *A. nigrofasciata* individuals collected from Lago Coatepeque by McMahan *et al.* (2014). Our final sample encompassed 318 *Amatitlania* individuals from 94 localities (Fig. 1; see also Supporting information, Data S1). We used two individuals of *Amphilophus citrinellus* (AcitTIPI.1 and AcitTIPI.2) and one individual of *Archocentrus centrarchus* (AcenTIPI.1) that we collected as outgroups. We also included up to 90 published outgroup sequences representing approximately 81 species/lineages of heroine cichlids (Říčan *et al.*, 2013) to obtain calibration points for phylogenetic analyses (see Supporting information, Data S1).

### DNA EXTRACTION, AMPLIFICATION, AND SEQUENCING

We extracted whole genomic DNA from tissue samples using DNeasy Blood & Tissue Kits (Qiagen). We sequenced the entire protein-coding mtDNA cytochrome *b* (cyt*b*) gene using the forward primer Glu18 5′-TAACCAGGACTAATGRCTTGAA-3′ (Unmack *et al.*, 2012) and reverse primer RF.Thr.48 5′-GCA GTAGGAGGGAATTTAACCTTCG-3′ (Unmack & Dowling, 2010). For a subset of individuals, we amplified the ribosomal protein S7 (*RPS7*) introns 1 and 2 ($N = 23$ and $N = 53$, respectively) using primers in

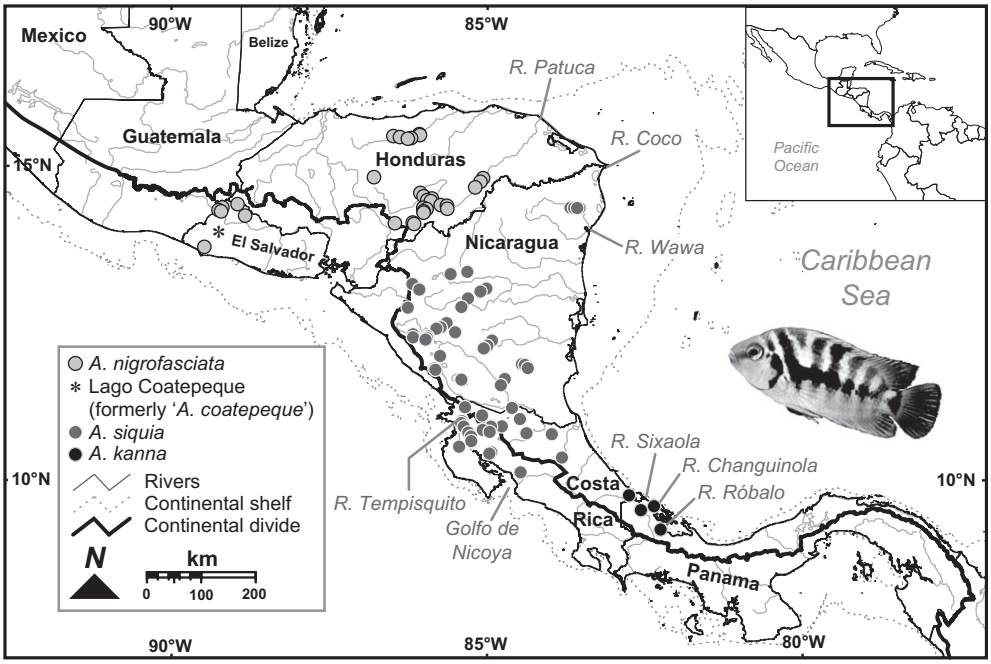


**Figure 1.** Map of Central American sampling localities for all *Amatitlania* samples examined in the present study. Sampling localities (dots) correspond to collections in the Supporting information (Data S1) and are coloured according to three nominal species of *Amatitlania* and one recently synonymized species (McMahan *et al.*, 2014; Schmitter-Soto, 2007a). Geopolitical boundaries (country names shown in bold), as well as the continental divide (thick black line), major river courses (grey lines), and the continental shelf represented by a 135 m bathymetric contour, are shown for context.

Chow & Takeyama (1998) and Unmack *et al.* (2012) via nested polymerase chain reactions (PCR). Final concentrations, cleaning and purification of PCR products, and the PCR reaction protocol were conducted *sensu* Unmack *et al.* (2012), except that we conducted the first 94 °C cycle for 3 min instead of 2 min, and the 72 °C cycle for 90 s instead of 60 s. For the nested PCRs, our first reaction was 10 μL (conditions above) and we subsequently diluted this to 1 : 99 before using 1 μL from the dilution as template for the second reaction. Sequences were obtained via cycle sequencing with Big Dye 3.1 dye terminator chemistry using 1/16th reaction size in accordance with the manufacturer's instructions (Applied Biosystems). We purified sequenced products using Sephadex columns (GE Healthcare) and ran them on an Applied Biosystems 3730*xl* automated sequencer.

### Sequence analysis

We edited sequences using SEQUENCHER, version 4.10.1 (Gene Codes Corporation). Mitochondrial cyt*b* sequences contained no gaps and were aligned by visual inspection in SEQUENCHER. We aligned nuclear sequences in MAFFT, version 6.850 (Katoh & Toh, 2008) using the local pair FFTS algorithm and default settings. GenBank accession numbers are provided for all sequences in the Supporting information (Data S1). *Amatitlania RPS7* intron sequences exhibited very limited genetic variation (< 0.072% overall mean *p*-distance < 0.48% overall mean *p*-distance, with outgroups) rendering them phylogenetically uninformative (see Supporting information, Fig. S1), and so we excluded them from our formal analyses but refer to them when interpreting our results in the Discussion. We collated three mtDNA datasets used in our analyses. First, we created a 'full-cyt*b*' dataset including 318 cyt*b* sequences from *Amatitlania*, plus 90 cyt*b* sequences from outgroup lineages. Second, we used TCS, version 1.21 (Clement, Posada & Crandall, 2000) to collapse identical cyt*b* ingroup sequences into haplotypes and generate statistical parsimony networks of ingroup haplotype clades (95% connection limit; outgroups: three haplotypes of *Cryptoheros panamensis*). We constructed a 192-sequence 'cyt*b* haplotype' dataset consisting of 106 *Amatitlania* cyt*b* haplotypes plus 86 outgroup sequences. Third, we created a 'reduced-cyt*b*' dataset for species delimitation analyses that included 20 ingroup samples (five samples per nominal species, plus five samples from the Lago Coatepeque *A. nigrofasciata* population formerly recognized as '*A. coatepeque*') and 18 outgroup samples included to meet modelling assumptions or obtain calibration points.

### Phylogeny estimation and sequence divergence

We estimated phylogenetic relationships among *Amatitlania* sequences and outgroup sequences in the full-cyt*b* and cyt*b* haplotype datasets using maximum likelihood (ML) and Bayesian inference analyses. We conducted ML searches in GARLI, version 2.0 (Zwickl, 2006) using default parameters, except we partitioned the data by codon position ({1 + 2}, 3) and unlinked parameters across data subsets. We assigned each data subset its 'best-fit' model of molecular evolution (Table 1) selected using the decision-theory algorithm of Minin *et al.* (2003) in JMODELTEST, version 2.1.4 (Darriba *et al.*, 2012). We used 500 ML bootstrap replicates and considered

**Table 1.** Sequence characteristics and 'best-fit' evolutionary models for mitochondrial DNA datasets used in the present study

| | *n* | bp | *S* (%) | Parsimony informative sites (%) | Parsimony uninformative sites (%) | Substitution model |
|---|---|---|---|---|---|---|
| Reduced-cyt*b* dataset | 38 | 1137 | 478 (42.0) | 398 (35.0) | 80 (7.0) | HKY+Γ+*I* |
| 1st + 2nd codon positions | 38 | 758 | 136 (17.9) | 97 (12.8) | 39 (5.1) | HKY+Γ+*I* |
| 3rd codon position | 38 | 379 | 342 (90.2) | 301 (79.4) | 41 (10.8) | GTR+Γ+*I* |
| cyt*b* haplotype dataset | 192 | 1137 | 567 (49.9) | 499 (43.9) | 68 (6.0) | HKY+Γ |
| 1st + 2nd codon positions | 192 | 758 | 195 (25.7) | 135 (17.8) | 60 (7.9) | HKY+Γ+*I* |
| 3rd codon position | 192 | 379 | 373 (98.4) | 365 (96.3) | 8 (2.1) | GTR+*I* |
| Full-cyt*b* dataset | 408 | 1137 | 567 (49.9) | 500 (44.0) | 67 (5.9) | TrN+Γ+*I* |
| 1st + 2nd codon positions | 408 | 758 | 195 (25.7) | 136 (17.9) | 59 (7.8) | HKY+Γ+*I* |
| 3rd codon position | 408 | 379 | 373 (98.4) | 365 (96.3) | 8 (2.1) | TrN+Γ+*I* |

bp, number of nucleotide base pairs; cyt*b*, cytochrome *b* gene; *n*, sample size; *S*, segregating sites, the number of variable sites in the alignment.

nodes with bootstrap proportions ≥70 as well supported (Hillis & Bull, 1993). We estimated Bayesian gene trees from all three datasets in the divergence dating analyses described below. We estimated average genetic distances within and between mtDNA clades and samples grouped by nominal species in MEGA, version 6 (Tamura *et al.*, 2013) based on the proportion of shared cyt*b* differences (*p*-distances).

### Bayesian species delimitation and divergence dating

We tested alternative hypotheses of species limits within *Amatitlania* using Bayes factor delimitation (BFD) (Grummer *et al.*, 2014). This method involves a four-step procedure: (1) individual assignment to species based on previous studies or exploratory analyses; (2) generation of different models specifying alternative groupings of individuals into species (e.g. by lumping or splitting taxa); (3) estimation of a species tree and marginal likelihood score for each model; and (4) Bayesian model selection using Bayes factors (Grummer *et al.*, 2014). The goal is to develop species delimitation models using objective criteria and rank them based on a metric of model evidence without making a priori assumptions about phylogenetic relationships (*sensu* O'Meara, 2010; Yang & Rannala, 2010).

Regarding steps (1) and (2), coalescent-based species delimitation analyses are at times circular in nature, with species delimitation hypotheses derived from phylogenies developed using the same molecular data (in whole or in part) used during hypothesis testing (Leaché & Fujita, 2010; Grummer *et al.*, 2014; Bagley *et al.*, 2015). Yet, previous taxonomic hypotheses have the advantage of providing a priori hypotheses of species limits based on external morphological data alone, hence avoiding such issues of circularity. We used current taxonomy as our null hypothesis for species delimitation. The three currently recognized species of *Amatitlania* were described in qualitative and phylogenetic studies of phenotypic characters (Schmitter-Soto, 2007a; Schmitter-Soto, 2007b) consistent with a morphological species concept (MSC) (Cronquist, 1978). The exception to this is that McMahan *et al.* (2014) synonymized '*A. coatepeque*' with *A. nigrofasciata* by applying a MSC with the additional requirement that each species be monophyletic relative to other species under a phylogenetic species concept (PSC) (Mayden, 1997). Operationally, applying the MSC is problematic because species may be morphologically cryptic and morphological descriptions are often subjective. Applying the PSC is also problematic because it effectively denies the possibility of species-level polyphyly as a result of incomplete lineage sorting (ILS), which affects mitochondrial and nuclear genotypes (Maddison, 1997; Funk & Omland, 2003). Thus, under a PSC, 'true' species with ILS can be mistakenly referred to synonymy. We avoided these issues by using a general lineage concept (GLC) of species as metapopulation lineages (*sensu* Mayden, 1997; de Queiroz, 2007) and conducting BFD using coalescent-based models that explicitly incorporated ILS.

We used BFD to test the null hypothesis against alternative models groupings individuals in the reduced-cyt*b* dataset into 1–4 species of *Amatitlania* recognized in previous studies based largely on a MSC (Bussing, 1998; Schmitter-Soto, 2007a; McMahan *et al.*, 2014). We evaluated five species delimitation models: (1) a 'one-species' model lumping ingroup samples into the type species, *A. nigrofasciata* (*sensu* Bussing, 1998; Allgayer, 2001); (2) a 'two-species allopatric' model consisting of *A. kanna* samples, plus a second, geographically allopatric lineage lumping the remaining samples into *A. nigrofasciata*; (3) a 'two-species disjunct' model consisting of *A. siquia*, plus a second lineage lumping remaining samples into *A. nigrofasciata*, yielding a geographically disjunct distribution of the latter lineage between Honduras and eastern Costa Rica; (4) a 'three-species' null model consisting of the three currently recognized species of *Amatitlania* (*sensu* Schmitter-Soto, 2007a; McMahan *et al.*, 2014); and (5) a 'four-species' model recognizing three nominal taxa plus Lago Coatepeque *A. nigrofasciata* as distinct species.

We ran competing species delimitation models under the multispecies coalescent model implemented in the *BEAST algorithm (Heled & Drummond, 2010) in BEAST, version 1.8.3 (Drummond *et al.*, 2012). This allowed us to simultaneously estimate the gene tree, species tree, and times to the most recent common ancestor ($t_{MRCA}$) for the samples. To ensure convergence, we ran three replicate searches on the reduced-cyt*b* dataset in BEAST [Markov chain Monte Carlo (MCMC) = $2 \times 10^7$, sampling every 4000 generations] for each model. Runs linked tree and clock models but partitioned the data into codon position subsets ({1 + 2}, 3) and unlinked site parameters across subsets. Site models for different subsets were set to best-fit models listed in Table 1. Analyses drew branch rates from an uncorrelated lognormal molecular clock (default settings), with rate variation following a birth-death tree prior. We calibrated each run with three fossil and biogeographical calibration points similar to those used by Chakrabarty (2006) and Říčan *et al.* (2013): (1) a normal distribution constraining the diversification of heroine cichlids (tribe Heroini), including outgroup and ingroup samples, to the minimum age of the

fossil †*Plesioheros chauliodus*, 39.9–48.6 Mya (mean in real space = 44.25, σ = 2.22); (2) a normal distribution for the basal split between three species of *Nandopsis* (*Nandopsis ramsdeni* + *Nandopsis tetracanthus* from Cuba vs. *Nandopsis haitiensis* from Hispaniola) correlated with the geological separation of the islands of Cuba and Hispaniola approximately 14–17 Mya (mean in real space = 15.5, σ = 0.764); and (3) a normal distribution constraining the separation of the Orinoco and Magdalena drainage basins to 10.2–11.8 Mya, which we applied to two samples of *Caquetaia* (mean in real space = 10.95, σ = 0.434). We summarized posterior distributions and ensured convergence and adequate effective sample sizes (ESS >> 200) in TRACER, version 1.6 (Rambaut *et al.*, 2013). We calculated a maximum clade credibility (MCC) tree annotated with median node ages from 5000 post-burn-in trees in TREEANNOTATOR, version 1.8.3. We archived our sequence alignments, input files, and gene tree and species tree results in Dryad (doi:10.5061/dryad.r1d8q).

We used two recently developed methods to estimate the log-marginal likelihoods of each model: path sampling (PS) and stepping-stone (SS) sampling (Xie *et al.*, 2011; Baele *et al.*, 2012). Simulation studies show that PS and SS marginal likelihood estimates are more accurate for Bayesian phylogenetic, demographic, and species delimitation models compared to other estimators (Baele *et al.*, 2012; Grummer *et al.*, 2014). After each *BEAST run, we ran PS and SS analyses for 100 steps each one million generations in length ($10^8$ total generations). We used a beta distribution of ~B(0.3, 1) to space out the path steps (Xie *et al.*, 2011). We calculated $2\log_e(B_{10})$ Bayes factors from the log-marginal likelihood scores and evaluated 'weight of evidence' of the models *sensu* Kass & Raftery (1995) using pairwise model comparisons. We relied on the *BEAST species tree from the best-supported model identified by BFD as our best estimate of the phylogeny and divergence times of *Amatitlania*. For comparative purposes, we also estimated time-calibrated Bayesian gene trees for the full-cyt*b* and cyt*b* haplotype datasets in BEAST using three replicate runs employing birth-death tree priors and the best-fit evolutionary models listed in Table 1; other priors were identical to those specified during *BEAST runs above.

### Historical demographical modelling

We inferred historical population dynamics of each species of *Amatitlania* identified during BFD (see Results) by running Bayesian skyline plot (BSP) models (Drummond *et al.*, 2005) in BEAST on corresponding samples in the full-cyt*b* dataset. We conducted three replicate runs (MCMC = $2 \times 10^8$, sampling every 4000 generations) with different starting seeds on each dataset. Each run partitioned the data into codon position subsets ({1 + 2}, 3), unlinked parameters across subsets, and employed a strict molecular clock. Site models were set to best-fit models selected in JMODELTEST (see Supporting information, Table S1) and the molecular clock was set to the 'standard' fish cyt*b* evolutionary rate of 1% Myr pairwise divergence (Bermingham, McCafferty & Martin, 1997), which may also be a reasonable approximation of mtDNA evolutionary rates in CA cichlids (Martin & Bermingham, 1998; Pérez *et al.*, 2007). We ensured model convergence and calculated posterior distributions of parameters in TRACER.

We cross-validated our BSP inferences by testing for genetic signals of past population growth, against a null hypothesis of a neutrally evolving population of constant size, using complementary neutrality tests. Specifically, we estimated Ramos-Onsins & Rozas' (2002) $R_2$ and Tajima's (1989) $D$ in DNASP (Librado & Rozas, 2009). We assessed significance and determined 95% confidence intervals for $R_2$ using coalescent simulations ($10^4$ replicates). We looked for agreement across BSP and neutrality test results as providing strong evidence for past population dynamics.

## RESULTS

### Phylogeny estimation and sequence divergence

Our final mtDNA datasets contained 1137 bp sequences and were highly informative (Table 1). The ML analysis of the cyt*b* haplotype dataset obtained a single 'best' gene tree with a log likelihood score (ln $L$) of −21170.5496. All *Amatitlania* haplotypes were resolved in a single monophyletic group, 'clade 1', with relatively shallow relationships and little clear geographical or taxonomic structuring (Fig. 2). Maximum pairwise mtDNA sequence divergence within *Amatitlania* based on $p$-distances was 2.6%, although we observed only 0.7–1.6% and 0.5–1.0% average sequence divergence among vs. within nominal taxa, respectively, and a mere 0.65% average sequence divergence within clade 1.

Rather than resolving the three morpho-species of *Amatitlania* as distinct monophyletic groups consistent with current taxonomy, we resolved them as polyphyletic with respect to one another (Fig. 2). The well-supported subclade '1-a' (86% bootstrap proportion) comprised *A. kanna* samples and one *Cryptoheros myrnae* sample (Amyrna208) from Pérez *et al.* (2007). We interpreted this *C. myrnae* sample as a potential case of interspecific hybridization (i.e. introgressed *Amatitlania* mtDNA within a *C. myrnae*

individual), or field misidentifications of samples by earlier workers, and we excluded it from subsequent analyses. Average sequence divergence within subclade 1-a (excluding Amyrna208) was the highest of any group, at 1.0%, and this *A. kanna* subclade was also up to 1.6% different on average from other samples of *Amatitlania*. We also resolved *A. nigrofasciata* and *A. siquia* as polyphyletic with respect to one another, being mixed throughout the tree.

Three other strongly statistically supported subclades in the ML tree (Fig. 2) were geographically cohesive, mostly taxonomically homogeneous (except clade '1-c'), and allopatrically distributed (Fig. 1). First, we obtained a small subclade '1-b' (94% bootstrap) of three haplotypes of *A. siquia* from the Río Tempisquito, a tributary of Golfo de Nicoya on the Pacific northwestern coast of Costa Rica. Second, subclade '1-c' (91% bootstrap) comprised a group of

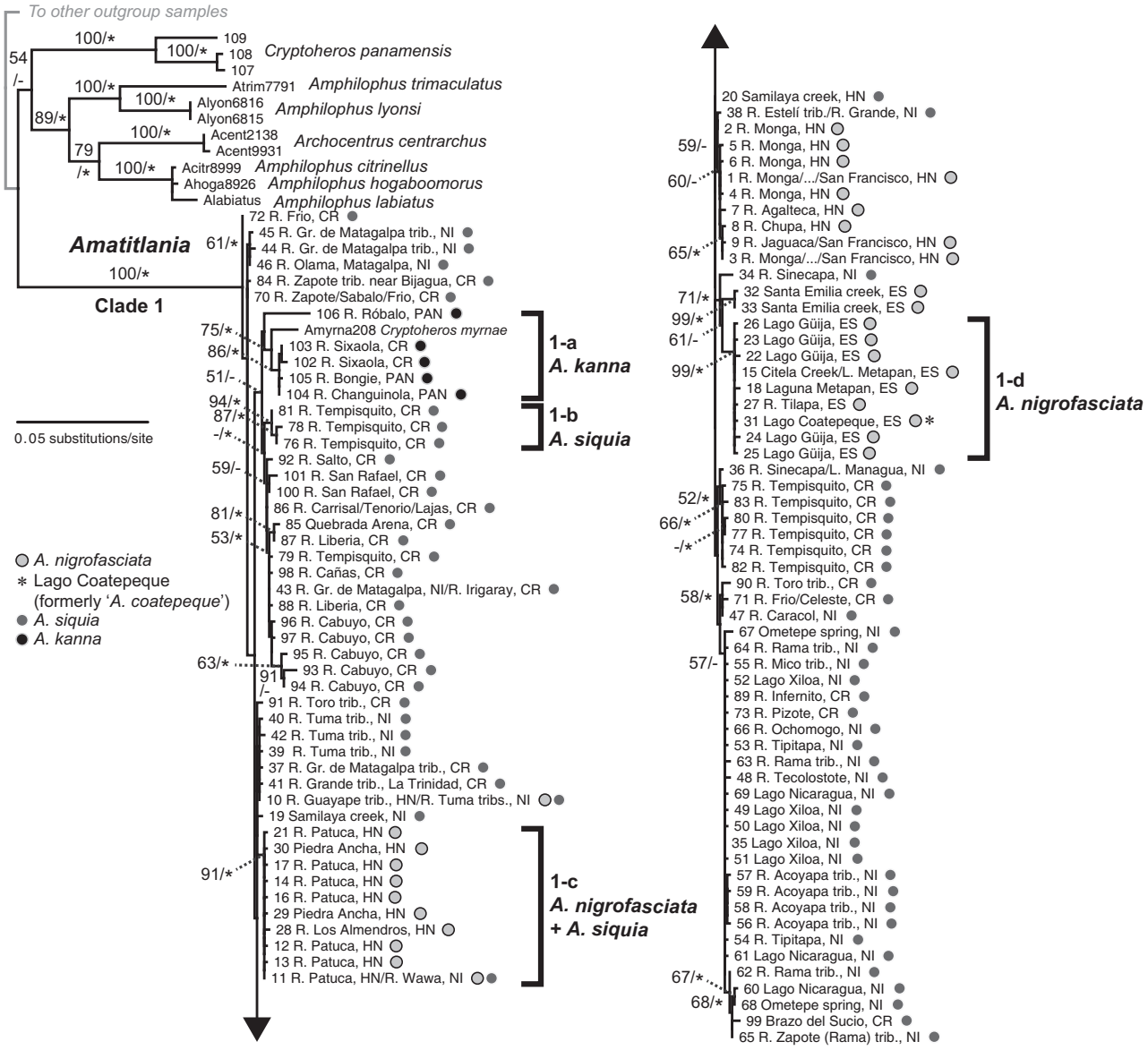


**Figure 2.** Maximum-likelihood (ML) tree generated from a GARLI (Zwickl, 2006) analysis of 106 haplotypes of *Amatitlania* and 86 outgroup sequences in the mtDNA cytochrome *b* (cyt*b*) haplotype dataset. Values at the left of nodes are given as ML bootstrap proportions (≥50%)/Bayesian posterior probabilities (with values ≥0.95 indicated by an asterisk). Each sample code is based on the haplotype number, followed by the locality name in the Supporting information (Data S1). Shaded circles to the right of the phylogeny show the nominal taxa represented by each sample, and major subclades 1-a to 1-d are indicated. CR, Costa Rica; ES, El Salvador; HN, Honduras; NI, Nicaragua.

10 haplotypes of *A. nigrofasciata* from the Río Patuca, Honduras, except that haplotype 11 was also shared by samples of *A. siquia* from three sites along the Río Wawa in north-eastern Nicaragua (Fig. 1; see also Supporting information, Data S1). Lastly, subclade '1-d' (99% bootstrap) comprised nine haplotypes of *A. nigrofasciata* from western El Salvador. Haplotypes within subclades 1-b, 1-c, and 1-d were very shallowly diverged from one another, with average within-group sequence divergence of < 0.5%.

The ML analysis of the full-cyt*b* dataset obtained a single best topology (ln $L = -21050.6617$); however, we do not present this topology because it is essentially identical to the cyt*b* haplotype ML topology. The best trees from our ML analysis were also very similar to Bayesian gene tree topologies estimated in BEAST in the full-cyt*b* (mean model ln $L = -21387.89$, ESS = 1193.88; Fig. S2) and cyt*b* haplotype (mean model ln $L = -21396.34$, ESS = 7665.78; Fig. S3) analyses, and qualitatively similar to the Bayesian tree estimated for the reduced-cyt*b* dataset by the best-supported *BEAST model identified during BFD (see Supporting information, Fig. S4). For example, all three BEAST topologies lent definitive support for *Amatitlania* clade 1 [Bayesian posterior probability (BPP) = 100%].

Our TCS parsimony networks yielded a pattern of relationships among cyt*b* haplotypes similar to those of the ML and Bayesian gene trees but gave a more detailed perspective of phylogenetic structure (Fig. 3). For example, *A. kanna* haplotypes in ML subclade 1-a were highly differentiated genetically in the network, being isolated from one another by up to 23 nucleotide differences, and differentiated from other *Amatitlania* haplotypes by up to 31 differences; one sample, haplotype 106 from the Río Róbalo, Panama, was even resolved as a singleton network. The parsimony network also revealed that haplotypes in subclades 1-a, 1-b, and 1-c were mostly diverged by 10–13 nucleotide differences. Contrasting these aspects of phylogenetic structuring, we also observed various star-like patterns in clade 1, with several tip haplotypes radiating from the inferred ancestral haplotype 35, as well as haplotypes 1, 10, 11, 15, and 43.

### Bayesian species delimitation and divergence dating

In our BFD results, PS and SS analyses yielded essentially identical log-marginal likelihood estimates supporting the same relative ranking of species delimitation models (Table 2). Bayes factors calculated from PS results [$2\log_e(B_{10}) = 19.795$] and SS results [$2\log_e(B_{10}) = 20.435$] lent decisive support in favour of model 2, our two-species allopatric model recognizing *A. nigrofasciata* and *A. kanna* lineages as distinct, over all other models. Model 4, a three-species model recognizing nominal *Amatitlania* species as distinct, was consistently the second best model but was not widely supported by Bayes factors. Model 3, a two-species disjunct model recognizing only *A. nigrofasciata* and *A. siquia* as distinct species, was consistently the least supported model (Table 2). The relaxed clock species tree for the preferred species delimitation model inferred using *BEAST (mean model ln $L = -7451.98$, ESS = 5363.79; see Supporting information, Fig. S4) dated the two BFD-inferred species of *Amatitlania* (BPP = 100%) as diverging at a mean time of 1.345 Mya [95% highest posterior density (HPD) = 2.378–0.504 Mya], corresponding to the early to mid-Pleistocene. Overall, the mean mtDNA evolutionary rate inferred for the cyt*b* gene was 0.0122 substitutions site$^{-1}$ Myr$^{-1}$. Although we take the *BEAST result as our best estimate of the date of lineage divergence within *Amatitlania*, the time-calibrated cyt*b* gene trees from BEAST yielded divergence dates that overlapped the *BEAST estimate but were predictably slightly older and more variable (gene divergence times are necessarily older than species tree divergence times) (Heled & Drummond, 2010). Ingroup $t_{MRCA}$ values from the full-cyt*b* and cyt*b* haplotype analyses in BEAST inferred that extant *Amatitlania* alleles coalesced to mean ages of 2.109 Mya (HPDs = 1.192–3.008 Mya) and 2.445 Mya (HPD = 1.559–3.504 Mya), respectively, in the late Pliocene to early Pleistocene. In both of the time-calibrated gene trees, *Amatitlania* subclades 1-a to 1-d were shallowly diverged from mid-late Pleistocene most recent common ancestors. The time-calibrated gene trees also resolved *Amatitlania* as having diverged from its sister clade (*Amphilophus* + *Archocentrus* + *Cryptoheros* cichlids; BPP = 93.4%) in the mid-Miocene.

### Historical demographical modelling

Bayesian demographic modelling strongly supported a history of population growth over the mid-late Pleistocene within the *A. nigrofasciata* lineage ($N = 313$ individuals) identified during BFD (Fig. 4A). The BSP reconstruction (ln $L = -2965.98$, ESS = 525.83) showed an almost exponential growth pattern for this lineage ever since approximately 200 000 years ago, yielding an approximately eight-fold increase in $N_e$ (effective population size). Population expansion in the *A. nigrofasciata* lineage was also supported by star-like haplotype connections in the parsimony network (Fig. 3), as well as positive $R_2$ (mean = 0.076, $P < 0.001$) and negative Tajima's $D$ (mean = −1.930, $P < 0.05$) statistics. The BSP

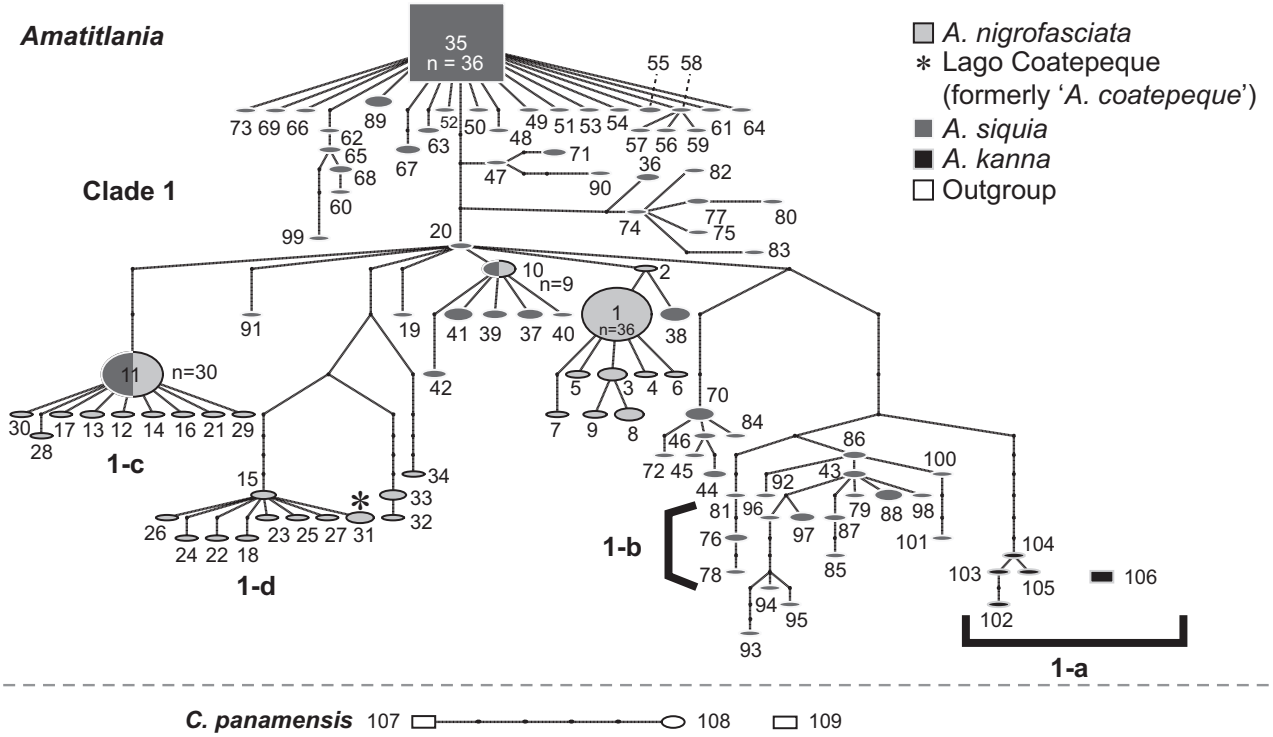


**Figure 3.** Statistical parsimony networks of putative *Amatitlania* and outgroup (*Cryptoheros panamensis*) cytochrome *b* (cyt*b*) haplotypes from Central America. Circles and rectangles represent haplotypes scaled according to their observed abundances and are colour-coded as in Figs 1, 2. Haplotype numbers are given beside each haplotype, and dots along branches between haplotypes represent single-step mutations, or unsampled haplotypes. Networks are separated based on a 95% parsimony criterion, and black bars indicate major maximum likelihood subclades 1-a to 1-d.

reconstruction for the *A. kanna* lineage ($N = 5$ individuals from clade 1-a) supported by BFD displayed a trend of population growth from approximately 700 000–350 000 years ago**.** during the mid-Pleistocene, predating that of the *A. nigrofasciata* lineage (Fig. 4B). However, *A. kanna* population expansion was not supported by the networks (no star-like connections), Tajima's $D$ (mean = −1.013, $P > 0.05$) or the $R_2$ statistic (mean $R_2 = 0.335$, $P > 0.05$).

## DISCUSSION

### Conflicting mitochondrial evidence and morpho-species boundaries in *Amatitlania*

Our results from phylogenetic analyses of convict cichlids based on comprehensive geographical and taxonomic sampling and data from a mitochondrial genome fragment are markedly at odds with the current morpho-species taxonomy of the group (*sensu* Schmitter-Soto, 2007a; McMahan *et al.*, 2014). Maximum likelihood, Bayesian, and parsimony phylogenetic analyses of mtDNA consistently obtained a single, relatively shallowly (maximum 2.6%) diverged clade of *Amatitlania*, in which all three currently recognized species in the genus were polyphyletic with respect to one another (Figs 2 and 3; Table 3; see also Supporting information, Fig. S1 and S2). The nuclear *RPS7* introns that we examined agreed with these mtDNA findings, exhibiting low genetic variability and yielding a single ingroup clade; indeed, the low number of variable sites rendered this nuclear locus insufficiently informative to phylogenetically differentiate any morpho-species of *Amatitlania* (see Supporting information, Fig. S1). This is unsurprising given the diminished mutation rates and four-times slower rates of lineage sorting of nuclear markers relative to mtDNA (Maddison, 1997; Funk & Omland, 2003), yet *RPS7* introns have previously proven useful for resolving phylogenetic relationships and delimiting species of Central American cichlid and poeciliid freshwater fishes (Chakrabarty, 2006; Říčan *et al.*, 2013; Bagley *et al.*, 2015). The only pattern of congruence between morpho-species and genetic data among our gene tree results was that we resolved *A. kanna* samples in a mtDNA subclade 1-a, which was genetically distinct, with the highest degree of sequence divergence from other

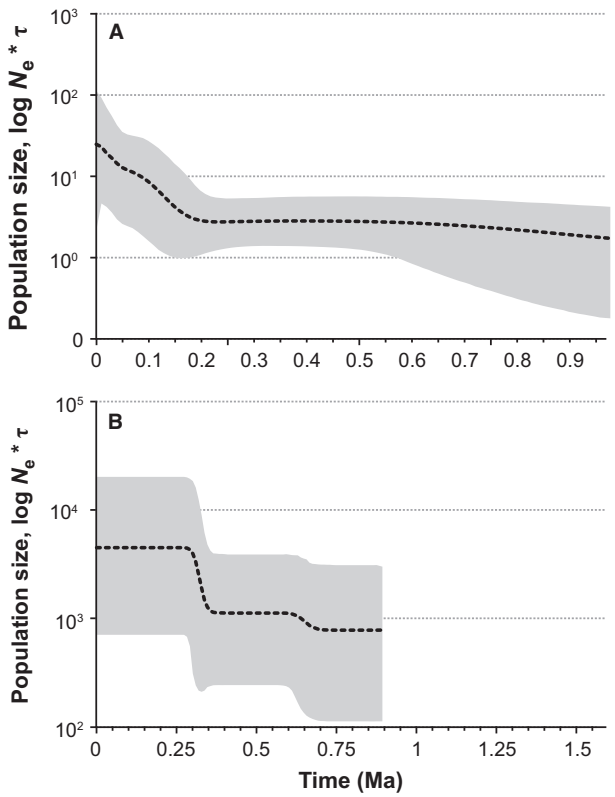


**Figure 4.** Results of Bayesian demographic models showing reconstructions of the demographic history of the two *Amatitlania* species inferred during our Bayes factor delimitation species delimitation analyses, including (A) the *A. nigrofasciata* lineage and (B) the *A. kanna* lineage. Bayesian skyline plot (Drummond *et al.*, 2005) reconstructions are presented with *x*-axes in units of time (Mya). Thick centrelines represent the mean values of $\log_{10}$ population sizes ($N_e * \tau$) and grey shading with thin black outlines indicates the upper and lower limits of the 95% highest posterior density of the estimates.

ingroup samples (Table 3). Subclade 1-a was strongly supported across phylogenetic gene tree analyses with different underlying assumptions and algorithms. Nevertheless, subclade 1-a fell within a broader clade of *A. nigrofasciata* and *A. siquia* samples, rendering *A. kanna* phylogenetically nested within other taxa. Overall, these results indicate that morpho-species of *Amatitlania* exhibit substantial species-level polyphyly; yet questions remain concerning the processes that are most likely responsible for these results.

Previous studies have identified several competing explanations for widespread species-level polyphyly, including mtDNA introgression, incomplete lineage sorting (Maddison, 1997; Maddison & Knowles, 2006) or imperfect taxonomy, all of which are known to impact phylogenetic inference and species delimitation (Funk & Omland, 2003; Fujita *et al.*, 2012). Distinguishing between introgression and ILS is difficult given that these processes produce similar phylogenetic signals, and rigorous tests of hybridization-mediated introgression require a well-resolved species tree and multilocus data (Joly, McLenachan & Lockhart, 2009; Kubatko, 2009). Although our data are insufficient for implementing such methods, the present data are also inconsistent with introgression as a potential cause of polyphyly in *Amatitlania*. Interspecific mtDNA hybridization events are inferred when one or more haplotypes fall out in geographically sympatric populations of distinct species that are divergent at nuclear loci or morphological characters (Funk & Omland, 2003). Against this expectation, a 98.1% majority of *Amatitlania* haplotypes were unique to a single nominal taxon, and only haplotypes 10 and 11 were shared between geographically allopatric populations identified in the field as *A. nigrofasciata* and *A. siquia* (Figs 2 and 3; see also Supporting information, Data S1). Also, there was no clear nuclear DNA variation among previously accepted, morphologically differentiated species of *Amatitlania*. We thus tentatively down-weight the hybridization hypothesis until the alternative hypothesis of ILS is rejected in favour of gene flow and hybridization-mediated introgression in *Amatitlania* based on rigorous modelling analyses of multiple unlinked nuclear loci.

Our BFD results, which employed Bayes factor comparisons of species tree models and accounted for topological uncertainty and ILS in a coalescent framework (Grummer *et al.*, 2014), decisively supported two lineages as distinct species within *Amatitlania* (two-species allopatric model, Table 2; see also Supporting information, Fig. S4). These two lineages correspond to *A. nigrofasciata*, the type species of the genus, and *A. kanna*, and they are diverged from one another by up to 2.6% pairwise mtDNA genetic distance (Table 3). The fact that *A. kanna* was statistically supported as a distinct species by BFD provides compelling evidence that ILS is the best explanation for the polyphyly of *A. kanna* in our gene tree analyses, which did not account for ILS. However, *A. siquia* was collapsed within the *A. nigrofasciata* lineage in the best-fit *BEAST model, and all models proposing three or four species, thus recognizing *A. siquia* as a distinct species, were strongly selected against in BFD (Table 2). Thus, we conclude that the polyphyly of *A. siquia* must be the result of factors other than ILS, most likely imperfect taxonomy. Our model lumping the three nominal species of *Amatitlania* into one species was also strongly selected against in BFD. Overall, the cumulative model probabilities of nonselected models (calculated from marginal likelihoods) amounted to only 0.0026; thus, the probability of the best-fit model was

**Table 2.** Posterior comparisons of evidence for different Bayesian species delimitation models for *Amatitlania*

| Model | Log-marginal likelihood | (1) | **(2)** | (3) | (4) | (5) | Model probability |
|---|---|---|---|---|---|---|---|
| (1) One-species | −7612.511 | – | **−20.435** | −2.457 | −12.413 | −8.519 | $5.016 \times 10^{-5}$ |
| **(2) Two-species allopatric** | **−7602.614** | **19.795** | **–** | **17.979** | **8.022** | **11.917** | **0.997** |
| (3) Two-species disjunct | −7611.516 | 1.992 | **−21.334** | – | −9.957 | −6.062 | $1.358 \times 10^{-4}$ |
| (4) Three-species | −7606.298 | 12.426 | **−13.437** | 7.897 | – | 3.895 | 0.025 |
| (5) Four-species | −7608.239 | 8.545 | **−14.418** | 6.916 | −0.9806 | – | 0.0036 |

Results include log marginal likelihoods calculated using path sampling (PS), as well as row-by-column $2\log_e(B_{10})$ Bayes factors [below diagonal, calculated from PS results; above diagonal, from stepping-stone (SS) results] for five models based on previous and current taxonomy of *Amatitlania*, as described in the text. Larger log marginal likelihoods, and larger and positive Bayes factors, indicate relatively greater model evidence. Boldface corresponds to the single best-supported model. Model probabilities are calculated based on differences between model log marginal likelihoods.

388 times higher than that of all four other models combined. In addition to support from BFD and genetic distances, the evolutionary distinctiveness of the two inferred species within *Amatitlania* is also supported by diagnostic alleles for each lineage, and the fact that no alleles were shared between the lineages because all haplotypes (haplotypes 102–106) of *A. kanna* were private alleles (Fig. 3). By contrast, as noted above, two haplotypes were shared between *A. nigrofasciata* and *A. siquia*.

### Phylogeography of *Amatitlania*

The two species of *Amatitlania* have experienced a geologically recent history of evolutionary independence. Relaxed-clock age estimates for the *Amatitlania* clade from the preferred *BEAST species tree model selected during BFD suggest that the two *Amatitlania* species most likely diverged approximately 1.3 Mya in the early Pleistocene, with Bayesian credible intervals spanning early-mid Pleistocene (see Supporting information, Fig. S4). These dates correspond to a period when global temperature and sea level were similar to their modern analogues (Zachos *et al.*, 2001; Miller *et al.*, 2005) and they are not correlated with any major geological events that could have caused the vicariant speciation of these lineages (Bagley & Johnson, 2014a). Also, no obvious geographical barriers are present today that could have promoted long-term, post-divergence isolation of *A. kanna* populations between the Río Sixaola, Costa Rica, and the Río Changuinola of northwestern Panama, from the *A. nigrofasciata* lineage. Nevertheless, *A. kanna* inhabits relatively short coastal rivers that drain a narrow Caribbean coastal plain with a very narrow continental shelf less than approximately 15–30 km in width. Thus, opportunities for historical interdrainage connections over the continental shelf to nearby river drainage networks have likely been limited, enhancing the historical isolation of *A. kanna* in these drainages (Unmack *et al.*, 2012; Unmack *et al.*, 2013).

Comparing the phylogeographic break witnessed in *Amatitlania* with patterns of genetic subdivision from other Central American taxa reveals that the Caribbean versant of Costa Rica presents an important area of regional intraspecific diversification. In particular, our finding that this region is phylogenetically diverged from outlying areas to the north is concordant with mtDNA phylogeography patterns in several codistributed taxa, including three frog species sharing phylogeographic breaks in the Limón region, Costa Rica [*Eleutherodactylus* (Crawford, 2003); *Oophaga pumilio* (Hagemann & Pröhl, 2007), and *Dendropsophus ebraccatus* (Robertson, Duryea & Zamudio, 2009)]; one freshwater fish lineage [*Bryconamericus* (Reeves & Bermingham, 2006)] in which populations diverge at the Sixaola–Changuinola drainage divide (Bagley & Johnson, 2014a); and three lineages of freshwater fishes showing local population differentiation at Bocas del Toro [*Rhamdia* (Perdices *et al.*, 2002); *Brachyhypopomus occidentalis* (Picq *et al.*, 2014); and *Poecilia* (Bagley *et al.*, 2015)]. These broadly congruent phylogeographic patterns coinciding with the deepest divergence between *Amatitlania* lineages imply that amphibian and freshwater fish species potentially responded in similar fashion to historical events in this region of the Caribbean coast. This hypothesis should be examined in more detail using tests for temporal phylogeographic congruence (Arbogast & Kenagy, 2001; Bagley & Johnson, 2014a; Bagley & Johnson, 2014b) that are improved by adding more data from codistributed species.

We also infer that the two species of *Amatitlania* may have experienced incongruent historical–demographic fluctuations over Pleistocene to recent timeframes. Within the *A. nigrofasciata* lineage, congruent results from BSP models and neutrality tests strongly supported a history of population expansion

**Table 3.** Mean pairwise genetic distances between nominal species of *Amatitlania*, as well as between clades based on phylogenetic results in Figs 2, 3

| | *Amatitlania nigrofasciata* | Lago Coatepeque population | *Amatitlania siquia* | *Amatitlania kanna* |
|---|---|---|---|---|
| *Amatitlania nigrofasciata* | 0.005 | 0.0026 | 0.0013 | 0.0028 |
| Lago Coatepeque population (formerly '*Amatitlania coatepeque*') | 0.009 | 0.000 | 0.0031 | 0.0037 |
| *Amatitlania siquia* | 0.007 | 0.013 | 0.006 | 0.0029 |
| *Amatitlania kanna* (subclade 1-a) | 0.016 | 0.021 | 0.016 | 0.010 |

Mean proportion of base pair differences per site (*p*-distances) within each group are shown along the diagonal; mean among-group *p*-distances are shown below the diagonal; and SEs calculated from 500 bootstrap pseudoreplicates are shown above the diagonal. Results are based on the full-cyt*b* sequence database. Grey cells indicate pairwise comparisons that were not conducted.

ever since approximately 200 kya in the late Pleistocene (Fig. 4A). The period of inferred expansion follows the mid-Pleistocene transition (0.8–1.2 Mya), when climatic extremes became enhanced and glaciations shifted from cycles of 41 000 years to cycles of 100 000 years (Clark *et al.*, 2006) and overlaps the last two glaciation cycles. By contrast, historical demographic tests do not strongly support population expansion in the *A. kanna* lineage. Although the BSP analysis inferred an *A. kanna* population expansion during the mid-Pleistocene (approximately Marine Isotope Stages 7–9), we failed to reject the null hypothesis of size-constancy based on $R_2$ and $D$ neutrality statistics. We cannot have as much confidence in the *A. kanna* BSP reconstruction as we have in that for the *A. nigrofasciata* lineage as a result of the small number of individuals and prevalence of singleton alleles sampled from the *A. kanna* lineage. Thus, sample size should be augmented to more confidently test the hypothesis of population expansion in this lineage in the future. Recent Bayesian demographic analyses of CA poeciliid fishes whose distributions overlap that of *Amatitlania* inferred bottleneck expansion events ever since approximately 40 kya during the last glacial cycle in the molly, *Poecilia gillii*, but rejected Pleistocene population expansions in other cases (e.g. *Alfaro cultratus*, *Xenophallus umbratilis*) (Bagley & Johnson, 2014b). Thus, combined with data from other CA phylogeography studies, our results suggest that Pleistocene climate change in Central America has affected different freshwater fish species to varying degrees over different timescales, resulting in genetic signatures of asynchronous population expansion events.

### Taxonomic recommendations and limitations

We follow the GLC, taking species to be metapopulation lineages evolving independently of other such aggregates (de Queiroz, 2007), and consider genealogical and statistical evidence from genetic loci adequate to diagnose distinct species. Thus, we used an objective coalescent-based species delimitation method (Fujita *et al.*, 2012) based on the GLC (Grummer *et al.*, 2014). Overall, our BFD results statistically reject the current status of *A. siquia* as a distinct evolutionary species, at the same time as also upholding the evolutionary distinctiveness of *A. nigrofasciata* and *A. kanna*. Previous morphological studies of the group supported all three currently recognized species of *Amatitlania* as diagnosable based on morphology (Schmitter-Soto, 2007a; McMahan *et al.*, 2014); however, our results suggest that only *A. kanna* is clearly diagnosable based on genetic and morphological characters. By contrast, *A. nigrofasciata* and *A. siquia* are each diagnosable in only one of these dimensions. One explanation for these findings is that the current taxonomic arrangement of *Amatitlania*, as reported by Schmitter-Soto (2007a), correctly diagnosed *A. kanna* but misdiagnosed allopatric morphological variation within the original type species of the genus, *A. nigrofasciata*, leading to 'oversplitting' of *A. nigrofasciata*. Nevertheless, we refrain from advocating formal taxonomic changes, given that our results are based largely on the mtDNA locus, and thus are subject to several potential limitations. Mitochondrial DNA has been shown to be a robust and rapidly evolving indicator of population history (Zink & Barrowclough, 2008), although evolutionary inferences from mtDNA may reflect the matrilineal history of the locus, rather than being representative of the broader evolutionary history of the species reflected by loci sampled from throughout the genome. As a result, additional nuclear markers might uncover a conflicting view of the evolution and species limits of *Amatitlania* (e.g. if our mtDNA sequences have been subject to the influence of undetected genetic processes such as sex-biased dispersal, selection, hybridization or hidden third-codon substitutions) (Ballard & Whitlock, 2004; Irwin, 2012). In

particular, it is not possible to discern to what extent our species tree reflects these confounding processes, although *BEAST species trees should yield more accurate branch lengths and divergence times than gene trees even when based on a single locus (Drummond *et al.*, 2012). Given these caveats, we recommend an analysis combining our results with independent molecular, morphological, and ecological data in an 'integrative taxonomy' framework (Dayrat, 2005; Fujita *et al.*, 2012) to further test the distinctiveness of *A. siquia* (i.e. putative diagnostic characters) in greater detail and provide a robust taxonomic revision of the group.

## ACKNOWLEDGEMENTS

We thank María Florencia Breitman, the editor, and two anonymous reviewers for constructive comments that improved earlier drafts of the manuscript. We are also grateful to Eduardo Castro Nallar, Joseph T. Nelson, Aaron H. Smith, and Eric P. van den Berghe for assistance with the fieldwork, and to María Florencia Breitman for insightful discussions on taxonomy. Honduran fieldwork was conducted under Instituto Nacional de Conservación y Desarrollo Forestal, Áreas Protegidas, y Vida Silvestre permit DVS-ICF-033-2009; Guatemalan fieldwork was conducted under Consejo Nacional de Áreas Protegidas permit 00088 from 2013; fieldwork in El Salvador was conducted under permit 027-2011 from the Ministerio de Medio Ambiente y Recursos Naturales; Nicaraguan sampling was conducted under Ministerio del Ambiente y Recursos Naturales permits DGPN/DB/DAP-IC-0008-2010 and DGPN/DB-IC-009-2012; and Costa Rican sampling was conducted under Ministerio de Ambiente Energía y Telecomunicaciones permits 030-2010-SINAC and 134-2012-SINAC. We thank J. Guevara Siquiera and Edilberto Duarte for help with obtaining the permits. Specimens were obtained through collections made under Brigham Young University (BYU) Institutional Animal Care and Use Committee (IACUC) approval #12-0701 and by Louisiana State University IACUC Protocol #09-002. This research was supported by a BYU Graduate Research Fellowship award to JCB, a BYU Mentoring Environment grant and US National Science Foundation (NSF) PIRE grant (OISE PIRE-0530267) to JBJ, an NSF Doctoral Dissertation Improvement Grant (DEB-1210883) to JBJ and JCB, a NSF grant DEB-1311408 to CDM, NSF grants DEB-0916695 and DEB-1354149 to PC, and a Guy D. Jordan Cichlid Research Fund award from the American Cichlid Association to MT.

## REFERENCES

**Agapow PM, Bininda-Emonds ORP, Crandall KA, Gittleman JL, Mace GM, Marshall JC, Purvis A. 2004.** The impact of species concept on biodiversity studies. *Quarterly Review of Biology* **79:** 161–179.

**Albert JS, Reis RE. 2011.** Introduction to Neotropical freshwaters. In: Albert JS, Reis RE, eds. *Historical biogeography of Neotropical freshwater fishes*. Berkeley, CA: University of California Press, 3–20.

**Allgayer R. 1994.** Description d'une espèce nouvelle du genre Archocentrus. *Revue Française des Cichlidophiles* **135:** 6–24.

**Allgayer R. 2001.** Description d'un genre nouveau, *Cryptoheros*, d'Amérique Centrale et d'une espèce nouvelle du Panama (Pisces: Cichlidae). *L'An Cichlidé* **1:** 13–20.

**Arbogast BS, Kenagy GJ. 2001.** Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography* **28:** 819–825.

**Avise JC. 2000.** *Phylogeography: the history and formation of species*. Cambridge, MA: Harvard University Press.

**Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012.** Improving the accuracy of demographic and molecular clock model comparison while accomodating phylogenetic uncertainty. *Molecular Biology and Evolution* **29:** 2157–2167.

**Bagley JC, Johnson JB. 2014a.** Phylogeography and biogeography of the lower Central American Neotropics: diversification between two continents and between two seas. *Biological Reviews* **89:** 767–790.

**Bagley JC, Johnson JB. 2014b.** Testing for shared biogeographic history in the lower Central American freshwater fish assemblage using comparative phylogeography: concerted, independent, or multiple evolutionary responses? *Ecology and Evolution* **4:** 1686–1705.

**Bagley JC, Mayden RL, Roe KJ, Holznagel W, Harris PM. 2011.** Congeneric phylogeography reveals polyphyly and novel biodiversity within black basses (Centrarchidae: *Micropterus*). *Biological Journal of the Linnean Society* **104:** 346–363.

**Bagley JC, Alda F, Breitman MF, Bermingham E, van den Berghe EP, Johnson JB. 2015.** Assessing species boundaries using multilocus species delimitation in a morphologically conserved group of Neotropical freshwater fishes, the *Poecilia sphenops* species complex (Poeciliidae). *PLoS ONE* **10:** e0121139.

**Bagley JC, Matamoros WA, Mcmahan CD, Tobler M, Chakrabarty P, Johnson JB. 2016.** Data from: Phylogeography and species delimitation in convict cichlids (*Cichlidae: Amatitlania*): implications for taxonomy and Plio–Pleistocene evolutionary history in Central America. *Dryad Digital Repository*. doi:10.5061/dryad.r1d8q.

**Ballard JWO, Whitlock MC. 2004.** The incomplete natural history of mitochondria. *Molecular Ecology* **13:** 729–744.

**Bermingham E, Martin AP. 1998.** Comparative mtDNA phylogeography of Neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology* **7:** 499–517.

**Bermingham E, McCafferty S, Martin A. 1997.** Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. In: Kocher T, Stepien C, ed. *Molecular systematics of fishes*. New York, NY: Academic Press, 113–126.

**Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007.** Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* **22:** 148–155.

**Bussing WA. 1976.** Geographic distribution of the San Juan ichthyofauna of Central America with remarks on its origin and ecology. In: Thorson TB, ed. *Investigations of the ichthyofauna of Nicaraguan lakes*. Lincoln, NE: University of Nebraska, 157–175.

**Bussing WA. 1998.** *Freshwater fishes of Costa Rica*, 2nd edn. San José, Costa Rica: Editorial de la Universidad de Costa Rica.

**Chakrabarty P. 2006.** Systematics and historical biogeography of Greater Antillean Cichlidae. *Molecular Phylogenetics and Evolution* **39:** 619–627.

**Chow S, Takeyama H. 1998.** Intron length variation observed in the creatine kinase and ribosomal protein genes of the swordfish *Xiphias gladius*. *Fisheries Science* **64:** 397–402.

**Clark PU, Archer D, Pollard D, Blum JD, Rial JA, Brovkin V, Mix AC, Pisias N, Roy M. 2006.** The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric $pCO_2$. *Quaternary Science Reviews* **25:** 3150–3184.

**Clement M, Posada D, Crandall KA. 2000.** TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657–1659.

**Crawford AJ. 2003.** Huge populations and old species of Costa Rican and Panamanian dirt frogs inferred from mitochondrial and nuclear gene sequences. *Molecular Ecology* **12:** 2525–2540.

**Cronquist A. 1978.** Once again, what is a species? In: Knutson LV, ed. *Biosystematics in agriculture*. Montclair, NJ: Allenheld Osmun, 3–20.

**Darriba D, Taboada GL, Doallo R, Posada D. 2012.** jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9:** 772.

**Dayrat B. 2005.** Towards integrative taxonomy. *Biological Journal of the Linnean Society* **85:** 407–415.

**Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005.** Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22:** 1185–1192.

**Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012.** Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29:** 1969–1973.

**Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C. 2012.** Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution* **27:** 480–488.

**Funk DJ, Omland KE. 2003.** Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics* **34:** 397–423.

**Grummer JA, Bryson RW Jr, Reeder TW. 2014.** Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology* **63:** 119–133.

**Günther A. 1867.** Memoir on the fishes of the States of Central America, founded upon specimens collected in the fresh and marine waters of various parts of that country by Messrs. Salvin and Godman and Capt. J.M. Dow. *Proceedings of the Zoological Society of London* **1866:** 600–609.

**Hagemann S, Pröhl H. 2007.** Mitochondrial paraphyly in a polymorphic poison frog species (Dendrobatidae; *D. pumilio*). *Molecular Phylogenetics and Evolution* **45:** 740–747.

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

**Hillis DM, Bull JJ. 1993.** An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42:** 182–192.

**Irwin DE. 2012.** Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. *American Naturalist* **180:** 35–49.

**International Union for Conservation of Nature and Natural Resources (IUCN). 2016.** *The IUCN Red List of Threatened Species version 2016-1*. Available at: http://www.iucnredlist.org (accessed on 26 April 2016).

**Joly S, McLenachan PA, Lockhart PJ. 2009.** A statistical approach for distinguishing hybridization and incomplete lineage sorting. *American Naturalist* **174:** e54–e70.

**Jordan DS, Evermann BW. 1898.** The fishes of North and Middle America: a descriptive catalogue of the species of fishes and fish-like vertebrates found in the waters of North America, north of the Isthmus of Panama. *United States National Museum Bulletin* **47:** 955–1936.

**Kass RE, Raftery AE. 1995.** Bayes factors. *Journal of the American Statistical Association* **90:** 773–795.

**Katoh K, Toh H. 2008.** Recent developments in the MAFFT multiple sequence alignment program. *Bioinformatics* **9:** 286–298.

**Kubatko LS. 2009.** Identifying hybridization events in the presence of coalescence via model selection. *Systematic Biology* **58:** 478–488.

**Kullander SO. 2003.** Family Cichlidae (Cichlids). In: Reis RE, Kullander SO, Ferraris CJ Jr, eds. *Check list of the freshwater fishes of South and Central America*. Porto Alegre: EDIPUCRS, 605–654.

**Leaché AD, Fujita MK. 2010.** Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London B* **277:** 3071–3077.

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

**Maddison WP. 1997.** Gene trees in species trees. *Systematic Biology* **46:** 523–536.

**Maddison WP, Knowles LL. 2006.** Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **55:** 21–30.

**Martin AP, Bermingham E. 1998.** Systematics and evolution of lower Central American cichlids inferred from

analysis of cytochrome *b* gene sequences. *Molecular Phylogenetics and Evolution* **9:** 192–203.

**Mayden RL. 1997.** A hierarchy of species concepts: the denouement in the saga of the species problem. In: Claridge MF, Dawah HA, Wilson MR, eds. *Species: the units of biodiversity*. London: Chapman & Hall, 381–424.

**McMahan CD, Matamoros WA, Barraza E, Kutz J, Chakrabarty P. 2014.** Taxonomic status of the Lago Coatepeque endemic convict cichlid *Amatitlania coatepeque* (Teleostei: Cichlidae). *Copeia* **2014:** 633–638.

**Miller RR. 1966.** Geographical distribution of Central American freshwater fishes. *Copeia* **1966:** 773–802.

**Miller KG, Kominz MA, Browning JV, Wright JD, Mountain GS, Katz ME, Sugarman PJ, Cramer BS, Christie-Blick N, Pekar SF. 2005.** The Phanerozoic record of global sea-level change. *Science* **310:** 1293–1298.

**Minin V, Abdo Z, Joyce P, Sullivan J. 2003.** Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology* **52:** 674–683.

**Moritz C, Faith DP. 1998.** Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology* **7:** 419–429.

**Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J. 2000.** Biodiversity hotspots for conservation priorities. *Nature* **403:** 853–858.

**O'Meara BC. 2010.** New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology* **59:** 59–73.

**Perdices A, Bermingham E, Montilla A, Doadrio I. 2002.** Evolutionary history of the genus *Rhamdia* (Teleostei: Pimelodidae) in Central America. *Molecular Phylogenetics and Evolution* **25:** 172–189.

**Pérez GAC, Říčan O, Ortí G, Bermingham E, Doadrio I, Zardoya R. 2007.** Phylogeny and biogeography of 91 species of heroine cichlids (Teleostei: Cichlidae) based on sequences of the cytochrome *b* gene. *Molecular Phylogenetics and Evolution* **43:** 91–110.

**Picq S, Alda F, Krahe R, Bermingham E. 2014.** Miocene and Pliocene colonization of the Central American Isthmus by the weakly electric fish *Brachyhypopomus occidentalis* (Hypopomidae, Gymnotiformes). *Journal of Biogeography* **41:** 1520–1532.

**Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP. 2006.** Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55:** 595–609.

**de Queiroz K. 2007.** Species concepts and species delimitation. *Systematic Biology* **56:** 879–886.

**Rambaut A, Suchard MA, Xie W, Drummond AJ. 2013.** Tracer, Version 1.6, Available at: http://tree.bio.ed.ac.uk/software/tracer

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

**Reeves RG, Bermingham E. 2006.** Colonization, population expansion, and lineage turnover: phylogeography of Mesoamerican characiform fish. *Biological Journal of the Linnean Society* **88:** 235–255.

**Reis RE, Kullander SO, Ferraris CJ Jr. 2003.** *Check list of the freshwater fishes of South and Central America*. Porto Alegre: EDIPUCRS.

**Říčan O, Piálek L, Zardoya R, Doadrio I, Zrzavý J. 2013.** Biogeography of the Mesoamerican Cichlidae (Teleostei: Heroini): colonization through the GAARlandia land bridge and early diversification. *Journal of Biogeography* **40:** 579–593.

**Robertson JM, Duryea MC, Zamudio KR. 2009.** Discordant patterns of evolutionary differentiation in two Neotropical treefrogs. *Molecular Ecology* **18:** 1375–1395.

**Rull V. 2011.** Neotropical biodiversity: timing and potential drivers. *Trends in Ecology and Evolution* **26:** 508–513.

**Schmitter-Soto JJ. 2007a.** A systematic revision of the genus *Archocentrus* (Perciformes: Cichlidae), with the description of two new genera and six new species. *Zootaxa* **1603:** 1–78.

**Schmitter-Soto JJ. 2007b.** Phylogeny of species formerly assigned to the genus *Archocentrus* (Perciformes: Cichlidae). *Zootaxa* **1618:** 1–50.

**Sites JW Jr, Marshall JC. 2003.** Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology and Evolution* **18:** 462–470.

**Tajima F. 1989.** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585–595.

**Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013.** MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30:** 2725–2729.

**Unmack PJ, Dowling TL. 2010.** Biogeography of the genus *Craterocephalus* (Teleostei: Atherinidae) in Australia. *Molecular Phylogenetics and Evolution* **55:** 968–984.

**Unmack PJ, Bagley JC, Adams MD, Hammer MD, Johnson JB. 2012.** Molecular phylogeny and phylogeography of the Australian freshwater fish genus *Galaxiella* (Teleostei: Galaxiidae), with an emphasis on dwarf galaxias (*G. pusilla*). *PLoS ONE* **7:** e38433.

**Unmack PJ, Hammer MP, Adams M, Johnson JB, Dowling TE. 2013.** The role of continental shelf width in determining freshwater phylogeographic patterns in south-eastern Australian pygmy perches (Teleostei: Percichthyidae). *Molecular Ecology* **22:** 1683–1699.

**Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. 2011.** Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* **60:** 150–160.

**Yang Z, Rannala B. 2010.** Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 9264–9269.

**Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001.** Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292:** 686–693.

**Zink RM. 2002.** Methods in comparative phylogeography, and their application to studying evolution in the North

American aridlands. *Integrative and Comparative Biology* **42:** 953–959.

**Zink RM, Barrowclough GF. 2008.** Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17:** 2107–2121.

**Zwickl DJ. 2006.** *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. Austin, TX: The University of Texas, Austin.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

**Figure S1.** Gene trees derived from ML analyses of nuclear *RPS7* data in GARLI, using evolutionary models selected for each intron using JMODELTEST. Bootstrap proportions ≥50% based on 500 Neighbour-joining pseudoreplicates are given to the left of nodes.
**Figure S2.** BEAST maximum clade credibility tree derived from the full-cyt*b* dataset.
**Figure S3.** BEAST maximum clade credibility tree derived from the cyt*b* haplotype dataset.
**Figure S4.** *BEAST species tree derived from the best-supported species delimitation model identified by Bayes factor delimitation.
**Table S1.** Best-fit evolutionary models for mtDNA datasets used during Bayesian demographic modelling analyses in the present study.
**Data S1.** Taxon list, locality (subpopulation) details, and GenBank accession numbers.

## SHARED DATA

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.r1d8q (Bagley *et al.*, 2016).