

RESEARCH ARTICLE

Open Access

# Paleoclimatic modeling and phylogeography of least killifish, *Heterandria formosa*: insights into Pleistocene expansion-contraction dynamics and evolutionary history of North American Coastal Plain freshwater biota

Justin C Bagley<sup>1\*</sup>, Michael Sandel<sup>3</sup>, Joseph Travis<sup>4</sup>, María de Lourdes Lozano-Vilano<sup>5</sup> and Jerald B Johnson<sup>1,2</sup>

## Abstract

**Background:** Climatic and sea-level fluctuations throughout the last Pleistocene glacial cycle (~130-0 ka) profoundly influenced present-day distributions and genetic diversity of Northern Hemisphere biotas by forcing range contractions in many species during the glacial advance and allowing expansion following glacial retreat ('expansion-contraction' model). Evidence for such range dynamics and refugia in the unglaciated Gulf-Atlantic Coastal Plain stems largely from terrestrial species, and aquatic species Pleistocene responses remain relatively uninvestigated. *Heterandria formosa*, a wide-ranging regional endemic, presents an ideal system to test the expansion-contraction model within this biota. By integrating ecological niche modeling and phylogeography, we infer the Pleistocene history of this livebearing fish (Poeciliidae) and test for several predicted distributional and genetic effects of the last glaciation.

**Results:** Paleoclimatic models predicted range contraction to a single southwest Florida peninsula refugium during the Last Glacial Maximum, followed by northward expansion. We inferred spatial-population subdivision into four groups that reflect genetic barriers outside this refuge. Several other features of the genetic data were consistent with predictions derived from an expansion-contraction model: limited intraspecific divergence (e.g. mean mtDNA p-distance = 0.66%); a pattern of mtDNA diversity (mean  $Hd$  = 0.934; mean  $\pi$  = 0.007) consistent with rapid, recent population expansion; a lack of mtDNA isolation-by-distance; and clinal variation in allozyme diversity with higher diversity at lower latitudes near the predicted refugium. Statistical tests of mismatch distributions and coalescent simulations of the gene tree lent greater support to a scenario of post-glacial expansion and diversification from a single refugium than to any other model examined (e.g. multiple-refugia scenarios).

**Conclusions:** Congruent results from diverse data indicate *H. formosa* fits the classic Pleistocene expansion-contraction model, even as the genetic data suggest additional ecological influences on population structure. While evidence for Plio-Pleistocene Gulf Coast vicariance is well described for many freshwater species presently codistributed with *H. formosa*, this species demography and diversification departs notably from this pattern. Species-specific expansion-contraction dynamics may therefore have figured more prominently in shaping Coastal Plain evolutionary history than previously thought. Our findings bolster growing appreciation for the complexity of phylogeographical structuring within North America's southern refugia, including responses of Coastal Plain freshwater biota to Pleistocene climatic fluctuations.

\* Correspondence: justin.bagley@byu.edu

<sup>1</sup>Department of Biology, Brigham Young University, 401 WIDB, Provo, UT 84602, USA

Full list of author information is available at the end of the article

## Background

Present-day distributions of many of Earth's biotas reflect the profound influence of climatic and sea-level fluctuations during the glacial-interglacial cycles of the Pleistocene (2.58-0.01 million years ago, Ma; [1]) [2-4]. Notably, eight glacial advances since 740 ka have produced more extreme 100,000-year environmental changes than those preceding 1 Ma [5]. Because glacial-interglacial transitions occurred in an evolutionary blink of an eye, many extant species have spent most of their recent evolutionary histories under glacial rather than short-lived (~10,000-20,000 years) interglacial conditions [6]. Therefore, extreme environmental changes associated with the Last Glacial Maximum (LGM; ~22-19 ka) may have been the most recent climatic factors that affected present-day patterns of biodiversity prior to the global proliferation of humans and anthropogenic environmental disturbance [2,6].

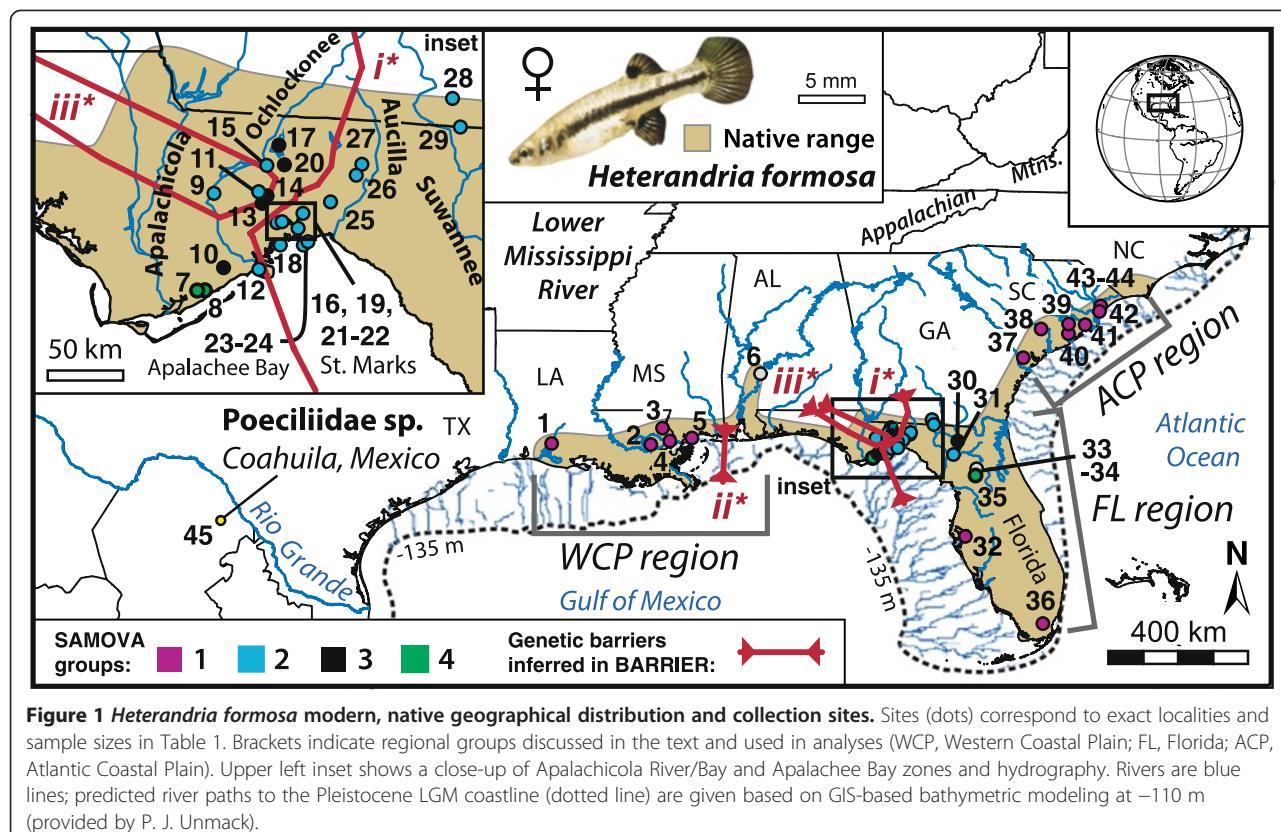
The biogeographical consequences of Pleistocene glaciations are generally thought of in terms of direct effects of glacial cover and temperature change on Northern Hemisphere biotas. This is because over 80% of glacial ice on Earth resided in the Northern Hemisphere during the LGM, resulting in the destruction and reorganization of continental landscapes in both the Old and New World e.g. [2,7]. While few species from glaciated areas persisted *in situ* by adapting to changing local environments, many were actively displaced (tracking suitable habitat), passively displaced (local extinction, regional extirpation), or went extinct [3,8]. Indeed, abundant paleontological and palynological evidence indicates that many terrestrial species of North America and Europe contracted their ranges during the LGM to one or more lower-latitude (or altitude) refugia, then expanded northward to recolonize much of their present-day ranges during the warmer Holocene interglacial [9-14]. This scenario is known as the 'expansion-contraction' model of Pleistocene biogeography and has reached paradigm status in biology [15].

Many patterns of genetic variation are consistent with predictions from a model of range contraction during glacial advance and rapid post-glacial expansion. For example, phylogeographical analyses across European flora and fauna have identified the Iberian, Italian and Balkan peninsulas as three main southern refugia from which postglacial recolonization of northern Europe took place [9,16,17]. In more recent work, the tools of phylogeography and ecological niche modeling e.g. [18] have been joined to refine models of Pleistocene expansion and contraction. Such analyses have revealed complicated patterns of northern and southern LGM refugia and complex histories of speciation, dispersal, secondary contact, and demographic expansion of species within the southern refugia themselves, particularly in Europe [15,18-24].

There is substantial geoscientific evidence that the glacial-interglacial cycles also had substantial effects in

subtropical and tropical zones, including refugial areas recognized today that were not covered by glaciers [7,19-21]. During the Last Interglaciation (LIG; 130-116 ka; including 'Sangamonian' stage, ~125 ka), wetter, warmer climate and sea levels ~4-6 m above present sea level (ASL) [25,26] created favorable conditions for expansions in many species ranges. Subsequent global cooling (~4-10°C cooler surface air temperatures [27]) and aridification during the LGM caused eustatic sea levels to drop ~120-130 m lower than today [28], exposing huge areas of continental shelf, areas large enough to have doubled the span of the Florida peninsula to 600 km wide. For freshwater species, this increase in LGM land area also increased river lengths (Figure 1), which likely increased population connectivity through river anastomoses. Such mixing of populations from formerly isolated drainages would have promoted dispersal and gene flow, and may have aided southward range shifts. Thus while glacial stages are perceived as periods of reduced population sizes and isolation in refugia [9], for some species, particularly those in fresh waters, they may have presented opportunities for coastal dispersal, leading to range expansion or shifting (contraction). The possibility for southern refugia to act as areas of long-term persistence, dispersal, and gene flow, rather than range expansion-contraction dynamics, is known as the 'southern crossroads' hypothesis [21].

The Gulf-Atlantic Coastal Plain of eastern North America presents an ideal system in which to study interactions between geology, climate and diversification of a unique biota. This unglaciated region has figured prominently in phylogeographic research, and has revealed several major genetic breaks and confirmed local glacial refugia for northern taxa and endemic coastal species [29,30]. Many Coastal Plain species possess evolutionary lineages that have diverged in an east-to-west direction across common biogeographic barriers, including peninsular Florida which separates the maritime Atlantic-vs.-Gulf coast biota, the Apalachicola River (Gulf-Atlantic slope drainages), Mobile Bay, and the Mississippi River [30-32]. Genetically identified breaks in this region also coincide with species distributional breaks [33] and thus likely reflect historical vicariance events, or recurrent vicariance coupled with extinction-recolonization of populations [29,31,34]. However, phylogeographical studies also indicate that species persisted through the LGM in refugia on either side of the Apalachicola basin in Florida and Texas-western Louisiana [29,31]. Between these regions, there are north-south-trending zones of hybridization and secondary contact, including one of Remington's 'suture zones' between Alabama and northwest Florida, indicating a meeting place for lineages undergoing northward postglacial expansion [30,35]. As is the case for the patterns in evidence for expansion-contraction dynamics in unglaciated regions of Europe [2,4,9], evidence for these refugia in North America



**Figure 1** *Heterandria formosa* modern, native geographical distribution and collection sites. Sites (dots) correspond to exact localities and sample sizes in Table 1. Brackets indicate regional groups discussed in the text and used in analyses (WCP, Western Coastal Plain; FL, Florida; ACP, Atlantic Coastal Plain). Upper left inset shows a close-up of Apalachicola River/Bay and Apalachee Bay zones and hydrography. Rivers are blue lines; predicted river paths to the Pleistocene LGM coastline (dotted line) are given based on GIS-based bathymetric modeling at -110 m (provided by P. J. Unmack).

has been developed mainly from terrestrial taxa, e.g. mints (e.g. *Conradina* [36,37]) and yellow-poplar (*Liriodendron tulipifera* [38]). These terrestrial data suggest that other codistributed species may have experienced similar range dynamics fitting the expansion-contraction model, possibly involving similar refugia. However, while regional genetic breaks are well documented, Pleistocene evolutionary responses of Coastal Plain aquatic biota remain relatively uninvestigated.

Here, we focus on least killifish, *Heterandria formosa* Girard 1859, a species which presents several advantages as a historical biogeographic study system that likely exhibited distinct response(s) to the climatic upheavals of the Late Pleistocene. First, these small (12–30 mm) livebearing freshwater fish (family Poeciliidae) are restricted to low-elevation Coastal Plain areas of subtropical humid climate zones (hot, wet summers and mild winters), and the tropical Everglades [39]. This suggests climate is likely an important factor limiting *H. formosa* distribution. This species is also intolerant to cold-seasonal temperatures, making it an ideal candidate for ecological niche modeling and for testing the expansion-contraction model. Second, *H. formosa* span the Florida peninsula, an important glacial-stage refugium [29,35]; thus, it is plausible that populations were displaced to warmer south-peninsula areas during the LGM. Third, *H. formosa* display ecological

characteristics consistent with the potential for rapid population expansion over evolutionary timescales, including (i) rapid reproduction and short generation time ( $T \approx 0.33$  yr [40]); (ii) a range of tolerances to different abiotic conditions [39,40], and (iii) female capacity for livebearing and sperm storage, suggesting individuals or groups of migrating females could found populations during rapid post-LGM expansions [41]. Fourth, *H. formosa* is endemic to the Gulf-Atlantic Coastal Plain and thus also offer the opportunity to test the generality of the standard regional vicariance model and phylogeographical breaks documented in other taxa.

A previous rangewide study by Baer [42] using allozymes inferred genetic barriers in *H. formosa* between the Western Coastal Plain (WCP) and Atlantic Coastal Plain (ACP) regions (Figure 1), around the Suwannee River, but not between north Florida and the ACP. Another break occurred between Gulf-draining Waccasassa and Withlacoochee Rivers, creating a clade of Louisiana plus south Florida samples. Baer [42] also hypothesized an important role for population expansion and cross-peninsula gene flow in influencing current levels of genetic variation—including recent founding of ACP populations after Early Pleistocene high seas ( $\leq 1.5$  Ma; ‘northeast colonization’ hypothesis), supported by lower genetic diversity and a lack of isolation-by-distance only among

ACP populations. Yet while Baer's [42] study presents a broad outline for understanding *H. formosa* historical biogeography, the limited levels of genetic variation he found and the more modest methods available at the time limited the hypotheses that could be tested. The current analytical framework in historical biogeography, including our ability to combine ecological niche modeling (e.g. to infer Pleistocene-Holocene distributions and generate spatially explicit hypotheses) with coalescent simulations (e.g. used to discriminate among historical scenarios [43,44]), provides a more rigorous framework for biogeographical inference in this species [45-47].

In this study, we integrate ecological niche modeling with paleoclimatic data, phylogeography, and coalescent simulations to infer the historical biogeography of *H. formosa*. Specifically, we use *H. formosa* to test the hypothesis that members of the Coastal Plain freshwater biota fit predictions of the expansion-contraction model (as opposed to southern crossroads, or regional vicariance and northeast colonization hypotheses) by testing for predicted patterns of Late Pleistocene range shifts, refugia and recolonization patterns and their spatial-genetic effects. In addition to analyzing new mtDNA and nDNA sequences, we re-analyze available allozyme data [42], which provide a novel temporal spectrum and improve our tests of genetic diversity predictions, including latitudinal patterns not previously investigated. We show that these new data, combined with paleoclimatic modeling in an integrative approach, support a hypothesis of Pleistocene range expansion-contraction dynamics in least killifish.

## Methods

### Sampling and laboratory methods

Rather than exhaustively sampling every *H. formosa* population rangewide, we conducted higher-density sampling near the middle of the range in northern Florida to get at finer-scale patterns of diversity and gene flow near the area where Baer [42] predicted the main genetic breaks to occur. We also sampled broadly to infer phylogeographical patterns and test rangewide models of population history. We collected 224 *H. formosa* individuals from 42 coastal sites across the species range (1-12 individuals/site; Figure 1; Table 1), which we assumed represented discrete populations. Specimens were collected using standard seining and electrofishing techniques, preserved in 95% ethanol, transported to the laboratory, and maintained at room temperature before DNA extraction. Studies of population mutation rate parameter  $\theta$  demonstrate that  $N \approx 8$  samples are typically sufficient for characterizing genetic diversity, though estimates are improved by recombination and more loci [48,49]. Thus, wherever possible during our analyses, we used samples that met this  $N \geq 8$  threshold-sampling criterion, or pooled samples across populations until this threshold was met. We included three samples of a novel

taxon (hereafter, 'Poeciliidae sp.') collected from Coahuila, Mexico in our analyses. Based on morphological variation we have uncovered while describing this recently discovered taxon, we hypothesize it is the lineage sister to *H. formosa* (MLLV, JCB, and JBJ, unpublished data). We tested this hypothesis through outgroup analysis by adding 72 sequences from 23 other poeciliid lineages retrieved from GenBank into our phylogenetic alignments [*Belonesox belizanus*; *Gambusia affinis*; *Limia dominicensis*, *L. melanogaster*, *L. tridens*, and *L. vittata*; *Pamphorichthys hollandi*; *Pseudoxiphophorus* (15 lineages); and *Xiphophorus helleri*; Additional file 1: Table S1]. This yielded a total of 24 possible outgroup lineages.

We isolated DNA using the Qiagen DNeasy96 tissue protocol (Qiagen Sciences, Maryland, USA). We amplified the mitochondrial cytochrome *b* (*cytb*) gene for each sample by PCR using Hrbek et al.'s [50] L14725 – H15982 primer pair. We also sequenced nuclear ribosomal protein *S7* (*RPS7*) introns 1 and 2 and exon 2 for subsamples of 1-3 *H. formosa* individuals/site for up to 4 random sites per region within each clade (based on our multilocus gene tree; see Results), and Poeciliidae sp. samples. We amplified *RPS7* using a nested PCR design, with primers 1 F – 3R.24 in the first reaction followed by internal primers 1 F.2 – 2R.67 and 2 F.2 – 3R in subsequent reactions [51]. Primers 1 F and 3R are from Chow and Takeyama [52]. Amplification conditions, sequencing reactions, clean-up, and sequence visualization followed Unmack et al. [51]. We aligned sequences manually while viewing electropherograms in Sequencher™ 4.10 (Gene Codes Corp.). All sequences generated in this study were deposited in GenBank (accession numbers: KF632895-KF633114, *cytb*, and KF633115-KF633133, *RPS7*). Alignment lengths were 1140 nucleotide base pairs (bp) for *cytb* and 876 bp for *RPS7*. We supplemented these data with orthologous ingroup sequences from Lake Pontchartrain, Louisiana ( $N = 1$ ) and Everglades, Florida ( $N = 1$ ) retrieved from GenBank (Additional file 1: Table S1). Thus, our sampling encompassed *H. formosa* from 44 georeferenced sites (Table 1). In total, the final *cytb* database we analyzed consisted of 223 *H. formosa* and 3 Poeciliidae sp. sequences that we collapsed into unique haplotypes in DnaSP 5.10 [53]. To create a multilocus alignment, we collated individuals with *cytb* and *RPS7* sequences and joined them with analogous sequences from potential outgroups. Except multilocus and haplotype-based tests, analyses used the full *cytb* database (sometimes divided among groups); iteratively adding/removing samples with missing data (usually <5 bp) had no qualitative effect on results.

We also obtained data for 11 nuclear protein-encoding allozyme loci described in [42] (excluding locus *Idh-2* due to missing data for several populations), representing 948 *H. formosa* from 34 sites across the range (a1-a34, Figure 2;  $N = 12-36$  individuals/site) that we georeferenced in decimal

**Table 1** *Heterandria formosa* collection sites, sample sizes and genetic diversity measures

Population (specimen no. code)	No.	Locality	Latitude (° N)	Longitude (° W)	N	Cytb haps.	Study region
<b>Bayou Nezpique (BNExLA)</b>	1	Bayou Nezpique, south of I-10, ~30 km east of Lake Charles, Cameron Par., LA	30.245	-92.627	2	22	WCP
<b>Lake Pontchartrain (PONxLA)</b>	2	Lake Pontchartrain, LA	30.189	-90.101	1	31	WCP
<b>Bogue Chitto (BOGxLA)</b>	3	Bogue Chitto R. overflow pool along LA Hwy 21 south of Sun, St. Tammany Par., LA	30.630	-89.897	2	30	WCP
<b>Porter's River (PORxLA)</b>	4	Porter's R. at Davis Landing Rd., St. Tammany Par., LA	30.318	-89.728	2	40, 41	WCP
<b>Wolf Creek Swamp (WCMSx)</b>	5	Wolf Cr. swamp at Belle Ferry Rd., Harrison Co., MS	30.387	-89.197	2	42	WCP
<b>Alabama River trib. (ALAxAL)</b>	6	Unnamed trib. to Alabama R. (Bailey Cr. origin), Wilcox Co., AL	31.981	-87.504	2	-	-
<b>Five Points (FIVxFL)</b>	7	Unnamed trib. at Five Points, Tate's Hell State Forest, Franklin Co., FL	29.893	-84.760	9	17, 29	FL
<b>Crooked River (CROxFL)</b>	8	Crooked R. at Tate's Hell State Forest, Franklin Co., FL	29.893	-84.731	10	17	FL
<b>Ochlockonee River (OCHxFL)</b>	9	Ochlockonee R. at Blountstown Hwy, Leon Co., FL	30.388	-84.665	2	1, 37	FL
<b>Womack Creek (WOMxFL)</b>	10	Womack Cr. at Apalachicola National Forest, Franklin Co., FL	30.009	-84.608	9	1, 3, 11-14	FL
<b>Moore Lake (MOOxFL)</b>	11	Moore Lake at Apalachicola National Forest, Leon Co., FL	30.392	-84.407	10	24	FL
<b>Hill Swale (HILxFL)</b>	12	Hill Swale, just north of Ochlockonee Bay, Wakulla Co., FL	29.100	-84.406	10	25, 26	FL
<b>Trout Pond (TROxFL)</b>	13	Trout Pond at Apalachicola National Forest, Leon Co., FL	30.335	-84.385	10	3	FL
<b>Cessna Pond (CESxFL)</b>	14	Cessna Pond at Apalachicola National Forest, Leon Co., FL	30.374	-84.360	9	3, 28	FL
<b>Little Lake (LITxFL)</b>	15	Little Lake Jackson, west of Hwy 27, northwest Tallahassee, Leon Co., FL	30.526	-84.359	6	2	FL
<b>Wakulla Springs (WAKxFL)</b>	16	Wakulla Springs at Edward Ball Wakulla Springs State Park, Wakulla Co., FL	30.235	-84.302	10	7, 9-10, 20	FL
<b>Lake Iamonia (IAMxFL)</b>	17	Lake Iamonia at CR 155, Leon Co., FL	30.627	-84.291	10	1	FL
<b>Shepherd Spring (SHExFL)</b>	18	Shepherd Spring at St. Marks National Wildlife Refuge, Wakulla Co., FL	30.123	-84.281	10	9-10, 27	FL
<b>McBride Slough (MCBxFL)</b>	19	McBride Slough at Bloxham Cutoff Rd, Wakulla Co., FL	30.239	-84.270	12	9-10, 43-44	FL
<b>Lake Overstreet (LOVxFL)</b>	20	Lake Overstreet at Alfred B. Maclay Gardens, Leon Co., FL	30.528	-84.257	10	1, 23	FL
<b>Newport Sulphur Spring (NEWxFL)</b>	21	Newport Sulphur Spring, Wakulla Co., FL	30.207	-84.179	10	2, 9, 15, 18	FL
<b>Natural Bridge (NATxFL)</b>	22	Natural Bridge at Natural Bridge Rd, Wakulla Co., FL	30.284	-84.152	9	2, 18-19	FL
<b>Gumbo Bayou (GAMxFL)</b>	23	Gumbo Bayou at Lighthouse Rd. west of Stony Bayou Pool, Wakulla Co., FL	30.122	-84.148	9	9, 15	FL
<b>Tram Road (TRAxFL)</b>	24	Tram Rd., 2 km east of Plum Orchard, Wakulla Co., FL	30.140	-84.121	9	9, 15-16	FL
<b>Wacissa River (WACxFL)</b>	25	Wacissa R., just north of Aucilla Wildlife Management Area, Jefferson Co., FL	30.341	-83.992	10	5-8	FL
<b>Buggs Creek borrow pit (BUGxFL)</b>	26	Buggs Creek borrow pit, FL	30.477	-83.844	2	35-36	FL
<b>Wolf Creek (WOLxFL)</b>	27	Wolf Creek at US 90, Jefferson Co., FL	30.532	-83.812	1	39	FL
<b>McKey Park (MCKxGA)</b>	28	McKey Park, Loundes Co., GA	30.864	-83.290	2	45	FL
<b>Bevil Creek (BEVxGA)</b>	29	Bevil Creek at Browns pond Loch Laurel Rd., GA	30.720	-83.244	2	3, 38	FL
<b>Robinson Creek (ROBxFL)</b>	30	Robinson Cr. at CR 131, Columbia Co., FL	30.321	-82.662	3	22	FL
<b>Ichetucknee blue hole (ICHxFL)</b>	31	Blue hole, Ichetucknee State Park, Columbia Co., FL	29.980	-82.758	2	32-33	FL
<b>Hillsborough River (HIRxFL)</b>	32	Hillsborough R. at Tampa, Hillsborough Co., FL	27.957	-82.465	11	22, 47	FL
<b>River Styx (RISxFL)</b>	33	River Styx at CR 346, Alachua Co., FL	29.517	-82.222	2	-	FL
<b>Saint Johns River (SJR2x)</b>	34	Unnamed trib. of the St. Johns R., FL	-	-	3	22	FL

**Table 1** *Heterandria formosa* collection sites, sample sizes and genetic diversity measures (Continued)

Newnan's Lake (NEWnxFL)	35	Newnan's Lake at southeast 16th Ave. boat ramp, Alachua Co., FL	29.637	-82.200	1	3	FL
Everglades (EVExFL)	36	Everglades drainage, Dade Co., FL	25.552	-80.564	1	30	FL
Bahama Swamp (BAHxSC)	37	Bahama Swamp off Kato Bay Rd., Jasper Co., SC	32.356	-81.044	2	34	ACP
Edisto River (CROxSC)	38	Edisto River, SC	33.085	-80.603	2	4, 46	ACP
Back River (BACxSC)	39	Back R. at SSR 503, east of Goose Creek, edge of US Navy station, SC	32.967	-79.938	2	22	ACP
Cooper River trib. (WAD1xSC)	40	Trib. to Cooper River at Wadboo, Berkeley Co., SC	33.201	-79.927	1	-	ACP
Waccamaw River (WAR1xSC)	41	Waccamaw River, SC	33.474	-79.188	1	-	ACP
Trib. to Waccamaw River (WAR2xSC)	42	Unnamed trib. to Waccamaw River, SC	33.506	-79.181	1	4	ACP
Lumber River (LUMxSC)	43	Lumber River, SC	33.661	-79.153	1	4	ACP
Cane Branch (CANxSC)	44	Cane Branch on farm road at Francis Marion National Forest, Berkeley Co., SC	33.185	-79.527	1	21	ACP

Site names (numbers, localities, geographical coordinates), sample sizes (N), regional groups (Figure 1), and *cytb* haplotypes are indicated. Abbreviations: AL Alabama, Ave avenue, Co County, Cr Creek, CR County Road, FL Florida, GA Georgia, Hwy highway, MS Mississippi, Par Parish, R River, Rd Road, SC South Carolina, Trib tributary.

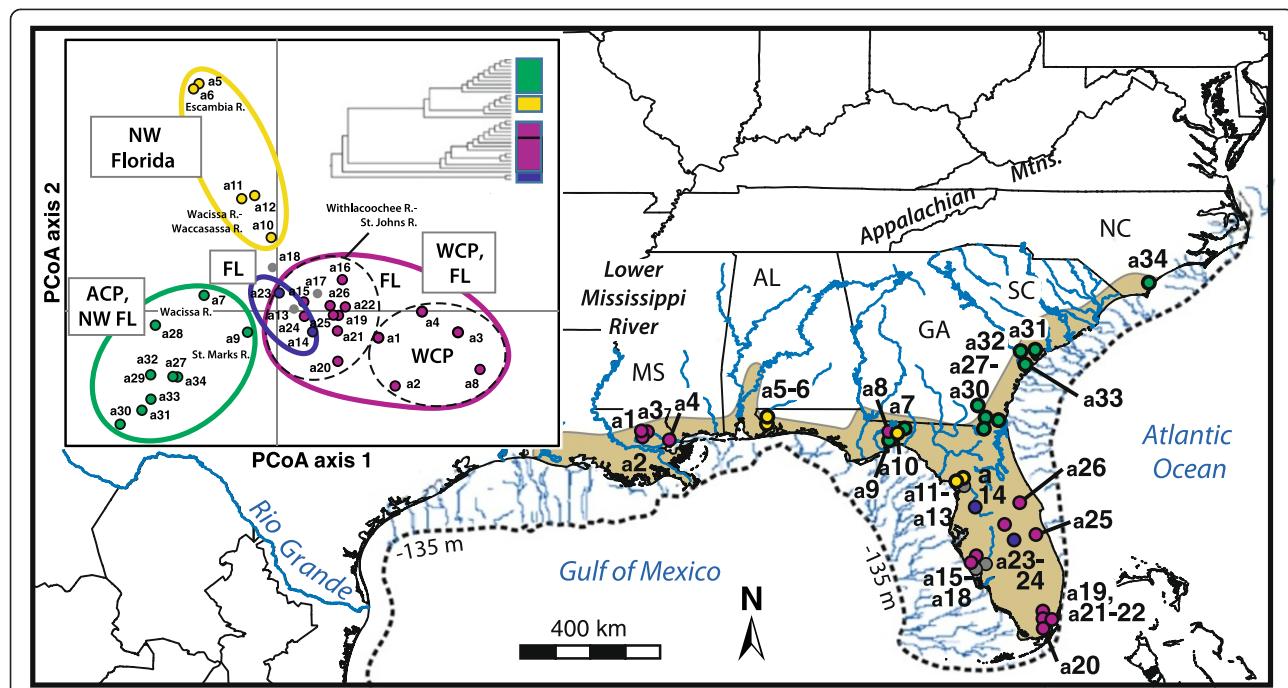
tx-variables in codes range 1-12, representing each individual sequenced for this study.

degrees. Putative loci included *Aat*, *Adk*, *Gpi-2*, *G3pdh*, *Idh-1*, *Ldh-2*, *Mdh-2*, *Mpi*, *Pgm-1*, *Pgm-2* and *6-Pgd*. Baer [42,54] showed these loci to be at Hardy-Weinberg equilibrium, so we included all 11 loci in analyses.

#### Ecological niche modeling

To test the prediction that *H. formosa* experienced a southward LGM range contraction and a northward range expansion from LGM to present, consistent with

the expansion-contraction model, we conducted ecological niche modeling using the maximum entropy (Maxent) approach implemented in MAXENT 3.3.3 k [55]. The Maxent method predicts locations of suitable habitat based on environmental-climatic characteristics and identifies abiotic and biotic factors that may limit species present-day distributions. We used Maxent because it exhibits high predictive performance [56].



**Figure 2** Nuclear allozyme phylogeography. Map of clades inferred from neighbor-joining analysis of unbiased Nei's *D* from 11 allozyme loci in [42]. Inset graphic: corresponding tree topology with assigned clade colors, as well as the results of principal coordinates analysis (PCoA) of allele frequency data for the populations, which confirm the neighbor-joining relationships.

We based our GIS-based niche-modeling analyses on 259 georeferenced *H. formosa* sites (Additional file 1: Figure S2) collated from our field collections (Table 1) and Global Biodiversity Information Facility (<http://www.gbif.org>) records. Sampling covered the species known distribution and was beyond sufficient given  $\geq 10$  sites permit accurate niche model construction [57]. From WorldClim (<http://www.worldclim.org>; [58]), we assembled GIS coverages with a resolution of 30 arc-seconds ( $1 \text{ km}^2$ ) for 19 bioclimatic environmental predictor variables (Additional file 1: Table S2) for present-day climates (1950–2000), plus analogous paleoclimatic data layers reconstructing past LGM (21 ka) and LIG (140–120 ka) environments. The LGM dataset (resampled from 2.5 arc-minutes resolution) was derived from the CCSM3 global circulation model of the Paleoclimate Modelling Intercomparison Project Phase II (PMIP2; [59]). The LIG dataset (originally 30-s resolution) was derived from climate simulations in Otto-Bliesner et al. [25]. Layers were clipped to an extent of  $24.5^\circ\text{N}$ – $35^\circ\text{N}$  and  $95^\circ\text{W}$ – $106^\circ\text{W}$  prior to analyses. Unfortunately, high-resolution data on freshwater environments is not available for all three time-periods modeled, although such data would aid niche-based modeling of freshwater organisms. However, the CCSM3 data estimate environments over continental shelf areas exposed during the LGM; thus, we were able to reproject our full models over subaerial LGM landmass within the modeled area, which was critical for testing the expansion-contraction model. We ran niche models using different combinations of predictor variables to broadly explore the utility of different variables and their effects on model prediction. Ultimately, we included all 19 data layers in the final modeling procedures, because models showed no evidence of over-fitting and MAXENT is robust to correlations among variables (Additional file 1: Supplement S1).

We predicted the current (0 ka) geographical distribution of *H. formosa*, by generating an ecological niche model in MAXENT using the program's default settings (e.g.  $10^4$  background points, logistic model with habitat suitability between 0 and 1), except we increased the number of iterations to 5000 to ensure convergence and we averaged results over 10 replicate cross-validation runs. All data were used for model testing and training. We reprojected the resulting niche models on LGM and LIG layers and interpreted areas with high-predicted LGM bioclimatic suitability as most-likely refugial areas. Under the expansion-contraction model, we expected areas of predicted occurrence to be reduced during the LGM relative to LIG and present-day, restricted to a southern refugium. By contrast, refugia located on either side of the Apalachicola River would support vicariance-northeast colonization, and a pattern of range stability through time would support southern crossroads. Model evaluation was based on threshold-independent measures of performance,

area under the Receiver Operating Characteristic curve (AUC) statistics; scores closer to 1 (maximum) indicate higher predictive ability and  $\text{AUC} > 0.5$  indicates better-than-random model prediction [56]. We generated response curves showing the impact of each variable alone on MAXENT prediction, and we studied 'multivariate environmental similarity surfaces' (MESS) output by the program to assess extrapolation and effects of novel environments on prediction.

### Population structure and genetic diversity

We evaluated DNA polymorphism levels by calculating segregating sites ( $S$ ),  $h$ ,  $Hd$ , and  $\pi$  in DnaSP for the full *H. formosa* *cytb* dataset, as well as populations and SAMOVA groups (see Results) that met a threshold-sampling criterion of  $N \geq 8$ . Due to many samples not meeting our  $N \geq 8$  threshold-sampling criterion, mtDNA data were not used to test predicted genetic diversity patterns. Watterson's estimator ( $\theta_W$ ; from  $S$ ) of  $\theta$  was calculated overall and for SAMOVA groups. Additionally, we calculated uncorrected pairwise mtDNA p-distances between individuals and groups in MEGA 5 [60].

We analyzed the mtDNA data to test for patterns of spatial-genetic subdivision predicted by the various phylogeographical hypotheses. The expansion-contraction model predicts genetic drift should have created distinct lineages subdivided between refugia—if multiple, allopatric refugia existed—especially over repeated glacial cycles [2,15]. However, if northward recolonization occurred from a single southern refuge via 'leading-edge colonization,' this should be reflected in a genetic distinction between the two areas, separating established (expanded) from blocked interior populations/lineages [2,9]. If the recolonization wave passed through spatial constraints, e.g. coastlines, we expect to identify most recent common ancestors (MRCA) through the coalescent and genetic distinctions that mark the location of barriers encountered outside of a refuge [61]. For what we term the 'vicariance-northeast colonization' hypothesis (uniting Coastal Plain vicariance e.g. [29,32] and northeast colonization [42] elements), we tested for genetic barriers in the Apalachicola River-region between ACP-FL and WCP samples, and between the Suwannee and Hillsborough Rivers along Florida's Gulf Coast (our sampling analog to the area of Baer's [42] Waccasassa-Withlacoochee break). However, we expected genetic barriers to be absent within the ACP [42]. By contrast, a southern crossroads scenario would be supported by a pattern of no phylogeographical structure [21].

We tested these predictions using Monmonier's [62] algorithm, implemented in BARRIER 2.2 [63], to identify genetic 'barriers' across a Delaunay triangulation network overlaying the sampling sites, based on maximum Tamura-Nei genetic distances calculated in Arlequin 3.5 [64] (1000 nonparametric permutations). To assess barrier

support, we calculated bootstrap proportions from 100 bootstrapped barriers generated in BARRIER using 100 bootstrapped Tamura–Nei distance matrices made in PopTools [65]; bootstrap values >50 indicated strong support. Next, we used spatial analysis of molecular variance (SAMOVA), implemented in SAMOVA 1.0 [66], to define population clusters maximizing the proportion of total genetic variance due to differences among geographical groups ( $\Phi_{CT}$ ). We performed SAMOVAs across  $K = 2\text{--}8$  groups using pairwise differences, drawing from any of 100 initial random conditions. We independently tested the grouping schemes indicated by the SAMOVA and BARRIER analyses by conducting analyses of molecular variance (AMOVA) in Arlequin (1000 nonparametric permutations). Because BARRIER and SAMOVA are sensitive to small sample sizes, we pooled adjacent sites until an  $N \geq 8$  threshold-sampling criterion was reached per population, yielding a 22-population subset.

We also conducted additional spatial analyses. Using AMOVA, we tested for significant genetic partitioning among drainages. Here, we capitalized on our dense north Florida sampling and split populations across seven drainages, including the Ochlockonee (collections 7–15, 17, and 20), St. Marks (16, 19, and 21–22), Apalachee plus Goose Creek Bays (18, 23–24), Aucilla (25–27), Suwannee (28–31), Hillsborough (32), and St. Johns Rivers (34–35). We employed GENALEX 6.3 [67] to examine predictions about patterns of isolation-by-distance, using Mantel tests [68] for correspondence between mtDNA genetic distances [ $F_{ST}/(1 - F_{ST})$ ] and geographic distances [ $\ln(\text{km})$ ] among sites at rangewide and FL-region levels ( $N \geq 8$  threshold-sampling;  $10^4$  permutations; relationships confirmed using linear regression). We used straight-line geographic distances, not river distances, because several sites were bodies of water (e.g. lakes) that are isolated from any nearby rivers. Based on our niche models (see Results), we predicted much of the range would lack isolation-by-distance; however, under vicariance-northeast colonization, only ACP populations should lack isolation-by-distance [42].

We tested for genetic diversity patterning consistent with the expansion-contraction model using analyses of the allozyme dataset. Coalescent theory predicts that northward postglacial expansion from reduced  $N_e$  should have created genetic signatures of population expansion in recently founded (recolonized) populations. These include excesses of rare alleles and low-frequency mutations, reduced genetic diversity (homozygosity, shorter demographic histories) resulting in negative latitudinal clines, and a lack of isolation-by-distance (reviewed in [4,9,69,70]). Conversely, long-term stable refugia should display isolation-by-distance and higher genetic diversity. We tested for predicted latitudinal clines in diversity by calculating expected heterozygosity ( $H_e$ ; Nei's expected heterozygosity) over all loci in GENALEX (from allele frequencies), then testing for

a negative relationship with latitude [4,9,17] using linear regression models in PAST 2.02 [71]. An absence of clinal diversity patterning would be more consistent with a southern crossroads scenario. We tested whether allozyme  $H_e$  and private allelic richness ( $h_p$ ) were greater within the niche modeling-inferred refugial populations (see Results) than throughout the putatively recolonized modern range using Mann-Whitney and Kolmogorov-Smirnov tests in PAST (due to non-normality, non-homogeneous variance). Using similar nonparametric tests, we tested whether WCP or ACP diversity was significantly lower than that of other populations combined, or populations within the putative refuge. And we tested for expected patterns of (lack of) isolation-by-distance using Mantel tests of unbiased Nei's  $D$  genetic distances and  $\ln$  geographic distances between populations, calculated in GENALEX at rangewide, WCP, ACP, FL, and within-refuge levels.

### Historical demography

We tested for population expansions predicted by the expansion-contraction model (as per above) at levels of regional groups, SAMOVA groups (see Results) and local populations ( $N \geq 8$  threshold size) using complementary *cyt b* analyses. We considered the expansion-contraction model 'strongly' supported where at least two statistical tests supported expansion. First, we estimated Fu's  $F_S$ , and  $R_2$ , and their 95% confidence intervals using coalescent simulations in DnaSP (testing significance with  $10^4$  replicates). To distinguish population expansion from purifying or positive selection, we tested each group, and the full *cyt b* database, for mtDNA neutrality using McDonald and Kreitman's [72] test and coalescent simulations of Fay and Wu's  $H$  [73] in DnaSP ( $10^4$  replicates). We used Poeciliidae sp. as the McDonald–Kreitman test outgroup; other potential outgroups from Additional file 1: Table S1 yielded identical results (unpublished data). Second, we conducted mismatch distribution tests for regions/groups with no apparent subdivision (WCP, ACP and SAMOVA-group levels) in Arlequin (1000 parametric bootstrapping iterations) to see if pairwise nucleotide difference frequencies rejected sudden- or spatial-expansion models (given similar results, only sudden-expansion results are presented). We calculated Harpending's raggedness index ( $r$  [70]) as a test statistic measuring goodness-of-fit to the mismatch distributions. We calculated the time in generations since population expansion ( $t$ ) using estimated expansion parameters and the equation  $t = 2 \mu t$  ( $\tau$  = coalescent time in mutations,  $\mu$  = mutation rate/locus/year based on our Bayesian *cyt b* rate estimate; see Results);  $t$  was converted to time in years using generation time. We expected mean  $t$  and/or 95% confidence intervals to overlap the LGM.

### Phylogenetic relationships and coalescent-dating analyses

We expected multiple Pleistocene refugia to be supported by allopatric clades/unique networks [4,15] with strong

nodal/parsimony support, whereas unique clades or intra-network structuring between refugia inferred from niche modeling and recolonized areas would support leading-edge effects [2,9]. Another phylogenetic prediction is that derived haplotypes (mutations at phylogeny/network tips) should become geographically localized outside of refugia with ancestral network populations found within refugia [61]. The placement of one or more network roots outside of the inferred refugia would be consistent with a model of allele surfing [61] or extinction of trailing expansion edges [9]. Here, southern crossroads is a null model with no obvious phylogeographical patterning within a single population-lineage.

We inferred relationships among *H. formosa* *cytb* haplotypes with nodal support using maximum-likelihood tree searches and bootstrap searches (500 pseudoreplicates) in GARLI 0.97 [74], partitioning the data by codon position, ((1 + 2), 3). We analyzed our multilocus dataset using similar GARLI runs partitioning the data into *cytb* codon-position subsets and an *RPS7*-gene subset, unlinking parameters across subsets. GARLI runs relied on DNA substitution models selected using the decision theory algorithm in DT-ModSel [75] (Supplement S1). We compared GARLI results to phylogenetic results from Bayesian coalescent-dating analyses (below) with similar site models. Because incomplete lineage sorting can obscure intraspecific phylogenetic relationships, we derived a 'species tree' for our *cytb* haplotypes using the 'Minimize Deep Coalescences' method [76]. Although crude, this method increases probability of obtaining accurate population trees using even a single locus and yields insights into the degree of lineage sorting [77]. We also inferred phylogenetic relationships among populations from the allozyme dataset. We imported unbiased Nei's *D* estimates (above) into PAUP\* 4.0b10 [78] and used them to construct a midpoint-rooted neighbor-joining tree with total-distance branch lengths. We tested the genetic distinctiveness of the resulting clades in multivariate space using principal coordinates analysis (PCoA) conducted in GENALEX. Last, we inferred networks of *cytb* and *RPS7* haplotype relationships with connections exceeding 95% parsimonious probability in TCS 1.21 [79].

We evaluated temporal diversification by estimating *H. formosa*-Poeciliidae sp. divergence, *H. formosa* population/clade divergence and other node ages using coalescent-dating analyses of our multilocus matrix in BEAST 1.74 [80]. To ensure convergence, we ran three independent searches (MCMC =  $10^8$ , sampled every 4000 generations; burn-in =  $10^7$ ) using identical priors and relaxed, uncorrelated lognormal molecular clocks. We linked tree models but partitioned the data into *cytb* codon-position subsets ((1 + 2), 3) and an *RPS7*-gene subset and unlinked clocks and parameters across subsets. Coalescent constant population size tree priors were

used. We set uniform priors on rates drawn on by the lognormal relaxed clock that, for *cytb*, spanned protein-coding mtDNA gene substitution rates reported for teleost fishes ('fish rate' =  $0.017\text{--}0.14 \times 10^{-8}$  subs/site/yr, per-lineage; refs. in [81,82]), and for *RPS7* spanned an arbitrary range of reasonable rates ( $1.0 \times 10^{-10}\text{--}1.0 \times 10^{-8}$  subs/site/yr) consistent with nDNA rates for freshwater fishes (unpublished data). By including *Poecilia* (subgenus *Limia*) outgroups in our analysis, we were able to add a calibration point constraining the split between *P. (L.) dominicensis* (from Cuba) and *P. (L.) vittata* (Hispaniola) to 17–14 Ma based on *Poecilia* (*L.*) phylogeny [83] and geological dates for the separation of Cuba and Hispaniola [84], following Doadrio et al. [85]. We calibrated this node using a lognormal prior (mean in real space = 1, log standard deviation = 1.25, offset = 14). We used similar calibrations to model basal *Pseudoxiphophorus* divergence between 11–5 Ma, following Agorreta et al. [86], and to model the tree root age (Poeciliidae) to 39.9 Ma with an extended tail (log standard deviation = 2.0) based on the oldest fossils known for the family from the Paleocene-Eocene Maiz Gordo and Lumbra formations, Argentina [87]. We summarized posterior parameter distributions and calculated parameter-trace effective sample sizes (ESS) in TRACER (<http://tree.bio.ed.ac.uk/software/tracer>). We archived our sequence alignments and maximum-likelihood and Bayesian tree results in TreeBASE (Submission 14718; <http://purl.org/phylo/treebase/phylows/study/TB2:S14718>).

### Hypothesis testing

We used coalescent simulations of mtDNA in Mesquite 2.73 [88] to statistically discriminate among phylogeographical scenarios representing the expansion-contraction model and two alternatives. This approach guards against the possibility that stochastic coalescences might make it more difficult to detect a signature of one or more models [43,44]. Population tree models in the simulations (described in Results, Supplement S1) spanned a total time ( $t_{\text{Total}}$ ) equal to tree depths of  $3.741 \times 10^6$  generations, derived by converting the mean intraspecific time to the most recent common ancestor ( $t_{\text{MRCA}}$ ) estimate from BEAST from absolute time to coalescent time (=1.247 Ma/T). We scaled branching intervals so that they summed to  $t_{\text{Total}}$ . We estimated overall  $N_e$  for simulations (assumed equal to ancestral  $N_e$ ) by summing female  $N_e$  ( $N_{ef}$ ) estimates from  $\theta$  values, calculated for each SAMOVA-inferred population using the equation  $\theta = 2N_{ef}\mu$  (for mtDNA,  $\mu = \text{cytb}$  rate estimated in BEAST). Thetas were derived from  $\theta_w$  (above), and from Bayesian  $\Theta$  estimates for SAMOVA groups, averaged from three replicate MIGRATE-N 3.1.3 [89] runs each using 1 long chain ( $3 \times 10^8$ ) and adaptive heating (chains = 1, 1.5, 3,  $10^4$ ; Supplement S1). We also estimated effective numbers of female migrants per generation ( $N_{ef}m$ ) from

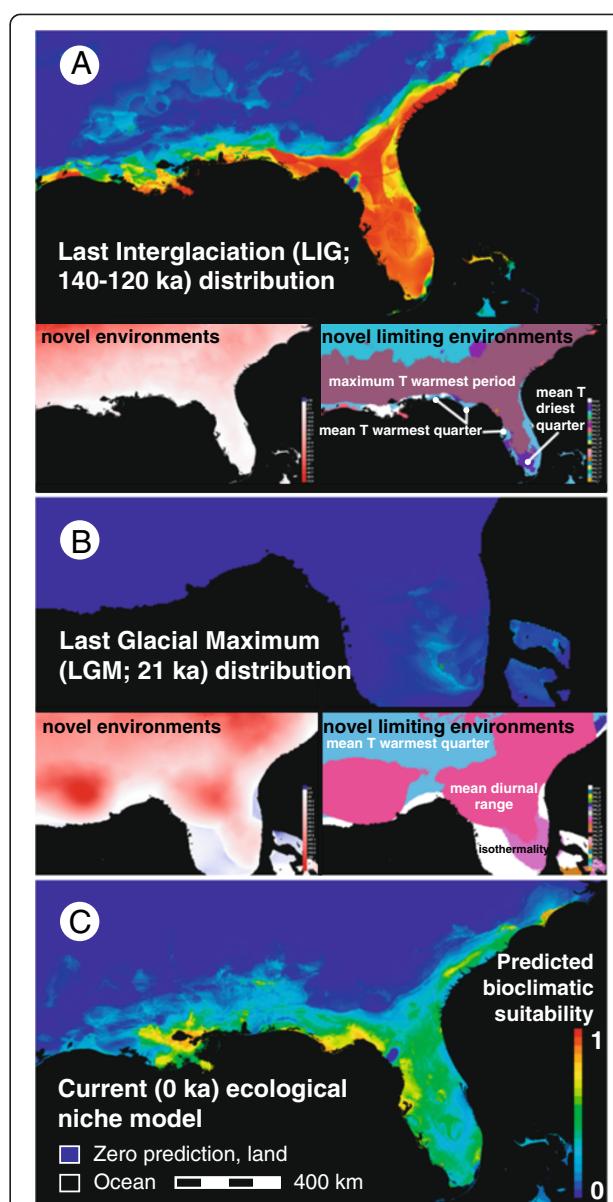
mtDNA in MIGRATE-N. Populations (internal branches) were modeled using groupings of subpopulations (areas, SAMOVA groups) appropriate to each hypothesis, and we scaled population-lineage sizes (internal branch widths) to their respective proportions of overall  $N_e$ , so they summed to overall  $N_e$  at all time points [90]. Population-lineage sizes were equally distributed among tip branches (subpopulations) evolving from them. We conducted hypotheses testing in Mesquite by simulating 1000 mtDNA gene genealogies within the expansion-contraction model under neutral coalescence. For these simulated gene trees, we calculated the ‘number of deep coalescences’ (nDC), a measure of gene tree-population tree discordance arising from incomplete lineage sorting [91]. We evaluated model fit by comparing nDC for the ‘best’ *cytb* maximum-likelihood gene tree to that from the simulated gene trees (treated as rooted). We rejected the expansion-contraction model in favor of an alternative by a one-tailed significance test ( $\alpha = 0.05$  level) [43].

## Results

### Ecological niche modeling

Ecological niche models predicting the species LIG, LGM, and present-day distributions are shown in Figure 3 and support expansion-contraction, not vicariance-northeast colonization or southern crossroads expectations. As expected, the paleoclimatic models predicted a wide LIG distribution (Figure 3A), drastic reduction in LGM bioclimatic suitability and southward range contraction to a refuge (Figure 3B), then northward recolonization of the modern range (Figure 3C). The putative refugium spanned roughly  $\geq 1500 \text{ km}^2$  area in an area of inland southwest Florida. The refugial area extended from the peninsula’s tip near Biscayne Bay ( $\sim 25.15^\circ\text{N}$ ) to its southwestern slope west of the Immokalee Rise (50 m ASL), north to Charlotte Harbor ( $\sim 28^\circ\text{N}$ ) and east across the Caloosahatchee Valley into south-central Florida (Figure 3B). The predicted bioclimatic suitability exceeded 0.079 (minimum logistic threshold). Additionally, we found that ACP and WCP-region suitability was similar across LIG/present-day interglacial models, as was a localized but distinct break in bioclimatic suitability recovered between the Suwannee and Withlacoochee Rivers.

Highly predicted (>80%) present and LIG ranges overlapped in patches around Lake Pontchartrain, Florida’s Apalachicola Bay-Apalachee Bay and western peninsula (Hillsborough River), and the northwest Atlantic Coast range margin. Under more conservative suitability criteria (logistic thresholds), present-day and LIG-modeled ranges covered >90% of the modern range. There were virtually no false negatives, but we obtained minor false positives where *H. formosa* does not occur (e.g. above the Fall Line, in the Bahamas).



**Figure 3** MAXENT reconstructions of Pleistocene Last Interglaciation (LIG) and Last Glacial Maximum (LGM), and present-day distributions. **A)** Reprojection of *H. formosa* ecological niche model (**C**) on paleoclimatic data layers representing LIG environments. **B)** Reprojection of *H. formosa* ecological niche model on colder/drier environmental conditions of the LGM, showing southward range contraction, possibly to ‘microrefugia’. **C)** Present-day ecological niche model. Colors represent logistic ecological niche model scores and range from 0 (dark blue, indicating unsuitable subaerial areas) to 1 (100% bioclimatic suitability, thus higher predicted probability of species occurrence).

Statistical model-evaluation analyses and comparisons of variable contributions supported these conclusions. Based on AUC scores from 10 replicates, models performed well with high discriminatory ability (LGM minimum training presence logistic threshold = 0.079,

$AUC = 0.920 \pm 0.016$ ; LIG minimum training presence logistic threshold = 0.022,  $AUC = 0.915 \pm 0.018$ ). Average test model gain ( $>1.5$ ) also indicated  $>4$ -fold better-than-random prediction. The principal environmental predictor of occurrence was precipitation during warmest quarter, followed by temperature seasonality (Additional file 1: Tables S3-S4). MESS diagrams showed near-coastal regions experienced more novel climates during the LGM than the LIG, relative to present day, causing model extrapolation (Figure 3A,B). Still, novel environments were not a major issue: predictor variables most responsible for environmental novelty (novel limiting environments), e.g. isothermality, contributed  $\leq 2.1\%$  to MAXENT model prediction.

#### Population structure and genetic diversity

We found substantial mean intraspecific mtDNA haplotype diversity ( $\pm$  standard deviation:  $H_d = 0.934 \pm 0.006$ ) but low mean nucleotide diversity ( $\pi = 0.0066 \pm 0.009$ ) and surprisingly shallow genetic divergence (mean cyt b p-distance = 0.66%, range: 0-1.61%), all of which are consistent with a recent, rapid population expansion for the species. Using DnaSP, we calculated an empirical estimate of  $\theta_W = 0.0087$  for all *H. formosa* populations combined. SAMOVA-group estimates of population mutation rate parameter were moderate, with  $\theta_W$  ranging 0.0031-0.0053 (Additional file 1: Table S5).

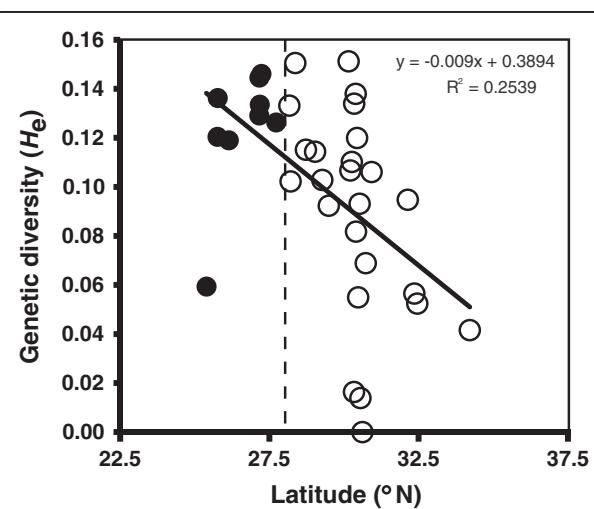
All three methods of spatial genetic analyses recovered three locations of genetic distinctions that defined four groups ( $K$ ) within the species range (Figure 1). The groupings were strongly supported by bootstrap analysis (BARRIER BP = 55-99); significant spatially explicit among-group variance partitioning (SAMOVA  $P < 0.0001$ ); and AMOVAs (Supplement S1). While SAMOVA and BARRIER analyses were not completely concordant, both algorithms recovered genetic barriers outside of the ecological niche modeling-inferred LGM refugium, and both inferred complex structuring between the Apalachicola delta and Aucilla River. These results are most consistent with leading-edge colonization or allele surfing (or other factors) during post-glacial range expansion. However, they do not firmly reject vicariance-northeast colonization: barriers were identified between WCP and ACP regions, but not FL and ACP; groups were divided north-to-south between the Suwannee and Hillsborough Rivers; and no ACP genetic barriers were inferred.

Most genetic barriers reflected breaks between drainages, and we also found significant among-drainage genetic partitioning in Florida (AMOVA: Among-group variation = 33.73%,  $\Phi_{CT} = 0.34$ ,  $P < 0.0001$ ). The most striking SAMOVA-inferred subdivision was isolation of this northern Florida area from all other populations (WCP, south Florida peninsula, ACP). However, within northern Florida, SAMOVA connected geographically

disjunct lakes/ponds (collections 13-15, 17, and 20), Womack Creek (10), and Robinson Creek (30), and linked disjunct Newnan's Lake (35) and Tate's Hell sites (7-8). BARRIER recovered contiguous groups (Figure 1) and split the WCP from other areas (barrier ii); however, given disparities with our other DNA and allozyme results, and barrier ii's placement in a sampling gap, we interpreted the WCP group as a manifestation of the limitations of our mtDNA sampling.

Mantel tests indicated that these conclusions were not unduly influenced by spatial autocorrelation in cyt b variation, whether rangewide ( $r = -0.007$ ,  $P = 0.492$ ; regression  $R^2 = 5 \times 10^{-5}$ ,  $t = -0.11$ ,  $P = 0.913$ ) or only within Florida ( $r = -0.054$ ,  $P = 0.391$ ; regression  $R^2 = 0.003$ ,  $t = -0.74$ ,  $P = 0.460$ ). MtDNA isolation-by-distance was thus overall lacking, reflecting low diversity and alleles shared among regions/populations; for example, haplotype 22 was common to all three regions and most ACP populations shared haplotype 4 (Table 1 and S6). We noted a similar lack of nDNA isolation-by-distance, with *RPS7* haplotype 2 present in each study region (Additional file 1: Table S7). These results agree with expansion-contraction predictions as well as the independently derived niche models.

Consistent with expansion from the niche-modeling inferred refugium, allozyme genetic diversity ( $H_e$ ) decreased with increases in latitude moving northward from the likely refugium ( $R^2 = 0.254$ ,  $t = -3.30$ ,  $P = 0.002$ ; Figure 4). Consistent with the inferred refugial location (Figure 3B), allozyme diversity was above average and statistically higher within, rather than outside, the putative refuge (WCP, ACP,



**Figure 4 Latitudinal cline in *Heterandria formosa* allozyme genetic diversity.** This linear regression model shows a strong negative relationship between allozymic genetic diversity ( $H_e$ ) and latitude ( $P = 0.002$ ). Black dots indicate genetic diversity of sites below 28°N latitude (dashed line; the temperate-tropical transition), with bioclimatic prediction above the logistic threshold in our Last Glacial Maximum model (Figure 3B), i.e. within the putative LGM refugium.

'outside' populations combined), as judged by  $H_e$  (Table 2). However,  $h_p$  was not statistically different within-vs.-outside the putative refuge (Table 2). Consistent with the mtDNA results, Mantel tests of the 11 allozyme loci revealed significant isolation-by-distance in the putative refuge ( $r = 0.360$ ,  $P = 0.014$ ; regression  $R^2 = 0.498$ ,  $t = 2.815$ ,  $P = 0.022$ ). We also found that isolation-by-distance was significant rangewide ( $r = 0.236$ ,  $P = 0.001$ ; regression  $R^2 = 0.056$ ,  $t = 5.77$ ,  $P < 0.0001$ ) and within regions (see Supplement S1 for details) except the ACP ( $r = 0.031$ ,  $P = 0.323$ ; regression  $R^2 = 0.001$ ,  $t = 0.18$ ,  $P = 0.855$ ). Allozyme polymorphism levels also supported the vicariance-northeastern colonization prediction of relatively lower ACP diversity (Table 2).

### Historical demography

Historical-demographic analyses of *cytb* variation strongly supported predicted patterns of population expansions among regions and population groups. The  $R_2$  statistic was significant ( $P < 0.0001$ ) and positive for the species, and for each regional group, SAMOVA group, and population analyzed (Table 3, S5). One positive Fu's  $F_S$  estimate was significant ( $P < 0.02$ ), indicating a genetic bottleneck in the WCP region (Table 3). Further testing rejected the possibility that  $R_2$  and  $F_S$  deviations from migration-drift equilibrium resulted from positive or purifying selection (full *cytb* database, McDonald-Kreitman test,  $P = 0.33$ ; mean  $H = -0.035$ , 95% confidence intervals: [-14.93, 5.45],  $P = 0.33$ ; see Additional file 1: Table S5 for other results). Mismatch distributions [70,92] were mostly unimodal or smoothly descending with modal numbers of differences between haplotypes ranging 0-2 for WCP, ACP, and SAMOVA groups (except small peaks of divergent haplotypes at low frequencies; Figure 5). Moreover, based on Harpending's  $r$ , we failed to reject models of sudden demographic expansion in all cases (Table 3), indicating the coalescent model accurately

accounted for observed polymorphism patterns. From  $\tau$ , we inferred a Late Pleistocene timing of intraspecific expansions (20.4-61.6 ka; Table 3), with the onset of expansion preceding the LGM by 0-40 ka, and overlapping estimates suggesting concurrent expansions since ~33-10 ka. Except in SAMOVA group 2, we failed to reject LGM expansions because confidence interval included values less than 22 ka (Table 3; Figure 5).

### Phylogenetic relationships and coalescent-dating analyses

Within our 223-*cytb* dataset, we found 47 unique *H. formosa* haplotypes distinguished by 59 segregating sites, and two unique Poeciliidae sp. haplotypes. Subsampling across regions yielded a final multilocus alignment of *cytb* and *RPS7* sequences for 17 *H. formosa* and 2 Poeciliidae sp., plus outgroups in Additional file 1: Table S1. Maximum-likelihood (Additional file 1: Figure S3) and Bayesian phylogenetic searches (Figure 6A; below) on the multilocus alignment recovered identical topologies with a monophyletic *H. formosa* comprised of a single lineage lacking strongly supported ( $BP \geq 70$  and  $PP \geq 95$ ) phylogenetic structure except a subclade (subclade 'a') representing the Ochlockonee River in part, Aucilla River, and the Suwannee River in part; otherwise, geographical structure was insignificant.

Several features of the sequence data supported predictions derived from the expansion-contraction model and were not consistent with the predictions of a vicariance-northeast colonization model. For one, no deep intraspecific east-west breaks were supported. This is not because there were no deep phylogenetic breaks detected; the predicted sister relationship between Poeciliidae sp. and *H. formosa* (maximum *cytb* p-distance = 9.01%) was strongly supported (Figure 6A, Additional file 1: Figure S3), indicating a deep split between Mexico and the Atlantic Plain. Our 'best' *cytb* haplotype gene tree ( $\ln L = -2000.8824$ ) revealed shallow variation and a close Ochlockonee-Aucilla-

**Table 2 Nonparametric tests of spatial-diversity predictions based on allozyme variation**

Comparison (group 1-vs.-group 2)	Mean 1	Mean 2	N	Mann-Whitney tests		Kolmogorov-Smirnov tests		
				U	z	P	D	P
$H_e$								
Within-vs.-outside putative refuge	0.122	0.089	34	59	-2.287	0.022*	0.55	0.016*
ACP-vs.-rest of range	0.071	0.108	34	36	-2.740	0.006**	0.69	0.003**
ACP-vs.-putative refuge	0.071	0.122	14	5	-3.065	0.002**	0.80	0.002**
WCP-vs.-rest of range	0.072	0.105	34	49	-1.559	0.119 ns	0.43	0.244 ns
WCP-vs.-putative refuge	0.072	0.122	16	8	-2.332	0.019*	0.70	0.0251*
$h_p$								
Within-vs.-outside putative refuge	0.036	0.015	34	99	-1.168	0.243 ns	0.41	0.141 ns
ACP-vs.-putative refuge	0.000	0.036	14				0.30	0.738 ns
WCP-vs. putative refuge	0.000	0.036	16				0.40	0.472 ns

$H_e$ , expected heterozygosity;  $h_p$ , private allelic richness; N, number of populations compared; \* $P < 0.05$ ; \*\* $P < 0.01$ ; ns, not significant.

**Table 3 Neutrality and mismatch distribution tests for population expansions within *Heterandria formosa***

Group	N	$F_S$	$R_2$	r	$\tau$	Expansion time (ka)
WCP region	8	<u>0.288**</u>	<b>0.201**</b>	<b>0.240 ns</b>	1.195 [0.000, 10.393]	24.5 [0.0, 213.4]
ACP region	9	<u>0.241</u>	<b>0.178**</b>	<b>0.169 ns</b>	1.498 [0.000, 3.119]	30.8 [0.0, 64.0]
SAMOVA group 1	10	<u>0.218</u>	<b>0.169**</b>	<b>0.061 ns</b>	1.672 [0.000, 9.023]	34.3 [0.0, 185.3]
SAMOVA group 2	115	<u>-0.267</u>	<b>0.090**</b>	<b>0.058 ns</b>	2.408 [1.256, 3.031]	49.4 [25.8, 62.2]
SAMOVA group 3	51	<u>-0.053</u>	<b>0.108**</b>	<b>0.095 ns</b>	0.992 [0.492, 1.615]	20.4 [10.1, 33.2]
SAMOVA group 4	23	<u>0.081</u>	<b>0.133**</b>	<b>0.482 ns</b>	3.000 [0.410, 3.172]	61.6 [8.4, 65.1]

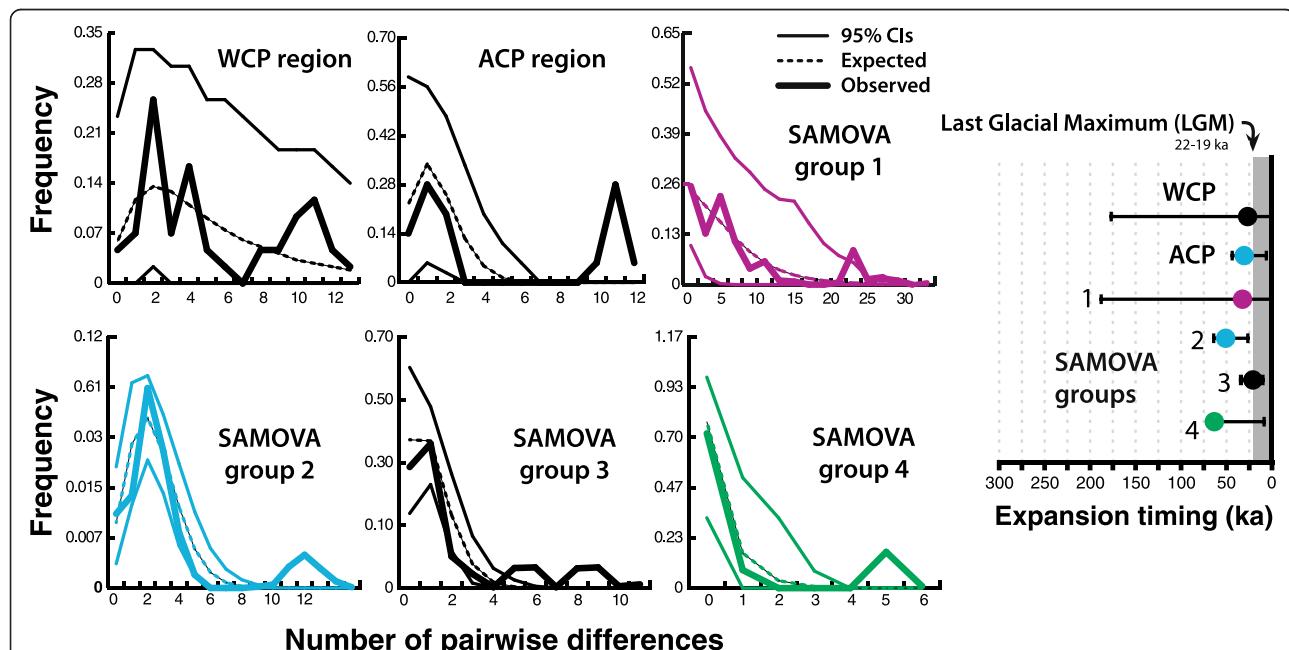
Sample sizes and coalescent simulation results for Fu's  $F_S$ ,  $R_2$ , Harpending's raggedness index ( $r$ ), and mutation time parameter  $\tau$  calculated from a sudden-expansion demographic model. Multiple tests supported population expansions for each group of populations (boldface); one test supported a population bottleneck (underlined). Population expansion dates (thousands of years ago, ka) were estimated from  $\tau$  using a per-locus mutation rate ( $\mu = 8.12 \times 10^{-6}$ ) converted from the BEAST *cyt b* estimate. Regional results are only presented for regions with no SAMOVA-inferred subdivision. Brackets contain 95% confidence intervals; ns, not significant; \*\*P < 0.0001 (P < 0.02 for Fu's  $F_S$ ).

Suwannee relationship, and supported differentiation (albeit incomplete) of SAMOVA groups 2 and 3. This topology differed greatly from the 'Minimize Deep Coalescences' topology, indicating substantial incomplete lineage sorting (Figure 6B).

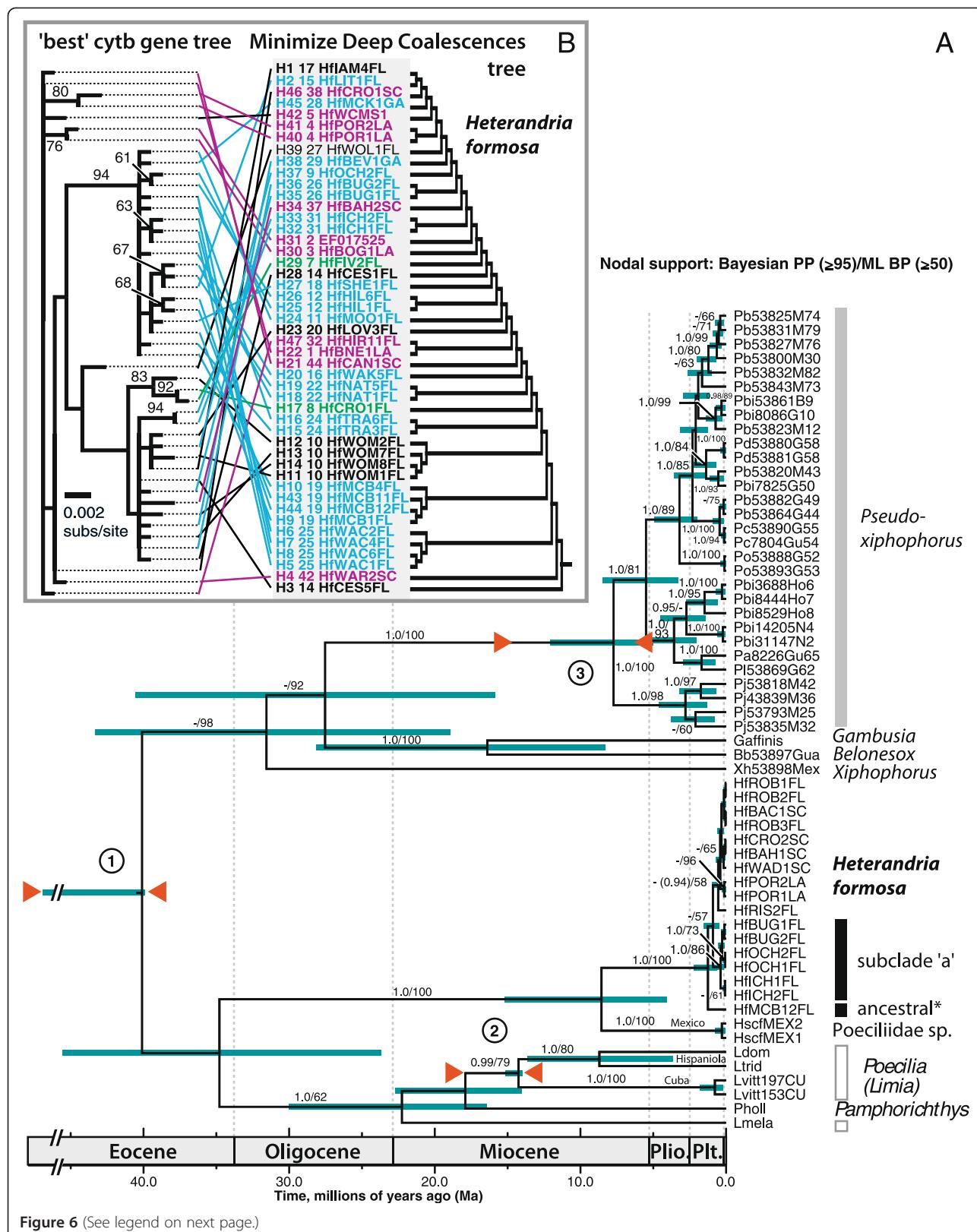
Consistent with shallow variation and a single population-lineage experiencing recent expansion-contraction dynamics, *cyt b* and nuclear *RPS7* haplotypes formed single networks (Figure 7) and *RPS7* sequences from the Apalachicola River-Apalachee Bay region were separated by only 1-3 mutations from haplotypes spread throughout the rest of the species range. Consistent with expansion-contraction, populations (alleles) in the putative LGM refuge (haplotypes 22, 30, and 47) were mostly recovered as ancestral/

basal in the inferred *cyt b* gene tree (100%) and network (66.7%). By contrast, outside of the putative LGM refugium (44 haplotypes) we found mostly derived populations (tip alleles) in the gene tree (81.8%) and network (70.5%). The *cyt b* network reflected the population structure inferred from the SAMOVA analysis and had two cases of star-like patterns with branches radiating out from inferred root-haplotype 10 (McBride Slough) and haplotype 3 (Cessna Pond, collection 14). Also, many (39%) network alleles were unsampled, indicating alleles potentially lost to glacial-stage extinctions.

Allozyme phylogenetic analysis recovered four shallow, spatially overlapping clades that were also genetically distinct during PCoA multivariate projection, with axes



**Figure 5** The six graphs at left show the frequency distributions of the numbers of pairwise *cyt b* differences among *H. formosa* individuals for each SAMOVA-inferred population group and regional group, with confidence intervals on expected values derived from parametric bootstrapping ( $10^4$  iterations) in Arlequin. Mean estimated timing of population expansions (circles) and 95% confidence intervals from parametric bootstrapping (bars) are plotted at right, with grey shading indicating time since onset of the LGM.



(See figure on previous page.)

**Figure 6 Chronogram of *Heterandria formosa* and related poeciliid species diversification, and gene tree results. A)** Chronogram resulting from Bayesian relaxed-clock coalescent-dating analysis in BEAST based on mitochondrial *cytb* and *RPS7* variation. Tip labels are sequence codes including population, site number, and specimen code for each individual sequenced (details in Figure 1 and Table 1). Node bars (dark blue) are 95% highest posterior densities for node ages. The analysis included three lognormally modeled fossil/biogeographic calibration points (red triangles enclose bounds of constraints). Mean node ages of interest are discussed in detail in the text. Nodal support values are of the form: Bayesian posterior probabilities (PP;  $\geq 95$ )/maximum-likelihood bootstrap proportions (BP;  $\geq 50$ ). **B)** Comparison between the 'best' gene tree of *H. formosa* *cytb* haplotypes and the 'Minimize Deep Coalescences' species tree (at right) inferred from the haplotype tree using Maddison and Knowles' [76] method, with BP  $\geq 50$  by each node.

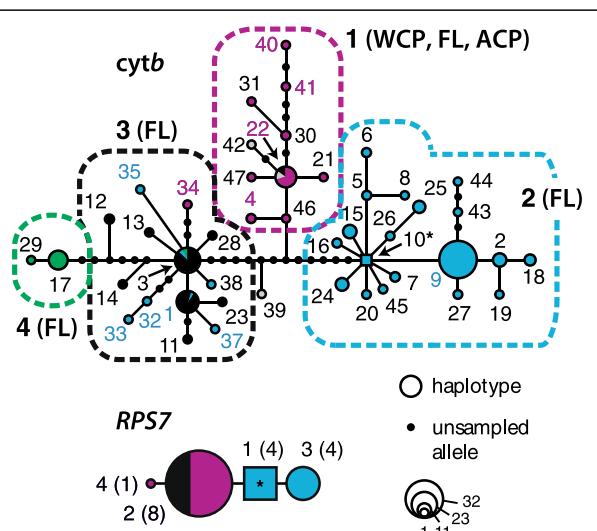
1 and 2 explaining 39.9% and 28.8% of the genetic variance, respectively (Figure 2). Consistent with expansion-contraction predictions, allozyme results corroborated SAMOVA-inferred population structure, with a north-south break between the Suwannee and Hillsborough Rivers, separating a group of northwestern plus south Florida populations from all others. Allozyme results also agreed with other phylogenetic results: the northwest Florida allozyme clade matched the strongly supported subclade 'a' in Figure 6A. Contrasting our DNA sequence-based results, and consistent with vicariance-northeast colonization, neighbor-joining and PCoA recovered ACP populations as a genetically and geographically distinct clade/group. However, against vicariance-northeast colonization expectations, intraspecific clades/groups were surprisingly not split east-to-west near the Apalachicola

River-region (Figure 2) or St. Mary's River, as suggested by [42].

The model resulting from multilocus coalescent-dating in BEAST (Figure 6A;  $\ln L \pm$  standard error =  $-10,054.902 \pm 0.118$ ) yielded a geometric mean  $t_{\text{MRCA}}$  of *H. formosa* samples dated to 1.247 Ma [0.620, 2.178] in the Early-Middle Pleistocene. The  $t_{\text{MRCA}}$  estimate for subclade 'a' was 0.377 Ma [0.153, 0.713]. Bayesian posterior distributions for these  $t_{\text{MRCA}}$  estimates strongly rejected LGM divergence ( $P < 0.05$ ; <5% of posterior samples estimated a  $t_{\text{MRCA}}$  for this node younger than 22 ka). Thus, the deepest population structuring (interpopulation divergence) within *H. formosa* appears to have initiated before the LGM. We estimated divergence between Poeciliidae sp. and *H. formosa* at 8.478 Ma [3.939, 14.913] in the Late Miocene-Early Pliocene; confidence intervals for this split were much wider than the  $t_{\text{MRCA}}$ s above but we still statistically rejected Pleistocene divergence ( $P < 0.05$ ). Parameter-trace ESS scores  $>600$  indicated thorough Markov chain mixing. Our BEAST runs estimated a geometric mean evolutionary *cytb* rate of  $7.12 \times 10^{-9}$  substitutions/site/yr, and an *RPS7* rate of  $1.01 \times 10^{-9}$  substitutions/site/yr.

#### Hypothesis testing

By summing regional group  $N_{\text{ef}}$  estimates based on empirical mtDNA  $\theta_W$  (Additional file 1: Table S1) and  $\Theta$  (range: 0.00106-0.00360) estimates, we respectively inferred overall  $N_{\text{ef}} \approx 6.05 \times 10^5$  and  $3.53 \times 10^6$ . We also inferred migration rates among SAMOVA groups (mean  $N_{\text{ef}}m$  range: 1.184-2.887); however, Bayesian posterior  $N_{\text{ef}}m$  distributions peaked at the lower limit of resolution, with 95% confidence intervals including zero. Setting ancestral  $N_e$  equal to the  $N_{\text{ef}}$  estimates above, we ran separate simulations testing three hypotheses modeled as hypothetical population trees (see Additional file 1: Supplement S1): (1) a null, 'fragmented ancestor' model representing the expansion-contraction model, with a single size-constant ancestral population evolving through the Pleistocene before experiencing a 90% LGM reduction in  $N_e$  (range contraction), and subsequent colonization and diversification of local subpopulations (15-0 ka; summing to current  $N_e$ ) while simultaneously expanding from a source population (south Florida refugium located by



**Figure 7 Statistical parsimony network relationships among *cytb* and *RPS7* haplotypes (numbered).** Networks were defined in TCS based on a 95% parsimony criterion. Network circles indicate haplotypes scaled according to their frequency (smallest colored circles = 1 sample; scale applies to *cytb* network; *RPS7* haplotype frequencies are given in parentheses), lines represent 1 mutation step between haplotypes, and dotted lines enclose distinct network regions (numbered, with regions in parentheses). Colors represent haplotype identities, by SAMOVA-inferred population groups ( $K=4$ ) shown in Figure 1.

ecological niche modeling; Figure 3B); (2) a ‘vicariance-northeast colonization’ model with a Early Pleistocene ancestor split basally east-west of Apalachicola River/Mobile Bay and ACP populations arising spontaneously post-LGM (15-0 ka) from a Florida refuge (colonized from a Saint Johns River source); and (3) a ‘four-refugia’ vicariance model with SAMOVA population groups (Figure 1, results below) diverging from a Pleistocene ancestor. The number of deep coalescences for our ‘best’ *cytb* gene tree fit into the population trees for each hypothesis was 9 for the null model, 14 for vicariance-northeast colonization, and 27 for four-refugia. Results of both simulation sets failed to reject the expansion-contraction hypothesis in favor of alternative models ( $P < 0.001$ ), thus our gene tree is more consistent with a scenario of LGM population contraction and postglacial re-expansion than either of the alternative vicariance models.

## Discussion

Our results failed to reject most expansion-contraction model predictions (66.7% of tests; summarized in Additional file 1: Table S8). Indeed, several independent lines of evidence from geospatial and genetic data give concordant results consistent with this model. Here, we explore these results.

Ecological niche modeling suggests that *H. formosa* underwent range contraction to a single LGM refugium in the southwest Florida peninsula, an area that underwent less environmental-climatic change than the northern parts of the species range (Figure 3B). Paleoclimatic data, climate models and pollen records indicate southwest Florida maintained ~0-4°C lower temperatures, wetter tropical climates (+400-800 mm/yr annual precipitation) and mixed southern pine/hardwood forest to open woodland vegetation during the LGM [11,27,93], relative to today. By contrast, LGM conditions in northern Florida and the ‘mainland’ Gulf-Atlantic plains were colder/drier with extreme ~8-10°C drops in winter temperatures, 400 mm/yr lower annual precipitation, and patchy northern forest and grasslands vegetation [27,93,94]. Combined with these data, our results strongly suggest *H. formosa* were extirpated from their northern range and survived the LGM in bioclimatically suitable peninsular areas, areas that, in fact, were the only parts of eastern North America that did not experience significantly colder winter and summer LGM temperatures [94].

While one paleobotanical synthesis suggested this putative refugial area was nothing more than arid sand-dune scrub at the time [13], other data [11] indicate that at this time south Florida was semiarid, containing suitable freshwater stream and wetland/marsh habitats where cold-intolerant freshwater organisms persisted. This deduction is consistent with pollen data showing aquatic macrophytes such as *Brasenia*, *Isoëtes*, *Myriophyllum*, *Typha* and

*Sparganium* occurring within the Florida peninsula during the LGM [95].

Comparing the predicted LGM refuge of *H. formosa* with its much wider modern distribution (Figures 1 and 3C), we suggest that *H. formosa* rapidly recolonized most or all of its preceding range to the north (e.g. of the LIG, Figure 3A) during postglacial climatic ‘amelioration.’ After 16-15 ka, eastern North American rivers underwent transitions from braided to meandering channel types, during which wetter-than-modern early Holocene conditions may have increased wetlands and river discharge [96]. Conditions from this time onward may thus have favored expansion into parts of the species range, even though temperature and precipitation did not stabilize near their modern levels until somewhat later, ~9-6 ka [93]. Such northward post-glacial expansions were unlikely to have been slowed by Younger Dryas cooling 12-11 ka (i.e. potentially negative effects on aquatic habitats), which did not affect areas below ~38°N latitude ([96], refs. therein). As a corollary, we reject vicariance-northeast colonization-predicted range dynamics and our results also seem to run counter to rangewide population stasis and connectivity expected under the southern crossroads hypothesis [21]; apparently, smaller areas of the North American Coastal Plain (south Florida) may have experienced prolonged ecological stability than in Mediterranean Basin coastal plains.

Ecological niche modeling assumes that relationships between present distributions, abiotic factors, and ecological interactions remain more or less at equilibrium, that species physiological limits have changed less than the magnitude of changes in abiotic variables, and that biotic features of a species’ niche affect historical distribution less than large-scale changes in abiotic conditions [18,22,23]. We cannot evaluate all of these assumptions, nor can we exclude possible bias due to coastline effects i.e. major influence of the Bermuda high-pressure zone on coastal precipitation, which was among the variables of highest (>30%) predictive importance (Additional file 1: Tables S3-S4). Nonetheless, our niche-based modeling results provide meaningful insights into *H. formosa* biogeography. Our paleoclimatic models yield highly accurate predictions, e.g. AUCs >0.90 and our results are consistent with existing ecophysiological data. For example, modern *H. formosa* populations reach critical thermal minima at ~10°C [97] and live close to this cold-season limit in northern parts of their range. Coinciding with this value, 88% of the *H. formosa* localities (Additional file 1: Figure S2) had mean temperature of coldest quarter values  $\geq 10^{\circ}\text{C}$  ( $\pm 0.1^{\circ}\text{C}$ ), and this variable had sharply higher effects on MAXENT prediction above 10°C (JCB, unpublished data). Given cold-season temperatures across the species entire northern range declined to -8 to 4°C during the LGM [27,93], it is reasonable to infer that glacial-stage

extirpation occurred passively in these areas, with drought and colder winter temperatures reducing survival. In addition, the assumption of ecological niche conservatism on these spatial and temporal scales seems justified for this species. Differences in thermal tolerances among *H. formosa* populations from different parts of their range are minimal (~2 °C; [97]) compared to the temperature differences induced by climatic oscillations of the last glacial cycle above; thus differential physiological adaptation to temperature regimes likely has not had a confounding effect on our inferences. Our analyses of DNA sequence data and re-analyses of nuclear allozyme loci reinforce the paleoclimatic data to support expansion-contraction dynamics. In particular, the negative latitudinal cline suggests a glacial refugium in south Florida and northward recolonization [4,9,15]. Admixture in south Florida due to secondary contact of diverged lineages, i.e. following 'vicariance' expected under vicariance-northeast colonization, or recent dispersal mediated by episodic connections (e.g. flooding, axial-valley connections) can be ruled out for two reasons. First, there is a distinct lack of phylogenetic divergence and, second, seven allozyme alleles in south Florida are private alleles not shared with populations in other areas. These data are also inconsistent with southern crossroads dynamics supported elsewhere, in which no clinal diversity patterns were witnessed [21]. That our phylogeographical results also support expansion-contraction dynamics agreeing with the niche models also suggests that the assumption of ecological niche conservatism through time is essentially correct for *H. formosa*.

Of course, it is difficult to discount the possibility of 'microrefugia' [98] in small habitat pockets elsewhere. This is because our niche-based models predict responses to prevailing macroclimatic conditions, not microclimatic conditions below the data-layer grain ( $1 \text{ km}^2$ ) or within local habitats. If any secondary refugium existed, the most likely candidate area would seem to have been the Tallahassee area, where we also found one private allozyme allele.

Several of our results agree with previous studies [42,54]. Similar to Baer [42] and despite extreme distances of >2000 km between our collection localities and those in the allozyme dataset, we find *H. formosa* are characterized by shallow genetic divergences and a single population-lineage (e.g. mean mtDNA p-distance = 0.66%; Figures 6 and 7). Combined with high  $H_d$  but low  $\pi$  estimates, this suggests recent postglacial expansion with too limited time for recovery to allow accumulation of large sequence divergence [29], and/or high historical gene flow levels. Also congruent with [42], we infer no genetic barriers within the Atlantic Coastal Plain. This is consistent with Baer's 'northeast colonization'

hypothesis (ACP populations colonized following an Early Pleistocene high-sea stand following 1.5 Ma), which he argued was also supported by lower allozyme diversity and a lack of ACP isolation-by-distance [42] (see below). *Heterandria formosa* also contains four significantly differentiated mtDNA population groups subdivided outside of the putative refugium, or between the refugium and recolonized areas (Figures 1, 2 and 7). The latter break, inferred between the Suwannee and Hillsborough Rivers (i.e. SAMOVA groups 1 vs. 2-4; Figures 1 and 7), corresponds roughly to a north-to-south break between Gulf-draining Waccasassa and Withlacoochee Rivers described by Baer [42]. This break creates a clade of Louisiana plus south Florida samples in our phylogenetic and multivariate ordination analyses of the allozyme-frequency data, and Baer's (Figure 2; though our results are based on different genetic distances). This Suwannee-Hillsborough break thus appears robust to different data types and methods and is shallow in both studies (e.g. ~0.6-0.8% mtDNA divergence) implying recently limited gene flow between these rivers. Perhaps more importantly, the Suwannee-Hillsborough break falls (with other genetic barriers) outside the predicted LGM refugium but is not correlated with deep phylogenetic divergence. This is consistent with the action of density-dependent processes, such as priority effects e.g. [2,9], or allele surfing [61], operating during leading-edge recolonization. Given their patchy distribution [39,42] and propensity for explosive population growth [99], it is reasonable to assume that *H. formosa* quickly fill new habitats once they arrive and that when local patches become extinct they are recolonized from nearby populations. Thus if northward postglacial expansion occurred through small numbers of colonists, alleles might have surfed along a relatively quickly expanding wave front. Our niche-based models suggest postglacial genetic divergence between these rivers may have partly been maintained by lack of predicted suitable habitat between these rivers (Figure 3).

By contrast, we find several points of nuclear-mitochondrial discordances, indicating differences between studies. For example, allozyme and mtDNA results yield mixed support for the expansion-contraction prediction that isolation-by-distance should occur within the refugium, but not outside of it, due to a longer demographic timeline and geographically restricted dispersal within the refuge. Allozymic variation exhibits nearly rangewide isolation-by-distance in our study and Baer's [42]. Baer [42] also inferred populations were at or near drift-migration equilibrium, using a 2-dimensional stepping stone model with cross-peninsula gene flow. Global  $N_e m$  estimates were high (>4, range: 2.5-37.5), suggesting local *H. formosa* populations have not evolved independently but that gene flow has importantly shaped their evolution [42]; and in a second study,  $N_e m$  within and across

two peninsular Florida rivers ranged >3–400 [54]. In Baer [42], ACP populations were exceptional, forming a relatively homogeneous monophyletic group with no isolation-by-distance consistent with recent colonization, consistent with our mtDNA and allozyme findings. However, contrasting these patterns, we find: a lack of mtDNA isolation-by-distance across the range;  $N_e m$  suggesting zero on-going mtDNA gene flow (see Supplement S1); and ACP populations as indistinct from other regions (i.e. nonmonophyletic; Figures 2, 6 and 7). Incongruent with Baer's [42] allozyme results, the historical signal of the mtDNA genome also does not indicate east-to-west genetic differentiation between WCP and ACP regions (Figure 1).

Such nuclear-mitochondrial discordances are traditionally explained by several factors. The first and most common explanation is balancing selection (e.g. during genetic hitchhiking) mainly on allozymes [100], some of which (particularly metabolic protein loci) are known to evolve non-neutrally. However, this is rejected outright by the data: Lewontin-Krakauer tests on the allozymes [42,54] and multiple tests of mtDNA neutrality (e.g. Additional file 1: Table S5) reject the hypothesis of selection operating in either genome, as does the concordant lack of deep phylogenetic structure within either marker-class. A second explanation is that these discordances arise from the contrasting modes of transmission, evolution, and resolution of these marker-types; this could plausibly explain a lack of isolation-by-distance in the mtDNA, but isolation-by-distance in the allozymes. Compared with haploid, maternally transmitted mtDNA, allozymes are diploid and evolve roughly 10× slower with ~4-fold larger  $N_e$  and coalescence times (independent of mutation rate), and higher gene-flow rates [29,101,102]. As a consequence, mtDNA provide a higher-resolution view of population structure, e.g. at finer spatial scales (supported by our data), and more likely reflect homogenizing effects of gene flow hence lack of isolation-by-distance. By contrast, due to lower resolution, allozymes may display isolation-by-distance in northern areas even when expansion occurred from a single southern refugium, if postglacial expansion(s) did not involve leptokurtic (long-distance) dispersal e.g. reviewed in [2,4,9]. Regarding discordant migration patterns, allozyme  $N_e m$  estimates and  $N_e$  estimates seem preferable in the case of *H. formosa* (mtDNA typically drastically underestimate  $N_e$  in high-gene flow species [29]). Still, their large coalescence times mean allozyme-inferred  $N_e m$  may reflect averages across transmission routes over four times the history of mtDNA, rather than current processes. Thus, we interpret our results and Baer's [42] as indicating that historical gene flow (high  $N_e m$ ) among populations nearing migration-drift equilibrium, and large  $N_e$ , has influenced *H. formosa* phyleogeographical structuring. This could partly explain the low observed genetic divergences, but can we reconcile this

scenario with the isolation-by-distance findings, taken at face value? Taking such a scenario as compatible with the expansion-contraction model assumes most population structure has arisen since the LGM and that sufficient time has passed for *H. formosa* allozyme loci to reach equilibrium. This is supported by the data if we assume  $Nm = 10$  (similar to [42], or higher [54]) and that populations must have achieved migration-drift equilibrium since the LGM, or  $G = 57,000$  generations ( $19 \text{ ka}^*T$ ). Given our  $N_{ef}$  ( $\sim 6 \times 10^5$ – $3 \times 10^6$ ) and  $\sim 1.2$  Ma intraspecific  $t_{\text{MRCA}}$  estimates, the number of generations required to approach equilibrium is recent,  $\sim 60,829$  generations, and since the LGM based on the equation  $G = 1/(2m + 1/(2N_e))$  [103]. Thus, the above interpretation seems reasonable in light of population genetics theory; equilibrium could have arisen even faster in this species, as postglacial recolonization and population expansion can rapidly generate population structuring [61].

Another factor potentially explaining our results is sex-biased migration. *Heterandria formosa* exhibit strongly female-biased sex ratios due to higher male mortality rates, as demonstrated by predation trials [104] and otolith ring counts (JT, personal observation). In light of this, female dispersal/gene flow should outpace male dispersal even assuming equal dispersal propensities and rates between sexes. Thus, aside from differences in marker resolution, female-biased dispersal has likely also contributed importantly to nuclear-mitochondrial discordance above, specifically lack of mtDNA isolation-by-distance. This is because, under female-based dispersal, females carry both mtDNA and male/female nuclear genes out of populations while males are effectively philopatric (geographically restricted), thus we expect matrilineal alleles to more likely homogenize among localities while nuclear genomes remain site-dependent [29]; this would result in a relatively weaker matrilineal signal of isolation-by-distance. Unfortunately, we cannot reliably empirically evaluate effects of female-biased survival and dispersal on our other genetic results given comparable metrics influenced by geographical distance (e.g. linearized  $F_{ST}$ ) are not estimable from the mtDNA and allozyme-frequency data, and low *RPS7*-gene sampling restricts possible mtDNA-nDNA comparisons. Last, consistent with the observation that deep coalescences appear to influence the *H. formosa* mtDNA/gene tree (e.g. Figure 6B), patterns of discordance between markers herein may to some extent reflect historical processes of incomplete lineage sorting in addition to the homogenizing effects of gene flow [significant when  $Nm > 4$  (reviewed in [29]), as in *H. formosa*]. In future studies, sampling multiple, unlinked nuclear loci will be necessary to tease apart the relative contributions of these two opposing forces within *H. formosa* using coalescent-based models.

In addition to the geospatial and genetic analyses above, our phylogenetic and historical-demographic results also met predictions of the expansion-contraction model. While our inferences regarding ancestral derived/populations were limited by reduced south Florida sampling, phylogenetic and spatial patterns of ancestral/derived populations (alleles) between the phylogeny and network analyses congruently follow expectations derived from the model (e.g. Figure 4, Additional file 1: Table S8), and star-like patterns in the network suggest a history of population expansions. Our argument that *H. formosa* fits the expansion-contraction model is also reinforced by mismatch analyses and mtDNA genetic equilibrium tests (Fu's  $F_S$ ,  $R_2$ ), which fail to reject demographic expansion models (e.g. for SAMOVA-inferred population groups; see text; Table 3, S5; Figure 5) and thus capture genetic signatures of predicted historical changes in  $N_e$  containing information about the record and timing of expansion [61,70,92]. The 95% confidence intervals on expansion parameter  $\tau$  suggest the onset of expansions has overlapped the LGM in most population groups and regions analyzed, and lower 95% confidence intervals extending into the present may be indications that expansion is on-going or ceased only recently, e.g. in north Florida (Table 3; Figure 5). Because balancing selection seems not to have been at play (as per above), low-frequency peaks in our mismatch distributions likely reflect secondary Holocene expansions rather than selection (Figure 5).

These demographic results generally agree with genetic patterns in European taxa (despite possibly earlier onset of *H. formosa* expansions) that are inferred to have experienced postglacial population expansions e.g. [4,17]; however, in our case, integrating niche-based modeling and genetic analyses provides an additional check that it is reasonable to conclude *H. formosa* expansion has been both spatial and demographic (Figure 3; Table 3). Interestingly, our mtDNA genetic equilibrium tests only point to recent population bottlenecking within WCP samples (Table 3) spanning extensive wetlands and swamps of the Mississippi River Delta, a major outlet of postglacial meltwater flow and alluvial deposition. This could reflect a small population surviving the LGM in one or more WCP microrefugia not identified in our paleoclimatic models; however, no evidence exists that a population-lineage experienced drift in isolation in the WCP since the LGM (e.g. Figure 6B). We thus interpret this bottleneck as a result of founder events during postglacial recolonization.

Although the distribution of *H. formosa* spans nearly the entire lowland Coastal Plain (Figure 1), its genetic patterns do not completely coincide with those described in several other species with overlapping present-day distributions. At least 20 species, including 50-67% of freshwater fishes and turtles examined to date, display

pronounced east-west Apalachicola phylogeographical [29-32] and distributional [33] breaks. In four freshwater fishes—bowfin (*Amia calva*) and three sunfishes (*Lepomis gulosus*, *Lepomis microlophus*, and *Lepomis punctatus*)—this pattern has been linked to vicariant isolation in upper reaches of different drainage basins that were reduced in size by eustatic high-stands 50-80 m ASL of the Pliocene-Early Pleistocene interglacials [105,106]. Thus, major phylogenetic divergence(s) should be expected in this region in as-yet unsampled freshwater taxa. Remarkably, however, neither patterns of multilocus phylogenetic structuring or mtDNA-population structuring, nor a coalescent simulation perspective on neutral mtDNA evolution, support such a pattern within *H. formosa*. Variation at protein loci in *H. formosa* also departs from this vicariance model, revealing shallowly diverged and geographically overlapping population groups/lineages without deep genetic breaks at predicted barriers. While the common pattern of Gulf drainages containing Atlantic-coast haplotypes (but not vice versa) is present in the allozymes (Figure 2), neither mtDNA nor allozyme data support clear east-west Apalachicola splits.

Our estimated  $t_{MRCA}$  suggests intraspecific diversification of *H. formosa* has coincided with Pleistocene diversification within *A. calva*, which is codistributed with *H. formosa* in the Coastal Plain but ranges much more widely across eastern North America (including the Mississippi and St. Lawrence River basins). This indicates that recent isolation within constraints of Pleistocene-Holocene (rather than earlier) drainage geomorphology, e.g. drainage divides, has influenced genetic variation of both *H. formosa* (i.e. genetic distinctiveness of mtDNA at AMOVAs structured by Florida drainages) and *A. calva*. However, additional studies will be necessary to test whether these species diversification has actually been synchronous, and to evaluate finer-scale patterns relative to drainage basins. It would also be interesting to examine whether Coastal Plain freshwater fish species that are presently codistributed with *H. formosa* responded to Late Pleistocene disruptions in climate and sea-level with similar range-shifting responses, possibly involving refugia; however, the only species in which effects of Pleistocene expansion-contraction dynamics have been rigorously tested is *H. formosa* (this study). Here, the eastern/Atlantic intraspecific lineages (Florida peninsula-Atlantic seaboard) present in many freshwater taxa will likely present good candidates for testing for range expansion-contraction dynamics similar to *H. formosa*. For example, the eastern/Atlantic lineages recovered by Bermingham and Avise [32] exhibit <2-4% genetic divergences, indicating Pleistocene  $t_{MRCA}$ s (assuming the standard 2%/Myr clock for vertebrate mtDNA [101]), and phylogroups or private alleles unique to the Florida peninsula. While *ad hoc* explanations that interspecific

differences in the position of the Apalachicola break (particularly geographical distributions of the eastern lineages) resulted from unique histories of dispersal and range-shifting in response to Pleistocene climatic fluctuations have been advanced [29,31,32], the actual pattern of such dynamics remains an unanswered question. Additional studies similar to ours of previously studied fish taxa, e.g. in [32], as well as understudied aquatic plants, insects, freshwater fishes (e.g. *Fundulus seminolis*, *Jordanella floridae*, etc.), and crayfishes from the Florida peninsula will therefore permit testing the generality of the scenario of historical range-shift and population dynamics uncovered within *H. formosa*, as well as general predictions of the expansion contraction model [15], within the Gulf-Atlantic Coastal Plain freshwater biota.

## Conclusions

Understanding historical responses of species to the climatic and sea-level fluctuations of the Late Pleistocene has emerged as a major goal of ecology, evolution and biogeography [2,4,15,29]. We find the evolutionary history of *Heterandria formosa* indicates southward LGM range contraction and large-scale postglacial expansion, as predicted under the Pleistocene expansion-contraction model [15]. Species-specific expansion-contraction dynamics may therefore have played a more general role in shaping the evolutionary history of Coastal Plain biota than previously thought. By adding to considerable variation surrounding common biogeographical themes recovered among even codistributed species in this region, our results bolster growing appreciation for the complexity of phylogeographical patterning in North America's southern refugia. Our study demonstrates the benefits of taking an integrative approach to biogeographical hypotheses generation and testing (e.g. more realistic hypotheses testing through niche modeling and statistical phylogeography) and showcases how similar approaches can be used by biologists in the future to gain further insights into the evolutionary history of the Gulf-Atlantic Coastal Plain biota, as well as the important role of this region as a dynamic refuge during the last glacial cycle. Indeed, a better understanding of the tempo and mode of species responses to the Pleistocene glacial-interglacial cycles will likely aid predicting and managing ecosystem responses to present-day and future climate change [6].

## Availability of supporting data

Sequences are deposited in GenBank (accession numbers: KF632895-KF633114, *cytb*, and KF633115-KF633133, *RPS7*). Sequence alignments and phylogenetic trees are available in the TreeBASE repository: <http://purl.org/phylo/treebase/phylows/study/TB2:S14718>.

## Additional file

**Additional file 1: Data supplement.** Contains supplementary **Tables (S1-S9)** and **Figures (S1-S3)** and an additional description of methods and results (Supplement S1).

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

MS and JCB conceived of the project. JCB directed project design and coordination, conducted molecular laboratory work, performed genetic and ecological niche modeling analyses, archived data in online repositories, and wrote the manuscript. MS conducted field and molecular laboratory work, discussed analyses, and revised the manuscript. MLLV and JT made field collections and revised the manuscript. JBJ discussed molecular laboratory work and helped revise and structure the manuscript. All authors have read and approved the final manuscript.

## Acknowledgements

We are grateful to M F. Breitman, E. Castro, and J. A. Wooten for helpful comments on this manuscript, and to A. L. Almendra for critical assistance with ecological niche modeling analyses. We thank P. J. Unmack for discussions and for graciously donating GIS data and samples from Hillsborough and St. Johns Rivers, Florida. Sampling was conducted under permits from Louisiana, Mississippi, Alabama, Georgia, Florida, and South Carolina issued to MS, JT and The University of Alabama Ichthyological Collection (UAIC). We thank A. Contreras-Balderas and G. Hubbard for assistance with field collections. Research was funded by Brigham Young University, including a Mentoring Environment Grant to JBJ that covered molecular sequencing costs; The University of Alabama; and The Florida State University.

## Author details

<sup>1</sup>Department of Biology, Brigham Young University, 401 WIDB, Provo, UT 84602, USA. <sup>2</sup>Monte L. Bean Life Science Museum, Brigham Young University, Provo, UT 84602, USA. <sup>3</sup>Department of Biological Science, Biodiversity & Systematics, The University of Alabama, Box 870345, Tuscaloosa, AL 35487, USA. <sup>4</sup>Department of Biological Science, The Florida State University, Tallahassee, FL 32306, USA. <sup>5</sup>Laboratorio de Ictiología, Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León, Monterrey, Nuevo León, México.

Received: 2 September 2012 Accepted: 13 September 2013

Published: 9 October 2013

## References

1. Gibbard PL, Head MJ, Walker MJC, and the Subcommission on Quaternary Stratigraphy: **Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma.** *J Quaternary Sci* 2009, 25:96–102.
2. Hewitt G: **The genetic legacy of the Quaternary ice ages.** *Nature* 2000, 405(6789):907–913.
3. Hewitt GM: **Ice ages: their impact on species distributions and evolution.** In *Evolution on Planet Earth*. Edited by Rothschild LJ, Lister AM. London, UK: Academic Press; 2003:339–361.
4. Hewitt GM: **Genetic consequences of climatic oscillations in the Quaternary.** *Philos T Roy Soc B* 2004, 359(1442):183–195.
5. EPICA community members: **Eight glacial cycles from an Antarctic ice core.** *Nature* 2004, 429:623–628.
6. Gates DM: **Climate Change and its Biological Consequences.** Sunderland, MA: Sinauer Associates; 1993.
7. Lomolino M, Riddle B, Whittaker R, Brown J: *Biogeography*. 4th edition. Sunderland, MA: Sinauer Associates; 2010.
8. Hof C, Levinsky I, Araújo MB, Rahbek C: **Rethinking species' ability to cope with rapid climate change.** *Glob Change Biol* 2011, 17:2987–2990.
9. Hewitt GM: **Some genetic consequences of ice ages, and their role in divergence and speciation.** *Biol J Linn Soc* 1996, 58(3):247–276.
10. Bennett KD, Tzedakis PC, Willis KJ: **Quaternary refugia of north European trees.** *J Biogeogr* 1991, 18(1):103–115.

11. Jackson ST, Webb RS, Anderson KH, Overpeck JT, Webb T III, Williams JW, Hansen BCS: Vegetation and environment in Eastern North America during the last glacial maximum. *Quat Sci Rev* 2000, 19:489–508.
12. Davis MB: Quaternary history and the stability of forest communities. In *Forest Succession*. Edited by West DC, Shugart HH, Botkin DB. New York, NY: Springer-Verlag; 1981:132–177.
13. Delcourt PA, Delcourt HR: *Vegetation maps for eastern North America: 40,000 yr. BP to the present*. In *Geobotany II*. Edited by Romans RC. New York, NY: Plenum; 1981:123–165.
14. FAUNMAP Working Group: Spatial response of mammals to late quaternary environmental fluctuations. *Science* 1996, 272(5268):1601–1606.
15. Provan J, Bennett KD: Phylogeographic insights into cryptic glacial refugia. *Trends Ecol Evol* 2008, 23(10):564–571.
16. Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF: Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* 1998, 7(4):453–464.
17. Hewitt GM: Post-glacial recolonization of European biota. *Biol J Linnean Soc* 1999, 68(1–2):87–112.
18. Waltari E, Hijmans RJ, Peterson AT, Nyari AS, Perkins SL, Guralnick RP: Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS One* 2007, 2(7):e563.
19. Gómez A, Lunt DH: Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In *Phylogeography in Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity*. Edited by Weiss S, Ferrand N. Dordrecht: Springer-Verlag; 2007:155–188.
20. Canestrelli D, Cimmaruta R, Nascetti G: Population genetic structure and diversity of the Apennine endemic stream frog, *Rana italica*—insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Mol Ecol* 2008, 17:3856–3872.
21. Porretta D, Canestrelli D, Urbanelli S, Bellini R, Schaffner F, Petric D, Nascetti G: Southern crossroads of the Western Palearctic during the Late Pleistocene and their imprints on current patterns of genetic diversity: insights from the mosquito *Aedes caspius*. *J Biogeogr* 2011, 38:20–30.
22. Svensson JC, Normand S, Kageyama M: Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *J Ecol* 2008, 96:1117–1127.
23. Flogaard C, Normand S, Skov F, Svensson JC: Ice age distributions of European small mammals: insights from species distribution modelling. *J Biogeogr* 2009, 36:1152–1163.
24. Dépraz A, Cordellier M, Haussler J, Pfenninger M: Postglacial recolonization at a snail's pace (*Trochulus villosus*): confronting competing refugia hypotheses using model selection. *Mol Ecol* 2008, 17:2449–2462.
25. Otto-Biesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A, CAPE Last Interglacial Project members: Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* 2006, 311:1751–1753.
26. Chen JH, Curran HA, White B, Wasserburg GJ: Precise chronology of the last interglacial period:  $^{234}\text{U}$ - $^{230}\text{Th}$  data from fossil coral reefs in the Bahamas. *Geol Soc Am Bull* 1991, 103:82–97.
27. Shin SI, Liu Z, Otto-Biesner B, Brady EC, Kutzbach JE, Harrison SP: A simulation of the last glacial maximum climate using the NCAR-CCSM. *Clim Dynam* 2003, 20(2–3):127–151.
28. Lambbeck K, Chappell J: Sea level change through the last glacial cycle. *Science* 2001, 292:679–686.
29. Avise JC: *Phylogeography: the History and Formation of Species*. Cambridge, MA: Harvard University Press; 2000.
30. Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS: Comparative phylogeography of unglaciated eastern North America. *Mol Ecol* 2006, 15(14):4261–4293.
31. Avise JC: Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos* 1992, 63(1):62–76.
32. Bermingham E, Avise JC: Molecular Zoogeography of Fresh-Water Fishes in the Southeastern United-States. *Genetics* 1986, 113(4):939–965.
33. Swift CC, Gilbert CR, Bortone SA, Burgess GH, Yerger RW: Zoogeography of the fresh water fishes of the south-eastern United States: Savannah River to Lake Ponchartrain. In *Zoogeography of North American Freshwater Fishes*. Edited by Hocutt CH, Wiley EO. New York, NY: Wiley; 1985:213–265.
34. Ronquist F: Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst Biol* 1997, 46(1):195–203.
35. Swenson NG, Howard DJ: Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *Am Nat* 2005, 166(5):581–591.
36. Edwards CE, Soltis DE, Soltis PS: Molecular phylogeny of *Conradina* and other scrub mints (Lamiaceae) from the southeastern USA: evidence for hybridization in Pleistocene refugia? *Syst Bot* 2006, 31:191–205.
37. Edwards CE, Soltis DE, Soltis PS: Using patterns of genetic structure based on microsatellite loci to test hypotheses of current hybridization, ancient hybridization and incomplete lineage sorting in *Conradina* (Lamiaceae). *Mol Ecol* 2008, 17:5157–5174.
38. Parks CR, Wendel JF, Sewell MM, Qiu Y-L: The significance of allozyme variation and introgression in the *Liriodendron tulipifera* complex (Magnoliaceae). *Am J Bot* 1994, 81:878–889.
39. Martin FD: *Heterandria formosa* Agassiz. In *Atlas of North American Freshwater Fishes*. Edited by Lee DS, Gilbert CR, Hocutt CH, Jenkins RE, McAllister DE, Stauffer JR Jr. Raleigh, NC: North Carolina State Museum of Natural History; 1980:547.
40. Leips J, Travis J: The comparative expression of life-history traits and its relationship to the numerical dynamics of four populations of the least killifish. *J Anim Ecol* 1999, 68(3):595–616.
41. Schrader M, Travis J, Fuller RC: Do density-driven mating system differences explain reproductive incompatibilities between populations of a placental fish? *Mol Ecol* 2011, 20:4140–4151.
42. Baer CE: Species-wide population structure in a southeastern U.S. freshwater fish, *Heterandria formosa*: gene flow and biogeography. *Evolution* 1998, 52(1):183–193.
43. Knowles LL, Maddison WP: Statistical phylogeography. *Mol Ecol* 2002, 11(12):2623–2635.
44. Nielsen R, Beaumont MA: Statistical inferences in phylogeography. *Mol Ecol* 2009, 18(6):1034–1047.
45. Knowles LL: Statistical phylogeography. *Annu Rev Ecol Evol Syst* 2009, 40:593–612.
46. Carstens BC, Richards CL: Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* 2007, 61(6):1439–1454.
47. Richards CL, Carstens BC, Knowles LL: Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *J Biogeogr* 2007, 34(11):1833–1845.
48. Pluzhnikov A, Donnelly P: Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics* 1996, 144:1247–1262.
49. Felsenstein J: Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Mol Biol Evol* 2006, 23:691–700.
50. Hrbek T, Seckinger J, Meyer A: A phylogenetic and biogeographic perspective on the evolution of poeciliid fishes. *Mol Phylogenet Evol* 2007, 43(3):986–998.
51. Unmack PJ, Bagley JC, Adams M, Hammer MP, Johnson JB: Molecular phylogeny and phylogeography of the Australian freshwater fish genus *Galaxiella*, with an emphasis on dwarf galaxias (*G. pusilla*). *PLoS One* 2012, 7(6):e38433.
52. Chow S, Takeyama H: Intron length variation observed in the creatine kinase and ribosomal protein genes of the swordfish *Xiphias gladius*. *Fish Sci* 1998, 64:397–402.
53. Librado P, Rozas J: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, 25(11):1451–1452.
54. Baer CF: Population structure in a south-eastern US freshwater fish, *Heterandria formosa*. II. Gene flow and biogeography within the St. Johns River drainage. *Heredity* 1998, 81:404–411.
55. Phillips SJ, Anderson RP, Schapire RE: Maximum entropy modeling of species geographic distributions. *Ecol Model* 2006, 190(3–4):231–259.
56. Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JMC, Peterson AT, Phillips SJ, Richardson KS, Schachter-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmerman NE: Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 2006, 29(2):129–151.
57. Stockwell DRB, Peterson AT: Effects of sample size on accuracy of species distribution models. *Ecol Model* 2002, 148(1):1–13.
58. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A: Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 2005, 25(15):1965–1978.

59. Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB, Kiehl JT, Large WG, McKenna DS, Santer BD, Smith RD: **The Community Climate System Model version 3 (CCSM3)**. *J Clim* 2006, **19**:2122–2143.
60. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods**. *Mol Biol Evol* 2011, **28**(10):2731–2739.
61. Excoffier L, Foll M, Petit RJ: **Genetic consequences of range expansions**. *Annu Rev Ecol Evol Syst* 2009, **40**:481–501.
62. Monmonier MS: **Maximum-difference barriers: an alternative numerical regionalization method**. *Geogr Anal* 1973, **3**:245–261.
63. Manni FE, Guerard E, Heyer E: **Geographical patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm"**. *Hum Biol* 2004, **76**:173–190.
64. Excoffier L, Lischer HEL: **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows**. *Mol Ecol Resour* 2010, **10**(3):564–567.
65. Hood GM: **PopTools, Version 3.0.3**; 2012. Available at <http://www.cse.csiro.au/poptools>.
66. Dupanloup I, Schneider S, Excoffier L: **A simulated annealing approach to define the genetic structure of populations**. *Mol Ecol* 2002, **11**(12):2571–2581.
67. Peakall R, Smouse PE: **GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research**. *Mol Ecol Notes* 2006, **6**:288–295.
68. Mantel N: **The detection of disease clustering and a generalized regression approach**. *Cancer Res* 1967, **27**:209–220.
69. Wakeley J: **The effects of subdivision on the genetic divergence of populations and species**. *Evolution* 2000, **54**(4):1092–1101.
70. Harpending HC: **Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution**. *Hum Biol* 1994, **66**(4):591–600.
71. Hammer Ø, Harper DAT, Ryan PD: **PAST: paleontological statistics software package for education and data analysis**. *Palaeontol Electron* 2001, **4**(1):1–9.
72. McDonald JH, Kreitman M: **Adaptive protein evolution at the Adh locus in Drosophila**. *Nature* 1991, **351**(6328):652–654.
73. Fay JC, Wu CI: **Hitchhiking under positive Darwinian selection**. *Genetics* 2000, **155**(3):1405–1413.
74. Zwickl DJ: **Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion**. PhD thesis. Austin: The University of Texas at Austin; 2006.
75. Minin V, Abdo Z, Joyce P, Sullivan J: **Performance-based selection of likelihood models for phylogeny estimation**. *Syst Biol* 2003, **52**(5):674–683.
76. Maddison WP, Knowles LL: **Inferring phylogeny despite incomplete lineage sorting**. *Syst Biol* 2006, **55**(1):21–30.
77. Knowles LL, Carstens BC: **Estimating a geographically explicit model of population divergence**. *Evolution* 2007, **61**(3):477–493.
78. Swofford D: **PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods), version 4.0**. Sunderland, MA: Sinauer Associates; 2001.
79. Clement M, Posada D, Crandall KA: **TCS: a computer program to estimate gene genealogies**. *Mol Ecol* 2000, **9**(10):1657–1659.
80. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees**. *BMC Evol Biol* 2007, **7**:214.
81. Waters JM, Burridge CP: **Extreme intraspecific mitochondrial DNA sequence divergence in *Galaxias maculatus* (Osteichthys: Galaxiidae), one of the world's most widespread freshwater fish**. *Mol Phylogenet Evol* 1999, **11**(1):1–12.
82. Burridge CP, Craw D, Fletcher D, Waters JM: **Geological dates and molecular rates: fish DNA sheds light on time dependency**. *Mol Biol Evol* 2008, **25**(4):624–633.
83. Hamilton A: **Phylogeny of *Limia* (Teleostei: Poeciliidae) based on NADH dehydrogenase subunit 2 sequences**. *Mol Phylogenet Evol* 2001, **19**(2):277–289.
84. Iturralde-Vinent MA: **Meso-Cenozoic Caribbean paleogeography: implications for the historical biogeography of the region**. *Int Geol Rev* 2006, **48**:791–827.
85. Doadrio I, Perea S, Alcaraz L, Hernandez N: **Molecular phylogeny and biogeography of the Cuban genus *Girardinus* Poey, 1854 and relationships within the tribe Girardinini (Actinopterygii, Poeciliidae)**. *Mol Phylogenet Evol* 2009, **50**:16–30.
86. Agorreta A, Dominguez-Dominguez O, Reina RG, Miranda R, Bermingham E, Doadrio I: **Phylogenetic relationships and biogeography of *Pseudoxiphophorus* (Teleostei: Poeciliidae) based on mitochondrial and nuclear genes**. *Mol Phylogenet Evol* 2013, **66**:80–90.
87. Pascual R, Bond M, Vučetić MG: **El Subgrupo Santa Bárbara (Grupo Salta) y sus vertebrados. Cronología, paleoambientes y paleobiogeografía**. San Luis, Argentina: VIII Congreso Geológico Argentino, Actas III; 1981:746–758.
88. Maddison WP, Maddison DR: **Mesquite: a modular system for evolutionary analysis**. Version 2.73; 2010. Available from <http://mesquiteproject.org>.
89. Beerli P, Felsenstein J: **Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach**. *P Natl Acad Sci USA* 2001, **98**(8):4563–4568.
90. Shepard DB, Burbink FT: **Phylogeographic and demographic effects of Pleistocene climatic fluctuations in a montane salamander**. *Plethodon fourchensis*. *Mol Ecol* 2009, **18**(10):2243–2262.
91. Maddison WP: **Gene trees in species trees**. *Syst Biol* 1997, **46**(3):523–536.
92. Rogers AR, Harpending H: **Population growth makes waves in the distribution of pairwise genetic differences**. *Mol Biol Evol* 1992, **9**(3):552–569.
93. Prentice LC, Bartlein PJ, Webb T III: **Vegetation and climate change in eastern North America since the last glacial maximum**. *Ecology* 1991, **72**:2038–2056.
94. Watts WA: **Late-Quaternary vegetation history at White Pond on the Inner Coastal Plain of South Carolina**. *Quat Res* 1980, **13**:187–199.
95. Sawada M, Viau AE, Gajewski K: **The biogeography of aquatic macrophytes in North America since the last glacial maximum**. *J Biogeogr* 2003, **30**:999–1017.
96. Leigh DS: **Terminal Pleistocene braided to meandering transition in rivers of the Southeastern USA**. *Catena* 2006, **66**:155–160.
97. Baer CF, Travis J: **Direct and correlated responses to artificial selection on acute thermal stress tolerance in a livebearing fish**. *Evolution* 2000, **54**(1):238–244.
98. Rull V: **Microrefugia**. *J Biogeogr* 2009, **36**:481–484.
99. Leips J, Travis J, Rodd FH: **Genetic influences on experimental population dynamics of the least killifish**. *Ecol Monogr* 2000, **70**:289–309.
100. Karl SA, Avise JC: **Balancing selection at allozyme loci in oysters: implications for nuclear RFLPs**. *Science* 1992, **256**:100–102.
101. Brown WM: **Evolution of animal mitochondrial DNA**. In *Evolution of Genes and Proteins*. Edited by Nei M, Koehn RK. Sunderland, MA: Sinauer Associates; 1983:62–88.
102. Moore WS: **Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees**. *Evolution* 1995, **51**:627–629.
103. Crow JF, Aoki K: **Group selection for a polygenic behavioral trait: estimating the degree of population subdivision**. *Proc Natl Acad Sci USA* 1984, **81**:6073–6077.
104. Richardson JML, Gunzburger MS, Travis J: **Variation in predation pressure as a mechanism underlying differences in numerical abundance between populations of the poeciliid fish *Heterandria formosa***. *Oecologia* 2006, **147**(4):596–605.
105. Cronin TM, Szabo BJ, Ager TA, Hazel JE, Owens JP: **Quaternary climates and sea levels of the U.S. Atlantic Coastal Plain**. *Science* 1981, **211**(4479):233–240.
106. Riggs SR: **Paleoceanographic model of Neogene phosphorite deposition, US Atlantic continental margin**. *Science* 1983, **223**:123–131.

doi:10.1186/1471-2148-13-223

**Cite this article as:** Bagley et al.: Paleoclimatic modeling and phylogeography of least killifish, *Heterandria formosa*: insights into Pleistocene expansion-contraction dynamics and evolutionary history of North American Coastal Plain freshwater biota. *BMC Evolutionary Biology* 2013 **13**:223.