



2007-12-05

Phylogeography of the Livebearer *Xenophallus umbratilis* (Teleostei: Poeciliidae): Glacial Cycles and Sea Level Change Predict Diversification of a Freshwater Tropical Fish

Carissa Poole Jones

Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Biology Commons](#)

BYU ScholarsArchive Citation

Jones, Carissa Poole, "Phylogeography of the Livebearer *Xenophallus umbratilis* (Teleostei: Poeciliidae): Glacial Cycles and Sea Level Change Predict Diversification of a Freshwater Tropical Fish" (2007). *All Theses and Dissertations*. 1565.

<https://scholarsarchive.byu.edu/etd/1565>

This Selected Project is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

**Phylogeography of the Livebearer *Xenophallus umbratilis* (Teleostei:
Poeciliidae): Glacial Cycles and Sea Level Change Predict
Diversification of a Freshwater Tropical Fish**

by

Carissa P. Jones

A selected project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biology

Brigham Young University

November 2007

Copyright © 2007 Carissa P. Jones

All Right Reserved

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a selected project submitted by

Carissa P. Jones

This selected project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date

Jerald B. Johnson, Chair

Date

Byron J. Adams

Date

Mark C. Belk

Date

Jack W. Sites, Jr.

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the selected project of Carissa P. Jones in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Jerald B. Johnson
Chair, Graduate Committee

Accepted for the Department

Date

Keith A. Crandall
Department Chair

Accepted for the College

Date

Rodney J. Brown
Dean, College of Life Sciences

ABSTRACT

**Phylogeography of the Livebearer *Xenophallus umbratilis* (Teleostei:
Poeciliidae): Glacial Cycles and Sea Level Change Predict
Diversification of a Freshwater Tropical Fish**

Carissa P. Jones

Department of Biology

Master of Science

The biogeography of Central America is considered a classical case study in understanding the impact of vicariant events on patterns of biotic dispersal. While many biogeographers have focused on community composition and geographic limits of species at broad scales across Central America, much less work has focused on post-colonization diversification patterns at finer scales. The livebearing freshwater fish *Xenophallus umbratilis* presents an ideal system for determining the impact of recent earth history events on biodiversity in northern Costa Rica. Here, we test the hypothesis that marine inundation of the San Carlos and Northern Limón basins during the Miocene has caused genetic fragmentation among *X. umbratilis* populations, despite contemporary freshwater connections across this region. To test this idea, we collected mitochondrial

(*cytb*) and nuclear (*Xmrk-2*) DNA sequence data from up to 162 individuals taken from 27 localities across northern Costa Rica. We employed a variety of analytical approaches, including: maximum parsimony (MP) and maximum likelihood (ML), analysis of molecular variance (AMOVA), nested clade phylogeographic analysis (NCPA), and demographic analysis of population size through time. We found four major clades within *X. umbratilis*, each geographically isolated with no shared haplotypes across drainages. Oddly, clades that occupy adjacent drainages are not always sister taxa in the phylogeny, suggesting that colonization in this species is more complex than a simple model of isolation by distance. All our results are consistent with the hypothesis that changes in sea level associated with glacial eustatic cycles have had an important effect in shaping diversification patterns in this species.

ACKNOWLEDGEMENTS

Jerald B. Johnson, Mark C. Belk, J. Jaime Zúñiga-Vega, Joshua Rasmussen, and Jared B. Lee helped with field collections and Peter J. Unmack helped with molecular laboratory procedures. We are grateful to Jack W. Sites, Jr., Byron J. Adams, Mark C. Belk, and Keith A. Crandall for valuable comments and insights, and to Jerald B. Johnson for invaluable assistance reviewing and revising every aspect of this manuscript. This research was funded in part by Brigham Young University through a Mentoring Environment Grant and a Graduate Research Fellowship Award, and through stipend support from the U.S. National Science Foundation (OISE-PIRE 0530267).

Table of Contents

Title page	i
Copyright page	ii
Graduate Committee Approval	iii
Final Reading Approval and Acceptance	iv
Abstract	v
Acknowledgments	vii
List of Tables	2
List of Figures	3
Introduction	4
Materials and Methods	6
<i>Geological Background</i>	6
<i>Geographic Sampling</i>	8
<i>DNA Extraction, Amplification, and Sequencing</i>	9
<i>Phylogenetic Analysis</i>	10
<i>Phylogeographic and Molecular Clock Analyses</i>	12
<i>Demographic and Population Diversity Analyses</i>	16
Results	17
<i>Phylogenetic Analysis</i>	17
<i>Phylogeographic and Molecular Clock Analyses</i>	18
<i>Demographic and Population Diversity Analyses</i>	20
Discussion	21
<i>Phylogeographic diversification: what is the impact of earth history events?</i>	22
<i>Future work and broader implications</i>	25
<i>Conclusions</i>	27
References	29
Tables	37
Figure Legends	44
Figures	46

List of Tables

Table 1. Collection site localities and tissue voucher numbers.....	37
Table 2. Analysis of Molecular Variance (AMOVA) table.....	39
Table 3. Nested Clade Phylogeographic Analysis (NCPA) table.....	40
Table 4. Results of the Bayesian analysis of coalescent time estimates using <i>Beast</i>	41
Table 5. Population diversity indices and mtDNA equilibrium test statistics.....	42
Table 6. Haplotype frequency chart.....	43

List of Figures

Figure Legends	44
Figure 1. Geographic map with <i>Xenophallus umbratilis</i> drainages and elevational information.....	46
Figure 2. Geographic sampling map and <i>Xenophallus umbratilis</i> distribution.....	47
Figure 3. Haplotype network and nesting structure.....	48
Figure 4. Strict consensus phylogeny generated by ML/MP analysis.....	49
Figure 5. Time to most recent common ancestor (T_{MRCA}) estimates plotted against known climatic and geologic events.....	50
Figure 6. Bayesian skyline plot showing effective population size over time.....	51

Introduction

The biogeography of Central America is viewed as a classical system for understanding the impact of vicariant events on continental patterns of biotic dispersal and community composition (Bermingham and Martin 1998, Knowlton 1993). While biogeographic patterns of South America and Central America have been well documented, both at the species level and higher (e.g. Rosen 1976; Wake & Lynch 1976; Duellman 1979; Gentry 1982; Savage 1982; Cadle 1985; Crother *et al.* 1992), few studies have focused on diversification within species to examine Central American biogeography at finer scales (but see Zamudio 1997; Garcia-Paris *et al.* 1997; Lowe *et al.* 2003). This lack of fine-scale examination is surprising because to explain current patterns of biodiversity in Central America it is essential to understand both patterns of colonization and the mechanisms responsible for evolutionary diversification following colonization.

Northern Costa Rica is an important region to examine the effects of recent vicariant events on patterns of diversification both among and within species. Much of Costa Rica is geologically young, and most Costa Rican species are also young or have recently dispersed there (Coates & Obando 1996; Gentry 1982). This makes it more likely that biogeographic approaches will be able to reveal the evolutionary history of species within the region.

Moreover, the spatial distribution of freshwater fishes permits strong inferences about the biotic and geological history of the study region because freshwater fishes depend on direct connections between drainage basins to disperse (Myers 1938), and because the history of drainage basin interconnections reflects the geological history of the landscape (Lundberg *et al.* 1998).

The geological history of northern Costa Rica over the last five millions years is not well understood, despite its importance in completing the land bridge between North and South America. What we do know is that much of this area occurs at a low elevation that has experienced repeated periods of marine flooding associated with global changes in sea level (Fig. 1). Hence, high sea levels could cause local extirpation of lowland freshwater populations while leaving higher elevations as isolated refugia. Upon subsequent return to lower sea levels, isolated refugia populations would be able to re-colonize the lowlands, potentially leaving the signature of allopatric fragmentation with deep levels of divergence. In addition, western Costa Rica possesses areas of intense volcanism while eastern Costa Rica is characterized by low-elevation areas with little elevation change from one hydrologic basin to the next (Fig. 1). These physiographic properties should lead to different patterns of divergence among lineages within species. For example, intense volcanism and uplift could physically isolate populations and/or cause extirpation. Similarly, small differences in elevation between drainages potentially could create patterns of dispersal through river capture due to differential erosion, river anastomosis due to reduced sea levels during glacial maxima, and/or pulses of freshwater during the rainy season that connect rivers along the coast (Myers 1938, 1966; Loftin 1965; Miller 1966; Bermingham & Martin 1998; Unmack 2001). Taxa that range over areas impacted by both changes in sea level and volcanic activity could potentially have strong phylogeographic signal that could then be interpreted in a biogeographic context.

In this study, we examine geographic patterns of molecular variation in the livebearing fish *Xenophallus umbratilis*. This species is ideally suited to examine the effect of local earth history events on diversification patterns following the emergence of the Central American land bridge. *Xenophallus umbratilis* has a range restricted primarily to northern Costa Rica (Fig. 1) spanning much of the area that would have been susceptible to flooding. Moreover, it currently occurs at elevations ranging from 35-600m in both lowland rivers and headland streams. It occurs on both sides of the volcanic cordilleras that form the continental divide (Bussing 1998). In addition, this species is limited to freshwater habitats so should closely track the geological history of the region. In this study we focus on three objectives. First, we identify major phylogenetic clades within *X. umbratilis* and look for congruence between these clades and several earth history events, including volcanic orogeny and isolation caused by changes in sea level. Second, we describe patterns of genetic structuring among populations within and between these clades, and changes in population size through space and time. Finally, given the taxonomic uniqueness of this species, we explore the implications of our findings on freshwater conservation efforts in northern Costa Rica.

Materials and Methods

Geological Background

We hypothesize that current patterns of genetic structuring in *Xenophallus umbratilis* are closely linked to the geological and climatic history of the northern Costa Rica area. Here we briefly summarize the major earth history events over the past five million years of this

area, and based on these events, generate specific predictions about patterns of genetic structuring we expect to in this species. Five million years ago Northern Costa Rica was composed of a coastal shelf interspersed by a few islands (Coates & Obando 1996). The remainder of the Costa Rican land mass emerged from the ocean over the next two million years and was coincident with volcanic activity that gave rise to the continental divide. This created a continuous land bridge across modern Panama through Costa Rica. However, a low elevation area remained in northern Costa Rica extending from modern Lake Nicaragua to the Caribbean Sea (Fig. 1; Coates & Obando 1996). Consequently, throughout the Pliocene and Pleistocene, changes in ocean levels associated with glacial eustatic cycles resulted in repeated periods of marine incursions across what is now northern Costa Rica and southern Nicaragua affecting both the northern Limon and San Carlos drainage basins (see Fig. 1; Haq *et al.* 1987; Potts & Behrensmeyer 1992; Gallup *et al.* 1994).

These two historical factors—marine incursions due to changes in sea level and volcanic formation of the central cordillera—are most likely to have shaped current patterns of genetic structuring in *X. umbratilis*. Although this species has been described as a secondary freshwater fish (meaning they are believed to derive from seawater fish that adapted to populate freshwater, Myers 1949), no tests have been done to validate or refute its putative salinity tolerance. If *X. umbratilis* is limited by salt water, or limited by ecological factors associated with salt water (i.e. predators, etc.), marine incursion events would result in isolation among drainages manifest by patterns of phylogenetic divergence associated with range-wide fragmentation. Hence, we predict an association between rises in sea level and phylogenetic divergence among *X. umbratilis* clades. We also predict that we will see higher

levels of population structuring among drainages that flow directly into a saline marine environment when compared to *X. umbratilis* populations that are interconnected by freshwater. In other words, *Xenophallus umbratilis* populations belonging to drainages that flow into the Caribbean Sea may have experienced more isolation than their inland counterparts that have freshwater connections. Finally, we test the hypothesis that the rise of the cordillera three million years ago has resulted in genetic fragmentation of populations that occur on either side of the continental divide.

Geographic Sampling

We sampled *Xenophallus umbratilis* from 27 localities in eight drainages throughout northern Costa Rica (Fig. 2). Our sampling spanned the entire range of this species in Costa Rica. *Xenophallus umbratilis* has also been reported in Lake Nicaragua and likely occupies the northern tributaries of the Rio San Juan, the river that delineates the border between Costa Rica and Nicaragua (Fig. 1; Bussing 1998). Our sampling includes all southern tributaries of the Rio San Juan, so although northern tributary samples from Nicaragua could expand our study, our collections appear to be sufficient to inform the phylogeographic hypotheses addressed here, and to provide a clear picture of *X. umbratilis* intraspecific diversity. Samples included in this study come from ethanol preserved field collections made in 1998, 2006, and 2007.

DNA Extraction, Amplification, and Sequencing

We collected DNA sequence data from up to eight individuals at each locality (see Table 1). These data were gathered using the following protocols. We extracted DNA from ethanol-preserved muscle tissue using Qiagen DNeasy Tissue Kits and DNeasy 96 (QIAamp DNA minikit; Qiagen, Hilden, Germany). All exemplars are currently housed in the BYU Monte L. Bean Life Science Museum (Table 1). We then amplified and sequenced two DNA regions from each individual: one mitochondrial (cytochrome *b* or *cytb*, an electron transport gene) and one nuclear (*Xmrk-2*, a tyrosene kinase receptor gene). We used bidirectional sequencing to amplify each target region. The entire *cytb* gene (1140 bp) was amplified using one forward primer (Glu31 5'- tgrcttgaaaaaccaccgttg-3') and one of two reverse primers (Poec15931 5'-ccctcgactttcggtttacaag-3'; Poec15940 5'-gactttaacctccgactttcg-3').

PCR amplification and sequencing conditions for *cytb* were as follows: after an initial 5 minutes at 94°C, reactions were amplified for 10 cycles at 94°C for 30 seconds, 53°C for 45 seconds, and 72°C for 90 seconds and for 24 cycles at 94°C for 30 seconds, 58°C for 45 seconds, and 72°C for 90 seconds. *Xmrk-2* was amplified using the primer set mrkA (5'-gctgagctctgatgacgacg-3') and mrkB (5'-ggacaggaggtgtacagtgcg-3') (Woolcock *et al.*, 1994). PCR amplification and sequencing conditions for *Xmrk-2* were as follows: after an initial 5 minutes at 94°C, reactions were amplified for 10 cycles at 94°C for 30 seconds, 53°C for 45 seconds, and 72°C for 90 seconds and for 24 cycles at 94°C for 30 seconds, 58°C for 45 seconds, and 72°C for 90 seconds.

We edited and aligned forward and reverse sequences for each gene fragment using Sequencher 3.1.1 (Gene Codes Corp., Inc., Ann Arbor, MI) and Bioedit 7.0.5 (Tom Hall, Ibis Therapeutics, Carlsbad, CA). The gene-coding nuclear region contained two indels, so to aid alignment we first translated the gene-coding nuclear region and then aligned the sequences by eye. Since *Xmrk-2* nuclear DNA sequences were much less variable than *cytb* sequences, we obtained one or two individuals for every clade found in a preliminary analysis on the mitochondrial gene in order to assess the congruence of deeper relationships within *X. umbratilis*. All sequences are archived in GenBank (accession numbers xxxxxx – xxxxxx).

Phylogenetic Analysis

To reconstruct evolutionary relationships within *X. umbratilis*, we used both maximum parsimony (MP) and maximum likelihood (ML) analyses. MP analysis was carried out using heuristic searches and TBR method implemented in PAUP* 4.10b (Swofford 1999), with 1000 random sequence additions to discover the most-parsimonious trees. We constructed strict consensus trees of 40 resulting trees that tied for the best score (tree length = 724). To investigate support for each node, we used bootstrapping with 1000 pseudoreplicates and 10 random sequence additions per pseudoreplicate (Felsenstein 1985). We considered nodes with bootstrap values above 70 to be strongly supported (Hillis & Bull 1993). Of the 1140 base pairs included in mtDNA phylogenetic analyses, 308 variable characters were parsimony informative and 102 variable characters were parsimony uninformative. For the nuclear dataset, we manually coded an indel 18 bp long (from bp 322-339) as one character (scored as present or absent) to give the indel no more weight than the point mutations found

elsewhere in the dataset. Of the 568 base pairs included in nDNA phylogenetic analyses, 31 variable characters were parsimony informative and 21 variable characters were parsimony uninformative. MP analysis resulted in 272 equally parsimonious trees with a length of 63.

To carry out ML analysis required identifying an appropriate model of molecular evolution. We did this using the program Modeltest 3.06 (Posada & Crandall 1998) which evaluates a set of 56 alternative substitution models. We used the Akaike Information Criterion (AIC) (Akaike 1974) and identified GTR + Γ as the best-fitting model for our mtDNA and HKY85 + I as the best-fitting model for our nDNA. Using the parameters described by Modeltest, we performed 10 random sequence additions with TBR branch swapping. ML analysis of mtDNA yielded one best tree with a $-\ln$ score of 4795.72, while ML analysis of nDNA resulted in one best tree with a $-\ln$ score of 1047.72. We used bootstrapping with 1000 pseudoreplicates and 10 random sequence additions per pseudoreplicate to investigate support for each node. We carried out all ML analysis in PAUP* 4.10b (Swofford 1999) except the nDNA bootstrapping. Bootstrapping on the nDNA dataset continually stalled in PAUP* and therefore we carried out 100 bootstrap pseudoreplicates in the program GARLI (Zwickl 2006). Since the implementation of this model is exactly equivalent to that in PAUP*, likelihood scores obtained by each program are directly comparable.

For phylogenetic analysis of the nuclear DNA, we utilized two outgroups (*Priapichthys annectens*, *Brachyrhaphis rhabdophora*). For the more complete mtDNA analyses, we included five outgroups (*Phallichthys amates*, *Brachyrhaphis rhabdophora*, *Priapichthys annectens*, *Alfaro cultratus*, and *Neoheterandria tridentiger*). Based on Hrbek *et al*'s (2007)

phylogeny of poeciliid fishes, we fixed *Alfaro cultratus* as the outgroup in mtDNA analyses and allowed the other taxa to float in both MP and ML analyses. Not only did this permit us to examine the monophyly of *X. umbratilis*, it also allowed us to evaluate Hrbek's assessment of *Priapichthys annectens* as the putative sister *X. umbratilis* rather than a *Neoheterandria* species, the genus in which *X. umbratilis* had previously been placed (Meek 1912). We did not combine the two partitions for analysis; instead, we explored congruence between the two by comparing the phylogenies generated by both parsimony and likelihood methods for each gene partition. Phylogenetic analysis was conducted using both *Xmrk-2* and *cytb* data, while all other analyses were performed on the mtDNA *cytb* data alone.

Phylogeographic and Molecular Clock Analyses

To explore the distribution and abundance of *X. umbratilis* populations through space and time, we employed both phylogeographic and molecular clock approaches. Our phylogeographic analysis focused on five *a priori* hypotheses that we modeled using an analysis of molecular variance (AMOVA) (Excoffier et al. 1992). These models reflect different ways of partitioning populations to test the amount of variation explained between geographic regions (Fig. 1). Cases where high amounts of variation can be explained between groups suggest an important historic barrier to gene flow exists. Our five models are as follows (see Fig. 1 for geographic reference): (1) each of the eight drainages as distinct groups (to test overall population structuring by drainage); (2) drainages that empty into the Caribbean versus those that drain to the Rio San Juan (to test the hypothesis that drainages flowing into the freshwater Rio San Juan are more connected to each other than to drainages

that flow into Caribbean Sea); (3) each of the two Caribbean drainages as distinct groups (to measure levels of gene flow through the ocean); (4) each of the five Rio San Juan drainages as distinct groups (to test the degree of isolation among drainages historically separated by marine incursions); and (5) the Tempisque drainage versus the Zapote drainage (to test the effect of the continental divide on gene flow). All of these analyses were executed in the program ARLEQUIN (Schneider et al. 2000) with significance assessed using 10^4 random permutations of the dataset.

We also used a Nested Clade Phylogeographic Analysis (NCPA) to explore patterns of phylogeographic structuring in our data (Templeton et al. 1995; Templeton 1998, 2001, 2004). Whereas the AMOVA approach allowed us to consider several hypotheses *a priori*, NCPA provides inferences about the history of populations based on the distribution of the observed data themselves. Hence, these two methods combined provide a stronger level of inference than either approach could reveal by itself. We performed this analysis with the program TCS 1.13 using predictions from the coalescent theory to calculate root probabilities and to infer relative haplotype ages (Clement *et al.* 2000). Two clades (Clade 1 and 2) were divergent enough from each other and from the two remaining clades (Clades 3 and 4) that TCS was unable to connect them at the 95% confidence level (17 steps) and it was uninformative to connect these through statistical parsimony, resulting in three networks (Fig. 3). Inferences at higher nesting levels were therefore made using the connections indicated by phylogenetic analysis using MP and ML (described above). We resolved a single ambiguous connection (a loop) in our network following the guidelines outlined in Crandall et al. (1994).

We nested the haplotype network by hand into increasingly inclusive clades following the rules described in Templeton et al. (1987) and Templeton and Sing (1993). Haplotypes fully nested within the network created a six-level hierarchy that included 19 one-step clades, nine two-step clades, five three-step clades, three four-step clades, two five-step clades, and the total cladogram. To calculate geographic distances between collecting sites we used river distances rather than direct linear distances; this provides a biologically more realistic measure for aquatic species. The NCPA calculates the clade distance, which measures the geographical spread of a clade, and the nested clade distance, which measures how a clade is geographically distributed relative to other clades in the same higher-level nesting category. We calculated both distances and tested them with 10^6 random permutations (Templeton & Sing 1993) implemented in the program Geodis 2.0 (Posada *et al.* 2000). We interpreted the results with the 2005 Inference Key for the Nested Haplotype Tree Analysis of Geographical Distance [<http://darwin.uvigo.es/software/geodis.html>].

In order to assess the relevance of geologic and climatic events to *X. umbratilis* divergences, we first established whether or not a strict molecular clock could be imposed on the mtDNA.

Using Modeltest 3.06, we identified the model of evolution for mtDNA (TrN + Γ). To establish whether the same substitution rates could be assumed across all *X. umbratilis* lineages, we generated likelihood scores with and without a molecular clock enforced in PAUP* 4.10b (Swofford 1999). Twice the difference between the two scores was tested against a null hypothesis (Chi-squared). Degrees of freedom were calculated as $n - 2$, where n is the number of taxa, following Huelsenbeck & Rannala (1997). We found that *X.*

umbratilis lineage divergence rates were consistent with a clock-like pattern (df=34, $\chi^2=24.464$, $P>0.25$). Hence, we made several inferences assuming a molecular clock.

To test haplotype divergence times against known earth history events, we used the program *Beast* (Drummond & Rambaut 2003), which uses a Bayesian MCMC algorithm, to estimate the time to the most recent ancestor (T_{MRCA}) for all *X. umbratilis* included in this study, as well as the time to most recent ancestor with its putative sister species, *P. annectens*. For these analyses we used the model of evolution identified by Modeltest 3.06 for the mtDNA data and a constant coalescent model for the prior estimate of population growth. In addition, we estimated the T_{MRCA} for divergences between clades revealed by phylogenetic analysis. A number of initial short runs were performed to optimize scale factors, and these optimized factors were then used to conduct multiple longer runs with chain length of 10^6 ; parameters were sampled every 1000 iterations after an initial burn-in of 10^6 . Under these conditions, the MCMC process performed well, achieving good stationarity and yielding large effective sample sizes (> 200) for estimates of time to most recent common ancestor. Since the mtDNA exhibited clock-like properties and we had no well-defined calibration points, we used a molecular clock as implemented for the related genus *Poeciliopsis* (1-2% per million years; Mateos *et al.* 2002; Reznick, Mateos & Springer, 2002; Hrbek, Seckinger, and Meyer 2007) to estimate temporal boundaries for divergence events within this lineage.

Demographic and Population Diversity Analyses

In order to assess the impact of historical isolation on genetic diversity within drainages, we employed a set of methods that require several assumptions, including a large, constant population size, no selection, infinite-sites, and no recombination. Since our dataset is mitochondrial DNA and does not appear to be saturated, no recombination and infinite-sites can safely be assumed. To test the other assumptions for our data set, we performed Tajima's D test (Tajima 1989), Fu and Li's D^* (Fu & Li 1993), Fu and Li's F^* test (Fu & Li 1993), and Fu's F_s test (Fu 1997) to examine whether samples from different localities are at genetic equilibrium with respect to mtDNA, considering that a significant deviation from mtDNA genetic equilibrium is presumably a result of recent population expansion or bottleneck in situations where no selective advantage among haplotypes exists (Rand 1996).

In addition, to evaluate the assumption of constant population size and to continue to explore the abundance of *X. umbratilis* through time, we explored the historical demography of *X. umbratilis* with a Bayesian skyline coalescent model using the program *Beast* (Drummond & Rambaut 2003) to examine changes in effective population size over time. For this analysis we used the model of evolution identified by Modeltest 3.06 for the mtDNA data and a constant coalescent model for the prior estimate of population growth. In addition, we estimated the T_{MRCA} for divergences between clades revealed by phylogenetic analysis. A number of initial short runs were performed to optimize scale factors, and these optimized factors were then used to conduct multiple longer runs with chain length of 10^6 ; parameters were sampled every 1000 iterations after an initial burn-in of 10^6 . Under these conditions, the

MCMC process performed well, achieving good stationarity and yielding large effective sample sizes (> 200).

To evaluate conservation priority, we measured mtDNA genetic diversity within major clades identified through phylogenetic analysis as the number of haplotypes, haplotype diversity (Hd), nucleotide diversity (both π and θ_w), and the number of polymorphic sites using the program DnaSP 4.0 (Rozas & Rozas 1999) available at <http://www.ub.es/dnasp>. Not only were estimates of nucleotide diversity useful for assessing conservation priority, comparing the present day genetic diversity (π) to the historical estimates of genetic diversity (θ_w) allowed us to assess the results of historical demography analyses aimed at reconstructing the effective population size over time of *X. umbratilis*.

Results

Phylogenetic Analyses

Xenophallus umbratilis is composed of four deeply divergent clades, with as much as 4.5% uncorrected sequence divergence among haplotypes (Fig. 4). As predicted, two of these clades (Clades 1 & 2) are geographically located in high elevation regions that would be potential refugia for *X. umbratilis* during high sea levels (Fig. 1). Coincidentally, these two clades are also defined by the most basal breaks in our phylogeny. Clade 1 is made up entirely of populations from the San Carlos drainage and is geographically centered in the range of *X. umbratilis*. Clade 2 consists mostly of individuals from the Parismina drainage,

which flows into the Caribbean and is at the southeastern limit of the distribution of *X. umbratilis*. This clade is also comprised of a population from the upper Sarapiquí drainage. Clade 3 is composed of populations from the lower Sarapiquí and Pocosal drainages which flank either side of the San Carlos drainage. Clade 4 presents the oddest geographic affinities of all groups. Not only is it made up of populations from three geographically juxtaposed drainages, the Tempisque and Zapote drainages (Clade 4c) and the Frio drainage (Clade 4a), it also includes populations in the Tortuguero drainage (Clade 4b) which is found at the complete opposite side of the range of *X. umbratilis*. Finally, populations on either side of the continental divide (Tempisque and Zapote drainages) are found in the same subclade (Clade 4c).

Phylogenetic relationships among major and minor clades were almost completely unambiguous. Both MP and ML strict consensus trees yielded completely congruent relationships within *X. umbratilis*. Nodal bootstrap values were high for mtDNA, although phylogenetic methods were not always able to resolve relationships among closely related haplotypes. Nodes for the nDNA phylogeny were largely unresolved and much less informative than the mtDNA phylogeny. However, those nodes supported by bootstrap values above 50% were completely consistent with the mtDNA results.

Phylogeographic and Molecular Clock Analyses

For each of the five *a priori* phylogeographic hypotheses we examined (see Methods), AMOVA recovered highly significant population structure F_{st} values (Table 2) indicating

little evidence for gene flow. Moreover, we found that four of the five models had the most genetic variation among the designated groups, with the remaining model showing most of the variation among populations within groups. We summarize each of these models below. First, the drainage by drainage analysis revealed that 93.72% of the total genetic variation was partitioned among the eight drainages indicating a high level of genetic partitioning across drainages. Second, we predicted that populations from drainages that empty into freshwater would be genetically distinct from drainages that empty into the sea. This was not the case. Comparing the San Juan (freshwater) drainages to the Caribbean (marine) drainages revealed most variation was partitioned among populations, within groups (76.16%) rather than between the freshwater and marine drainage groups (22.67%). Our third and fourth models were designed to test the hypothesis that drainages without freshwater connections would show lower levels of gene flow than drainages with freshwater connections. We found that this was also not true. Indeed, variation within each of those partitions is almost equivalent—the freshwater drainages had 93.47% of the variation among groups, while the marine drainages had 95.78% of the variation among groups. Finally, the fifth model evaluated the extent to which the continental divide is a barrier to gene flow. We found that 70.54% of the variation was found among groups on either side of the divide, with 27.40% of the variation within populations, while only 2.06% of the variation was among populations, within groups, suggesting the continental divide is an important barrier to gene flow. The NCPA corroborated the findings from the AMOVA. NCPA revealed significant geographic structuring among *X. umbratilis* haplotypes at varying geographic scales across the distribution of this species (Table 3). In particular, there was a strong pattern of allopatric fragmentation across the deepest nesting levels.

We estimated coalescent times for all nodes found in phylogenetic analysis in order to assess the relevance of known earth history events such as eustatic sea-level changes to divergences within lineages of *X. umbratilis* (Table 4). With a mutation rate of 1-2%, the estimate of T_{MRCA} was 4.392 mya (95% credibility interval between 7.059 and 2.299 mya). Thus, patterns of haplotypes diversity strongly suggest a most recent common ancestor during the late Miocene/early Pliocene most likely prior to both the closure of the Central American land bridge and the uplift of the Central Cordilleras of Costa Rica. Using *Beast*, we estimated the T_{MRCA} for *X. umbratilis* and its putative sister *P. annectens* at 18.94 mya (95% credibility interval between 28.98 and 10.38 mya). We mapped the T_{MRCA} estimates for each of the nodes found in phylogenetic analysis onto a phylogeny against scaled geological and climactic events (Fig. 5) and found that mean T_{MRCA} estimates for all major divergences within *X. umbratilis* coincided closely with interglacial periods.

Demographic and Population Diversity Analysis

The Bayesian skyline analysis revealed a recent bottleneck event between 0.3-0.4 mya (Fig. 6). Fu and Li's F , Fu's F_s , and Tajima's D -statistic for the overall system were significantly positive, while for most clades the values were not significant (Table 5). These tests show that the species is at mutation-migration-drift genetic disequilibrium with respect to mtDNA alleles. Clade 1 has significantly negative test results, indicating either recent population expansion or positive (diversifying) selection. The positive test results for the overall system could indicate either a recent population bottleneck or negative (balancing) selection.

Considerable diversity exists within each clade. Estimates of present day genetic diversity (π) are much smaller than the historical estimates of genetic diversity (θ_w ; Table 5), indicative of a recent reduction in effective population size. Clade 4 has the highest π and θ_w estimates (Table 5), yet is geographically fragmented across four drainages and is made up of three sub-clades, so we would expect more diversity relative to the other three clades. Relative to its sample size, Clade 3 has high estimates of diversity. Genetic analysis revealed that among the 162 individuals included, 36 haplotypes were observed (Table 6). No haplotypes were shared between drainages, although within drainages some populations had haplotypes in common. Fully half of the haplotypes were found in only one individual.

Discussion

Our results point to a strong influence of earth history events on diversification patterns in *X. umbratilis*. This species is composed of four distinct and well-supported phylogenetic clades that cleanly map to distinct river basins. Moreover, the timing of divergence among these clades is coincident with known earth history events, including periods of marine inundation of lowland areas in northern Costa Rica and the uplift of the Costa Rican central cordillera associated with volcanic activity. Yet, some unusual patterns also exist in our data, including genetic affinities between non-adjacent river basins, suggesting that patterns of colonization or local extinction are more complex than a simple isolation by distance scenario. Below we explore in detail patterns of congruence and discord between our initial hypotheses and our observed results.

Phylogeographic diversification: what is the impact of earth history events?

We initiated this study with several hypotheses focused on the impact of vicariant events on current patterns of diversification in *X. umbratilis*. Here we consider each of these hypotheses in turn. First, almost all of our findings are consistent with the hypothesis that marine transgressions during interglacial maxima in the late Pleistocene could have isolated *X. umbratilis* in high-elevation refugia. Deep divergences revealed in our phylogenetic analyses point to widespread vicariant events resulting in fragmentation of populations. Each clade in our phylogeny is geographically isolated and genetically distinct, indicating historical events have strongly influenced genetic patterns within *X. umbratilis*. Indeed, the two most divergent clades (Clades 1 & 2; namely, the San Carlos, Parismina, and upper Sarapiquí drainages) are found in the high-elevation areas we predicted would be above the highest interglacial sea levels. This pattern is consistent with the idea that rising sea levels during eustatic cycles isolated *X. umbratilis* in highland refugia. Conversely, populations at low elevations (namely, those populations within the San Carlos and Northern Limon basins, consisting of the Pocosol, Tortuguero, and lower Sarapiquí drainages) have more recently diverged and may have experienced extirpation followed by subsequent recolonization. Interestingly, Clades 4b and 4c (the Tempisque, Zapote, and Frio drainages) occupy high elevation areas, but are relatively young. One explanation for this is that they occupy areas along the highly active volcanic arc that characterizes Costa Rica's continental divide, and may have experienced extirpation due to volcanism and subsequent recolonization. While we were limited to less precise divergence time estimates since we had no well-defined

calibration points for the area, the coincidence of every T_{MRCA} estimate within *X. umbratilis* lineages with a putative marine incursion is highly consistent with the eustatic cycle refugia hypothesis. In addition, NCPA ascribed much of the genetic diversity within *X. umbratilis* to allopatric fragmentation. The signature of allopatric fragmentation is spread over several different nesting levels, indicating that there may have been a series of vicariant events leading to allopatry rather than an isolated occurrence. This evidence for a series of allopatric speciation events over time is consistent with expectations of the eustatic cycle refugia hypothesis. In addition, the Bayesian skyline plot indicates that *X. umbratilis* populations experienced a bottleneck, a hypothesis that is confirmed by nucleotide diversity and neutrality tests. The bottleneck occurred at about 0.3-0.4 million years ago, which is concordant with the time estimates for the most recent haplotype diversification within *X. umbratilis* lineages. The timing of this bottleneck also coincides with the marine incursion that occurred during the Mindel-Riss interglacial period (0.2-0.4 mya), although other factors such as the increased uplift and volcanism that characterized the Pleistocene (Coates & Obando 1995) could have played a role as well. It is likely that the effective population size of *X. umbratilis* was negatively impacted by other interglacial periods, but the most recent bottleneck was severe enough that it potentially obscured any previous signal.

Second, contrary to our expectations, we found that movement of *X. umbratilis* has not only been limited by marine barriers, it also seems to have been limited by large freshwater connections through the Rio San Juan (Figure 2). Our AMOVA results indicate that drainages separated by freshwater show similar levels of genetic isolation as drainages separated by saltwater, suggesting that the Rio San Juan may be an effective barrier against

migration for *X. umbratilis*. It is curious to note that within the Tempisque drainage, NCPA of four populations within a few kilometers of each other revealed significant isolation by distance. It is evident that *X. umbratilis* is highly selective of its environment and prefers small, vegetated or rocky streams (Bussing 1998, personal observation), but it is also possible that *X. umbratilis* individuals are not prone to intentionally migrate long distances, even within their native drainages. Indeed, the AMOVA analysis showed that each drainage is essentially unique, a pattern we would normally only see between completely unconnected drainages. If *X. umbratilis* individuals do indeed avoid migration, then this in combination with isolation due to sea level fluctuations could explain much of the spatial patterning evident in the phylogeny. In other words, marine inundation events would have caused extirpation of lowland populations and isolation of highland populations. Then with subsequent recessions in sea levels recolonization of lowland environments would have occurred slowly and perhaps not at all through larger freshwater habitats. The current distribution of this fish suggests that some ecological barriers in the lower reaches of drainages might currently exist (see Figure 2). Some populations of *X. umbratilis* co-occur with the major poeciliid predator (*Parachromis dovii*), so predation alone may not be a limiting ecological factor.

Finally, we found that the rise of the continental divide (3-2 Mya, Collins *et al.* 1995) was not an important vicariant event that caused a deep split among *X. umbratilis* lineages. Instead, our phylogenetic reconstruction and NCPA both indicate that the Zapote/Tempisque divide is a relatively young split. In fact, the T_{MRCA} estimate for Clade 4c is more recent than one million years ago. Yet, even though the Tempisque and Zapote groups have only

recently diverged, our AMOVA analysis revealed that almost all the variation is either among groups or among populations, indicating that the continental divide, while only recently crossed, has been an effective barrier to migration. The most likely method of dispersal across the continental divide from the Zapote to Tempisque drainage is that of a headwater river capture event caused by local volcanic uplift after the continental divide had already risen. The hypothesis of a headwater river capture event is consistent with Smith and Bermingham's (2005) findings that this area is marked by less species turnover across the divide than further south along the continental divide.

Future work and broader implications

Our findings point to several lines of research that should yield interesting results. First, some unusual patterns in our phylogeographic data deserve additional attention. For example, we found that haplotypes of fish from locality 27 (found in the Tempisque drainage) clearly cluster with fish from the Lake Arenal Basin (in Clade 1) rather than with fish from the upper Tempisque drainage (in Clade 4). This anomaly is best explained by a man-made tunnel between Lake Arenal (San Carlos drainage) and Rio Santa Rosa (lower eastern portion of the Tempisque drainage) which redirected the flow of Lake Arenal from the San Carlos drainage (toward the Caribbean) to the Tempisque drainage (toward the Pacific). Our study provides the first genetic evidence that the tunnel built in the 1970's has resulted in the potential mixing of fishes between these basins. Another interesting relationship is found among the eastern *X. umbratilis* populations. Although upper and lower Sarapiquí populations are in the same drainage, upper Sarapiquí populations are more closely

related to populations found in the Parismina drainage (2-4 mutational steps, Clade 2), which drains into the Caribbean, than they are to lower Sarapiquí populations (20+ mutational steps, Clade 3). Given the relatively flat topography of eastern Costa Rica and the geographic proximity of the headland streams in the Sarapiquí and Parismina drainages, the most probable explanation is that there has been a fairly recent river capture from the Parismina drainage into the Sarapiquí drainage. Indeed, even though eight individuals were sampled from the upper Sarapiquí population, only one haplotype was present, which could indicate a recent introduction there. Since this introduction is probably much older than the introduction of population 27 into the Tempisque drainage just discussed, it would be interesting to further examine the population genetics of these two systems. More extensive sampling is necessary to discern whether the recent introduction of *X. umbratilis* into the lower Tempisque drainage and/or the upper Sarapiquí drainage has led to any measurable admixture with the natural upper Tempisque populations or the lower Sarapiquí populations. The introduction of previously isolated populations into a different drainage may provide an ideal system for future studies to discover more about the biology of *X. umbratilis*, specifically exploring the propensity of this species to migrate.

Besides perhaps providing an ideal system for studying progressive admixture of previously genetically distinct populations, the phylogeographic hypotheses developed in this study of *X. umbratilis* are easily testable across other freshwater lineages and indeed, should also be evaluated in order to reliably discern recent events in the geological history of the Central American landscape. In particular, species that share a similar range throughout northern Costa Rica (e.g. *Brachyrhaphis holdridgei*) would be ideal for evaluating the

phylogeographic hypothesis of marine inundation resulting in higher elevation refugia for freshwater species.

Our findings also have important implications with the respect to conservation of this species. Phylogenetic analysis of *Xenophallus umbratilis* reveals that while it is monophyletic, there is little resolution among outgroups with *cytb* data and we are unable to support or refute *Priapichthys annectens* as the closest sister species. However, our results do support Hrbek *et al*'s (2007) finding that *Xenophallus umbratilis* does not belong in the genus *Neoheterandria*. Hence, *Xenophallus* is therefore a monotypic genus endemic to Costa Rica, with its sole species *X. umbratilis* possessing high levels of variation despite its highly specialized habitat preferences (Bussing 1998, personal observation). As such, this fish is a prime candidate for conservation consideration. The four major clades found in phylogenetic analysis were assessed for conservation priority by measuring population diversity parameters. Considerable diversity exists within each clade, and each clade is distinct. Conservation priority should therefore be spread over all four clades, which geographically signifies the entire Costa Rican range of *X. umbratilis*. Conservation measures are not only important to preserve *X. umbratilis* diversity in this area, but to preserve our ability to test phylogeographic hypotheses across other taxa. For example, our findings indicate that *X. umbratilis* populations were historically fragmented and to a large degree have remained isolated since then. Considering the almost complete isolation of populations and the evidence for recent diversification, *X. umbratilis* therefore provides a unique system for examining local adaptation. Ecological conditions vary across the range of *X. umbratilis*, making it highly likely that independent evolution has and will continue to occur.

Conclusions

The most striking result from this study is the link between late Pleistocene changes in sea level and clade diversification in *X. umbratilis*. Our findings indicate that *X. umbratilis* populations were historically fragmented and to a large degree have remained isolated. Moreover, we find strong evidence that this species has experienced a substantial recent bottleneck, but is apparently in a period of demographic recovery. Diversity and population size appear to be increasing, and the past 0.3 million years seem to have been characterized by favorable conditions for *X. umbratilis*. Hence, we show that the closure of the land bridge in northern Costa Rica has been important not only as a conduit for faunal exchange between North and South America, but also for its direct impact on diversification patterns of freshwater species that occupy this unique part of the world.

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.
- Anderson EP, Pringle CM, Rojas M (2002) Transforming tropical rivers: an environmental perspective on hydropower development in Costa Rica. *Aquatic Conservation*, 16, 679-693.
- Anderson S (2007) 2005 Minerals Yearbook: Central America. US Geologic Survey, 6.1-6.19.
- Bermingham E, Martin AP (1998) Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology*, 7, 499–517.
- Bermingham E, Avise JC (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics*, 113, 939-965.
- Bussing WA (1976) Geographic distribution of the San Juan ichthyofauna of Central America with remarks on its origin and ecology. In: *Investigations of the Ichthyofauna of Nicaraguan Lakes* (ed. Thorson TB), pp. 157-175. University of Nebraska, Lincoln, Nebraska.
- Bussing WA (1998) Freshwater fishes of Costa Rica. 46 (Suppl. 2), *Revista de Biología Tropical*.
- Cadle JE (1985) The neotropical colubrid snake fauna (Serpentes: Colubridae): lineage components and biogeography. *Systematic Zoology*, 34, 1–20.
- Castillo LE., Ruepert C, and Solis E (2000) Pesticide residues in the aquatic environment of banana plantation areas in the north Atlantic zone of Costa Rica. *Environmental Toxicology and Chemistry*, 19, 1942–1950.

- Chernoff B (1982) Character variation among populations and the analysis of biogeography. *American Zoologist*, 22, 425-439.
- Coates AG, Obando JA (1996) The geologic evolution of the Central America Isthmus. In: *Evolution and Environment in Tropical America* (eds. Jackson JBC, Budd AF, Coates AG), pp. 21–56. Chicago University Press, Chicago, IL.
- Cockerham CC, Weir BS (1993) Estimation of Gene Flow from F-Statistics. *Evolution*, 47, 855-863.
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, 9, 1657–1660.
- Collins LC, Coates AG, Obando JA (1995) Timing and rates of emergence of the Limon and Bocas del Toro Basins. In: *Geologic and tectonic development of the Caribbean Plate boundary in southern Central America* (ed. Mann P), pp. 263-289. Geological Society of America Special Paper 295, Boulder.
- Cracraft J (1994) Species Diversity, Biogeography, and the Evolution of Biotas. *American Zoologist*, 34, 33-47.
- Crandall KA (1994) Intraspecific cladogram estimation: accuracy at higher levels of divergence. *Systematic Biology*, 43, 222–235.
- Crother BI, Campbell JA, Hillis DM (1992) Phylogeny and historical biogeography of the palm pitvipers, genus *Bothriechis*: biochemical and morphological evidence. In: *Biology of the pitvipers* (eds. Campbell JA, Brodie Jr) pp. 1-20. Selva, Tyler.
- Cusack D, Dixon L (2006) Community-based ecotourism and sustainability cases in Bocas del Toro Province, Panama and Talamanca, Costa Rica. *Journal of Sustainable Forestry*, 22, 157-182.

- Doan, D B (1999) U.S. Geological Survey Minerals Yearbook. U.S. Department of the Interior, U.S. Geological Survey. 9.1-9.3
- Drummond AJ, Rambaut A (2006) *BEAST* version 1.4.1 (computer program). Available: <http://Beast.bio.ed.ac.uk/>
- Duellman WE (1979) The South American herpetofauna: a panoramic view. Monographs of the Museum of Natural History, University of Kansas, 7, 1–28.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics*, 131, 479-491.
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39, 783–791.
- Fitzsimmons K (2000) Future trends of tilapia aquaculture in the Americas. In: *Tilapia Aquaculture in the Americas* (eds. Costa-Pierce BA and Rakocy JE), 2nd edn. pp. 252–264. The World Aquaculture Society, Baton Rouge, Louisiana, United States.
- Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147, 915–925.
- Gallup CD, Edward RL, Johnson RG (1994) The Timing of High Sea Levels Over the Past 200,000 Years. *Science*, 263, 796-799.
- García-París M, Good DA, Parra-Olea G, Wake DB (2000) Biodiversity of Costa Rican salamanders: Implications of high levels of genetic differentiation and phylogeographic structure for species formation. *Proceedings of the National Academy of Sciences*, 97, 1640-1647.

- Gentry AH (1982) Neotropical floristic diversity: phylogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Annals of the Missouri Botanical Garden*, 69, 557–593.
- Haq BU, Hardonbol J, Vail PR (1987) Chronology of fluctuating sea levels since the Triassic. *Science*, 235, 1156-1167.
- Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, 42, 182-192.
- Hrbek T, Seckinger J, Meyer A (2007) A phylogenetic and biogeographic perspective on the evolution of poeciliid fishes. *Molecular Phylogenetics and Evolution*, 43, 986-998.
- Huelsenbeck JP, Rannala B (1997) Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science*, 276, 227-232.
- Kluge AG (1989) A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). *Systematic Zoology*, 38, 7–25.
- Loftin HG (1965) *The geographical distribution of freshwater fishes in Panama*. 1st edn. Florida State University, Florida.
- Lowe AJ, Jourde B, Breyne P, *et al.* (2003) Fine-scale genetic structure and gene flow within Costa Rican populations of mahogany (*Swietenia macrophylla*). *Heredity*, 90, 268–275.
- Lucinda PHF and Reis RE (2005) Systematics of the subfamily Poeciliinae Bonaparte (Cyprinodontiformes: Poeciliidae), with an emphasis on the tribe Cnesterodontinin Hubbs. *Neotropical Ichthyology*, 3, 1-60.
- Lundberg JG, Marshall LG, Guerrero J, *et al.* (1998) The stage for Neotropical diversification: a history of tropical South American rivers. In: *Phylogeny and classification*

of Neotropical fishes. (eds. Malabra LR, Rreis RE, Vari RP, Lucena ZM, Lucena CAS), pp. 13-48. Edipuers, Porto Alegre, Brazil.

- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209-220.
- Martin AP, Bermingham E (1998) Systematics and evolution of lower Central American cichlids inferred from analysis of cytochrome *b* gene sequences. *Molecular Phylogenetics and Evolution*, 9, 192-203.
- Mateos M, Sanjur OI, Vrijenhoek RC (2002) Historical biogeography of the livebearing fish genus *Poeciliopsis* (Poeciliidae: Cyprinodontiformes). *Evolution*, 56, 972-984.
- Meek SE (1912) New species of fishes from Costa Rica. *Field Museum Natural History, Zoological Series*, 10, 69-75.
- Miller RR (1966) Geographical distribution of Central American freshwater fishes. *Copeia*, 4, 773–802.
- Murphy WJ, Collier GE (1996) Phylogenetic relationships within the Aplocheiloid fish genus *Rivulus* (Cyprinodontiformes, Rivulidae): implications for Caribbean and Central American Biogeography. *Molecular Biology and Evolution*, 13, 642-649.
- Myers GS (1938) Fresh-water fishes and West Indian zoogeography. *Annual Report Smithsonian Institution*, 1937, 339-364.
- Myers GS (1949) Salt-tolerance of fresh-water fish groups in relation to zoogeographical problems. *Bijdragen tot de Dierkunde*, 28, 315-322.
- Myers GS (1966) Derivation of the freshwater fish fauna of Central America. *Copeia*, 4, 766-773.

- Knowlton N (1993) Sibling species in the sea. *Annual Review of Ecology and Systematics*, 24, 189-216.
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 14, 817-818.
- Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the nested cladistic analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, 9, 487-488.
- Rambaut A, Drummond AJ (2006) Tracer version 1.3 (computer program). Available: <http://Beast.bio.ed.ac.uk/>
- Rand DM (1996) Neutrality tests of molecular markers and the connections between DNA polymorphism, demography, and conservation biology. *Conservation Biology*, 10, 665-671.
- Reznick DN, Mateos M, Springer MS (2002) Independent origins and rapid evolution of the placenta in the fish genus *Poeciliopsis*. *Science*, 298, 1018-1020.
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9, 552-569.
- Rosen DE (1976) A vicariance model of Caribbean biogeography. *Systematic Zoology*, 24, 431-464.
- Rosen DE (1978) Vicariant patterns and historical explanation in biogeography. *Systematic Zoology*, 27, 159-188.
- Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, 15, 174-175.
- Sanderson MJ (2006) r8s, version 1.71: user's manual. Section of evolution and ecology. Davis, CA: University of California.

- Savage JM (1982) The enigma of the Central American herpetofauna: dispersals or Vicariance? *Annals of the Missouri Botanical Gardens*, 69, 464-547.
- Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN ver. 2.0: a software for population genetic data analysis. Genetics and Biometry Laboratory, Univ. of Geneva, Geneva, Switzerland.
- Schulz R (2004) Field studies on exposure, effects, and risk mitigation of aquatic nonpoint-source insecticide pollution. *Journal of Environmental Quality*, 33, 419-448.
- Smith SA, Bermingham E (2005) The biogeography of lower Mesoamerican freshwater fishes. *Journal of Biogeography*, 32, 1835-1854.
- Swofford, DL (1999) PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b8. Sinauer, Sunderland, Massachusetts.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595.
- Templeton AR (1998) Nested clade analyses of phylogeographic data: testing methods about gene flow and population history. *Molecular Ecology*, 7, 381–397.
- Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology*, 10, 779–791.
- Templeton AR (2002) ‘Optimal’ randomization strategies when testing the existence of a phylogeographic structure: a reply to Petit and Grivet. *Genetics*, 161, 473–475.
- Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, 13, 789–809.

- Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, 117, 343-351.
- Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, 132, 619-633.
- Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, 140, 767-782.
- Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetic*, 134, 659-669.
- Unmack PJ (2001) Biogeography of Australian freshwater fishes. *Journal of Biogeography*, 28, 1053-1089.
- Wake DB, Lynch JF (1976) The distribution, ecology and evolutionary history of plethodontid salamanders in tropical America. *Natural History Museum LAC Science Bulletin*, 25, 1-65.
- Zamudio KR (1997) Phylogeography of the bushmaster (*Lachesis muta*: Viperidae): implications for neotropical biogeography, systematics, and conservation. *Biological Journal of the Linnean Society*, 62, 421-442.
- Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, University of Texas at Austin.

Table 1. Locality information for each collection site and voucher numbers for tissue samples. Bold voucher numbers indicate individuals sequenced for both mtDNA and nDNA, while regular font voucher numbers indicate individuals sequenced for mtDNA only. Sample size per population for both genetic markers is indicated.

Drainage	Locality ID	GPS	Tissue Sample BYU Voucher Numbers	Sample Size mt(nDNA)
Tempisque	1	10° 39' 28" N 85° 05' 25" W	120200 , 120201-120205	6 (1)
	2	10° 41' 27" N 85° 05' 01" W	118731, 118736, 118747, 231581 , 231584, 232623, 232659, 232678	8 (1)
	3	10° 41' 14" N 85° 04' 00" W	231648, 231653, 231663-231664, 231674-231675, 232665, 232667	8
	4	10° 41' 17" N 85° 04' 34" W	118764, 118778, 118812, 118813, 118816, 118834, 231773, 231777	8
	27	10° 29' 12" N 84° 59' 12" W	231207 , 231208, 218725, 231169, 231172, 231184, 231190, 231197	8 (1)
Zapote	5	10° 43' 27" N 85° 03' 59" W	232294, 232295, 232300, 232303, 232305, 232317, 232542, 232545	8
	6	10° 43' 40" N 85° 03' 53" W	120197, 120198 , 119623-119627, 119855	8 (1)
	7	10° 43' 53" N 85° 03' 19" W	232132, 232154, 232079, 232084, 232087 , 232141, 232147, 232155	8 (1)
Frio	8	10° 38' 41" N 84° 49' 20" W	135256, 135257, 135260, 135261	4
Pocosol	9	10° 43' 39" N 84° 33' 30" W	119098, 119100, 119109, 232170 , 232173 , 232183, 232186, 232188	8 (2)
Sarapiquí	10	10° 31' 28" N 84° 01' 53" W	135247	1
	11	10° 23' 35" N 83° 58' 05" W	232242 , 232239 , 232240, 232254, 232257, 232259, 232276, 235410	8 (2)
	12	10° 12' 40" N 83° 53' 07" W	232374 , 232378, 232372, 232375, 232385, 232386, 232388, 232416	8 (1)
Tortuguero	13	10° 15' 34" N 83° 48' 44" W	135239 , 135240, 135241 , 135242- 135246	8 (2)
	14	10° 21' 19" N 83° 44' 15" W	120199	1 (1)
Parismina	15	10° 17' 22" N 83° 36' 36" W	120206 , 120207	2 (2)
	16	10° 11' 52" N 83° 39' 07" W	135232 , 135234 , 135235-135238	6 (2)
San Carlos	17	10° 21' 06" N 84° 35' 17" W	120208 , 120209	2 (2)
	18	10° 32' 55" N 84° 58' 51" W	232458, 232460 , 231111, 231152, 231155, 231158, 232583, 232609	8 (1)
	19	10° 33' 35" N 84° 58' 11" W	120195 , 120196-120238, 135228- 123231	8 (1)

San Carlos (continued)	20	10° 33' 37" N 84° 56' 25" W	232872, 232873 , 231837, 231838, 231844, 231851, 231866, 231911	8 (2)
	21	10° 30' 23" N 84° 50' 45" W	119031, 119057, 232340, 232350	4
	22	10° 30' 05" N 84° 50' 26" W	119620-119622	3
	23	10° 29' 32" N 84° 50' 09" W	231406, 231412, 230974, 230975, 231002, 231004, 231011, 231013	8
	24	10° 28' 25" N 84° 49' 20" W	231469, 231467, 231551, 231546	4
	25	10° 26' 06" N 84° 43' 24" W	135248-135251	4
	26	10° 25' 39" N 84° 45' 08" W	135252-135255	4

Table 2. Results from the analysis of molecular variance (AMOVA). AMOVA was used to evaluate five *a priori* biogeographic hypotheses (see Methods section for details). Significant structuring was found in all comparisons, shown here as Fst values, where a value of 1.000 indicates complete isolation and 0.000 indicates no population structuring whatsoever. Here, the term ‘groups’ refers to the collection of populations being compared under each hypothesis. The source of variation from among groups, within groups/among populations, and within populations is given as a percentage for each comparison.

Comparison (number of groups)	Source of variation (percentage)			Fst
	Among groups	Within groups, among populations	Within populations	
Drainage by drainage over entire range (8)	93.72%	4.96%	1.32%	0.987**
Caribbean vs. San Juan drainages (2)	22.67%	76.16%	1.16%	0.988**
San Juan drainage by drainage (5)	93.47%	5.63%	0.90%	0.991**
Caribbean drainage by drainage (2)	95.78%	1.02%	3.20%	0.968**
Across Continental Divide (2)	70.54%	2.06%	27.40%	0.726**

** P < 0.001

Table 3. Inferences about population histories generated from the nested clade phylogeographic analysis (NCPA). We report inferences only for clades in which the observed χ^2 value was significantly greater than the χ^2 value generated under the random permutation test.

Clade	χ^2 probability	Chain of inference ¹	Inference
1-1	0.0000	1-2-11-12-NO	Contiguous range expansion within Arenal Basin
1-5	0.0037	1-2-3-4-NO	Restricted gene flow, isolation by distance within the Tempisque drainage
2-1	0.0000	1-19-NO	Allopatric fragmentation between the Arenal Basin and the San Carlos drainage
2-2	0.0000	1-19-NO	Allopatric fragmentation between the Parismina drainage and the upper Sarapiquí drainage
2-3	0.0000	1-2-11-12-13-14-NO	Past fragmentation between the Tempisque and Zapote drainages
3-1	0.0014	1-19-20-2-11-12-13-YES	Long distance colonization possibly coupled with subsequent fragmentation or past fragmentation, followed by range expansion, between the Frio and Tortuguero drainages
3-2	0.0008	1-2-11-12-13-14-NO	Long distance colonization and/or past fragmentation (not necessarily mutually exclusive) between Lower Sarapiquí and the Pocosal drainage
4-1	0.0000	1-19-NO	Allopatric fragmentation between Clade 3, Clade 4a/b, and Clade 4c
5-1	0.0000	1-19-NO	Allopatric fragmentation between Clades 3 & 4 and Clade 2
Total	0.0000	1-19-NO	Allopatric fragmentation between Clades 2, 3, & 4 and Clade 1

¹ based on Templeton's (2005) inference key.

Table 4. Results of the Bayesian analysis of coalescent time estimates using *Beast*. Mean and error estimates are based on the results of six separate runs of 10 million generations and a burn-in of 100,000. HPD = highest posterior density, ESS = effective sample size, T_{MRCA} labels correspond with those mapped onto the phylogeny in Figure 4.

Statistic	Mean when mutation rate is 1% per million years	Mean when mutation rate is 2% per million years	Combined Mean	Combined 95% HPD lower	Combined 95% HPD upper	Combined ESS
Posterior	-5015.528	-4986.583	-5000.408	-5026.305	-4976.136	1674.869
treeLikelihood	-4859.04	-4859.197	-4859.101	-4870.16	-4849.248	2381.96
coalescent	-156.487	-127.386	-141.307	-161.42	-133.059	1295.193
treeModel.rootHeight	38.001	19.016	28.1	15.474	43.433	2582.758
Constant.popSize	15.85	7.932	11.725	5.416	19.635	3667.899
T_{MRCA} 1	0.225	0.113	0.167	0.047	0.333	1246.942
T_{MRCA} 2	0.46	0.232	0.342	0.122	0.628	1569.906
T_{MRCA} 3	1.035	0.524	0.769	0.335	1.299	1820.188
T_{MRCA} 4	0.553	0.277	0.41	0.132	0.796	2889.334
T_{MRCA} 5	1.416	0.712	1.048	0.482	1.769	2298.014
T_{MRCA} 6	0.387	0.191	0.285	0.054	0.611	2953.186
T_{MRCA} 7	2.516	1.26	1.859	0.914	3.075	3035.626
T_{MRCA} 8	0.496	0.245	0.364	0.117	0.698	1274.643
T_{MRCA} All	5.946	2.977	4.392	2.299	7.059	2630.947
T_{MRCA} <i>P. annectens</i>	25.707	12.795	18.935	10.38	28.98	1800.842

Table 5. Population diversity indices and mtDNA equilibrium test statistics for *Xenophallus umbratilis*. Populations were divided into four geographically isolated sets that correspond to the four clades recovered in the phylogenetic analysis (see Fig. 3). Summary statistics and equilibrium tests were run on each of these clades and on the entire data set overall. Population diversity measures are defined as follows: n, sample size; H, number of haplotypes; M, total number of mutations; S, number of segregating (polymorphic) sites; π , average pairwise nucleotide diversity per site; θ_w , effective population size by mutation rate (for haploids, $2N_e\mu$). The equilibrium tests evaluate whether or not groups are at mutation-migration-drift equilibrium: significant negative values indicate either recent population expansion or positive (diversifying) selection; significant positive values indicate either a recent population bottleneck or negative (balancing) selection.

Sample Set	Population diversity estimates					mtDNA equilibrium tests statistics			
	n	H	M(S)	π	θ_w	Fu and Li's F^*	Fu's F_s	Fu and Li's D^*	Tajima's D
Clade 1	61	9	9(9)	0.00068	2.137	-1.69630	-6.997*	-1.81287	-1.69301†
Clade 2	16	4	4(4)	0.00140	1.205	0.51954	0.824	0.25371	1.00231
Clade 3	17	7	8(8)	0.00191	2.366	-1.20360	-1.311	-1.33662	-0.28379
Clade 4	68	15	23(23)	0.00391	4.817	-0.12986	-0.967	-0.03979	-0.23017
All	162	36	104(102)	0.02830	18.036	1.99824*	11.549‡	1.06623	2.38721*

‡ P < 0.01; * P < 0.05; † P < 0.10

Table 6. Distribution of *Xenophallus umbratilis* mtDNA haplotypes identified by drainage basin and collection sites within drainages. Collection site identification numbers (ID) correspond to those presented in Figure 2. Values in the matrix identify the number of individuals from each population with a particular haplotype; empty cells indicate zeros.

Drainage	ID	mtDNA Haplotypes																																									
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36						
Tempisque	1	2	3	1																																							
	2	3			4	1																																					
	3	1	6					1																																			
	4	2	6																																								
	27																																						8				
Zapote	5						6	2																																			
	6						6	1	1																																		
	7						6	2																																			
Frio	8									3	1																																
Pocosol	9										7	1																															
Sarapiquí	10													1																													
	11														5	1	1	1																									
	12																			8																							
Tortuguero	13																		5	3																							
	14																									1																	
Parismina	15																								1	1																	
	16																								2		4																
San Carlos	17																												2														
	18																																					7	1				
	19																																					4	2	1	1		
	20																																					5			3		
	21																																					4					
	22																																					3					
	23																																					6	1				1
	24																																					4					
	25																																					3				1	
	26																																					4					

Figure Legends

Figure 1. Map indicating the major drainages occupied by *Xenophallus umbratilis* and showing elevational clines across these drainages. Areas susceptible to flooding during high sea levels are indicated by different shades of grey. The continental divide is indicated by dashed line running the length of the land mass. Comparison of this map to Figure 2 reveals that much of the current distribution of *X. umbratilis* occurs at sufficiently low elevations to be impacted by rises in sea level. Note that during the gradual rise of this land bridge much of this area would have been at even lower elevations.

Figure 2. Map showing the geographic sampling of *Xenophallus umbratilis* populations for this study. Samples were taken from 27 localities in eight drainages throughout the northern provinces of Costa Rica. Solid circles indicate locations where *X. umbratilis* were present; open circles indicate sites where *X. umbratilis* were not present. The southern range of *X. umbratilis* is indicated by the grey area (after Bussing 1998). The northern range fades from grey to white to show that the extent to which *X. umbratilis* is distributed in Nicaragua is largely unknown. The continental divide is symbolized by the dotted line.

Figure 3. Haplotype network and nesting structure for 36 *Xenophallus umbratilis* haplotypes identified in this study. Circle sizes are proportional to haplotype frequency. Each line indicates one substitution between haplotypes and small empty circles represent haplotypes assumed to exist but that were not sampled. Nested clades are designated using a two-number system, with the first number refers to nesting level, and the second number identifying individual clades. Dashed lines indicate that the probability of parsimony did not exceed 95% for connections between clades 4-1 and 4-2 or between clades 5-1 and 5-2. Hence, links between these clades are inferred from the phylogeny presented in Figure 4.

Figure 4. Strict consensus tree of maximum parsimony (MP) and maximum likelihood (ML) phylogenies generated from the mitochondrial DNA data set. Drainages and assigned clades are labeled to the right of the phylogeny. Only nodal bootstrap values above 70 percent are indicated.

Figure 5. The time to most recent common ancestor (T_{MRCA}) for each of the clades found in phylogenetic analysis as estimated in the program *Beast* (Drummond and Rambaut 2003). These divergence times are here mapped onto the phylogeny against scaled geological and climactic events. The horizontal grey rectangles represent the 95% credibility intervals for each time estimate. Vertical grey rectangles with dotted line boundaries indicate sea level highs that coincide with a mean estimate of divergence within lineages of *X. umbratilis*. Eustatic cycles are taken from Haq et al. (1987) and Gallup et al. (1994).

Figure 6. Bayesian skyline analysis revealing a constant population size historically, with a recent bottleneck event about 0.35 mya. Light grey lines indicate the upper and lower 95% confidence intervals.

Figure 1.

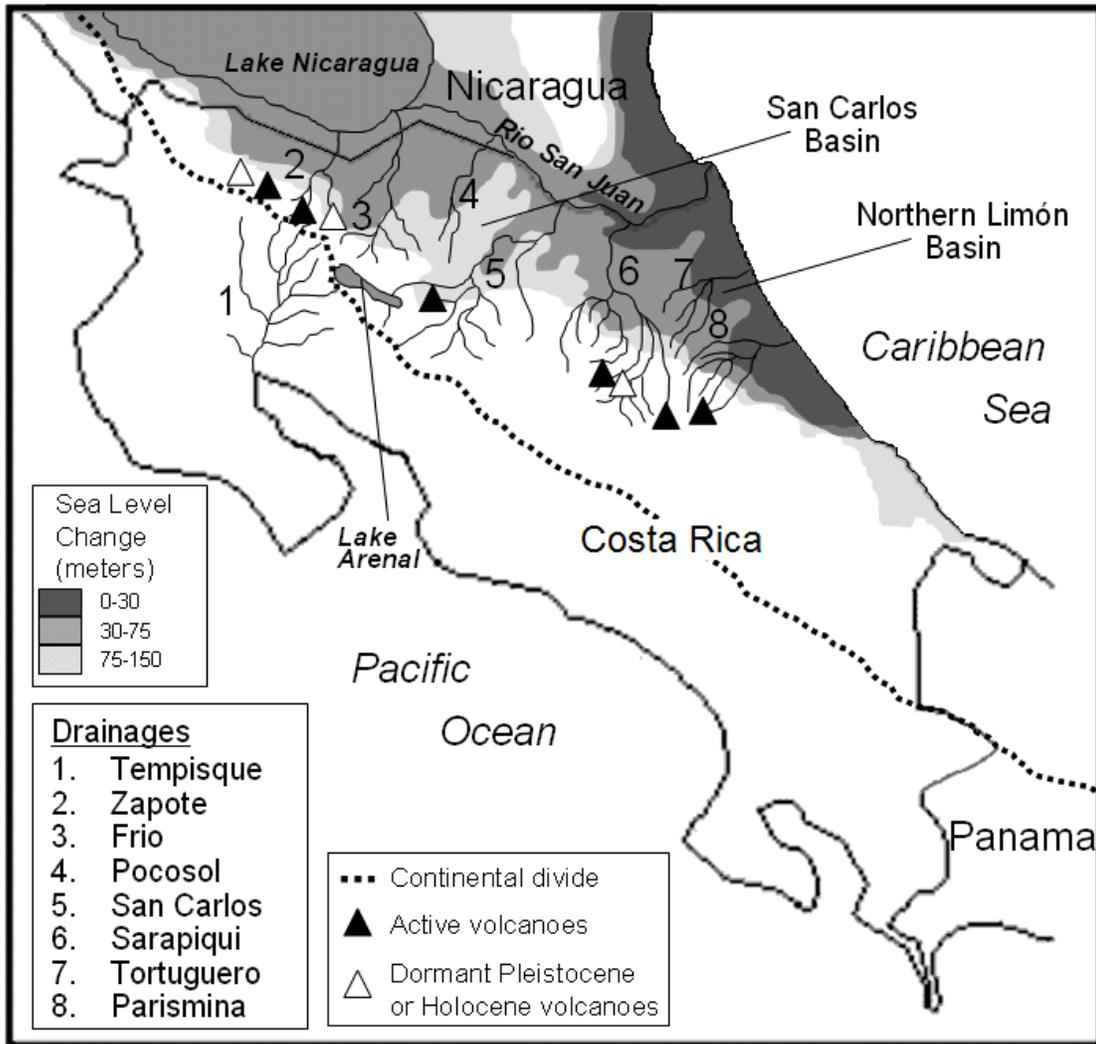


Figure 2.

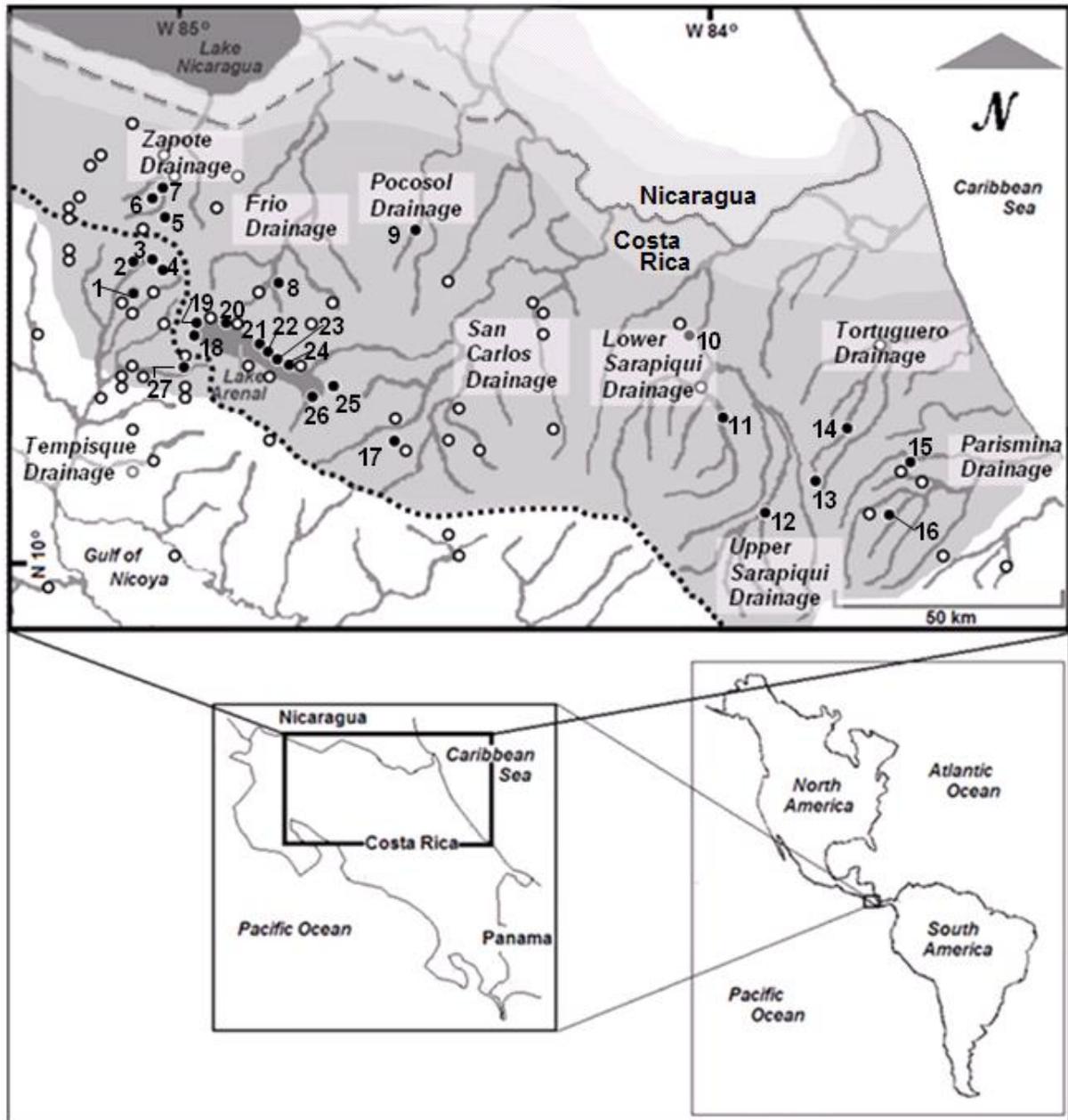


Figure 4.

mtDNA MP/ML Strict Consensus

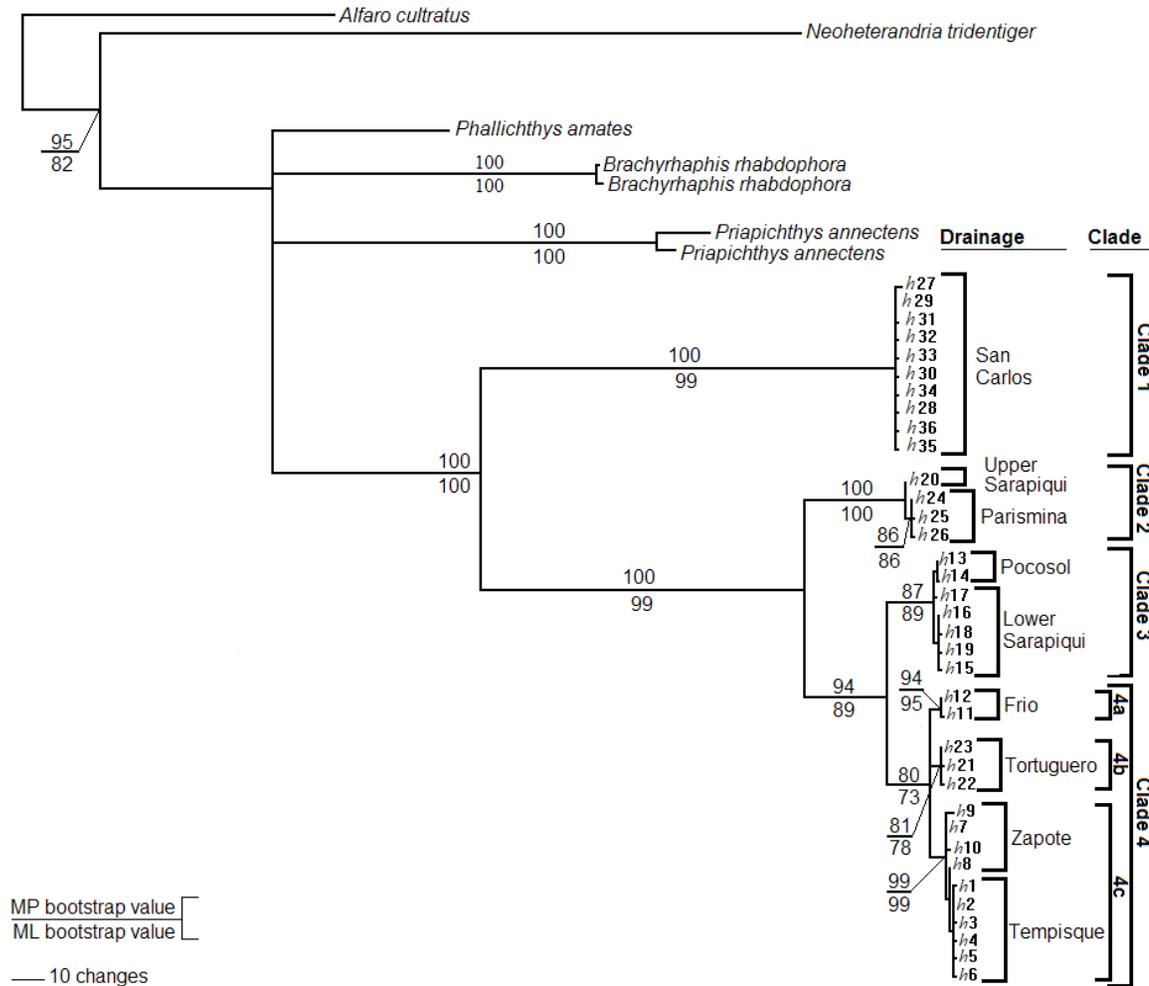


Figure 5.

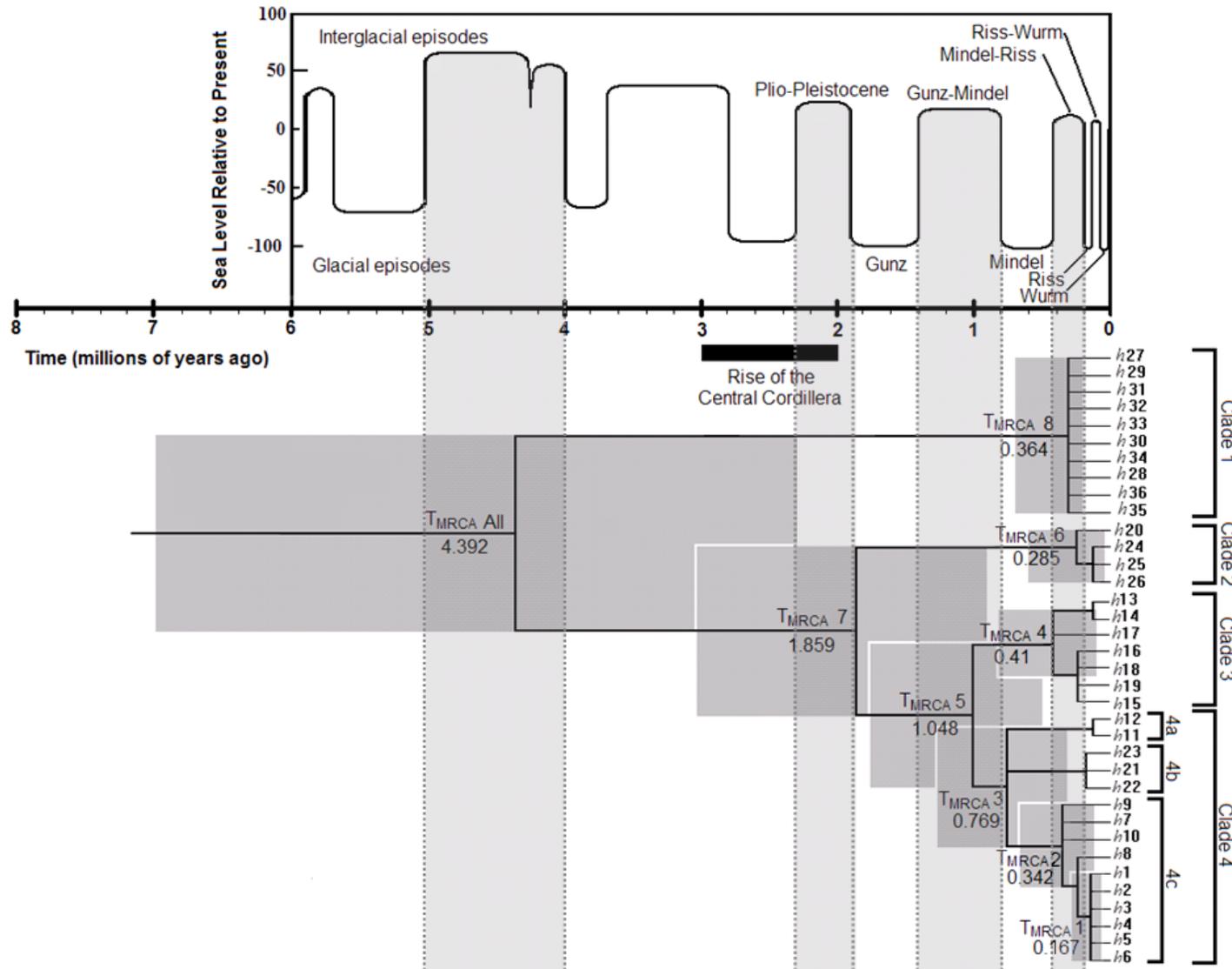


Figure 6.

