Phylogenetic divergence in leatherside chub (*Gila copei*) inferred from mitochondrial cytochrome *b* sequences

JERALD B. JOHNSON* ‡ and STEVE JORDAN†

*Department of Biology, University of Vermont, Burlington, VT 05405, USA †Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA

Abstract

We examined intra-specific phylogenetic relationships in leatherside chub, *Gila copei*. The complete mitochondrial (mt) cytochrome *b* gene (1140 bp) was sequenced for 30 individuals from 10 populations that span the geographical distribution of this species. Traditional phylogenetic analyses revealed two deeply divergent and evolutionarily distinct mtDNA clades that are geographically separated in northern and southern drainage basins. Inter-population sequence variation between clades ranged from 7.7 to 8.1%. The northern clade was genetically more similar and phylogenetically more closely related to the selected out-group *Lepidomeda m. mollispinus* than to the southern clade, suggesting that the taxonomy of this species may require revision. Sequence variation among populations within clades ranged from 0 to 0.3% in the north and from 0 to 0.7% in the south. Statistical parsimony was used to construct phylogenetic networks of haplotypes within clades. Nested clade analysis revealed that geographical fragmentation has played an important role in genetic structuring within northern and southern clades.

Keywords: conservation genetics, cytochrome *b*, intra-specific phylogeny, nested clade analysis, species boundaries, statistical parsimony

Received 1 October 1999; revision received 31 December 1999; accepted 20 January 2000

Introduction

Molecular systematics is now recognized as an integral part of conserving rare species (Haig 1998; Soltis & Gitzendanner 1999). By delineating boundaries between evolutionary lineages, phylogenetic hypotheses focus attention on genetically distinct populations or groups of populations that may require special protection or warrant independent management strategies (Moritz 1994). Further, phylogenetic reconstruction identifies the relationships between these lineages and reveals the extent to which cohesive evolutionary groups have diverged. Such information is critical to planning conservation strategies that seek to maintain genetic diversity, preserve adaptive potential, and protect evolutionary processes.

In this study, we examined intra-specific divergence in leatherside chub (*Gila copei*), a threatened fish endemic to the Bonneville Basin and upper Snake River drainages of western North America. This biogeographic distribution is thought to coincide with the hydrological history of Pleistocene Lake Bonneville (Smith 1978), suggesting a recent expansion (approximately 14 500 years ago) of populations from the Bonneville Basin to the upper Snake River (Jarrett & Malde 1987). Historically, leatherside chub were common and widespread throughout this range (Sigler & Sigler 1996). However, declines in distribution and abundance of this species now threaten its long-term survival (Sigler & Sigler 1996; Wilson and Belk, in press). Understanding the extent of phylogenetic differentiation among extant populations will be an integral part of protecting the evolutionary potential and genetic diversity of leatherside chub. Hence, the objectives of our study were to identify phylogenetic and geographical relationships among leatherside chub haplotypes; to infer processes responsible for observed genetic differentiation; and to provide an evolutionary framework in which conservation decisions for this species can be evaluated.

© 2000 Blackwell Science Ltd
Methods

Field sampling

Leatherside chub were sampled from 10 sites in the four major drainage basins where this fish now occurs (Snake, Bear, Utah Lake and Sevier) – collectively, these samples span the current geographical distribution of the species (Fig. 1). From each collecting site we examined three fish for a total of 30 individuals, a conservative sample size that reflects the uncertain conservation status of this species. Historically, leatherside chub may have occupied two additional drainage basins (Weber and Salt Lake) located between the current populations; despite extensive sampling from suitable habitats in these drainages, extant populations have not been found.

DNA extraction, amplification and sequencing

Genomic DNA was extracted from ethanol-preserved muscle tissue using a proteinase K digestion, ammonium acetate separation of proteins, and cold-ethanol precipitation. We amplified the cytochrome b gene by PCR using primers LA and HA (Dowling & Naylor 1997). The thermal profile (94 °C for 1 min; 48 °C for 1 min; 72 °C for 2 min) was repeated 35 times. Purified double-stranded DNA was used as template (150 ng) for automated sequencing reactions. We sequenced the entire gene (1140 bp) in both forward and reverse directions for each fish sampled using external primers LA and HA and internal primers LD (Dowling & Naylor 1997) and INH (designed for this study: 5’-GGGTGTGGGAYCCSGTYTCGT-3’). Sequences are available from GenBank (accession numbers AF270885–AF270914).
Data analysis

Cytochrome b sequences were aligned using Sequencher version 3.0 (Gene Codes Corp., Ann Arbor, Michigan, USA). Basic sequence attributes, including raw haplotype divergences, codon site-specific mutation rates, base composition bias, and amino acid substitutions were calculated using MEGA (Kumar et al. 1993) and PAUP* version 4.0b2 (Swofford 1999). An appropriate maximum likelihood model for these data was selected using the model-fitting routine in Frati et al. (1997). The resulting negative log-likelihood scores were compared using a likelihood ratio test (Swofford et al. 1996). Based on this analysis, we selected a general time-reversible model of molecular evolution (GTR: –ln likelihood = 2186.35), the simplest model statistically indistinguishable (\(\chi^2 = 4.84, P = 0.09\)) from the most likely model evaluated (GTR + I + G: –ln likelihood = 2183.94).

We employed two suites of techniques to examine phylogeographic relationships among leatherside chub mitochondrial (mt) DNA haplotypes: maximum parsimony/maximum likelihood analyses (Swofford et al. 1996) and statistical parsimony coupled with nested clade analysis (Templeton et al. 1992; Templeton et al. 1995). These techniques are designed to detect haplotype associations at different levels of phylogenetic divergence. Maximum likelihood and maximum parsimony are well-suited for the analysis of bifurcating phylogenetic groups with many synapomorphic traits, but have difficulty resolving relationships of closely related haplotypes (Crandall 1994). Statistical (or TCS) parsimony (Templeton et al. 1992) allows haplotypes with extremely low divergences to be organized into cladograms by calculating the maximum number of mutations that can be present in a haplotype network with a 95% probability that no multiple hits have occurred (see Crandall 1996). Out-group weights can be estimated for each haplotype to identify probable roots in the network (Castelloe & Templeton 1994). Haplotype networks can then be organized into a series of nesting clades (Crandall 1996) that are used to identify geographical associations among haplotypes (Templeton 1998).

We first calculated phylogenetic relationships among leatherside chub haplotypes using PAUP* 4.0b2 (Swofford 1999). Equally weighted maximum parsimony analysis was conducted using the heuristic search option (starting tree generated through 10 replications of random stepwise addition, with TBR branch swapping) and a strict consensus was constructed to reconcile equally parsimonious topologies. Support for each node of the resulting tree was evaluated by 1000 bootstrap replicates (Felsenstein 1985). We also conducted a heuristic search using the maximum likelihood criterion with the general time-reversible model of molecular evolution (Frati et al. 1997). Nodal support for the resulting tree was evaluated by 100 bootstrap replicates. In all analyses, the plagopterin fish *Lepidomeda m. mollispinus* was included as an out-group based on the phylogenetic hypothesis of Simons & Mayden (1997).

We used two software packages to complete the statistical parsimony and nested clade analyses. PAUP* 1.0 (D. Posada 1999, http://bioag.byu.edu/zooloogy/crandall_lab/programs.htm) uses a coalescence model to calculate the probability that there are no hidden mutations in a network of haplotypes (i.e. that the network is parsimonious). cxpars 2.0 (Posada et al. 2000) tests for genetic structuring by evaluating the geographical distribution of haplotypes in a nested clade framework (Templeton & Sing 1993). We estimated geographical distances by river connections between sites and linked drainage systems based on late-Pleistocene water connections (Jarrett & Malde 1987; Currey 1990). We constructed haplotype networks by hand and calculated out-group weights as outlined in Crandall (1994). Finally, we explored the nested clade analysis using the inference key of Templeton (1998), which summarizes patterns of haplotype distribution expected when genetic structuring is the result of restricted gene flow, vicariant fragmentation events, or past range expansions.

Results

Molecular evolution of the cytochrome b gene in leatherside chub consisted exclusively of point mutations. Substitutions were synonymous in all but nine cases, and the majority of these involved the exchange of structurally similar amino acids. Of the 120 observed mutations (out of 1140 bp), 9, 1 and 110 were at the first, second, and third codon positions, respectively. Ninety-one of these mutations were parsimony-informative. Nucleotide base frequencies did not differ among taxa (\(\chi^2 = 1.81, P = 0.09\)), with average frequencies of \(A = 0.25, C = 0.29, G = 0.17\) and \(T = 0.29\).

Sequence divergence between individuals from the northern drainage basins (Snake and Bear) and individuals from the southern drainage basins (Utah Lake and Sevier) ranged from 7.7 to 8.1%. This contrasted sharply with the low variation found among individuals within stream drainage basins in the north (0–0.3%) and in the south (0–0.7%). Unexpectedly, we also found that leatherside chub from northern streams showed lower genetic distances from the selected out-group *Lepidomeda m. mollispinus* (3.7%) than from con-specifics in southern streams (8.0%).

Maximum parsimony and maximum likelihood consistently found two strongly supported, evolutionarily distinct clades of leatherside chub (Fig. 2). The northern clade comprised haplotypes found only in the Snake and Bear River drainages while the southern clade comprised haplotypes found only in the Utah Lake and Sevier River

Drainages. Phylogenetic analyses using maximum parsimony and maximum likelihood produced similar cladograms — both show complete phylogenetic separation of the northern and southern groups. A total of 68 characters unambiguously supported monophyly of the southern clade, while 19 characters unambiguously supported monophyly of the northern clade. However, neither maximum parsimony nor maximum likelihood analyses provided adequate resolution of haplotype associations within the northern and southern clades (Fig. 2).

In contrast, statistical parsimony successfully resolved relationships among haplotypes within clades. Haplotypes that differed by as many as 14 single mutational steps could be parsimoniously connected at the 95% confidence level — this resulted in the construction of two networks (Fig. 3), one comprising four northern clade haplotypes and the other comprising 10 southern clade haplotypes. Within the northern clade, we found no geographical overlap between Snake River and Bear River haplotypes; a Bear River haplotype (B3) occupied the single interior position in the network and consequently had the highest out-group weight (0.75). Within the southern clade, we found no geographical overlap between Utah Lake and Sevier River haplotypes. Out-group weights found among southern clade haplotypes were more uniform than in the north, with an upper value of 0.30 shared by haplotypes U3 and S3.

Within the northern clade, nested clad analysis revealed significant geographical isolation between chub haplotypes nested in clades 1-1 and 1-2, but no geographical association among haplotypes within clade 1-2 (Fig. 3). Within the southern clade, significant geographical associations between haplotypes were found among one-step clades nested in clade 2-2 and among two-step clades nested in the entire cladogram. In all other comparisons, we failed to reject the null hypothesis of no geographical association among haplotypes.

© 2000 Blackwell Science Ltd, Molecular Ecology, 9, 1029–1035
Discussion

Leatherside chub is composed of two evolutionarily distinct groups of leatherside chub: northern clade and southern clade. Each line in the network represents a single mutational change between haplotypes; the dashed line identifies an ambiguous connection of three mutational steps between U1 and two other haplotypes. 0 indicates a haplotype not present in the sample but inferred to be intermediate between two nearest-neighbour haplotypes. Haplotypes correspond to those given in Fig. 2.

Davis & Nixon (1992) suggest that evolutionary lineages become candidates for species when at least one character state is fixed in one group and absent in the other. Although inadequate sampling can falsely designate species candidates, our sample sites are geographically widespread and our data overwhelmingly satisfy the candidate criteria. Templeton (1994) argued that evolutionary lineages defined as species must (1) be phylogenetically distinct, (2) show no recent gene flow, and (3) demonstrate ecological or demographic limitations to reproduction. For the mitochondrial genome, our study satisfies criteria 1 and 2. Some evidence suggests that criterion 3 might also be met as differences in reproductive timing have been noted between Sulphur Creek in the northern clade and Main Creek in the southern clade (Johnson et al. 1995). Avise & Ball (1990) require concordant divergence among clades for two or more unlinked markers. Their standard highlights an important caveat with our data—mitochondrial introgression or differences in lineage sorting rates could result in discordance between the haplotype gene tree presented here and the underlying species tree (Pamilo & Nei 1988; Moore 1995). Hence, our study invites similar analyses using nuclear gene loci or other independently segregating characters.

Additional analyses support the existence of two major evolutionary lineages of leatherside chub. Our parsimony analysis resulted in an unrooted network of three distinct taxa: two leatherside chub clades and the selected out-group *Lepidomeda m. mollispinus*. Such a network can be rooted—revealing the relationships of its constituent members—using a maximum likelihood analysis enforcing a molecular clock (Felsenstein 1983; Lewis 1998). We ran this post hoc analysis (fit of the molecular clock model was statistically indistinguishable from the GTR model evaluated above; likelihood ratio test, $\chi^2 = 16.04$, $P = 0.25$) and found that the northern clade was more closely related to *L. m. mollispinus* than to the southern clade. This finding is confirmed by a broader phylogenetic analysis that included all extant *Lepidomeda* species (Dowling et al., unpublished data) which showed the northern clade was most closely related to an unresolved trichotomy of *L. m. mollispinus*, *L. m. pratensis* and *L. albivallis*. *L. vittata* was sister to this group, and our southern leatherside chub clade was basal to the entire group. However, since this phylogeny is also derived from cytochrome *b* sequences, the need for corroborative evidence from independently segregating characters remains.

The phylogeography of leatherside chub is enigmatic if we assume a sister relationship between the northern clade and *L. mollispinus + L. albivallis*. *L. mollispinus* occurs in the Virgin River system (tributary to the Colorado River) of extreme southern Utah, *L. albivallis* is restricted to the pluvial White River of southern Nevada, and the northern leatherside chub clade occurs far to the north in southern...
Idaho and Wyoming (Fig. I). Geographically intermediate to these groups is the more distantly related southern leatherside chub clade. Dowling et al. (unpublished data) proposes a Late Miocene connection from upper Snake River, through the Bonneville Basin, to pluvial White River, and into the Colorado River that could account for the observed disjunct distributions. This hypothesis is in contrast to a more traditional view that the majority of the upper Snake River fish fauna were introduced from the Bonneville Basin in the late Pleistocene (Hubbs & Miller 1948).

Nested clade analysis provided further insight into factors responsible for the distribution of genetic variation among leatherside chub populations within each clade. In both northern and southern clades, the inference key of Templeton (1998) indicated that genetic structuring within each cladogram was due primarily to geographical fragmentation. In the northern clade, this pointed to allopatric fragmentation between the Bear River and Snake River haplotypes. In the southern clade, we found genetic subdivision among haplotypes restricted to disjunct drainage basins, suggesting that geographical barriers have restricted gene flow with the most likely mechanism being the late Pleistocene recession of ancient Lake Bonneville (Currey 1990). One exception pointed to long-distance colonization between populations in Salina Creek (Sevier River drainage) and Spanish Fork River (Utah Lake drainage). The conservation implications of our findings are clear. Failure to recognize two evolutionarily distinct groups of leatherside chub populations could unwittingly lead to the loss of biodiversity (May 1990). If fish from northern and southern clades are capable of reproducing, the homogenizing effects of introgression via translocations between clades could severely limit the adaptive potential of these genetically distinct clades. Perhaps the most intriguing question that arises from our results is why such genetically distinct clades have eluded systematic consideration to be a single cohesive species. The out-group sequence. P. Lewis gave insightful feedback on maximum likelihood models in phylogenetic reconstruction. The manuscript was improved by comments from T.E. Dowling, K.A. Crandall and the Gotelli/Brody laboratory group. This work is dedicated to the memory of Danny Harris, a graduate colleague of ours while at Brigham Young University.

References


