Genetics and Molecular Biology, 31, 1 (suppl), 343-351 (2008)

www.sbg.org.br

Research Article

Phylogeny of the Serrasalmidae (Characiformes) based on mitochondrial DNA sequences

Guillermo Ortí, Arjun Sivasundar[1], Kelly Dietz and Michel Jégu[2]

School of Biological Sciences, University of Nebraska, Lincoln, NE, USA.

Abstract

Previous studies based on DNA sequences of mitochondrial (mt) rRNA genes showed three main groups within the subfamily Serrasalminae: (1) a "pacu" clade of herbivores (*Colossoma*, *Mylossoma*, *Piaractus*); (2) the "*Myleus*" clade (*Myleus*, *Mylesinus*, *Tometes, Ossubtus*); and (3) the "piranha" clade (*Serrasalmus*, *Pygocentrus*, *Pygopristis*, *Pristobrycon*, *Catoprion*, *Metynnis*). The genus *Acnodon* was placed as the sister taxon of clade (2+3). However, poor resolution within each clade was obtained due to low levels of variation among rRNA gene sequences. Complete sequences of the hypervariable mtDNA control region for a total of 45 taxa, and additional sequences of 12S and 16S rRNA from a total of 74 taxa representing all genera in the family are now presented to address intragroup relationships. Control region sequences of several serrasalmid species exhibit tandem repeats of short motifs (12 to 33 bp) in the 3' end of this region, accounting for substantial length variation. Bayesian inference and maximum parsimony analyses of these sequences identify the same groupings as before and provide further evidence to support the following observations: (a) *Serrasalmus gouldingi* and species of *Pristobrycon* (non-*striolatus*) form a monophyletic group that is the sister group to other species of *Serrasalmus* and *Pygocentrus*; (b) *Catoprion*, *Pygopristis*, and *Pristobrycon striolatus* form a well supported clade, sister to the group described above; (c) some taxa assigned to the genus *Myloplus* (*M. asterias, M tiete, M ternetzi, and M rubripinnis*) form a well supported group whereas other *Myloplus* species remain with uncertain affinities (d) *Mylesinus, Tometes* and *Myleus setiger* form a monophyletic group.

Key words: piranhas, pacus, D-loop, phylogeny, Bayesian inference.

Received: September 13, 2006; Accepted: April 19, 2007.

Introduction

Piranhas and pacus (Serrasalmids) form a distinctive assemblage of characiform fishes. For a long time, they were considered a subfamily within the family Characidae. Recent phylogenetic studies of these fishes, however, strongly suggest that Characidae is non-monophyletic and that serrasalmids are not closely related to taxa originally placed in the subfamily Characinae, or other characid subfamilies (Zanata, 2000), but rather that they may be more closely related to Anostomoidea (Calcagnotto *et al.*, 2005). All these arguments support the separate family status of piranhas and pacus; their relationships to other families within the order Characiformes, however, remain uncertain (Ortí and Meyer, 1997; Calcagnotto *et al.*, 2005; Hubert *et al.*, 2005). Species of the Serrasalmidae are endemic to the Neotropics and are distributed widely in all the major river systems of South America. At least 60 species (in 15 genera) have been recognized. This family includes the well-known piranhas, notorious from accounts of their group-predatory behavior, the seed-eating tambaquí, which is highly regarded as a food species, and the pacus. Several serrasalmid species are of economic importance and are used in aquaculture (Junk, 1984; Marshall, 1995; Araujo-Lima and Goulding, 1997).

Characteristic features of serrasalmids include a compressed body, a long dorsal fin with more than 16 rays and the presence of sharp serrae arising from modification of abdominal scales. The number of these serrae is variable, ranging from 6 to 9 in *Acnodon* to over 60 in *Piaractus*.

Serrasalmids occupy diverse habitats from lowland floodplains and flooded forests to upstream habitats in the headwater regions of river systems (Lowe-McConnel, 1975; Géry, 1977, 1984). They also display a range of trophic specializations, with three general feeding habits: carnivory, frugivory and lepidophagy (feeding on the scales of other fishes). Feeding habit is reflected in the mor-

Send correspondence to Guillermo Orti. School of Biological Sciences, University of Nebraska, 314 Manter Hall, Lincoln, 68588 NE, USA. E-mail: gorti1@unl.edu.

[1]Current address: Stanford University Hopkins Marine Station, Pacific Grove, CA, USA.

[2]Current address: Museum National d'Histoire Naturelle, Ichtyologie, Paris, France.

phology and patterns of dentition found among these taxa (Goulding, 1980).

Carnivorous serrasalmids usually have one row of tricuspid teeth on each jaw, while frugivores have two series of incisor or molariform teeth on the premaxilla, one row of teeth on the dentaries, and often a pair of symphyseal teeth. The lepidophagous taxa have tuberculated teeth located on the outer side of the premaxilla that are used to remove scales from other fish. Not all species are specialists however, and their feeding habit varies with age and food availability (Nico and Taphorn, 1988; Winemiller, 1989; Leite and Jégu, 1990). The arrangement and morphology of teeth have been the main characters traditionally used in serrasalmid classification.

Eigenmann (1915) erected the subfamilies Serrasalminae, containing six genera with one row of teeth on each jaw, and Mylinae, with nine genera having two rows of teeth on the premaxilla. The monotypic, lepidophagous genus *Catoprion* was included in the Mylinae. Classifications that followed also were based largely on dental morphology (Norman, 1929; Gosline, 1951; Géry 1977), and differed mainly in the assignment of ranks for some taxa (*e.g.* genera changed to subgenera).

In the first cladistic treatment of serrasalmid systematics, Machado-Allison (1983) inferred the presence of two lineages, labeled A and B (Figure 1a), which correspond to the Mylinae and Serrasalminae of Eigenmann, respectively, but including the genera *Catoprion* and *Metynnis* with the piranha clade. The first test of this hypothesis with molecular data (Ortí *et al.*, 1996) used mitochondrial DNA (mtDNA) sequences, and recovered a phylogeny of the group containing three or four distinct lineages rather than two (Figure 1b) based on fragments of the 12S and 16S rRNA genes. Relatively low levels of sequence divergence among the rRNA genes, however, resulted in poor resolution within these groups, and a representative of the genus *Pygopristis* was not included in that study. The mtDNA data strongly suggested that *Pristobrycon* includes two components: *Prystobrycon striolatus*, closely allied to *Catoprion*, and the other species of *Pristobrycon*, more closely related to *Serrasalmus* and *Pygocentrus*. A recent phylogenetic study of species of *Serrasalmus* and *Pygocentrus* (Hubert *et al.*, 2007) based on mitochondrial control region sequences provided higher resolution for this group. Within the "*Myleus* clade," mtDNA data (Ortí *et al.*, 1996) were not able to resolve with confidence the relationships among the included taxa, but also did not support the monophyly of *Myleus* or the subspecies designations proposed by Géry (1972, 1977): *Myleus*, *Myloplus*, *Prosomyleus*, and *Paramyloplus*. A morphological reassessment of elements included in *Myleus* (Jégu and Santos; 2002; Jégu *et al.*, 2003) proposed the recognition of *Myleus setiger* (formerly *Myleus pacu*) as the only valid representative of the genus and moved the other components to the genus *Myloplus* (originally erected by Gill, 1896). We follow these taxonomic recommendations in this study.

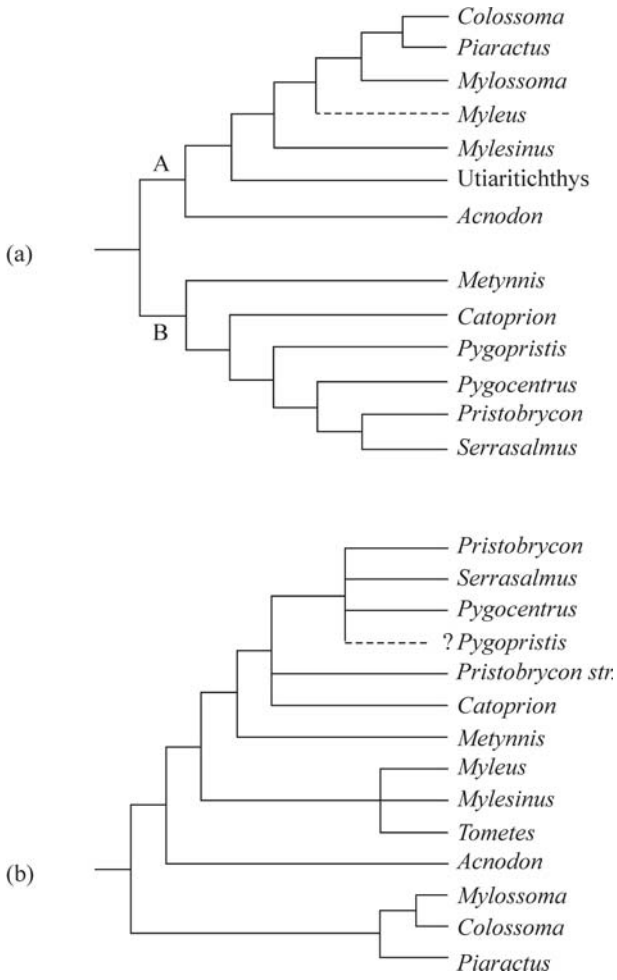

Figure 1 - Previous phylogenetic hypotheses for the Serrasalmidae: (a) phylogenetic relationships within the subfamily Serrasalminae proposed by Machado-Allison (1983) based on morphological characters and (b) Ortí *et al.* (1996) based on partial mtDNA sequences of the 12S and 16S ribosomal RNA genes.

Most recently, a multi-gene assessment of characiform phylogeny based on mitochondrial (16S and cytochrome b) and nuclear DNA (4 fragments) supported the distinctive groping of serrasalmids among charciforms (Calcagnotto *et al.*, 2005). But since that study focused on higher-level relationships among characiforms, it included only 12 serrasalmid taxa, and obtained inconclusive results for within family relationships.

The current study aims to evaluate the previous findings with an extended data set, and also employ a more variable molecular marker to resolve relationships at the shallower nodes within each of the groups. The taxonomic sampling of the 12S and 16S mtDNA sequence data set is here extended from 34 to a total of 74 serrasalmid taxa (including *Pygopristis*). In an attempt to increase resolution among closely related species, 44 sequences from the mitochondrial control region (D-loop) representing all genera in

the family are employed. Albeit based on mtDNA sequence data only, this study represents the most comprehensive molecular systematic treatment of this group to date.

Methods

Taxon sampling

Representatives of all serrasalmid genera were sampled from their natural habitat and also obtained from commercial sources (aquarium trade). Several specimens per genus, and in some cases more than one specimen per species were used to confirm taxonomic identifications and also to control for intraspecific variation. Outgroup taxa were chosen from the Anostomoidea and Cynodontidae based on a recent analysis of characiform relationships that suggest a close relationship of these groups to serrsalmids (Calcagnotto *et al.*, 2005). A complete list of taxa used for this study, their associated Genbank accession numbers, their source and (when present) voucher information are presented as Supplementary Material (Table S1).

DNA amplification and sequencing

Genomic DNA was isolated from ethanol-preserved muscle tissue by proteinase K / SDS dissolution, followed by phenol/chloroform extraction and ethanol precipitation (Sambrook *et al.*, 1989). Segments of the small (12S) and large (16S) subunits of the ribosomal RNA mitochondrial genes were amplified by PCR in 50 µL reactions containing 10 µL dNTPs (1 mM), 5 µL reaction buffer (10X), 2 µL $MgCl_2$ (50 mM), 2 µL of each primer (10 µM), 0.5 µL (2.5 U) of Taq DNA polymerase (Gibco BRL), 2 µL of template DNA (100 ng/µL) and 26.5 µL H_2O. PCR conditions were as follows: 94 °C (3 min), 30 cycles of 94 °C (1 min), 57 °C (1 min), 72 °C (1 min), followed by 72 °C (2 min). Primers used for PCR and sequencing of the 12S fragment were L1091 and H1478 (Kocher *et al.*, 1989), and for the 16S fragment, 16Sar-L and 16Sbr-H (Palumbi *et al.*, 1991). These primers amplify fragments of the 12S and 16S rRNA genes corresponding to positions 1091-1478 and 2510-3059 in the human mitochondrial genome, respectively (Anderson *et al.*, 1981). Sequences of the 12S and 16S fragments published by Ortí *et al.* (1996) for 31 serrasalmin taxa were obtained from GenBank, and included in the phylogenetic analyses.

The mitochondrial D-loop region was amplified by PCR in 50 µL reactions containing 10 µL dNTPs (1 mM) , 5 µL reaction buffer (10X), 2 µL $MgCl_2$ (50 mM), 2 µL of each primer (10 µM), 0.5 µL (2.5 U) of Taq DNA polymerase (Gibco BRL), 2 µL of template DNA (100 ng/µL) and 26.5 µL H_2O. PCR conditions were as follows: 94 °C (3 min), 10 cycles of 94 °C (1 min), 53 °C (1 min), 72 °C (1 min), 10 cycles of 94 °C (1 min), 51 °C (30 s), 72 °C (1 min), 10 cycles of 94 °C (1 min), 50 °C (30 s), 72 °C (1 min), followed by 72 °C (2 min). The primers used for PCR and sequencing were designed for this study: F-TTF (5'-GCCTAAGAGCATCGGTCTTGTAA) and F-12R (5'-GTCAGGACCATGCCTTTGTG). Additional internal primers used for sequencing were F-TTF2 (5'-CTAACT CCCAAAGCTAGTATT), F-12R2 (5'-CTACACTAGCT ACAACTATATAA), PM-DLF3 (5'-TAATGCATATTA TCCTTGAT) and F-DLR3 (5'-GTTTTGGGGTTTGACA GGA). These sequences consist of the complete control region (approximately 1100 bp) along with the flanking tRNA genes - about 20 bp of tRNA Thr (3' half), the complete tRNA Pro (approximately 70 bp), and about 65 bp of tRNA Phe (almost complete).

All samples were sequenced using the BigDye Terminator cycle sequencing reaction kit (Applied Biosystems Inc.) on an automated DNA sequencer (Applied Biosystems 310 or an MJ Research Basestation) following manufacturer's instructions. All templates were sequenced completely in both directions. The nucleotide sequence data determined for the present paper were deposited in GenBank (see Supplementary Material, Table S1).

Phylogenetic analysis

All sequences were aligned with Clustal X (Thompson *et al.* 1997) using default parameters. Each fragment (12S, 16S, and D-loop) was aligned separately and the ribosomal gene alignments were subsequently verified using the secondary structure models described by Ortí *et al.* (1996). Alignment gaps that were inserted by ClustalX in putative stem regions that may imply disruption of hairpin structure were moved to contiguous loops or non-paired regions. D-loop sequences were compiled into two separate groups due to alignment ambiguities when alI sequences were aligned together. The two groups (the 'piranha clade' and the rest) differ substantially in total length for this fragment. Micro or minisatellite repeats within the variable control region were identified and excised from the sequences before the alignment. Indels for all resulting alignments were coded for phylogenetic analysis following the modified complex method described by Müller (2006) implemented in the program SeqState (Müller, 2005).

Alignments for each fragment were analyzed initially by the neighbor joining method (NJ: Saitou and Nei, 1987) to control for potential sequencing errors. Sequences that were found misplaced in the resulting tree were re-sequenced or eliminated from subsequent analyses. Given the degree of redundancy in taxonomic sampling, errors can be detected when sequences from putative congeneric or conspecific specimens are not placed together in the tree. Some exceptions to this procedure are discussed below. After the preliminary NJ analyses, the ribosomal 12S and 16S fragments were gauged for congruence in phylogenetic signal by the incongruence length difference (ILD) test (Farris *et al.*, 1994; Farris *et al.*, 1995), implemented as the partition homogeneity test in PAUP* 4.0 (Swofford, 2000). This test showed no significant difference among the two partitions, and the 12S and 16S sequences

were concatenated for all further analyses. The D-loop data were compiled into 2 separate data sets, one for the 'piranha clade' and one for the remaining taxa of the family.

Tree searches were performed using PAUP* version 4.0b4a (Swofford, 2000), MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001, Ronquist and Huelsenbeck, 2003), and TreeFinder (Jobb, 2006). Maximum parsimony (MP) analyses performed in PAUP* used heuristic searches starting with stepwise addition trees and replicated 100 times, with each replicate starting with random input order of sequences. Branch swapping was performed by the tree-bisection-reconnection (TBR) method. The consistency index (CI) and the rescaled consistency index (RC) were computed for the best trees. Bootstrap values (BV) were used to estimate confidence in the resulting topology and were based on 100 replicates of heuristic search with starting trees obtained by stepwise addition. MP analyses were applied to the DNA sequence data alone or in combination with the coded indel characters. Modeltest 3.7 (Posada and Crandall, 1998) was used to determine the optimal model of nucleotide evolution for each data set. Maximum likelihood (ML) searches were performed with TreeFinder specifying the model determined by Modeltest. Bayesian inference (BI) was performed by running 4 MCMC chains simultaneously for 1 million generations, sampling every 100 steps (*i.e.*, saving a total of 10,000 trees and parameter sets). MrBayes 3.1 by default runs two such MCMC chains simultaneously (Nruns = 2) and independently for each run, starting from different random trees. The value of 1 million generations was determined by examination of the average standard deviation of split frequencies (as they approach zero). At least two independent runs were performed to check for convergence. After each run, stationarity was verified by examination of the plot of generation versus the log probability of the data (the log likelihood values) and the burnin value was determined to summarize the results. This value was typically less than 2000 samples, but a conservative value of 5000 was usually chosen. The DNA data were analyzed under the 6-parameter model (Nst = 6) with invariant sites and rate variation (rates = invgamma). Indels were coded as a second partition for the Bayesian analyses under the Standard model of evolution allowing for among site rate variation (rates = gamma), and both partitions were unlinked for the analysis (unlink statefreq = (all) revmat = (all) shape = (all) pinvar = (all)), allowing each to have its own rate (prset applyto = (all) ratepr = variable). A consensus tree was computed using the sumt command and the posterior probabilities (PP) were obtained directly from the frequency of each partition among the post-burnin trees.

Results

12S and 16S data

Mitochondrial rRNA sequences from a total of 74 serrasalmin taxa plus 9 outgroup species were collected (total = 83). The total length of the combined 12S plus 16S alignment was 890 bp (347 bp of 12S and 543 bp of 16S). Length variation among sequences resulted in 58 additional indel characters coded by the modified complex method, 14 of which used a step matrix and the rest were unordered for MP analyses. None of the step matrices coded by SeqState was internally inconsistent. Of the total 948 characters, 580 were constant, 104 were variable but uninformative for parsimony, and 264 were informative. Pairwise sequence divergence ranged from 0 to 0.105 (uncorrected "p" or proportion of sites that differ) among the ingroup taxa, and from 0.055 to 0.144 between serrasalmids and the outgroup taxa. A total of 10 ingroup taxa were excluded from further analyses because their sequences were identical to another taxon that was included (see Supplementary Material, Table S1). Therefore a final data set of 73 taxa was used for phylogenetic inference.

Model parameters for the mtDNA sequence data suggest significant levels of among site rate variation (proportion of invariable sites = 0.51, and alpha = 0.56), typical of ribosomal DNA data. Indel characters also exhibited significant among-site rate variation (alpha = 1.0). The result obtained by Bayesian analysis (Figure 2) agrees with previous results based on mtDNA (Figure 1b) and with the other inference methods used in this study. The Serrasalmidae form a distinct, strongly supported monophyletic group, containing three main clades: (1) a "pacu" clade, comprised of *Colossoma*, *Mylossoma* and *Piaractus* that is the sister group to the other serrasalmids, (2) the *Myleus* clade, containing *Myleus*, *Mylesinus*, *Tometes* and *Ossubtus*; and (3) the "piranha" clade, with the genera *Serrasalmus*, *Pristobrycon*, *Pygocentrus*, *Pygopristis*, *Catoprion* and *Metynnis.* The analyses are not conclusive with regard to the placement of *Acnodon* and also do not support the monophyly of the two species (*A. normani* and *A. oligacanthus*) included in the study. Results from ML analysis are almost identical to the BI tree (Figure 2) differing only in the branching pattern with each major clade (mostly shown as polytomies in Figure 2). ML results also place *A. oligacanthus* within the *Myleus* clade, separate from *A. normani* (outside of the *Myleus* clade). Maximum parsimony analyses yielded 740 equally parsimonious trees (L = 1224, CI = 0.43, RC = 0.32), a strict consensus of which recovers the monophyly of the piranha and the pacu clades, but not of the *Myleus* clade. MP bootstrap analysis yields BV = 56 and 95 for these two clades, respectively, and no support for the *Myleus* clade. Interestingly, the monophyly of *Acnodon* was recovered in several (but less than 50%) of the 740 equally parsimonious trees. *A posteriori* reweighting of characters (based on the RC, Farris, 1969, Carpenter, 1988) reduced the number of MP trees to 81, a strict consensus of which shows the same relationships among the main groups that were obtained with BI (Figure 2) but with a monophyletic *Acnodon* (*A. normani* + *A. oligacanthus*) as the sister group to the *Myleus* clade.

Figure 2 - Phylogeny of the Serrasalmidae obtained with Bayesian inference based on 12S and 16S mtDNA sequences (890 bp) and indel characters (58 characters). The combined data were partitioned into two categories, each with its independent model (Nst = 6 and rates = pinvar for DNA, and the standard model with rates = gamma for indel characters, see text for more details). Numbers next to nodes are posterior probabilities (the lower value of two independent runs are shown) reflecting partition frequencies obtained in the majority rule-consensus of all post-burnin trees. The three main clades are highlighted: (1) pacu clade, (2) *Myleus* calde, (3) piranha clade.

Relationships among taxa within these three groups are relatively well resolved for the pacu and piranha groups but not among taxa in the *Myleus* clade. Within the pacu group, the sister group relationship between *Colossoma* and *Mylossoma* is recovered by ML analysis, and supported by a PP = 0.98 (Figure 2), but it does not receive support from MP bootstrap analysis. The five species of *Metynnis* included in the study form a strong monophyletic group (PP = 1.0, BV = 92) and the genus is well supported (PP = 1.0 and BV = 78) as the sister group to all other taxa in the piranha clade (Figure 2). *Pristobrycon striolatus* forms a distinct taxon, branching off next in the piranha clade, and quite separate from the other putative *Pristobrycon* species (*P. serrulatus and P. eigenmanni*) that group tightly with *Serrasalmus* and *Pygocentrus,* the most derived group within the piranha clade. This result is robust in all methods of analysis and was reported before (Ortí *et al.*1996). The two specimens assigned to *Pygopristis denticulatus* used in this study do not form a monophyletic group in any analysis and their sequence divergence is 0.025 (uncorrected "p" value), a relatively large difference for an intraspecific comparison. As reported earlier (Orti *et al.*, 1996) the relationships among species of *Serrasalmus*, *Pygocentrus* and *Pristobrycon*, are not resolved by the 12S and 16S sequence data.

Mitochondrial control region data

Complete mitochondrial control region sequences from 45 serrasalmid taxa were collected. In agreement with *a priori* expectations (*e.g.*, Meyer, 1993), D-loop sequences display much higher levels of variation than the rRNA genes, with sequence divergences ranging from 0.017 to 0.256 (uncorrected "p" divergence) among serrasalmid taxa. However, this higher level of polymorphism also resulted in problematic sequence alignment. Up to about 500 bp of the 5' end of the control region (Domain I, Brown *et al.*, 1986) contained tandem repeats in several taxa examined (Table 1). The repeated motifs ranged from 12 bp to 33 bp and some repeat patterns were imperfect (some repeats slightly different to the others). These motifs were repeated up to 17 times (in the case of *Catoprion*) and they accounted for most of the variation in length of the amplified fragments. Tandem repeat regions were not used for phylogenetic analyses because their homology among divergent serrasalmid taxa was impossible to assess. After excluding the repeated regions, the D-loop sequences were aligned separately for the "piranha" clade (18 taxa) and the other groups (27 taxa), resulting in alignment lengths of 1130 bp and 1180 bp, respectively. Indels were coded as above, resulting in the addition of 86 or 119 characters for each data set, respectively.

Figure 3 shows the BI tree for the pacu and the *Myleus* clades based on the control region data. The data set analyzed consisted of 1299 total characters, of which 583 were constant, 214 were variable but parsimony-uninformative and 502 were parsimony-informative. MP analysis recov-

Table 1 - Tandem repeats in the 3' region of the Control Region of serrasalmid taxa.

Taxon	Repeat motif	Number of repeats[1]	Length of tandem repeat region
Catoprion	AGTACATATGTATATAGTACATCATGGTTT	17 (p)	510 bp
Pristobrycon striolatus 224,225	AGTACATATTATGTATATAGTACATGATGGTTT	3 or 7 (p)	99 to 231 bp
Metynnis hypsauchen	ATGGTGATCTAAGTACATAATAGTTATATAGTACATA	3 (i)	111 bp
19 *Metynnis*	ATGGTGATCTAAGTACATTATATGTATATAGTACATA	4 (i)	148 bp
20 *Metynnis*	ATGATCTAAATACATTATATGTATATAGTACATA	4 (i)	136 bp
Serrasalmus rhombeus 222, 220	ATGGTGATCTAAGTACATTATATGTATATAGTACATA	3 to 5 (i)	111 to 185 bp
Serrasalmus spilopleura 139	GGCGCCCCACAT	5 (p)	60 bp

[1]Number of perfect (p) or imperfect (i) repeats.

ered a single tree (L = 1941, CI = 0.55, RC = 0.34) that is almost identical to the BI tree, differing only in the branching order among the more derived taxa. MP bootstrap analysis results agree well with PPs obtained with BI (Figure 3). In agreement with the 12S and 16S data, *Mylossoma* and *Colossoma* are placed as sister genera, but unlike the rRNA genes, control region data provide strong support for this relationship (PP = 1.0 and BV = 92). The monophyly of both species of *Acnodon* also is supported strongly by the control region data (posterior prob = 1.0 and MP bootstrap support = 98) and *Acnodon* is placed as the sister group of *Myleus*, *Myloplus*, *Mylesinus*, *Tometes*, and *Ossubtus.* Among these taxa there is not much resolution, except to support a basal position of *Myloplus rhombiodalis* and a robust clade composed of *Myloplus* species, *M. rubripinnis*, *M. asterias*, *M. tiete*, and *M. ternetzi,* that forms the sister group to the rest of the taxa.

Figure 4 shows the BI tree for the piranha clade obtained with control region sequences. The data set analyzed consisted of 1216 total characters (1130 bp and 86 indel characters), of which 697 were constant, 189 were variable but parsimony-uninformative and 330 were parsimony-informative. MP analysis recovered a single tree (L = 976, CI = 0.58, RC = 0.44) that is almost identical to the BI tree, differing only in the branching order among *Serrasalmus*, *Pygocentrus* and (non-striolatus) *Pristobrycon* taxa. MP bootstrap analysis results agree well with PPs obtained with BI (Figure 4). The presence of three divergent groups of piranhas is well supported: (1) the genus *Metynnis*, (2) the *Catoprion-Pygopristis-Pristobrycon striolatus* group and (3) the *Serrasalmus-Pygocentrus*-group (*Pristobrycon* species other than *Pristobrycon striolatus,* such as *P. serrulatus* and *P. eigenmanni*, are here assigned to *Serrasalmus*). The control region data resolved with high confidence (PP = 1.0 and BV = 100) the relationships among *Catoprion*, *P. striolatus*, and *Pygopristis* that were not fully resolved in the 12S and 16S tree (Fig 2). Within the *Serrasalmus* clade, there is weak evidence (PP = 0.53) for affinities among *Pygocentrus* and one group of *Serrasalmus* species (*S. manueli, S. maculatus,* and *S. rhombeus*), and somewhat higher support (PP = 0.83 and BV = 54) for affinities among putative *Pristobrycon* (non-striolatus) with a different group of *Serrasalmus* species (*S. gouldingi, S. serrulatus,* and S. *eigenmanni*).

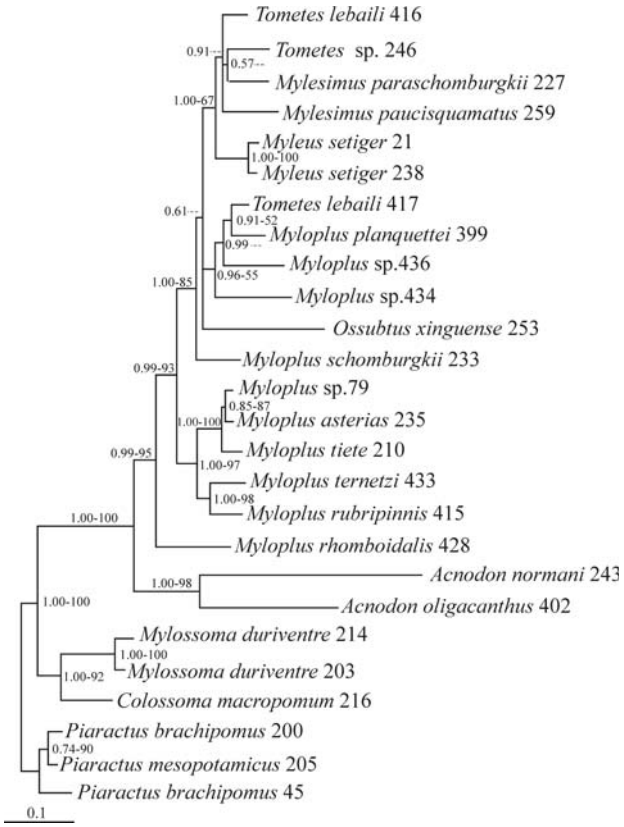

Figure 3 - Phylogeny for the pacu and *Myleus* groups, obtained with Bayesian inference based on control region mtDNA sequences (1180 bp) and indel characters (119 characters). The combined data were partitioned into two categories, each with its independent model (Nst = 6 and rates = pinvar for DNA, and the standard model with rates = gamma for indel characters, see text for more details). Numbers next to nodes are posterior probabilities followed by a dash and bootstrap values (BV) from the maximum parsimony analysis.

Discussion

This study represents the most complete molecular systematic treatment of serrasalmids to date. Building on

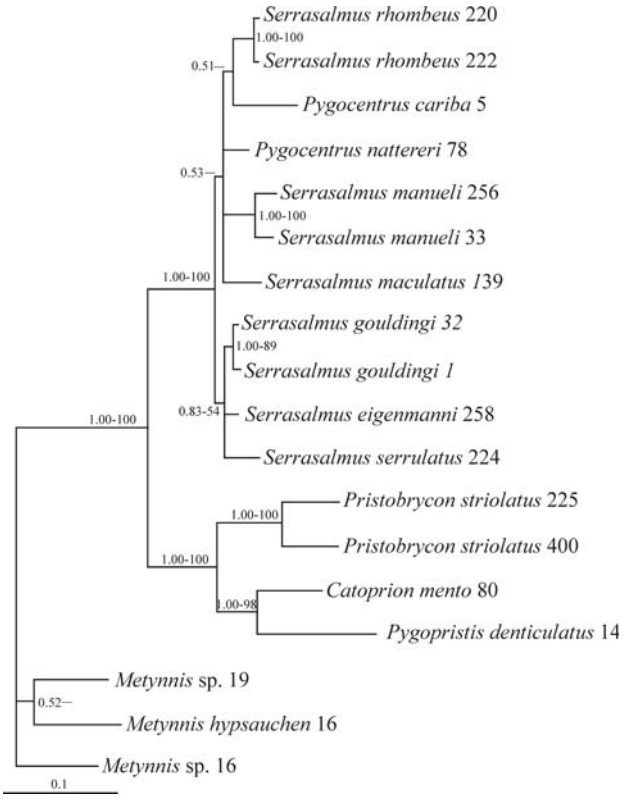

Figure 4 - Phylogeny of the piranha clade, obtained with Bayesian inference based on control region mtDNA sequences (1216 total characters, 1130 bp and 86 indel characters). The combined data were partitioned into two categories, each with its independent model (Nst = 6 and rates = pinvar for DNA, and the standard model with rates = gamma for indel characters, see text for more details). Numbers next to nodes are posterior probabilities followed by a dash and bootstrap values (BV) from the maximum parsimony analysis.

the previous study by Ortí *et al.* (1996), analyses presented here include additional taxa from all genera of the subfamily and add a new molecular marker (complete mtDNA control region sequences) to recover phylogenetic patterns or population structure among serrasalmids. As expected, the higher level of variation in the control region compared to the 12S and 16S mitochondrial rRNA genes provides better resolution of the relationships within each of the main clades (Figures 3 and 4). In agreement with previous estimates of rates of substitution-control region rates were found to range between 2.8 (Cann *et al.*, 1984) to 5 times (Aquadro and Greenberg, 1983) the rate of the rest of the mtDNA genome-the higher rate observed for control region among serrasalmid taxa provided additional characters for phylogenetic inference of closely related species. The study also documents complex patterns of variation involving tandem repeats in the mitochondrial control region. This phenomenon has been generally accepted to account for size variation among vertebrate mitochondrial genomes (reviewed by Hoelzel, 1993, Rand, 1993) and has been was documented in fishes before (*e.g.* Bentzen *et al.*, 1998).

Although based on a single molecular marker (mtDNA), the results of this study carry several taxonomic implications. Most notably, many of the generic designations in the family seem to lack support or are clearly contradicted by the data. Some of these conclusions are not new: *Pristobrycon striolatus* has previously been regarded as quite distinct from its congeners (Machado-Allison *et al.*, 1989), differing in several morphological aspects and its well-supported grouping with *Catoprion* and *Pygopristis* is consistent with the finding of Ortí *et al.* (1996). Our present results confirm this observation and therefore we prefer to restrict *Pristobrycon* to the single species *P. striolatus*, and place all other taxa previously assigned to this genus in *Serrasalmus*. According to the classification of Géry (1977), the genus *Serrasalmus* contained the subgenera *Pygopristis*, *Pristobrycon*, *Pygocentrus*, *Taddyella* and the nominate subgenus *Serrasalmus*; *Serrasalmus* (*Pristobrycon*) *striolatus* was noted to resemble closely the subgenus *Pygopristis*. This observation is well supported by our molecular analysis of control region data, as this species forms a clade with *Catoprion* and *Pygopristis* (Figure 4), and is not closely related to the other specimen putatively assigned to *Pristobrycon* (#224 designated *Serrasalmus serrulatus* here) in the rRNA tree (Figure 2). Based on various morphological characters, *Serrasalmus gouldingi* is distinct from other members of the genus (Machado-Allison and Fink, 1996). In this analysis, it was found to be more closely related to the remaining *Pristobrycon* than it is to other species of *Serrasalmus*. This group containing *S. gouldingi, S. eigenmanni* and *S. serrulatus* is the sister group to the *Serrasalmus-Pygocentrus* clade. The genus *Serrasalmus* contains within it the genus *Pygocentrus*. Results from analysis of control region sequences of a dense taxonomic sampling for *Serrasalmus* and *Pygocentrus* provides strong evidence for the monophyly of *Pygocentrus* but its relationship to diverse components of *Serrasalmus* remains unresolved (Hubert *et al.*, 2007). Some of the poor resolution obtained in our study is evidently the consequence of poor taxonomic sampling.

Some authors (*e.g.* Géry, 1977) have recognized the existence of four subgenera within *Myleus*, namely *Myloplus*, *Paramyloplus*, *Prosomyleus* and the nominate subgenus *Myleus*, within this genus. These subgeneric distinctions have been, as with all previous classifications, based primarily on dental morphology. Other authors, however, rejected these subgeneric distinctions due to the lack of autapomoprhies (Machado-Allison and Fink, 1995). The monophyly of subgenera within *Myleus* is not supported by analyses of mtDNA data. Analysis of the *Myleus* group reveals the polyphyly of the formerly designated genus *Myleus* and supports the taxonomic rearrangement proposed by Jégu and Dos Santos (2002) and Jégu *et al.* (2003), but relationships among the various components of this group remain tentative. The group formed by *Myleus*

setiger with *Mylesinus* and *Tometes* is relatively well-supported (PP = 1.00, BV = 67, Figure 3) suggesting strong affinities of *Myleus* with species designated to these genera. A robust group of *Myloplus* species (*M. rubripinnis*, *M. asterias, M. tiete*, and *M. ternetzi*) is also well supported by the control region data.

As these analyses have shown, there are several taxonomic inconsistencies in this subfamily. While this study represents the most comprehensive molecular systematic treatment of this group, and utilizes a highly variable mtDNA marker to provide resolution of shallow nodes, placement of some taxa remains uncertain. In order to provide a strong foundation for taxonomic revision of the group, future studies would benefit from utilizing dense taxonomic sampling, nuclear gene sequences, together with mtDNA and morphological characters to help resolve some of these ambiguous relationships.

Acknowledgments

Many colleagues provided tissue samples for this study, but in particular we would like to thank Jorge Porto and Paulo Petry for their contributions. This work is part of Arjun Sivasundar's MS thesis. Financial support came from the National Science Foundation (USA) through a CAREER Grant (DEB 9985045) to G.O.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, *et al.* (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457-465.

Aquadro CF and Greenberg BD (1983) Human mitochondrial DNA variation and evolution: Analysis of nucleotide sequences from seven individuals. Genetics 103:287-312.

Araujo-Lima C and Goulding M (1997) So Fruitful a Fish: Ecology, Conservation, and Aquaculture of the Amazon's Tambaqui. Columbia University Press, New York, 191 pp.

Bentzen P, Wright JM, Bryden LT, Sargent M and Zwanenburg KC (1998) Tandem repeat polymorphism and heteroplasmy in the mitochondrial control region of redfishes (Sebastes, Scorpaenidae). J Hered 89:1-7.

Brown GG, Gadaleta G, Pepe G, Saccone C and SbisE (1986) Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J Mol Biol 192:503-511.

Calcagnotto D, Schaefer SA and DeSalle R (2005) Relationships among characiform fishes inferred from analysis of nuclear and mitochondrial gene sequences. Mol Phylogenet Evol 36:135-153.

Cann RL, Brown WM and Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106:479-499.

Carpenter J (1988) Choosing among equally parsimonious cladograms. Cladistics 4:291-296.

Eigenmann C (1915) The Serrasalminae and Mylinae. Ann Carnegie Mus Pittsburgh 9:266-272.

Farris JS (1969) A successive approximations approach to character weighting. Syst Zool 18:374-385.

Farris JS, Källersjö M, Kluge AG and Bult C (1994) Testing significance of incongruence. Cladistics 10:315-319.

Farris JS, Källersjö M, Kluge AG and Bult C (1995) Constructing a significance test for incongruence. Syst Biol 44:570-572.

Géry J (1972) Poissons Characoïdes des Guyanes. I. Généralités. II. Famille des Serrasalmidae. Zool Verhand 122:1-250.

Géry J (1977) Characoids of the World. T.F.H. Publications Inc, Neptune City, 672 pp.

Géry J (1984) The fishes of Amazonia. In: Sioli H (ed) The Amazon, Limnology and Landscape Ecology of a Mighty Tropical River and its Basin. Junk Publishers, Dordrecht, pp 343-370.

Gosline W (1951) Notes on the characoid fishes of the Subfamily Serrasalminae. Proc Cal Acad Sci ser 4 27:17-64.

Goulding M (1980) The Fishes and the Forest: Explorations in Amazonian Natural History. University of California Press, Berkeley, 280 pp.

Hoelzel AR (1993) Evolution by DNA turnover in the control region of vertebrate mitochondrial DNA. Curr Opin Genetics 3:891-895.

Hubert N, Bonillo C and Paugy D (2005) Does elision account for molecular saturation: Case study based on mitochondrial ribosomal DNA among Characiform fishes (Teleostei, Ostariophysii). Mol Phylogenet Evol 35:300-308.

Hubert N, Duponchelle F, Nuñez J, Garcia-Dávila C, Paugy D and Renno J-F (2007) Phylogeography of the piranha genera *Serrasalmus* and *Pygocentrus*: Implications for the diversification of the Neotropical ichthyofauna. Mol Ecol 16:2115-2136.

Huelsenbeck JP and Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754-755.

Jégu M and Dos Santos GM (1990) Description d'*Acnodon senai* n. sp. du Rio Jari (Brésil, Amapá) et redescription d'*A.normani* (Teleostei, Serrasalmidae). Cybium 14:187-206.

Jégu M and Dos Santos GM (2002) Révision du statut de *Myleus setiger* Müller and Troschel, 1844 et de *Myleus knerii* (Steindachner, 1881) (Teleostei, Characidae, Serrasalminae) avec une description complémentaire des deux espèces. Cybium 26:33-57.

Jégu M, Keith P and Le Bail PY (2003) *Myloplus planquettei* n. sp. (Teleostei, Characidae, Serrasalminae), une nouvelle espèce de grand Serrasalminae phytophage du bouclier guyanais (Guyane française). Rev Suisse Zool 110:823-853.

Jobb G (2006) TREEFINDER, v. of May 2006. Munich, Germany, http://www.treefinder.de.

Junk WJ (1984) Ecology, fisheries and fish culture in Amazonia. In: Sioli H (ed) The Amazon, Limnology and Landscape Ecology of a Mighty Tropical River and its Basin. Dr W Junk Publishers, Dordrecht, pp 443-476.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX and Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals. Proc Natl Acad Sci USA 86:6196-6200.

Leite RG and Jégu M (1990) Food habits of two species of *Acnodon* (Characiformes, Serrasalmidae) and scale-eating habits of *Acnodon normani*. Cybium 14:353-360.

Lowe-McConnel RH (1975) Fish Communities in Tropical Freshwaters: Their Distribution, Ecology and Evolution. Longman, London, 337 pp.

Machado-Allison A (1983) Estudios sobre la sistemática de la subfamilia Serrasalminae (Teleostei, Characidae). Parte II. Discusión sobre la condición monofilética de la subfamilia. Acta Biol Venez 11:145-195.

Machado-Allison A and Fink WL (1995) Sinopsis de las Especies de la Subfamilia Serrasalminae Presentes en la Cuenca del Orinoco. Serie Peces de Venezuela. Museo de Biologia, Caracas, 89 pp.

Machado-Allison A and Fink WL (1996) Los Peces Caribes de Venezuela: Diagnosis, Claves, Aspectos Ecologicos y Evolutivos. Universidad Central de Venezuela, Caracas, 149 pp.

Machado-Allison A, Fink WL and Antonio ME (1989) Revisión del género *Serrasalmus* Lacepede, 1803 y géneros relacionados en Venezuela: I. Notas sobre la morfología y sistemática de *Pristobrycon striolatus* (Steindachner, 1908). Acta Biol Venez 12:140-171.

Marshall E (1995) Homely fish draws attention to Amazon deforestation. Science 267:814.

Meyer A (1993) Evolution of mitochondrial DNA of fishes. In: Hochachka PW and Mommsen P (eds) Molecular Biology Frontiers, Biochemistry and Molecular Biology of Fishes. Elsevier Press, Amsterdam, pp 1-38.

Müller K (2005) SeqState - Primer design and sequence statistics for phylogenetic DNA data sets. Appl Bioinf 4:65-69.

Müller K (2006) Incorporating information from length-mutational events into phylogenetic analysis. Mol Phylogenet Evol 38:667-676.

Nico L and Taphorn DC (1988) Food habits of piranhas in the low llanos of Venezuela. Biotropica 20:311-321.

Norman JR (1929) The South American characid fishes of the subfamily Serrasalmoninae with a revision of the genus *Serrasalmus* Lacepede. Proc Zool Soc London 52:661-1044.

Ortí G and Meyer A (1997) The radiation of characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. Syst Biol 46:75-100.

Ortí G, Petry P, Porto JIR, Jégu M and Meyer A (1996) Patterns of nucleotide change in mitochondrial ribosomal RNA genes and the phylogeny of piranhas. J Mol Evol 42:169-182.

Palumbi S, Martin A, Romano S, McMillan WO, Stice L and Grabowski G (1991) The Simple Fool's Guide to PCR. University of Hawaii, Honolulu, 46 pp.

Posada D and Crandall KA (1998) Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817-818.

Rand DM (1993) Endotherms, ectotherms, and mitochondrial genome-size variation. J Mol Evol 37:281-295.

Ronquist F and Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Swofford DL (2000) PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), v. 4. Sinauer Assoc, Sunderland.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876-4882.

Winemiller KO (1989) Ontogenetic diet shifts and resource partitioning among piscivorous fishes in the Venezuelan Llanos. Env Biol Fishes 26:177-199.

Zanata AM (2000) Estudo das relações Filogenéticas do gênero *Brycon* Müller and Troschel, 1844 (Characidae, Characiformes). Ph.D. Thesis, Universidade de São Paulo, São Paulo.

Supplementary Material

The following online material is available for this article:

- Table S1: Specimen and sequence information

This material is available as part of the online article from http://www.scielo.br/gmb.

Associate Editor: Antônio Sole-Cava