Molecular Phylogenetics and Evolution 59 (2011) 177–194

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Molecular systematics of the neotropical shovelnose catfish genus *Pseudoplatystoma* Bleeker 1862 based on nuclear and mtDNA markers

L.F. Carvalho-Costa [a], N.M. Piorski [b], S.C. Willis [c], P.M. Galetti Jr. [d], G. Ortí [e,*]

[a] Center for Agrarian and Environmental Sciences, Universidade Federal de Maranhão, Chapadinha, MA, Brazil
[b] Department of Oceanography and Limnology, Universidade Federal de Maranhão, São Luis, MA, Brazil
[c] School of Biological Sciences, University of Nebraska-Lincoln, USA
[d] Department of Genetics and Evolution, Universidade Federal de São Carlos, São Carlos, SP, Brazil
[e] Department of Biological Sciences, The George Washington University, 2023 G Street, NW, Washington, DC, USA

ARTICLE INFO

Article history:
Received 26 March 2010
Revised 27 January 2011
Accepted 2 February 2011
Available online 16 February 2011

Keywords:
Species tree
Coalescent analysis
Cytochrome b
Rag1
S7
Intron

ABSTRACT

Pseudoplatystoma is a commercially important genus of Neotropical migratory catfishes widely distributed in all major river basins of South America. Historically, only three species were recognized, but a recent revision proposed eight putative morphospecies for the genus. A molecular study based on mitochondria DNA (mtDNA) provided support for recognition of only some of the species and raised questions about species boundaries in this group. We present a more encompassing analysis based on mtDNA (cytochrome b, 818 bp) and nuclear DNA-based phylogenies (Rag1 intron 1, 664 bp and S7 intron 1, 635 bp) for a more extensive sampling (279 individuals from 42 localities) of all putative species in all major river basins. Patterns generated by individual gene genealogies and a multispecies coalescent analysis provided evidence to suggest recognition of only four distinct species in this genus: *Pseudoplatystoma magdaleniatum*, *Pseudoplatystoma corruscans*, *Pseudoplatystoma tigrimun* (*sensu lato*) and *Pseudoplatystoma fasciatum* (*sensu lato*). The species phylogeny places *P. magdaleniatum* as the sister group to all the other species in the genus, but the relationships among *P. fasciatum s.l*, *P. tigrimum s.l.*, and *P. corruscans* could not be resolved with confidence.

1. Introduction

The Neotropics harbour the most diverse ichthyofauna of the world, representing 46% of all freshwater fish species and 10% of extant vertebrates (Vari and Malabarba, 1998; Lundberg et al., 2000; Reis et al., 2003). Among the freshwater fishes, catfishes form one the most diverse groups, including the endemic family Pimelodidae, composed by more than 90 species and 30 genera (Lundberg and Littmann, 2003). *Pseudoplatystoma* Bleeker 1862 (Bleeker, 1862) is an economically important pimelodid genus from South America, with species reaching large size. Their piscivorous and migratory habits make them key top-down regulators of the trophic structure of all major river drainages (Burgess, 1989; Reid, 1983; Barthem and Goulding, 1997), but current anthropogenic activities in most of South America, such as damming and overfishing, are reducing the effective size of their populations and raising conservationist concerns.

A recent revisionary study of *Pseudoplatystoma* (Buitrago-Suárez and Burr, 2007) proposed that eight species should be recognized in this genus. Before this publication, most taxonomists recognized only three, that could be distinguished mainly by coloration pattern (Lundberg and Littmann, 2003; Burgess, 1989): (i) *Pseudoplatystoma corruscans* (Spix and Agassiz, 1829), the spotted surubim or "pintado", distributed in the São Francisco and Paraná–Paraguay–Uruguay basins; (ii) *Pseudoplatystoma fasciatum* (Linnaeus, 1766), the striped catfish ("bagre rayado" or "surubim/cachara/pintadillo"), is widely distributed in the Amazonas, Magdalena, Orinoco, Paraná–Paraguay, and drainages from Guyana and Suriname and Northeastern Brazil; and (iii) *Pseudoplatystoma tigrinum* (Valenciennes, 1840), the tiger surubim or "caparari", distributed in both Amazonas and Orinoco basins. External morphological differences among these species involve minor variation in body shape so the species are usually recognized by coloration (dark vertical bars, loops, or spots), as described. In their recent study, Buitrago-Suárez and Burr (2007), recognize eight species based on these characters plus evidence from skeletal anatomy, vertebral numbers, and geographic distribution (namely, river basin). On this basis, the widely distributed striped surubim

* Corresponding author. Address: 2023 G Street NW, 411 Lisner Hall, Washington DC 20052, USA. Fax: +1 202 994 6100.
E-mail addresses: lfcc@ufma.br (L.F. Carvalho-Costa), stuartcwillis@gmail.com (S.C. Willis), galettip@ufscar.br (P.M. Galetti), gorti@gwu.edu, guilleorti@gmail.com (G. Ortí).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.02.005

previously assigned to *P. fasciatum* has been split into five species: *P. fasciatum* (*sensu stricto*): restricted to Guyana and Suriname river basins; *Pseudoplatystoma punctifer* (Casteunau, 1855): present in the Amazonas basin and, presumably, in Northeastern Brazil drainages; *Pseudoplatystoma reticulatum* (Eigenmann and Eigenmann, 1889): distributed widely in the Paraná–Paraguay system and presumably also restricted to an area of the central Amazonas river (near Manaus); *Pseudoplatystoma orinocoense*: restricted to the Orinoco basin; and *Pseudoplatystoma magdaleniatum*: distributed only in the Magdalena basin. The tiger catfish (*P. tigrinum*) populations from the Orinoco drainage were assigned to a new species (*Pseudoplatystoma metaense*), restricting the name *P. tigrinum* to the tiger catfish of the Amazonas basin. The species status of *P. corruscans* and its geograpbhic distribution remained unchanged.

Two years later, Torrico et al. (2009) assessed the taxonomy and phylogeny of *Pseudoplatystoma* based on mitochondrial DNA (mtDNA) markers (cytochrome *b* and control region sequences). Their study supported four clades of mitochondrial haplotypes, corresponding to: (i) *P. tigrinum*, (i) *P. corruscans*, (iii) *P. reticulatum* + *P. punctifer* + *P. fasciatum,* and (iv) *P. magdaleniatum* (following the taxonomy proposed by Buitrago-Suárez and Burr). In addition, for some of the included species the mtDNA haplotypes did not form monophyletic groups, casting doubt about their taxonomic status; however, many important issues could not be addressed by this study due to incomplete taxonomic sampling and evidence limited to a single molecular locus (mtDNA).

In the present study, we present new evidence to assess the taxonomy and systematics of this group by using two nuclear genes in addition to mtDNA markers. We include specimens obtained from an extensive sampling of all major river basins in South America where *Pseudoplatystoma* is distributed, covering some critical drainages that were not sampled in the previous studies. In addition to analyzing single gene genealogies, we use a recently developed coalescent approach to estimate a species trees from multilocus data (Heled and Drummond, 2010).

2. Materials and methods

Pseudoplatystoma specimens were sampled from all major river basins where the genus occurs, except the Gurupi and Munin rivers (Northeastern Brazil) and some drainages from Guyana and Suriname (Table 1, Fig. 1). Samples were identified based on the old taxonomy, the only one available at the time of sampling. Considering that the new taxonomy is generally concordant with species distribution (river basin), we reassigned species names based on their geographic distribution (Buitrago-Suárez and Burr, 2007). This allowed comparison of the old and new taxonomies in the light of the molecular phylogenies unraveled in our study. Voucher information is provided in Appendix D. DNA was extracted from fin clips or muscle tissues following a rapid salt-extraction protocol (Aljanabi and Martinez, 1997) or using the DNeasy Blood and Tissue Kit (*QIAGEN* Inc.).

The complete mitochondrial Cytochrome *b* gene (Cytb, 1140 bp) and two nuclear gene fragments containing intron sequences were analyzed in this study. The nuclear genes targeted were Rag1 (recombination activating gene intron one; Rag1int1, 800 bp) and S7 ribosomal gene (first intron, S7int1, 1000 bp). All gene fragments were amplified via PCR (polymerase chain reaction), in a final reaction volume of 30 μl containing 10–50 ηg of DNA, 200 μM of dNTPs (dATP, dGTP, dCTP and dTTP), 1× PCR buffer, 0.2 μM of each primer, 1.5–2.0 mM of $MgCl_2$ and 1 U of Taq polymerase (Invitrogen Life Technologies). All reactions were conducted in either PTC100 (MJ Research) or Mastercycler (Eppendorf) thermo cyclers, following the amplification program: 94 °C for 5 min, 35 cycles of 94 °C for 30 s, annealing temperature (Table 2) for 35 s, 72 °C for 40 s and a final extension step (72 °C for 5 min). PCR primers for Cytb and S7int1 were available from the literature (Table 2), while those for Rag1int1 were designed using exons 1 and 2 sequences available on GenBank for other fishes. PCR products were purified (Illustra GFX PCR DNA and Gel band purification, GE HealthCare) and sequenced on a MegaBACE 1000 (GE HealthCare) using EtDye kit (GE HealthCare). All fragments were directly sequenced using forward and reverse primers. Sequences were checked on Codoncode 3.0 (Codoncode Corp.) and alignedusing Clustalw (Thompson et al., 1994) with Bioedit (Hall, 1999). Heterozygous individuals for a single nucleotide site at nuclear loci were resolved manually. Otherwise, haplotype phases were determined using Phase 2.1 (Stephens et al., 2001; Stephens and Donnelly, 2003) that implements a Bayesian method for haplotype reconstruction. Posterior probabilities higher than 0.9 were accepted if repeated throughout five runs of 500 iterations each. Possible instances of recombination events on nuclear alelles were assessed with Rdp3 3.34 (Martin et al., 2005) using the RDP and Bootscan/RECSCAN methods (Salminen et al., 1995; Padidam et al., 1999), the method applied in the program GENECONV (Padidam et al., 1999; Sawyer, 1989), the MaxChi method (Maynard-Smith, 1992; Posada and Crandall, 2001), and the 3SEQ approach (Posada and Crandall, 2001).

Overall variation among DNA sequences was characterized using Mega 4.0 (Tamura et al., 2007) and DnaSP 4.10 (Rozas et al., 2003). Uncorrected *p*-distances were calculated with Mega and neutrality tests (Tajima's *D*, F^*, and D^*) (Tajima, 1989; Fu and Li, 1993) were conducted using DnaSP 4.10. Maximum likelihood (ML) gene genealogies were estimated for each gene separately using Treefinder (Jobb, 2008). Each data set was reduced to distinct haplotypes or alleles for each gene before phylogenetic analysis (the full list of "collapsed" alleles/haplotypes is shown in Appendices A–C). The evolutionary model for each gene also was estimated using Treefinder, based on the Akaike Information Criterion (AIC). Cytochrome b sequences were partitioned by grouping 1st and 2nd codon positions into one partition and 3rd codon positions into another, each fitted with an independent model. Sequences from *Brachyplatystoma* and *Zungaro* were used as outgroup. Statistical support for nodes was estimated by bootstrap analysis with 1000 pseudoreplicates (Felsenstein, 1985). Topological tests were conducted for each gene genealogy *a posteriori* to compare the best tree to alternative hypotheses using the Shimodaira and Hasegawa (SH) and the Approximately Unbiased tests (AU) Shimodaira and Hasegawa, 1999; Shimodaira, 2002; both testing procedures are available in Treefinder and use the RELL technique (Kishino et al., 1990).

The species tree was estimated under a Bayesian Markov chain Monte Carlo method for the multispecies coalescent with the *Beast protocol (Heled and Drummond, 2010) implemented in the program BEAST v 1.5.3 (Drummond and Rambaut, 2007a). All sequences obtained in the study for all genes and individuals were used as input for this analysis. This includes redundant haplotypes and alleles. For the nuclear genes, both alleles for each individual were included in the input file, even for homozygous individuals. The species from which each allele/haplotype was obtained was used as the trait value for the species tree and specified using the program BEAUTi, a simple user interface for creating input files to run BEAST. Other parameters for this analysis included specifying a relaxed clock and coding for ploydy types (haploid for mtDNA and diploid autosomal for the nuclear genes). Population sizes were assumed to be constant and all other priors used default conditions. The xml file used as input is available upon request from the corresponding author. Two runs, each of 10 million generations, were conducted and stationarity was checked using the program TRACER (Drummond and Rambaut, 2007b), a graphical tool for visualization and diagnostics of MCMC output.