Nucleotide diversity of *Hemigrammus* cf. *marginatus* (Characiformes, Characidae) in the upper Paraná river floodplain

Thiago Cintra Maniglia¹²*, Alberto José Prioli¹, Thais Souto Bignotto³, Talge Alex Boni¹, Sônia Maria Alves Pinto Prioli¹ and Carla Simone Pavanelli¹

¹Núcleo de Pesquisa em Limnologia Ictiologia e Aquicultura, Universidade Estadual de Maringá, Av. Colombo, 5790, 87020-900, Maringá, Paraná, Brazil. ²Universidade Paranaense, Cascavel, Paraná, Brazil. ³Programa de Pós-graduação em Melhoramento Genético, Universidade Estadual de Maringá, Maringá, Paraná, Brazil. *Author for correspondence. E-mail: tmaniglia@gmail.com

**ABSTRACT.** Characidae is the largest and more diversified family from Characiformes and presents several classification problems, with several genera currently allocated as *incertae sedis*, such as the genus *Hemigrammus*. The upper Paraná river floodplain is an environment with high fish diversity. There is at least one species of *Hemigrammus*, however there are divergences among some authors about the number and the identification of the species from this genus. Therefore the goal of this study was to characterize, using a molecular approach, individuals of *Hemigrammus* from the upper Paraná river floodplain and to compare them with individuals from the type locality of *Hemigrammus marginatus*, since this is the only species distributed in this floodplain. For this, the DNA was extracted and a partial region from the mitochondrial genes ATPase 6 and ATPase 8 were amplified and sequenced. The results evidenced the existence of two species of *Hemigrammus* in the floodplain, although impossible to be distinguished only through morphological traits. High nucleotide diversity among individuals from the upper Paraná river in relation to those from the type locality was also observed, indicating that both species of *Hemigrammus* present in the upper Paraná river floodplain are not *Hemigrammus marginatus*.

**Keywords:** DNA, Mitochondrial, ATPase 6, ATPase 8, nucleotidic diversity.

Diversidade nucleotídica de *Hemigrammus* cf. *marginatus* (Characiformes, Characidae) na planície de inundação do alto rio Paraná

RESUMO. Characidae é a maior e mais diversificada família de Characiformes e apresenta vários problemas de classificação, com inúmeros géneros alocados atualmente como *incertae sedis*, dentre estes *Hemigrammus*. A planície de inundação do alto rio Paraná é um ambiente com alta diversidade de peixes. Existe neste ambiente pelo menos uma espécie de *Hemigrammus*, entretanto, existem divergências entre alguns autores quanto ao número e a identificação das espécies deste gênero. Portanto, o objetivo deste trabalho foi realizar a caracterização molecular de indivíduos de *Hemigrammus* da planície de inundação do alto rio Paraná e compará-los com exemplares da localidade-tipo de *Hemigrammus marginatus*, tendo em vista ser esta a única espécie identificada de *Hemigrammus* com distribuição na referida planície. Para isso, foi extraído o DNA, amplificado e sequenciado uma região parcial dos genes mitocondriais ATPase 6 e ATPase 8. Os resultados demonstraram a existência de duas espécies de *Hemigrammus* na planície de inundação do alto rio Paraná, embora impossível de serem diferenciadas apenas pelos caracteres morfológicos utilizados atualmente. Alta diversidade nucleotídica entre os exemplares do alto rio Paraná em relação aos da localidade-tipo também foi observada, indicando que ambas as espécies de *Hemigrammus* presente na planície de inundação do alto rio Paraná não são da espécie *Hemigrammus marginatus*.

**Palavras-chave:** DNA, Mitochondrial, ATPase 6, ATPase 8, diversidade nucleotídica.

Introduction

The order Characiformes is one of the largest group of freshwater fish worldwide including 18 families, about 270 genera and at least 1,674 species (NELSON, 2006). These organisms present a wide diversity of forms, inhabiting in various types of aquatic environments. Characidae is the largest and more diversified family, with 12 subfamilies, 16 genera and more than 962 species (NELSON, 2006). From the 165 genera, 88 are allocated as *incertae sedis* (LIMA et al., 2003). Great part of these genera belonged to the subfamily Tetragonopterinae, originally proposed by Günther (1864). The genera *incertae sedis* within Characidae comprise about 620 species, 43 from the genus *Hemigrammus*, Gill, 1858 (LIMA et al., 2003).

Among the species of *Hemigrammus*, *H. marginatus* Ellis, 1911 presents wide geographical distribution and is found in the watershed of the rivers São Francisco, Itapicuru (type locality), Paraná, Paraguai, Guaporé, Amazonas and Orinoco (LIMA et al., 2003). The
taxonomy of the genus *Hemigrammus* from the upper Paraná river is problematic, both regarding the number of species as the identification. According to Agostinho et al. (1997, 2004) there are two species of *Hemigrammus* in the upper Paraná river floodplain, *H. marginatus* and *Hemigrammus* sp. Otherwise, Graça and Pavanelli (2007) recorded only the species *H. marginatus* in this same floodplain. Nevertheless, Portela-Castro and Júlio Junior. (2002) registered cytogenetic divergences between the populations of *H. marginatus* from the upper Paraná river floodplain with individuals from the São Francisco river, suggesting that they may be distinct species. Recently, Marinho et al. (2008), described one new species of *Hemigrammus*, named *H. parana*, present at the upper Paraná river.

The DNA polymorphism, resulting from mutation and recombination that may accumulate due to geographical and reproductive isolation, associated to morphological and ecological studies have assisted the recognition of the genetic and ecological diversity and the evolutionary process in different taxonomic groups. This approach reveals the evolutionary path from several species, establishing kinship levels and mapping genetic traits from each group.

In the last decades, mainly with the advent of PCR (Polymerase Chain Reaction), studies using molecular markers led the population genetics to achieve a huge impact on biology (SUNNUCKS, 2000). The molecular markers also aid the systematic of fishes (PRIOLI et al., 2002). For the studies concerning the speciation and population differentiation, the analyses of mitochondrial DNA sequences (mtDNA) are an excellent starting point (AVISE, 2004).

The mitochondrial genes ATPase 6 and ATPase 8 have as characteristic the ability to accumulate nucleotide substitutions that allow detecting genetic variations between species, or even between populations from a same species (BERMINGHAM; MARTIN, 1998; MACHORDON; DOADRIO, 2001; PERDICES; DOADRIO, 2001; WONG et al, 2004). Despite being a coding region, the rate of nucleotide substitution of the region that comprises these genes is comparable or even superior in some cases to the control region (*D-loop*), considered the most variable portion of the mtDNA (FROUFE et al., 2005; FAULKS et al., 2008).

The recognition of cryptic species is one of the significant applications of molecular markers (SOLÉ-CAVA, 2001). Therefore, molecular analyses, comparing the populations from the type locality with other watersheds, may provide important information to solve taxonomical problems. In this way, the present study compared, using molecular techniques, individuals of *Hemigrammus cf. marginatus* from the type locality with individuals from the upper Paraná river floodplain, in order to assist the correct identification of this species.

**Material and methods**

For this study, we collected 11 specimens of *Hemigrammus* sp. in the upper Paraná river floodplain, near Porto Rico, Paraná State (22°45'S and 53°30'W), and two specimens of *H. marginatus* in Itapicuru river, Bahia State (between 10°30'S and 11°05'S and 40°08'W and 40°30'W (Figure 1). The individuals were preserved in alcohol 96% until DNA extraction.

![Figure 1. Map of South America, indicating the two sites of collection of specimens of *Hemigrammus*. (1) upper Paraná river floodplain (Paraná State); (2) Itapicuru river (Bahia State).](image)

Total DNA was extracted based on the phenol-chloroform method, according to Sambrook et al. (1989), with some modifications. After the extraction, the DNA was quantified in 0.8% agarose gel, compared with the φ phage DNA with known concentration. The DNA samples, after extraction, were stored in a freezer -20°C.

The genes ATPase 6 and ATPase 8 and the part of the genes tRNA^Thr^ and COIII were amplified via PCR with the same amplification conditions proposed by Prioli et al. (2002). The temperature patterns of amplification were as follows: an initial cycle at 94°C
for four minutes, followed by 40 cycles of 15 seconds at 94°C, 30 seconds at 59°C and two minutes at 72°C, ending with a cycle of 10 minutes at 72°C. The amplification was confirmed by electrophoresis on 1% agarose gel comparing with known quantities of 100 bp DNA ladder (Invitrogen). The pair of primers used for the amplification was the L8331 (5'-AAAGCRTYRGCCTTTTAAGC-3’) and H9236 (5'-GTTAGTGGTCAKGGGGTCGGRT-3’).

After the amplification, the fragment was again amplified with the primer L8331 for the sequencing reaction. Approximately 50 ng of DNA from each reaction were directly used in sequencing reactions using MegaBace automatic sequencer (Amersham), according to manufacturer instructions.

The sequences were aligned using Clustal W program (THOMPSON et al., 1994) and edited with the program Bioedit (HALL, 1999). From the comparisons between the pairs, we accomplished the phylogenetic analyses. The choice of the evolutionary model, using the procedures Akaike Information Criterion corrected (AICc) and Bayesian Information Criterion (BIC), was performed with the programs Paup 4.0b4 (SWOFFORD, 2002) and Modeltest 3.0 (POSADA; CRANDALL, 1998). The molecular index $(F_\infty)$ was calculated using the program Arlequin 3.11 (EXCOFFIER et al., 2005). The nucleotide diversity matrix and the dendrograms Neighbor-Joining and Maximum Likelihood were made with the program Paup 4.0b4. The Bayesian dendrogram was undertaken with 300,000 generations with the program Mr. Bayes 3.0 (HUELSENBECK; RONQUIST, 2001). The dispersal graph at main coordinates was built from the matrix and the dendrograms Neighbor-Joining and Maximum Likelihood were built from this model.

The values of $F_a$ for the three groups of Hemigrammus were 0.81 between the two groups from the upper Paraná river, Hemigrammus sp.1 and Hemigrammus sp.2, 0.87 between Hemigrammus sp.1 and H. marginatus and 0.93 between Hemigrammus sp.2 and H. marginatus. These values indicate that at least 81% from genetic diversity was found among individuals from different groups.

The results were employed to generate the bars graphic with the mean values of nucleotide diversity of the species obtained at GenBank and the specimens of Hemigrammus.

**Results**

The amplification of the partial sequence of the genes ATPase 6 and ATPase 8 produced a fragment with approximately 1,000 base pairs (bp). Nevertheless, after the sequences edition, we obtained a partial fragment of 398 bp for all individuals. This fragment corresponds to a sequence of 132 amino acids. We verified 55 nucleotide substitutions. From these, 25 are exclusive of the individuals of H. marginatus from Itapicuru river, and 20 of the individuals from the upper Paraná river floodplain. At this environment, there were two groups of individuals with several exclusive substitutions in each one. Therefore, these specimens were separated into two groups, Hemigrammus sp.1 and Hemigrammus sp.2. We registered eight exclusive substitutions for Hemigrammus sp.1 and 13 for Hemigrammus sp.2. Moreover, 44 amino acid substitution were recorded; 22 were exclusive of H. marginatus from Itapicuru river, nine of Hemigrammus sp.2 and eight of Hemigrammus sp.1.

The sequence of 398 bp presented a proportion of nucleotide bases of $A = 0.3174; C = 0.2253; G = 0.1215$ and $T = 0.3357$, a $T_i/T_v$ ratio of 2.8164, the Nst was 2 with gamma distribution value of 0.2005. Thus, the analysis to select the evolutionary model revealed that the HKY + G is the model that presents the highest probability of explaining the nucleotide substitutions for the individuals of Hemigrammus here analyzed. Consequently, the distance matrix and the dendrograms Neighbor-Joining and Maximum Likelihood were built from this model.

The Table 1 contains the mean values of nucleotide diversity calculated with the model HKY + G, among the three groups of Hemigrammus previously mentioned and the external group (Outgroup) A. mexicanus. The lowest and the highest diversity value among the three groups of Hemigrammus was 0.066 and 0.11, respectively. These values associated to the high $F_a$ values evidence the genetic differentiation existing among the groups.

**Table 1.** Matrix of the mean values from genetic distances between the individuals of Hemigrammus from rivers Paraná (1- Hemigrammus sp.1 and 2- Hemigrammus sp.2) and Itapicuru (3- H. marginatus) and the outgroup (4- A. mexicanus) calculated with the model HKY + G, from the partial fragment of 398 bp of the genes ATPase 6, 8.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.066</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.372</td>
<td>0.348</td>
<td>0.334</td>
</tr>
</tbody>
</table>
The mean values of nucleotide diversity from the three groups of *Hemigrammus* and the species obtained at GenBank are represented in the Figure 2. The selected species were: *Brycon petrosus* (AF412681), *Brycon oligolepis* (AF412670), *Brycon obscurus* (AF412667), *Brycon meeki* (AF412658), *Brycon argenteus* (AF412653), *Bryconamericus scopiferus* (AF412627), *Bryconamericus emperor* (AF412623), *Roeboides dayi* (AF040527), *Roeboides meeki* (AF040524), *Roeboides occidentalis* (AF040521), *Roeboides guatemalensis* (AF040517), *Prochilodus lineatus* (AF281839), *Prochilodus nigricans* (AF281842), *Prochilodus magdalenae* (AF281847) and *Prochilodus mariae* (AF281844).

The groupings Neighbor-Joining, Maximum Likelihood, Bayesian and the dispersal graph at main coordinates corroborated the formation of three groups (Figures 3A, 3B and 4). One group was formed with the individuals from Itapicuru river, and two with the individuals from upper Paraná river. High bootstrap values provide consistence to these groupings.

**Figure 2.** Mean values of nucleotide diversity, calculated with the model HKY + G, from the partial fragment of 398 bp of the genes ATPase 6, 8 for different species from the genera *Hemigrammus*, *Brycon*, *Bryconamericus*, *Roeboides* and *Prochilodus*.

**Figure 3.** Neighbor-Joining dendrogram (A) and Maximum Likelihood (B), built with the model HKY + G, from nucleotide sequences of the partial genes ATPase 6 and ATPase 8 from individuals of *Hemigrammus* sp. from the Paraná river watershed, and individuals of *H. marginatus* from the Itapicuru river watershed. One individual of *Astyanax mexicanus* was included in the analysis as an external group. The bootstrap analyses were based on 10,000 resamplings.

**Figure 4.** Bayesian dendrogram constructed with the model HKY + G, from nucleotide sequences of the partial genes ATPase 6 and ATPase 8 from individuals of *Hemigrammus* sp. from the Paraná river watershed, and individual of *H. marginatus* from the Itapicuru river watershed. The dendrogram was built with 300,000 generations, under the simulations of Markov Chain Monte Carlo Simulation (MCMC). The first 20,000 generations were not included in the analyses. The values represent the probabilities of the groupings remain after 300,000 generations.
The dispersal graph at main coordinates, built with the two major eigenvectors, also evidenced the existence of the same three groups of *Hemigrammus* (Figure 5).

Figure 5. Dispersal graph at main coordinates of the individuals from the three groups of *Hemigrammus* from the upper Paraná river basin and Itapicuru river basin, based on the eigenvectors, form the matrix of genetic 6,03 7,57 distance calculated with the model HKY + G.

Discussion

Numerous studies were already made, using the genes ATPase 6 and ATPase 8 with fish at different taxonomical level, with different approaches. At lower taxonomic levels, with closer phylogenetic relationships, such as populations from a same species, subspecies or species complex, the values of nucleotide diversity ranges from 0.002 to 0.054 (FAULKS et al., 2008; KINZIGER et al., 2007; MACHORDON; DOADRIO, 2001; PERDICES; DOADRIO, 2001; SIVASUNDAR et al., 2001). Among species of a same genus, the values observed in literature are between 0.015 and 0.13 (PERDICES; DOADRIO, 2001; MACHORDON; DOADRIO, 2001; SIVASUNDAR et al., 2001; FROUFE et al., 2005; REID; WILSON, 2006). For higher taxonomic levels, e.g., between species of different genera, the values of genetic diversity are also greater, varying between 0.062 and 0.16 (SIVASUNDAR et al., 2001; FROUFE et al., 2005). Comparing the nucleotide diversity between the species of *Hemigrammus* examined in the present study and the species of Characiformes obtained at Genbank, it was possible to observe that the mean diversity of *Hemigrammus* (0.091) is the second highest, only smaller than between *Brycon* species (Figure 2). This result indicates that the diversity registered among the individuals of *Hemigrammus* is in accordance to the diversity found among different species of Characiformes. Among the five species of *Brycon*, the values of nucleotide diversity ranged from 0.015 to 0.155. For two species of *Bryconamericus* the value was 0.056. With five species of *Roeboides*, the values were at least 0.010 and at most 0.128. Among four species of *Prochilodus* the values varied between 0.015 and 0.051.

The values obtained for the groups of *Hemigrammus* in the present study were 0.066 between the two groups from the upper Paraná river and 0.097, and 0.11 between *H. marginatus* and *Hemigrammus* sp.1, and *Hemigrammus* sp.2, respectively. Based on the comparisons performed both with published studies as with the sequences available at GenBank, the values obtained for *Hemigrammus* are quite superior to those found among individuals within a species, subspecies or species complex, and below those established among species of different genera, but consistent with the values between different species of a same genus. Therefore, the obtained values indicate that the three groups observed in this study represent three distinct species from a same genus.

The type locality of *H. marginatus* is the Itapicuru river, thus, the two specimens taken in this river must be identified as *H. marginatus*. In this way, we did not register any exemplar of *H. marginatus* in the upper Paraná river. As expected, the two groups found at the upper Paraná river are of the genus *Hemigrammus*, but certainly they are not *H. marginatus*. This result is in accordance to those obtained by Portela-Castro and Júlio Junior. (2002) that suspected that specimens of *Hemigrammus* from the upper Paraná river would not be *H. marginatus*.

Since it was not possible to identify the exemplars from the upper Paraná river, they were classified as *Hemigrammus* sp.1 and *Hemigrammus* sp.2. Although *H. parana* is exclusive from the influence area of the Ilha Solteira reservoir, upper Paraná river, according to Marinho et al. (2008), we should not reject the hypothesis that one of the two species found at the upper Paraná river floodplain is *H. parana*. Further surveys, with new sampling areas must be accomplished in order to test this hypothesis and to determine the geographical distribution of species of *Hemigrammus* identified for the upper Paraná river floodplain.

Conclusion

In this study, we evidenced the efficient role of the genes ATPase 6 and ATPase 8 in detecting of genetic diversity between closely related species of
fish, including when only morphological traits are not enough for the correct identification. Meanwhile the use of molecular markers without the help of taxonomy is not enough to understand the biological diversity. In this way, the results indicate the need of taxonomic researches for the genus *Hemigrammus* in the upper Paraná river floodplain.

The upper Paraná river floodplain is an environment with high biological diversity. However, the present study indicated that this diversity is still underestimated, probably due to the high degree of spatial complexity of this ecosystem. The identification of new species, including endemic ones, emphasizes the importance of this ecosystem as a priority area in the conservation of Neotropical biodiversity.

**Acknowledgements**

The authors thank to Dra. Ângela Maria Zanata for gathering *Hemigrammus marginatus* at Itapicuru river; to Nupélia-UEM for logistic support and to PELD/CNPq and CAPES for financial support.

**References**


Received on March 17, 2009.
Accepted on July 26, 2010.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.