

INSTITUTO NACIONAL DE PESQUISAS DA AMAZÔNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA, CONSERVAÇÃO E
BIOLOGIA EVOLUTIVA

VALIDAÇÃO DO *DNA BARCODING* COMO IDENTIFICADOR DE ESPÉCIES: UM
ESTUDO DE AMPLA AMOSTRAGEM COM O GÊNERO *Pseudoplatystoma*
(SILURIFORMES; PIMELODIDAE) NA AMAZÔNIA

MAURO DE FREITAS ORTIZ

Manaus, Amazonas

Março, 2010

MAURO DE FREITAS ORTIZ

VALIDAÇÃO DO *DNA BARCODING* COMO IDENTIFICADOR DE ESPÉCIES: UM
ESTUDO DE AMPLA AMOSTRAGEM COM O GÊNERO *Pseudoplatystoma*
(SILURIFORMES; PIMELODIDAE) NA AMAZÔNIA

ORIENTADOR: JOSÉ ANTÔNIO ALVES GOMES

CO-ORIENTADOR: TOMAS HRBEK

Dissertação apresentada ao
Programa de Pós-Graduação em
Genética, Conservação e Biologia
Evolutiva do Instituto Nacional de
Pesquisas da Amazônia como parte
dos requisitos para obtenção do
título de mestre em Genética,
Conservação e Biologia Evolutiva.

Manaus, Amazonas

Março, 2010

Ficha Catalográfica

O77

Ortiz, Mauro de Freitas

Validação do DNA Barcode como identificador de espécies: um estudo de ampla amostragem com o gênero *Pseudoplatystoma* (Siluriformes, Pimelodidae) na Amazônia / Mauro de Freitas Ortiz. --- Manaus : [s.n.], 2010.

xv, 73 f. : il. color.

Dissertação (mestrado)-- INPA, Manaus, 2010

Orientador : José Antonio Alves- Gomes

Co-orientador : Tomas Hrbek

Área de concentração : Genética, Conservação e Biologia Evolutiva

1. *Pseudoplatystoma*. 2. Genética de populações. 3. Marcadores genéticos. 4. DNA Barcoding. 5. Pimelodidae. I. Título.

CDD 19. ed. 597.520415

Sinopse:

Foram realizados testes para validação do *DNA Barcoding* como identificador de espécies usando-se como modelo bagres amazônicos do gênero *Pseudoplatystoma*. Em um primeiro momento o *DNA Barcoding* foi testado com espécimens referenciais e depois então este foi aplicado a um número maior de espécimens. O *DNA Barcoding* conseguiu diferenciar todas as espécies de *Pseudoplatystoma* e ficou evidente a presença do *barcoding gap*, assim demonstrando seu potencial na identificação de espécies, em especial de peixes amazônicos.

Palavras-chave: citocromo oxidase I, bagres neotropicais, código de barras de DNA.

Dedico esta dissertação
aos meus pais, Deomar e
Marlene, à minha irmã
Simone e a todos os ateus.

Agradecimentos

Primeiramente, quero deixar claro que esta dissertação não tem nenhum tipo de envolvimento com qualquer tipo de entidade superior onipresente, onisciente, onipotente ou com qualquer tipo de atributo sobrenatural. Assim como nenhum tipo de relação com credices, superstições ou qualquer tipo de simpatia ou crença que ainda não tenha sido devidamente justificada segundo os critérios do método científico.

Agradeço ao INPA e em especial à coordenação e secretaria do PPG-GCBEv por todo o suporte durante os dois anos de mestrado.

À CAPES pela concessão da bolsa de estudos durante este período.

Ao Dr. José Antônio Alves-Gomes e Dr. Tomas Hrbek pela orientação.

À Dra. Izeni Pires Farias pelas sugestões e a acolhida em seu laboratório.

Aos professores do programa pelas disciplinas ofertadas.

À Dra. Lúcia Rapp, Dra. Izeni Farias e Dr. Jorge Porto pelas contribuições na aula de qualificação.

Aos avaliadores de minha dissertação, Dra. Maria Cláudia Gross, Dra. Camila Ribas e Dr. Jorge Porto (titulares) e Dra. Marina Anciães e Dr. Charles Clement (suplentes) pelas contribuições para a versão final da dissertação.

Aos colegas da turma Alexandre, Daniela, Deyla, Edson, Edvaldo, Gabriela, Graciela, Joel, Mellina e em especial Bárbara e Mariana que de alguma forma ou outra ajudaram a enfrentar essa empreitada.

Aos colegas de laboratório LTBM Andréa, Adriel, Carol, Fabiola, Giuliano, Iamili, Karen, Keroleny, Kyara, Larissa, Lucy, Naiara, Rosa, Saulo, Vanessa e em especial Jacqueline e Giselle por toda a ajuda e companhia nesses dois anos que compartilhamos a bancada.

Aos participantes do Projeto PIRADA que diretamente ou indiretamente ajudaram na obtenção de muitas das amostras utilizadas neste trabalho.

Aos colegas Legais Adam, Daniela, Deyla, Eduardo, Edvaldo, Fabiola, Ígor, Kelmer, Marina, Mário, Natasha, Nicole, Pedro, Rafaela, Valeria, Waleska pela

companhia e ajuda nas diversas vezes que levei placas de sequenciamento para UFAM.

Aos amigos e colegas Camila Burlamaqui, Camila Ritter, Carlos S., Izaura, Leandra, Leila, Alexandra, Marlos, Jarbas, Patrícia, Carlos N., Gabriela Moraes, Maní, Beto, pela companhia e conversas.

Àqueles que comigo vieram de tão longe Cristian, Anderson, Ígor e Daiani.

A Bethânia pelo incentivo desde o início. Por toda o apoio no difícil início em Manaus e grande amizade. Pela certeza de que muitas dessas realizações foram possíveis por sua causa. Um carinhoso muito obrigado!

Ao Dr. Gilson Volpato pelos fantásticos ensinamentos em seus cursos.

Ao Dr. Élgion Loreto que me acolheu no seu laboratório durante a graduação e me ensinou muito do que eu sei, possibilitando assim grande parte da minha independência durante este mestrado e pela compreensão a minha escolha do incerto.

E em especial à minha maravilhosa família pelo apoio incondicional e principalmente pela compreensão na minha escolha de estudar e trabalhar tão longe de casa.

Peço desculpas àquelas pessoas que eu eventualmente tenha esquecido de citar.

Certamente, este mestrado só foi concluído, especialmente, em função da ajuda destas pessoas. Muito obrigado!

“Ou o século XXI é dedicado aos valores humanos, morais e éticos... ou de nada valeram os avanços tecnológicos conquistados até aqui.”

Gilson Volpato

RESUMO

O *DNA Barcoding* é uma metodologia na qual se pressupõe que um pequeno fragmento de DNA pode ser utilizado como um identificador padronizado e único, tal como um código de barras, para identificação de espécies, servindo tanto na identificação de espécies já descritas, como auxiliando na descoberta de novas espécies. O principal pressuposto para a sua efetividade é de que as divergências intraespecíficas sempre sejam menores que as interespecíficas, o que se chama de *barcoding gap*. A monofilia recíproca das espécies seria então o principal determinante para a existência deste. Entre os fatores que irão determinar a existência da monofilia recíproca e consequentemente eficiência do *DNA Barcoding* estão a diversidade intraespecífica e a estruturação populacional. Poucos testes empíricos para avaliação do *DNA Barcoding* foram realizados no intuito de validá-lo. Neste trabalho foi usado um gênero de peixes neotropicais como modelo em testes de validação do *DNA Barcoding*. Num primeiro momento foi referendado um banco de dados que continha apenas sequências de amostras com referenciais em imagens. Este obteve 100% de identificação correta entre as identificações *a priori* e o resultado da identificação *a posteriori* do *DNA Barcoding*. Desta forma podemos usar o *DNA Barcoding* para identificação *a posteriori* de amostras sem referenciais em imagens ou até mesmo de origem desconhecida com confiança. Então, num segundo momento, foram acrescidas as sequências de amostras sem referências em imagens e o banco de dados somou 229 sequências. Este obteve 91,26% de identificação correta entre as espécies identificadas *a posteriori* pelas sequências e a identificação atribuída *a priori*. O aparente resultado de ineficiência do *DNA Barcoding* foi atribuído a erros de identificação *a priori* em virtude da confiabilidade do primeiro banco, aos altos valores de confiança que sustentam os clados na árvores de distâncias e a clara presença do *barcoding gap*. Todas as espécies de *Pseudoplatystoma* mostraram-se monofiléticos e as divergências intraespecíficas sempre foram inferiores as interespecíficas, ou seja, foi verificado o *barcoding gap* em todos os casos. Desta forma, este trabalho contribui na validação da eficácia do *DNA Barcoding* na identificação de espécies, em especial de peixes.

ABSTRACT

DNA Barcoding is a methodology in which one assumes that a small fragment of DNA can be used as a standard and unique identifier, such as a barcode for species identification, serving both in the identification of species already described, and in assisting in discovery of new species. The main precondition for its effectiveness is that intraspecific differences are always smaller than interspecific, which is called the barcoding gap. The reciprocal monophyly of species would then be the main determinant for its existence. Among the factors that will determine the existence of reciprocal monophyly and hence efficiency of DNA Barcoding are intraspecific diversity and population structure. Few empirical tests of DNA Barcoding were undertaken to validate it and its premises. In this study we used a genus of Neotropical fishes as a model for validating DNA Barcoding. In the first series of tests we used a database that contained only sequences from samples with references to images. This yielded 100% correct identification identifications between a priori identification and a posteriori identification based on DNA Barcoding. Thus we can confidently use DNA Barcoding for the a posteriori identification of samples without reference to images or even samples of unknown origin. In the second set of tests unreferenced sequences were added to database; the sequence database totaled 229 samples. In this study we obtained 91.26% correct identification of a priory identified species. The apparent inefficiency of DNA Barcoding has been attributed to errors in identifying specimens a priori because of the reliability of the first database, the high confidence values that support the clades in the first analysis, and the clear presence of the barcoding gap. All species of *Pseudoplatystoma* proved to be monophyletic and intraspecific differences were always lower than the interspecific, i.e. the barcoding gap was observed in all cases. Thus, this study contributes to the validation of the efficacy of DNA Barcoding in identifying species, especially fishes, in the Amazon basin.

SUMÁRIO

LISTA DE TABELAS.....	xi
LISTA DE FIGURAS.....	xiii
INTRODUÇÃO.....	1
DNA Barcoding.....	1
<i>Pseudoplatystoma</i>	6
OBJETIVOS.....	7
MATERIAIS E MÉTODOS.....	7
Material de estudo.....	7
Isolamento e amplificação do DNA.....	11
Purificação e sequenciamento do DNA.....	11
Análise das sequências.....	12
Validação do DNA Barcoding.....	13
RESULTADOS.....	14
Banco de Dados.....	14
Diversidade Genética.....	18
Validação do DNA Barcoding.....	21
<i>Pseudoplatystoma</i>	29
DISCUSSÃO.....	30
Banco de Dados.....	30
Diversidade Genética.....	31
Validação do DNA Barcoding.....	32
Teste para validação.....	34
<i>Pseudoplatystoma</i>	38
CONCLUSÕES.....	39
REFERÊNCIAS BIBLIOGRÁFICAS.....	40
ANEXOS.....	46

LISTA DE TABELAS

Tabela 1. Localização dos pontos de amostragem e o número de amostras sequenciadas para cada espécie (banco padrão).....	9
Tabela 2. Localização dos pontos de amostragem e o número de amostras sequenciadas para cada espécie (banco completo).....	10
Tabela 3. Distância média entre as espécies de <i>Pseudoplatystoma</i> e em relação aos grupos externos obtida no programa MEGA (banco padrão).....	14
Tabela 4. Composição nucleotídica do gene COI (mínima, média, máxima e desvio padrão) do banco padrão.....	16
Tabela 5. Valores médios das distâncias intra e interespecíficas do gene COI (distância mínima, média, máxima e desvio padrão) do banco padrão calculados no BOLD (K2P).....	17
Tabela 6. Composição nucleotídica do gene COI (mínima, média, máxima e desvio padrão) do banco completo.....	18
Tabela 7. Índices de polimorfismo de DNA das quatro espécies de <i>Pseudoplatystoma</i> analisadas e testes de neutralidade. N (tamanho da amostra), H (número de haplótipos), S (sítios polimórficos), ETA (número de mutações), HD (diversidade haplotípica), PI (diversidade nucleotídica), K (médias das diferenças par-a-par).....	20

Tabela 8. Valores médios das distâncias intra e interespecíficas do gene COI (distância mínima, média, máxima e desvio padrão) do banco completo calculados no BOLD (K2P).....20

Tabela 9. Distância máxima intra e interespecífica e desvio padrão, intervalos de confiança de 95% e 99% para subamostragens com 5, 10 e todas as amostras do banco com referências em imagens. Ao lado do nome das espécies aparece o número amostral total e ainda os intervalos de confiança (IC) (banco padrão).....23

Tabela 10. Distância média intra e interespecífica e desvio padrão, intervalos de confiança de 95% e 99% para subamostragens com 5, 10 e todas as amostras do banco com referências em imagens. Ao lado do nome das espécies aparece o número amostral total e ainda os intervalos de confiança (IC) (banco padrão).....24

Tabela 11. Distância máxima intra e interespecífica e desvio padrão, intervalos de confiança de 95% e 99% para subamostragens com 5, 10 e todas as amostras. Ao lado do nome das espécies aparece o número amostral total e ainda os intervalos de confiança (IC) (banco completo).....27

Tabela 12. Distância média intra e interespecífica e desvio padrão, intervalos de confiança de 95% e 99% para subamostragens com 5, 10 e todas as amostras. Ao lado do nome das espécies aparece o número amostral total e ainda os intervalos de confiança (IC) (banco completo).....28

LISTA DE FIGURAS

Figura 1. Analogia entre o <i>DNA Barcode</i> e o código de barras convencional.....	3
Figura 2. Esquema do <i>barcoding gap</i> . A distribuição da variação intraespecífica é mostrada em vermelho e da interespecífica em amarelo. (A) Mundo ideal para <i>DNA Barcoding</i> , com distribuições discretas e sem sobreposição. (B) Uma versão alternativa do mundo com significativa sobreposição, sem <i>gap</i> (Fonte: Meyer e Paulay, 2005).....	4
Figura 3. Exemplares do gênero <i>Pseudoplatystoma</i> , <i>P. fasciatum</i> (acima) e <i>P. tigrinum</i> (abaixo) (Modificado de Santos <i>et al.</i> , 2006).....	6
Figura 4. Mapa das localidades de coleta na bacia amazônica (banco padrão). Localidades: 1 – Coari, 2 – Manaus, 3 – Tabatinga, 4 – Tefé, 5 – Boa Vista, 6 – Eirunepé, 7 – Porto Velho, 8 – Manicoré, 9 – Guajará Mirim, 10 – São Gabriel da Cachoeira, 11 – Lábrea.....	8
Figura 5. Mapa das localidades de coleta na bacia amazônica (banco completo). Localidades: 1 – Belém, 2 – Coari, 3 – Manaus, 4 – Santarém, 5 – Tabatinga, 6 – Tefé, 7 – São Félix do Araguaia, 8 – Itaipava, 9 – Boa Vista, 10 – Rio Iriri, 11 – Eirunepé, 12 – Porto Velho, 13 – Manicoré, 14 – Guajará Mirim, 15 – Riberalta, 16 – São Gabriel da Cachoeira, 17 – Boca do Acre, 18 – Ipiranga, 19 – Lábrea, 20 – Rio Branco, 21 – Jacareacanga, no alto a esquerda – Rio Magdalena.....	9
Figura 6. Árvore Neighbour-joining da divergência nucleotídica das sequências do COI entre três espécies do gênero <i>Pseudoplatystoma</i> e dois grupos externos, usando o modelo de distância Kimura-2-parâmetros com análise de bootstrap 1000 réplicas.....	15

Figura 7. Distribuição das distâncias intra (A) e interespecíficas (B) do gene COI do banco padrão.....	17
Figura 8. Gráfico do número de substituições nucleotídicas (transições – s, e transversões – v) contra a distância genética do gene COI.....	19
Figura 9. Distribuição das distâncias das espécies de <i>Pseudoplatystoma</i> intra (A) e interespecíficas (B) do gene COI do banco completo.....	21
Figura 10. Representação do <i>barcoding gap</i> entre as espécies de <i>Pseudoplatystoma</i> . A esquerda os resultados dos teste com reamostragens de 5 indivíduos e à direita de 10 indivíduos (banco padrão).....	22
Figura 11. Representação do <i>barcoding gap</i> entre as espécies de <i>Pseudoplatystoma</i> . A esquerda os resultados dos teste com reamostragens de 5 indivíduos e à direita de 10 indivíduos (banco completo).....	26
Figura 12. Representação do <i>barcoding gap</i> entre as espécies de <i>Pseudoplatystoma</i> . Acima reamostragens de uma população da cada espécie (<i>P. fasciatum</i> e <i>P. tigrinum</i>) e abaixo reamostragens com duas populações de cada. A esquerda os resultados dos teste com reamostragens de 5 indivíduos e à direita de 10 indivíduos (banco completo).....	29
Figura 13. Diagrama dos testes de monofilia e estruturação dos grupos mostrando que o acréscimo de amostras ou pontos amostrais sempre será do sentido de falsear as hipótese de que os grupos sejam monofiléticos ou estruturados.....	35
Figura 14. Diagrama do desenho amostral do teste de validação do <i>DNA Barcoding</i>	36

INTRODUÇÃO

DNA Barcoding

Estamos em uma corrida entre a descrição da biodiversidade da Terra e sua extinção (Lawton *et al.*, 1998). Estima-se que a maior parte das espécies ainda não foi descrita (May, 1988). Citando como exemplo os peixes, os vertebrados de maior sucesso evolutivo da Terra (ao menos quando considerado o número de espécies), vemos uma impressionante diversidade. Atualmente existem cerca de 31400 espécies registradas no *FishBase* (10/2009), a maior base de dados online com informações sobre peixes, e este número continua aumentando. Estima-se que cerca de 13000 sejam de água doce, sendo que destas mais de 2400 estão presentes na província amazônica, das quais mais de 2000 são endêmicas (Lévêque *et al.*, 2008). Por outro lado, se ainda mal foram descritas todas as espécies do nosso planeta, as taxas de extinção estão sendo elevadas vertiginosamente. Hoje elas já são cerca de 100 vezes maiores do que no passado e as previsões indicam que ainda serão de 1000 a 10000 vezes maiores no futuro (Millennium Ecosystem Assessment, 2005).

A necessidade de se poder identificar espécies de uma forma mais rápida e eficiente também está vinculada a uma série de situações relacionadas à biologia, desde a questão de saúde, de pragas da agricultura, de identificação de espécies exóticas, de monitoramento ambiental, de manejo de fauna, no turismo, no controle de comércio de produtos biológicos e/ou organismos protegidos e de manejo de estoques (Hebert *et al.*, 2003).

Assim, diante de um cenário como este, de grande biodiversidade e eminente perigo de sua perda, um método rápido e padronizado que auxilie na caracterização dessa biodiversidade é essencial. Foi então com esse intuito que o “código de barras de DNA” ou *DNA Barcoding* foi proposto (Hebert *et al.*, 2003). Este pressupõe que um pequeno fragmento de DNA pode ser utilizado como um identificador

padronizado e único, tal como, um código de barras para identificação de espécies, servindo tanto na identificação de espécies já descritas como auxiliando na descoberta de novas espécies (Figura 1). Para animais, o fragmento em questão é uma região de 648 pares de bases do gene mitocondrial citocromo oxidase subunidade I (COI) (Hebert *et al.*, 2003). O gene COI codifica parte de uma enzima terminal da cadeia respiratória da mitocôndria. Para outros organismos, somente agora que os melhores segmentos e genes para serem usados como *DNA Barcoding* estão sendo estabelecidos, como nas plantas (Kress *et al.*, 2005), macroalgas (Saunders, 2005), fungos (Summerbell *et al.*, 2005), protozoários (Scicluna *et al.*, 2006) e bactérias (Sogin *et al.*, 2006).

O principal pressuposto para a efetividade do *DNA Barcoding* é de que as divergências intraespecíficas sempre sejam menores que as interespecíficas, o que se chama de *barcoding gap* (Figura 2). A monofilia recíproca das espécies seria o principal determinante para a existência deste. De fato, diversos estudos já demonstraram a aplicação com sucesso do *DNA Barcoding* nos mais diversos grupos: mamíferos (Clare *et al.*, 2007; Borisenko *et al.*, 2008), aves (Kerr *et al.*, 2007; Kerr *et al.*, 2009), peixes (Ward *et al.*, 2005; Steinke *et al.*, 2009), insetos (Hebert *et al.*, 2009) e nemátodos (Elasser *et al.*, 2009). Estes resultados, em geral, apontam para resoluções acima de 95% na identificação das espécies. Os resultados do projeto *Fish Barcode of Life*, por exemplo, indicam que o *DNA Barcoding* possibilitou a identificação de cerca de 93% das espécies de água doce estudadas até o momento (Ward *et al.*, 2009). Inclusive, diversas novas espécies foram confirmadas por análises taxonômicas integradas (Ward *et al.*, 2009). No entanto, existem outros estudos que apontam para uma menor eficiência do *DNA Barcoding* na identificação dos organismos, como foram nos casos de alguns lepidópteros (Kandul *et al.*, 2004; Elias *et al.*, 2007; Wiemers e Fiedler, 2007) e dípteros (Meier *et al.*, 2006).

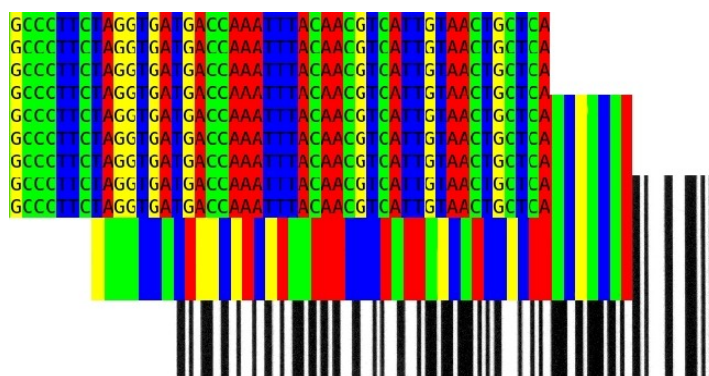


Figura 1. Analogia entre o *DNA Barcode* e o código de barras convencional.

Em virtude destes resultados díspares e das formas de interpretá-los, desde sua proposição, o debate acerca de sua eficiência e funcionalidade é intenso (Moritz e Cicero, 2004; Meyer e Paulay, 2005; Dasmahapatra e Mallet, 2006; Hickerson *et al.*, 2006; Meier *et al.*, 2006; Elias *et al.*, 2007; Wiemers e Fiedler, 2007). Para os propositores e defensores da abordagem, os casos de não resolução geralmente são atribuídos a problemas taxonômicos, que podem ser tanto na própria identificação dos espécimens como na classificação das espécies, como nos casos de espécies crípticas quando tende-se a subestimar o número de espécies, ou ainda espécies polimórficas em que o número seria superestimado. Alguns casos também seriam em função de especiações muito recentes, de forma que as espécies ainda estariam compartilhando haplótipos (Hebert *et al.*, 2009; Lukhtanov *et al.*, 2009). Já para os críticos, a abordagem possuiria limitações intrínsecas que comprometeriam sua funcionalidade de maneira significativa (Meyer e Paulay, 2005; Dasmahapatra e Mallet, 2006; Hickerson *et al.*, 2006; Wiemers e Fiedler, 2007). O que pode se ver é que muitas das críticas são em função das amostragens dos trabalhos (Meyer e Paulay, 2005; Dasmahapatra e Mallet, 2006; Hickerson *et al.*, 2006).

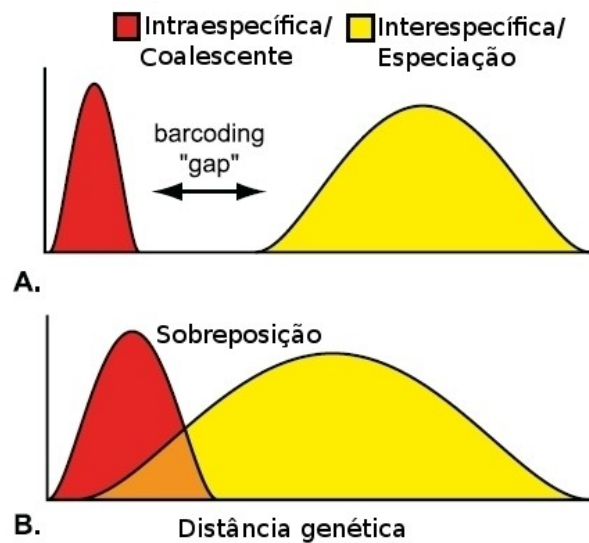


Figura 2. Esquema do barcoding gap. A distribuição da variação intraespecífica é mostrada em vermelho e da interespecífica em amarelo. (A) Mundo ideal para DNA Barcoding, com distribuições discretas e sem sobreposição. (B) Uma versão alternativa do mundo com significativa sobreposição, sem gap (Fonte: Meyer e Paulay, 2005).

Entre os fatores que irão determinar a existência da monofilia recíproca e consequentemente eficiência do *DNA Barcoding* estão a diversidade intraespecífica e a estruturação populacional. A diversidade intraespecífica é inversamente proporcional ao tamanho do *barcoding gap* e esta pode até mesmo fazer com que este seja ausente, no caso de uma elevada diversidade intraespecífica. A estrutura populacional também pode afetar de maneira importante o *barcoding gap* na medida que esta irá afetar a diversidade intraespecífica. Por isso, o papel da amostragem para efetividade do *DNA Barcoding* é de grande importância (Meyer e Paulay, 2005). Isso porque tanto o número amostral e o número de localidades de coleta pode interferir nos resultados uma vez que a subamostragem de espécimens pode subestimar as divergências intraespecíficas e superestimar as divergência interespecíficas, facilitando a visualização do *barcoding gap*.

Outra variável importante para a existência do *barcoding gap* é o tamanho efetivo populacional (N_e). O N_e afeta o tempo de divergência que levará a especiação entre duas populações (Palumbi *et al.*, 2001). Desta maneira, N_e maiores fazem com

que as populações levem mais tempo para se diferenciarem totalmente. Isso significa dizer que populações maiores demoram mais para apresentar a monofilia recíproca. Assim, populações grandes podem exigir muito tempo para separação total e o consequente surgimento do *barcoding gap*, característico da monofilia recíproca.

Nesse contexto da aplicabilidade do *DNA Barcoding*, o conceito de espécie adotado é fundamental. Este conceito é amplamente debatido e ainda muito controverso (Sites e Marshall, 2003; de Queiroz, 2005; Hey, 2006). O *DNA Barcoding* assume que as espécies são monofiléticas (conceito filogenético de espécies). Isso traz sérias consequências na sua aplicação e no modo de interpretação dos resultados. Assim, quando se atribui o sucesso ou fracasso do *DNA Barcoding*, isso estará diretamente ligado as exigências do conceito de espécie adotado. Mas quando assume-se para o *DNA Barcoding* o conceito filogenético de espécies, os problemas da falta de monofilia poderão ser de grande relevância, visto que parte das espécies de fato não o são (Funk e Omland, 2003). Na Amazônia em especial, os grande N_e e as estruturas populacionais podem resultar na falta de monofilia como no caso de alguns peixes amazônicos, *Potamotrygon* (Toffoli *et al.*, 2008), *Cichla* (Willis *et al.*, 2008) e *Symphysodon* (Farias e Hrbek, 2008).

Normalmente os estudos que fazem uso do *DNA Barcoding* amostram um grande número de espécies e poucos espécimens, geralmente entre 5 a 12 amostras por espécie e muitas vezes sendo estas de uma ou poucas localidades ao longo da área de distribuição das mesmas (Dasmahapatra e Mallet, 2006; Hickerson *et al.*, 2006; Hebert *et al.*, 2009). Esse delineamento acaba suscitando dúvidas acerca de resultados positivos do uso dessa abordagem, em geral atribuindo-se parte de seu sucesso como consequência da amostragem insuficiente de espécies do taxa e ao longo da distribuição das espécies, de forma que a divergência intraespecífica fosse subestimada, aumentando a aparente acurácia do *DNA Barcoding* (Dasmahapatra e Mallet, 2006; Hickerson *et al.*, 2006; Kelly *et al.*, 2007).

Poucos testes empíricos para avaliação do *DNA Barcoding* foram realizados

no intuito de validar o *DNA Barcoding*. Entre eles pode-se citar a busca por um número mínimo de exemplares na utilização do *DNA Barcoding* (Matz e Nielsen, 2005; Zhang *et al.*, 2009) e teste da eficiência em espécies irmãs (Kelly *et al.*, 2007 ; Tavares e Baker, 2008). Porém, até o momento, poucos estudos avaliaram de fato a resolução do *DNA Barcoding* em estudos de ampla amostragem em espécies próximas. Assim, um trabalho que realizasse amplas amostragens em espécies próximas ao longo de toda suas respectivas distribuições poderia ajudar no melhor entendimento acerca da eficiência e funcionalidade do *DNA Barcoding*.

Pseudoplatystoma

Dessa forma, utilizou-se quatro espécies do gênero de peixes neotropicais *Pseudoplatystoma* (Figura 3) como modelo neste estudo, onde foram amostrados um número maior de exemplares do que é normalmente realizado de forma a também contemplar grande parte da distribuição desses peixes ao longo da bacia Amazônica. Pertencem à classe Actinopterygii, ordem Siluriformes, família Pimelodidae, subfamília Sorubiminae.

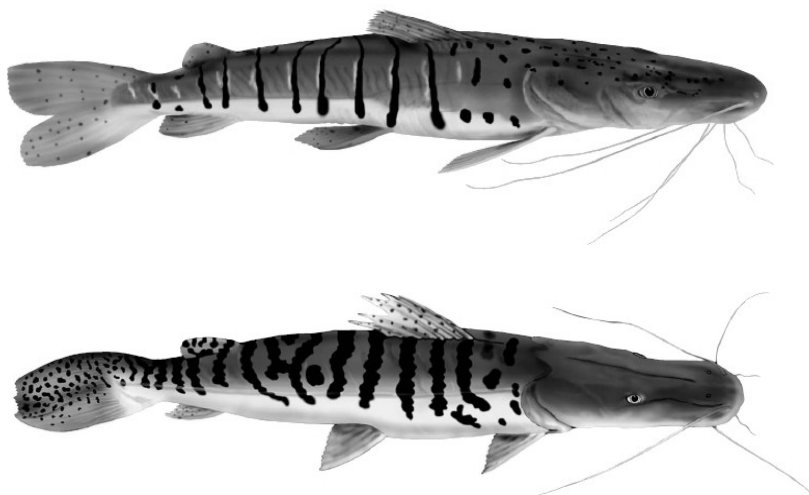


Figura 3. Exemplares do gênero *Pseudoplatystoma*, *P. fasciatum* (acima) e *P. tigrinum* (abaixo) (Modificado de Santos *et al.*, 2006).

O gênero *Pseudoplatystoma* é composto por grandes bagres migradores de

distribuição continental que possuem grande importância ecológica e econômica. Acerca de sua taxonomia, por um longo período o gênero foi dividido em três espécies, quando recentemente um trabalho de revisão morfológica dividiu o grupo em oito espécies (Buitrago-Suárez e Burr, 2007). Entretanto, mais recentemente a primeira filogenia molecular do gênero mostrou que este é composto por seis espécies (Torrico *et al.*, 2009), sendo uma espécie encontrada na bacia do Rio Magdalena (*P. magdaleniatum*), uma na bacia do Rio Orinoco (*P. orinocoense*), uma na bacia do Rio São Francisco (*P. corruscans*), duas na bacia do Paraná (*P. corruscans* e *P. reticulatum*) e três na bacia do Rio Amazonas (*P. reticulatum*, *P. tigrinum* e *P. fasciatum*).

OBJETIVOS

O objetivo geral deste trabalho foi realizar um teste de validação da ferramenta molecular *DNA Barcoding* como identificador da espécie usando o gênero *Pseudoplatystoma* como modelo. Os objetivos específicos foram: (1) testar a performance da ferramenta *DNA Barcoding* em *Pseudoplatystoma*, (2) verificar a correspondência entre a identificação molecular e a morfológica, feita *a priori* e (3) determinar a variabilidade genética das espécies de *Pseudoplatystoma* amostradas na Amazônia.

MATERIAIS E MÉTODOS

Material de estudo

Para validação do *DNA Barcoding*, foram inicialmente utilizadas 78 amostras de tecido do gênero *Pseudoplatystoma* coletadas em mercados de peixes ao longo de 11 pontos da Bacia do Rio Amazonas (Tabela 1, Figura 4), para as quais se obteve registros fotográficos dos exemplares. Em seguida foram acrescentadas amostras sem referências dos espécimens em imagens, somando-se um total de

229 amostras de tecido do gênero *Pseudoplatystoma*. Estas foram coletadas em mercados de peixes ao longo de 21 pontos da Bacia do Rio Amazonas e um ponto da Bacia do Rio Magdalena (Tabela 2, Figura 5). Estas amostras (barbilhão, nadadeira ou músculo) foram acondicionadas em álcool 70% e levados para o laboratório onde foram mantidas à 20°C negativos até o isolamento do DNA total. O material biológico está depositado no Banco de Tecidos e de DNA do Laboratório Temático de Biologia Molecular (LTBM), Instituto Nacional de Pesquisas da Amazônia (INPA). Os espécimens foram identificados segundo a chave de Buitrago-Suárez e Burr (2007) e fotografados para corroborar as identificações *a priori* usados em um primeiro momento (fotografias anexadas no *BOLD*, *Barcode of Life Data Systems*).

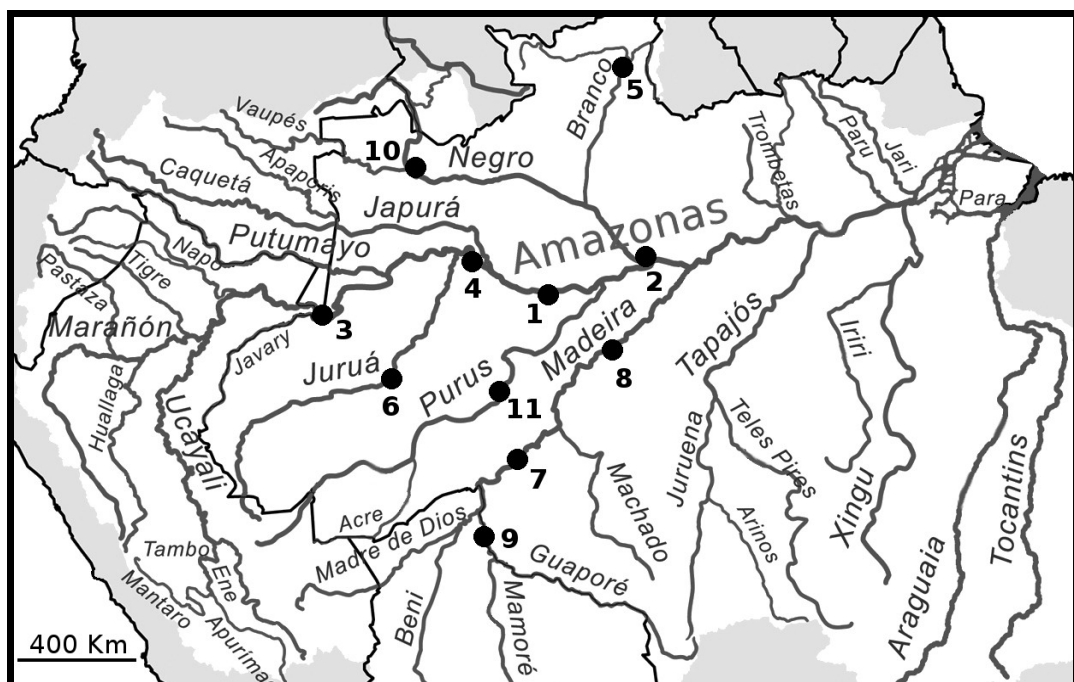


Figura 4. Mapa das localidades de coleta na bacia amazônica (banco padrão). Localidades: 1 – Coari, 2 – Manaus, 3 – Tabatinga, 4 – Tefé, 5 – Boa Vista, 6 – Eirunepé, 7 – Porto Velho, 8 – Manicoré, 9 – Guajará Mirim, 10 – São Gabriel da Cachoeira, 11 – Lábrea.

Tabela 1. Localização dos pontos de amostragem e o número de amostras sequenciadas para cada espécie (banco padrão).

Rio	Localidade	Número de mostras por espécie		
		<i>P. fasciatum</i>	<i>P. tigrinum</i>	<i>P. reticulatum</i>
Rio Amazonas	Coari	1	1	
	Manaus	5	1	
	Tabatinga	1	1	
	Tefé	4	3	
Rio Branco	Boa Vista	3	2	
Rio Juruá	Eirunepé	3		
Rio Madeira	Porto Velho	2	1	
Rio Madeira	Manicoré	1	2	
Rio Mamoré	Guajará Mirim	1		
Rio Negro	São Gabriel da Cachoeira	5		3
Rio Purus	Lábrea	35	3	
Total (78)		61	14	3

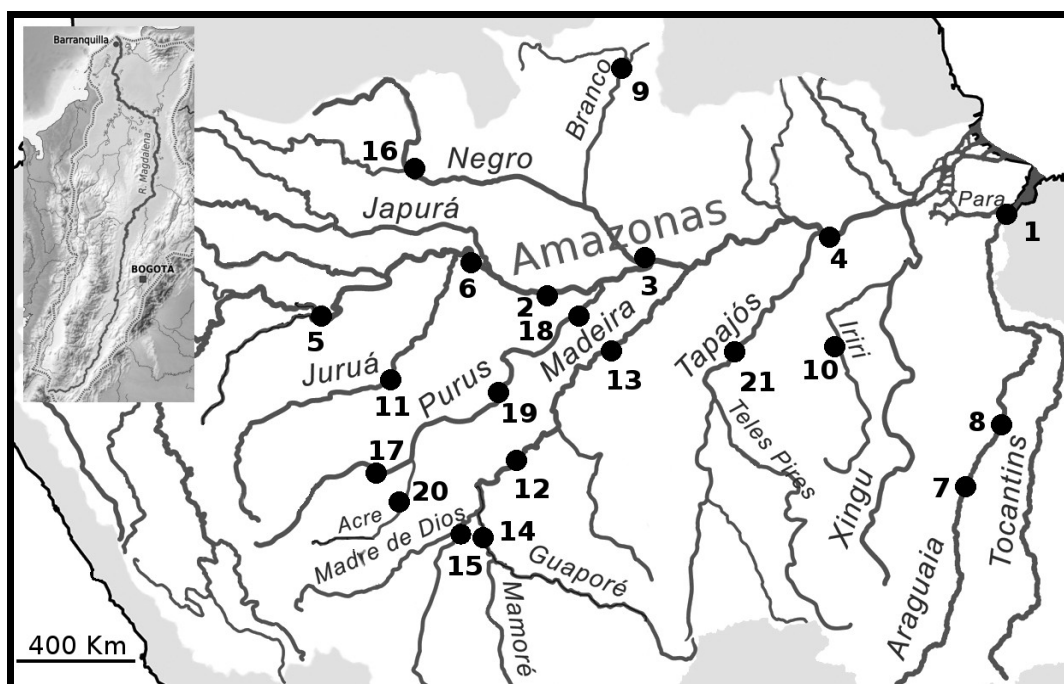


Figura 5. Mapa das localidades de coleta na bacia amazônica (banco completo). Localidades: 1 – Belém, 2 – Coari, 3 – Manaus, 4 – Santarém, 5 – Tabatinga, 6 – Tefé, 7 – São Félix do Araguaia, 8 – Itaipava, 9 – Boa Vista, 10 – *Rio Iriti*, 11 – Eirunepé, 12 – Porto Velho, 13 – Manicoré, 14 – Guajará Mirim, 15 – Riberalta, 16 – São Gabriel da Cachoeira, 17 – Boca do Acre, 18 – Ipiranga, 19 – Lábrea, 20 – Rio Branco, 21 – Jacareacanga, no alto a esquerda – *Rio Magdalena*.

Tabela 2. Localização dos pontos de amostragem e o número de amostras sequenciadas para cada espécie (banco completo).

Rio	Localidade	Número de mostras por espécie				
		<i>P. fasciatus</i>	<i>P. tigrinus</i>	<i>P. reticulatus</i>	<i>P. magdaleniatum</i>	<i>H. platyrrhynchos</i> <i>S. planiceps</i>
Rio Amazonas	Belém	1	1			
	Coari	1	2			
	Manaus	13	9		2	2
	Santarém	10	5			
	Tabatinga	6	6			
	Tefé	6	9			
Rio Araguaia	São Félix do Araguaia	3				
Rio Araguaia	Itaipava	5				
Rio Branco	Boa Vista	5	3			
Rio Iriri		10				
Rio Juruá	Eirunepé	6	1			
Rio Madeira	Porto Velho	5	6			1
Rio Madeira	Manicoré	3	4			
Rio Mamoré	Guajará Mirim	5	4			
	Ribeiratta	1	1			
Rio Negro	São Gabriel da Cachoeira	5		3		
Rio Purus	Boca do Acre	6	2			
	Ipiranga	1				
	Lábrea	35	3		2	1
	Rio Branco	4	8			
Rio Téles Pires	Jacareacanga	10				
Rio Magdalena					9	
Desconhecido		2	3	7		
Total		143	67	10	9	4

Isolamento e amplificação do DNA

Para isolamento de DNA foi utilizado um método com uso de fenol-clorofórmio modificado de Sambrook *et al.* (1989). A quantificação do DNA foi determinada em gel de agarose 0,8%, corado com brometo de etídio e fotografada com o sistema de foto-documentação (*Eagle Eye* - Stratagene). Um fragmento da região do gene mitocondrial citocromo oxidase subunidade 1 (COI) foi amplificado pela reação em cadeia da polimerase com um volume total de 50 µL com 10-100 ng de DNA total, 1.5 mM MgCl₂, 1x *Taq buffer*, 200 µM dNTPs, 0.5 µM de cada *primer* e 0.5 unidades de *Taq DNA Polymerase* em termociclador GeneAmp™ PCR System 2400 (Applied Biosystems). Os *primers* utilizados foram: BOL-COI F3: 5' TCA ACY AAT CAY AAA GAT ATY GGC CAC 3' e COI R3: 5' ACT TCY GGG TGR CCR AAR AAT CA 3' modificados de Ward *et al.* (2005). No primer BOL-COI F3 foi inserida uma cauda M13 na extremidade 5'. A amplificação foi realizada em 35 ciclos com as seguintes temperaturas: desnaturação a 94°C por 1 minuto, anelamento a 55°C por 40 segundos e extensão a 72°C por 1 minuto. Após, a eficiência da amplificação foi verificada em gel de agarose 0,8% com marcador de peso molecular.

Purificação e sequenciamento do DNA

Todos os produtos amplificados foram purificados com um kit comercial *GFX™ PCR DNA and Gel Band Purification Kit* (Amersham Pharmacia Biotech Europe GmbH, Freiburg, Germany) de acordo com o protocolo do fabricante. O produto purificado (região COI do DNA mitocondrial) foi sequenciado no analisador automático de DNA *ABI 3730 DNA Analyzer* (Applied Biosystems) através do método de Sanger *et al.* (1977), com terminadores marcados com fluorescência. Para a realização das reações de sequenciamento foi utilizado o *BigDye® Terminator v3.1 Cycle Sequencing Kit* (Applied Biosystems). Para as reações de sequenciamento foi utilizado o *primer* M13 e eventualmente o *primer* R3.

Análise das sequências

As sequências nucleotídicas foram então submetidas à análise de confiança nos programas do pacote *Staden Package* (Staden, 1996) para obtenção de uma sequência consenso. Esta foi submetida ao programa BLAST (Altschul *et al.*, 1997) para confirmação e comparação da região sequenciada com os dados disponíveis através do *National Center for Biotechnology Information* (Genbank, <http://www.ncbi.nlm.nih.gov>), assim como também foram comparadas com os dados disponíveis através do *Barcoding of Life Database* (BOLD) (Ratnasingham e Hebert, 2007). Para edição das sequências foram utilizados os programas GENEDOC version 2.6.001 (Nicholas e Nicholas, 1997) e BioEdit 7.0.9 (Hall, 1999). Para o alinhamento das sequências foi utilizado o programa ClustalW (Thompson *et al.*, 1994).

Para a determinação dos grupos específicos, utilizou-se o método de Agrupamento de Vizinhos (*Neighbour-joining*, NJ) (Saitou e Nei, 1987) sob o modelo *Kimura-2-parameter* (K2P) (Kimura, 1980) para correção das distâncias nucleotídicas, e a confiabilidade de cada nó foi estimada pelo método de *bootstrap*, com 1000 réplicas (Felsenstein, 1985). Esse modelo K2P fornece uma melhor medição quando as distâncias genéticas são baixas (Nei e Kumar, 2000). O algoritmo de distância NJ é frequentemente usado para o *DNA Barcoding* porque é simples e rápido, o que é uma vantagem para a análise de grandes volumes de dados, além de o objetivo do *DNA Barcoding* ser o de identificar através da similaridade das sequências e não fornecer uma hipótese filogenética definitiva. Foram usadas as espécies irmãs do gênero *Pseudoplatystoma*, *Hemisorubim platyrhynchos* e *Sorubimichthys planiceps*, como grupos externos.

Foram calculadas as divergências par-a-par intra e interespecífica entre as sequências (distância K2P), assim como foram calculadas a distância média e desvio padrão intra e interespecíficos. Os valores de divergência foram plotadas em um gráfico de frequência para visualização. Também foram calculadas a distância média e desvio padrão intra e interespecíficos. As análises filogenéticas e cálculos

de distância foram realizadas com auxílio dos programas MEGA 4 (Tamura *et al.*, 2007). Além destas análises em programas individualizados foram utilizadas em paralelo as ferramentas do BOLD: *Distance Summary* e *Sequence Composition* (Ratnasingham e Hebert, 2007).

Validação do DNA Barcoding

Para validação do *DNA Barcoding* foram realizados três testes e estes foram reproduzidos para cada um dos dois bancos. Os testes visaram verificar a relação das amostras (teste da monofilia), a estruturação em função dos pontos de coleta (teste da estruturação) e as divergências intra e interespecíficas (teste do *barcoding gap*). Nos testes foram calculadas as distâncias par-a-par entre as sequências das amostras de *Pseudoplatystoma* e feitos testes de reamostragens com diferentes números de amostras.

Teste 1: 5 amostras ao acaso/1000 permutações.

Teste 2: 10 amostras ao acaso/1000 permutações.

Teste 3: todas as amostras.

No Teste 1 foram usadas reamostragens ao acaso das distâncias de até 5 amostras da cada espécie. No Teste 2, foram usadas reamostragens ao acaso das distâncias de até 10 amostras da cada espécie. No Teste 3, foram usadas as distâncias de todas as amostras da cada espécie. Esses testes foram realizados com auxílio de programas feitos na linguagem de programação R (R Development Core Team, 2005) e foram calculados os seguintes índices: média da distância máxima e desvio padrão, intervalos de confiança de 95% e 99%, média e desvio padrão, intervalos de confiança de 95% e 99% e o *barcoding gap* (Anexo B). Em todos os testes foram utilizadas 1000 permutações e para plotagem das distâncias

foi usadas a função *locfit*.

RESULTADOS

Banco de Dados

O banco construído como padrão de sequências de amostras com referências em imagens somou 78 sequências, 61 de *P. fasciatum*, 14 de *P. tigrinum* e três de *P. reticulatum*. Este obteve 100% de identificação correta entre imagens e sequências. Foi então obtida uma árvore Neighbour-joining da divergência nucleotídica das 78 sequências de *Pseudoplatystoma* e os dois grupos externos, na qual todas as espécies formaram grupos monofiléticos (Figura 6 e Tabelas 2 e 3).

Tabela 3. Distância média entre as espécies de *Pseudoplatystoma* e em relação aos grupos externos obtida no programa MEGA (banco padrão).

	[1]	[2]	[3]	[4]	[5]
[1] <i>P. fasciatum</i>	-				
[2] <i>P. tigrinum</i>	0,023	-			
[3] <i>P. reticulatum</i>	0,023	0,018	-		
[4] <i>S. planiceps</i>	0,107	0,109	0,106	-	
[5] <i>H. platyrhynchos</i>	0,143	0,141	0,143	0,149	-

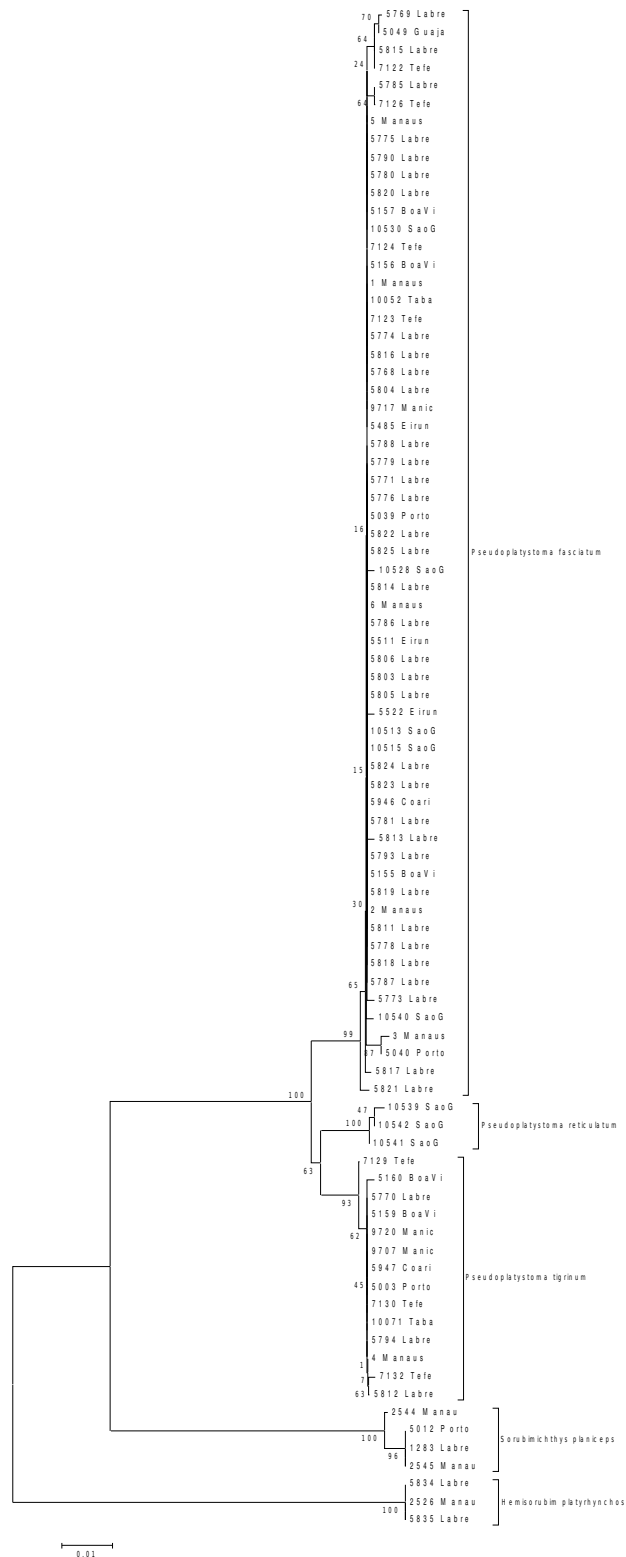


Figura 6. Árvore Neighbour-joining da divergência nucleotídica das sequências do COI entre três espécies do gênero *Pseudoplatystoma* e dois grupos externos, usando o método Kimura-2-parâmetros com análise de bootstrap com 1000 réplicas.

Das 78 sequências, 72 possuem um tamanho de 652 pares de bases e seis sequências são fragmentos menores (>620 pb). Não foram verificadas inserções, deleções, códigos de parada ou alterações nas sequências de aminoácidos. A seguir os números de acesso do *GenBank*: GU570706 a GU570942, e do *BOLD*: OMFAF001 a OMFAF237.

A composição nucleotídica média do gene COI dos 78 indivíduos de *Pseudoplatystoma* foi de: 27,71% para citosina, 27,95% para timina, 27,04% para adenina e 17,31% para guanina. O índice médio de GC foi de 45,02%, sendo 57%, 42,83% e 34,78% os valores médios para primeira, segunda e terceira base dos códons, respectivamente (Tabela 4).

Tabela 4. Composição nucleotídica do gene COI (mínima, média, máxima e desvio padrão) do banco padrão.

	Mín.	Méd.	Máx.	D.P.
G %	16,72	17,31	17,94	0,028
C %	27,21	27,71	27,91	0,015
A %	26,25	27,04	27,81	0,030
T %	27,65	27,95	28,66	0,018
GC %	44,53	45,02	45,44	0,019
GC % Codon Pos 1	56,04	57	57,08	0,023
GC % Codon Pos 2	42,44	42,83	44,02	0,025
GC % Codon Pos 3	34,05	34,78	35,89	0,047

Os resultados do *BOLD* mostraram para o banco padrão uma distância nucleotídica intraespecífica média de 0,098% e a interespecífica de 2,26% no gênero *Pseudoplatystoma* (Tabela 5). Um gráfico mostrando as frequências das distâncias pode ser observado na Figura 7.

Tabela 5. Valores médios das distâncias intra e interespecíficas do gene COI (distância mínima, média, máxima e desvio padrão) do banco padrão calculados no BOLD (K2P).

	N	Comparações	Dist. Mín.	Dist. Méd.	Dist. Máx.	D.P.
Intraespecífica	78	1924	0,0000	0,0009	0,0077	0,00003
Interespecífica	78	1079	0,0147	0,0226	0,0299	0,00005

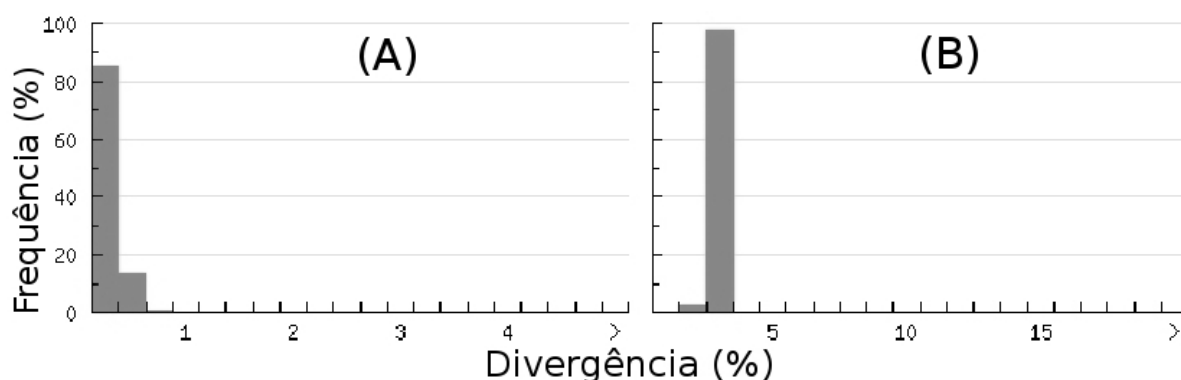


Figura 7. Distribuição das distâncias intra (A) e interespecíficas (B) do gene COI do banco padrão.

Em um segundo momento foram acrescentadas as sequências de amostras sem referências em imagens e o banco de dados somou 229 sequências (151 novas sequências). Este obteve 91,26% (20 de 229) de identificação correta entre as espécies identificadas *a posteriori* pelas sequências e a identificação atribuída *a priori*. O aparente resultado de ineficiência (8,74%) do *DNA Barcoding* foi atribuído a erros de identificação *a priori* em virtude da confiabilidade do primeiro banco, aos altos valores de confiança que sustentam os clados na árvores de distâncias e a clara presença do *barcoding gap*, assim pode-se afirmar que os 8,74% de falta de correspondência não seriam explicados pela ineficiência da identificação do *DNA Barcoding* ou compartilhamento de haplótipos entre as espécies. A identificação *a*

priori errada foi corrigida para as análises.

Com essa confirmação da identificação das espécies *a posteriori* somaram-se as seguintes quantidades de sequências para cada espécie no banco completo: 143 de *P. fasciatum*, 63 de *P. tigrinum*, 10 de *P. reticulatum* e nove de *P. magdaleniatum*. A árvore Neighbour-joining das 229 sequências de *Pseudoplatystoma*, mais os dois grupos externos também apresentou todos os grupos monofiléticos (Anexo A).

Diversidade Genética

Foram obtidas 237 sequências, 229 do gênero *Pseudoplatystoma* e oito dos dois grupos externos. Destas, 200 com 652 pares de bases e 37 sequências são fragmentos menores (>517 pb). Não foram verificadas inserções, deleções, códigos de parada ou alterações nas sequências de aminoácidos intragenéricas. A seguir os números de acesso do *GenBank*: GU570706 a GU570942, e do *BOLD*: OMFAF001 a OMFAF237.

A composição nucleotídica do gene COI dos 229 indivíduos de *Pseudoplatystoma* foi de: 27,67% para citosina, 27,98% para timina, 26,94% para adenina e 17,4% para guanina. O índice médio de GC foi de 45,07%, sendo 56,95%, 42,82% e 35,04% os valores médios para primeira, segunda e terceira base dos códons respectivamente (Tabela 6). Não foi observado saturação entre as substituições nucleotídicas com a distância genética (Figura 8).

Tabela 6. Composição nucleotídica do gene COI (mínima, média, máxima e desvio padrão) do banco completo.

	Mín.	Méd.	Máx.	D.P.
G %	16,37	17,4	18,87	0,03
C %	27,21	27,67	28,1	0,01
A %	25,46	26,94	27,81	0,03
T %	27,61	27,98	29,15	0,01
GC %	44,13	45,07	46,32	0,02
GC % Codon Pos 1	54,45	56,95	57,08	0,02
GC % Codon Pos 2	42,02	42,82	44,08	0,02
GC % Codon Pos 3	34,05	35,04	38,65	0,05

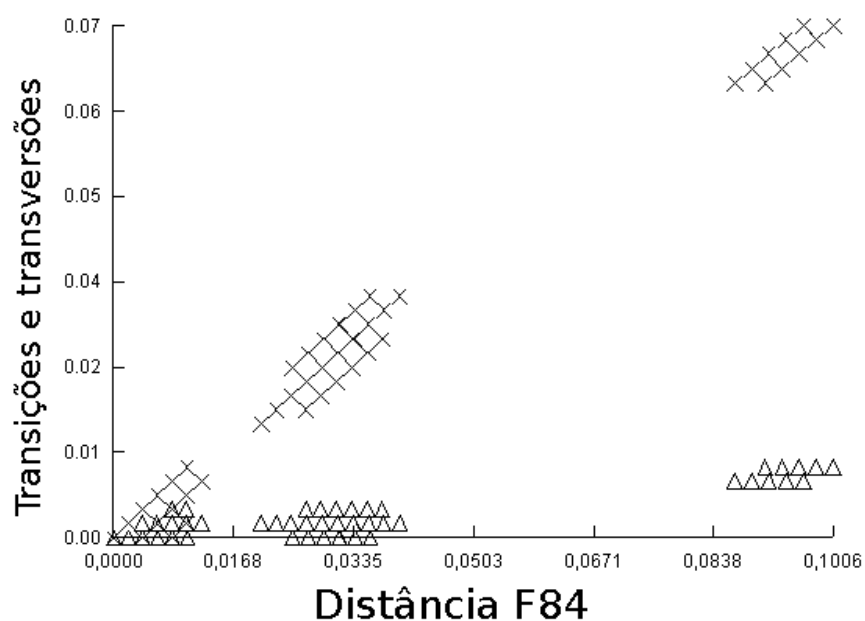


Figura 8. Gráfico do número de substituições nucleotídicas (transições – s, e transversões – v) contra a distância genética do gene COI (banco completo).

Foram encontrados 36 haplótipos entre as 229 sequências de *Pseudoplatystoma*, 29 sítios polimórficos e 30 mutações. A diversidade haplotípica variou de 0,3713 a 0,5691 e a diversidade nucleotídica de 0,00085 a 0,00119. Estas informações mais detalhadas podem ser visualizadas na Tabela 7, assim como as médias das diferenças par-a-par e os resultados dos testes *D* de Tajima e *F_s* de Fu.

Tabela 7. Índices de polimorfismo de DNA das quatro espécies de *Pseudoplatystoma* analisadas e testes de neutralidade (banco completo). N (tamanho da amostra), H (número de haplótipos), S (sítios polimórficos), ETA (número de mutações), HD (diversidade haplotípica), PI (diversidade nucleotídica), K (médias das diferenças par-a-par).

Espécie	N	H	S	ETA	HD	PI	K	D de Tajima	Fs de Fu
<i>P. fasciatum</i>	143	23	19	20	0,5691	0,00119	0,671	-2,25969*	-19,55*
<i>P. tigrinum</i>	67	8	7	7	0,3713	0,00088	0,436	-1,76036	-6,477
<i>P. reticulatum</i>	10	3	2	2	0,5111	0,00094	0,55556	-0,69098	-0,691
<i>P. magdaleniatum</i>	9	2	1	1	0,5556	0,00085	0,55556	1,40117	1,015
TOTAL	229	36	29	30					

*Resultado significativo

Os resultados do BOLD mostraram uma distância nucleotídica intraespecífica média de 0,10% e a interespecífica de 3,07% no gênero *Pseudoplatystoma* (Tabela 8). Um gráfico mostrando as frequências das distâncias pode ser observado na Figura 9.

Tabela 8. Valores médios das distâncias intra e interespecíficas do gene COI (distância mínima, média, máxima e desvio padrão) do banco completo calculados no BOLD (K2P).

	N	Comparações	Dist. Mín.	Dist. Mín.	Dist. Mín.	D.P.
Intraespecífica	229	12445	0,0000	0,0010	0,0077	0,00001
Interespecífica	229	13661	0,0146	0,0307	0,0877	0,00016

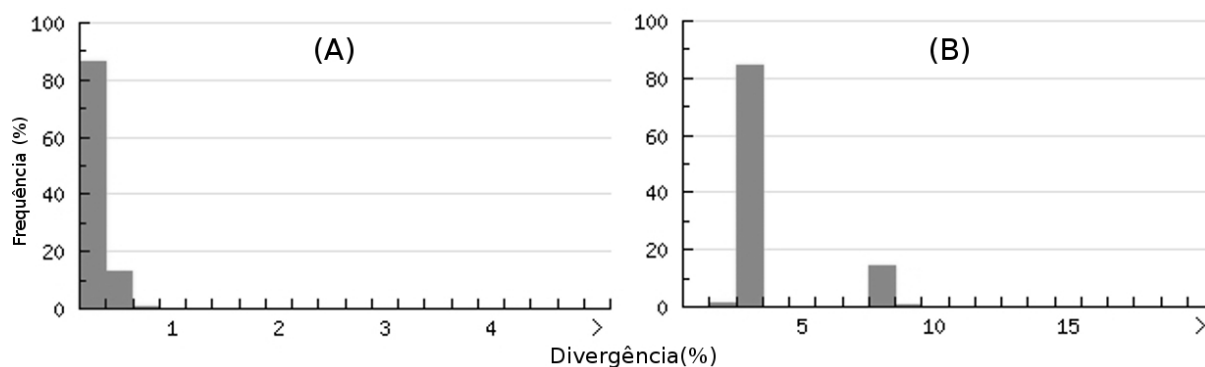


Figura 9. Distribuição das distâncias das espécies de *Pseudoplatystoma* intra (A) e interespecíficas (B) do gene COI do banco completo.

Validação do *DNA Barcoding*

Nos testes realizados para a validação do *DNA Barcoding* como o banco padrão todas espécies de *Pseudoplatystoma* sempre mostraram-se monofiléticas e as divergências intraespecíficas sempre foram inferiores às interespecíficas.

Foram obtidos os valores para os testes: distância máxima, desvio padrão e intervalos de confiança de 95% e 99% (Tabela 9), média da distância, desvio padrão e intervalos de confiança de 95% e 99% (Tabela 10) e o *barcoding gap* (Figura 10).

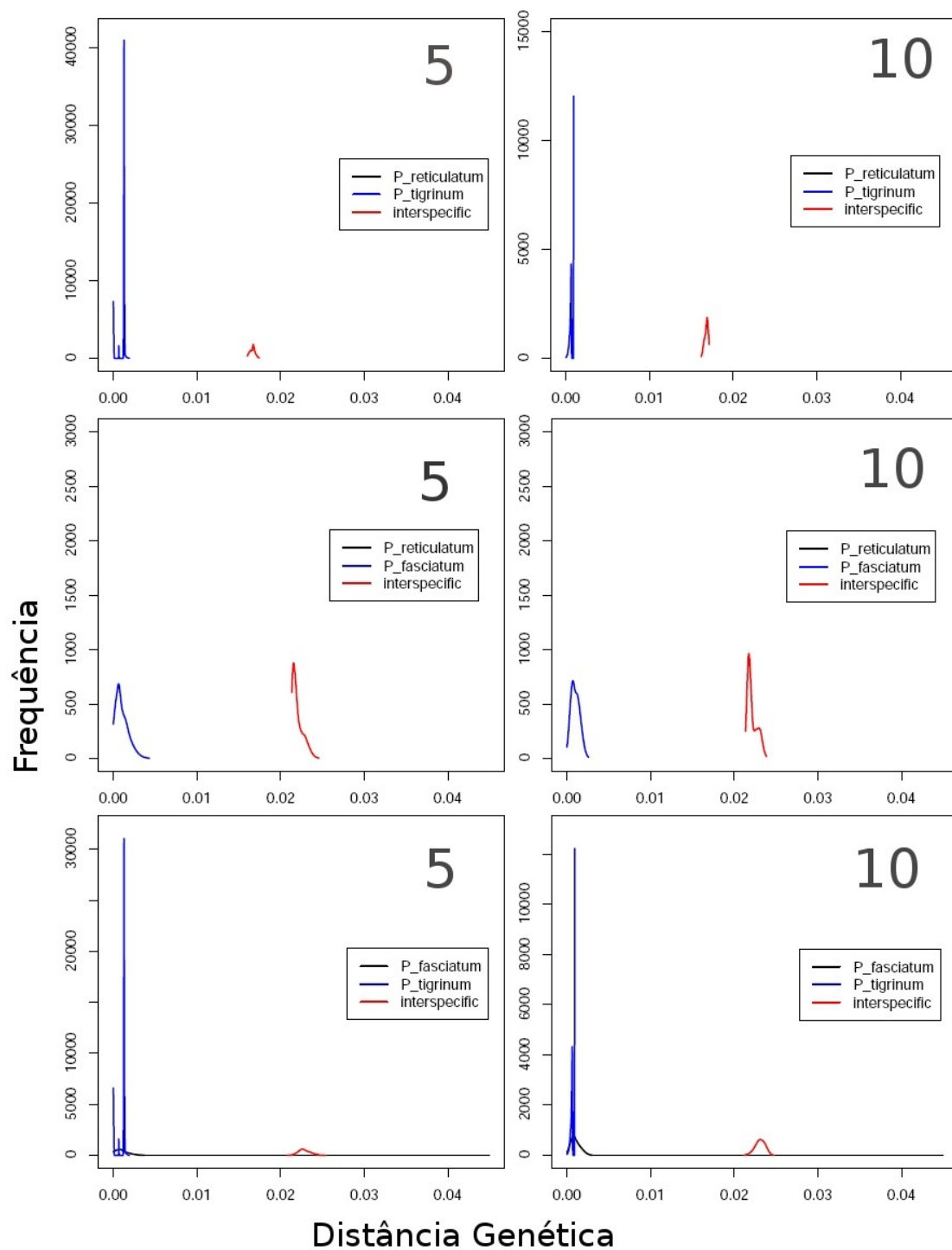


Figura 10. Representação do *barcoding gap* entre as espécies de *Pseudoplatystoma* (frequência absoluta de amostragens de distâncias par-a-par). A esquerda os resultados dos teste com reamostragens de 5 indivíduos e à direita de 10 indivíduos (banco padrão).

Tabela 9. Distância máxima intra e interespecífica e desvio padrão, intervalos de confiança de 95% e 99% para subamostragens com 5, 10 e todas as amostras do banco com referências em imagens. Ao lado do nome das espécies aparece o número amostral total e ainda os intervalos de confiança (IC) (banco padrão).

Teste 1 (N = 5)		<i>P. fasciatum</i>	<i>P. tigrinum</i>	<i>P. reticulatum</i>
<i>P. fasciatum</i> (61)	Média	0,0023 ± 0,0018		
	IC 95%	(0,0000 – 0,0062)		
	IC 99%	(0,0000 – 0,0065)		
<i>P. tigrinum</i> (14)	Média	0,0250 ± 0,0017	0,0016 ± 0,0011	
	IC 95%	(0,0219 – 0,0281)	(0,0000 – 0,0032)	
	IC 99%	(0,0219 – 0,0291)	(0,0000 – 0,0032)	
<i>P. reticulatum</i> (3)	Média	0,0232 ± 0,0015	0,0175 ± 0,0008	0,0000
	IC 95%	(0,0214 – 0,0264)	(0,0163 – 0,0185)	(0,0000 – 0,0000)
	IC 99%	(0,0214 – 0,0264)	(0,0163 – 0,0185)	(0,0000 – 0,0000)
Teste 2 (N = 10)				
<i>P. fasciatum</i> (61)	Média	0,0039 ± 0,0018		
	IC 95%	(0,0065 – 0,0015)		
	IC 99%	(0,0000 – 0,0077)		
<i>P. tigrinum</i> (14)	Média	0,0267 ± 0,0015	0,0029 ± 0,0006	
	IC 95%	(0,0242 – 0,0291)	(0,0016 – 0,0032)	
	IC 99%	(0,0234 – 0,0298)	(0,0015 – 0,0032)	
<i>P. reticulatum</i> (3)	Média	0,0242 ± 0,0014	0,0182 ± 0,0004	0,0000
	IC 95%	(0,0214 – 0,0264)	(0,0168 – 0,0185)	(0,0000 – 0,0000)
	IC 99%	(0,0214 – 0,0278)	(0,0163 – 0,0185)	(0,0000 – 0,0000)
Teste 3 (todas)				
<i>P. fasciatum</i> (61)	Média	0,0084 ± 0,0000		
	IC 95%	0,0084 – 0,0084		
	IC 99%	0,0084 – 0,0084		
<i>P. tigrinum</i> (14)	Média	0,0293 ± 0,0000	0,0032 ± 0,0000	
	IC 95%	(0,0293 – 0,0293)	(0,0032 – 0,0032)	
	IC 99%	(0,0293 – 0,0293)	(0,0032 – 0,0032)	
<i>P. reticulatum</i> (3)	Média	0,0278 ± 0,0000	0,0185 ± 0,0000	0,0000
	IC 95%	(0,0278 – 0,0278)	(0,0185 – 0,0185)	(0,0000 – 0,0000)
	IC 99%	(0,0278 – 0,0278)	(0,0185 – 0,0185)	(0,0000 – 0,0000)

Tabela 10. Distância média intra e interespecífica e desvio padrão, intervalos de confiança de 95% e 99% para subamostragens com 5, 10 e todas as amostras do banco com referências em imagens. Ao lado do nome das espécies aparece o número amostral total e ainda os intervalos de confiança (IC) (banco padrão).

Teste 1 (N = 5)		<i>P. fasciatum</i>	<i>P. tigrinum</i>	<i>P. reticulatum</i>
<i>P. fasciatum</i> (61)	Média	0,0010 ± 0,0008		
	IC 95%	(0,0000 – 0,0025)		
	IC 99%	(0,0000 – 0,0034)		
<i>P. tigrinum</i> (14)	Média	0,0229 ± 0,0007	0,0007 ± 0,0005	
	IC 95%	(0,0216 – 0,0240)	(0,0000 – 0,0013)	
	IC 99%	(0,0212 – 0,0247)	(0,0000 – 0,0018)	
<i>P. reticulatum</i> (3)	Média	0,0220 ± 0,0007	0,0166 ± 0,0003	0,0000
	IC 95%	(0,0214 – 0,0232)	(0,0160 – 0,0171)	(0,0000 – 0,0000)
	IC 99%	(0,0214 – 0,0239)	(0,0160 – 0,0175)	(0,0000 – 0,0000)
Teste 1 (N = 10)				
<i>P. fasciatum</i> (61)	Média	0,0010 ± 0,0005		
	IC 95%	(0,0003 – 0,0019)		
	IC 99%	(0,0000 – 0,0023)		
<i>P. tigrinum</i> (14)	Média	0,0231 ± 0,0006	0,0007 ± 0,0002	
	IC 95%	(0,0221 – 0,0240)	(0,0003 – 0,0009)	
	IC 99%	(0,0213 – 0,0244)	(0,0003 – 0,0009)	
<i>P. reticulatum</i> (3)	Média	0,0222 ± 0,0006	0,0168 ± 0,0002	0,0000
	IC 95%	(0,0214 – 0,0233)	(0,0163 – 0,0171)	(0,0000 – 0,0000)
	IC 99%	(0,0214 – 0,0235)	(0,0162 – 0,0171)	(0,0000 – 0,0000)
Teste 1 (todas)				
<i>P. fasciatum</i> (61)	Média	0,0010 ± 0,0000		
	IC 95%	(0,0010 – 0,0010)		
	IC 99%	(0,0010 – 0,0010)		
<i>P. tigrinum</i> (14)	Média	0,0229 ± 0,0000	0,0007 ± 0,0000	
	IC 95%	(0,0229 – 0,0229)	(0,0007 – 0,0007)	
	IC 99%	(0,0229 – 0,0229)	(0,0007 – 0,0007)	
<i>P. reticulatum</i> (3)	Média	0,0230 ± 0,0000	0,0169 ± 0,0000	0,0000
	IC 95%	(0,0230 – 0,0230)	(0,0169 – 0,0169)	(0,0000 – 0,0000)
	IC 99%	(0,0230 – 0,0230)	(0,0169 – 0,0169)	(0,0000 – 0,0000)

Quando acrescentadas mais sequências, formando o banco completo, foram obtidos os os valores para os testes: distância máxima, desvio padrão e intervalos de confiança de 95% e 99% (Tabela 11), média da distância, desvio padrão e intervalos de confiança de 95% e 99% (Tabela 12) e o *barcoding gap* (Figura 11). Novamente nos testes de validação do *DNA Barcoding* todas as espécies de *Pseudoplatystoma* sempre mostraram-se monofiléticas, não houve estruturação em função dos pontos de coletas (Figura 12) e as divergências intraespecíficas sempre foram inferiores as interespecíficas (Tabelas 11 e 12).

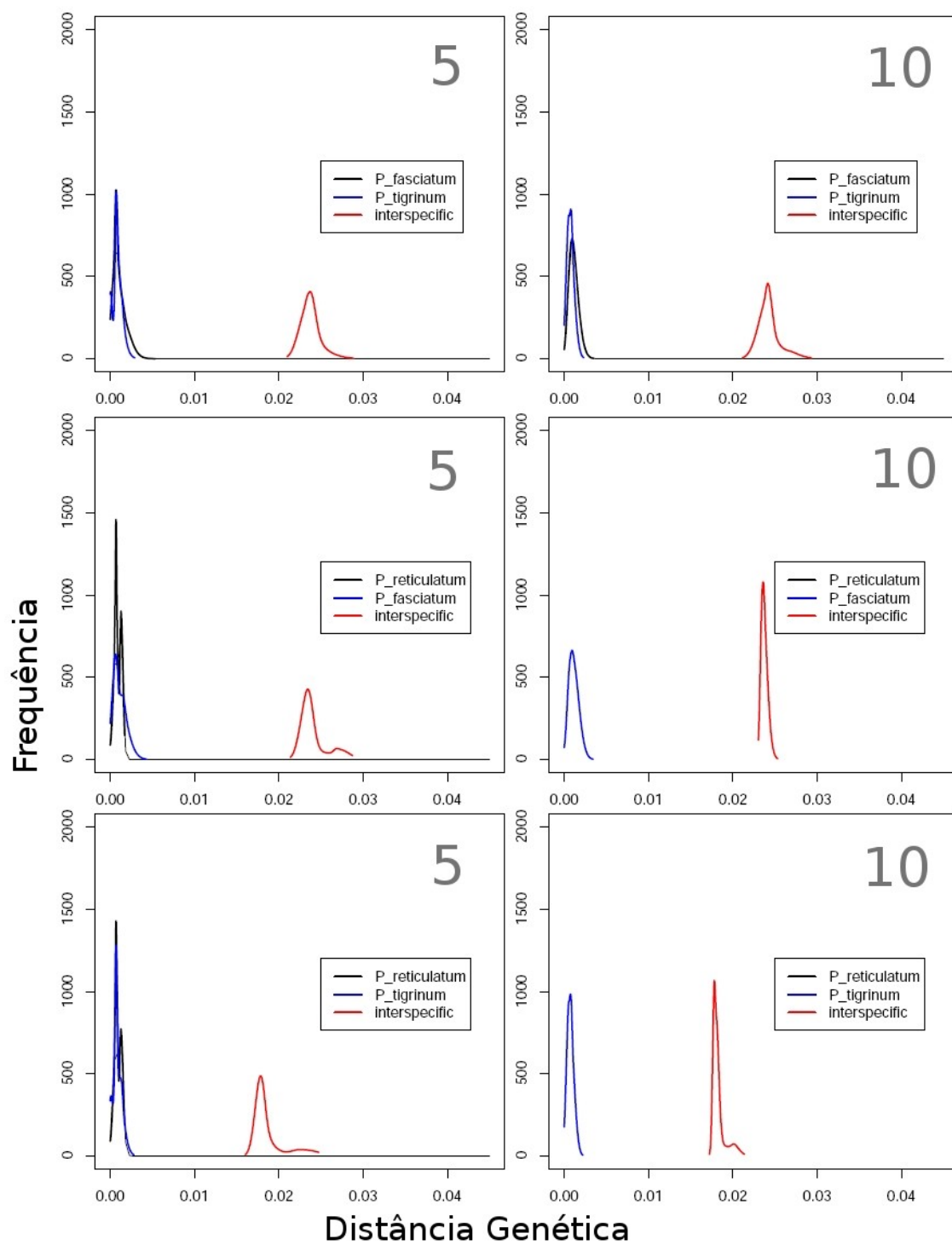


Figura 11. Representação do *barcoding gap* entre as espécies de *Pseudoplatystoma* (frequência absoluta de amostragens de distâncias par-a-par). À esquerda os resultados dos teste com reamostragens de 5 indivíduos e à direita de 10 indivíduos (banco completo).

Tabela 11. Distância máxima intra e interespecífica e desvio padrão, intervalos de confiança de 95% e 99% para subamostragens com 5, 10 e todas as amostras. Ao lado do nome das espécies aparece o número amostral total e ainda os intervalos de confiança (IC) (banco completo).

Teste 1 (N = 5)		<i>P. fasciatum</i>	<i>P. tigrinum</i>	<i>P. reticulatum</i>	<i>P. magdaleniatum</i>
<i>P. fasciatum</i> (143)	Média	0,0025 ± 0,0017			
	IC 95%	(0,0000 – 0,0062)			
	IC 99%	(0,0000 – 0,0065)			
<i>P. tigrinum</i> (76)	Média	0,0259 ± 0,0021	0,0017 ± 0,0014		
	IC 95%	(0,0226 – 0,0298)	(0,0000 – 0,0039)		
	IC 99%	(0,0219 – 0,0319)	(0,0000 – 0,0052)		
<i>P. reticulatum</i> (10)	Média	0,0266 ± 0,0022	0,0209 ± 0,0024	0,0021 ± 0,0010	
	IC 95%	(0,0239 – 0,0301)	(0,0180 – 0,0264)	(0,0000 – 0,0034)	
	IC 99%	(0,0221 – 0,0326)	(0,0169 – 0,0264)	(0,0000 – 0,0034)	
<i>P. magdaleniatum</i> (9)	Média	0,0772 ± 0,0023	0,0818 ± 0,0035	0,0826 ± 0,0020	0,0015 ± 0,0001
	IC 95%	(0,0738 – 0,0815)	(0,0774 – 0,0886)	(0,0795 – 0,0850)	(0,0015 – 0,0015)
	IC 99%	(0,0738 – 0,0830)	(0,0774 – 0,0920)	(0,0777 – 0,0850)	(0,0015 – 0,0015)
Teste 2 (N = 10)					
<i>P. fasciatum</i> (143)	Média	0,0040 ± 0,0018			
	IC 95%	(0,0015 – 0,0072)			
	IC 99%	(0,0000 – 0,0081)			
<i>P. tigrinum</i> (76)	Média	0,0279 ± 0,0023	0,0029 ± 0,0013		
	IC 95%	(0,0247 – 0,0321)	(0,0000 – 0,0052)		
	IC 99%	(0,0238 – 0,0339)	(0,0000 – 0,0058)		
<i>P. reticulatum</i> (10)	Média	0,0271 ± 0,0015	0,0213 ± 0,0016	0,0034 ± 0,0000	
	IC 95%	(0,0260 – 0,0296)	(0,0190 – 0,0239)	(0,0034 – 0,0034)	
	IC 99%	(0,0243 – 0,0311)	(0,0190 – 0,0257)	(0,0034 – 0,0034)	
<i>P. magdaleniatum</i> (9)	Média	0,0791 ± 0,0024	0,0845 ± 0,0035	0,0840 ± 0,0000	0,0015 ± 0,0000
	IC 95%	(0,0756 – 0,0828)	(0,0791 – 0,0904)	(0,0840 – 0,0840)	(0,0015 – 0,0015)
	IC 99%	(0,0738 – 0,0840)	(0,0774 – 0,0922)	(0,0840 – 0,0840)	(0,0015 – 0,0015)
Teste 3 (todas)					
<i>P. fasciatum</i> (143)	Média	0,0090 ± 0,0000			
	IC 95%	(0,0090 – 0,0090)			
	IC 99%	(0,0090 – 0,0090)			
<i>P. tigrinum</i> (76)	Média	0,0379 ± 0,0000	0,0061 ± 0,0000		
	IC 95%	(0,0379 – 0,0379)	(0,0061 – 0,0061)		
	IC 99%	(0,0379 – 0,0379)	(0,0061 – 0,0061)		
<i>P. reticulatum</i> (10)	Média	0,0311 ± 0,0000	0,0268 ± 0,0000	0,0034 ± 0,0000	
	IC 95%	(0,0311 – 0,0311)	(0,0268 – 0,0268)	(0,0034 – 0,0034)	
	IC 99%	(0,0311 – 0,0311)	(0,0268 – 0,0268)	(0,0034 – 0,0034)	
	Média	0,0824 ± 0,0000	0,0942 ± 0,0000	0,0840 ± 0,0000	0,0015 ± 0,0000

Tabela 12. Distância média intra e interespecífica e desvio padrão, intervalos de confiança de 95% e 99% para subamostragens com 5, 10 e todas as amostras. Ao lado do nome das espécies aparece o número amostral total e ainda os intervalos de confiança (IC) (banco completo).

Teste 4 (N = 5)		<i>P. fasciatum</i>	<i>P. tigrinum</i>	<i>P. reticulatum</i>	<i>P. magdaleniatum</i>
<i>P. fasciatum</i> (143)	Média	0,0011 ± 0,0008			
	IC 95%	(0,0000 – 0,0026)			
	IC 99%	(0,0000 – 0,0033)			
<i>P. tigrinum</i> (76)	Média	0,0237 ± 0,0013	0,0007 ± 0,0006		
	IC 95%	(0,0219 – 0,0264)	(0,0000 – 0,0019)		
	IC 99%	(0,0213 – 0,0276)	(0,0000 – 0,0023)		
<i>P. reticulatum</i> (10)	Média	0,0242 ± 0,0018	0,0189 ± 0,0022	0,0009 ± 0,0004	
	IC 95%	(0,0220 – 0,0277)	(0,0169 – 0,0240)	(0,0000 – 0,0016)	
	IC 99%	(0,0217 – 0,0284)	(0,0163 – 0,0243)	(0,0000 – 0,0017)	
<i>P. magdaleniatum</i> (9)	Média	0,0750 ± 0,0020	0,0800 ± 0,0031	0,0804 ± 0,0020	0,0009 ± 0,0001
	IC 95%	(0,0728 – 0,0786)	(0,0759 – 0,0864)	(0,0769 – 0,0832)	(0,0006 – 0,0009)
	IC 99%	(0,0724 – 0,0799)	(0,0756 – 0,0878)	(0,0765 – 0,0835)	(0,0006 – 0,0009)
Teste 5 (N = 10)					
<i>P. fasciatum</i> (143)	Média	0,0011 ± 0,0006			
	IC 95%	(0,0003 – 0,0022)			
	IC 99%	(0,0000 – 0,0028)			
<i>P. tigrinum</i> (76)	Média	0,0243 ± 0,0014	0,0007 ± 0,0004		
	IC 95%	(0,0224 – 0,0272)	(0,0000 – 0,0015)		
	IC 99%	(0,0216 – 0,0282)	(0,0000 – 0,0019)		
<i>P. reticulatum</i> (10)	Média	0,0237 ± 0,0004	0,0183 ± 0,0009	0,0009 ± 0,0000	
	IC 95%	(0,0232 – 0,0245)	(0,0176 – 0,0203)	(0,0009 – 0,0009)	
	IC 99%	(0,0230 – 0,0250)	(0,0174 – 0,0208)	(0,0009 – 0,0009)	
<i>P. magdaleniatum</i> (9)	Média	0,0762 ± 0,0022	0,0819 ± 0,0029	0,0816 ± 0,0000	0,0009 ± 0,0000
	IC 95%	(0,0730 – 0,0798)	(0,0765 – 0,0871)	(0,0816 – 0,0816)	(0,0009 – 0,0009)
	IC 99%	(0,0728 – 0,0806)	(0,0762 – 0,0883)	(0,0816 – 0,0816)	(0,0009 – 0,0009)
Teste 6 (todas)					
<i>P. fasciatum</i> (143)	Média	0,0012 ± 0,0000			
	IC 95%	(0,0012 – 0,0012)			
	IC 99%	(0,0012 – 0,0012)			
<i>P. tigrinum</i> (76)	Média	0,0281 ± 0,0000	0,0009 ± 0,0000		
	IC 95%	(0,0281 – 0,0281)	(0,0009 – 0,0009)		
	IC 99%	(0,0281 – 0,0281)	(0,0009 – 0,0009)		
<i>P. reticulatum</i> (10)	Média	0,0245 ± 0,0000	0,0209 ± 0,0000	0,0009 ± 0,0000	
	IC 95%	(0,0245 – 0,0245)	(0,0209 – 0,0209)	(0,0009 – 0,0009)	
	IC 99%	(0,0245 – 0,0245)	(0,0209 – 0,0209)	(0,0009 – 0,0009)	
<i>P. magdaleniatum</i> (9)	Média	0,0777 ± 0,0000	0,0884 ± 0,0000	0,0816 ± 0,0000	0,0009 ± 0,0000
	IC 95%	(0,0777 – 0,0777)	(0,0884 – 0,0884)	(0,0816 – 0,0816)	(0,0009 – 0,0009)
	IC 99%	(0,0777 – 0,0777)	(0,0884 – 0,0884)	(0,0816 – 0,0816)	(0,0009 – 0,0009)

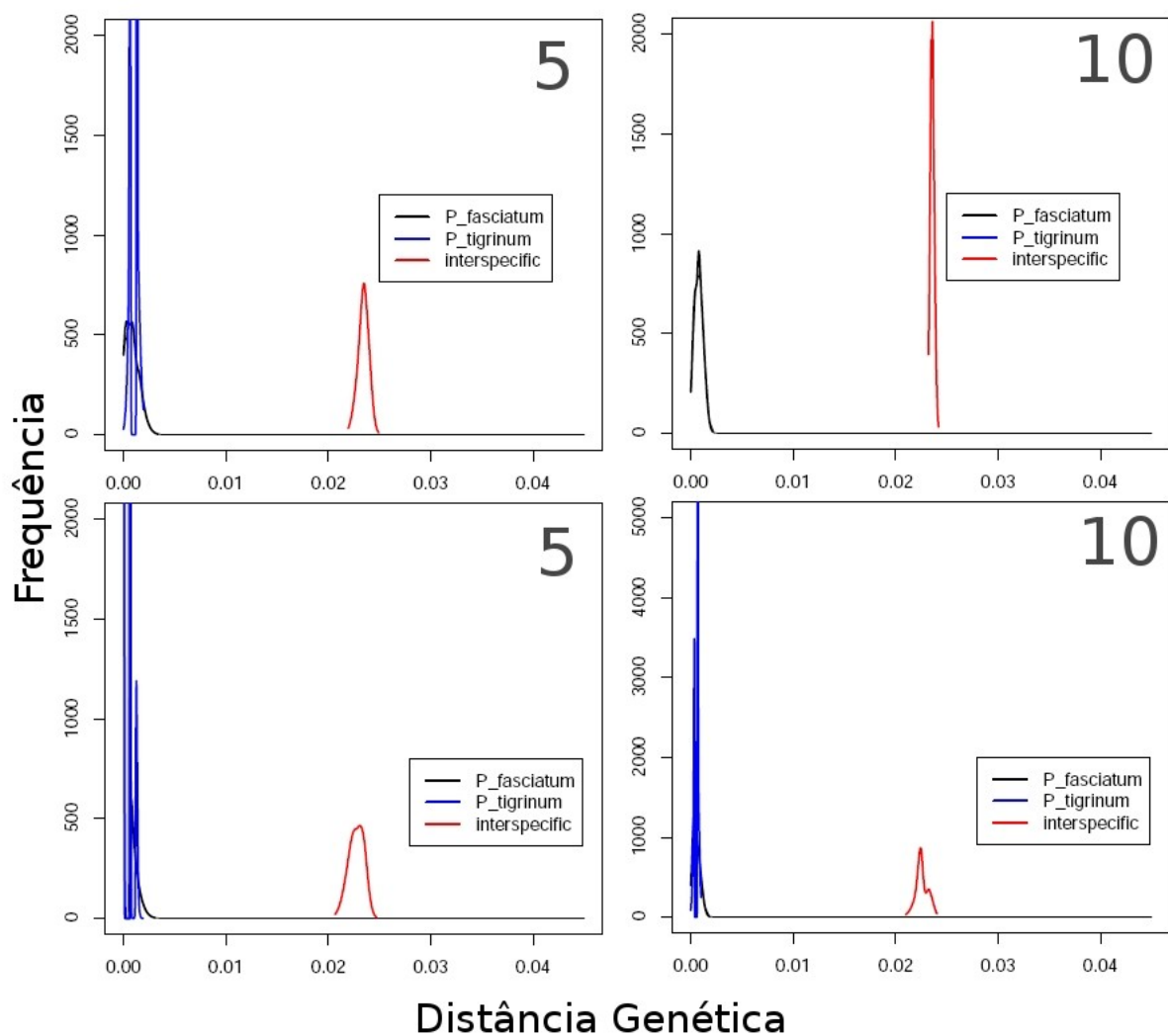


Figura 12. Representação do *barcoding gap* entre as espécies de *Pseudoplatystoma* (frequência absoluta de amostragens de distâncias par-a-par). Acima reamostragens de uma população da cada espécie (*P. fasciatum* e *P. tigrinum*) e abaixo reamostragens com duas populações de cada. A esquerda os resultados dos teste com reamostragens de 5 indivíduos e à direita de 10 indivíduos (banco completo).

Pseudoplatystoma

Nenhuma potencial nova espécie de *Pseudoplatystoma* foi encontrada. Quanto a ocorrências das espécies, *P. fasciatum* foi encontrada em toda a área amostrada na Bacia Amazônica. *Pseudoplatystoma tigrinum* foi encontrada

principalmente ao longo da calha principal do Rio Amazonas, assim como, nos Rios Madeira, Purus, Juruá e Branco. *Pseudoplatystoma reticulatum* foi encontrada apenas no Rio Negro. E *P. magdaleniatum* apenas no Rio Magdalena, da bacia de mesmo nome.

DISCUSSÃO

Banco de Dados

Primeiramente foi construído e analisado um banco dados padrão compostos apenas por sequências com referenciais em imagens dos indivíduos, o qual pode ser usado como banco de referência seguro para identificação *a posteriori* de novos espécimens. Este obteve 100% de identificação correta entre as identificações *a priori* e o resultado da identificação *a posteriori* do *DNA Barcoding*. Assim, pode-se concluir que este foi efetivo na identificação das espécies do gênero *Pseudoplatystoma*. Desta forma, podemos usar o *DNA Barcoding* para identificação *a posteriori* de amostras sem referenciais em imagens ou até mesmo de origem desconhecida com confiança.

O banco de dados completo com todas as amostras, com e sem referências em imagens, obteve 91,26% de identificação correta entre as espécies identificadas *a posteriori* pelas sequências e a identificação atribuída *a priori*. Em virtude da confiabilidade do primeiro banco e também dos altos valores de confiança que sustentam os clados na árvores de distâncias pode-se afirmar que os 8,74% de não correspondência seriam melhor explicados por erros na identificação *a priori* dos espécimens e não pela ineficiência da identificação do *DNA Barcoding* ou compartilhamento de haplótipos entre as espécies.

A incongruência nesse caso pôde ser testada em virtude de um conjunto de amostras referencias, o que permitiu explicá-la. Os trabalhos que usam *DNA Barcoding* na identificação de espécies podem ter os mesmos tipos de problemas na identificação *a priori* dos espécimens. Isso vem a corroborar que parte dos

resultados de ineficiência na identificação por *DNA Barcoding* está nos bancos de amostras identificadas *a priori*, e possivelmente, também associado à não monofilia de algumas espécies. Assim, um teste de validação com amplas amostragem seria uma possibilidade interessante em casos de estudos mais detalhados.

Diversidade Genética

A composição nucleotídica dos 652 pares de bases do gene COI de *Pseudoplatystoma* é muito próxima com a relatada para teleósteos. As proporções nucleotídicas mantêm as mesmas proporções destes, com a base timina sendo a mais comum, depois citosina, adenina e por último guanina. O conteúdo de GC, considerando o banco completo, foi de 45,07% comparado aos 47,1% relatados para os teleósteos (Ward *et al.*, 2005). Os valores de GC nos códons também apresentam um padrão de diminuição do seu conteúdo no sentido da primeira para a terceira base.

Os resultados das distâncias intraespecíficas de 0,103% ficaram abaixo da média de 0,34% constatada por Ward (2009) para os peixes, o que evidencia uma baixa variabilidade genética nas espécies do gênero, embora, o autor ressalte que esse valor mais elevado em seu estudo quando comparada com as aves (0,22%), pode ser ao menos em parte por erros taxonômicos (não reconhecimento de complexos de espécies crípticas) e/ou alguns erros de identificação. No entanto, também podem ser explicadas por diferenças nas taxas evolutivas (Kumar, 2005) ou nas diferentes histórias evolutivas das espécies (Ward, 2009).

Considerando o banco completo, as diversidades haplotípicas e em especial as nucleotídicas foram muito baixas, destacando a diferença quando comparadas com as diversidades, de uma espécie do mesmo gênero, de *P. corruscans* ($h = 0,814$; $\pi = 0,007$) e *P. reticulatum* ($h = 0,952$; $\pi = 0,007$) (Bignotto *et al.*, 2009). No entanto, muito provavelmente estas diferenças sejam muito mais relacionadas as diferenças entre os marcadores do que diferenças entre as espécies ou populações.

Os resultados dos testes D de Tajima e F_s de Fu indicam algum tipo de

desequilíbrio apenas para espécie *P. fasciatum*. Isso pode sugerir uma expansão populacional recente. As implicações disso na eficiência do *DNA Barcoding* ainda não são avaliadas, no entanto essa expansão populacional recente pode fazer com que a diversidade intraespecífica observada seja menor do que a esperada de modo a facilitar a existência do *barcoding gap*.

Validação do *DNA Barcoding*

O *DNA Barcoding* é uma ferramenta molecular poderosa com muito potencial na identificação de espécies. No entanto a maior parte dos trabalhos ainda estão restritos a região neártica, embora em todas as regiões do mundo os trabalhos sejam crescentes. A região neotropical, a mais biodiversa, ainda apresenta poucos trabalhos representativos, como aves da Argentina (Kerr *et al.*, 2009), morcegos da Guiana (Clare *et al.*, 2007) e peixes do México e Guatemala (Valdez-Moreno *et al.*, 2009). A Amazônia, em especial, ainda é uma grande lacuna nas amostragens do *DNA Barcoding* e sem dúvida será um grande desafio para este.

Neste trabalho foram realizados testes de validação do *DNA Barcoding* onde foram testadas monofilia dos grupos, estruturação das populações e o *barcoding gap* em um gênero de grande bagres amazônicos. Todas as espécies de *Pseudoplatystoma* analisadas no presente trabalho sempre mostraram-se monofiléticos e às divergências intraespecíficas sempre foram inferiores as interespecíficas, ou seja, foi verificado o *barcoding gap* em todos os casos. Por isso, a identificação das espécies de *Pseudoplatystoma* por *DNA Barcoding* mostrou-se eficiente neste caso. Quando avaliada a variação ao longo da distribuição das espécies, não houve nenhum tipo de estruturação em função dos pontos de coletas. Por isso conclui-se que as populações de *Pseudoplatystoma* são panmíticas.

Deste forma este trabalho contribui na validação da eficácia do *DNA Barcoding* na identificação de espécies, em especial de peixes. Isso porque que, este foi efetivo na diferenciação das espécies do gênero *Pseudoplatystoma*, sendo

identificadas quatro espécies de um total de seis espécies propostas para o gênero. Tal resultado mostra que é possível a identificação de espécies próximas mesmo com uma ampla amostragem ao longo das distribuições das espécies, visto que foram coletadas amostras de duas espécies em 21 pontos ao longo da bacia amazônica. As divergências intra e interespecíficas foram 0,10% e 3,08% respectivamente, o que representa uma grandeza de 30 vezes, semelhante a de 25 vezes reportada por Ward *et al.* (2005). Desta forma, esses resultados também indicam que o *barcoding gap* pode existir mesmo com uma ampla amostragem e independentemente do nível de relação entre as espécies estudadas, neste caso o *barcoding gap* não seria um artefato de amostragem insuficiente ao longo do taxa.

Os resultados mostram que um critério para delimitação das espécies baseado em um limiar (2% ou 5%) não se aplicaria neste caso. Visto que há divergências, tanto intraespecíficas quanto interespecíficas, abaixo desses valores no caso das espécies *P. tigrinum* e *P. reticulatum* que possuem valores de 1,6% de divergência (Tabela 12). Neste caso será a presença do *barcoding gap*, o fato determinante que diferencia uma espécie da outra e o critério de divergência de 10x entre divergências intraespecíficas e interespecíficas não se aplicaria.

É sabidamente conhecido que a quantidade de amostras é uma variável que influencia na eficiência do *DNA Barcoding*. Este trabalho pode ajudar no melhor entendimento do tamanho dessa influência. O *barcoding gap* foi verificado em todos os testes realizados, no entanto percebeu-se que este pode ter uma redução de cerca de dez vezes dependendo da amostragem, ou seja, isso quer dizer um *barcoding gap* que em geral pode ter uma grandeza de 30 vezes entre divergência intraespecífica e interespecífica, pode ser reduzido para menos de 3 vezes com o aumento da amostragem. As implicações disso nos resultados do *DNA Barcoding* sem dúvidas são grandes.

Em geral, são usados como padrão cinco indivíduos por espécie no uso do *DNA Barcoding* (Ward *et al.*, 2009), no entanto muitas espécies, se não a maioria, possuem o *DNA Barcode* de menos de cinco espécimens, muitas ainda com apenas

um ou dois espécimens. As implicações disso são importantes na medida que a divergência intraespecífica pode ser subestimada e a divergência interespecífica superestimada gerando um aparente *barcoding gap* inexistente ou simplesmente maior que o real.

Hoje, por exemplo, já são mais de 9000 espécies de peixes que possuem *DNA Barcode*, com uma média de seis espécimens por espécie (BOLD, consultado em Janeiro de 2010). Este esforço se deve principalmente pelo *Barcoding Fish* (*FishBOL*), um grande esforço internacional para catalogação de todas as espécies de peixes por *DNA Barcoding*. Na medida que os bancos de *DNA Barcode* fiquem completos, de forma a representar todas as espécies com uma média de cinco espécimens por espécie, possivelmente isso não será o suficiente para catalogação de toda a biodiversidade, visto que a diversidade intraespecífica ainda estará subestimada para muitas espécies. Uma maneira de toda a biodiversidade ter maior chance de ser catalogada por completo seria a avaliação mais refinada das variações intraespecíficas. Isso, no entanto, requereria amostragens maiores tanto do número de espécimens, quanto mais pontos ao longo da distribuição das espécies em muitos dos casos.

Teste para validação

Devido a necessidade de dar maior confiabilidade aos resultados do *DNA Barcoding*, propõe-se aqui um teste para validação deste. O principal objetivo desse teste é falsear a hipótese de que um grupo seja monofilético. Além de testar a monofilia também será testado se há estruturação de populações e presença ou não do *barcoding gap*. A lógica do teste visa sempre falsear a hipótese de que um grupo seja monofilético com o aumento do número de amostras de um mesmo ponto amostral inicial ao longo da distribuição da espécie e/ou do aumento dos pontos amostrais. Caso uma determinada espécie também apresente estruturação em função dos pontos amostrais a hipótese dessa estruturação também será testada (Figura 13).

A importância de se testar a estruturação das populações é devida ao fato de que quando se amostra apenas um ponto ao longo da distribuição de uma dada espécie, a sua divergência intraespecífica pode estar sendo subestimada e consequentemente superestimando-se o *barcoding gap*. Assim, para amostrar-se de uma maneira representativa a diversidade intraespecífica, precisar-se-ia amostrar um número suficiente de indivíduos e em caso da existência de estrutura populacional, estes indivíduos precisam ser amostrados em diferentes localidades. Para o *DNA Barcoding*, a amostragem de múltiplas localidades distantes entre si teria vantagem em relação a amostrar-se na mesma localidade, porque aumentam-se as chances de encontrar mais alelos nas espécies (diversidade gênica) e caso exista estrutura (se estes alelos estão estruturados geograficamente), a amostragem do outro ponto ajudaria a resolver um problema de subamostragem.

Como não há um número mágico, mas em geral se usam cinco amostras por espécie, este será o número inicial padrão para o teste. Assim, um novo ponto de amostragem necessariamente significará mais cinco amostras desse ponto, da mesma forma que mais amostras de um mesmo ponto também significarão mais cinco amostras desse mesmo ponto.

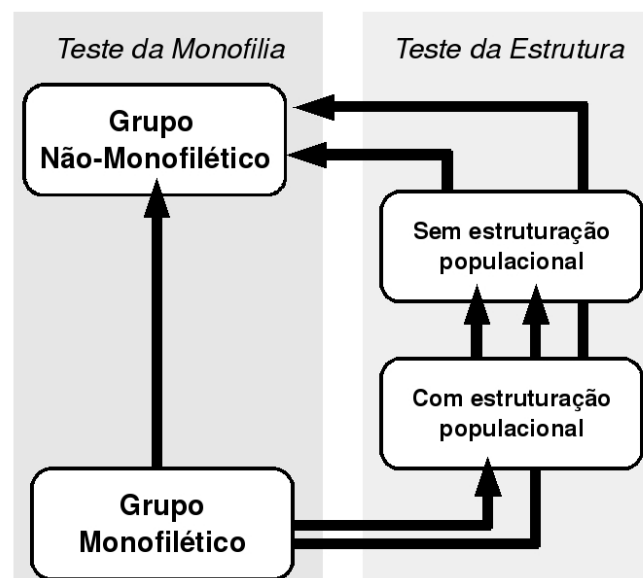


Figura 13. Diagrama dos testes de monofilia e estruturação dos grupos mostrando que o acréscimo de amostras ou pontos amostrais sempre será no sentido de falsear a hipótese de que os grupos

sejam monofiléticos ou estruturados.

O teste dar-se-ia da seguinte forma: inicialmente amostra-se UM ponto ao longo da distribuição de pelo menos DUAS espécies. Como resultado inicial há duas possibilidades, os grupos podem ser MONOFILÉTICOS ou NÃO-MONOFILÉTICOS. Se os grupos apresentarem-se NÃO-MONOFILÉTICOS, independente de amostragens de outras localidades ou mais indivíduos da mesma localidade, estes permanecerão NÃO-MONOFILÉTICOS, assim não cabendo ampliar a amostragem para negar a monofilia do grupo. O fato que explicaria tal resultado possivelmente estaria no recente tempo de divergência destas espécies ou ao fato do grupo necessitar de uma revisão. Funk e Omland (2003) avaliaram que 23% das espécies de metazoários não são monofiléticas. Para o caso do grupo apresentar-se MONOFILÉTICO, seria necessária a ampliação da amostragem. Há duas possibilidades, amplia-se o número de amostra do mesmo ponto ou amostra-se outro ponto. Esta última possibilidade é mais poderosa, porque além de testar a monofilia será testado se há estruturação. Seriam então amostrado DOIS pontos ao longo da distribuição de pelo menos DUAS espécies (Figura 14).

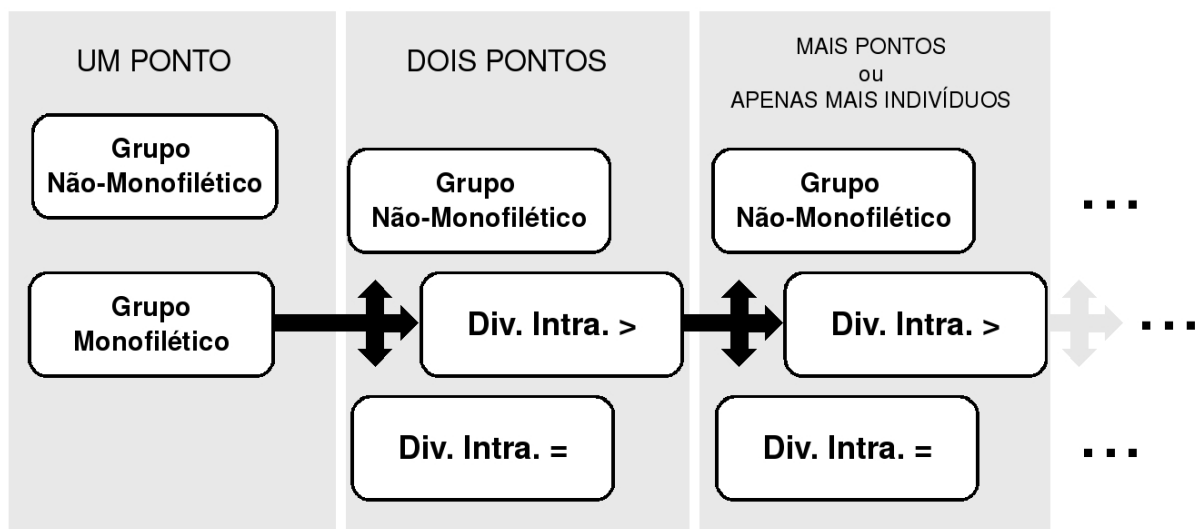


Figura 14. Diagrama do desenho amostral do teste de validação do *DNA Barcoding*.

Como resultado da amostragem de dois pontos os grupos que eram MONOFILÉTICOS podem se mostrar que de fato eram NÃO-MONOFILÉTICOS ou continuariam MONOFILÉTICOS. Preferencialmente, novos pontos deveriam ser o mais distantes possíveis do ponto inicial. No caso dos grupos manterem-se MONOFILÉTICOS, estes poderiam apresentar ou não estruturação em função dos pontos de coleta. No caso de não apresentarem estruturação e não apresentarem mudança significativa da divergência intraespecífica não seria mais necessária a ampliação da amostragem. Isso indicaria que seria muito provável que por mais pontos e exemplares que se amostrasse, a diferença não mudaria significativamente para revelar que os grupos seriam de fato NÃO-MONOFILÉTICOS. Apresentando estruturação em função dos pontos de coleta e mudança significativa da divergência intraespecífica deveria-se novamente aumentar-se a amostragem. Agora o teste seria em função de explicar a mudança na divergência, se esta seria em função do novo ponto de amostragem ou apenas em função das novas amostras. Assim, seriam acrescidas novas amostras dos mesmos dois pontos. Os grupos mantendo-se MONOFILÉTICOS poderiam não mais apresentar mudança de divergência, e assim concluiria-se que a variação anterior foi em função dos pontos e não em função do número de amostras, o que então pediria novos pontos. Ou se o aumento de amostras dos mesmos pontos mudasse a divergência, poderia se concluir que esta mudou em função do aumento de amostras, o que pediria mais amostras dos mesmos pontos. Esse processo de acréscimo de novos pontos de amostragem e/ou novas amostras continuariam até a divergência não mudar significativamente.

A conclusão do teste é que se houvesse uma estruturação em função dos pontos importaria onde coletar, mas caso não houvesse não importaria o local nem o número de pontos de coleta. Outra conclusão seria acerca do tamanho amostral, como resultado poderia se chegar a um tamanho mínimo de amostras dependente da variável diversidade intraespecífica.

A proposição deste teste não tem o intuito de sugerir que para toda catalogação de espécies por *DNA Barcoding* haveria de ser realizado tal avaliação. Em geral o uso do *DNA Barcoding* tem por objetivo revelar um padrão geral da

biodiversidade conhecida e principalmente da desconhecida. Assim, esta proposição se aplica a abordagens mais detalhadas que podem auxiliar na identificação de padrões de diversidade e estruturação das espécies e como forma de validação mais apurada de uso do *DNA Barcoding*.

Finalmente, o que se conclui com esse estudo é que o *DNA Barcoding* é uma ferramenta muito útil de grande potencial para a caracterização da biodiversidade. No entanto, os seus resultados devem ser interpretados com cuidado levando-se em conta que estes estarão sujeitos a variáveis como tamanho amostral, diversidade intraespecífica, estruturação populacional e monofilia das espécies.

Pseudoplatystoma

A filogenia do gênero *Pseudoplatystoma* foi recentemente publicada (Torrico *et al.*, 2009), no entanto ela ainda deixa algumas questões em aberto. Devido a lacunas na amostragem, em especial na Amazônia Oriental e no Rio Negro (local tipo da espécie *P. reticulatum*), poderia se perguntar se ainda haveriam mais espécies que foram não identificadas nessas regiões. Se em função da baixa divergência em relação as outras espécies, *P. reticulatum* poderia ser considerada uma espécie. Visto que Buitrago-Suarez e Burr (2007) e Torrico *et al.* (2009) não amostraram no Rio Negro, a distribuição de *P. reticulatum* é real? Esta espécie estaria de fato no Rio Negro? Ou seriam duas espécies?

Os resultados do presente trabalho apontam para resultados que podem ajudar a responder tais questões. Foram identificadas quatro espécies que mostram-se monofiléticas, são elas: *P. fasciatum*, *P. tigrinum*, *P. reticulatum* e *P. magdaleniatum*. Desta forma nenhuma nova espécie foi identificada dentre as quatro analisadas de um total de seis propostas para o gênero, visto que foram inclusas neste trabalho amostram da Amazônia oriental e do Rio Negro, regiões estas não amostradas anteriormente. Além disso, há o fato da variação fenotípica no padrão de manchas que poderia suscitar a presença de mais espécies, mas este fenômeno de fato mostrou-se ser apenas variação fenotípica intraespecífica.

Interessantemente, *P. reticulatum* foi encontrada no Rio Negro, local onde seu holótipo foi descrito, sendo esta então a primeira confirmação genética da espécie na bacia da descrição original, visto que Buitrago-Suarez e Burr (2007) e Torrico *et al.* (2009) não amostraram tal espécie no Rio Negro. Isso já responde a pergunta se de fato esta espécie estaria no Rio Negro, embora ainda não deixe completamente clara sua exata relação com a população da Bacia do Paraná.

Quanto as áreas de distribuição, ficou claro a distribuição mais ampla de *P. fasciatum* em relação a *P. tigrinum*. O primeiro se distribui por toda a bacia amazônica, enquanto o segundo não foi encontrado em importantes tributários do Rio Amazonas, como o Negro, Tapajós, Xingu e Tocantins. Aparentemente há uma substituição de *P. tigrinum* por *P. reticulatum* no Rio Negro.

CONCLUSÕES

Conclui-se que o *DNA Barcoding* pode ser efetivo na identificação de espécies e que o *barcoding gap* pode existir mesmo com uma ampla amostragem e independentemente do nível de relação entre as espécies estudadas, não sendo um artefato de amostragem insuficiente ao longo da distribuição do taxa. Mas apesar de sua eficiência, esta pode ser influenciada significativamente por variáveis como tamanho amostral, diversidade intraespecífica, estruturação populacional e monofilia das espécies.

Por fim, verificou-se que o *DNA Barcoding* foi eficiente para identificação do gênero *Pseudoplatystoma* sendo este composto por três espécies na Amazônia (*P. fasciatum*, *P. tigrinum* e *P. reticulatum*) e observou-se uma alta correspondência entre os clados identificados *DNA Barcoding* e as espécies identificadas morfologicamente *a priori*.

REFERÊNCIAS BIBLIOGRÁFICAS

Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25: 3389–3402.

Bignotto, T. S.; Prioli, A. J.; Prioli, S. M.; Maniglia, T. C.; Boni, T. A.; Lucio, L. C.; Gomes, V. N.; Prioli, R. A.; Oliveira, A.; Julio-Junior, H. F.; Prioli, L. M. 2009. Genetic divergence between *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* (Siluriformes: Pimelodidae) in the Parana River Basin. *Braz. J. Biol.*, 69(2): 681–689.

Borisenko, A.V.; Lim, B.K.; Ivanova, N.V.; Hanner, R.H.; Hebert, P.D.N. 2008. DNA barcoding in surveys of small mammal communities: a field study in Suriname. *Mol. Ecol. Res.*, 8(3): 471–479.

Buitrago-Suárez, U.A.; Burr, B.M. 2007. Taxonomy of the catfish genus *Pseudoplatystoma* Bleeker (Siluriformes: Pimelodidae) with recognition of eight species. *Zootaxa*, 1512: 1–38.

Clare, E.L.; Lim, B.K.; Engstrom, M.D.; Eger, J.L.; Hebert, P.D.N. 2007. DNA barcoding of Neotropical bats: species identification and discovery within Guyana. *Mol. Ecol. Notes*, 7: 184–190.

Dasmahapatra, K.K.; Mallet, J. 2006. DNA barcodes: recent successes and future prospects. *Heredity*, 97: 254–255.

de Queiroz, K. 2005. Different species problems and their resolution. *Bioessays*, 27(12): 1263–1269.

Elasser, S.C.; Floyd, R.; Hebert, P.D.N.; Schulte-Hostedde, A.I. 2009. Species identification of North American guinea worms (Nematoda: Dracunculus) with DNA barcoding. *Mol. Ecol. Res.*, 9: 707–712.

Elias, M.; Hill, R.I.; Willmott, K.R.; Dasmahapatra, K.K.; Brower, A.V.Z.; Mallet, J.; Jiggins, C.D. 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. *P. Roy. Soc. B-Biol. Sci.*, 274, 2881–2889.

Farias, I.P.; Hrbek, T. 2008. Patterns of diversification in the discus fishes (Symphysodon spp. Cichlidae) of the Amazon basin. *Mol. Phylogenet. Evol.*, 49(1):

32-43.

Felsenstein, J.F. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39: 783–791.

Funk, D.J.; Omland, K.E. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Syst.*, 34: 397–423.

Hall, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41: 95-98.

Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; DeWaard, J.R. 2003. Biological identifications through DNA barcodes. *P. Roy. Soc. B-Biol. Sci.*, 270: 313–321.

Hebert, P.D.N.; deWaard, J.R.; Landry, J.F. 2009. DNA barcodes for 1/1000 of the animal kingdom. *Biol. Lett.*, Online.

Hey, J. 2006. On the failure of modern species concepts. *Trends Ecol. Evol.*, 21(8): 447-450.

Hickerson, M.J.; Meyer, C.P.; Moritz, C. 2006. DNA Barcoding Will Often Fail to Discover New Animal Species over Broad Parameter Space. *Syst. Biol.*, 55(5):729–739.

Kandul, N.P.; Lukhtanov, V.A.; Dantchenko, A.V.; Coleman, J.W.S.; Sekercioglu, C.H.; Haig, D.; Pierce, N.E. 2004. Phylogeny of *Agrodiaetus* Hubner 1822 (Lepidoptera: Lycaenidae) inferred from mtDNA sequences of COI and COII and nuclear sequences of EF1-alpha: karyotype diversification and species radiation. *Syst. Biol.*, 53: 278–298.

Kelly, R.P.; Sarkar, I.N.; Eernisse, D.J.; DeSalle, R. 2007. DNA barcoding using chitons (genus *Mopalia*). *Mol. Ecol. Notes*, 7: 177–183.

Kerr, K.C.R.; Lijtmaer, D.A.; Barreira, A.S.; Hebert, P.D.N.; Tubaro, P.L. 2009. Probing Evolutionary Patterns in Neotropical Birds through DNA Barcodes. *PLoS ONE*, 4(2): e4379.

Kerr, K.C.R.; Stoeckle, M.Y.; Dove, C.J.; Weigt, L.A.; Frances, C.M.; Hebert, P.D.N. 2007. Comprehensive DNA barcode coverage of North American birds. *Mol. Ecol. Notes*, 7: 535–543.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, 16: 111–120.

Kress, W.J.; Wurdack, K.J.; Zimmer, E.A.; Weigt, L.A.; Janzen, D.H. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA*, 102: 8369–8374.

Kumar, S. 2005. Molecular clocks: four decades of evolution. *Nature Rev. Genet.*, 6: 654–662.

Lawton, J.H.; Bignell, D.E.; Bolton, B.; Bloemers, G.F.; Eggleton, P.; Hammond, P.M.; Hodda, M.; Holt, R.D.; Larsen, T.B.; Mawdsley, N.A.; Stork, N.E.; Srivastava, D.S.; Watt, A.D. 1998. Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, 391: 72–75.

Leveque, C.; Oberdorff, T.; Paugy, D.; Stiassny, M.L.J.; Tedesco, P.A. 2008. Global diversity of fish (Pisces) in freshwater. *Hydrobiologia*, 595: 545–567.

Lukhtanov, V.A.; Sourakov, A.; Zakharov, E.V.; Hebert, P.D.N. 2009. DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce the success of species identification. *Mol. Ecol. Res.*, 9: 1302–1310.

Matz, M.V.; Nielsen, R. 2005. A likelihood ratio test for species membership based on DNA sequence data. *Phil. Trans. R. Soc. B*, 360, 1969–1974.

May, R. M. 1988. How Many Species are There on Earth? *Science*, 241: 1441–1449.

Meier, R.; Shiyang, K.; Gaurav, V.; Ng, P.K.L. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst. Biol.*, 55: 715–728.

Meyer, C.P.; Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.*, 3: e422.

Millennium Ecosystem Assessment. 2005. *Ecosystems and Human Well-Being: Biodiversity Synthesis*. World Resources Institute, Washington, DC. 86 pp.

Moritz, C.; Cicero, C. 2004. DNA barcoding: promise and pitfalls. *PLoS Biol.*,

2: e354.

Nei, M.; Kumar, S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York, USA.

Nicholas, K.B.; Nicholas Jr, H.B. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors.

Palumbi, S.R.; Cipriano, F.; Hare, M.P. 2001. Predicting nuclear gene coalescence from mitochondrial data: The three-times rule. *Evolution*, 55(5): 859–868.

R Development Core Team. 2005. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.

Ratnasingham, S.; Hebert, P.D.N. 2007. BOLD: the barcode of life data system (www.barcodinglife.org). *Mol. Ecol. Notes*, 7: 355–364.

Saitou, N.; Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4: 406–425.

Sambrook, J.; Fritsch, E.F.; Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Springs Harbor Laboratory Press, Cold Springs Harbor, NY, USA. 1659p.

Sanger, F.; Nicklen, S.; Coulson, A.R. 1977. DNA Sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 74: 5463-5467.

Santos, G.M.; Ferreira, E.J.G.; Zuanon, J. *Peixes comerciais de Manaus*. 1.ed. Manaus, AM: IBAMA/AM, Provárzea, 2006. v. 1. 144p.

Saunders, G.W. 2005. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Phil. Trans. R. Soc. B*, 360: 1879–1888.

Scicluna, S.M.; Tawari, B.; Clark, C.G. 2006. DNA barcoding of Blastocystis. *Protist*, 157: 77–85.

Sites, J.W.; Marshall, J.C. 2003. Delimiting species: a Renaissance issue in systematic biology. *Trends Ecol. Evol.*, 18(9): 462–470.

Sogin, M.L.; Morrison, H.G.; Huber, J.A.; Welch, D.M.; Huse, S.M.; Neal, P.R.;

Arrieta, J.M.; Herndl, G.J. 2006. Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc. Natl. Acad. Sci. USA*, 103: 12115–12120.

Staden, R. 1996. The Staden sequence analysis package. *Mol. Biotechnol.*, 5: 233–241.

Steinke, D.; Zemlak, T.S.; Hebert, P.D.N. 2009. Barcoding Nemo: DNA-based identifications for the ornamental fish trade. *PLoS One*, 4: e6300.

Summerbell, R.C.; Lévesque, C.A.; Seifert, K.A.; Bovers, M.; Fell, J.W.; Diaz, M.R.; Boekhout, T.; de Hoog, G.S.; Stalpers, J.; Crous, P.W. 2005. Microcoding: the second step in DNA barcoding. *Philos. T. R. Soc. B*, 360: 1897–1903.

Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, 24: 1596–1599.

Tavares, E.S.; Baker, A.J. 2008. Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. *BMC Evol. Biol.*, 8: 81.

Thompson, J.D.; Higgins, D.G.; Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22: 4673–4680.

Toffoli, D.; Hrbek, T.; de Araujo, M.L.G.; de Almeida, M.P.; Charvet-Almeida, P.; Farias, I.P. 2008. A test of the utility of DNA barcoding in the radiation of the freshwater stingray genus *Potamotrygon* (Potamotrygonidae, Myliobatiformes). *Genet. Mol. Biol.*, 31(1): 324–336.

Torrico, J.P.; Hubert, N.; Desmarais, E.; Duponchelle, F.; Rodriguez, J.N.; Montoya-Burgos, J.; Davila, C.G.; Carvajal-Vallejos, F.M.; Grajales, A.A.; Bonhomme, F.; Renno, J.F. 2009. Molecular phylogeny of the genus *Pseudoplatystoma* (Bleeker, 1862): Biogeographic and evolutionary implications. *Mol. Phylogenet. Evol.*, 51(3): 588–594.

Valdez-Moreno, M.; Ivanova, N.V.; Elías-Gutiérrez, M.; Contreras-Balderas, S.; Hebert, P.D.N. 2009. Probing diversity in freshwater fishes from Mexico and Guatemala with DNA barcodes. *J. Fish Biol.*, 74(2): 377–402.

Ward, R.D.; Zemlak, T.S.; Innes, B.H.; Last, P.R.; Hebert, P.D.N. 2005. DNA barcoding Australia's fish species. *Philos. T. R. Soc. B*, 360: 1847–1857.

Ward, R.D. 2009. DNA barcode divergence among species and genera of birds and fishes. *Mol. Ecol. Res.*, 9(4): 1077–1085.

Ward, R.D.; Hanner, R.; Hebert, P.D.N. 2009. The campaign to DNA barcode all fishes, FISH-BOL. *J. Fish Biol.* 74(2): 329–356.

Wiemers, M.; Fiedler, K. 2007. Does the DNA barcoding gap exist? — a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front. Zool.*, 4: 8.

Willis, S.C.; Nunes, M.S.; Montana, C.G.; Farias, I.P.; Lovejoy, N.R. 2007. Systematics, biogeography, and evolution of the neotropical peacock basses *Cichla* (Perciformes : Cichlidae). *Mol. Phylogenet. Evol.*, 44(1): 291–307.

Xia, X.; Xie, Z.; 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered.*, 92: 371–373.

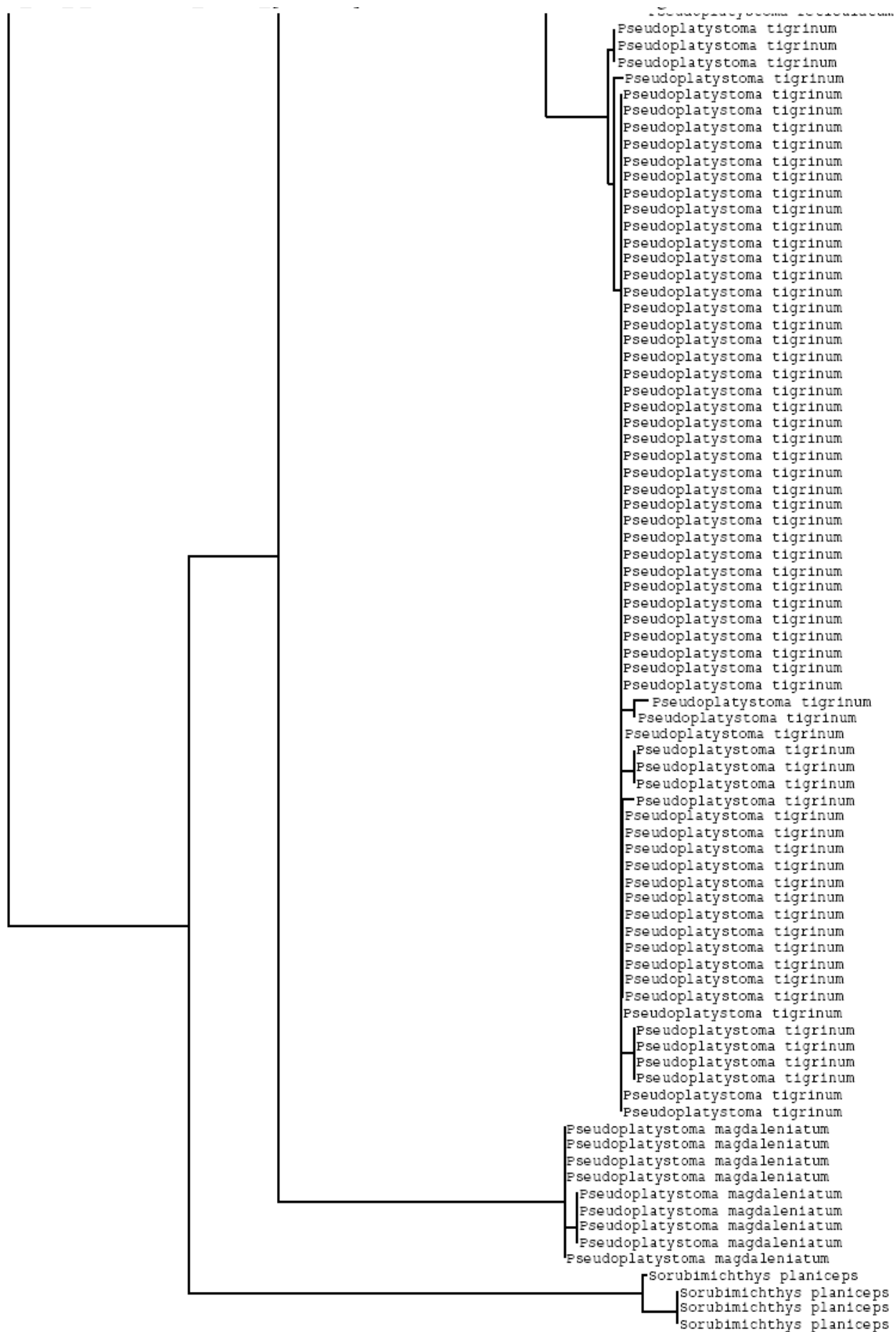
Zhang, A.B.; He, L.J.; Crozier, R.H.; Muster, C.; Zhu, C.-D. 2009. Estimating sample sizes for DNA barcoding, *Mol. Phylogenet. Evol.*, In Press.

ANEXOS

Anexo A. Árvore Neighbour-joining da divergência nucleotídica das sequências do COI entre as quatro espécies do gênero *Pseudoplatystoma* e dois grupos externos, usando o modelo de distância Kimura-2-parâmetros com análise de bootstrap 1000 réplicas.

[illegible]

[illegible]



Anexo B. Funções usadas para os cálculos.

Função (*BarcodeDistance.r*)

```
BarcodeDistance <- function(sp1,sp2,subsample=5,permutations=1000,dist.model="K80")
{

library(ape)
library(seqinr)

#---initialize variables---

sp1.num <- length(sp1)
sp2.num <- length(sp2)

# if(sp1.num < subsample) {sp1.subsample = sp1.num} else {sp1.subsample = subsample}
# if(sp2.num < subsample) {sp2.subsample = sp2.num} else {sp2.subsample = subsample}

sp1.subsample <- ifelse(sp1.num < subsample, sp1.num, subsample)
sp2.subsample <- ifelse(sp2.num < subsample, sp2.num, subsample)

cutoff.95 <- as.integer(permutations*0.95)
cutoff.99 <- as.integer(permutations*0.99)
cutoff.05 <- as.integer(permutations*0.05)
cutoff.01 <- as.integer(permutations*0.01)

sp1.max1 <- rep(NA, permutations)
sp1.mean1 <- rep(NA, permutations)
sp2.max1 <- rep(NA, permutations)
sp2.mean1 <- rep(NA, permutations)
inter.max1 <- rep(NA, permutations)
inter.mean1 <- rep(NA, permutations)

#----calculate metrics----
for (i in 1:permutations)
{

#----permute sequences----
sp1 <- as.alignment(sp1)
seq.perm <- sample(sp1$seq,replace=F)
sp1$seq <- seq.perm
sp1 <- as.DNABin(sp1)
sp1.sub <- sp1[c(1:sp1.subsample),]

sp2 <- as.alignment(sp2)
seq.perm <- sample(sp2$seq)
sp2$seq <- seq.perm
sp2 <- as.DNABin(sp2)
sp2.sub <- sp2[c(1:sp2.subsample),]

#-----
write.dna(sp1.sub,file='mydata.fas',format='fasta')
write.dna(sp2.sub,file='mydata.fas',format='fasta',append=T)
inter.sub <- read.dna(file='mydata.fas',format='fasta')
sequences <- length(inter.sub)

#-----calculate DNA distance-----
x <- dist.dna(sp1.sub,model=dist.model)
sp1.max1[i] <- max(x)
```



```

sp1.mean1[i] <- mean(x)

x <- dist.dna(sp2.sub,model=dist.model)
sp2.max1[i] <- max(x)
sp2.mean1[i] <- mean(x)

y <- rep(NA, sp1.subsample*sp2.subsample)
z <- 0
x <- dist.dna(inter.sub,model=dist.model,as.matrix=T)
for (j in (sp1.subsample+1):sequences)
{
  for (k in 1:sp1.subsample)
  {
    z <- z+1
    y[z] <- x[j,k]
  }
}
inter.max1[i] <- max(y)
inter.mean1[i] <- mean(y)
}

#---generate output---
barcode.distance <- data.frame(sp1.mean1,sp1.max1,sp2.mean1,sp2.max1,inter.mean1,inter.max1)
return(barcode.distance)
}

```

Sumário (*BarcodeDistanceSummary.r*)

```

BarcodeDistanceSummary <-
function(sp1,sp2,subsample=5,permutations=1000,dist.model="K80",barcode.distance)
{

library(ape)
library(seqinr)

#---initialize variables----

sp1.num <- length(sp1)
sp2.num <- length(sp2)

# if(sp1.num < subsample) {sp1.subsample = sp1.num} else {sp1.subsample = subsample}
# if(sp2.num < subsample) {sp2.subsample = sp2.num} else {sp2.subsample = subsample}

sp1.subsample <- ifelse(sp1.num < subsample, sp1.num, subsample)
sp2.subsample <- ifelse(sp2.num < subsample, sp2.num, subsample)

cutoff.95 <- as.integer(permutations*0.95)
cutoff.99 <- as.integer(permutations*0.99)
cutoff.05 <- as.integer(permutations*0.05)
cutoff.01 <- as.integer(permutations*0.01)

sp1.max1 <- barcode.distance$sp1.max1
sp1.mean1 <- barcode.distance$sp1.mean1
sp2.max1 <- barcode.distance$sp2.max1
sp2.mean1 <- barcode.distance$sp2.mean1

```

```
inter.max1 <- barcode.distance$inter.max1
inter.mean1 <- barcode.distance$inter.mean1
```

```
#---summarize distances---
sp1.max <- mean(sp1.max1)
sp1.max.sd <- sd(sp1.max1)
sp1.max.sort <- sort(sp1.max1,decreasing=F)
sp1.max.c95.upper <- sp1.max.sort[cutoff.95]
sp1.max.c95.lower <- sp1.max.sort[cutoff.05]
sp1.max.c99.upper <- sp1.max.sort[cutoff.99]
sp1.max.c99.lower <- sp1.max.sort[cutoff.01]
sp1.mean <- mean(sp1.mean1)
sp1.mean.sd <- sd(sp1.mean1)
sp1.mean.sort <- sort(sp1.mean1,decreasing=F)
sp1.mean.c95.upper <- sp1.mean.sort[cutoff.95]
sp1.mean.c95.lower <- sp1.mean.sort[cutoff.05]
sp1.mean.c99.upper <- sp1.mean.sort[cutoff.99]
sp1.mean.c99.lower <- sp1.mean.sort[cutoff.01]
```

```
sp2.max <- mean(sp2.max1)
sp2.max.sd <- sd(sp2.max1)
sp2.max.sort <- sort(sp2.max1,decreasing=F)
sp2.max.c95.upper <- sp2.max.sort[cutoff.95]
sp2.max.c95.lower <- sp2.max.sort[cutoff.05]
sp2.max.c99.upper <- sp2.max.sort[cutoff.99]
sp2.max.c99.lower <- sp2.max.sort[cutoff.01]
sp2.mean <- mean(sp2.mean1)
sp2.mean.sd <- sd(sp2.mean1)
sp2.mean.sort <- sort(sp2.mean1,decreasing=F)
sp2.mean.c95.upper <- sp2.mean.sort[cutoff.95]
sp2.mean.c95.lower <- sp2.mean.sort[cutoff.05]
sp2.mean.c99.upper <- sp2.mean.sort[cutoff.99]
sp2.mean.c99.lower <- sp2.mean.sort[cutoff.01]
```

```
inter.max <- mean(inter.max1)
inter.max.sd <- sd(inter.max1)
inter.max.sort <- sort(inter.max1,decreasing=F)
inter.max.c95.upper <- inter.max.sort[cutoff.95]
inter.max.c95.lower <- inter.max.sort[cutoff.05]
inter.max.c99.upper <- inter.max.sort[cutoff.99]
inter.max.c99.lower <- inter.max.sort[cutoff.01]
inter.mean <- mean(inter.mean1)
inter.mean.sd <- sd(inter.mean1)
inter.mean.sort <- sort(inter.mean1,decreasing=F)
inter.mean.c95.upper <- inter.mean.sort[cutoff.95]
inter.mean.c95.lower <- inter.mean.sort[cutoff.05]
inter.mean.c99.upper <- inter.mean.sort[cutoff.99]
inter.mean.c99.lower <- inter.mean.sort[cutoff.01]
```

```
#---generate output---
metrics <-
c('num.indiv','num.sample','max','max.sd','max.c95.upper','max.c95.lower','max.c99.upper','max.c99.lower','mean',
  'mean.sd','mean.c95.upper','mean.c95.lower','mean.c99.upper','mean.c99.lower')
sp1.barcode <-
c(sp1.num,sp1.subsample,sp1.max,sp1.max.sd,sp1.max.c95.upper,sp1.max.c95.lower,sp1.max.c99.upper,sp1.max.c99.lower,sp1.mean,sp1.mean.sd,sp1.mean.c95.upper,sp1.mean.c95.lower,sp1.mean.c99.upper,sp1.mean.c99.lower)
sp2.barcode <-
c(sp2.num,sp2.subsample,sp2.max,sp2.max.sd,sp2.max.c95.upper,sp2.max.c95.lower,sp2.max.c99.upper,sp2.max.c99.lower,sp2.mean,sp2.mean.sd,sp2.mean.c95.upper,sp2.mean.c95.lower,sp2.mean.c99.upper,sp2.mean.c99.lower)
```

```

ax.c99.lower,sp2.mean,sp2.mean.sd,sp2.mean.c95.upper,sp2.mean.c95.lower,sp2.mean.c99.upper,sp2.mean.c99.lower)
inter.barcode <-
c(NA,NA,inter.max,inter.max.sd,inter.max.c95.upper,inter.max.c95.lower,inter.max.c99.upper,inter.max.c99.lower,inter.mean,inter.mean.sd,inter.mean.c95.upper,inter.mean.c95.lower,inter.mean.c99.upper,inter.mean.c99.lower)

barcode.distance.summary <- data.frame(metrics,sp1.barcode,sp2.barcode,inter.barcode)

return(barcode.distance.summary)
}

```

Reamostragens (*Surubim_resample.r*)

```

rm(list=ls())
setwd('diretorio')
source('BarcodeDistance.r')
source('BarcodeDistanceSummary.r')

library(ape)
library(seqinr)
library(locfit)

#---initialize variables---
sp1.name <- 'arquivofasta1'
sp2.name <- 'arquivofasta2'
subsample <- 5
permutations <- 1000
dist.model <- 'K80' # this is K2P

#---read data---
sp1 <- read.dna(file=paste(sp1.name,'COI.fas',sep='_',collapse=""),format='fasta')
sp2 <- read.dna(file=paste(sp2.name,'COI.fas',sep='_',collapse=""),format='fasta')

#---calculate distances----
barcode.distance <- BarcodeDistance(sp1,sp2,subsample,permutations,dist.model)
barcode.distance.summary <-
BarcodeDistanceSummary(sp1,sp2,subsample,permutations,dist.model,barcode.distance)

#---save results----
save(barcode.distance,file=paste('Barcode',sp1.name,sp2.name,permutations,dist.model,'permutations.Rdata',sep='_',collapse=""))
write.csv(barcode.distance,file=paste('Barcode',sp1.name,sp2.name,subsample,permutations,dist.model,'permutations.csv',sep='_',collapse=""))
write.table(barcode.distance,file=paste('Barcode',sp1.name,sp2.name,subsample,permutations,dist.model,'permutations.txt',sep='_',collapse=""))

save(barcode.distance.summary,file=paste('Barcode',sp1.name,sp2.name,permutations,dist.model,'results.summary.Rdata',sep='_',collapse=""))
write.csv(barcode.distance.summary,file=paste('Barcode',sp1.name,sp2.name,subsample,permutations,dist.model,'results.summary.csv',sep='_',collapse=""))
write.table(barcode.distance.summary,file=paste('Barcode',sp1.name,sp2.name,subsample,permutations,dist.model,'results.summary.txt',sep='_',collapse=""))

#---plot permutation data----

#sp1.den <- density(barcode.distance$sp1.mean1)
#sp2.den <- density(barcode.distance$sp2.mean1)
#inter.den <- density(barcode.distance$inter.mean1)

```

```

sp1.den <- locfit(~barcode.distance$sp1.mean1)
sp2.den <- locfit(~barcode.distance$sp2.mean1)
inter.den <- locfit(~barcode.distance$inter.mean1)

plot(sp1.den,type="l",ylim=range(0,2000),xlim=range(0,0.045),main="",xlab="",ylab="",lty=1,lwd=2,col="black")
lines(sp1.den,xlab="",ylab="",lty=1,lwd=2,col="black")
lines(sp2.den,xlab="",ylab="",lty=1,lwd=2,col="blue")
lines(inter.den,xlab="",ylab="",lty=1,lwd=2,col="red")
title(main="Barcoding",ylab="Frequency",xlab="Genetic distance")
legend(0.025,1200,c(sp1.name,sp2.name,"interspecific"),lty=c(1,1,1),lwd=c(2,2,2),col=c("black","blue","red"))

dev.copy2pdf(file=paste('Barcode',sp1.name,sp2.name,subsample,permutations,dist.model,'graph.pdf',sep='_',collapse=""))

dev.copy2eps(file=paste('Barcode',sp1.name,sp2.name,subsample,permutations,dist.model,'graph.eps',sep='_',collapse=""))

```

Considerando populações

Função (*BarcodeDistance2.r*)

```

BarcodeDistance2 <- function(sp1.1,sp1.2,sp2.1,sp2.2,subsample=5,permutations=1000,dist.model="K80")
{
  library(ape)
  library(seqinr)

  #---initialize variables---

  sp1.1.num <- length(sp1.1)
  sp1.2.num <- length(sp1.2)
  sp2.1.num <- length(sp2.1)
  sp2.2.num <- length(sp2.2)

  subsample.1.1 <- ifelse(sp1.1.num < sp1.2.num, trunc(subsample/2), subsample - trunc(subsample/2))
  subsample.1.2 <- subsample - subsample.1.1
  subsample.2.1 <- ifelse(sp2.1.num < sp2.2.num, trunc(subsample/2), subsample - trunc(subsample/2))
  subsample.2.2 <- subsample - subsample.2.1

  # if(sp1.num < subsample) {sp1.subsample = sp1.num} else {sp1.subsample = subsample}
  # if(sp2.num < subsample) {sp2.subsample = sp2.num} else {sp2.subsample = subsample}

  sp1.1.subsample <- ifelse(sp1.1.num < subsample.1.1, sp1.1.num, subsample.1.1)
  sp1.2.subsample <- ifelse(sp1.2.num < subsample.1.2, sp1.2.num, subsample.1.2)
  sp1.subsample <- sp1.1.subsample + sp1.2.subsample
  sp2.1.subsample <- ifelse(sp2.1.num < subsample.2.1, sp2.1.num, subsample.2.1)
  sp2.2.subsample <- ifelse(sp2.2.num < subsample.2.2, sp2.2.num, subsample.2.2)
  sp2.subsample <- sp2.1.subsample + sp2.2.subsample

  cutoff.95 <- as.integer(permutations*0.95)
  cutoff.99 <- as.integer(permutations*0.99)

```

```

cutoff.05 <- as.integer(permutations*0.05)
cutoff.01 <- as.integer(permutations*0.01)

sp1.max1 <- rep(NA, permutations)
sp1.mean1 <- rep(NA, permutations)
sp2.max1 <- rep(NA, permutations)
sp2.mean1 <- rep(NA, permutations)
inter.max1 <- rep(NA, permutations)
inter.mean1 <- rep(NA, permutations)

#----calculate metrics----
for (i in 1:permutations)
{

#----permute sequences----
sp1.1 <- as.alignment(sp1.1)
seq.perm <- sample(sp1.1$seq,replace=F)
sp1.1$seq <- seq.perm
sp1.1 <- as.DNABin(sp1.1)
sp1.1.sub <- sp1.1[c(1:sp1.1.subsample),]
sp1.2 <- as.alignment(sp1.2)
seq.perm <- sample(sp1.2$seq,replace=F)
sp1.2$seq <- seq.perm
sp1.2 <- as.DNABin(sp1.2)
sp1.2.sub <- sp1.2[c(1:sp1.2.subsample),]
sp2.1 <- as.alignment(sp2.1)
seq.perm <- sample(sp2.1$seq)
sp2.1$seq <- seq.perm
sp2.1 <- as.DNABin(sp2.1)
sp2.1.sub <- sp2.1[c(1:sp2.1.subsample),]
sp2.2 <- as.alignment(sp2.2)
seq.perm <- sample(sp2.2$seq)
sp2.2$seq <- seq.perm
sp2.2 <- as.DNABin(sp2.2)
sp2.2.sub <- sp2.2[c(1:sp2.2.subsample),]

#---is there are better way?----

write.dna(sp1.1.sub,file='mydata.fas',format='fasta')
write.dna(sp1.2.sub,file='mydata.fas',format='fasta',append=T)
sp1.sub <- read.dna(file='mydata.fas',format='fasta')

write.dna(sp2.1.sub,file='mydata.fas',format='fasta')
write.dna(sp2.2.sub,file='mydata.fas',format='fasta',append=T)
sp2.sub <- read.dna(file='mydata.fas',format='fasta')

write.dna(sp1.sub,file='mydata.fas',format='fasta')
write.dna(sp2.sub,file='mydata.fas',format='fasta',append=T)
inter.sub <- read.dna(file='mydata.fas',format='fasta')
sequences <- length(inter.sub)

#----calculate DNA distance-----
x <- dist.dna(sp1.sub,model=dist.model)
sp1.max1[i] <- max(x)
sp1.mean1[i] <- mean(x)

x <- dist.dna(sp2.sub,model=dist.model)
sp2.max1[i] <- max(x)
sp2.mean1[i] <- mean(x)

```

```

y <- rep(NA, sp1.subsample*sp2.subsample)
z <- 0
x <- dist.dna(inter.sub,model=dist.model,as.matrix=T)
for (j in (sp1.subsample+1):sequences)
{
  for (k in 1:sp1.subsample)
  {
    z <- z+1
    y[z] <- x[j,k]
  }
}

inter.max1[i] <- max(y)
inter.mean1[i] <- mean(y)

}

#---generate output---
barcode.distance <- data.frame(sp1.mean1,sp1.max1,sp2.mean1,sp2.max1,inter.mean1,inter.max1)

return(barcode.distance)

}

```

Sumário (*BarcodeDistance2Summary.r*)

```

BarcodeDistance2Summary <-
function(sp1.1,sp1.2,sp2.1,sp2.2,subsample=5,permutations=1000,dist.model="K80",barcode.distance)
{
  library(ape)
  library(seqinr)

  #---initialize variables---

  sp1.1.num <- length(sp1.1)
  sp1.2.num <- length(sp1.2)
  sp2.1.num <- length(sp2.1)
  sp2.2.num <- length(sp2.2)

  subsample.1.1 <- ifelse(sp1.1.num < sp1.2.num, trunc(subsample/2), subsample - trunc(subsample/2))
  subsample.1.2 <- subsample - subsample.1.1
  subsample.2.1 <- ifelse(sp2.1.num < sp2.2.num, trunc(subsample/2), subsample - trunc(subsample/2))
  subsample.2.2 <- subsample - subsample.2.1

  # if(sp1.num < subsample) {sp1.subsample = sp1.num} else {sp1.subsample = subsample}
  # if(sp2.num < subsample) {sp2.subsample = sp2.num} else {sp2.subsample = subsample}

  sp1.1.subsample <- ifelse(sp1.1.num < subsample.1.1, sp1.1.num, subsample.1.1)
  sp1.2.subsample <- ifelse(sp1.2.num < subsample.1.2, sp1.2.num, subsample.1.2)
  sp1.subsample <- sp1.1.subsample + sp1.2.subsample
  sp2.1.subsample <- ifelse(sp2.1.num < subsample.2.1, sp2.1.num, subsample.2.1)
  sp2.2.subsample <- ifelse(sp2.2.num < subsample.2.2, sp2.2.num, subsample.2.2)
  sp2.subsample <- sp2.1.subsample + sp2.2.subsample

  cutoff.95 <- as.integer(permutations*0.95)
  cutoff.99 <- as.integer(permutations*0.99)

```

```

cutoff.05 <- as.integer(permutations*0.05)
cutoff.01 <- as.integer(permutations*0.01)

sp1.max1 <- barcode.distance$sp1.max1
sp1.mean1 <- barcode.distance$sp1.mean1
sp2.max1 <- barcode.distance$sp2.max1
sp2.mean1 <- barcode.distance$sp2.mean1
inter.max1 <- barcode.distance$inter.max1
inter.mean1 <- barcode.distance$inter.mean1

#---summarize distances----

sp1.max <- mean(sp1.max1)
sp1.max.sd <- sd(sp1.max1)
sp1.max.sort <- sort(sp1.max1,decreasing=F)
sp1.max.c95.upper <- sp1.max.sort[cutoff.95]
sp1.max.c95.lower <- sp1.max.sort[cutoff.05]
sp1.max.c99.upper <- sp1.max.sort[cutoff.99]
sp1.max.c99.lower <- sp1.max.sort[cutoff.01]
sp1.mean <- mean(sp1.mean1)
sp1.mean.sd <- sd(sp1.mean1)
sp1.mean.sort <- sort(sp1.mean1,decreasing=F)
sp1.mean.c95.upper <- sp1.mean.sort[cutoff.95]
sp1.mean.c95.lower <- sp1.mean.sort[cutoff.05]
sp1.mean.c99.upper <- sp1.mean.sort[cutoff.99]
sp1.mean.c99.lower <- sp1.mean.sort[cutoff.01]

sp2.max <- mean(sp2.max1)
sp2.max.sd <- sd(sp2.max1)
sp2.max.sort <- sort(sp2.max1,decreasing=F)
sp2.max.c95.upper <- sp2.max.sort[cutoff.95]
sp2.max.c95.lower <- sp2.max.sort[cutoff.05]
sp2.max.c99.upper <- sp2.max.sort[cutoff.99]
sp2.max.c99.lower <- sp2.max.sort[cutoff.01]
sp2.mean <- mean(sp2.mean1)
sp2.mean.sd <- sd(sp2.mean1)
sp2.mean.sort <- sort(sp2.mean1,decreasing=F)
sp2.mean.c95.upper <- sp2.mean.sort[cutoff.95]
sp2.mean.c95.lower <- sp2.mean.sort[cutoff.05]
sp2.mean.c99.upper <- sp2.mean.sort[cutoff.99]
sp2.mean.c99.lower <- sp2.mean.sort[cutoff.01]

inter.max <- mean(inter.max1)
inter.max.sd <- sd(inter.max1)
inter.max.sort <- sort(inter.max1,decreasing=F)
inter.max.c95.upper <- inter.max.sort[cutoff.95]
inter.max.c95.lower <- inter.max.sort[cutoff.05]
inter.max.c99.upper <- inter.max.sort[cutoff.99]
inter.max.c99.lower <- inter.max.sort[cutoff.01]
inter.mean <- mean(inter.mean1)
inter.mean.sd <- sd(inter.mean1)
inter.mean.sort <- sort(inter.mean1,decreasing=F)
inter.mean.c95.upper <- inter.mean.sort[cutoff.95]
inter.mean.c95.lower <- inter.mean.sort[cutoff.05]
inter.mean.c99.upper <- inter.mean.sort[cutoff.99]
inter.mean.c99.lower <- inter.mean.sort[cutoff.01]

#---generate output----

metrics <-

```

```

c('num.indiv','num.sample','max','max.sd','max.c95.upper','max.c95.lower','max.c99.upper','max.c99.lower','mean','
mean.sd','mean.c95.upper','mean.c95.lower','mean.c99.upper','mean.c99.lower')

sp1.barcode <-
c(sp1.1.num+sp1.2.num,sp1.subsample,sp1.max,sp1.max.sd,sp1.max.c95.upper,sp1.max.c95.lower,sp1.max.c99.
upper,sp1.max.c99.lower,sp1.mean,sp1.mean.sd,sp1.mean.c95.upper,sp1.mean.c95.lower,sp1.mean.c99.upper,sp
1.mean.c99.lower)

sp2.barcode <-
c(sp2.1.num+sp2.2.num,sp2.subsample,sp2.max,sp2.max.sd,sp2.max.c95.upper,sp2.max.c95.lower,sp2.max.c99
.upper,sp2.max.c99.lower,sp2.mean,sp2.mean.sd,sp2.mean.c95.upper,sp2.mean.c95.lower,sp2.mean.c99.upper,
sp2.mean.c99.lower)

inter.barcode <-
c(NA,NA,inter.max,inter.max.sd,inter.max.c95.upper,inter.max.c95.lower,inter.max.c99.upper,inter.max.c99.lower,in
ter.mean,inter.mean.sd,inter.mean.c95.upper,inter.mean.c95.lower,inter.mean.c99.upper,inter.mean.c99.lower)

barcode.distance.summary <- data.frame(metrics,sp1.barcode,sp2.barcode,inter.barcode)

return(barcode.distance.summary)

}

```

Reamostragens (*Surubim_resample2.r*)

```

rm(list=ls())
setwd('C:/Documents and Settings/Mauro/Desktop/Pseudoplatystoma')
source('BarcodeDistance2.r')
source('BarcodeDistance2Summary.r')

library(ape)
library(seqinr)
library(locfit)

#---initialize variables---

sp1.name <- 'P_fasciatum'
sp2.name <- 'P_tigrinum'
subsample <- 10
permutations <- 1000
dist.model <- 'K80' # this is K2P

#---read data---

sp1.1 <- read.dna(file=paste(sp1.name,'1_COI.fas',sep='_',collapse=""),format='fasta')
sp1.2 <- read.dna(file=paste(sp1.name,'2_COI.fas',sep='_',collapse=""),format='fasta')
sp2.1 <- read.dna(file=paste(sp2.name,'1_COI.fas',sep='_',collapse=""),format='fasta')
sp2.2 <- read.dna(file=paste(sp2.name,'2_COI.fas',sep='_',collapse=""),format='fasta')

#---calculate distances---

barcode.distance <- BarcodeDistance2(sp1.1,sp1.2,sp2.1,sp2.2,subsample,permutations,dist.model)

barcode.distance.summary <-
BarcodeDistance2Summary(sp1.1,sp1.2,sp2.1,sp2.2,subsample,permutations,dist.model,barcode.distance)

#---save results---

save(barcode.distance,file=paste('Barcode2',sp1.name,sp2.name,permutations,dist.model,'permutations.Rdata',s

```



```

ep='_ ',collapse="))

write.csv(barcode.distance,file=paste('Barcode2',sp1.name,sp2.name,permutations,dist.model,'permutations.csv',
sep='_ ',collapse="))

write.table(barcode.distance,file=paste('Barcode2',sp1.name,sp2.name,permutations,dist.model,'permutations.txt',
sep='_ ',collapse="))

save(barcode.distance.summary,file=paste('Barcode2',sp1.name,sp2.name,permutations,dist.model,'results.sum
mary.Rdata',sep='_ ',collapse="))

write.csv(barcode.distance.summary,file=paste('Barcode2',sp1.name,sp2.name,permutations,dist.model,'results.s
ummary.csv',sep='_ ',collapse="))

write.table(barcode.distance.summary,file=paste('Barcode2',sp1.name,sp2.name,permutations,dist.model,'results.
summary.txt',sep='_ ',collapse="))

#---plot permutation data----

#sp1.den <- density(barcode.distance$sp1.mean1)
#sp2.den <- density(barcode.distance$sp2.mean1)
#inter.den <- density(barcode.distance$inter.mean1)

sp1.den <- locfit(~barcode.distance$sp1.mean1)
sp2.den <- locfit(~barcode.distance$sp2.mean1)
inter.den <- locfit(~barcode.distance$inter.mean1)

plot(sp1.den,type="l",ylim=range(0,5000),xlim=range(0,0.045),main="",xlab="",ylab="",lty=1,lwd=2,col="black")
lines(sp1.den,xlab="",ylab="",lty=1,lwd=2,col="black")
lines(sp2.den,xlab="",ylab="",lty=1,lwd=2,col="blue")
lines(inter.den,xlab="",ylab="",lty=1,lwd=2,col="red")
title(main="Barcoding",ylab="Frequency",xlab="Genetic distance")
legend(0.028,2000,c(sp1.name,sp2.name,"interspecific"),lty=c(1,1,1),lwd=c(2,2,2),col=c("black","blue","red"))

dev.copy2pdf(file=paste('Barcode2',sp1.name,sp2.name,permutations,dist.model,'graph.pdf',sep='_ ',collapse="))
dev.copy2eps(file=paste('Barcode2',sp1.name,sp2.name,permutations,dist.model,'graph.eps',sep='_ ',collapse="))

```