

UEA

UNIVERSIDADE
DO ESTADO DO
AMAZONAS

**UNIVERSIDADE DO ESTADO DO AMAZONAS
ESCOLA DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA E RECURSOS
NATURAIS DA AMAZÔNIA**

ELISAMA FRANCO BEZERRA

**CÓDIGODE BARRA DE DNA COMO FERRAMENTA NA TAXONOMIA DE AVES
AMAZÔNICAS**

**MANAUS
2013**

ELISAMA FRANCO BEZERRA

**CÓDIGODE BARRA DE DNA COMO FERRAMENTA NA TAXONOMIA DE AVES
AMAZÔNICAS**

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia e Recursos Naturais da Amazônia da Universidade do Estado do Amazonas (UEA), como parte dos requisitos para obtenção do título de mestre em Biotecnologia e Recursos Naturais.

Orientador: Prof Dr Cleiton Fantin
Co-Orientadora: Prof^a Dr^a Camila C. Ribas

MANAUS
2013

ELISAMA FRANCO BEZERRA

**CÓDIGO DE BARRA DE DNA COMO FERRAMENTA NA TAXONOMIA DE AVES
AMAZÔNICAS**

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia e Recursos Naturais da Amazônia da Universidade do Estado do Amazonas (UEA), como parte dos requisitos para obtenção do título de mestre em Biotecnologia e Recursos Naturais.

Data da aprovação ___/___/___

Banca Examinadora:

**MANAUS
2013**

Ficha Catalográfica
(Catalogação na fonte realizada pela Biblioteca Central – UEA)

Bezerra, Elisama F.

Bxxxxd Códigode barra de DNA Como Ferramenta Na
Taxonomia De Aves Amazônicas / Elisama F. Bezerra. -
Manaus: UEA, 2013.

98 f.; il. color.

Dissertação (Mestrado em Biotecnologia) —
Universidade do Estado do Amazonas (UEA), 2013.

Orientador: Prof Dr Cleiton Fantin

Co-Orientadora: Profª Drª Camila C. Ribas

1. *Barcode* – Aves – Estudo taxonômicos 2. Aves –
Amazônia 3. Coleção de tecidos e zoológicos – Genética de
populações

CDU xxxxxxxx

Dedicatória.
A minha família pelo amor incondicional

*“Ora, a fé é o firme fundamento das coisas que se esperam,
e a prova das coisas que se não vêem”.*
Aos Hebreus; Cap. 11 verso 1

AGRADECIMENTOS

A Deus, minha rocha e fortaleza.

À CNPq pelo auxílio financeiro.

Ao Programa de Pós-Graduação em Biotecnologia e Recursos Naturais da Amazônia MBT/UEA.

Ao Laboratório Temático de Biologia Molecular LTBM/INPA, pelo apoio logístico.

Aos meus orientadores Dr. Cleiton Fantin e Dra. Camila Ribas pelos ensinamentos nestes dois anos.

Ao Mario Cohn-Haft, sempre presente desde a graduação, que me ensinou o amor pelas aves.

Aos meus pais (Alírio e Sônia) pelo apoio e incentivos aos estudos, muitas vezes sem entender porque passo tanto tempo no Laboratório procurando o que ninguém consegue ver.

Aos meus irmãos Filipe, Gisele e Kássia pelas brincadeiras e as brigas sem elas não somos irmãos. Infinitamente amo minha família.

Aos amigos que conheci no MBT/UEA que fizeram não só das aulas, mas todos os momentos que compartilhamos juntos únicos e inesquecíveis.

Aos amigos do LTBM/INPA por me ensinarem além dos procedimentos técnicos a importância da amizade. As reuniões do café filosófico, ou nissin miojo compartilhado onde neste momento sabíamos que estávamos em família. Em especial a Adriel Lira, Jaqueline Fortuna, Paola Castro e Saulo por suas orientações científicas e conselhos.

RESUMO

O Código de barra de DNA tem sido utilizado com sucesso na identificação da avifauna global com mais de 90% de acerto. A identificação correta das espécies é importante para a conservação e quantificação da biodiversidade. Com isso este trabalho tem como objetivo gerar um banco de dados de sequências do código de barras de DNA relacionadas à *vouchers* depositados na Coleção de Aves do Instituto Nacional de Pesquisa da Amazônia (INPA) para confirmar a confiabilidade do BOLD (The Barcode of Life Data System) na identificação das espécies de aves amazônicas com a taxonomia vigente. Foi selecionada uma amostra por espécie da Coleção de Tecidos do Instituto Nacional de Pesquisa da Amazônia (INPA), totalizando 422 amostras de aves. O gene utilizado como código de barra de DNA é o gene mitocondrial Citocromo C Oxidase 1 (COI). Na validação do gene COI do código de barra de DNA como ferramenta taxonômica para identificação de espécies de aves da Amazônia, obtivemos 72,9% de espécies identificadas corretamente. Dentre as 407 espécies com DNA sequenciado, 90 são sequências inéditas para o BOLD. Em 20 amostras foram identificados erros taxonomicos e somente duas amostras o BOLD identificou errado. O Código de barra de DNA interage entre a Taxonomia Clássica e Molecular para uma identificação precisa das espécies. Os resultados do presente trabalho demonstraram que como ferramenta taxonômica, o DNA *barcode* pode ser usado para a identificação molecular de aves da Amazônia. Apesar da falta de sequências de DNA de aves amazônicas no banco de dados do BOLD, as espécies estudadas neste trabalho apresentaram de 93,5% a 100% de confiabilidade, demonstrando o poder resolutivo desta ferramenta.

ABSTRACT

The DNA barcode has been successfully used in identifying the global avifauna with over 90% accuracy. Correct identification of species is important for the conservation of biodiversity and quantification. Thus this work aims to generate a database of sequences DNA barcode related vouchers deposited in the Collection of Birds from the National Institute for Amazonian Research (INPA) to confirm the reliability of BOLD (The Barcode of Life Data System) in identifying the species of Amazonian birds with the current taxonomy. A sample was selected by species of Tissue Collection of the National Institute for Amazonian Research (INPA), totaling 422 samples from birds. The gene used as the DNA barcode is a mitochondrial cytochrome c oxidase gene 1 (COI). Validation of the gene COI barcode DNA as a tool for taxonomic identification of bird species in the Amazon, we had 72.9% of species identified correctly. Among the 407 species with DNA sequenced, 90 are unpublished sequences for BOLD. In 20 samples were indentificados taxonomic errors and only two samples identified the BOLD wrong. The DNA barcode interacts between the Classical and Molecular taxonomy for accurate identification of species. The results of this study demonstrate that taxonomic tool as the DNA barcode can be used for molecular identification of birds in the Amazon. Despite the lack of DNA sequences of Amazonian birds in the BOLD database, the species studied in this work showed 93.5% to 100% reliability, demonstrating the resolving power of this tool.

LISTA DE TABELAS

CAPÍTULO I

- Tabela 1. Tabela da representatividade de Ordens, Famílias, Gêneros e Espécies de aves amazônicas utilizadas para o presente trabalho.....32
- Tabela 2. Lista de espécies de aves amostradas na Amazônia que apresentaram erro de identificação taxonômica e a identificação obtida de acordo com a porcentagem de similaridade das sequências de COI no BOLD (*The Barcode of Life Data System*).....35
- Tabela 3. Lista de espécies de aves amostradas na Amazônia que apresentaram trocas de identificação nas etiquetas do tubo de tecido e identificação obtida de acordo com a porcentagem de similaridade das sequências de COI no BOLD (*The Barcode of Life Data System*) (erro tipo II).....36

LISTA DE FIGURAS

INTRODUÇÃO

Figura 1 – Modelo do *barcoding* gap. (A) Distribuições intraespecíficas e interespecíficas não se sobrepõem. (B) Sobreposição dos níveis intraespecífica e interespecífica (Fonte: MEYER E PAULAY 2005).....18

CAPÍTULO I

Figura 1. Mapa de distribuição dos pontos de coletas das amostras de tecido de aves da coleção de aves do INPA.....33

Figura 2. Composição média das bases nitrogenadas da região do Código de barra de DNA de aves coletadas na Amazônia.....35

Figura 3. Árvore de *Neighbour-joining* da Família Tyrannidae com valores de bootstrap, no ramo em destaque as espécies com trocas de identificação em campo que ficaram fora do clado da família.....

Figura 4. Árvore de *Neighbor-joining* da divergência nucleotídica das sequências de COI por família de aves amostrada na Amazônia com valores de *bootstrap* e régua de distância genética. A espessura dos ramos representa a quantidade de táxons por Família.....38

Figura 5. Árvore de *Neighbour-joining* com o valor de *bootstrap* da Família Rhynchocyclidae, em destaque as espécies de aves com identificação diferente do BOLD (*The Barcode of Life Data System*).....39

Figura 6. Frequência da distância genética intragenérica e intergenérica dos 78 generos de aves amostradas em 20 localidades na Amazônia brasileira.....41

Figura 7. Fluxograma de ligação entre coleções zoológicas e coleções de tecido através do método barcode.45

SUMÁRIO

Bxxxxd	4
CDU xxxxxxxx.....	4
LISTA DE TABELAS	10
LISTA DE FIGURAS	11
1. INTRODUÇÃO	12
2. REFERENCIAL TEÓRICO.....	13
2.1 Taxonomia de aves	13
2.2 Taxonomia Moderna	14
2.3 Código de Barra de DNA	16
2.4 Código de barra de DNA e Aves.....	20
3. OBJETIVOS	23
3.1 Geral	23
3.2 Específicos	23
3. REFERÊNCIAS BIBLIOGRÁFICAS	24
CAPÍTULO I	29
DNA Barcode: Eficácia do banco de dados do BOLD na identificação molecular de espécies de aves coletadas Amazônica	29
Resumo.....	30
1. INTRODUÇÃO	30
2. MATERIAL E MÉTODOS.....	31
2.1 Amostragem.....	31
.....	33
Figura 1. Mapa de distribuição dos pontos de coletas das amostras de tecido de aves da coleção de aves do INPA.....	33
2.2. Extração e sequenciamento de DNA.....	33
2.3. Análises das sequências de DNA	34
3. RESULTADOS	34
4. DISCUSSÃO	42
5. CONCLUSÃO	46
6. REFERENCIAS BIBLIOGRAFICAS	47
ANEXO	51

1. INTRODUÇÃO

Desde a antiguidade a humanidade se preocupa em diferenciar as espécies, sendo o pioneiro Aristóteles (384-322 a.C.), que separou as espécies de acordo com sua complexidade, distribuídas em dois grupos maiores, animais com sangue e os sem sangue, posteriormente em grupos mais amplos como: I) Vivíparos- são animais quentes e úmidos que produzem descendentes iguais aos pais (Ex: Homem, Cavalo e Golfinho); II) Ovíparos- são animais quentes e secos que produzem ovos perfeitos (Ex: Aves, Répteis); III) Externamente Vivíparos- animais frios e úmidos (cobras e peixes cartilagosos); IV) Externamente Ovíparos – animais frios e secos com ovo imperfeito (Crustáceos e cefalópodes); V) Insetos- Animais imperfeitos e frios que produzem larvas em vez de ovos e VI) Testáceo – intermediários entre plantas e animais, que não produzem vida e se reproduzem espontaneamente. No caso das plantas, foram divididas em Árvore, Arbusto, Erva e Hortaliça (ABREU, 1994). Portanto, pode-se dizer que teve assim início a taxonomia.

O termo taxonomia se origina do grego, *taxis* (ordem), e *nomos* (lei, norma) e foi usado pela primeira vez em 1735, com a publicação da versão inicial da obra *Systema Naturae*, pelo cientista e médico sueco Karl Von Linné, e assim tornando-se conhecida no domínio da biologia (AGANETTE et al., 2010). Com mais de 250 anos a classificação taxonômica desenvolvida por Karl Linné no Século XVIII permanece até hoje, e é à base da identificação de espécies, classificando os seres vivos em uma hierarquia, a qual se inicia com os Reinos que são divididos em Filos, e então em Classes, Ordens, Famílias, Gêneros e Espécies (GODFRAY, 2002). Desde sua criação, o sistema de Linné tem sofrido diversas modificações, estando atualmente composto por um conjunto de princípios e regras complexas, organizados por uma comissão e publicados na forma de códigos (RAPINI, 2004).

Estima-se que o número de espécies de eucariotos no planeta seja de aproximadamente 8,7 milhões, sendo que aproximadamente 2,2 milhões são de espécies marinhas e 1,2 milhões destas espécies já foram catalogadas (ou descritas). Portanto, cerca de 86% das espécies na Terra e 91% das espécies do oceano ainda aguardam descrição taxonômica (MORA et al., 2011). Segundo este mesmo autor poderia levar até 1.200 anos e exigiria 303 mil taxonomistas, a um custo aproximado de 364 milhões de dólares para descrever todas as espécies do planeta. Assim sendo, nota-se um avanço lento nas descrições taxonômicas, enquanto as extinções têm acontecido rapidamente, como consequências disso,

várias espécies nunca chegaram a ser conhecidas para a ciência. O trabalho de levantamento taxonômico exige muito conhecimento na distribuição, morfologia e ecologia de cada táxon, tempo para coletas de espécimes, comparações de indivíduos depositados em museus e levantamento bibliográfico e financiamento para todos estes itens, contudo tem-se observado um avanço das técnicas moleculares, que estão fornecendo meios para auxiliar e acelerar o processo de identificação das espécies, revelando ser bastante eficaz na identificação de espécies de diversos grupos (AMORIM 1997; HEBERT ET al., 2003a; MORITZ e CICERO 2004; RAPINI 2004; DAYRAT 2005).

2. REFERENCIAL TEÓRICO

2.1 Taxonomia de aves

As classificações de aves anteriores aos anos 90 eram feitas somente com análises morfológicas. O trabalho de Sibley et al. (1988) foi o primeiro a sugerir uma nova classificação para aves baseada na hibridização DNA-DNA. Também conhecida como Taxonomia de *Sibley-Ahquist*, confirmou dados morfológicos tradicionais e trouxe mudanças radicais, algumas delas abandonadas depois de novas técnicas e descobertas com DNA (ERICSON et., 2006; HACKETT et al., 2008).

A principal dificuldade da contagem de espécies de aves do mundo é decidir, se consideram espécies ou subespécies. Cerca de 75% das espécies de aves são politípicas, isto é, são compostas por mais de uma subespécie; conseqüentemente, o número de subespécies é mais elevado que o número de espécies (MAYR, 1946). A utilização indiscriminada e sem critérios da categoria subespecífica trouxe consigo diversos problemas na delimitação dos táxons levando a várias interpretações na literatura dos limites taxonômicos de espécies (ALEIXO, 2007).

O Brasil tem a segunda mais rica avifauna do mundo, em 2005 estimava-se que o número de espécies estava entre 1.696 e 1.731 (MARINI e GARCIA, 2005). De acordo com o Comitê Brasileiro de Registro Ornitológico, o Brasil possui 1.895 espécies de aves, onde 1.217 destas são espécies da Amazônia brasileira (CBRO, 2011). De 1997 a 2010 foram descobertas 16 novas espécies de aves no Brasil (CBRO, 2011), sendo que análises filogenéticas foram utilizadas em só uma espécie, mas somente para relacionar parentesco.

Ainda há discussão entre os ornitólogos sobre a utilização de ferramentas moleculares na identificação de novas espécies, como é o caso de *Schiffornis turdina* (NYÁRI, 2007) e *Willisornis poecilinotus* (ISLER E WHITNEY, 2011) onde foram feitas análises moleculares e morfológicas. Em *S. turdina* a análise molecular indicou sete espécies, apesar das semelhanças morfológicas e vocais, por isso não foi aceito. Já em *W. poecilinotus* as análises moleculares confirmaram as diferenças vocais. Donegan et al. (2011) fizeram uma nova proposta para dividir *Schiffornis turdina*, utilizando os dados morfológicos, vocais e os dados genéticos de Nyári (2007) sendo aceitas quatro novas espécies das sete proposta por Nyári (2007). O nível de confiança em espécies suportadas pelo uso múltiplo de dados é muito maior do que para as espécies apoiadas por apenas um tipo, portanto é fundamental que as revisões taxonômicas utilizem o maior número de caracteres possível (ALEIXO, 2007).

2.2 Taxonomia Moderna

A discussão de atualizar a taxonomia não é uma idéia nova. Conforme a modernização em outras áreas biológicas foi sendo questionado se a utilização de somente um método, no caso análise de caracteres morfológicos, para descrever as espécies era suficiente. Como no caso dos microrganismos, somente as diferenças morfológicas não poderiam ser incluídas nos Reinos existente na época (animalia e plantae) passando a ser divididos em três reinos, em seguida as análises moleculares (RNA16S) revelaram novas relações evolutivas, permitindo a divisão em cinco reinos (WOESE et al., 1990).

Desde a década de 1970 a análise do DNA mitocondrial (DNAm) tem se estabelecido como uma poderosa ferramenta para estudos evolutivos em animais. O genoma mitocondrial dos animais é constituído por uma molécula de DNA circular pequena, com conteúdo gênico conservado (apenas 37 genes), taxa de substituição rápida, e não possui DNA repetitivo, *transposons*, *intros* ou pseudogenes (MORITZ et al., 1987). A combinação dessas características faz com que o DNAm seja amplamente utilizado em estudos de caracterização de populações, subespécies e espécies, além de estudos de caráter evolutivo e filogenético (MORITZ et al., 1987; HARRISON, 1989; AMORIM, 1997; OKUMUS e ÇİFTCI 2003).

A amplificação do DNA e sequenciamento automático durante a década de 80 levou ao desenvolvimento de várias classes de marcadores de DNA (OKUMUS e ÇİFTCI 2003). O uso de marcadores moleculares possibilita a avaliação da variabilidade genética existente dentro e entre espécies distintas (BERED et al., 1997). Estas análises podem ser direcionadas

para inferências filogenéticas, determinação de estruturas populacionais, identificação de linhagens, determinação da variação genética e na caracterização e identificação de espécies, ou seja, na taxonomia (OKUMUS e ÇİFTCI 2003; LIU e CORDES, 2004). Dentre os vários marcadores moleculares o DNA mitocondrial é o mais utilizado, por ter sequências de DNA altamente conservadas dentro da espécie e variada entre espécies e por já existirem vários *primers* universais (OKUMUS e ÇİFTCI 2003). Estes marcadores moleculares têm ajudado a decifrar árvores filogenéticas, como no caso de aves e peixes, onde foram revelados novos rearranjos taxonômicos (HACKETT et al., 2008; JAVONILLO et al., 2009).

Tautz et al. (2003) propõem que o DNA assumira um papel central na taxonomia, como principal característica para diferenciar as espécies. Logo em seguida foi desenvolvido *Molecular Operational Taxonomic Units* (MOTU) para agrupar táxons morfologicamente crípticos usando similaridade genética, sendo definido geneticamente em uma unidade taxonômica (BLAXTER et al., 2004). O MOTU Agrupa sequências que diferem por um número máximo de bases em uma região de 500 pares de bases do RNAr 18S (BLAXTER et al., 2004; VOGLER e MONAGHAN, 2007).

Godfray (2002) sugere uma nova taxonomia, mas sem descartar os conhecimentos adquiridos nesses 259 anos de Taxonomia Clássica. Denominada como Taxonomia Unitária, ela junta todos os dados descritos da espécie como: as primeiras descrições da espécie, genoma, proteínas conhecidas, aspectos biológicos, ecológicos e fotos. Todos estes dados estariam disponíveis na internet para serem discutidos entre os taxonomistas, a disposição dos pesquisadores e ao público leigo. Bisby et al. (2002) cita alguns sites que começaram a organizar dados taxonômicos: Species 2000 (<http://www.sp2000.org>), o Sistema Integrado de Informação Taxonômica (<http://www.itis.usda.gov>), Iniciativa Global de Taxonomia (GTI) da Convenção sobre Diversidade Biológica (<http://www.biodiv.org/decisions> em COP6/V8) e o Global Biodiversity Information Facility (GBIF; <http://www.gbif.org>).

Em 2006 foi criado o consórcio EDIT (European Distributed Institute of Taxonomy), que tem como objetivos: 1) unir todas as informações taxonômicas das espécies em um único banco de dados e 2) Gerar profissionais na área de taxonomia. O EDIT segue a base da taxonomia unitária onde os dados das espécies estão disponíveis em rede, mas no momento somente as instituições Europeias estão no consórcio, o que limita o número de espécies.

Dayrat (2005) ressalta que um sistema de identificação baseado em DNA só funciona se todas as espécies já descritas tenham sua sequência no banco de dados, caso isso não

ocorra tem-se um banco de dados incompleto, dando ao usuário a ilusão de estar descrevendo uma nova espécie, contudo trata-se apenas de uma sequência de espécie não cadastrada.

Assim, a taxonomia moderna enfrenta dois grandes desafios. Primeiro, o desafio qualitativo que implica em chegar a um consenso científico sobre a categoria base em torno do qual a taxonomia é construída - as espécies e, assim, melhorar a delimitação das espécies. O segundo, um desafio quantitativo, o grande número de espécies na terra que exigem descoberta e descrição (PADIAL et al., 2010). Com isso uma abordagem integrativa à taxonomia é necessária, visto que a biologia de espécies é complexa e assim devemos estudar espécies a partir de perspectivas múltiplas e complementares (DAYRAT, 2005; PADIAL et al., 2010). A taxonomia do DNA (*DNA taxonomy*) utiliza o DNA como referência taxonômica, ou seja, mais uma característica de separação na análise de identificação das espécies (TAUTZ et al., 2003). Além disso, o nosso nível de confiança na identificação de espécies suportadas por diferentes tipos de dados é muito maior do que para as espécies apoiadas por apenas um tipo (DAYRAT, 2005; PADIAL et al., 2010).

Entretanto, com a recente inclusão da técnica molecular para identificação de espécies, algumas questões permanecem em discussão como, por exemplo: como fazer amostragem completa de um grupo de espécies? Em que escala geográfica devem ser feitas as análises? Qual gene e números de *locos* a serem usados? Qual o limite e conceito de espécie a ser usado? Que níveis de incongruência entre morfológico e genético são válidos? (VOGLER e MONAGHAN 2007).

2.3 Código de Barra de DNA

Em maio de 2004 foi estabelecido o consórcio CBOL (Consortium for the *Barcode* of Life) com apoio financeiro da Fundação Alfred P. Sloan. Esse consórcio envolve a colaboração internacional de museus, herbários, coleções biológicas e especialistas em genômica, taxonomia, eletrônica e computação. Inclui mais de 120 organizações de 45 nações (RATNASINGHAM e HEBERT 2007). O objetivo do CBOL é acelerar a compilação dos códigos de barras de DNA das espécies de animais e plantas, estabelecendo assim um banco público de sequências correlacionadas a espécimes (*vouchers*) e promover o desenvolvimento de um equipamento portátil para identificação por meio do código de barras da vida.

Inicialmente, foi feito um banco de sequências do Código de barra de DNA para várias espécies existentes, denominado de BOLD (*The Barcode of Life Data System*), que é uma

plataforma da bioinformática que permite associar outros tipos de dados as amostras. Dentre esses dados, destacam-se: I) fotos de espécimes; II) Informações de campo (ponto de coleta, data da coleta, coletor, etc); III) número de espécimes e instituição no qual esse material foi depositado; IV) dados taxonômicos e V) informações moleculares que foram utilizada no depósito da sequência com eletroferogramas e *primers*. Para preencher esse banco de dados, têm sido utilizadas preferencialmente amostras que foram identificadas por taxonomistas de museus ou de outras instituições (RATNASINGHAM e HEBERT 2007). Vale ressaltar que esses dados podem ser visualizados no site www.barcodeoflife.org tanto por cientistas quanto para o público leigo. Hoje, o BOLD tem mais de 5.000 usuários registrados e armazena os registros de código de barras de mais de 1 milhão espécimes, representando aproximadamente 171.817 espécies. Há 5.327 espécies de aves com sequencias de DNA, das quais 4.548 possuem código de barras e somente 446 são do Brasil (BODSYSTEMS, 2013).

Segundo Herbet et al. (2003a) outros benefícios esperados com a utilização de Código de barra de DNA são: a identificação de espécies crípticas, o descobrimento de novas espécies, a identificação de formas juvenis e adultas de uma mesma espécie e a identificação de espécies a partir de fragmentos de material biológico provenientes de apreensão ilegal ou de acidentes de avião. A proposta é de que qualquer pessoa, em qualquer lugar, a qualquer momento, seja apta a identificar rapidamente e com eficiência a espécie de um determinado espécime ou um fragmento de material biológico independentemente da condição de preservação (STOECKLE et al., 2005).

Herbert et al. (2003) avaliam o potencial do Citocromo C Oxidase 1 (COI) como uma ferramenta para a taxonomia de espécies em diversos filos animais (Annelida, Chordata, Echinodermata, Nematoda, Platelminto, Artrópoda, Mollusca), apontando 96% de certeza na identificação das espécies estudadas. O fragmento do gene mitocondrial Citocromo Oxidase I (COI ou COX1) é constituído de 1000 pares de bases (pb), onde na media 650pb são utilizados para identificação taxonômica.

Hebert et al. (2004b), estudando o uso de Código de barra de DNA na identificação de aves norte americanas, propõem que a diferença genética interespecífica é 10 vezes maior do que intraespecífica, de tal maneira que formaria claramente um *gap*, permitindo assinalar um espécime desconhecido à sua espécie com uma taxa de erro insignificante, sendo este o principal pressuposto para a eficiência do Código de barra de DNA (HEBERT et al., 2004b; RATNASINGHAM e HEBERT, 2007).

A escolha do citocromo oxidase I como código de barras deve-se ao fato de sua taxa de evolução molecular permitir distinguir espécies próximas e também grupos filogenéticos dentro da mesma espécie (HEBERT, ET al., 2003a); e os iniciadores universais para este gene estão bem estabelecidos, permitindo a amplificação do mesmo em quase todos os filos animais (FOLMER ET al., 1994). O COI possui variabilidade semelhante à encontrada em outros genes codificadores de proteína, os genes codificadores de proteína mitocondriais contêm maior variabilidade do que os genes ribossomais e, portanto, seriam melhores para distinguir espécies próximas. Além disso, a comparação de sequências de genes codificadores de proteínas é mais fácil, pois elas geralmente não apresentam inserções ou deleções, que são comuns em genes ribossomais. Dentre as razões para utilizar o COI como código de barras estão : I) fácil de ser sequenciado para vários táxons usando um número pequeno de *primers*; II) suas sequências são facilmente alinhadas para comparações e III) é informativo para distinguir espécies geneticamente próximas de animais (vários invertebrados e vertebrados) (STOECKLE et al., 2005). Além disso, o gene COI possui 15 sítios variáveis, o que permite 1 bilhão de diferentes combinações de bases, o suficiente para gerar padrões de código de barra ao nível de espécies (HEBERT et al., 2003).

O *Barcode* tem sido utilizado com sucesso na identificação de vertebrados (BARBER e BOYCE, 2006; ORTIZ, 2010; HENRIQUES, 2010), invertebrados (KAMARUZZANAN et al., 2011; VERSHININA e LUKHTANOV, 2010), répteis (NARO-MACIEL ET AL, 2010) e fungos (ROBIDEAU et al., 2011), com resolução acima de 90% de identificação. Também pode ser útil como fonte de informação de caráter, ou seja, como componente da descrição de espécies e como prova potencial para uma maior atribuição taxonômica ou re-atribuição, como por exemplo; o nível de gênero ou família (DeSALLE et al., 2005; GOLDSTEIN e DeSALLE, 2010), conservação, manejo de pragas, forense e saúde (RUBINOFF, 2006b; CARVALHO et al., 2008; KOSMANN, 2009; DINCA et al 2011). A plataforma de dados do *barcode* (BOLD) já é utilizada para ajudar a descrever a diversidade de espécies onde somente a morfologia é insuficiente (CHEN et al 2011; NARO-MACIEL et al, 2010).

O *Barcode* é um grande avanço, não apenas pela identificação, mas a utilização completa de três inovações taxonômicas: 1) A molecularização - o uso de um marcador molecular como um discriminador; 2) informatização - utilização de suportes informáticos e 3) padronização - o mesmo gene e metodologia para todo o organismo (CASIRAGHI et al., 2010a). Sendo que hoje para plantas e fungos são utilizados outros genes, mas com a mesma metodologia.

Por outro lado, há controvérsias em utilizar apenas um gene para a identificação de todos os organismos, sendo o principal argumento que um único fragmento possui uma história única, sujeita a retenção de polimorfismos ancestrais (MORITZ e CICERO 2004; MEYER e PAULAY 2005). Este fato faz com que espécies que divergiram recentemente em alguns casos ainda compartilhem os mesmos haplótipos, o que torna difícil sua identificação (BLAXTER 2004; MORITZ e CICERO 2004; HEBERT et al., 2009; LUKHTANOV et al., 2009).

Espécies estreitamente relacionadas como espécies-irmãs, híbridas, espécies que divergiram recentemente e crípticas apresentam baixa divergência interespecífica podendo ocorrer sobreposição das divergências (MORITZ e CICERO 2004; MEYER e PAULAY 2005; STOECKLE et al., 2005). Meyer e Paulay (2005) confirmam que a sobreposição pode gerar falsos positivos e negativos, indo contra o principal pressuposto para a efetividade do Código de barras de DNA, de que as divergências intraespecíficas sempre sejam menores que as interespecíficas (Fig.1) (HEBERT et al., 2004a; HEBERT et al., 2004b). Somente o surgimento do *gap* não é garantia de uma nova espécie, podendo ser erroneamente interpretado os tempos de divergência entre populações recentemente isoladas com linhagens reprodutivamente isoladas (HICKERSON et al., 2006).

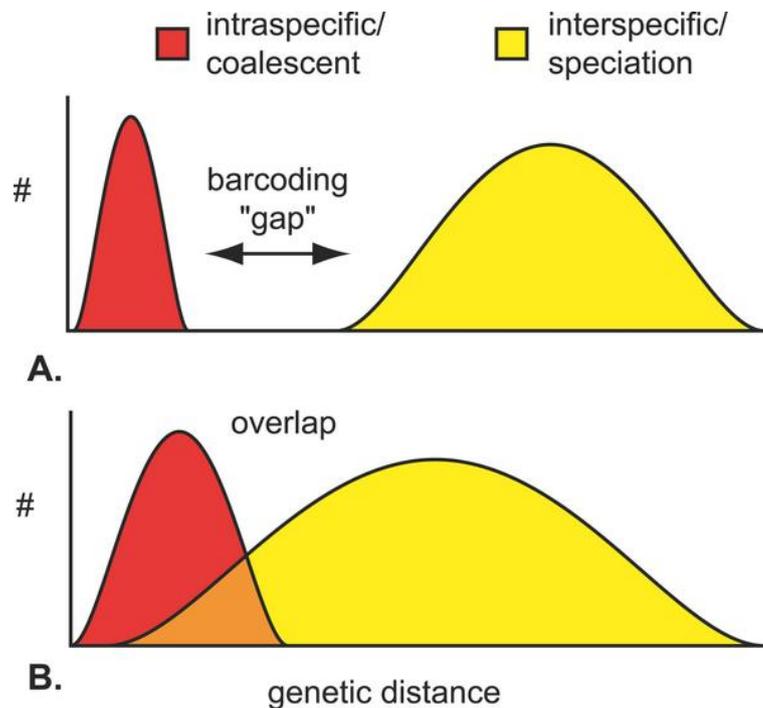


Figura 1. Modelo do *barcoding* gap. (A) Distribuições intraespecíficas e interespecíficas não se sobrepõem. (B) Sobreposição dos níveis intraespecífica e interespecífica. (Fonte: MEYER E PAULAY 2005)

Moritz e Cicero (2004) e Stoeckle et al. (2005) destacam que os maiores obstáculos encontrados na completa aplicação do Códigode barra de DNA são: a) grupos que possuem baixa diversidade de sequências; b) espécies que divergiram recentemente; c) detecção de espécies híbridas e os d) pseudogenes nucleares *numts* (*Nuclear mitochondrial-like sequences*).

As cópias de pseudogenes podem ser amplificadas por PCR junto com o DNAm_t, este fato prejudica a separação do DNA mitocondrial de *numts*, podendo levar a uma inferência filogenética equivocada (THALMANN, et al., 2004; RUBINOFF e HOLLAND 2005). Como as taxas de evolução são mais lentas no DNA nuclear do que no DNA mitocondrial, se as sequências de *Numts* e de DNA mitocondrial forem utilizadas simultaneamente em análises filogenéticas, relações errôneas podem ser propostas e espécies crípticas subestimadas (MARTINS et al., 2007).

Hebert et al. (2009) defendem que a não resolução com *barcode* deve-se a problemas taxonômicos na espécie, podendo ser espécies crípticas ou polimórficas, que pode sub ou superestimar os dados, em alguns casos pode ser especiação recente, ou seja, cada caso deve ser investigado individualmente. Por exemplo, o gênero de aves *Sporophila* (*Aves*) sofreu diversificação recente, havendo compartilhamento de háplotipos, fazendo com que o *barcode* não consiga diferenciar as espécies (KERR et al., 2009; CAMPAGNA et al., 2010)

2.4 Código de barra de DNA e Aves

A primeira pesquisa com Código de barra de DNA em aves foi realizada para testar sua eficácia em discriminar espécies de vertebrados, esta escolha baseou-se em que as aves constituem um dos maiores grupos e mais bem estudados da classe dos vertebrados (HEBERT et al., 2004b). Neste trabalho foram avaliadas 260 espécies de aves Norte-Americanas e todas as espécies apresentaram código de barras COI diferentes, sendo que 90% das espécies apresentaram distância interespecífica 10 vezes maior que a intraespecífica. Os resultados identificaram quatro prováveis novas espécies de aves Norte-Americanas, sugerindo que uma pesquisa global levará ao reconhecimento de muitas espécies de aves adicionais (HEBERT et al., 2004b).

Hebert et al. (2004b) analisando aves Norte-Americanas destaca que as glaciações recentes podem ter diminuído a variabilidade genética dentro das espécies através da redução das populações ou pode ter aumentado a variação entre espécies pela extinção de muitos

táxons irmãos. Por outro lado, analisar as espécies que ocorrem nos trópicos é um desafio, devido à maior diversidade genética intraespecífica, o que faz com que as análises precisem ser taxonomicamente mais amplas e que haja necessidade de estender a área de estudo, além das regiões geográficas focais para garantir que potenciais *táxons* irmãos sejam avaliados, podendo assim ser discriminados (MORITZ e CICERO 2004). Contudo, Lijtmaer et al. (2011) ressalta que a maior variação intraespecífica nos trópicos levaria a um aumento de espécies sinalizadas pelo Códigode barra de DNA, o que levaria a uma análise mais profunda e, conseqüentemente, aumentaria a sua utilidade como ferramenta para a descoberta de espécies novas, sem comprometer a sua eficiência para identificação das espécies.

Kerr, et al. (2007) fizeram uma análise mais abrangente da avifauna da América do Norte totalizando 643 espécies, representando 93% da avifauna dos EUA e Canadá, o trabalho mostra que 94% das espécies apresentaram grupos distintos de código de barras de DNA. Em 6% das espécies houve compartilhamento ou sobreposição no código de barras de DNA que correspondem a pequenos conjuntos de espécies estreitamente relacionadas, a maioria dos quais hibridizam regularmente. Apesar das restrições, Kerr et al., (2007) afirmam que o Códigode barra de DNA pode ser utilizado na identificação de diversos grupos de aves.

A primeira pesquisa realizada com aves tropicais utilizando Códigode barra de DNA, foi com 16 espécies da Família *Thamnophilidae* (VILAÇA et al., 2006). As divergências obtidas entre espécies (4.8 a 15.6%), intraespecífica (max 1.1%) e as congênicas (7.9 a 11.1%) foram maiores das observadas por HEBERT et al., (2004b), demonstrando que como imaginado, as espécies de aves tropicais possuem maiores divergências dos que as Norte-Americanas. Embora tenha sido um estudo preliminar, as sequências do gene COI apresentaram potencial para ser usadas como “Códigode barra de DNA” também em aves tropicais, identificando com êxito as espécies de *Thamnophilidae*.

Tavares e Baker (2008) demonstraram que é possível diferenciar espécies-irmãs utilizando barcode, onde 60 espécies-irmãs de aves foram reconhecidas como reciprocamente monofiléticas com um bootstrap de 55 a 100%. Contrariando as críticas, um código de barras de DNA único é uma forma rápida de descobrir linhagens monofiléticas dentro de uma metapopulação que pode conter espécies crípticas desconhecidas. Os autores ressaltam que é importante testar essas divisões com vários genes independentes em uma abordagem coalescente para prevenir à inferência de subdivisão populacional devido à restrição a dispersão do sexo feminino.

Em 2009, Kerr e colaboradores compararam padrões de diversidade genética da avifauna Norte-Americana com a Neotropical utilizando o Códigode barra de DNA. No total foram analisadas 500 espécies de aves coletadas na Argentina, sendo este o primeiro estudo com um grande número de aves Neotropicais. Os resultados demonstraram que 98,8% das espécies foram identificadas corretamente, apenas nove espécies não foram identificadas, destas seis pertencem ao sub-grupo de *Sporophila*, que são conhecidos por hibridizar e por terem divergido recentemente, ou seja, ainda compartilham os mesmos haplótipos. A avifauna da Argentina apresentou divergência intraespecífica de 3,3% mostrando divergências maiores que o limite de 2,4% sugerido por Hebert et al. (2004b) e maiores dos que as Norte-Americanas (2,7%). O mesmo ocorreu com a distância entre as espécies congênica na América do Norte (4,3%) e na Argentina (6,2%). Apesar das divergências, as sequências de COI do DNAm_t comprovam a eficiência do código de barra de DNA para a identificação das espécies em ambas as configurações, 94% das aves da América do Norte e 98% das aves Argentina foram identificadas a um nível de espécie.

No mesmo ano, Gonçalves (2009) analisou o potencial do código de barra de DNA em 228 espécies de aves Neotropicais coletadas no Brasil, e assumiu o valor mínimo de 2% de divergência interespecífica, que foi proposto por Herbert et al. (2004b). Utilizando este limite, 92,5% das espécies foram identificadas corretamente. Seus resultados também mostraram que houve sobreposição dos valores de distância intra e interespecífica. Dentro das amostras que não puderam ser identificadas, 4,3% apresentaram alta diversidade intraespecífica e 3,2% correspondem a pares de espécies com baixa divergência. A autora sugere algumas hipóteses: a) alta diversidade intraespecífica pode indicar a existência de complexos de espécies crípticas, principalmente pelo fato das amostras serem de regiões diferentes do Brasil; b) as espécies com baixa divergência podem ser consequência de diversos processos evolutivos, como hibridação ou pouco tempo de divergência. Estando de acordo com as críticas de que o Código de barra de DNA não pode diferenciar espécies crípticas, híbridas e de baixa divergência (MORITZ e CICERO, 2004; MEYER e PAULAY, 2005).

Tavares et al. (2011), pesquisando 561 aves neotropicais de diferentes países (Brasil, Guiana, Argentina, Chile e México) identificaram que 93% das espécies apresentaram sequências de código de barra de DNA diferentes, podendo ser utilizadas para identificação. Algumas espécies do gênero *Trogon*, demonstraram distância interespecífica mais alta de 19 a 20%, indicando um índice mais elevado de evolução ou divergências antigas. Ao comparar espécies de áreas biogeográficas diferentes, 16 espécies de localidade de endemismo, ou

ecorregiões, apresentaram variação genética intra-específica profunda (1,54 a 13.7%). Padrões comuns e divergentes das distâncias genéticas observadas dentro e entre espécies estreitamente relacionadas sugerem que múltiplos processos geográficos moldaram a distribuição dos táxons na região Neotropical.

LIJTMAR et al. (2011) para testar as hipóteses de refúgio e estabilidade climática sobre os padrões de diversificação das espécies aviárias utilizou o banco de dados do BOLDSYSTEM de aves, totalizando: 50% da avifauna do Neotrópicos (Argentina), cerca de 52% de espécies do Paleártico e 93% do Neártico. As análises sugerem que as espécies de aves da Argentina (15%) e no Paleártico (10%) são mais velhas do que no Neártico. Estima-se que quase um terço das espécies-irmãs apareceram apenas nos últimos 700.000 anos no Pleistoceno médio a tardio no Neártico.

3. OBJETIVOS

3.1 Geral

Gerar um banco de dados de sequências do código de barras de DNA relacionadas à *vouchers* depositados na Coleção de Aves do Instituto Nacional de Pesquisa da Amazônia (INPA) para verificar a correlação entre a diversidade genética e a taxonomia vigente.

3.2 Específicos

- 1- Obter as sequências do gene *Citocromo oxidase 1* (COI) do DNA mitocondrial de das 407 espécies e sub espécies de aves Amazônicas representadas na Coleção de Aves do INPA;
- 2- Testar se o banco de dados do Código de barra de DNA é capaz de identificar essas espécies corretamente;
- 3- Verificar se há uma boa correlação entre as distâncias genéticas para COI e os gêneros atualmente reconhecidos;
- 4- Depositar os dados de sequenciamento de DNA obtidos e imagens dos *vouchers* associados a cada sequência no banco de dados BOLD.

3. REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU, S. Y. “O Método de Aristóteles para o Estudo dos Seres Vivos”. **Revista da SBHC**, n 11, 35-40, 1994.
- AGANETTE, E; ALVARENGA, L.; SOUZA, R.R.. Elementos constitutivos do conceito de Taxonomia. **Inf. & Soc.:** Est., João Pessoa, v. 20, n. 3, p. 77-93, set./dez. 2010. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/3994>>. Acesso em: 02 Agosto. 2011.
- ALEIXO, A. “Conceitos de espécie e o eterno conflito entre continuidade e operacionalidade: uma proposta de normatização de critérios para o reconhecimento de espécies pelo Comitê Brasileiro de Registros Ornitológicos.” **Revista Brasileira de Ornitologia** 15, n. 2: 297-310. 2007.
- AMORIM, D. S. “**Elementos Básicos de Sistemática Filogenética**”. 2 Edição. Ribeirão Preto: Holos, Editora e Sociedade Brasileira de Entomologia, 1997.
- BARBER, P., BOYCE, S. Estimating diversity of Indo-Pacific coral reef stomatopods through *DNA barcoding* of stomatopod larvae. **Proc. R. Soc. B**, 273, pp. 2053-2061. 2006.
- BERED, F., N.J.F., e CARVALHO, F. I. Marcadores Moleculares e sua Aplicação no Melhoramento Genético de Plantas. **Ciências Rural** , 27 (3), 513-520, 1997.
- BISBY, F. A.; SHIMURA, J.; RUGGIERO, M.; EDWARDS, J. e HAEUSER ,C. “Taxonomy, at the click of a mouse Informatics and taxonomy are working together to achieve more than either could alone.” **Nature** 418 (Jully): 367. 2002.
- BLAXTER, M.L. “The promise of a DNA taxonomy.” **Phil. Trans. R. Soc. Lond. B** (359),: 669–679. 2004.
- CBRO. Comitê Brasileiro de Registros Ornitológicos. Disponível em <http://www.cbro.org.br> Acessado em 23/09/11 . 2011.
- CAMPAGNA, L., LIJTMAER, D. A., Kerr, K., BARREIRA, A. S., HEBERT, P., LOUGHEED, S.. DNA Barcode provide new evidence of a recent radiation in the genus *Sporophila* (Aves: Passeriformes). **Molecular Ecology Resources**, 10, pp. 449-458. 2010.
- CARVALHO, D. C., SEERIG, A., MELO, D. C., SOUSA, A. B., PIMENTA, D., OLIVEIRA, D. A.. Identificação molecular de peixes: o caso do Surubim (*Pseudoplatystoma spp.*). **Rev Bras Reprod Anim**, 32 (4), 215-219. 2008.
- CASIRAGHI, M., LABRA, M., FERRI, E., GALIMBERTI, A., DE MATTIA, F. . *DNA barcoding*: a six-question tour to improve users’ awareness about the method. **Briefings In Bioinformatics**. , 11 (4), pp. 440-453. 2010a.
- CHEN, J., LI, Q., KONG, L., YU, H.. How BarcodeComplement Taxonomy and Explore Species Diversity: The Case Study of a Poorly Understood Marine Fauna. **Plos One**, 6 (6), 1-9. 2011

DAYRAT, B.. “Towards integrative taxonomy.” **Biological Journal of the Linnean Society** 85: 407-415. 2005.

DeSALLE, R., EGAN, M. G.; SIDDALL, M. . The unholy trinity: taxonomy, species delimitation and *DNA barcoding*. **Phil. Trans. R. Soc. B**, 360, 1905-1916. 2005.

DINCA, V., ZAKHAROV, E. V., HEBERT, P. D.; VILA, R. . DNA barcode reference library for a country’s butterfly fauna reveals high performance for temperate Europe. **Proc. R. Soc. B**, 278, 347-355. 2011.

DONEGAN, T., QUEVEDO, A., MCMULLAN, M., & SALAMAN, P.. Revision of the status of bird species occurring or reported in Colombia 2011. **Conservación Colombiana**, 15. (31 October) 2011.

ERICSON, P. G., ZUCCON, D., OHLSON, J. I., JOHANSSON, U. S., ALVARENGA, H., & PRUM, R. O. . Higher-level phylogeny and morphological evolution of tyrant Xycatchers, cotingas, manakins, and their allies (Aves: Tyrannida). **Molecular Phylogenetics and Evolution**, 40, 471-483. 2006.

EDIT,. European Distributed Institute of Taxonomy. (<http://www.e-taxonomy.eu/>). Accessed em 28/07/11. 2011.

FOLMER, O., BLACK, M., HOEH, W., LUTZ, R., VRIJENHOEK, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. **Mol. Mar. Biol. Biotechnology**, n. 3, p. 294–299. 1994.

GODFRAY, H. C. J. “Challenges for taxonomy.” **Nature** 417 (May): 17-19. 2002.

GONÇALVES, P. F.M.. “**O potencial do Código de barra de DNA na identificação de espécies de aves neotropicais.**” Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências, na área de Biologia/Genética., 97p. 2009.

HACKETT, S. J., KIMBALL, R. T., REDDY, S., BOWIE, R. C., BRAUN, E. L., BRAUN, M. J., et al. . A Phylogenomic Study of Birds Reveals Their Evolutionary History. **Science** **320**, 1763-1767. 2008.

HAJIBABAEI, M.; S., GREGORY A.C.; HEBERT, P. D.N. e HICKEY, D.I A.. “*DNA barcoding*: how it complements taxonomy, molecular phylogenetics and population genetics.” **TRENDS in Genetics** 23, n. 4 : 167-172. 2007.

HICKERSON, M. J., MEYER, C. P.; MORITZ, C.. *DNA barcoding* Will Often Fail to Discover New Animal Species over Broad Parameter Space. **Syst. Bio.**, 55 (5), 729-739. 2006.

HEBERT, P. D. N, CYWINSKA, A.; BALL, S. L; De WAARD J. R. “Biological identifications through DNA Barcode.” **Proc R Soc Lond B** (270):313-322. 2003a.

HEBERT, P. D. N.; RATNASINGHAM, S.; DeWAARD, J. R.. “*Barcoding* animal life: cytochrome c oxidase subunit 1 divergences among closely related species.” **Proc Biol Sci** (270) : 96-99. 2003b.

HEBERT, P. D. N.; STOECKLE, M. Y.; ZEMLAK, T. S. and FRANCIS, C. M.. “Identification of Birds through DNA Barcode.” **PLoS BIOLOGY** (2), no. 10 : 1657-1663. 2004b.

HEBERT, P. D.N.; PENTON, E.H; BURNS, J. M.; JANZEN, D.H. and HALLWACHS, W. “Ten species in one: *DNA barcoding* reveals cryptic.” **PNAS**, 12-October: 1-6. 2004a.

HENRIQUES, J. **Identificação Molecular (DNA Barcode) Dos Peixes Da Bacia Do Rio Ribeira De Iguape E Dos Rios Costeiros Do Estado De São Paulo.** *Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas (Zoologia) do Instituto de Biociências de Botucatu, para obtenção do título de Mestre.* 2010.

HARRISON, R. Animal Mitochondrial DNA as a Genetic Marker in Population and Evolutionary Biology. *TREE* , 4 (7). 1989.

ISLER, M.L.; WHITNEY. B. M. Species Limits in Antbirds (Thamnophilidae): The ScaleBacked Antbird (*Willisornis poecilinotus*) Complex. **The Wilson Journal of Ornithology**, 123(1):1-14. 2011.

KAMARUZZANAN, B., JONH, A., ZALEHA, K., & JALAL, K. Molecular Phylogeny of Horseshoe Crab. **Asian Journal of Biotechnology**, pp. 1-8. 2011.

KERR, K.C. R.; STOECKLE, M.Y.; DOVE, C.J.; WEIGT, L. A; FRANCIS, C. M. e HEBERT. P.D. N. “Comprehensive DNA Barcode coverage of North American Birds.” **Molecular Ecology Notes** (7): 535-543. 2007.

KERR, K. C., BIRKS, S. M., KALYAKIN, M. V., RED'KIN, Y. A., KOBLIK, E. A., HEBERT, P. D. Filling the gap - COI *barcode* resolution in eastern Palearctic birds. **Frontiers in Zoology**, 6 (29), 1-13. 2009.

KOSMANN, C. (2009). **Código De Barras (DNA Barcode) De Dípteros De Interesse Forense.** *Dissertação apresentada à Coordenação do Programa de Pós-Graduação em Ciências Biológicas, Área de concentração em Entomologia, da Universidade Federal do Paraná, como requisito parcial à obtenção do grau de Mestre* , 74. 2009.

JAVONILLO, R., MALABARBA, L. R., WEITZMAN, S.,; BURNS, J. R. “Relationships among major lineages of characid fishes (Teleostei: Ostariophysi: Characiformes), based on molecular sequence data”. **Molecular Phylogenetics and Evolution** , 1-14. 2009.

LIU, Z., e CORDES, J. “DNA marker technologies and their applications in aquaculture genetics”. **Aquaculture** , 238, 1-37. 2004.

LIJTMAYER, D. A.,; KERR, K.C. R.; BARREIRA, A.S.; HEBERT, P. D. N. AND TUBARO, P.L.. “DNA Barcode Libraries Provide Insight into Continental Patterns of Avian Diversification.” **PLoS ONE** 6(7): e20744. doi:10.1371/journal.pone.0020744 6, n. 7 (July). 2011.

LUKHTANOV, V. A., SOURAKOV, A., ZAKHAROV, E. V., & HEBERT, P. D. *DNA barcoding* Central Asian butterflies: increasing geographical dimension does not significantly reduce the success of species identification. **Molecular Ecology Resources** , 9, 1302–1310. 2009.

MARINI, M. Â. e GARCIA, F. I. “Conservação de aves no Brasil.” (1), **Megadiversidade** no. 1 : 95-102. 2005.

MARTINS, J. JR. ; SOLOMON, S.E; MIKHEYEV, A.S.; MUELLER, U.G. ; ORTIZ, A. e BACCI, M. Jr. "Blackwell Publishing Ltd Nuclear mitochondrial-like sequences in ants: evidence from *Atta cephalotes* (Formicidae: Attini)." **Insect Molecular Biology** (16), no. 6: 777-784. 2007.

MAYR, E. (1946). The Number Of Species Of Birds. **The Auk** , 63, 64-69.

MEYER, C. P. e PAULAY, G. "DNA barcoding: Error Rates Based on Comprehensive Sampling." **PLoS Biology** 3, n. 12 (December): p422. 2005.

MORA, C.; TITTENSOR, D.P.; ADL, S.; SIMPSON, A.G. B and WORM, B. "How Many Species Are There on Earth and in the Ocean?" **PLoS Biology**. 9, no. 8 (August): 1-8. 2011.

MORITZ, C. e CICERO, C. "DNA barcoding: Promise and Pitfalls." **PLoS Biology**, 1-Outober: 1529-1531. 2004.

MORITZ, C., DOWLING, T., BROWN, W. Evolution Of Animal Mitochondrial Dna: Relevance For Population Biology And Systematics. **Ann. ReI". Ecol. Syst.** , 18, pp. 269-292. 1987.

NARO-MACIEL, E., REID, B., FITZSIMMONS, N. N., LE, M., DESALLE, R., & AMATO, G. Barcode for globally threatened marine turtles: a registry approach to documenting biodiversity. **Molecular Ecology Resources**, 10, 252–263. 2010.

NYARI, A., Phylogeographic patterns, molecular and vocal differentiation, and species limits in *Schiffornis turdina* (Aves), **Molecular Phylogenetics and Evolution**. 154–164, 2007.

ORTIZ, M. **Validação Do DNA barcoding Como Identificador De Espécies: Um Estudo De Ampla Amostragem Com O Gênero Pseudoplatystoma (SILURIFORMES; PIMELODIDAE) Na Amazônia**. Dissertação apresentada ao Programa de PósGraduação Genética, Conservação e Biologia Evolutiva do Instituto Nacional de Pesquisas da Amazônia para obtenção do título de mestre em Genética Conservação e Biologia Evolutiva. 2010

OKUMUS, I., e ÇIFTCI, Y. Fish Population Genetics and Molecular Markers: II- Molecular Markers and Their Applications in Fisheries and Aquaculture. **Turkish Journal of Fisheries and Aquatic Sciences**, 3, 51-79 . 2003.

PADIAL, J.M.; MIRALLES, A.; RIVA, I.D.L.; VENCES, M. "RTehvievew integrative future of taxonomy." **Frontiers in Zoology** (7), no. 16: 1-14. 2010.

RAPINI, A. "MODERNIZANDO A TAXONOMIA." **Biota Neotropica** 4, no. 1,2004.

RATNASINGHAM, S. e HEBERT, P.N. "BOLD: The Barcode of Life Data System (www.barcodinglife.org)." **Molecular Ecology Notes**,: 1-10. 2007.

ROBIDEAU, G., & AL, E. DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. **Molecular Ecology Resources**, 11, pp. 1002-1011. 2011.

RUBINOFF, D. e HOLLAND, B.S. "Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference." **Syst. Biol.** 54, (6): 952-961. 2005.

RUBINOFF, D.. *DNA barcoding* Evolves into the Familiar. **Conservation Biology**, 20 (5), 1548-1549. 2006a.

RUBINOFF, D.. Utility of Mitochondrial Barcode in Species Conservation. **Conservation Biology**, 20 (4), pp. 1026-1033. 2006b.

STOECKLE, M.; WAGGONER, M. P.E.; AUSUBEL, J. “O Códigode Barra da Vida ilustrado: Objetivos, Justificativa, Resultados. V 1.3 . (<http://barcoding.si.edu>) acessado em 12/10/11. 2005.

TAVARES, E. S.; GONÇALVES, P.; MIYAKI, C.Y. e BAKER, A.J. “DNA Barcode Detects High Genetic Structure within Neotropical Bird Species.” **PLoS ONE** 6, n. 12 (December): 1-13, 2011.

TAVARES, E. S., BAKER, A. J. Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. **BMC Evolutionary Biology**, 8 (81), 1-14.2008.

THALMANN, O.; HEBLER, J.; POINAR, H.N ; PÄÄBO, S. and VIGILANT, L. “Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes.” **Molecular Ecology** 13: 321–335. 2004.

TAUTZ, D., ARCTANDER, P., MINELLI, A., THOMAS, R. H., VOGLER, A. P. “A plea for DNA taxonomy”. **TRENDS in Ecology and Evolution** Vol.18 No.2: 70-74. 2003.

VILAÇA, S.T.,; LACERDA, D.R.; SARI, E.H. R. e SANTOS, F.R.. “DNA-based identification applied to Thamnophilidae (Passeriformes) species: the first barcodes of Neotropical birds.” **Revista Brasileira de Ornitologia** 14, n. 1 (março): 7-13. 2006.

VOGLER, A. P., e MONAGHAN, M. Recent advances in DNA taxonomy. **J Zool Syst Evol Res**, 45 (1), 1-10. 2007.

WOESE, C. R., KANDLER, O., & WHEELIS, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Nati. Acad. Sci. USA* , 87, 4576-4579. 1990.

CAPÍTULO I

**DNA *Barcode*: Eficácia do banco de dados do BOLD
na identificação molecular de espécies de aves
coletadas Amazônica**

DNA Barcode: Eficácia do banco de dados do BOLD na identificação molecular de espécies de aves coletadas Amazônica

BEZERRA¹, F. E.; FANTIN¹, C.; RIBAS², C.

1-Universidade do Estado do Amazonas- UEA

2- Instituto Nacional de Pesquisa da Amazônia-INPA

Resumo

O Códigode barra de DNA, tem sido utilizado com sucesso na identificação da avifauna global com mais de 90% de acerto. Banco de dados como o BOLD é importante para fazer esta integração com os dados genéticos, diante destes fatos, a presente pesquisa visa testar a eficácia do banco de dados do BOLD em identificar espécies de aves da Amazônia. Os resultados do presente trabalho demonstraram que existe sobreposição na distância intergenérica e intragenérica correspondentes das aves amazônicas, é esperado um *gap* entre as distâncias intra e interespecífica para espécies, mas ainda não está definido se este *gap* se estende para as distancias intragenérica e família. As espécies apresentaram de 93,5% a 100% de confiabilidade, demonstrando o poder resolutivo desta ferramenta taxonômica, podendo ser usado para a identificação molecular de aves da Amazônia. Dentre as 407 espécies sequenciadas, 90 são sequências inéditas no BOLD.

1. INTRODUÇÃO

O Códigode barra de DNA ou “códigode barra de DNA” é um método que utiliza um fragmento curto de DNA mitocondrial como padrão para todas as espécies partindo do princípio de que cada espécie tem sua própria sequência, ou seja, um conjunto único de pares de bases (HEBERT et al., 2003a; HEBERT et al., 2003b; HEBERT et al., 2004a; BLAXTER 2004). Além disso, o padrão de variação das sequências pode assinalar uma espécie nova e espécies crípticas (HAJIBABAEI et al., 2007). A primeira pesquisa com DNA *barcode* em aves revelou quatro prováveis novas espécies de aves Norte-Americanas, sugerindo que uma pesquisa global levará ao reconhecimento de muitas espécies de aves adicionais (HEBERT et al., 2004b). Desde então, o DNA barcode está sendo utilizado com sucesso na identificação de aves, com mais de 90% de acerto (HEBERT et al., 2004b; VILAÇA, et al., 2006; KERR et al., 2007; GONÇALVES, 2009; KERR et al., 2009; TAVARES, et al., 2011; LIJTMAER et al. 2011). Além de diferenciar espécies irmãs (TAVARES e BAKER, 2008) e o COI junto com outros genes está sendo utilizado para a filogenia de aves (ARBELÁ EZ-CORTÉS et al., 2012).

O BOLD (*The Barcode of Life Data System*) integra dados morfológicos, moleculares e geográficos. Portanto, a Taxonomia Integrativa visa delimitar as unidades da diversidade da vida a partir de perspectivas múltiplas e complementares como a filogeografia comparativa,

morfologia, genética de populações, ecologia, comportamento, desenvolvimento, etc. (DAYTRAT, 2005). Banco de dados como o BOLD é importante para fazer esta integração com os dados genéticos, como neste trabalho, e além de ser utilizado para ajudar a descrever a diversidade de espécies onde somente a morfologia é insuficiente (CHEN et al 2011; NAROMACIEL et al., 2010).

Analisar geneticamente as espécies que ocorrem nos trópicos é um desafio, devido à maior diversidade genética, o que faz com que as análises precisem ser taxonomicamente mais amplas e há necessidade de estender a área de estudo além das regiões geográficas focais, para garantir que potenciais táxons irmãos sejam avaliados, podendo assim ser discriminados (MORITZ e CICERO 2004). Contudo, Lijtmaer et al. (2011) ressaltam que a maior variação intraespecífica nos trópicos levaria a um aumento de espécies sinalizadas pelo Códigode barra de DNA, o que acarretaria em uma análise mais profunda e, conseqüentemente, aumentaria a sua utilidade como ferramenta para a descoberta de espécies novas, sem comprometer a sua eficiência para identificação das espécies.

Apesar do grande número de pesquisas com o Códigode barra de DNA em aves (HEBERT et al., 2004b; VILAÇA, et al., 2006; KERR et al., 2007; KERR et al., 2009; TAVARES, et al., 2011; LIJTMAER et al. 2011), ainda existem poucas realizadas na Amazônia (GONÇALVES, 2009; TAVARES et al., 2011). Testes de desempenho do *barcode* em diversos grupos biológicos têm mostrado diferenças marcantes no sucesso de identificação, havendo necessidade de estudos com objetivo de testar o método em grupos com taxonomia sólida (DINCA et al., 2011). Diante destes fatos, a presente pesquisa visa testar a eficácia do banco de dados do BOLD em identificar espécies de aves da Amazônia utilizando o gene Citocromo C Oxidase 1 (COI) como código de barra de DNA.

2. MATERIAL E MÉTODOS

2.1 Amostragem

Foi selecionada uma amostra por espécie da Coleção de Tecidos do Instituto Nacional de Pesquisa da Amazônia (INPA), totalizando 422 espécies de aves. Estão representadas em 64 Famílias, 236 Gêneros e 19 Ordens (Tab. 1). Os indivíduos foram identificados no momento da coleta através da análise morfológica pelos ornitólogos e os espécimes *vouchers* estão depositados na Coleção de Aves do INPA. Os tecidos musculares estão armazenados em

álcool 100%. Para as análises intra e intergenéricas foram utilizados 78 gêneros com mais de uma amostra, totalizando 273 amostras. As amostras foram coletadas em 20 pontos da Amazônia nos anos 2003 a 2011 (Fig. 2).

Tabela 1. Tabela da representatividade de Ordens, Famílias, Gêneros e Espécies de aves amazônicas utilizadas para o presente trabalho.

Ordens	Família	Gênero	Espécies
Accipitriformes	2	2	2
Apodiformes	1	13	19
Caprimulgiformes	2	6	10
Charadriiformes	2	4	4
Columbiformes	1	3	3
Coraciiformes	2	3	5
Cuculiformes	1	3	3
Falconiformes	1	1	1
Galbuliformes	3	11	22
Galliformes	2	6	7
Gruiformes	3	3	4
Opisthocomiformes	1	1	1
Passeriformes	33	154	294
Pelecaniformes	1	2	2
Piciformes	4	12	23
Psittaciformes	1	7	9
Strigiformes	1	2	3
Tinamiformes	1	2	3
Trogoniformes	1	1	7
Total	63	236	422

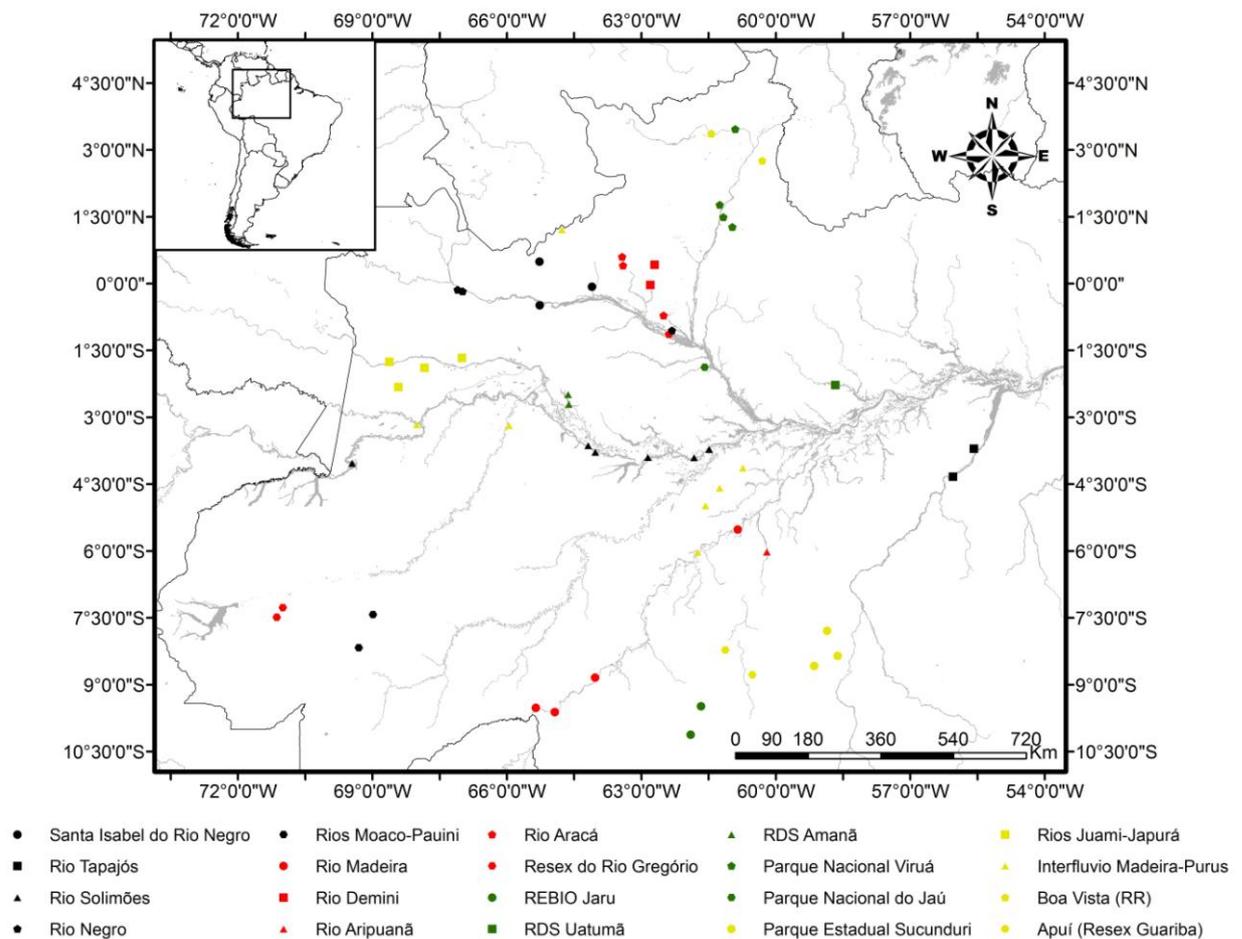


Figura 1. Mapa de distribuição dos pontos de coletas das amostras de tecido de aves da coleção de aves do INPA.

2.2. Extração e sequenciamento de DNA.

A extração foi feita com o kit do fabricante PROMEGA *Wizard Genomic DNA*. A amplificação do gene mitocondrial citocromo c oxidase subunidade I (COI) foi realizada por reação da cadeia de polimerase (PCR) com os *primers* LTyr (Direito) 5'-TGTA AAAAGGWCTACAGCCTAACGC-3' e o reverso COI908aH2- 5'-GTRGCNGAYGTRAARTATGCTC-3' (GONÇALVES, 2009). Também foram utilizados os *primers* Birds F1/F2 (HEBERT et al., 2004b) e M13 (IVANOVA et al., 2007) em 20µL de reação com os seguintes reagentes: 3,8µL de H₂O, 4 µL de tampão (*Flexi Buffer-5x Green GoTaq*), 2,0 µL do *primer* LTyr/ COI908aH2 (10mn); 4,0 µl de MgCl₂ (25mM), 2,0 µL de DNTP (1µm), 0,2 µL de Taq DNA Polimerase (Invitrogen- 5U/µl). A amplificação foi realizada com ciclos de desnaturação a 95°C por 5 minutos (min.), no anelamento 36 ciclos a 94°C por 40 segundos (seg.), a 50°C por 40 seg., 72°C por 1 min., e na extensão a 72°C por 7 min. (TAVARES et al., 2011). A purificação do PCR foi realizada com PEG 8000 (20% NaCl

2,5M). Para a reação de sequenciamento foi utilizado o Kit BigDye Terminator v3.1 Cycle Sequencing (Applied Biosystems) e os *primers* LTyr e o reverso COI908aH2 de acordo com Platt et al (2007). As amostras foram sequenciadas no analisador automático de DNA ABI 3130xl DNA Analyzer (Applied Biosystems) de acordo com o protocolo do fabricante.

2.3. Análises das sequências de DNA

As sequências foram alinhadas nos programas BIOEDIT 7.0.9 (HALL, 1999) e CLUSTALW (THOMPSON et al., 1996). Para a determinação dos grupos específicos, foi utilizado o método de agrupamento de vizinhos Neighbour-joining (NJ) (SAITOU e NEI 1987) sob o modelo de distância genética Kimura-2-parameter (K2P) (KIMURA, 1980) utilizando o software MEGA5 (Molecular Evolutionary Genetics Analysis) (TAMURA et al., 2011). A análise de confiabilidade de cada nó foi realizado pelo método de *bootstrap*, 1000 réplicas (FELSENSTEIN, 1985). Foi feita a comparação das sequências nucleotídicas de todas as espécies com as sequências do banco de dados do BOLD (The Barcode of Life Data System), para verificar a identificação correta das espécies.

3. RESULTADOS

Das 422 amostras obtidas na Coleção, havia efetivamente 407 espécies diferentes, destas 297 foram identificadas corretamente com o BOLD (72.97%). Nenhuma sequência apresentou *numts* e a maioria (407 espécies) não teve sequências repetidas com outras espécies. Somente 13 espécies apresentaram sequências repetidas devido erros de identificação, sendo que três espécies foram identificadas com seu sinônimo ou estavam com a nomenclatura desatualizada sendo, portanto, contadas duas vezes.

Das 407 espécies identificadas 90 espécies não possuem sequências depositadas (22.11%) e outras nove espécies não dispõem de sequência pública (2.21%). Apenas 29 amostras apresentaram erros de identificação (6.87%), sendo 20 erros de identificação taxonômica (Tab.2), seis foram trocas de identificação nas etiquetas de coleta (Tab. 3) e duas espécies apresentaram erros de identificação no banco de dados do BOLD.

O tamanho das sequências foi de 650pb, a composição média das bases nitrogenadas da região do Códigode barra de DNA de aves apresentou Citosina com a maior média (31.8%) (Fig. 2). O gene COI apresentou media de transição e transversã das sequencias foram de 1,4% (59,00/52,00) com 275 sítios polimórficos e 516 mutações.

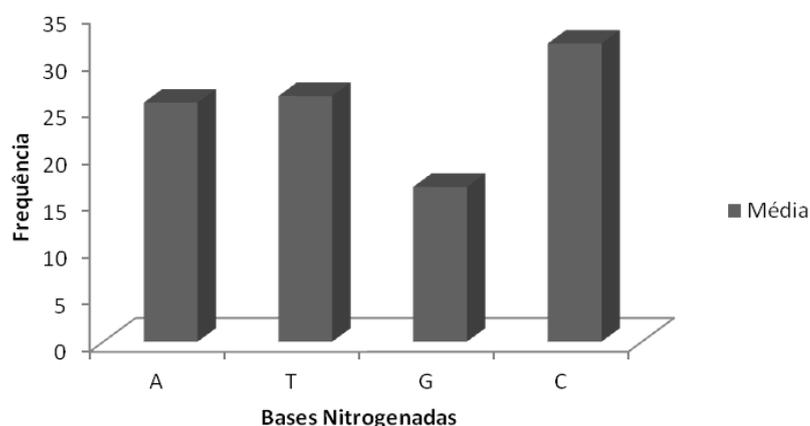


Figura 2. Composição média das bases nitrogenadas da região do Código de barra de DNA de aves coletadas na Amazônia.

A porcentagem de identificação da espécie no BOLD variou de 93.57% a 100% de similaridade de sequências. As espécies que apresentaram erros de identificação foram reidentificadas através da análise dos *vouchers* depositados na Coleção de Aves/INPA. Assim os erros foram classificados em três tipos: I) erro de identificação taxonômica (equivoco na identificação dos espécimes na Coleção de Tecidos e não nas sequências obtidas), II) troca de identificação nas etiquetas no momento de coleta dos tecidos e III) erro de identificação do BOLD.

Tabela 2. Lista de espécies de aves amostradas na Amazônia que apresentaram erro de identificação taxonômica e a identificação obtida de acordo com a porcentagem de similaridade das sequências de COI no BOLD (*The Barcode of Life Data System*).

N ^o do Tecido	Identificação em campo	Espécie identificada pelo BOLD	% de similaridade
A 1800	<i>Caprimulgus maculicaudus</i>	<i>Caprimulgus cayennensis</i>	99.69
A 2216	<i>Sporophila castaneiventris</i>	<i>Sporophila minuta</i>	99
A 2408	<i>Deconychura longicauda</i>	<i>Deconychura stictolaema</i>	100
A 3208	<i>Xiphorhynchus obsoletus</i>	<i>Xiphorhynchus elegans</i>	98
A 984	<i>Hylophilus semicinereus</i>	<i>Hylophilus pectoralis</i>	93.06
A 2086	<i>Chlorestes notata</i>	<i>Hylocharis cyanus</i>	99
A 2304	<i>Xiphorhynchus pardalotus</i>	<i>Xiphorhynchus obsoletus</i>	100
A 2500	<i>Thamnophilus stictocephalus</i>	<i>Thamnophilus amazonicus</i>	100
A 2504	<i>Elaenia chiriquensis</i>	<i>Elaenia parvirostris</i>	99.85
A 1661	<i>Pernostola minor</i>	<i>Myrmoborus myotherinus</i>	97.64
A 3184	<i>Thamnomanes saturninus</i>	<i>Glyphorhynchus spirurus</i>	94.99
A 961	<i>Lathrotriccus euleri</i>	<i>Cnemotriccus fuscatus</i>	99.7
A 2377	<i>Philydor ruficaudatum</i>	<i>Automolus ochrolaemus</i>	99.67

A 2180	<i>Cranioleuca vulpina</i>	<i>Synallaxis gujanensis</i>	100
A 167	<i>Philydor ruficaudatus</i>	<i>Philydor erythrocerum</i>	98.6
A 1155	<i>Catharus ustulatus</i>	<i>Catharus minimus</i>	99.85
A 2190	<i>Rhynchocyclus olivaceus</i>	<i>Tolmomyias sulphurescen</i>	100
A 691	<i>Hemitriccus minor pallens</i>	<i>Lophotriccus galeatus</i>	96.46
A 1842	<i>Atalotriccus pilaris</i>	<i>Hemitriccus margaritaceiventer</i>	98.53
A 5349	<i>Tachyphonus coronatus</i>	<i>Tachyphonus cristatus</i>	

Paras espécies que apresentaram o erro do tipo II, foram refeitos da extração ao sequenciamento do DNA confirmando o erro (Tab.3). Em seguida foi feita a reidentificação do *voucher* das amostras confirmando que a identificação em campo estava correta e que houve uma troca na identificação dos tubos de tecidos.

Tabela 3. Lista de espécies de aves amostradas na Amazônia que apresentaram trocas de identificação nas etiquetas do tubo de tecido e identificação obtida de acordo com a porcentagem de similaridade das sequências de COI no BOLD (*The Barcode of Life Data System*) (erro tipo II).

N ^o do Tecido	Identificação em campo	Espécie identificada pelo BOLD	% de similaridade
A 2812	<i>Mionectes oleagineus</i>	<i>Myrmotherula surinamensis</i>	99.84
A 404	<i>Phaeomyias murina</i>	<i>Pteroglossus viridis</i>	99.7
A 134	<i>Psophia crepitans</i>	<i>Alopochelidon fucata</i>	91.98
A 3086	<i>Myiarchus tuberculifer</i>	<i>Cercomacra sp.</i>	100
		<i>Cercomacra laeta</i>	99.85
A 5201	<i>Ramphocaenus melanurus</i>	<i>Cercomacra laeta</i>	9.36
A 633	<i>Euphonia plúmbea</i>	<i>Glyphorynchus spirurus</i>	98.92

As espécies que apresentaram erros do tipo I e II, após a análise da Árvore de *Neighbor-joining* ficaram distribuídas na família correta ou em ramos separados da família, como por exemplo, as amostras com trocas de identificação de tubos de tecidos em campo (Tipo II)(A404; A1689 e A3086) que foram identificadas como espécies da família Tyrannidae ficaram em ramos separados da família referida (Fig. 3).

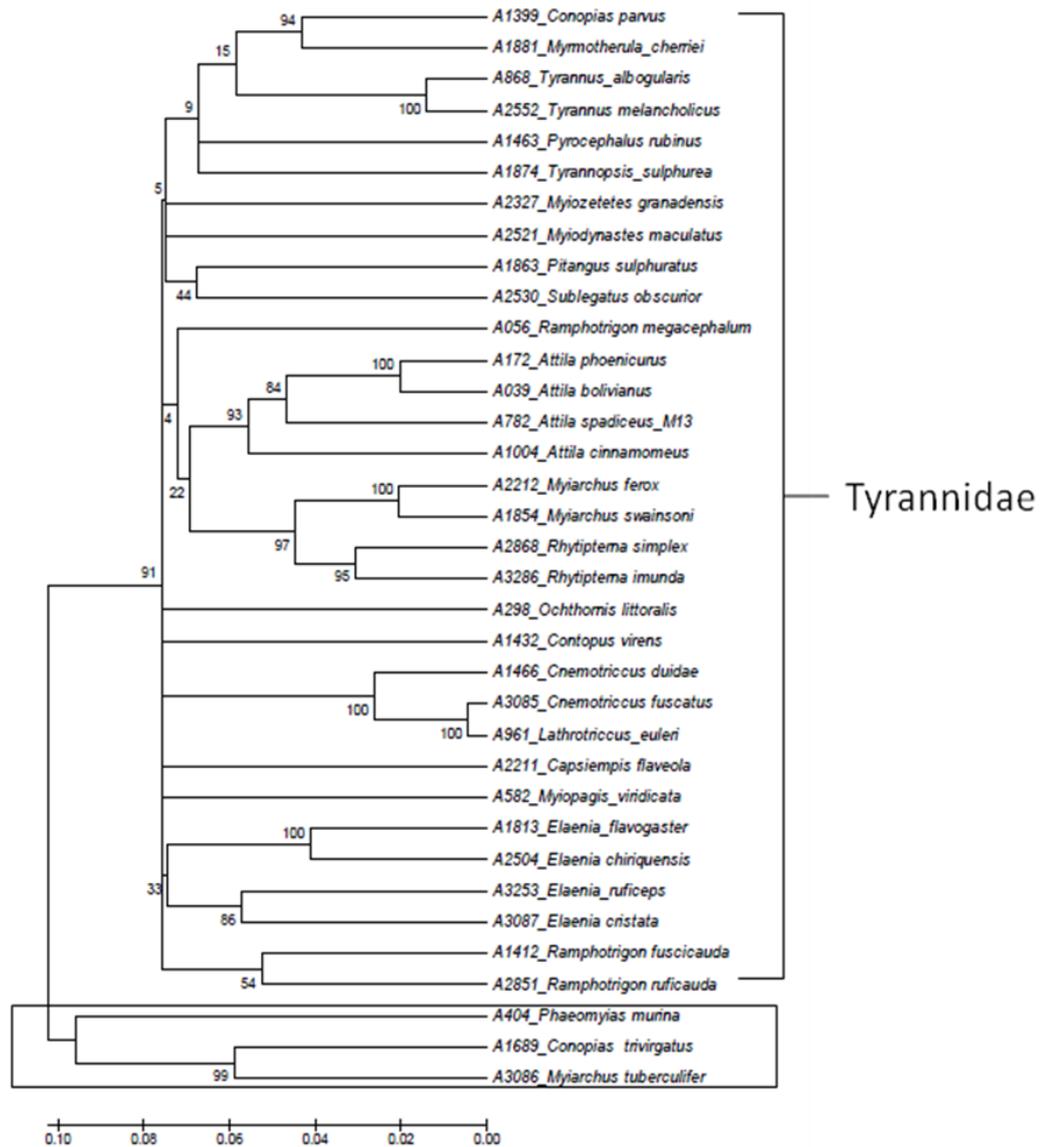


Figura 3. Árvore de *Neighbour-joining* da Família Tyrannidae com valores de bootstrap, no ramo em destaque as espécies com trocas de identificação em campo que ficaram fora do clado da família.

Nas árvores de NJ a maioria das amostras pertencentes à mesma família ficaram no mesmo clado, com exceção de algumas espécies que apesar de estarem corretas ficaram em ramos separados. As espécies pertencem às famílias Thraupidae, Icteridae, Polioptilidae, Hirundinidae, Rhynchocyclidae, Tyrannidae (Fig. 4). Oito gêneros não apresentaram monofilia *Tachyphonus*, *Basileuterus*, *Myrmeciza*, *Mymoththerula*, *Cercomacra*, *Hypocnemis*, *Hemitriccus* e *Sporophila* (Fig. 1 Árvore em anexo). O gênero *Dolospingus* possui somente uma espécie (*Dolospingus fringiloides*) que ficou no clado das *Sporophila* (Fig. 1 Árvore em anexo).

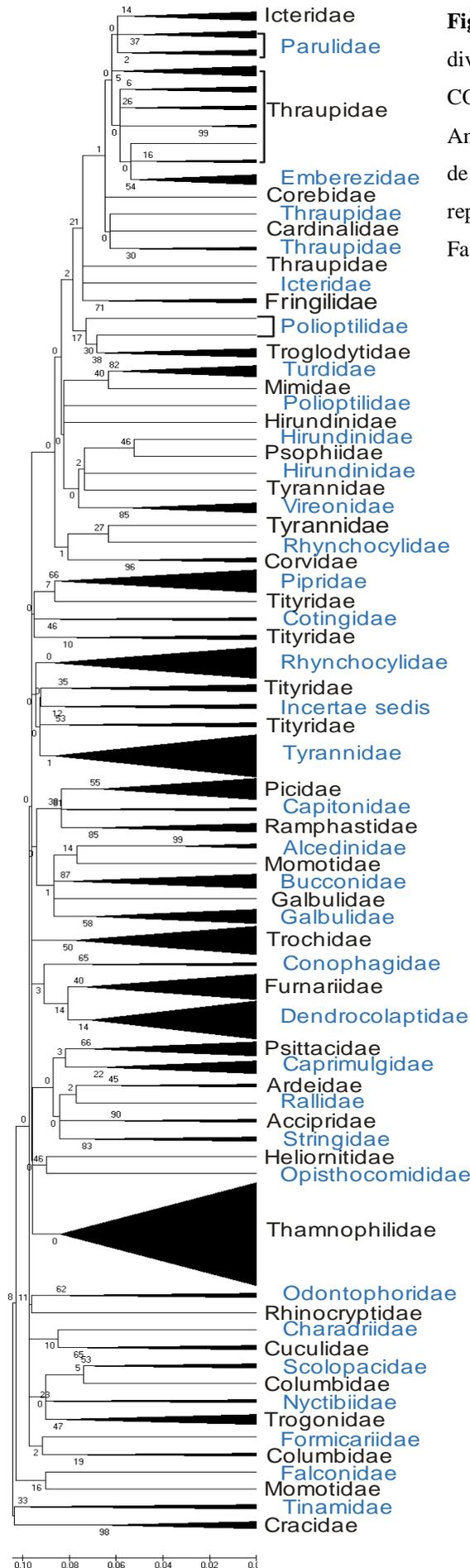


Figura 4. Árvore de *Neighbor-joining* da divergência nucleotídica das sequências de COI por família de aves amostrada na Amazônia com valores de *bootstrap* e régua de distância genética. A espessura dos ramos representa a quantidade de táxons por Família.

A Família Rhynchocyclidae apresentou três espécies identificadas com erro taxonômico. O erro foi identificado quando as sequências foram comparadas com as do BOLD e depois confirmadas com o *voucher*. As amostras **A691**, **A1842** e **A2190**, identificadas previamente como *Hemitriccus minor pallens*, *Atalotriccus pilaris* e *Rhynchocyclus olivaceus* respectivamente, aplicando as técnicas de identificação BOLD e árvore NJ ficaram nos ramos e correspondem respectivamente a *Lophotriccus galeatus*, *Hemitriccus margaritaceiventer* e *Tolmomyias sulphurescens* e suporte com alto valor de *bootstrap* (Fig. 5; tab. 2).

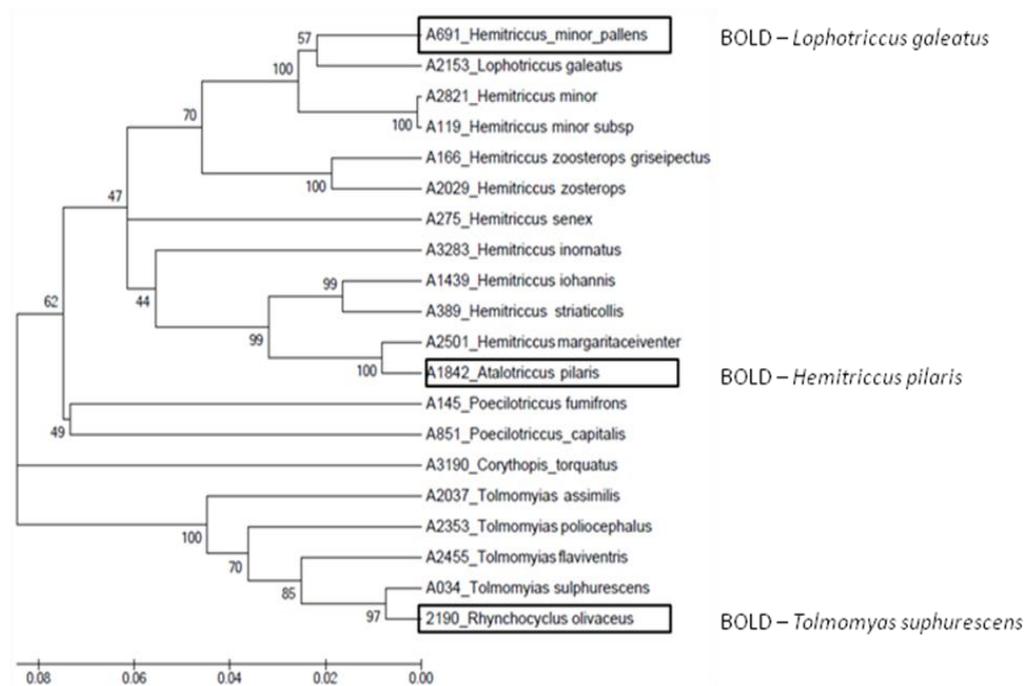


Figura 5. Árvore de *Neighbour-joining* com o valor de *bootstrap* da Família Rhynchocyclidae, em destaque as espécies de aves identificação diferente do BOLD (*The Barcode of Life Data System*).

Os erros de identificação do BOLD ocorreram pela identificação de duas ou mais espécies para uma única sequência de DNA. A sequência **A1022-** *Columbina passerina* apresentou 100% de similaridade com duas espécies (*C. passerina* e *Heliomaster longirostris*) de ordens e Família diferentes. A sequência **A1155-** identificada no campo como *Catharus ustulatus* foi tombada na Coleção de Tecidos do INPA como *Catharus minimus* e pelo BOLD apresentou três identificações com similaridade parecida de espécies do mesmo gênero *C. minimus* (99, 85%), *C. bicknelli* e *C. guttatus* (ambas com 99,84%). Sendo confirmada com vouchers que a amostra é *Catharus minimus*.

As espécies que não possuem sequências depositadas foram utilizadas para analisar a confiabilidade de o banco de dados do BOLD. Foram feitas duas análises: I) verificação da correspondência entre a identificação do banco de dados da Coleção de Tecidos com a categoria espécie-específica do BOLD, e II) verificação da correspondência entre a identificação do banco de dados da Coleção de Tecidos com a categoria geral do BOLD (que inclui todas as amostras, sequências com menores de que 500pb e amostra identificada até o nível de gênero).

Na primeira análise o banco de dados do BOLD foi aprovado com êxito (100% de confiabilidade) devido não ter identificado outra espécie presente no banco de dados. Na segunda análise, nove espécies não foram identificadas até o nível de gênero corretamente, mas identificadas corretamente até a família. Destas, quatro amostras são dos gêneros identificados morfologicamente como *Harpia*, *Liosceles*, *Odontorchilus* e *Opisthocomus*, que não possuem espécies depositadas, e cinco amostras dos gêneros *Herpsilochmus*, *Nyctibius* (duas amostras), *Pharomachrus* e *Pygiptila* que possuem poucas espécies depositadas para comparação.

A média das distâncias intergenérica e intragenérica foram 19% e 7% respectivamente (Fig.6). No gráfico de distribuição das distâncias intergenérica e intragenérica podem ser observadas a sobreposição das distâncias de 9 a 17%, com exceção de 6 e 20% e a falta de *gap*.

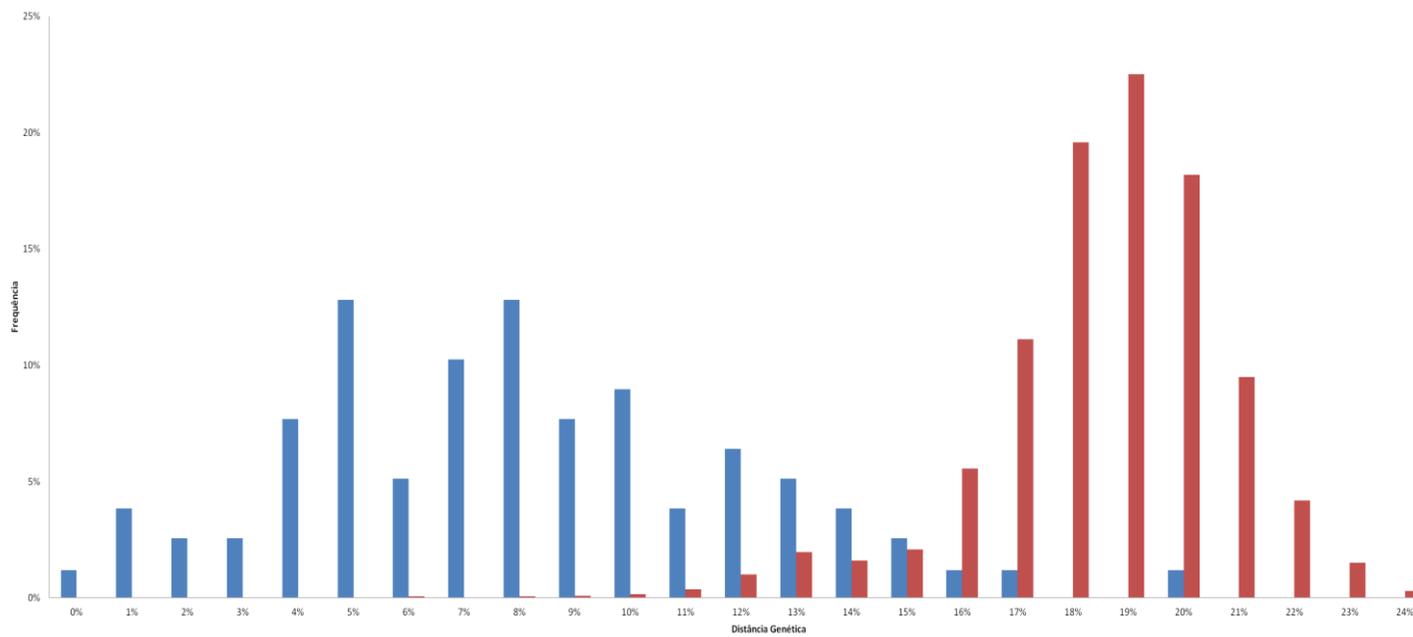


Figura 6. Frequência da distância genética intragenérica e intergenérica dos 78 generos de aves amostradas em 20 localidades na Amazônia brasileira.

4. DISCUSSÃO

O banco de Códigode barra de DNA gerado neste trabalho é uma das poucas iniciativas brasileiras para identificação molecular de aves depositadas em coleção científica. Este banco pode ser utilizado no futuro como fonte de identificação de amostras sem referencial de *vouchers* ou com origem desconhecida que forem tombados no INPA. O tamanho das sequencias (650pb) está de acordo com a proposta de Hebert et al. (2003a) para identificação de espécies. A composição média das bases nitrogenadas está similar com as encontradas em aves neotropicais (GONÇALVES, 2009).

Na validação do Códigode barra de DNA como ferramenta taxonômica para identificação de espécies de aves da Amazônia, obtivemos 72.9% de espécies identificadas corretamente, representando 24,4% da avifauna da Amazônia Brasileira (correspondente a 1.217 espécies) (CBRO 2011). As poucas pesquisas realizadas com aves neotropicais, incluindo amostras da Amazônia com sequências depositadas no BOLD, alcançaram cerca de 90% de espécies identificadas corretamente (TAVARES e BAKER, 2008; GONÇALVES, 2009).

Dentre as 407 espécies estudadas, 90 são sequências inéditas no BOLD. O sistema de identificação baseado em DNA só funciona se todas as espécies já descritas tiverem suas sequências no banco de dados (DAYRAT, 2005) e esta pesquisa visa também à complementação deste banco de sequências de aves amazônicas.

Utilizamos espécies que não tinham sequência depositada no BOLD para testar a confiabilidade do banco. Na categoria de identificação ao nível de espécie, o BOLD indicou como incapaz de combinar a sequência com o banco de dados existente, ou seja, não indicou a sequência de DNA de outra espécie presente no banco de dados, sendo confiável para identificação neste nível. Stoeckle e Kerr (2012) confirmam qualidade do banco de dados de aves, e afirma que os erros de sequenciamento de DNA encontrado não afeta a precisão de identificação da espécie pelo BOLD. Já para identificação ao nível de gênero o BOLD foi capaz de identificar corretamente em 91% dos casos. O mesmo ocorre em insetos, mesmo se a espécie comparada não tiver no banco de dados, o BOLD atribui corretamente ao nível de gênero (WILSON, et al., 2011).

O Códigode barra de DNA interage entre a Taxonomia Clássica e Molecular para uma identificação precisa das espécies, comparando o resultado das sequências aos *vouchers* no caso de sequências dúbias (cerca de 7% neste trabalho) demonstrando a valorização das

coleções e museus (HEBERT et al., 2003a; 2003b; RATNASINGHAM e HEBERT, 2007), além da importância da Taxonomia Integrativa.

Alguns dos benefícios esperados com a utilização do Código de barra de DNA são a identificação de espécies morfológicamente parecidas, a identificação de formas juvenis e adultas de uma mesma espécie (HEBERT et al., 2003a; 2003b; 2004a; 2004b). Em aves, mostramos que as espécies identificadas erroneamente em campo por terem duas ou mais espécies morfológicamente parecidas, são de fato do mesmo gênero ou pelo menos da mesma família, mas apenas com o Código de barra de DNA foi possível corrigir a identificação. Outro problema na identificação morfológica de aves está na dificuldade de identificar indivíduos jovens demais que não possuem ainda plumagem definitiva. Este foi o caso da amostra **A2500**, onde um indivíduo jovem foi confundido em campo e tombado como *Thamnophilus stictocephalus*, e depois de comparar sua sequência no banco de dados do BOLD observamos que na realidade trata-se de *T. amazonicus*. Apesar das espécies identificadas erradas em campo terem sido revisadas no momento do tombamento, *T. amazonicus* permaneceu como *T. stictocephalus*.

Outra justificativa para utilizar o Código de barra de DNA na identificação de espécies de aves pode ser exemplificada através da família Rhynchocyclidae, que apresenta espécies com sutis diferenças morfológicas. As espécies *Lophotriccus galeatus*, *Hemitriccus margaritaceiventer* e *Tolmomyias sulphureus*, equivocadamente identificadas, são consideradas evolutivamente jovens (TELLO e BATES 2007), portanto possuem baixa divergência genética (MEYER e PAULAY 2005) e ainda não acumularam diferenças morfológicas suficientes para serem identificadas com precisão apenas com esses caracteres (ALEIXO, 2007).

Em relação aos oito gêneros que na presente pesquisa demonstraram não serem monofiléticos, diferentes processos podem causar compartilhamento de haplótipos e a falta de monofilia, como táxons erroneamente divididos, separação de linhagens incompleta e hibridização (FUNK e OMLAND, 2003; CAMPAGNA et al., 2010). Separação de linhagens incompleta é mais comum em aves que não apresentam monofilia, isto ocorre quando táxons que divergiram recentemente ainda não tiveram diferenças acumuladas nos *locos* analisados (FUNK e OMLAND, 2003). A monofilia só ocorre quando a taxa está reprodutivamente isolada por longos períodos, antes disso, ocorre processo evolutivo de polifilia, parafilia e por último a monofilia (ALEIXO, 2007). Estudos estatísticos mostram que os testes de monofilia precisam ser revistos uma vez que podem ser altamente tendenciosos (DeSALLE et al., 2005; ROSENBERG, 2007).

Gêneros estudados neste trabalho, como por exemplo, *Hemitricus* e *Lophotriccus* demonstraram ser não monofiléticos em outras pesquisas com os genes ND2, Citocromo B e FIB7 (COHN- HAFT, 2000; TELLO e BATES 2007). Também não há monofiletismo do gênero *Sporophila* em análises de COI devido ao compartilhamento de haplótipos, e assim o *barcode* não consegue diferenciar pelo menos seis espécies do gênero (KERR et al., 2009; CAMPAGNA et al., 2010). Também foi confirmado o compartilhamento de um ancestral comum entre os gêneros *Sporophila*, *Oryzoborus* e *Dolospingus*, diante desse fato foi sugerida a unificação dos gêneros (ROBBINS et al., 2005). Os não monofiléticos também podem ser decorrentes de amostragem insuficiente (FUNK e OMLAND, 2003). Portanto, uma amostragem apropriada é essencial para o desenvolvimento de estudos filogenéticos (AMORIM, 1997; FUNK e OMLAND, 2003,) e alguns autores (MEYER e PAULAY, 2005; GOLDSTEIN e DeSALLE 2010) defendem que para análise com Códigode barra de DNA o conceito de monofilismo não é apropriado, já que não reflete a filogenia.

Os resultados do presente trabalho demonstraram que existe sobreposição na distância intergenérica e intragenérica correspondentes das aves amazônicas, é esperado um *gap* entre as distâncias intra e interespecífica para espécies (HEBERT et al., 2004b), mas ainda não está definido se este *gap* se estende para as distancias intragenérica e família. Já foi observado sobreposição no nível de espécies para aves neotropicais (1 à 17%) (GONÇALVES, 2009). Nesta pesquisa as distâncias intragenéricas são maiores do que Hebert et al. (2003a) para aves norte-americanas (0 a 17%), Vilaça et al. (2006) com a Família *Thamnophilidae* (7.9 a 11.1%) e similares as distâncias intraespecífica (1 à 25%) de aves neotropicais de Gonçalves (2009). Ambos os trabalhos haviam uma amostragem maior de indivíduos por espécie (2-5), sendo mais apropriado para uma melhor visualização da variação das distância (MEYER e PAULAY, 2005; DeSALLE et al., 2005; HAJIBABAEI et al., 2007) e um melhor agrupamento das espécies em suas respectivas famílias na árvore de *Neighbor-joining* (GOLDSTEIN e DeSALLE, 2010).

Os dados publicados com a técnica *barcode* na identificação de espécie utilizando o BOLD só podem ser aceitos se a similaridade for maior que 97% (RATNASINGHAM e HEBERT 2007). Na presente pesquisa as espécies foram identificadas corretamente com similaridade a partir de 93,57%. Os trópicos possuem uma maior biodiversidade genética gerando uma maior variabilidade intraespecífica, o que faz com que as análises precisem ser taxonomicamente mais amplas e que haja necessidade de estender a área de estudo, além das regiões geográficas focais, para garantir que potenciais táxons irmãos sejam avaliados, podendo assim ser discriminados (MORITZ e CICERO 2004).

O BOLD usa o sistema “Hidden-Markov Model” baseado em um modelo probabilístico para alinhar sequências parecidas em um grande banco de dados (EDDY, 1998 RATNASINGHAM e HEBERT 2007). A similaridade das sequências limitada à diagnose de espécies que divergiram recentemente (TAUTZ et al., 2003; VOGLER e MONAGHAN 2007). Este pode ser o caso de *Catharus bicknelli*, que era uma subespécie de *C. minimus* e foi elevada a espécie devido às diferenças morfológicas, de vocalização, ecológicas e genéticas (OUELLET, 1993). Mais tarde a análise genética completa do gênero mostrou que *C. minimus* é espécie irmã de *C. fuscescens* e que *C. bicknelli* compartilha um ancestral comum com as espécies irmãs (OUTLAW et al., 2003). Neste caso, o BOLD ao comparar a sequência indicou que poderia ser qualquer uma das três espécies. Poderendo também haver um engano na identificação da espécie com a sequência depositada no BOLD usada para fazer as comparações, e não pela limitação da similaridade, como por exemplo, na amostra **A1022** que teve duas identificações com similaridade idêntica no BOLD, uma correta *Columbina passerina* e outra errada *Heliomaster longirostris*.

Apesar das limitações encontradas, o Códigode barra de DNA foi uma ferramenta taxonômica capaz de identificar espécies de aves da Amazônia com confiabilidade. As informações deste trabalho serão úteis para aperfeiçoar o banco de dados do BOLD, aumentando ainda mais a confiabilidade do Códigode barra de DNA para as aves da Amazônia. Por último, o Códigode barra de DNA não deve ser utilizado como ferramenta única para identificação de espécies, mas ainda é válida e precisa (RUBINOFF e HOLLAND 2005), e parte da Taxonomia Integrativa (DeSALLE et al., 2005; HAJIBABAEI, et al., 2007; VOGLER e MONAGHAN 2007; GOLDSTEIN e DeSALLE 2010).

A partir dos dados obtidos com o *Barcode* foi possível interligar as duas coleções (Tecidos e Zoológica) (Fig.7). As sequências de DNA tiveram êxito no teste de identificação de espécies de Hebert et al. (2003a), não havendo sequências idênticas em duas espécies.

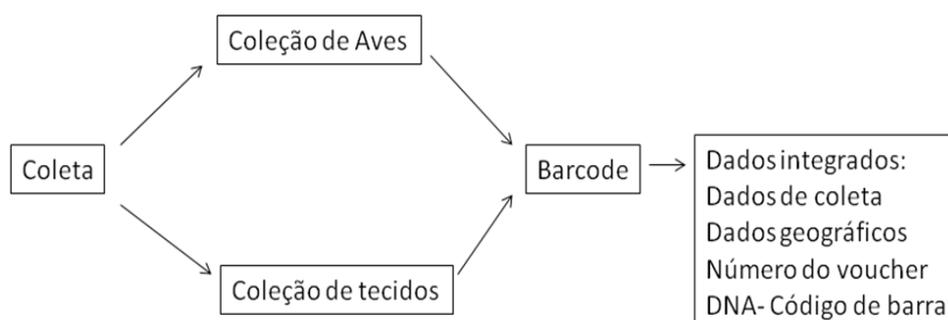


Figura 7. Fluxograma de ligação entre coleções zoológicas e coleções de tecido através do método barcode.

Quando um pesquisador deseja analisar geneticamente uma espécie, ele procura as coleções de tecidos para verificar a disponibilidade de tecidos da espécie, caso haja a espécie desejada e a coleção apresentar erros taxonômicos poderá induzir a uma pesquisa equivocada. De tal forma a presente pesquisa sugere a utilização do DNA *barcode* para integrar as coleções de tecidos e zoológicos, formando um inventário *online* na instituição que possua ambas as coleções.

5. CONCLUSÃO

Obtivemos as sequências de DNA das 422 amostras de tecidos de aves coletadas na Amazônia brasileira. O Código de barra de DNA diferenciou espécies biologicamente juvenis, morfologicamente parecidas e espécies evolutivamente jovens; e forneceu a identificação molecular de 407 espécies depositadas no Banco de Tecidos da Coleção de Aves do INPA até 2011.

Como ferramenta taxonômica, o *barcode* pode ser usado para a identificação molecular de aves da Amazônia. Apesar da falta de sequências amazônicas no banco de dados do BOLD, as espécies estudadas neste trabalho apresentaram de 93,5% a 100% de confiabilidade, demonstrando o poder resolutivo desta ferramenta. Nos casos necessários, a integração dos dados moleculares com os *vouchers* associados solucionou os problemas de identificação encontrados.

A utilização da árvore de *Neighbor-joining* e o uso do critério monofilético para o *barcode* ainda gera muita discussão e dúvidas entre os pesquisadores, necessitando assim de novas idéias para aperfeiçoar a metodologia.

A região neotropical possui a maior biodiversidade de aves do planeta, sendo que no Brasil concentra-se aproximadamente 1.763 espécies, destas no banco de dados do BOLD há somente sequências de 446 espécies. Este trabalho está acrescentando sequências de 90 espécies.

6. REFERENCIAS BIBLIOGRAFICAS

- ARBELÁ EZ-CORTÉ S, E., NAVARRO-SIGU ENZA, A., & GARCÍA-MORENO, J. Phylogeny of woodcreepers of the genus *Lepidocolaptes* (Aves, Furnariidae), a widespread Neotropical taxon. **Zoologica Scripta**, pp. 1-11. 2012.
- ALEIXO, A. “Conceitos de espécie e o eterno conflito entre continuidade e operacionalidade: uma proposta de normatização de critérios para o reconhecimento de espécies pelo Comitê Brasileiro de Registros Ornitológicos.” **Revista Brasileira de Ornitologia** 15, n. 2: 297-310. 2007.
- AMORIM, D. S. “**Elementos Básicos de Sistemática Filogenética**”. 2 Edição. Ribeirão Preto: Holos, Editora e Sociedade Brasileira de Entomologia, 1997.
- BLAXTER, M.L. “The promise of a DNA taxonomy.” **Phil. Trans. R. Soc. Lond. B** (359): 669–679. 2004.
- BOLDSYSTEMS- The Barcode of Life Database <http://www.boldsystems.org/>
- CAMPAGNA, L., LIJTMAYER, D. A., Kerr, K., BARREIRA, A. S., HEBERT, P., LOUGHEED, S. Barcode provide new evidence of a recent radiation in the genus *Sporophila* (Aves: Passeriformes). **Molecular Ecology Resources**, 10, pp. 449-458. 2010.
- CBRO. Comitê Brasileiro de Registros Ornitológicos. Disponível em <http://www.cbro.org.br> Acessado em 23/09/11. 2011.
- CHEN, J., LI, Q., KONG, L., YU, H.. How Barcode Complement Taxonomy and Explore Species Diversity: The Case Study of a Poorly Understood Marine Fauna. **Plos One** , 6 (6), 1-9. 2011.
- COHN-HAFT, M.. **A case study in Amazonian biogeography: Vocal and DNA-sequence variation in *Hemitriccus flycatchers***. Ph.D.dissertation, Louisiana State University, Baton Rouge. 2000.
- DAYRAT, B.. “Towards integrative taxonomy.” **Biological Journal of the Linnean Society** 85: 407-415. 2005
- DeSALLE, R., EGAN, M. G.,; SIDDALL, M. . The unholy trinity: taxonomy, species delimitation and *DNA barcoding*. **Phil. Trans. R. Soc. B**, 360, 1905-1916. 2005.
- DINCA, V., ZAKHAROV, E. V., HEBERT, P. D.,; VILA, R. . Complete Barcode reference library for a country’s butterfly fauna reveals high performance for temperate Europe. **Proc. R. Soc. B**, 278, 347-355. 2011.
- EDDY, S. R. Profile Hidden Markov models. **Bioinformatics Review** , 13 (9), pp. 755-763. 1998.
- FUNK, D. J., e OMLAND, K. E.. SPECIES-LEVEL PARAPHYLY AND POLYPHYLY: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA. **Annu. Rev. Ecol. Evol. Syst.**, 34, pp. 397-423. 2003.

FELSENSTEIN, J.F. "Confidence limits on phylogenies: an approach using the bootstrap". **Evolution**, v. 39, p. 783-791. 1985.

GOLDSTEIN, P. Z., & DeSALLE, R.. Integrating Barcode data and taxonomic practice: Determination, discovery, and description. WILEY Periodicals, Inc. **Bioessays**, 33, 135-147. 2010.

GONÇALVES, P. F.M. "**O potencial do Barcode na identificação de espécies de aves neotropicais.**" Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências, na área de Biologia/Genética., 97p. 2009.

HALL, T. A. "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT". **Nucleic Acids Symp.**, n. 41, p. 95-98. 1999.

HAJIBABAEI, M.; S., GREGORY A.C.; HEBERT, P. D.N. e HICKEY, D.I A.. "*DNA barcoding*: how it complements taxonomy, molecular phylogenetics and population genetics." **Trends In Genetics** 23, n. 4: 167-172. 2007.

HEBERT, P. D. N, CYWINSKA, A.; BALL, S. L; De WAARD J. R. "Biological identifications through Barcodes." **Proc R Soc Lond B** (270): 313-322. 2003a.

HEBERT, P. D. N.; RATNASINGHAM, S.; DeWAARD, J. R.. "*Barcoding* animal life: cytochrome C oxidase subunit 1 divergences among closely related species." **Proc Biol Sci** (270): 96-99. 2003b.

HEBERT, P. D. N.; STOECKLE, M. Y.; ZEMLAK, T. S. and FRANCIS, C. M.. "Identification of Birds through Barcodes." **PLoS BIOLOGY** (2), no. 10 : 1657-1663. 2004b.

HEBERT, P. D.N.; PENTON, E.H; BURNS, J. M.; JANZEN, D.H. and HALLWACHS, W. "Ten species in one: *DNA barcoding* reveals cryptic." **PNAS**, 12-October: 1-6. 2004a.

IVANOVA, N., ZEMLAK, T. S., HANNER, R. H., & HEBERT, P. D. Universal primer cocktails for fish DNA barcoding. **Molecular Ecology Notes** , 7, pp. 544-548. 2007.

KERR, K.C. R.; STOECKLE, M.Y.; DOVE, C.J.; WEIGT, L. A; FRANCIS, C. M. e HEBERT. P.D. N. "Comprehensive Barcode coverage of North American Birds." **Molecular Ecology Notes** (7): 535-543. 2007.

KERR, K. C., BIRKS, S. M., KALYAKIN, M. V., RED'KIN, Y. A., KOBLIK, E. A., HEBERT, P. D. Filling the gap - COI *barcode* resolution in eastern Palearctic birds. **Frontiers in Zoology**, 6 (29), 1-13. 2009.

KIMURA, M. "A Simple Method for Estimating Evolutionary Rates of Base Substitutions Through Comparative Studies of Nucleotide Sequences." **Journal of Molecular Evolution** (11): 111- 120. 1980.

LIJTMAYER, D. A.; Kerr, K.C. R.; Barreira, A.S.; Hebert, P. D. N. and TUBARO, P.L.. “Barcode Libraries Provide Insight into Continental Patterns of Avian Diversification.” **PLoS ONE** 6 (7): e20744. doi:10.1371/journal.pone.0020744 6, n. 7 (July). 2011.

MEYER, C. P. e PAULAY, G. “DNA barcoding: Error Rates Based on Comprehensive Sampling.” **PLoS Biology** 3, n. 12 (December): p422. 2005.

MORITZ, C. e CICERO, C. “DNA barcoding: Promise and Pitfalls.” **PLoS Biology**, 1-October: 1529-1531. 2004.

NARO-MACIEL, E., REID, B., FITZSIMMONS, N. N., LE, M., DESALLE, R., & AMATO, G. DNA barcode for globally threatened marine turtles: a registry approach to documenting biodiversity. **Molecular Ecology Resources**, 10, 252–263. 2010.

OUTLAW, D. C., VOELKER, G., MILA, B.,; GIRMAN, D. J. Evolution Of Long-Distance Migration In And Historical Biogeography Of Catharus Thrushes: A Molecular Phylogenetic Approach. **The Auk**, 120 (2), pp. 299-310. 2003.

OUELLET, H. Bicknell's Thrush: taxonomic status and distribution. **Wilson Bull** 105:545–572. 1993.

PLATT, A.R.; WOODHALL, R.W.; GEORGE J.R. “Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol”. **Biotechniques** n. 43, p.58-62. 2007.

RATNASINGHAM, S.; HEBERT, P.N. “BOLD: The *Barcode* of Life Data System (www.barcodinglife.org).” **Molecular Ecology Notes**, 1-10. 2007.

ROBBINS, M. B., BRAUN, M., HUDDLESTON, C. J., FINCH, D. W., & MILENSKY, C. M. First Guyana records, natural history and systematics of the White-naped Seedeater *Dolospingus fringilloides*. **Ibis**, pp. 334-341. 2005.

ROSENBERG, N. A.. Statistical Tests For Taxonomic Distinctiveness From Observations Of Monophyly. **Evolution**, pp. 317-323. 2007.

RUBINOFF, D. e HOLLAND, B.S. “Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference.” **Syst. Biol.** 54, (6): 952-961. 2005.’

SAITOU, N.; NEI, M.. “The *Neighbor-joining* Method: A New Method for Reconstructing Phylogenetic Trees.” **Mol. Biol. Evol.** 4,(4): 406-425. 1987.

SANGER, F., NICKLEN, S., COULSON, A.R. “DNA Sequencing with chain terminating inhibitors”. **Proc. Natl. Acad. Sci. USA**, n. 74, p. 5463-5467. 1977.

STOECKLE, M. Y., & KERR, K. C. Frequency Matrix Approach Demonstrates High Sequence Quality in Avian Barcode and Highlights Cryptic Pseudogenes. **PLOS ONE**, 7 (8), pp. 1-9.2012.

TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M. and KUMAR, S. MEGA 5. Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. **Molecular Biology and Evolution** doi: 10.1093/molbev/msr121. 2011.

TAVARES, E. S., BAKER, A. J. Single mitochondrial gene barcode reliably identify sister-species in diverse clades of birds. **BMC Evolutionary Biology**, 8 (81), 1-14.2008.

TAVARES, E. S.; GONÇALVES, P.; MIYAKI, C.Y. e BAKER, A.J. “Barcode Detects High Genetic Structure within Neotropical Bird Species.” **PLoS ONE** 6, n. 12 (December): 1-13, 2011.

TAUTZ, D., ARCTANDER, P., MINELLI, A., THOMAS, R. H., VOGLER, A. P. “A plea for DNA taxonomy”. **TRENDS in Ecology and Evolution** Vol.18 No.2: 70-74. 2003.

TELLO, J. G., e BATES, J. M. Molecular Phylogenetics Of The Tody-Tyrant And Flatbill Assemblage Of Tyrant Flycatchers (Tyrannidae). **The Auk**, 1. 2007.

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. “CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weight matrix choice”. **Nucleic Acids Res.**, v. 22, 4673–4680. 1994.

VILAÇA, S.T.,; LACERDA, D.R.; SARI, E.H. R. e SANTOS, F.R.. “DNA-based identification applied to Thamnophilidae (Passeriformes) species: the first barcodes of Neotropical birds.” **Revista Brasileira de Ornitologia** 14, n. 1 (março): 7-13. 2006.

VOGLER, A. P., e MONAGHAN, M. Recent advances in DNA taxonomy. **J Zool Syst Evol Res**, 45 (1), 1-10. 2007.

WILSON, J., ROUGERIE, R., SCHONFELD, J., JANZEN, D., HALLWACHS, W., HAJIBABAEI, M., When species matches are unavailable are Barcodecorrectly assigned to higher taxa? An assessment using sphingid moths. **BMC Ecology**, pp. 11-18. 2011.

ANEXO