José Salvatore Leister Patané

Filogenia Molecular e Biogeografia das Espécies e
Subespécies do Gênero *Ramphastos* (Piciformes:
Ramphastidae)

São Paulo

2007

José Salvatore Leister Patané

Filogenia Molecular e Biogeografia das Espécies e
Subespécies do Gênero *Ramphastos* (Piciformes:
Ramphastidae)

Tese apresentada ao Instituto de
Biociências da Universidade de
São Paulo, para a obtenção de
Título de Doutor em Ciências, na
Área de Genética e Biologia
Evolutiva.

Orientadora: Profa. Dra. Anita
Wajntal

São Paulo

2007

Ficha Catalográfica

Patané, José Salvatore Leister

Filogenia Molecular e Biogeografia das
Espécies e Subespécies do Gênero *Ramphastos*
(Piciformes: Ramphastidae).

145 páginas

Tese (Doutorado) - Instituto de Biociências da
Universidade de São Paulo. Departamento de
Genética e Biologia Evolutiva.

1. *Ramphastos* 2. Filogenia 3. Evolução
molecular I. Universidade de São Paulo. Instituto
de Biociências. Departamento de Genética e
Biologia Evolutiva.

Comissão Julgadora:

Prof. Dr. Eduardo Eizirik

Profa. Dra. Elizabeth Höfling

Prof. Dr. Mercival R. Francisco

Prof. Dr. Sergio R. Mاتيoli

Profa. Dra. Anita Wajntal
Orientadora

Aos meus pais Dalva e Giuseppe,
sempre comigo.

“Vivendo e aprendendo...”
(Ditado popular)

Agradecimentos

À Profa Dra. Anita Wajntal pela orientação e pela confiança em mim depositada. Acima de tudo, lhe sou grato pela paciência e compreensão dedicada, pois estou certo de que sem isso eu não teria concluído a tese a tempo.

À Profa. Cristina Yumi Miyaki pela permissão de uso da infra-estrutura do Laboratório de Genética e Evolução Molecular de Aves ao longo dos anos em que desenvolvi este trabalho.

À FAPESP, pela concessão da bolsa de estudos, sem a qual este estudo não teria sido possível, e ao assessor anônimo pelos comentários e sugestões que ajudaram a direcionar o estudo.

Ao CNPq pelo suporte financeiro dedicado à infra-estrutura do Laboratório de Genética e Evolução Molecular de Aves.

À National Science Foundation pelo suporte financeiro aos laboratórios do Field Museum of Natural History, onde trabalhei por seis meses.

Aos colegas de laboratório Adriana Oliveira-Marques, Camila Ribas, Cibele Biondo, Érica S. Tavares, Erwin T. Grau, Fábio S. R. Amaral, Fernando d'Horta, Fernando Nodari, Flávia T. Presti, Gustavo S. Cabanne, Polyana Andrioni, Priscila F. Gonçalves, Renato K. Kimura, Rodrigo O. Pessoa, e Tania Matsumoto - pelo agradável convívio e importante apoio que me ofereceram sempre que necessário, além da troca diária de experiências e das discussões científicas tão fundamentais ao crescimento de qualquer profissional.

Ao Prof. Dr. Luís Fábio Silveira, do Departamento de Zoologia do Instituto de Biociências da USP, pelo auxílio com a obtenção de amostras de tecidos importantes no campo, bem como pelas valiosas sugestões e discussões que enriqueceram meu conhecimento sobre os tucanos e as aves em geral.

Ao Prof. Dr. Alexandre Aleixo, do Museu Paraense Emílio Goeldi, que coletou muitas das amostras de tecidos utilizadas neste trabalho. Nenhum estudo desse tipo seria completo sem a coleta e correta identificação de espécimes no campo.

Aos profissionais, colegas, e amigos próximos do Field Museum of Natural History, em Chicago, Illinois (EUA), que foram muito importantes para mim profissional e/ou

pessoalmente. Menciono especialmente Jason Weckstein, John Bates, Shannon Hackett, David Willard, Sushma Reddy, Steffen Pauls, Paul Velazco, Kevin Feldheim, Erin Sackett, Steffi Kautz e Lydia Smith.

Aos colegas de outros laboratórios (e grandes camaradas!) pela companhia no IB, pelas proveitosas discussões, e pelo apoio nos momentos difíceis de conclusão da tese, em especial Enéas Carvalho, Felipe F. Curcio, Frank Rheindt, Guilherme R. R. Brito, Lincoln S. Rocha, Marcos Figueiredo, Renata C. A. Ghilardi, Renata Moretti, Roberto Takata, Sabrina Baroni, Sergio L. Pereira, Talitha Pires... a lista poderia se estender mais e mais...

À Profa. Dra. Célia P. Koiffmann, coordenadora do programa do departamento de Genética e Biologia Evolutiva, e membro-titular do conselho departamental.

Ao pessoal da secretaria do departamento de Genética e Biologia Evolutiva; sem a ajuda deles, tudo teria sido muito mais difícil.

Aos funcionários da seção de pós-graduação do Instituto de Biociências da USP, que também me ajudaram bastante ao longo de todo o doutorado.

À Oxford University Press©, por me permitir utilizar e modificar algumas das figuras existentes no livro “Toucans, Barbets and Honeyguides” (autores: Lester Short & Jennifer Horne), publicado em 2001.

Aos meus familiares, meus pais em especial, por todo o apoio que me ofereceram durante todo esse tempo.

Muito obrigado a todos!

CAPÍTULO 1 - INTRODUÇÃO..... 3

1.1 Introdução geral 3

- 1.1.1 - O gênero *Ramphastos* Linnaeus, 1758 5
- Caracterização das espécies e subespécies do gênero 7

1.2 Considerações históricas e científicas sobre os métodos utilizados 13

- 1.2.1 - Sequenciamento de DNA 13
 - Polymerase Chain Reaction (PCR) 14
 - A técnica de sequenciamento 15
- 1.2.2 - DNA nuclear 16
- 1.2.3 - DNA mitocondrial 16
- 1.2.4 - Estrutura secundária de RNAs transportadores e ribossômicos 17
- 1.2.5 - Inferências filogenéticas 17
 - Máxima Parcimônia (MP) 18
 - Métodos probabilísticos 19
 - O uso de modelos evolutivos 19
 - Máxima Verossimilhança (MV) 20
 - Análise Bayesiana (AB) 23
 - Medidas de suporte 28
 - Bootstrap* paramétrico 30
- 1.2.6 - Datação de divergências 31
- 1.2.7 - Estudos biogeográficos 34
 - Biogeografia histórica e ecológica 34
 - Hipóteses para diversificação nos neotrópicos 35
 - Hipóteses ligadas a ciclos paleoclimáticos do Terciário/Quaternário 36
 - Hipóteses paleogeográficas 37
 - Hipóteses híbridas 37
 - Hipótese dos gradientes ecológicos 38
 - Hipótese da estabilidade climática 38

1.3 Objetivos 38

1.4 Material e Métodos..... 38

- 1.4.1 - Táxons estudados 38
- 1.4.2 - Regiões sequenciadas 40
- 1.4.3 - Extração de DNA 40
 - Protocolo I 40
 - Protocolo II 41
- 1.4.4 - Qualidade e concentração do DNA 41
- 1.4.5 - Amplificação dos fragmentos 41
- 1.4.6 - Purificação dos fragmentos 44
 - Protocolo I (Exo-SAP) 44
 - Protocolo II (Gelase) 44
- 1.4.7 - PCR para sequenciamento 44
- 1.4.8 - Purificação e precipitação para sequenciamento 45
 - Protocolo I (LGEMA) 45
 - Protocolo II (FMNH) 45
- 1.4.9 - Sequenciamento 45
 - Protocolo I (LGEMA) 45
 - Protocolo II (FMNH) 46
- Referências bibliográficas - Capítulo 1 46

CAPÍTULO 2 - ASPECTOS DA EVOLUÇÃO MOLECULAR DO GENE 12S NAS ESPÉCIES DO GÊNERO *RAMPHASTOS*..... 61

2.1 Introdução.....	63
2.2 Objetivos	65
2.3 Material e Métodos.....	65
2.4 Resultados	71
2.4.1 - Evitando artefatos de amplificação.....	71
2.4.2 - Estrutura primária.....	71
2.4.3 - Estrutura secundária	72
2.6 Discussão	82
2.6.1 - Estrutura primária do 12S.....	82
2.6.2 - Estrutura secundária do 12S	83
2.6.3 - Estrutura secundária e reconstruções filogenéticas	85
2.6.4 - Evolução molecular em alças e hastes.....	88
Referências bibliográficas - Capítulo 2	92

CAPÍTULO 3 - REAVALIAÇÃO FILOGENÉTICA DAS ESPÉCIES E SUBESPÉCIES DO GÊNERO *RAMPHASTOS* E CONSIDERAÇÕES BIOGEOGRÁFICAS 97

3.1 Introdução.....	99
3.2 Objetivos	102
3.3 Material e Métodos.....	102
3.4 Resultados	110
3.5 Conclusões.....	127
3.5.1 - Aspectos moleculares.....	127
3.5.2 - R. v. ariel SE x R. v. ariel N.....	128
3.5.3 - O clado R. vitellinus ssp. é monofilético ou parafilético?	129
3.5.4 - O posicionamento dos táxons R. dicolorus e R. sulfuratus	130
3.5.5 - R. v. citreolaemus.....	130
3.5.6 - Ausência de monofiletismo recíproco em subespécies de R. vitellinus ssp. e R. tucanus ssp.	131
3.5.7 - Digressões biogeográficas.....	131
3.5.8 - R. v. ariel: considerações taxonômicas.....	136
Referências bibliográficas - Capítulo 3	137

CAPÍTULO 4 - CONCLUSÕES..... 142

Referências bibliográficas - Capítulo 4	145
---	-----

Capítulo 1 - Introdução

1.1 Introdução geral

O presente trabalho versa sobre relações filogenéticas das espécies do gênero *Ramphastos*, representado pelos tucanos. Estas aves, que apresentam um bico grande e recurvado característico, estão distribuídas ao longo da região neotropical, do México à Argentina. São parentes próximos dos araçarís (gêneros *Aulacorhynchus*, *Andigena*, *Selenidera* e *Pteroglossus*; este último *sensu* Kimura *et al.*, 2004), formando com estes a família Ramphastidae. Esta família, juntamente com os capitães [famílias Megalaimidae, Lybiidae, Capitonidae e Semnornithidae; (*South American Checklist Committee*)], os pica-paus (família Picidae), e os “honeyguides” africanos (família Indicatoridae), formam a ordem Piciformes (*sensu* Short & Horne, 2001).

Os piciformes são aves corpulentas, com pescoço vigoroso e bico robusto. Os pés são zigodáctilos, com dois artelhos virados para a parte anterior (dedos II e III), e dois para a posterior (I e IV). Esta condição também ocorre nas ordens Psittaciformes (papagaios e araras), Cuculliformes incluindo Musophagidae (anus, cucos e afins; turacos) e Strigiformes (corujas) (Hafer, 1974; Short & Horne, 2001). No entanto, evidências indicam que o caráter é homoplástico, já que esses grupos aparentemente não formam uma unidade monofilética em relação a outras ordens dentro de Neoaves (van Tuinen *et al.*, 2000; Johansson *et al.*, 2001; Espinosa de los Monteros, 2000, 2003; Johansson & Ericson, 2003; Fain & Houde, 2004; Poe & Chubb, 2004; Ericson *et al.*, 2006; Gibb *et al.*, 2006). Ainda com relação à zigodactilia em Piciformes, a ordem dos tendões que servem os 4 artelhos do pé é diferente em relação às demais aves zigodáctilas (fig. 2 em Swierczewski & Raikow, 1981). Nos pés há, também, um processo articular desenvolvido na tróclea do artelho IV, ligando este ao tarso-metatarso.

A estrutura do palato também é diferenciada. Com relação ao sistema digestório, os adultos não possuem cecos intestinais (com exceção do capitonídeo *Megalaima franklinii*). São bons escaladores e empoleiram-se com facilidade; a prole nasce em cavidades

existentes em árvores, ou então em ninhos suspensos, como membros da família Indicatoridae (Swierczewski & Raikow, 1981; Short & Horne, 2001).

Dentro da ordem, diferentes trabalhos mostram que a família Indicatoridae é mais próxima de Picidae (pica-paus), enquanto capitães, araçaris e tucanos estariam mais estreitamente relacionados entre si (Sibley & Ahlquist, 1990; Johansson *et al.*, 2001; Johansson & Ericson, 2003; Pritchko & Moore, 2003; Fain & Houde, 2004). Os capitonídeos são pantropicais, abundantes na Ásia e na África, mas com relativamente poucas espécies nos neotrópicos (Sick, 1997). Possuem cerdas na base do bico, pescoço robusto, e cabeça relativamente grande (Sick, 1997; Short & Horne, 2001). Os capitonídeos do Novo Mundo apresentam cerdas relativamente menores e dimorfismo sexual relacionado à plumagem em todas as espécies.

Os tucanos e araçaris provavelmente dividiram um ancestral comum com os capitães do Novo Mundo (Burton, 1984; Prum, 1988; Sibley & Ahlquist, 1990; Lanyon & Hall, 1994; Barker & Lanyon, 2000; Johansson *et al.*, 2001, Johansson & Ericson, 2003; Moyle, 2004). Os capitães seriam, desta forma, parafiléticos (figura 1.1). Baseando-se em evidências filogenéticas, a tendência nomenclatural atual é admitir cinco famílias distintas, englobando capitães, araçaris e tucanos (figura 1.1). Tal separação é ratificada pela *South American Checklist Committee* (SACC). Segundo este comitê, tais famílias seriam: Megalaimidae (capitães da Ásia), Lybiidae (capitães da África), Capitonidae (capitães do Novo Mundo), Semnornithidae (um gênero apenas, *Semnornis*, de relações incertas quanto aos outros “capitonídeos” e tucanos) e Ramphastidae (araçaris e tucanos).

Figura 1.1 - Relações filogenéticas entre os capitães, tucanos e araçaris, de acordo com a literatura (ver página anterior). A família Ramphastidae é aqui considerada de acordo com a *South American Checklist Committee*, englobando os araçaris e tucanos.

As características que reúnem tucanos e araçaris na família Ramphastidae são: a presença de um bico relativamente longo e recurvado, colorido, duro e cortante, porém leve e poroso; nas laterais das regiões mediana e distal da maxila superior estão presentes protuberâncias cortantes, denominadas tômios; a língua é longa e filamentosa; em geral, não há dimorfismo sexual ligado à plumagem; o palato é peculiar, apresentando um vômer pequeno e truncado; dormem com a cauda projetada anteriormente (movimento possível graças à estrutura única das vértebras caudais), cobrindo o dorso e a cabeça; procriam em ninhos abandonados, buracos ou fendas existentes em árvores; alimentam-se de frutas, mas são predadores ativos ocasionalmente, alimentando-se de ovos e jovens de outras espécies de aves (Haffer, 1974; Sick, 1997; Höfling, 1991, 1998; Short & Horne, 2001).

1.1.1 - O gênero Ramphastos Linnaeus, 1758

Este gênero é representado pelos tucanos, que se diferenciam dos araçaris por várias características: as narinas ficam escondidas sob a base da mandíbula superior; a coloração da região dorsal, abdômen, cauda e coxa é predominantemente negra, em algumas espécies com tonalidade marrom do vértice ao dorso superior; a garganta e o peito apresentam coloração branca ou amarela (com matizes diferentes, dependendo da espécie); o peito inferior apresenta uma faixa vermelha de largura variável, podendo ser até ausente em alguns táxons; a primeira e segunda rêmiges primárias possuem extremidades prolongadas estreitas; a cauda é quadrada e relativamente pequena; as coberteiras inferiores da cauda são sempre vermelhas, com as superiores variando entre branco, vermelho, e diferentes tonalidades de amarelo; os membros posteriores (com exceção da coxa, coberta por plumagem) são em tonalidade azulada. Não há dimorfismo sexual entre os tucanos, exceto pelo tamanho geral e comprimento do bico (Haffer, 1974; Short & Horne, 2001).

A taxonomia de *Ramphastos* foi discutida por Haffer (1974), que reconheceu a existência de sete espécies. Estas, por sua vez, foram consideradas pelo autor como pertencentes a dois agrupamentos naturais, com base no tipo de vocalização e na morfologia dorsal da maxila superior: os *croakers* (com vocalização semelhante a um coxo, e presença de uma quilha ou canal longitudinal no dorso do bico, englobando as espécies *R. brevis*, *R. dicolorus*, *R. sulfuratus*, *R. toco* e *R. vitellinus*.), e os *yelpers* (de vocalização aguda, e parte dorsal da maxila superior arredondada - espécies *R. ambiguus* e *R. tucanus*). Dentre estas

sete espécies, apenas duas seriam monotípicas (*R. brevis* e *R. dicolorus*), as restantes sendo divididas em subespécies (Haffer, 1974; Short & Horne, 2001). O dendrograma com as relações de parentesco propostas por Haffer (1974) está mostrado na figura 1.2.

Figura 1.2 - Dendrograma com as relações de parentesco entre as espécies de *Ramphastos*, segundo a hipótese de Haffer (1974). A barra em vermelho indica os *croakers*, e a barra em azul, os *yelpers*.

Weckstein (2005), analisando um conjunto de três genes mitocondriais (citocromo-oxidase I, citocromo-b, e NADH-desidrogenase 2, num total de 2468 sítios de DNA), corroborou em alguns aspectos as inferências propostas por Haffer, mas também obteve incongruências: os *yelpers* formaram um clado monofilético, assim como seis das espécies propostas por Haffer também formaram; por outro lado, *R. vitellinus* (um *croaker*) seria parafilético (com a inserção do táxon *R. brevis* no grupo), o mesmo sendo verdadeiro para o componente filogenético dos *croakers* (já que *R. toco*, um dos *croakers*, se mostrou basal dentro do gênero). Uma das espécies (*R. vitellinus ariel*), que apresenta distribuições disjuntas na mata atlântica e na região de Belém, se mostrou parafilética, estando essas duas populações separadas na topologia final obtida (Weckstein, 2005): o táxon *R. v. ariel* do

sudeste da América do Sul seria basal às subespécies restantes de *R. vitellinus*, enquanto que a população de *R. v. ariel* de Belém seria mais próxima às subespécies de *R. vitellinus* da bacia amazônica.

Caracterização das espécies e subespécies do gênero

Haffer (1974) e Short & Horne (2001) são bastante congruentes quanto à subdivisão do gênero *Ramphastos*, reconhecendo basicamente o mesmo número de espécies e subespécies. A seguir são apresentados os táxons de acordo com Short & Horne (2001):

R. dicolorus - monotípico (sudeste da América do Sul): a vocalização é do tipo *croaker*. Possui o culmen do bico com canal alargado. O bico é amarelo-esverdeado, com exceção da base da maxila superior e da mandíbula, que são pretos, e da região dos tômios, que é avermelhada. A pálpebra é azulada. A região perioftálmica é vermelha. Não há tons marrons na plumagem dorsal. A garganta, o pescoço e o peito superior são alaranjados. Existe uma faixa estreita amarela entre o peito superior e o inferior. O peito inferior é vermelho. A parte superior do abdômen é vermelha. As supracaudais também são vermelhas.

Ramphastos dicolorus

© Oxford University Press, 2001

R. sulfuratus - 2 subespécies (norte da Colômbia e América Central): a vocalização é do tipo *croaker*. O culmen do bico possui uma quilha. O bico é predominantemente verde-amarelado, com faixa alaranjada na lateral inferior da maxila superior, e vermelho na ponta

do bico. A mandíbula apresenta porção azulada na parte imediatamente posterior à ponta do bico. A pálpebra e a região perioftálmica são verde-amareladas ou verde-azuladas. A fronte e a face dorsal do pescoço são marrons. A garganta, o pescoço e peito superior são amarelos de forma uniforme. Há uma eventual faixa tênue avermelhada entre o peito superior e o inferior. O peito inferior e o abdômen são pretos. As supracaudais são brancas. A subespécie *R. s. brevicarinatus*, que ocorre da Guatemala e Honduras até o norte da Colômbia, é um pouco menor e possui uma faixa vermelha, tênue mas relativamente mais conspícua, entre o peito superior e inferior, do que a subespécie *R. s. sulfuratus*, que ocorre até o México.

Ramphastos sulfuratus

© Oxford University Press, 2001

R. brevis - monotípico (noroeste da Colômbia e do Equador): a vocalização é do tipo *croaker*. O culmen do bico possui quilha. Há uma diagonal lateral marrom-escura ou preta no bico, cruzando a maxila superior a partir da base superior da mesma, até aproximadamente metade da extensão total da maxila superior. O restante da maxila superior é de coloração amarelada. A mandíbula é marrom-escura ou preta. A pálpebra e a região perioftálmica são verdes ou verde-amareladas. A fronte e a face dorsal do pescoço são marrons. A garganta, o pescoço e o peito superior possuem uma coloração amarelada uniforme, com uma faixa branca tênue, seguida de vermelha, entre o peito superior e o inferior. O peito inferior e o abdômen são pretos. As supracaudais são brancas.

Ramphastos brevis

© Oxford University Press, 2001

R. vitellinus - 4 subespécies (Amazônia, Maranhão e Piauí, e sudeste da América do Sul):

1. *R. v. vitellinus* (leste da Amazônia, margem norte do Rio Amazonas): a vocalização é do tipo *croaker*. O culmen do bico apresenta um canal longitudinalmente. O bico é todo preto com exceção da base, que é azul. A pálpebra e a região perioftálmica são azuis. Não há tons marrons na plumagem dorsal. A garganta é branca. O pescoço e peito superior apresentam gradiente crescente de amarelo-alaranjado em direção à parte inferior. Há a presença de uma faixa estreita branca entre o peito superior e o inferior. O peito inferior é vermelho. O abdômen é preto. As supracaudais são de coloração vermelha.

2. *R. v. ariel* (região de Belém, na margem sul do Rio Amazonas; Maranhão e Piauí; sudeste da América do Sul): a vocalização é do tipo *croaker*. O culmen do bico possui canal longitudinal. O bico é todo preto com exceção da base da maxila superior e da mandíbula, que são amarelos, apresentando uma faixa azul longitudinal estreita na maxila superior (na junção desta com a mandíbula), e triangular no culmen do bico. A pálpebra e a região perioftálmica são vermelhas. Há ausência de tons marrons na plumagem dorsal. A garganta, o pescoço e o peito superior apresentam coloração amarelo-escura. Existe uma faixa amarela estreita entre o peito superior e o inferior. O peito inferior é vermelho. O abdômen é preto. As supracaudais são vermelhas.

3. *R. v. culminatus* (oeste da Amazônia): a vocalização é do tipo *croaker*. O culmen do bico apresenta um canal longitudinal. No bico ocorrem faixas laterais pretas envolvendo quase toda a mandíbula e a maxila superior. A mandíbula é azul na parte proximal e amarela na

distal. As partes proximal, superior e distal da maxila superior são amarelas. A pálpebra é azul. A região perioftálmica é azul-clara. Não há tons marrons na plumagem dorsal. A garganta, o pescoço e o peito superior são brancos ou ligeiramente coloridos de amarelo. Existe uma faixa vermelha estreita entre o peito superior e o inferior. O peito inferior e o abdômen são pretos. As supracaudais são amarelas.

4. *R. v. citreolaemus* (noroeste da Colômbia e da Venezuela): a vocalização é tipo *croaker*. O culmen do bico tem um canal longitudinal. O bico apresenta faixas laterais pretas envolvendo parte da mandíbula e da maxila superior. A mandíbula e a maxila superior são amarelas na parte proximal, ambas apresentando coloração azul entre a parte proximal e a faixa preta lateral; a cor é amarela nas regiões distal e dorsal da maxila superior, assim como na distal da mandíbula. A pálpebra e a região perioftálmica são azuis. Há ausência de tons marrons na plumagem dorsal. A garganta e o pescoço são brancos. O peito superior apresenta um gradiente crescente de amarelo em direção à parte inferior. Há uma faixa vermelha estreita entre o peito superior e o inferior. O peito posterior e o abdômen são pretos. As supracaudais são amarelo-claras.

Ramphastos v. ariel

Ramphastos v. vitellinus

Ramphastos v. culminatus

Ramphastos v. citreolaemus

© Oxford University Press, 2001

R. toco - 2 subespécies (predominantemente áreas abertas da América do Sul): a vocalização é do tipo *croaker*. O culmen do bico possui quilha. O bico é laranja, sendo avermelhado tanto na região que engloba os tómos, como na quilha dorsal. A base da maxila superior e da mandíbula são de coloração preta. A parte distal da maxila superior

apresenta uma mancha preta característica. A pálpebra é azul-escura. A região perioftálmica é amarelo-alaranjada. A plumagem dorsal é preta. A garganta e o pescoço são brancos. O peito superior é branco, podendo apresentar tom levemente amarelado. Há indivíduos que apresentam uma faixa vermelha, bastante tênue, entre o peito superior e o inferior. O abdômen é preto. As supracaudais são brancas. As duas subespécies são bastante similares, sendo que em *R. t. albogularis* (que ocorre no centro e centro-sudeste da América do Sul) há uma tendência à presença de um bico ligeiramente menor, peito com ainda menos amarelo, e faixa vermelha ventral praticamente ausente, em comparação com *R. t. toco* (que ocorre nas Guianas e no nordeste do Brasil) - mas ainda assim com muita variação ocorrendo em ambas as subespécies.

R. tucanus - 3 subespécies (Amazônia):

1. *R. t. tucanus* (margem superior do Rio Amazonas a leste e região de Belém): a vocalização é do tipo *yelper*. O bico não possui quilha ou canal. A coloração é azul na base da mandíbula. A base da maxila superior apresenta a região lateral inferior de cor azul. A região lateral superior, o dorso e a parte distal da maxila superior são amarelos. Há faixas laterais avermelhadas cobrindo boa parte da maxila superior e da mandíbula; ambas apresentam uma faixa preta entre estas e a base do bico. A pálpebra e a região perioftálmica são azuis. Não ocorrem tons marrons na plumagem dorsal. A garganta é branca. O peito superior é branco, eventualmente colorido de amarelo. Há uma faixa vermelha entre o peito superior e o inferior. O peito posterior e o abdômen são pretos. As supracaudais são amarelas.

2. *R. t. cuvieri* (oeste da Amazônia): semelhante à anterior, exceto por apresentar tamanho do bico e do corpo maiores, supracaudais amarelo-escuros (podendo às vezes apresentar pontas avermelhadas), e faixa lateral do bico predominantemente preta, sem presença de tons vermelhos.

3. *R. t. inca* (leste do Peru e da Bolívia): plumagem e medidas morfológicas similares à subespécie *R. t. cuvieri*, exceto por apresentar supracaudais com tom de amarelo mais escuro e maior quantidade de pontas avermelhadas, e faixa lateral do bico preta com manchas de vermelho.

Ramphastos t. tucanus

Ramphastos t. cuvieri

© Oxford University Press, 2001

R. ambiguus - 3 subespécies (noroeste da América do Sul; América Central):

1. *R. a. ambiguus* (contorna as encostas do leste da cadeia dos Andes oriental): a vocalização é do tipo *yelper*. O culmen do bico não possui quilha ou canal. Há uma diagonal lateral marrom-escura cruzando a maxila superior a partir da base superior do bico até mais da metade da extensão total da maxila. O restante da maxila superior é amarela. A mandíbula é marrom. A pálpebra e a região perioftálmica são azuis. A fronte e a face dorsal do pescoço são marrons. A garganta, o pescoço e o peito superior são uniformemente amarelos. Há uma faixa branca estreita, seguida de uma vermelha, entre o peito superior e o inferior. O peito inferior e o abdômen são pretos. As supracaudais são brancas.

2. *R. a. abbreviatus* (extremo noroeste da América do Sul): semelhante à anterior, mas tamanho do corpo e do bico ligeiramente menores, e com pálpebra e região perioftálmica verde-amarelados.

3. *R. a. swainsonii* (região trans-andina da América do Sul a noroeste, e América Central): semelhante a *R. a. ambiguus*, mas com pálpebra e região perioftálmica verde-amarelados e diagonal da maxila superior, assim como da mandíbula, em tom marrom.

Ramphastos a. swainsonii

Ramphastos a. abbreviatus

Ramphastos a. ambiguus

© Oxford University Press, 2001

1.2 Considerações históricas e científicas sobre os métodos utilizados

1.2.1 - Seqüenciamento de DNA

O desenvolvimento de técnicas voltadas à determinação da seqüência dos aminoácidos que compõem um polipeptídeo deu-se entre as décadas de 40-60 do séc. XX (Chadarevian, 1999). No entanto, devido à degeneração do código genético (onde um aminoácido específico pode ser determinado por mais de um códon), o seqüenciamento de proteínas não revela de forma inequívoca a cadeia de DNA a partir da qual foi sintetizada (não extraindo, portanto, o máximo de informação genética possível que pode ser analisada).

O advento posterior de métodos de seqüenciamento de DNA (Maxam & Gilbert, 1977; Sanger *et al.*, 1977) permitiu a análise de genes ao nível dos seus nucleotídeos constituintes. Por esta razão, tal ferramenta tem sido utilizada em muitas áreas de pesquisa, tais como estudos genômicos em geral (mecanismos de regulação genômica, detecção de mutações causadoras de doenças, elaboração de sondas de hibridação específicas -

Rumjanek, 2001), evolução molecular (elucidação de processos geradores da variabilidade observada atualmente em seqüências nucleotídicas/polipeptídicas, estudos de novos alelos e *loci*, investigação de convergência e seleção - Graur & Li, 2000; Nei & Kumar, 2000; Felsenstein, 2003; Hedrick, 2005) e genética de populações (análise de variação geográfica, fluxo gênico, hibridação, variabilidade e estudos genéticos voltados à conservação - Felsenstein, 2003; Avise, 2004; Hedrick, 2005).

O principal uso do seqüenciamento para o presente estudo é a detecção e utilização de sítios de DNA informativos para reconstruções filogenéticas.

Polymerase Chain Reaction (PCR)

O passo crucial prévio para a realização do seqüenciamento é a presença de muitas cópias da região de DNA de interesse. Uma das maneiras de se obter essas réplicas é através da clonagem gênica, onde o fragmento a ser seqüenciado é ligado a um vetor de replicação autônoma em hospedeiros como células bacterianas ou leveduras, do qual ele pode ser posteriormente isolado (Thompson *et al.*, 1991). Como esses vetores podem atingir um número grande de réplicas por célula, e como bactérias e leveduras podem crescer indefinidamente em laboratório, uma grande quantidade de clones da molécula de interesse é gerada.

Outra forma de se obter uma grande quantidade de cópias (e que foi aqui utilizada) é através da amplificação por reação em cadeia da DNA polimerase (PCR, do inglês *Polymerase Chain Reaction* - Saiki *et al.*, 1985; Scharf *et al.*, 1986; Mullis & Faloona, 1987). Este método simula *in vitro* a replicação que ocorre naturalmente em células vivas; a partir de DNA genômico total, do uso de oligonucleotídeos iniciadores que flanqueiam a região de interesse (*primers*), da presença dos 4 desoxirribonucleotídeos, e de uma DNA polimerase especial capaz de resistir a altas temperaturas (*Taq* polimerase), são geradas cópias da região-alvo em ritmo exponencial, a partir de ciclos sucessivos de desnaturação do DNA, hibridação dos *primers*, e extensão da nova fita de DNA. Após 25-40 ciclos, são obtidas cópias em grande número, de maneira a ser possível a checagem do sucesso da reação pela presença de uma banda em um gel de agarose ou poliacrilamida submetido a um diferencial de potencial elétrico (Saiki *et al.*, 1985; Scharf *et al.*, 1986; Mullis & Faloona, 1987).

Em alguns casos, bandas extras podem ser encontradas; no caso do DNA nuclear, isso pode significar amplificação coincidente de um fragmento que não possui sequência semelhante e/ou ausência de relação funcional com o DNA alvo, a presença de um pseudogene (gene derivado da mesma sequência ancestral que deu origem ao DNA de real interesse, e que pode ou não ter perdido a função devido a um relaxamento da pressão seletiva, tornando-se sujeito a substituições ou *indels*), ou a detecção dos dois alelos de um heterozigoto para aquele *locus* (Graur & Li, 2000). Para o caso do DNA mitocondrial, valem as mesmas interpretações (Sorenson & Quinn, 1998; Bensasson *et al.*, 2001), com exceção da última, já que esta é uma cadeia haplóide - em compensação, um fenômeno conhecido por heteroplasmia, ou presença de mais de um genoma mitocondrial num mesmo indivíduo, é conhecida para um número pequeno de casos (Scheffler, 1999).

A técnica de seqüenciamento

Uma técnica de seqüenciamento amplamente usada é uma variação do método originalmente proposto por Sanger e col. (1977). Ela envolve o emprego de análogos químicos dos nucleotídeos (dideoxirribonucleotídeos, ddNTPs), os quais impedem o crescimento da cadeia de DNA em formação a partir do momento em que são incorporados à cadeia recém-sintetizada, e que são marcados por um tipo de fluorescência que emite luz numa faixa espectral específica quando excitados pelo feixe de laser de um seqüenciador, ou decodificada a partir da passagem por um capilar (estes correspondem a dois modos de leitura padrão em modelos de seqüenciadores automáticos). A cor da fluorescência desses nucleotídeos modificados é única para cada um dos 4 tipos possíveis (ddATP, ddCTP, ddGTP, ddTTP). Assim como na PCR, a síntese é desencadeada por um *primer*; a DNA polimerase então prossegue ao longo da cadeia, incorporando nucleotídeos normais, até que um dideoxirribonucleotídeo seja incorporado, parando o alongamento da cadeia imediatamente, e emitindo então a fluorescência que indica a qual dos 4 tipos ele pertence. Como a quantidade de cópias do DNA de interesse é muito alta na reação, é praticamente impossível que alguma posição específica da sequência não apresente um dideoxirribonucleotídeo marcado em algum ponto das cadeias em crescimento, ocorrendo, pelo contrário, a síntese de várias cadeias que param num mesmo ponto (isto sendo verdade para todas as posições). Desta forma obtém-se, ao final das reações de síntese, cadeias de

DNA com uma sequência de bandas coloridas separadas por tamanho, apresentando sempre a mesma cor para uma mesma posição. Atualmente, os próprios sequenciadores automáticos traduzem esta sequência de cores na sequência correta de bases de DNA (Sanger *et al.* 1977; Rumjanek, 2001).

1.2.2 - DNA nuclear

O tamanho do genoma nuclear (valor C) varia nos diferentes grupos de organismos. Valores C em eucariotos são em geral muito maiores que em procariotos, mas há exceções (Graur & Li, 2000). Entretanto, muitas vezes não há essa correlação entre complexidade estrutural e tamanho do genoma, o que é conhecido como paradoxo do valor C (Graur & Li, 2000). O DNA apresenta-se compactado na forma de um ou mais cromossomos, sendo que em eucariotos a maior parte do DNA não é transcrita ou codificada. Neste grupo de organismos, os pré-RNAs dos genes que codificam proteínas apresentam íntrons, sequências não traduzidas que se intercalam com aquelas que permanecem após edição do RNA, os éxons, que serão traduzidos em proteínas (Griffiths *et al.*, 1999; Graur & Li, 2000). Diferentes íntrons apresentam taxas diferentes de evolução, sendo alguns mais informativos para cladogêneses recentes, e outros que discriminam divergências mais remotas (Graur & Li, 2000).

1.2.3 - DNA mitocondrial

O DNA mitocondrial é uma molécula linear em muitas plantas e fungos, mas em metazoários é uma molécula circular, haplóide, de herança materna, que em geral não sofre recombinação (Scheffler, 1999; Avise, 2004). O fato de essa molécula ser haplóide e herdada uniparentalmente faz com que o tempo médio de coalescência dessa molécula como um todo seja quatro vezes superior quando comparada a genes nucleares (Avise, 2004). Há também evidências de que o sistema de reparo de DNA seja ausente ou defeitivo na mitocôndria (Scheffler, 1999), o que aumentaria ainda mais os danos ao DNA. Diferentes genes da molécula apresentam taxas de evolução distintas, refletindo assim as restrições específicas a que estão submetidos (Scheffler, 1999).

O genoma mitocondrial das aves é semelhante, quanto aos produtos sintetizados, aos outros vertebrados (Desjardins & Morais, 1990; Scheffler, 1999). Ele é responsável pela

síntese de 12 proteínas, 2 RNAs ribossômicos e 14 RNAs transportadores. Há pelo menos quatro arranjos gênicos diferentes nos grupos de aves existentes (Desjardins & Morais, 1990; Mindell *et al.*, 1998), sendo todos eles variantes em relação ao tipo encontrado em mamíferos e *Xenopus laevis* (Roe *et al.*, 1985).

É importante a utilização de marcadores que segregam de forma independente, como é o caso do uso em conjunto de seqüências mitocondriais e nucleares, para se testar se as topologias sugeridas por cada um são congruentes, e em caso negativo tentar entender quais processos estariam envolvidos na resolução diferencial entre ambos (Avise, 2004).

1.2.4 - Estrutura secundária de RNAs transportadores e ribossômicos

Muitos estudos filogenéticos se utilizam de moléculas de RNA, que possuem estrutura secundária decorrente de suas funções como arcabouço de ribossomos (RNAr) ou de ligação e transporte de aminoácidos à cadeia polipeptídica a ser formada durante a tradução (RNAt), acarretando em regiões de pareamento intra-molecular, e regiões de fita simples (Houde *et al.*, 1997). Isso acarreta dois grandes problemas: correlação de estados entre sítios nessas regiões de pareamento, o que viola a premissa de independência assumida por métodos de reconstrução filogenética; e ambigüidades no alinhamento (caso este não tenha sido baseado nessa estrutura secundária). Este último problema pode ser parcialmente resolvido com a comparação com modelos de estrutura secundária da mesma molécula em organismos filogeneticamente próximos (Houde *et al.*, 1997; Mindell *et al.*, 1997; Espinosa de los Monteros, 2003), já que a partir daí pode-se observar quais as regiões estruturais mais conservadas e usá-las como pontos fixos do alinhamento, sejam estas posições de fita simples ou dupla (Espinosa de los Monteros, 2003).

1.2.5 - Inferências filogenéticas

Priorizou-se aqui o uso de métodos que utilizam a informação a partir de cada caráter individualmente, sendo as sucessivas posições de um dado alinhamento consideradas diferentes caracteres (cada um com o mesmo conjunto de estados, A, C, G ou T/U - no caso de máxima parcimônia, eventos de inserção/deleção, ou *indels*, podem ser classificados como 5º estado), em detrimento a métodos de reconstrução baseados em matrizes de distâncias ou outros procedimentos. As análises mais embasadas segundo esse

critério adotado, seja em pilares filosóficos ou estatísticos, são a máxima parcimônia (MP), máxima verossimilhança (MV) e análise bayesiana (AB).

Máxima Parcimônia (MP)

A MP tem suas origens nos conceitos de Hennig publicados em alemão (Hennig, 1950), posteriormente ampliados e traduzidos para o inglês (Hennig, 1966), e também numa sugestão de Edwards e Cavalli-Sforza (1963) acerca das propriedades que um critério filogenético analítico deva possuir. O conceito fundamental de Hennig é que táxons que compartilham estados derivados (apomórficos) devem ser descendentes de um ancestral comum que apresentava estados mais primitivos (plesiomorfias), que teriam sofrido mutações até adquirirem os estados atuais nos descendentes.

Em linhas gerais, o princípio de máxima parcimônia no contexto filogenético estabelece que o número de mudanças de estados numa dada filogenia deve ser o menor possível dada a observação dos estados atuais nos táxons analisados. Assim sendo, a MP pode ser interpretada como um tipo de derivação, dentro do ramo da biologia evolutiva, do princípio filosófico da navalha de Ockham, formulado pelo filósofo inglês William de Ockham (Graur & Li, 2000). Tal princípio advoga que a explicação mais provável para um determinado conjunto de observações deve ser a mais simples, envolvendo o menor número possível de premissas. Edwards e Cavalli-Sforza (1963) também sugeriram que a melhor estimativa de relações de parentesco deve ser aquela que envolva a menor quantidade total de evolução.

Camin & Sokal (1965) foram os primeiros a atribuir a palavra “parcimônia” a um método de reconstrução filogenética baseada em estados discretos. A primeira aplicação do método para seqüências moleculares foi com o uso de proteínas (Eck & Dayhoff, 1966); posteriormente (Fitch, 1977) a MP passou a ser inferida também a partir de nucleotídeos.

De acordo com este método, para cada topologia possível relacionada a um determinado número de táxons, a evolução mais parcimoniosa para cada caráter é inferida através de um algoritmo que minimiza o número de passos necessários para explicar os estados observados nos ramos terminais (Fitch, 1971), e este processo é então realizado para todas as posições do alinhamento. Ao final, a árvore com menor número de passos dos caracteres como um todo é escolhida como árvore mais parcimoniosa. Este método tem

premissas simples, é intuitivo e é computacionalmente mais rápido que métodos baseados em probabilidades (descritos a seguir). Entretanto, a reconstrução topológica por MP pode não representar de maneira adequada o modo de evolução de seqüências: o método assume implicitamente que a taxa de evolução é homogênea ao longo de toda a seqüência; ramos longos (ou seja, porções da árvore que acumularam um número de substituições maior que outros ramos) podem se atrair uns aos outros devido a paralelismos e convergências (Felsenstein, 1978; Hendy & Penny, 1989; Bergsten, 2005); similaridade de composição nucleotídica em algumas das seqüências analisadas pode fazer com estas se atraiam; o número de substituições inferidas é sempre uma subestimativa do valor real para seqüências que divergiram há um tempo relativamente longo (Jukes & Cantor, 1969; Kimura, 1980); a inferência do número de substituições em cada ramo da árvore pode ser ambígua em muitos casos (Felsenstein, 2003); e o método não oferece tratabilidade estatística, ou seja, não há uma maneira direta de se obter esperanças matemáticas e variâncias a partir do número mínimo de passos (Nei & Kumar, 2000; Felsenstein, 2003).

Métodos probabilísticos

O uso de modelos evolutivos

Como mostrado acima, uma das características da reconstrução filogenética por MP é a subestimação do número de passos evolutivos, sendo a magnitude do desvio proporcional ao tempo de separação entre duas linhagens, o que pode acarretar em problemas não só quanto ao cálculo de tamanhos de ramo, como também em relação à própria topologia em si (Felsenstein, 2003). Jukes e Cantor (1969) desenvolveram um modelo envolvendo correção de Poisson para contornar esse problema. Entretanto, esse modelo não engloba uma série de premissas provenientes das observações de padrões de evolução molecular, como por exemplo diferenças nas taxas entre transições (mudanças entre purinas - A e G - ou entre pirimidinas - C e T) e transversões (transformações entre uma purina e uma pirimidina). Tais diferenças foram abordadas no modelo proposto por Kimura (1980), onde há um parâmetro que define a taxa de transição, e outro a de transversão. Outro pressuposto modelado posteriormente (Felsenstein, 1981; Hasegawa *et al.*, 1985) foi a quantidade diferencial de cada uma das bases, que até então eram assumidas como tendo freqüências idênticas e constantes ao longo da árvore. Com o tempo, modelos

com cada vez mais parâmetros foram sendo descritos (Liò & Goldman, 1998; Felsenstein, 2003), sendo inclusive abordado o uso de funções que contabilizam a variação de taxas ao longo da cadeia nucleotídica (Golding, 1983; Nei & Gojobori, 1986; Jin & Nei, 1990; Yang, 1994b; Yang & Kumar, 1996).

Uma forma mais apropriada de se tratar a evolução de um grupo de seqüências, computacionalmente falando, é através do uso de matrizes que englobam todos os parâmetros mencionados acima. Um pressuposto na utilização dessas matrizes é a propriedade de Markov, em que a chance de um nucleotídeo assumir um determinado estado no futuro depende apenas do seu estado atual, e portanto elas são mais apropriadamente denominadas matrizes de Markov (Liò & Goldman, 1998). Usualmente, essas matrizes assumem que o processo evolutivo é homogêneo (as taxas dos diferentes tipos de substituições são constantes em todos os ramos de uma dada topologia e através do tempo), estacionário (as frequências de cada base estão no equilíbrio decorrente do processo homogêneo subjacente) e reversível (a chance de uma determinada base mudar para outra é a mesma que no sentido oposto) - Liò e Goldman, 1998; Jayaswal e col., 2005. Irreversibilidade, assim como outros fatores, pode ser implementada nesse modelo geral (Barry & Hartigan, 1987; Jayaswal *et al.*, 2005), embora existam complicações quanto à tratabilidade matemática (Felsenstein, 2003).

Outro pressuposto assumido por essa classe de matrizes (e portanto extendidos aos métodos de reconstrução filogenética que se utilizam desses modelos) é o de que a evolução de cada posição nucleotídica é independente das outras, e que todos os caracteres são provenientes de uma mesma distribuição subjacente (Felsenstein, 1981, 2003), suposições estas que podem também ser relaxadas através do uso de um maior número de parâmetros na modelização markoviana.

Máxima Verossimilhança (MV)

A MV é uma adaptação proposta por Edwards & Cavalli-Sforza (1963) aplicada ao campo da filogenética molecular, de um método estatístico geral desenvolvido por Fisher (1922, 1956); estes autores utilizaram o princípio da MV para calcular uma árvore a partir de dados de frequência gênica. Felsenstein (1981) posteriormente adaptou o método para estimativas filogenéticas baseadas em sítios individuais, dado um alinhamento prévio. De

forma geral, quando se realiza algum tipo de amostragem de dados (D) para se testar alguma hipótese (H), o que se pretende descobrir é a probabilidade daquela hipótese em face dos dados obtidos ($P[H|D]$). De acordo com o teorema de Bayes, esse valor é proporcional ao produto das probabilidades $P[D|H]$ e $P[H]$; esta última é a probabilidade *a priori* da hipótese. Portanto, ao termos em mãos estes dois últimos valores, podemos estimar um valor proporcional à probabilidade posterior $P[H|D]$. Assumir probabilidades *a priori* para hipóteses, no entanto, é motivo de controvérsias (Fisher, 1922; Felsenstein, 2003); mas à medida que a amostragem de dados aumenta, a probabilidade *a priori* tende a perder peso na equação, tornando a correlação entre $P[H|D]$ e $P[D|H]$ menos dependente do valor de $P[H]$. A quantidade $P[D|H]$ é denominada de verossimilhança da hipótese H. Apesar da correlação, é importante salientar que $P[D|H]$ não equivale à probabilidade da hipótese em face dos dados ($P[H|D]$), ou probabilidade posterior da hipótese; ela é na verdade a probabilidade dos dados, dada a hipótese. A alcunha de verossimilhança é importante para ressaltar que esta grandeza não equivale a um valor de probabilidade propriamente dito (Fisher, 1922). A razão para essa diferenciação decorre do seguinte aspecto: quando se varia os valores dos parâmetros da hipótese a ser testada, tem-se um gráfico de valores de probabilidade para essas diferentes quantidades; esse gráfico, no entanto, não representa uma distribuição, já que nas distribuições estatísticas o que varia são justamente os dados, e estima-se a probabilidade de se obtê-los fixando-se os valores dos parâmetros da distribuição; como mencionado, na verossimilhança ocorre o inverso: o conjunto de dados é fixo, sendo plotadas as probabilidades do mesmo para diferentes valores dos parâmetros da hipótese; portanto esse gráfico não indica as probabilidades de se obter dados mutuamente exclusivos; assim sendo, a área sob a curva também não precisa ser igual a 1 (como ocorre com distribuições probabilísticas) - Fisher, 1922; Felsenstein, 2003.

O ponto de máxima verossimilhança é simplesmente o ponto de máximo global desse gráfico (Felsenstein, 1981). No caso de reconstruções filogenéticas, $P[D|H]$ é traduzido como $P[\text{alinhamento} | t, v_i, s(p_1, p_2, \dots, p_n)]$, onde a “hipótese” é desmembrada em 3 variáveis: t representa a topologia, v_i são os tamanhos de ramo, s é o modelo de evolução, e p_j são os parâmetros desse modelo (para um total de N parâmetros). Para cada árvore, os valores de t , v_i e p_j que em conjunto apresentarem o maior valor de verossimilhança para o

alinhamento obtido, serão os valores de máxima verossimilhança para essa hipótese (Felsenstein 1981).

A verossimilhança de uma dada topologia é calculada da seguinte maneira: para cada posição do alinhamento, calcula-se as probabilidades de cada um dos 4 nucleotídeos (ou estados) ocuparem os nós internos, multiplicada pela probabilidade de cada uma dessas bases sofrer mutação através de um ramo com determinado comprimento e gerar cada um dos 4 estados na outra ponta do ramo (sempre assumindo reversibilidade); esse passo é repetido sucessivamente através de todos os ramos da filogenia, até que se atinja um nucleotídeo observado de um ramo terminal, e esse procedimento é realizado até que todas as probabilidades de transição markovianas, para todos os ramos internos e externos, tenham sido calculadas. Esse algoritmo é repetido para todas as posições do alinhamento, assumindo-se os pressupostos de independência e distribuição idênticas em cada sítio, e para todos os ramos da topologia (embora essas premissas possam ser relaxadas, ver acima).

Os logaritmos das verossimilhanças dos sítios individuais são então somados para uma mesma árvore (pois a multiplicação dos valores brutos leva a números cada vez mais próximos de zero, acarretando em erros de arredondamento derivados da limitação física de estocagem dos resultados parciais, que ocupam sempre um tamanho fixo de *bits* para cada tipo de computador - Huelsenbeck & Ronquist, 2005). Os tamanhos de ramo são otimizados para cada árvore examinada. O processo, idealmente, deve ser reiterado algumas vezes, até que haja estabilização dos tamanhos de ramo na árvore com um todo (Felsenstein, 1981). Depois, todo o processo é reiniciado para outras topologias, e dentre elas, se escolhe a de MV. Os valores das probabilidades de transição markovianas relativas ao modelo escolhido podem ser estipuladas de antemão ou estimadas por máxima verossimilhança em conjunto com a otimização de tamanhos de ramos (Felsenstein, 1981; Swofford, 2002).

Um grande atrativo que a MV incorpora é, como mencionado, o uso de modelos evolutivos. Por incorporar diferentes parâmetros explicitamente, esse método pode refletir uma aproximação do verdadeiro processo molecular gerador do padrão filogenético a ser estimado - mas isto apenas se o modelo for uma generalização razoável do processo que de fato ocorreu na natureza, durante a evolução do grupo (Liò & Goldman, 1998).

Do ponto de vista das aplicações estatísticas em geral, estimações usando o princípio da máxima verossimilhança são consistentes (tendem ao valor verdadeiro da população à

medida que mais dados são incorporados), robustos (violações do modelo podem não afetar sensivelmente a grandeza estimada) e eficientes (possuem a menor variância possível em torno do valor correto do parâmetro) - Fisher, 1956. Sob determinadas condições, está demonstrado que essas mesmas características se aplicam à MV enquanto estimador de árvores filogenéticas (Felsenstein, 2003).

Entretanto, as premissas para o uso de MV em filogenética nem sempre são garantidas (sítios evoluindo independentemente, seguindo uma mesma distribuição, e uso de um modelo minimamente adequado), o que faz com que o sucesso na reconstrução topológica e de tamanhos de ramo não seja garantida (Felsenstein, 2003; Jayaswal *et al.*, 2005). Outro porém da técnica é a parte computacional: no caso da MP, para efeito de comparação, calcula-se para cada topologia o menor número de passos; depois, dentre todas as topologias, é escolhida aquela de MP. Por outro lado, para MV, não há um algoritmo que calcule diretamente os melhores tamanhos de ramo para cada árvore, sendo este um processo em si mais demorado que o simples cálculo do número de passos em MP, além de não ser exato - ademais, para cada nó interno, todas as 4 bases são testadas como possíveis estados ancestrais para o cálculo da verossimilhança total, sendo isto feito para cada sítio (Felsenstein, 1981).

Análise Bayesiana (AB)

A análise bayesiana busca a probabilidade da hipótese, dadas as verossimilhanças das diferentes árvores e suas respectivas probabilidades *a priori*. É de fato a estimativa que se deseja obter ao analisar-se os dados ($= P[H|D]$; ver acima). Há, entretanto, muita discussão quanto ao uso de probabilidades *a priori* (Fisher, 1922; Holder & Lewis, 2003; Beaumont & Rannala, 2004), embora o peso destas seja cada vez menor à medida que a quantidade de dados aumenta (ver acima).

A implementação da análise bayesiana se dá, na prática, pelo método denominado *Markov-chain Monte Carlo* (MCMC), baseada no trabalho de Metropolis e col. (1953), e mais tarde modificada por Hastings (1970). O algoritmo Metropolis-Hastings, aplicado à análise bayesiana, usa cadeias de Markov como mecanismo de ligação entre o estado atual e novos estados propostos (estado = topologia, tamanhos de ramos e parâmetros do modelo), sendo os valores de probabilidades *a priori* ($P[H]$) desses parâmetros obtidos por números

aleatórios sorteados a partir de distribuições previamente estipuladas. Os valores de $P[D|H]$ x $P[H]$ para a nova árvore proposta, e para a atual, são calculados (já que estes são proporcionais às respectivas probabilidades posteriores), e uma razão entre as duas é obtida; se o numerador (probabilidade posterior da nova árvore) for maior que o denominador (árvore atual), o passo é aceito, e então outro estado será proposto a esta nova árvore; caso o valor do denominador seja maior, é sorteado um número aleatório entre 0 e 1, a partir da função geradora de probabilidades *a priori* para as topologias (que pode ser uma distribuição uniforme, por exemplo, ou qualquer outra). Se este valor for menor ou igual ao valor da razão, a nova árvore (de menor probabilidade) é aceita; caso contrário, a cadeia permanece no estado em que está, e um novo estado será então proposto.

Dessa forma, a chance de uma árvore proposta ser aceita é proporcional à diferença de probabilidade entre as duas; se uma determinada árvore tem probabilidade alta, ela tenderá a ser visitada mais vezes, e topologias com baixa probabilidade serão visitadas menos; isto indica que as probabilidades posteriores das topologias podem ser estimadas a partir da frequência com que tais estados são aceitos (Felsenstein, 2003; Huelsenbeck & Ronquist, 2005). Entretanto, para que essa aproximação seja aceitável, a amostragem dos estados deve ser realizada na região de equilíbrio estacionário (ou seja, na região do espaço amostral multi-paramétrico onde a chance de se chegar a um determinado estado a partir de um outro estado específico seja a mesma que no sentido oposto), mas não há uma maneira de se dizer com certeza se o ponto de equilíbrio em questão é um ótimo global ou uma região sub-ótima (Felsenstein, 2003; Holder & Lewis, 2003; Beaumont & Rannala, 2004; Huelsenbeck & Ronquist, 2005).

Para aumentar a chance de se chegar à verdadeira região estacionária, dois tipos de cadeias de Markov são usadas concomitantemente: uma cadeia “fria” e uma (ou mais) cadeia(s) “quente(s)”. A primeira tende a aceitar estados não muito diferentes de geração para geração, o que gera uma tendência dessa cadeia a ficar isolada em picos de probabilidade posterior, com pouca chance de cruzar os vales; já as cadeias “quentes” tendem a aceitar estados relativamente mais alterados, o que permite “saltos” no espaço paramétrico multi-dimensional, facilitando assim a busca em locais mais longínquos do mesmo (isso é feito elevando-se a razão entre a árvore proposta e a atual, a um número entre

0 e 1; desta maneira, a razão passa a ter um valor maior do que se não houvesse a exponenciação, e portanto a chance de se aceitar um estado pior aumenta).

Existe, a cada geração, uma probabilidade de troca de estados entre os dois tipos de cadeia de Markov. Assim sendo, de tempos em tempos a cadeia fria aceita o estado atual da cadeia quente, saltando para um outro ponto, a seguir estabilizando-se novamente como cadeia fria, em uma outra área do espaço paramétrico (Larget & Simon, 1999; Archibald *et al.*, 2003; Altekari *et al.*, 2004).

As amostragens de estados são sempre da cadeia “fria”, que é mais estável e tende a chegar a um pico de probabilidades mais normalmente que as cadeias “quentes”, ali permanecendo - além do mais, é só a partir dela que se pode atingir equilíbrio estacionário após um longo número de gerações, pois a razão entre as probabilidades dos estados proposto e atual não é alterada pelas exponenciações que ocorrem nas cadeias quentes. As amostragens da cadeia fria ocorrem em intervalos (por exemplo, a cada 1.000 iterações da cadeia), para que o problema da correlação entre os estados sucessivos da cadeia seja minimizada (já que estados propostos são em geral o resultado de algum tipo de perturbação dos parâmetros do estado anterior). Uma maneira ainda mais garantida de se evitar máximos locais é a implementação de diferentes cadeias em paralelo, cada uma começando a partir um estado diferente (Huelsenbeck & Ronquist, 2005).

A partir da região de equilíbrio, pode-se gerar um intervalo (ex: 95%) de interesse para as variáveis contínuas (taxas de transição e transversão, frequências evolutivas, tamanhos de ramo, heterogeneidade de taxas ao longo da cadeia de DNA, etc.), e um consenso de maioria relativo às partições topológicas. Abandonando-se o pressuposto de homogeneidade, a MCMC pode inclusive estimar taxas de evolução diferentes para cada ramo (Drummond *et al.*, 2006), como mencionado na seção “O uso de modelos evolutivos”, acima. Para a geração dessas estatísticas descritivas, descarta-se a série inicial de estados que ainda não teriam alcançado a fase estacionária (período de *burn-in*).

Huelsenbeck e col. (2001) e Huelsenbeck e Ronquist (2005) acreditam que a frequência de amostragem de uma dada árvore analisada por MCMC seja realmente equivalente à sua probabilidade posterior, e que portanto esse valor representa a chance daquela topologia ser correta. Entretanto, outros trabalhos mostram que os valores de suporte bayesiano para ramos individuais são superiores aos gerados por outros índices,

sendo a magnitude da superestimação maior que as subestimativas observadas para outros índices, como *bootstrap* (Suzuki *et al.*, 2002; Simmons *et al.*, 2004). Assim, deve-se encarar com cautela estimativas de probabilidade posterior, ao menos com relação ao suporte de ramos.

É arriscado dizer de antemão qual dos três métodos acima descritos estima com maior precisão a topologia baseada num conjunto de dados; Huelsenbeck (1995) comparou o desempenho de diferentes métodos filogenéticos analisando simulações em uma topologia não-enraizada de quatro táxons, onde os tamanhos de ramos variavam consideravelmente. Os resultados mostraram que MV foi superior a outros métodos testados, inclusive MP (AB não foi incluída no estudo), mas que todos eles resgataram a topologia correta em 95% das vezes. Outros estudos (Hillis *et al.*, 1992; Sanson *et al.*, 2002) mostraram, através da evolução *in vitro* de seqüências nucleotídicas, que para divergências relativamente baixas, ambos os métodos inferem a verdadeira topologia com alta precisão.

O fato de alinhamentos em geral envolverem a presença de *indels* em maior ou menor grau torna importante a consideração de posições do alinhamento que apresentem lacunas (*gaps*) para alguns táxons (Giribet & Wheeler, 1999). A partir do momento em que um alinhamento é inferido como sendo o mais correto, pode-se admitir que os *gaps* presentes sejam resultado do processo evolutivo que gerou as seqüências observadas. Partindo desse pressuposto, *gaps* podem constituir uma fonte valiosa de informação filogenética; métodos que não os incluam nas análises podem vir a reconstituir a árvore filogenética com menor fidelidade, já que eles deixam de lado parte da informação histórica (Giribet & Wheeler, 1999). A MP utiliza mais informação, sob esse aspecto, que MV ou BA, pois implementações destes dois últimos métodos se restringem apenas ao uso dos 4 estados usuais (A, C, G e T), devendo-se optar, nesses métodos, entre a exclusão das posições nucleotídicas em que apareça um *gap* para todos os táxons, ou então considerar a presença de *gaps* como “missing data” (ou seja, eles são interpretados como falta de resolução na determinação da base); assim sendo, a MP pode ser o método de reconstrução mais eficiente em alguns casos, justamente devido à inclusão de *gaps* na análise (Giribet & Wheeler, 1999).

Quando há variação de taxas ao longo da molécula, a reconstrução por MV através de modelos que ignoram essa variação se mostra inconsistente, sendo que a MP nestes casos tende a resgatar a verdadeira topologia com maior probabilidade (Yang, 1996, 1997; Siddall, 1998). Este é um caso particular dentre outros mostrando que premissas “erradas”, que subestimam o número total de substituições (nesse caso, MP), tendem a resgatar a árvore verdadeira com maior probabilidade que com o uso de modelos mais adequados (Yang, 1997; Pol & Siddall, 2001); isso pode acontecer quando ramos longos são realmente vizinhos na árvore verdadeira (zona de Farris - Siddall, 1998), e nesse caso o comportamento da parcimônia de atrair ramos longos faz com que este método resgate a topologia com maior êxito do que a MV.

Por outro lado, quando na árvore verdadeira os ramos longos são separados por ramos curtos (zona de Felsenstein - Huelsenbeck & Hillis, 1993), essa mesma tendência da MP tenderá a inferir a árvore errada, enquanto que a MV com modelos adequados reconstrói a topologia verdadeira com maior eficiência (Felsenstein, 2003). Em conjuntos de dados onde as frequências nucleotídicas não são estacionárias, não há homogeneidade de taxas e o princípio de reversibilidade é violado, a MV apresenta dificuldade em estimar a árvore correta (Jayaswal *et al.*, 2005). Um trabalho recente (Philippe *et al.*, 2005) mostra que na presença de heterotaquia (termo que se refere à variação das taxas sítio-específicas no tempo), a eficácia da MV diminui, independentemente do grau de variação entre as linhagens; por outro lado, a precisão da MP pode aumentar ou diminuir quando os níveis de heterotaquia aumentam, dependendo dos tamanhos relativos dos ramos (Philippe *et al.*, 2005).

No caso de AB, como mencionado acima, se o conjunto de dados for relativamente grande, as probabilidades *a priori* deixam de ter tanta importância para a estimativa da probabilidade posterior da árvore; nesse caso, é de se esperar a convergência de AB e MV para uma mesma topologia (Felsenstein, 2003). A vantagem do uso de AB sobre a MV, além do fato da primeira gerar ao mesmo tempo uma topologia e índices de suporte de ramos, é que também são indicados os valores obtidos de todos os parâmetros (incluindo tamanhos de ramos), em todas as árvores da região de equilíbrio, o que permite a descrição de intervalos de confiança para os mesmos. Mas além da controvérsia do uso de probabilidades *a priori* (que, como já mencionado, pode não representar efetivamente um

problema), há discussão quanto à interpretação do índice de suporte de ramos por AB, conforme abordado acima.

O ponto a ser frisado aqui, após esse exame de alguns dos problemas existentes, é que não se deve assumir que qualquer um desses métodos seja superior a todos os outros sob quaisquer condições; e mesmo conhecendo o comportamento diferencial desses métodos sob determinadas circunstâncias, não há como saber por antecipação se os dados coletados estão sob influência dessas condições; por conseguinte, não há como antever qual método filogenético se sairá melhor numa dada análise.

Medidas de suporte

No caso de MP e MV, foi utilizado o *bootstrap* não-paramétrico (Efron, 1979) como forma de medida de suporte dos diferentes ramos. Tal procedimento é uma alternativa estatística, baseada em re-amostragem de dados, para o cálculo de valores de parâmetros, realização de testes de hipótese, estimativa de intervalos e/ou inferência, quando não se conhece a distribuição subjacente aos dados, ou quando o tratamento matemático do(s) parâmetro(s) não é trivial (Efron, 1979; Manly, 1997).

O método se baseia no princípio de que a distribuição de médias amostrais tende ao valor da verdadeira média populacional com maior precisão (ou seja, com intervalo de confiança mais reduzido) que a estimativa a partir de uma amostra única (Efron, 1979; Sokal & Rohlf, 1995). A justificativa para o uso deste método é a existência de eventuais empecilhos no levantamento de novas amostragens. As premissas necessárias são que o conjunto de dados em mãos seja representativo da variabilidade populacional real, e que esses dados sejam independentes e identicamente distribuídos.

O procedimento envolve o sorteio com re-amostragem do conjunto de dados inicial, formando um novo conjunto com o mesmo número de elementos do original; o processo é repetido centenas ou milhares de vezes, e para cada conjunto de dados re-amostrados (ou pseudo-réplicas) o parâmetro de interesse é calculado. Ao final das análises de todas as pseudo-réplicas, um histograma de frequências das estimativas do parâmetro pode então ser gerado, apresentando magnitude e níveis de variação próximos a aqueles que seriam obtidos caso novos conjuntos de dados fossem amostrados a partir de uma suposta distribuição subjacente (Efron, 1979).

No caso de inferências filogenéticas, cada pseudo-réplica tem o mesmo número de caracteres da matriz de dados sequenciados, com cada coluna do alinhamento original sendo sorteada com possibilidade de re-amostragem (supondo que os táxons estejam na horizontal, e os caracteres na vertical da matriz); centenas a milhares de pseudo-réplicas podem ser assim geradas. Para cada uma delas, a melhor topologia de acordo com um mesmo critério de otimização é inferida, e para cada ramo existente nesta árvore é atribuído o valor 1, ou 0 caso o referido ramo não esteja presente. Ao final da análise de todas as pseudo-réplicas, uma árvore de consenso de maioria, considerando a presença (ou não) de tal ramo no conjunto total de topologias examinadas, é computada, e este valor pode então ser usado como medida de congruência interna dos dados; este índice é assim calculado para todos os ramos da árvore (Felsenstein, 1985).

O valor do índice de *bootstrap* não deve ser interpretado como equivalente ao *p-value* estatístico comumente utilizado, já que são testadas várias hipóteses ao mesmo tempo (cada hipótese = um ramo diferente), e portanto incorreria-se no problema do teste múltiplo, perdendo-se o poder da inferência real de cada ramo da filogenia individualmente (Felsenstein, 1985). Outro problema do *bootstrap* não-paramétrico é que ele é incapaz de reduzir eventuais enviesamentos dos dados (Efron, 1979; Manly, 1997). No caso de ramos longos em MP, por exemplo, em que a inclusão de mais pares de bases não soluciona o problema de inconsistência, os índices de *bootstrap* simplesmente refletiriam esse viés (Swofford *et al.*, 1996); no entanto, pode-se argumentar que nesse caso o viés é devido aos dados, e não ao método. Hillis e Bull (1993), em estudos de simulação, obtiveram que valores de *bootstrap*, sob certas circunstâncias, tendem a ser medidas conservativas de acurácia: proporções $\geq 70\%$ em geral corresponderam ao clado “real” (com probabilidade $\geq 95\%$). No entanto, tal resultado deve ser interpretado com cautela, pois as condições de simulação empregadas não foram muito realistas - os próprios autores mostraram que quando estas não eram atendidas, valores de *bootstrap* $\geq 50\%$ poderiam ser sobrestimativas de acurácia; assim sendo, índices de *bootstrap* não refletiriam acurácia topológica (Hillis & Bull, 1993).

Por outro lado, aplicações do *bootstrap* a alguns conjuntos de dados reais mostraram que, para clados bastante corroborados, os índices de *bootstrap* seriam subestimativas tanto de acurácia quanto de repetibilidade (Sanderson & Wojciechowski, 2000). De forma geral,

portanto, medidas de *bootstrap* não devem ser encaradas como índices estatísticos triviais. Sanderson (1989) menciona que mesmo que a interpretação estatística do *bootstrap* seja complexa, o método ainda é útil pois estabelece uma maneira sistemática de medir a robustez dos dados quando estes são perturbados. O *bootstrap* pode ser aplicado tanto a reconstruções topológicas por MP quanto por MV.

No caso da AB, como já mencionado, a medida de suporte utilizada foi um consenso de maioria com as árvores da região de equilíbrio da MCMC.

***Bootstrap* paramétrico**

Um dos métodos mais justificados, tanto do ponto de vista biológico quanto estatístico, para comparar hipóteses distintas quanto à evolução molecular, é o *bootstrap* paramétrico (Huelsenbeck & Crandall, 1997; Huelsenbeck & Rannala, 1997). O método tem esse nome pois diversos conjuntos de alinhamentos são gerados, baseados nas características de evolução molecular do alinhamento original, assim como no caso do *bootstrap* não-paramétrico (embora esta seja uma suposição implícita no caso do *bootstrap* não-paramétrico, ver acima). Entretanto, na versão paramétrica as colunas do alinhamento não são reamostradas do conjunto de dados original, mas sim geradas (“evoluídas”) segundo uma topologia específica, através de um dado modelo markoviano, e a partir de uma única seqüência inicial que possua as mesmas freqüências de bases que os espécimes estudados*. Os sítios do alinhamento vão sendo formados até que se atinja o número de colunas equivalente ao alinhamento original (Huelsenbeck & Crandall, 1997; Huelsenbeck & Rannala, 1997).

Dentre as duas hipóteses a serem testadas, uma deve ser considerada a hipótese nula; a topologia e modelo markoviano associados à hipótese nula são então usados para a geração de um conjunto de alinhamentos (~100 ou mais) por *bootstrap* paramétrico. Após a geração desses conjuntos de dados, cada um tem para si otimizados os parâmetros do modelo (o mesmo que foi utilizado para os dados originais) para as duas hipóteses diferentes. Uma estatística biológica de interesse é então computada para cada hipótese (ex:

* Isso caso se assuma que o processo é estacionário na faixa de tempo que abrange a evolução desses táxons a partir de um ancestral comum. Por outro lado, poderia-se assumir que o processo é homogêneo mas não estacionário, ou seja, a seqüência ancestral pode não ter as mesmas freqüências de bases que os terminais, mas nesse caso deve-se de alguma maneira fornecer quais seriam as freqüências desse ancestral, o que é difícil de se estimar.

número de passos, valor de verossimilhança, taxa de transição/transversão, valor α da distribuição gama de taxas, etc.), e uma diferença entre esses valores é registrada (δ). Ao final da análise de todos os alinhamentos simulados, obtém-se uma distribuição dos valores de δ ; por fim, o valor de δ para os dados originais é comparado com a distribuição obtida, e se o valor é significativo, então conclui-se que a hipótese alternativa é significativamente superior à hipótese nula (Huelsenbeck & Crandall, 1997; Huelsenbeck & Rannala, 1997).

1.2.6 - Datação de divergências

É importante o conhecimento das datas aproximadas de divergências em uma dada filogenia, pois estas ajudam na inferência de processos que possam ter contribuído na história evolutiva daquele grupo de organismos (Bromham & Penny, 2003; Donoghue & Moore, 2003). Para se estabelecer pontos aproximados de divergência, deve-se aceitar algumas premissas, dependendo do método. Um dos caminhos possíveis é assumir que a taxa de evolução das seqüências ao longo do tempo é praticamente constante, premissa esta conhecida como hipótese do relógio molecular (Zuckerkandl & Pauling, 1962, 1965).

Para verificação de conformidade ou não com essa hipótese, realiza-se um teste hierárquico de razão de verossimilhanças (hLRT - do inglês *hierarchical Likelihood Ratio Test*) envolvendo a árvore de verossimilhança obtida com restrição do relógio molecular embutida e a árvore sem restrições. Esse teste hierarquizado é justificado, dado que a árvore na qual se assume constância de taxas é igual à topologia sem restrições em todos os aspectos (topologia, modelo e valores dos parâmetros desse modelo), a não ser pela restrição do emprego de relógio molecular. Essa restrição faz com que apenas (n-1) ramos variem de forma independente na topologia. Então, nesse caso, o valor do teste da razão é comparado com uma tabela de χ^2 com (n-2) graus de liberdade, que é o número total de parâmetros livres entre as duas árvores (2n-3 ramos, para n terminais, na árvore sem restrição de taxa homogênea, menos n-1 ramos na árvore restrita - Posada, 2003). Caso a hipótese de relógio molecular não seja rejeitada, procede-se à estimativa de datações, fixando-se um dos nós a partir de informações externas ao conjunto de dados (por exemplo, fósseis ou processos paleobiogeográficos específicos).

No caso da hipótese de relógio molecular ser rejeitada globalmente, existem formas alternativas de se determinar datações assumindo relógios locais. Uma delas (Takezaki *et*

al., 1995) é o uso de dois testes concomitantemente, que devem revelar resultados similares: o *two-cluster test* e o *branch-length test*. Ambos os testes são implementados numa árvore gerada previamente por neighbor-joining. O primeiro testa se dois grupos de clados derivados a partir de um mesmo nó interno apresentam ou não desvio significativo da premissa de uma mesma taxa para ambos, enquanto o segundo testa se a distância entre a raiz e cada um dos táxons terminais não desvia significativamente da média dessa mesma distância baseada no número total de táxons. Em caso de um determinado clado (seja ele uma sequência única ou um grupo de táxons) apresentar desvio significativo, ele deve ser retirado da análise, e então a linearização (ou transformação da topologia original em uma árvore ultramétrica) da nova árvore sem esses táxons heterogêneos torna-se possível.

Um raciocínio análogo foi desenvolvido através do uso de quartetos em conjunção com o princípio de máxima verossimilhança (Rambaut & Bromham, 1998). Tal método admite a incorporação de duas taxas diferenciais por quarteto analisado, excluindo das análises os quartetos que possuam heterogeneidade significativa. A partir do emprego desta técnica, inferiu-se com alto grau de corroboração que muitas das ordens de aves existentes atualmente teriam surgido antes do limite Cretáceo-Terciário (Rambaut & Bromham, 1998), conforme previamente proposto por Cooper e Penny (1997). Entretanto, para a análise, é necessário que sejam estipuladas duas datas de divergência para cada quarteto (Rambaut & Bromham, 1998), o que pode não ser aplicável em muitos casos.

Métodos como os descritos acima indicam (ou não) a exclusão de táxons para as análises de datações, o que é discutível. Além disso, em muitos casos apenas desvios dramáticos da hipótese de taxas constantes são detectados; por fim, a exclusão só faz sentido se a heterogeneidade não afetar muitas linhagens, caso contrário uma grande parte dos táxons pode vir a ser excluída (Welch & Bromham, 2005).

Uma outra maneira de considerar heterogeneidade de taxas de evolução é considerando-se a existência de mais de uma taxa ao longo da topologia sem evocar a necessidade de exclusão de clados. Alternativas possíveis são análises não-paramétricas (Sanderson, 1997; Britton *et al.*, 2002, 2006), semi-paramétricas (Sanderson, 2002) e bayesianas (Larget & Simon, 1999; Thorne *et al.* 1998; Huelsenbeck *et al.*, 2000; Kishino *et al.* 2001).

A técnica de NPRS (*Non-Parametric Rate Smoothing*) é uma alternativa não-paramétrica que envolve o cálculo, para cada ramo da topologia, de uma estimativa da taxa local de evolução molecular, com posterior minimização das diferenças entre essa taxa local e as taxas dos ramos que são descendentes imediatos. Assim sendo, variações bruscas de taxas ao longo da topologia são mais penalizadas (Sanderson, 1997).

A análise semi-paramétrica (PL - *Penalized Likelihood*) de Sanderson (2002) combina o fator de penalização descrito para o método NPRS com valores individuais de verossimilhança de taxas para cada ramo. A justaposição desses dois critérios (verossimilhanças individuais e função de penalidade) se dá através de um parâmetro de validação múltipla (*cross-validation*), responsável pela determinação do melhor equilíbrio entre os dois fatores. Quanto maior o valor desse parâmetro (portanto sendo maior a penalidade atribuída a variações de taxas entre ramos ancestrais e descendentes como um todo), menor o desvio em relação à premissa de relógio. O modelo é semi-paramétrico pois não chega a utilizar matrizes markovianas individuais para cada ramo (Sanderson, 2002).

O método PATHd8 busca minimizar a variação de taxas entre dois ramos descendentes de um mesmo ancestral (Britton *et al.*, 2002, 2006), através de uma média entre os tamanhos de todos os ramos das duas sub-árvores (que podem até ser ramos terminais) derivadas dessa linhagem, levando-se em consideração um ou mais nós da topologia com com datação fornecida.

Os três métodos acima utilizam como base para os cálculos uma árvore previamente obtida por algum método filogenético, não incluindo a informação presente no alinhamento das seqüências.

Já a análise bayesiana de taxas é um método totalmente paramétrico, que engloba o uso de matrizes markovianas com valores de parâmetro diferentes para cada ramo da filogenia, estipulando probabilidades de datações *a priori* (Larget & Simon, 1999; Thorne *et al.* 1998; Huelsenbeck *et al.*, 2000; Kishino *et al.* 2001), e assumindo como estimativas de taxas aquelas obtidas por amostragem a partir da zona de equilíbrio da MCMC. Na AB, a informação do alinhamento é levada em conta, ao contrário dos dois métodos anteriores.

De forma geral, PL possui performance superior à NPRS (Sanderson, 2002). O método PATHd8 é recomendado pelos autores em casos com muitos terminais e muitos nós com datação, e tende a ser mais eficiente quando as taxas ao longo da topologia não

desviam muito de uma taxa média (Britton *et al.*, 2002, 2006), além de ser uma técnica com execução computacional mais rápida, já que as médias para os tamanhos de ramo são sempre locais, ao contrário de PL, que visa uma minimização global das variações de taxas. Um estudo recente (Ericson *et al.*, 2006), com 91 táxons englobando a maioria dos grupos de aves e 23 pontos de calibração ao longo da topologia, mostrou que na maioria dos casos as datações realizadas por PL são mais antigas que as obtidas com PATHd8. Quanto à análise bayesiana, no entanto, a comparação torna-se mais complexa, pois os métodos diferem substancialmente na maneira de lidar com a variação de taxas, estimação dos tamanhos de ramos, uso da informação de fósseis para datação de divergências, e no tipo de estatísticas que são geradas após as análises (Welch & Bromham, 2005).

1.2.7 - Estudos biogeográficos

Biogeografia histórica e ecológica

A Biogeografia é a disciplina que estuda as distribuições dos organismos (Fosberg, 1976; Morrone & Crisci, 1995; Crisci, 2001). Ela pode ser subdividida em 2 campos principais: biogeografia ecológica e biogeografia histórica. A primeira está focada em processos ecológicos ocorrendo em escalas pequenas de tempo e espaço, enquanto a última lida com processos evolutivos ocorrendo ao longo de milhões de anos, e em larga escala espacial (Nelson & Platnick, 1981; Morrone & Crisci, 1995; Brown & Lomolino, 1998). Apesar do reconhecimento de que ambos os campos são importantes para o entendimento das distribuições de organismos, e de que na verdade não existe uma delimitação clara entre as duas abordagens, é ainda incomum a integração entre os dois campos em análises biogeográficas (Smith, 1989; Morrone & Crisci, 1995; Johnson *et al.*, 1998; Wiens & Donoghue, 2004).

A fim de se obter um panorama mais claro acerca de processos biogeográficos, portanto, deve-se recorrer a uma interação entre a biogeografia histórica e a ecológica (Morrone & Crisci, 1995; Wiens & Donoghue, 2004). Uma apreciação da importância desses dois elementos é a inferência de relações históricas entre áreas de endemismo: um cladograma de áreas, obtido por consenso a partir de filogenias para diferentes grupos de organismos, não depende apenas da história das conexões geológicas entre as áreas, mas também da história das conexões entre habitats similares: como exemplos opostos, uma

mesma região geológica pode apresentar habitats diferentes, onde cada um se assemelha ecologicamente a habitats de outras áreas (Wiens & Donoghue, 2004); e regiões localizadas em diferentes continentes podem apresentar as mesmas condições ecológicas apesar de habitadas por táxons completamente distintos (Morrone & Crisci, 1995).

Além disso, a localização temporal de eventos no passado também desempenha um papel fundamental (Donoghue & Moore, 2003). Isto porque congruência entre diferentes filogenias não necessariamente implica que elas tenham se originado de um mesmo processo, já que padrões filogenéticos concordantes podem estar associados a eventos ocorridos em tempos distintos (Donoghue & Moore, 2003).

Hipóteses para diversificação nos neotrópicos

A região neotropical apresenta uma grande representatividade mundial quanto ao número de espécies, tanto animais quanto vegetais (Myers *et al.*, 2000). Quanto às aves, a biodiversidade é a mais alta do mundo, com um total de mais de 3.750 espécies (Stotz *et al.*, 1996), inseridas em 90 famílias diferentes, 28 das quais são endêmicas. Como o número total de espécies de aves no mundo é superior a 9.700 espécies (Sibley & Monroe, 1990), cerca de mais de 1/3 dessas espécies se encontra nos neotrópicos.

Haffer (1969, 1974) sugeriu, como hipótese unificadora para a alta biodiversidade dos neotrópicos (mais especificamente a Amazônia), que o padrão atual das áreas de endemismo seria resultante de um processo comum: a partir das distribuições de diferentes organismos e de dados geológicos, ele postulou a existência de ciclos climáticos pretéritos influenciando ciclos de retração e posterior re-expansão das matas, o que levaria ao desenvolvimento de novas espécies durante os períodos de retração por isolamento reprodutivo (ou, ao contrário, ao desenvolvimento de novas espécies durante os períodos de *expansão* das matas, no caso de organismos associados a áreas abertas). Esta hipótese ficou conhecida como teoria dos refúgios (Haffer, 1969; 1974; Vanzolini & Williams, 1970).

Cracraft (1985), baseando-se nos dados de Haffer (1969, 1974), delimitou grandes áreas de endemismo utilizando apenas dados de distribuições de aves. Posteriormente, Cracraft e Prum (1988) inferiram as relações históricas entre essas áreas a partir da comparação das filogenias de 4 grupos de aves não-passeriformes, sempre sob o pressuposto de que a evolução de grupos distintos teria sido influenciada pelos mesmos

fatores subjacentes. Entretanto, Bates e col. (1998), baseando-se na distribuição de 1.717 espécies e subespécies de passeriformes, mostraram que apesar da obtenção de um único cladograma de áreas mais parcimonioso, as histórias individuais de muitas famílias se mostravam diferentes. Isto sugere que a relação histórica entre as áreas de endemismo por si só não explica a evolução nos diferentes grupos (Bates *et al.*, 1998). Devido a resultados como esse, e também ao fato da importância dos ciclos climáticos pretéritos não ser considerada unanimidade (enfraquecendo portanto a teoria dos refúgios - Colinvaux *et al.*, 2001), passou-se a reconhecer definitivamente que a existência das áreas de endemismo atuais não pode ser explicada por um único processo. Apesar do reconhecimento de que um único mecanismo não deva ser o responsável pela grande diversificação observada, diferentes teorias (não mutuamente exclusivas) foram levantadas ao longo do tempo para explicar a evolução de grupos de organismos, principalmente na Amazônia, sendo as principais as seguintes:

Hipóteses ligadas a ciclos paleoclimáticos do Terciário/Quaternário

- Hipótese dos refúgios - conforme descrito acima, populações separadas em diferentes áreas, devido a retrações florestais durante os ciclos paleoclimáticos do Terciário, poderiam evoluir independentemente em alopatria (Haffer, 1969, 1974, 1997; Vanzolini & Williams, 1970; Prance, 1973; Brown Jr., 1975).
- Hipótese do resfriamento climático - não teria havido fragmentação florestal em uma paisagem dominada por savanas. O resfriamento climático seria o fator decisivo para o rearranjo da vegetação na Amazônia durante os períodos glaciais do Pleistoceno (Bush, 1994; Bush & Oliveira, 2006).
- Hipótese da perturbação da vicariância - não teria havido retrações proeminentes nas áreas de floresta do Cenozóico; nas épocas glaciais, o principal fator gerador de vicariância teria sido regiões com heterogeneidade de gás carbônico (Colinvaux, 1998; Colinvaux *et al.*, 2000; Colinvaux & Oliveira, 2001).
- Hipótese da densidade do dossel - simulações sugeriram que não teria havido tempo para a invasão de vegetação de área aberta nos períodos glaciais. No entanto, teriam ocorrido reduções e acréscimos repetidos na densidade do dossel florestal sobre grandes áreas, que poderiam desempenhar um papel na vicariância (Cowling *et al.*, 2001).

Hipóteses paleogeográficas

- Hipótese do museu - a especiação ocorreria primariamente devido a processos dinâmicos atuando nas regiões montanhosas (ex: soerguimento dos Andes); com o tempo, as espécies originadas nesses habitats ocupariam também altitudes mais baixas; estas, então, funcionariam como um “museu” de espécies relativamente antigas (Fjeldså, 1994; Roy *et al.*, 1997; Moritz *et al.*, 2000; Rahbek & Graves, 2001; Stephen & Wiens, 2003; Smith *et al.*, 2007).
- Hipótese dos rios como barreiras - os rios seriam a fonte responsável pelo isolamento e evolução de espécies-irmãs localizadas em margens opostas (Wallace, 1852; Sick, 1967; Hershkovitz, 1977).
- Hipótese dos arcos - estruturas tectônicas dômicas em profundidade que subdividiriam uma bacia em diferentes sub-bacias, poderiam ter causado a vicariância entre diferentes grupos de organismos, levando à especiação (Patton *et al.*, 1997, 2000; Patton & Da Silva, 1998).
- Hipótese do soerguimento dos Andes - espécies que antes ocupavam terras baixas a oeste e noroeste da Amazônia teriam sido fragmentadas com o soerguimento das cordilheiras andinas (Chapman, 1917).

Hipóteses híbridas

- Hipótese da tricotomia basal - incursões marinhas de grande porte, durante o Terciário, teriam isolado 3 grandes regiões: o planalto das Guianas ao norte, o planalto Brasileiro ao sul, e a parte oriental dos Andes, em soerguimento, a oeste. A partir dessas áreas, populações ancestrais teriam se especiado por vicariância (Bates, 2002).
- Hipótese rios-refúgios - populações amazônicas teriam sido isoladas e evoluído em áreas delimitadas pela presença de rios de grande porte por um lado, e por regiões de terra firme ecologicamente desfavoráveis adjuntas às nascentes de rios, durante os períodos glaciais (Haffer, 1993a, 1993b).
- Hipótese do Lago - a maior parte da Amazônia teria sido coberta por um imenso lago durante o Plio-Pleistoceno (Klammer, 1984; Marroig & Cerqueira, 1997), que teria favorecido vicariância entre populações ao longo de sua margem.

Hipótese dos gradientes ecológicos - as pressões de seleção nos extremos opostos de um gradiente ecológico levariam à especiação por parapatría (Endler, 1977, 1982; Smith *et al.*, 2001).

Hipótese da estabilidade climática - em um ambiente fisicamente estável, úmido e relativamente quente, a taxa de extinção seria menor que a taxa de especiação, resultando num grande acúmulo de diversidade* (Sanders, 1968, 1969; Sanders & Hessler, 1969).

1.3 Objetivos

- Reavaliar as relações de parentesco entre as espécies do gênero *Ramphastos*;
- Abordar aspectos de evolução molecular do grupo;
- Estimar os períodos de divergência entre as linhagens;
- Associar as datações obtidas com processos paleogeográficos.

1.4 Material e Métodos

1.4.1 - Táxons estudados

Na tabela 1.1, a seguir, estão detalhadas as informações relativas a cada uma das amostras estudadas.

* Esta hipótese foi originalmente desenvolvida para organismos marinhos bentônicos.

Tabela 1.1: Amostras utilizadas no presente estudo.

	Identificação	Material	nº de coleção	LGEMA	Fonte	Localidade	Data da coleta	nº GenBank
1	<i>R. dicolorus 1</i>	sangue	-	R-210	Laboratório de Genética e Evolução Molecular de Aves	SC: Lages	-	-
2	<i>R. dicolorus 2</i>	tecido	B-282	-	University of Kansas	Paraguai: Caazapá	-	-
3	<i>R. s. sulfuratus</i>	tecido	B-2007	-	University of Kansas	México: Campeche	1998	-
4	<i>R. s. brevicarinatus</i>	tecido	B-28577	-	Louisiana Museum of Natural History	Panamá: Colón	-	-
5	<i>R. toco albogularis 1</i>	tecido	-	R-264	Laboratório de Genética e Evolução Molecular de Aves	SP: Uru	2004	-
6	<i>R. toco albogularis 2</i>	tecido	-	R-265	Laboratório de Genética e Evolução Molecular de Aves	SP: Uru	2004	-
7	<i>R. toco albogularis 3</i>	tecido	B-1477	-	Louisiana Museum of Natural History	-	-	-
8	<i>R. toco albogularis 4</i>	tecido	B-38206	-	Louisiana Museum of Natural History	Bolívia: Santa Cruz	1999	-
9	<i>R. brevis 1</i>	tecido	B-12175	-	Louisiana Museum of Natural History	Equador: Pichincha	1987	-
10	<i>R. brevis 2</i>	tecido	B-34977	-	Louisiana Museum of Natural History	Equador: Pichincha	-	-
11	<i>R. v. ariel SE 1</i>	tecido	B-35555	R-132	Museu Paraense Emílio Goeldi	SP: SW de Sete Barras	2000	-
12	<i>R. v. ariel SE 2</i>	sangue	-	R-213	Laboratório de Genética e Evolução Molecular de Aves	SC: Camboriú	-	-
13	<i>R. v. citreolaemus 1</i>	pele de museu	190822	-	Field Museum of Natural History	Equador: Antioquia	1947	-
14	<i>R. v. citreolaemus 2</i>	pele de museu	190823	-	Field Museum of Natural History	Equador: Antioquia	1947	-
15	<i>R. v. citreolaemus 3</i>	pele de museu	352027	-	Field Museum of Natural History	-	1961	-
16	<i>R. v. vitellinus 1</i>	tecido	B-35638	R-136	Museu Paraense Emílio Goeldi	PA: baixo rio Curuá	2000	-
17	<i>R. v. vitellinus 2</i>	tecido	B-35646	R-137	Museu Paraense Emílio Goeldi	PA: baixo rio Curuá	2000	-
18	<i>R. v. culminatus 1</i>	tecido	B-35652	R-139	Museu Paraense Emílio Goeldi	AM: NE de Careiro	2000	-
19	<i>R. v. culminatus 2</i>	tecido	B-35721	R-145	Museu Paraense Emílio Goeldi	AM: baixo rio Solimões	2000	-
20	<i>R. v. ariel N 1</i>	tecido	B-35667	R-143	Museu Paraense Emílio Goeldi	PA: NE de Belém	2000	-
21	<i>R. v. ariel N 2</i>	tecido	B-35668	R-144	Museu Paraense Emílio Goeldi	PA: NE de Belém	2000	-
22	<i>R. t. tucanus 1</i>	tecido	B-1356	-	University of Kansas	Guiana: Reserva Iwokrama	-	-
23	<i>R. t. tucanus 2</i>	tecido	B-35550	-	Museu Paraense Emílio Goeldi	PA: NE de Belém	-	-
24	<i>R. t. cuvieri 1</i>	tecido	B-35665	R-141	Museu Paraense Emílio Goeldi	AM: NE de Careiro	2000	-
25	<i>R. t. cuvieri 2</i>	tecido	B-35666	R-142	Museu Paraense Emílio Goeldi	AM: NE de Careiro	2000	-
26	<i>R. a. swainsonii 1</i>	tecido	B-2309	-	Louisiana Museum of Natural History	Panamá: Darién	-	-
27	<i>R. a. swainsonii 2</i>	tecido	B-11712	-	Louisiana Museum of Natural History	Equador: Esmeraldas	-	-
28	<i>R. a. ambiguus</i>	tecido	ANSP-4465	-	Academy of Natural Sciences	Equador: Zamora-Chinchi	1992	-
29	<i>Baillonius bailloni</i>	tecido	B-25891	-	Louisiana Museum of Natural History	Paraguai: Caazapá	1995	-
30	<i>Pteroglossus inscriptus</i>	tecido	B-8819	-	Louisiana Museum of Natural History	Bolívia: Pando	1986	-
31	<i>Selenidera reinwardtii</i>	tecido	B-27756	-	Louisiana Museum of Natural History	Perú: Loreto	-	-
32	<i>Andigena cucullata</i>	tecido	B-1273	-	Louisiana Museum of Natural History	Bolívia: La Paz	1981	-
33	<i>Aulacorhynchus prasinus</i>	tecido	B-1373	-	Louisiana Museum of Natural History	Panamá: Darién	1982	-
34	<i>Pteroglossus azarae flavirostris</i>	-	-	-	-	-	-	NC008549
35	<i>Dryocopus pileatus</i>	-	-	-	-	-	-	NC008546
36	<i>Ninox novaeseelandiae</i>	-	-	-	-	-	-	AY309457

1.4.2 - Regiões seqüenciadas

Foram obtidas seqüências de 11 regiões, sendo três não codificadoras (região controladora da mitocôndria, e íntrons 5 e 7 do gene β -fibrinogênio), duas que transcrevem RNA mas não traduzem proteínas (12S e 16S ribossomais), e seis que traduzem (genes mitocondriais citocromo-b, NADH-desidrogenase II, NADH-desidrogenase III, ATPase 6, ATPase 8 e citocromo c - oxidase I), para um total de 30 táxons (a não ser no caso do citocromo-b, para o qual foram seqüenciados 33 indivíduos). Foram deixados de fora das análises a NADH-desidrogenase III (por apresentar bandas duplas em amplificações, com os indivíduos apresentando uma ou outra) e o íntron 5 do β -fibrinogênio (pois não apresentou resolução topológica alguma entre os diferentes táxons), utilizando-se portanto os outros 9 fragmentos para o presente estudo.

1.4.3 - Extração de DNA

Protocolo I

O protocolo de extração utilizado foi modificado a partir de Brufford e col. (1992).

- Misturou-se 300 μ L de TNE (Tris-HCl 50mM, pH 7.5; NaCl 100mM; EDTA 5mM, pH 7.5), 30 μ L de Tris-HCl 1M, pH 7.5, 8 μ L de SDS 25% e 20 μ L de proteinase K (25 mg/mL) em um microtubo de 1,5mL;
- Colocou-se uma pequena fração da amostra e misturou-se por inversão;
- Incubou-se a 37 °C por 18 horas ou por 55 °C por 4 horas;
- Acrescentou-se 1 volume de fenol:clorofórmio:álcool isoamílico (25:24:1), misturou-se e centrifugou-se a 12000rpm por 10 minutos;
- Transferiu-se a fase aquosa para um novo tubo;
- Adicionou-se 2 volumes de etanol 95%, invertendo o tubo algumas vezes até o DNA precipitar;
- Centrifugou-se a 12000rpm por 5 minutos;
- Descartou-se o sobrenadante;
- Lavagem com etanol 70% (~ 300- 800mL);
- Descartou-se o etanol e deixou-se secar à temperatura ambiente ou em bomba de vácuo, por aproximadamente 30 minutos;

- Ressuspendeu-se em TE (Tris 10 mM; EDTA 1 mM), ~50-200µL.

Protocolo II

Algumas das extrações de DNA foram realizadas usando-se kits de extração da *QIAGEN*, entre elas aquelas realizadas para três indivíduos (13-15, tabela 1), pois estas foram obtidas de peles depositadas em museu. Como muita quantidade de DNA é degradada durante a preparação dessas peles nos museus, e como as datas de coleta são relativamente antigas, optou-se pelo uso de *kits*.

1.4.4 - Qualidade e concentração do DNA

As amostras de DNA foram submetidas à eletroforese em gel de agarose 1,0% (p/v), com tampão TBE 0.5x contendo 0,5 mg/mL de brometo de etídio, com marcador "100 bp DNA ladder" (GibcoBRL). A voltagem utilizada foi de 5 Volts/cm. A concentração foi medida por fragmentos padrões de 250, 125 e 62,5 ng/µL, pipetados no mesmo gel.

1.4.5 - Amplificação dos fragmentos

As reações de PCR foram preparadas para um volume final de 10µL (LGEMA - Laboratório de Genética e Evolução Molecular de Aves - USP) ou 25µL (FMNH - *The Field Museum of Natural History* - Chicago, Illinois-EUA) por microtubo:

FMNH

- 14,375µL de água mili-Q
- 2,5µL de Tampão 10x;
- 2,5µL de dNTP mix (2mM);
- 2 µL de MgCl₂ (25mM);
- 1,25µL de *primer L* (10 µM);
- 1,25µL de *primer H* (10 µM);
- 0,125µL de *Taq* DNA polimerase (5 U/µL);
- 1µL de DNA (1-20ng);

LGEMA

- 4,9µL de água mili-Q;
- 1µL de Tampão 10x;
- 1µL de dNTP mix (2mM);
- 1µL de *primer L* (10 µM);
- 1µL de *primer H* (10 µM);
- 0,1µL de *Taq* DNA polimerase (5 U/µL);
- 1µL de DNA (1-20ng);

Na tabela 1.2, a seguir, estão os *primers* utilizados e suas respectivas seqüências.

Tabela 1.2: *Primers* utilizados para amplificação das seqüências. Legenda - L: cadeia leve da mitocôndria; H: cadeia pesada.

Região	Direção	Nome	Seqüência	Fonte
12S	L	L _{ph} e1248	5' - AAA GCA TGG CAC TGA AGA YGC	Tavares, E. S. (com. pess.)
	H	12SH1843	5' - CCG CCA AGT CCT TAG AGT TT	Tavares, E. S. (com. pess.)
	L	12SL1735	5' - GGA TTA GAT ACC CCA CTA TGC	Miyaki et al., 1998
	H	12SH2170	5' - AGG GTG ACG GGC GGT ATG TAC G	Miyaki et al., 1998
16S	L	16SL2702	5' - CCT ACC GAG CTG GGT GAT AGC TGG TT	Miyaki et al., 1998
	H	16SH3309	5' - TGC GCT ACC TTC GCA CGG T	Miyaki et al., 1998
Região controladora	L	L538	5' - CCT CTG GTT CCT ARG TCA GG	modificado de Delport et al., 2002
	H	PROHSS	5' - GCT TTG GGA GTT GGA GAT AAA GG	Cabanne, G. S. (com. pess.)
ATPases 6 e 8	L	CO2GQL	5' - GGA CAA TGC TCA GAA ATC TGC GG	González et al., 2003
	H	CO3HMH	5' - CAT GGG CTG GGG TCR ACT ATG TG	González et al., 2003
COI	L	L6625	5' - CTA GAG GAG CCT GTT CTA TAA TCG	Hafner et al., 1994
	H	H7005	5' - CTT TCA GGT GTA AGC TGA RTG C	Hafner et al., 1994
Cyt-b	L	L14841	5' - GCT TCC ATC CAA CAT CTC AGC ATG ATG	Kocher et al., 1989
	H	TOUCCBH	5' - GAG AAR RAT GGG TGR AAT GG	Weckstein, 2005
	L	TOUCCBL	5' - CTT CCT NCT NCC ATT CCT AAT YRC AGG	Weckstein, 2005
	H	H16065	5' - GGA GTC TTC AGT CTC TGG TTT ACA AGA C	Helm-Bjornowski & Cracraft, 1993
ND2	L	L5215	5' - TAT CGG GCC CAT ACC CCG AAA AT	Hackett, 1996
	H	H5776TOUC	5' - GGC TGA RYA GGC MTC AAC CAR AC	Weckstein, 2005
	L	L5758TOUC	5' - TGN GAG ATR GAG GAG AAR GC	Weckstein, 2005
	H	H6313	5' - CTC TTA TTT AAG GCT TTG AAG GC	Sorenson et al., 1999
Fib-7	L	FibL2	5' - CTT CTG AGT AGG CAG AAC TT	Moyle, 2004
	H	Fib-B17U	5' - GGA GAA AAC AGG ACA ATG ACA ATT CAC	Pyrcchitto & Moore, 1997

As ampliações foram realizadas em termociclador TC2400 (Perkin-Elmer), no LGEMA, ou no modelo *DNA Engine Dyad Peltier Thermal Cycler* (MJ Research), no FMNH. Os ciclos de PCR foram semelhantes para cada fragmento amplificado, mudando-se apenas a etapa referente à hibridação dos *primers*. A seguir estão descritas as informações para cada região:

12s (12SL1735/12SH2170)

95°C ----- 30"
61°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

12s (Lphe1248/12SH1843)

95°C ----- 30"
54°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

16s (16SL2702/16SH3309)

95°C ----- 30"
55°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

Região controladora (CRL538/PROHSS)

95°C ----- 30"
54°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

Cyt-b (L14841/TOUCCBH)

95°C ----- 30"
58°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

Cyt-b (TOUCCBL/H16065)

95°C ----- 30"
56°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

COI (L6625/H7005)

95°C ----- 30"
58°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

ND2 (L5215/H5776TOUC)

95°C ----- 30"
58°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

ND2 (L5758TOUC/H6313)

95°C ----- 30"
60°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

Fib-7 (FibL2/FibI7U)

95°C ----- 30"
60°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

ATPases (CO2GQL/CO3HMH)

95°C ----- 30"
58°C ----- 30"
72°C ----- 40"
Repetir 40x
72°C ----- 2'

A quantidade de 2µL da reação foi carregada em gel de agarose 1,0% com marcador de peso molecular. A eletroforese indicou se o tamanho dos fragmentos correspondiam ao tamanho esperado, além de mostrar se teria havido alguma contaminação.

1.4.6 - Purificação dos fragmentos

Protocolo I (Exo-SAP)

Cerca de 5µL do produto da PCR foram misturados com 1µL de exonuclease (10U/µL) e 1µL de fosfatase alcalina de camarão (1U/µL) e incubados por 1h00 a 37 °C, seguido de 10 minutos a 80 °C.

Protocolo II (Gelase)

Quando observado mais de uma banda ou algum rastro na PCR, todo o volume da reação foi carregado em gel tipo “low-melt” 1% e submetido à eletroforese por 15 minutos. A seguir, as bandas de interesse foram cortadas e colocadas em microtubos. As amostras foram então aquecidas por 15 minutos a 70 °C, e depois resfriadas até 45 °C; após esse período, 1µL da enzima gelase foi colocado; as amostras foram então incubadas por ~ 4hrs a 45°C, e posteriormente, guardadas a 4 °C.

1.4.7 - PCR para seqüenciamento

Para cada amostra, a reação deu-se da seguinte maneira:

<u>LGEMA</u>	<u>FMNH</u>
- 2µL de Big Dye (Applied Biosystems);	- 3µL de água;
- 1,2µL de um dos primers (10µM);	- 3,3 µL de tampão p/ Big Dye;
- DNA amplificado (~2-4 µL);	- 1µL de Big Dye (Applied Biosystems);
	- 0,6µL de um dos primers (10µM);
	- DNA amplificado (~1-2 µL);

As reações foram conduzidas no termociclador TC2400 (Perkin-Elmer), no LGEMA, ou no *DNA Engine Dyad Peltier Thermal Cycler* (MJ Research), no FMNH, com

1 ciclo de 1 minuto a 96 °C, e 35 ciclos de: 95 °C por 10 segundos; 50 °C por 5 segundos; e 60 °C por 4 minutos.

1.4.8 - Purificação e precipitação para seqüenciamento

Protocolo I (LGEMA)

Este protocolo constou de uma lavagem inicial com isopropanol 75%, vortexação, 10 minutos à temperatura ambiente, centrifugação por 10 minutos e retirada do sobrenadante; posteriormente, uma nova lavagem foi realizada com 250 µL de etanol 70%, seguida de vortexação e nova centrifugação por 10 minutos. Retirou-se, então, o excesso de sobrenadante e procedeu-se à secagem (temperatura ambiente, estufa ou centrífuga à vácuo). Ao término, as amostras foram mantidas a -20 °C até o momento de suas aplicações no gel de seqüenciamento.

Protocolo II (FMNH)

Os produtos purificados foram, inicialmente, submetidos a uma lavagem com etanol 100% + 125mM EDTA. Então, foram submetidos à centrifugação por 30 minutos a 2500g, e depois retirado o sobrenadante. Uma nova lavagem e centrifugação foram efetuados, só que desta vez apenas com etanol 70% e sem EDTA, e com período de centrifugação de 15 minutos. Novamente submeteu-se à secagem dos produtos, podendo esta ser por uma das maneiras já mencionadas no “Protocolo I”.

1.4.9 - Seqüenciamento

Protocolo I (LGEMA)

O processo de preparação para a leitura automática contou com a lavagem das placas de vidro do sequenciador com detergente especial (Alconox 1%), seguido por repetidas lavagens em água corrente, e, por fim, água milli-Q quente. Após secagem, as placas foram posicionadas num suporte especial, colocando-se entre elas os espaçadores laterais. Seguiu-se então à preparação do gel:

- Misturou-se 10,8g de uréia, 0,4g de resina especial, 3mL de acrilamida e 11mL de água em agitador magnético, até que a uréia se dissolva completamente;

- Foram adicionados 3mL de TBE 10x e água mili-Q até a solução atingir 30mL;
- A solução foi então filtrada e deixada na geladeira até o momento da aplicação;
- Adicionou-se 2mL de TEMED e 150μL de persulfato de amônia;
- Aplicou-se rapidamente entre as placas de vidro;

O produto da precipitação foi ressuspenso em 1,6μL de dextran azul, EDTA 25μM e formamida (1:5), com desnaturação posterior das duplas fitas a 96 °C por 2 minutos. Após a polimerização do gel, as amostras foram aplicadas. O seqüenciador utilizado foi o modelo ABI-PRISM 377 (Perkin-Elmer).

Protocolo II (FMNH)

Foi aplicado a cada uma das amostras 10μL de Formamida “Hi-Di”, e então cada uma foi transferida para uma placa de seqüenciamento com entrada para 96 reações. O seqüenciador utilizado, ABI 3730, lê as amostras através do uso de capilares. Em cada um dos 96 capilares, existe um polímero especial que é usado como matriz de leitura para o detector de fluorescência. Após efetuada a troca desse polímero, de 2 placas com água (uma delas responsável pelo descarte de polímero já utilizado, durante o seqüenciamento, e a outra para lavagem dos capilares, entre as corridas), e da placa com tampão de corrida, prosseguiu-se à leitura automática das seqüências através do seqüenciador ABI 3730.

Referências bibliográficas - Capítulo 1

- Altekar, G.; Dwarkadas, S.; Huelsenbeck, J. P.; Ronquist, F. (2004): Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20:407-415.
- Archibald, J. K.; Mort, M. E.; Crawford, D. J. (2003): Bayesian inference of phylogeny: a non-technical primer. *Taxon* 52:187-191.

- Avice, J.C. (2004): *Molecular markers, natural history, and evolution* - 2^a edição. Sinauer Associates, Inc., Sunderland, MA.
- Barker F. K.; Lanyon, S. M. (2000): The Impact of Parsimony Weighting Schemes on Inferred Relationships among Toucans and Neotropical Barbets (Aves: Piciformes). *Mol. Phyl. Evol.* 15(2): 215–234.
- Barry, D.; Hartigan, J. A. (1987): Statistical analysis of hominoid molecular evolution. *Statistical Science* 2: 191-210.
- Bates, J. M. (2002): Avian diversification in Amazonia: evidence for historical complexity and avicariance model for a basic diversification pattern. *Bol. Mus. Para. Emílio Goeldi, Ser. Zoologia*.
- Bates, J. M.; Hackett, S. J.; Cracraft, J. (1998): Area-relationships in the neotropical lowlands: an hypothesis based on raw distributions of passerine birds. *J. Biogeogr.* 25: 783-793.
- Beaumont, M. A.; Rannala, B. (2004): The Bayesian revolution in genetics. *Nat. Rev. Genet.* 5: 251-261.
- Bensasson, D.; Zhang, D. X.; Hartl, D. L.; Hewitt, G. M. (2001): Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. & Evol.* 16(6): 314-321.
- Bergsten, J. (2005): A review of long-branch attraction. *Cladistics*. 21: 163-193.
- Britton, T.; Oxelman, B.; Vinnersten, A.; Bremer., K. (2002): Phylogenetic dating with confidence intervals using mean path lengths. *Mol. Phyl. Evol.* 24:58-65.
- Britton, T.; Anderson, C. L.; Jacquet, D.; Lundqvist, S.; Bremer., K. (2006): Estimating divergence times in large phylogenetic trees: (submitted to Systematic Biology). PATHd8 users manual version 1.0 - Website: www.math.su.se/PATHd8
- Bromham, L.; Penny, D. (2003): The modern molecular clock. *Nature Rev. Genet.* 4: 216-224.
- Brown, J. H.; Lomolino, M. V. (1998): *Biogeography* - 2^a edição. Sinauer, Sunderland, MA.
- Brown Jr., K. S. (1975): Geographical patterns of evolution in Neotropical Lepidoptera. Systematics and derivation of known and new Heliconiini (Nymphalidae: Nymphalinae). *J. Entomology B* 44: 201–242.

- Brufford, M. W.; Hanotte, O.; Brookfield, J. F. Y.; Burke, T. (1992): Single locus and multilocus DNA fingerprinting. In: C. A. R. Hoelzel (ed.), *Molecular genetics analysis of populations: a practical approach*. Oxford University Press, New York, pp. 225-269.
- Burton, P. J. K. (1984): Anatomy and evolution of the feeding apparatus in the avian orders Coraciiformes and Piciformes. *Bull. Br. Mus. (Nat. Hist.) Zool.* 47: 331-443.
- Bush, M. B. (1994): Amazonian speciation - a necessarily complex model. *J. Biogeogr.* 21: 5-17.
- Bush, M.B.; Oliveira, P. E. (2006): The rise and fall of the Refugial Hypothesis of Amazonian Speciation: a paleoecological perspective. *Biota Neotrop.* 6(1): 17 págs.
- Camin, J. H.; Sokal, R. R. (1965): A method for deducing branching sequences in phylogeny. *Evolution* 19: 311-326.
- Chadarevian, S. (1999): Protein sequencing and the making of molecular genetics. *Trends Biochem. Sci.* 24: 203-206.
- Chapman, F.M. (1917) The distribution of bird-life in Colombia - a contribution to a biological survey of South America. *Bull. Am. Mus. Nat. Hist.* 36, 1-729.
- Colinvaux, P. A. (1998): A new vicariance model for Amazonian endemics. *Global Ecol. Biogeogr. Lett.* 7: 95-96.
- Colinvaux, P.; Oliveira, P. E. (2001): Amazon plant diversity and climate through the Cenozoic. *Palaeogeogr. Palaeoclimat. Palaeoecol.* 166: 51-63.
- Colinvaux, P. A.; Oliveira, P. E.; Bush, M. B. (2000): Amazonian and Neotropical plant communities on glacial time scales - the failure of the aridity and refuge hypotheses. *Quaternary Sci. Rev.*, 19: 141-169.
- Colinvaux, P. A.; Irion, G.; Räsänen, M. E.; Bush, M. B.; Nunes de Mello, J. A. S. (2001): A paradigm to be discarded - geological and paleoecological data falsify the Haffer & Prance refuge hypothesis of Amazonian speciation. *Amazoniana*, 16, 609-646.
- Cooper, A.; Penny, D. (1997): Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. *Science* 275: 1109-1113.
- Cowling, S. A.; Maslin, M. A.; Sykes, M. T. (2001): Paleovegetation simulations of lowland Amazonia and implications for neotropical allopatry and speciation. *Quaternary Res.* 55: 140-149.

- Cracraft, J. (1985): Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. *In: Neotropical ornithology* (Eds.: Buckley, P. A.; Foster, M. S.; Morton, E. S.; Ridgely, R. S.; Buckley, F. G). Ornithological Monographs, n° 36: 49-84. Washington.
- Cracraft, J.; Prum, R. O. (1988): Patterns and processes of diversification: Speciation and historical congruence in some neotropical birds. *Evolution* 42: 603-620.
- Crisci, J. V. (2001): The voice of historical biogeography. *J. Biogeogr.* 28: 157-168.
- Delport, W.; Ferguson, J. W. H.; Bloomer (2002): Characterization and evolution of the mitochondrial DNA control region in hornbills (Bucerotiformes). *J. Mol. Evol.* 54: 794-806.
- Desjardins, P.; Morais, R. (1990): Sequence and gene organization of the chicken mitochondrial genome - a novel gene order in higher vertebrates. *J. Mol. Biol.* 212: 599-634.
- Donoghue, M. J.; Moore, B. R. (2003): Toward an integrative historical biogeography. *Integr. Comp. Biol.* 43: 261-270.
- Drummond, A. J.; Ho, S. Y. W.; Phillips, M. J.; Rambaut, A. (2006): Relaxed phylogenetics and dating with confidence. *PLOS Biology* 4(5): 699-710.
- Eck, R. V.; Dayhoff, M. O. (1966): *Atlas of protein sequence and structure*. National Biomedical Research Foundation, Silver Spring, MD.
- Edwards, A. W. F.; Cavalli-Sforza, L. L. (1963): The reconstruction of evolution. *Ann. Hum. Genet.* 27: 105-106.
- Efron, B. (1979): Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7: 1-26.
- Endler, J. A. (1977): *Geographic variation, speciation, and clines*. Princeton University Press, Princeton, N. J.
- Endler, J. A. (1982): Problems in distinguishing historical from ecological factors in biogeography. *Am. Zool.* 22: 441-452.
- Ericson, P. G. P.; Anderson, C. L.; Britton, T.; Elzanowski, A.; Johansson, U. S.; Källersjö, M.; Ohlson, J. I.; Parsons, T. J.; Zuccon, D.; Mayr, G. (2006): Diversification of Neoaves – integration of molecular sequences data and fossils. *Biol. Lett.* 2(4): 543-547.

- Espinosa de los Monteros, A. (2000): Higher-level phylogeny of Trogoniformes. *Mol. Phyl. Evol.* 14(1): 20-34.
- Espinosa de los Monteros, A. (2003): Models of the Primary and Secondary Structure of the 12S RNA of Birds: A Guideline for Sequence Aligning. *DNA Sequence* 14: 241-256.
- Fain, M. G.; Houde, P. (2004): Parallel radiations in the primary clades of birds. *Evolution* 58(11): 2558-2573.
- Felsenstein, J. (1978): Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27: 401-410.
- Felsenstein, J. (1981): Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17: 368-376.
- Felsenstein, J. (1985): Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.
- Felsenstein, J. (2003): *Inferring Phylogenies*. Sinauer Associates, Inc., Sunderland, MA.
- Fisher, R. A. (1922): On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. London A* 222: 309-368.
- Fisher, R. A. (1956): *Statistical Methods and Scientific Inference*. Oliver & Boyd, Edinburgh.
- Fitch, W. M. (1971): Toward defining the course of evolution: minimum change for a specified tree topology. *Syst. Zool.* 20: 406-416.
- Fitch, W. M. (1977): Phylogenies constrained by the crossover process as illustrated by human hemoglobins and a thirteen-cycle, eleven-amino-acid repeat in human apolipoprotein A-I. *Genetics* 86: 623-644.
- Fjeldså, J. (1994) Geographical patterns for relict and young species of birds in Africa and South America and implications for conservation priorities. *Biodiv. Conserv.* 3: 207-226.
- Fosberg, F. R. (1976): Geography, ecology, and biogeography. *Annals of the Association of American Geographers* 66:117-28.
- Gibb, G. C.; Kardailsky, O.; Kimball, R.; Braun, E.; Penny, D. (2006): Mitochondrial Genomes and Avian Phylogeny: Complex Characters and Resolvability Without

- Explosive Radiations. *Molecular Biology and Evolution*. MBE Advance Access published online on October 24, 2006.
- Giribet, G.; Wheeler, W. C. (1999): On gaps. *Mol. Phyl. Evol.* 13(1): 132-143.
- Golding, G. B. (1983): Estimates of DNA and protein sequence divergence: an examination of some assumptions. *Mol. Biol. Evol.* 1: 125-142.
- González, M. A.; Eberhard, J. R.; Lovette, I. J.; Olson, S. L.; Bermingham, E. (2003): Mitochondrial DNA phylogeography of the bay wren (Troglodytidae: *Thryothorus nigricapillus*) complex. *Condor* 105: 228-238.
- Graur, D.; Li, W.-H. (2000): *Fundamentals of molecular evolution* - 2^a edição. Sinauer Associates, Inc., Sunderland, MA.
- Griffiths, A. J. F.; Gelbart, W. M.; Miller, J. H.; Lewontin, R. C. (1999): *Modern genetic analysis*. W. H. Freeman and Company, New York, NY.
- Hackett, S. J. (1996): Molecular phylogenetics and biogeography of tanagers in the genus *Ramphocelus* (Aves). *Mol. Phyl. Evol.* 5: 368-382.
- Haffer, J. (1969): Speciation in Amazonian forest birds. *Science* 165: 131-137.
- Haffer, J. (1974): *Avian speciation in tropical South America*. Coleção “Publications of the Nuttall Ornithological Club”, nº 14, Cambridge, MA.
- Haffer, J. (1993a): On the ‘river effect’ in some forest birds of southern Amazonia. *Bol. Mus. Para. Emílio Goeldi, Zoologia* 8: 217-245.
- Haffer, J. (1993b): Time's cycle and Time's arrow in the history of Amazonia. *Biogeographica* 69, 15-45.
- Haffer, J. (1997): Alternative models of vertebrate speciation in Amazonia: An overview. *Biodiversity and Conservation* 6: 451-476.
- Hafner, M. S.; Sudman, P. D.; Villablanca, F. X.; Spradling, T. A.; Demastes, J. W.; Nadler, S. A. (1994): Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* 265: 1087-1090.
- Hasegawa, M.; Kishino, H.; Yano, T. (1985): Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22: 160-174.
- Hastings, W. K. (1970): Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97-109

- Hedrick, P. W. (2005): *Genetics of populations* - 3a edição. Jones and Bartlett Publishers, Sudbury, MA.
- Helm-Bychowski, K; Cracraft, J. (1993): Recovering phylogenetic signal from DNA sequences - relationships within the corvine assemblage (Class Aves) as inferred from complete sequences of the mitochondrial DNA cytochrome-b gene. *Mol. Biol. Evol.* 10(6): 1196-1214.
- Hendy, M. D.; Penny, D. (1989): A framework for the study of evolutionary trees. *Syst. Zool.* 38: 297-309.
- Hennig, W. (1950): *Grundzuge einer theorie der phylogenetischen systematik*. Deutscher Zentralverlag, Berlin.
- Hennig, W. (1966): *Phylogenetic systematics*. University of Illinois Press, Urbana, IL.
- Hershkovitz, P. (1977): *Living New World Monkeys (Platyrrhini)*, vol. 1. Chicago: University of Chicago Press. Chicago, IL.
- Hillis, D. M.; Bull, J. J. (1993): An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology*, 42 (2): 182-192.
- Hillis, D. M.; Bull, J. J.; White, M. E.; Badgett, M. R.; Molineux, I. J. (1992): Experimental phylogenetics: generation of a known phylogeny. *Science* 255: 589-592.
- Höfling, E. (1991): Etude comparative du crane chez des Ramphastidae (Aves, Piciformes). *Bonner Zoologische Beiträge* 42(1): 55-65.
- Höfling, E. (1998): Comparative cranial anatomy of Ramphastidae and Capitonidae. *Ostrich* 69: 389-389.
- Holder, M.; Lewis, P. O. (2003): Phylogeny estimation - traditional and bayesian approaches. *Nat. Rev. Genet.* 4: 275-284.
- Houde, P.; Cooper, A.; Leslie, E.; Strand, A. E.; Montaña, G. A. (1997): Phylogeny and evolution of 12S rDNA in Gruiformes (Aves). In: *Avian Molecular Evolution and Systematics* (Ed.: Mindell, D. P.), p. 121-158. Academic Press, San Diego, CA.
- Huelsenbeck, J. P. (1995): Performance of phylogenetic methods in simulation. *Syst. Biol.* 44(1): 17-48.
- Huelsenbeck, J. P.; Hillis, D. M. (1993): Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* 42: 247-264.

- Huelsenbeck, J. P.; Crandall, K. A. (1997): Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.* 28: 437–466.
- Huelsenbeck, J. P.; Rannala, B. (1997): Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. *Science* 276: 227–232.
- Huelsenbeck, J. P.; Ronquist, F. (2005): Bayesian analysis of molecular evolution using Mr. Bayes. *In: Statistical methods in molecular evolution* (Ed.: Nielsen, R.), p. 183-232. Springer Science+Business Media, Inc. New York, NY.
- Huelsenbeck, J. P.; Rannala, B.; Masly, J. P. (2000): Accomodating phylogenetic uncertainty in evolutionary studies. *Science* 288: 2349-2350.
- Huelsenbeck, J. P.; Ronquist, F.; Nielsen, R.; Bollback, J. P. (2001): Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310-2314.
- Jayaswal, V.; Jermiin, L. S.; Robinson, J. (2005): Estimation of phylogeny using a general Markov model. *Evol. Bioinf. Online* 1: 62-80.
- Jin, L.; Nei, M. (1990): Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.* 7: 82-102.
- Johansson, U. S.; Ericson, P. G. P. (2003): Molecular support for a sister group relationship between Pici and Galbulae (Piciformes sensu Wetmore 1960). *J. Avian Biol.* 34: 185-197.
- Johansson, U. S.; Parsons, T. J.; Irestedt, M.; Ericson, P. G. P. (2001): Clades within "higher land birds," evaluated by nuclear DNA sequences. *J. Zool. Syst. Evol. Research* 39: 37-51.
- Johnson, D. D. P.; Hay, S. I.; Rogers, D. J. (1998): Contemporary environmental correlates of endemic bird areas derived from meteorological satellite sensors. *Proc. R. Soc. Lond. B* 265: 951-959.
- Jukes, T. H., Cantor, C. R. (1969): Evolution of protein molecules. *In: Mammalian Protein Metabolism, Vol. III* (Ed.: Munro, H. N.), p. 21-132. Academic Press, New York, NY.
- Kimura, M. (1980): A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.

- Kimura, R. K.; Pereira, S. L.; Grau, E. T; Höfling, E.; Wajntal, A. (2004): Genetic distances and phylogenetic analysis suggest that *Baillonius* Cassin, 1867 is a *Pteroglossus* Illiger 1811. *Orn. Neotrop.* 15 : 527-537.
- Kishino, H.; Thorne, J. L.; Bruno, W. J. (2001): Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.*, 18(3): 352-361.
- Klammer, G. (1984): The relief of the extra-Andean Amazon basin. In: SIOLI, H. (ed.): The Amazon. Limnology and landscape ecology of a mighty tropical river and its basin. Dordrecht: Dr. W. Junk Publ., págs. 47-83.
- Kocher, T. D.; Thomas, W. K.; Meyer, A.; Edwards, S. V.; Pääbo, S.; Villablanca, F. X.; Wilson, A. C. (1989): Dynamics of mitochondrial DNA evolution in animals - amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA* 86: 6196-6200.
- Lanyon, S. M.; Hall, J. G. (1994): Reexamination of barbet monophily using mitochondrial-DNA sequence data. *Auk* 111(2): 389-397.
- Larget, B.; Simon, D. L. (1999): Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750-759.
- Liò, P.; Goldman, N. (1998): Models of molecular evolution and phylogeny. *Genome Res.* 8: 1233-1244.
- Manly, B. F. J. (1997): *Randomization, bootstrap and Monte Carlo methods in biology* - 2^a edição. CRC Press LLC, Boca Raton, FL.
- Marroig, G.; Cerqueira, R. (1997): Plio-Pleistocene South American history and the Amazon Lagoon hypothesis: a piece in the puzzle of Amazonian diversification. *J. Comp. Biol.* 2: 103-119.
- Maxam, A. M.; Gilbert, W. (1977): A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* 74: 560-564.
- Metropolis, N.; Rosenbluth, A.E.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. (1953): Perspective on "Equation of state calculations by fast computing machines". *J. Chem. Phys.* 21: 1087-1092.

- Mindell, D. P.; Sorenson, M. D.; Huddleston, C. J.; Miranda, H. C.; Knight, A.; Sawchuck, S. J.; Yuri, T. (1997): Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. *In: Avian Molecular Evolution and Systematics* (Ed.: Mindell, D. P.), p. 213–247. Academic Press, San Diego, CA.
- Mindell, D. P.; Sorenson, M. D.; Dimcheff, D.E. (1998): Multiple independent origins of mitochondrial gene order in birds. *Proc. Natl. Acad. Sci. USA* 95(18): 10693-10697.
- Miyaki, C. Y.; Mاتيoli, S. R.; Burke, T.; Wajntal, A. (1998): Parrot evolution and paleogeographical events - mitochondrial DNA evidence. *Mol. Biol. Evol.* 15: 544-551.
- Moritz, C.; Patton, J. L.; Schneider, C. J.; Smith, T. B. (2000): Diversification of rainforest faunas: an integrated molecular approach. *Annu. Rev. Ecol. Syst.* 31: 533–563.
- Morrone, J. J.; Crisci, J. V. (1995): Historical biogeography: introduction to methods. *Ann. Rev. Ecol. Syst.* 26: 373-401.
- Moyle, R.G. (2004): Phylogenetics of barbets (Aves: Piciformes) based on nuclear and mitochondrial DNA sequence data. *Mol. Phyl. Evol.* 30: 187-200.
- Mullis, K. B.; Faloona, F. A. (1987): Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology* 155: 335-350.
- Myers, N.; Mittermeier, R. A.; Mittermeier, C. G.; da Fonseca, G. A. B.; Kent, J. (2000): Biodiversity hotspots for conservation priorities. *Nature* 403(24): 853-858.
- Nei, M.; Gojobori, T. (1986): Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3: 418-426.
- Nei, M.; Kumar, S. (2000): *Molecular evolution and phylogenetics*. Oxford University Press, Inc., New York, NY.
- Nelson, G.; Platnick, N. I. (1981): *Systematics and biogeography: cladistics and vicariance*. Columbia University Press, New York, NY.
- Patton, J. L.; Da Silva, M. N. F. (1998): Rivers, refuges, and ridges - the geography of speciation of Amazonian mammals. In: Howard, D. J. & S. H. Berlocher (eds.): *Endless Forms - Species and Speciation* (págs. 202-213). Oxford University Press. Nova York, NY.
- Patton, J. L.; Da Silva, M. N.; Lara, M. C.; Mustrangi, M. A. (1997): Diversity, differentiation and the historical biogeography of nonvolant small mammals of the

- Neotropical forests. In: Laurance, W. F. & R. O. Bierregaard Jr. (eds.): Tropical forest remnants: Ecology, management and conservation of fragmented communities (págs. 455-465). University of Chicago Press. Chicago, IL.
- Patton, J. L.; Da Silva, M. N.; Malcolm, J. R. (2000): Mammals of the Rio Juruá and the evolutionary and ecological diversification of Amazonia". *Bull. Amer. Mus. Nat. Hist.* 244: 1-306.
- Philippe, H.; Zhou, Y.; Brinkmann, H.; Rodrigue, N.; Delsuc, F. (2005): Heterotachy and long-branch attraction in phylogenetics. *BMC Evol. Biol.* 5(50): 1-8.
- Poe, S.; Chubb, A. L. (2004): Birds in a bush: five genes indicate explosive evolution of avian orders. *Evolution.* 58(2): 404-415.
- Pol, D.; Siddall, M. E. (2001): Biases in maximum likelihood and parsimony: a simulation approach to a 10-taxon case. *Cladistics* 17: 266-281.
- Posada, D. (2003): Selecting models of evolution. In: The phylogenetic handbook: A Practical Approach to DNA and Protein Phylogeny (Eds.: M. Salemi & A.-M. Vandamme). Cambridge University Press, CA.
- Prance, G. T. (1973): Phytogeographic support for the theory of Pleistocene forest refuges in the Amazon basin, based on evidence from distribution patterns in Caryocaraceae, Chrysobalanaceae, Dichapetalaceae and Lecythidaceae". *Acta Amazônica* 3: 5-28.
- Prum, R. O. (1988): Phylogenetic interrelationships of the barbets (Aves: Capitonidae) and toucans (Aves: Ramphastidae) based on morphology with comparisons to DNA-DNA hybridization. *Zool. J. Linn. Soc.* 92: 313-343.
- Prychitko, T. M.; Moore, W. S. (1997): The utility of DNA sequences of an intron from the beta-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). *Mol. Phyl. Evol.* 8: 193-204.
- Prychitko, T. M.; Moore, W. S. (2003): Alignment and Phylogenetic Analysis of β -Fibrinogen Intron 7 Sequences Among Avian Orders Reveal Conserved regions Within Intron. *Mol. Biol. Evol.* 20(5): 762-771.
- Rahbek, C.; Graves, G. R. (2001): Multiscale assessment of patterns of avian species richness. *Proc. Natl. Acad. Sci. USA* 98: 4534-4539.

- Rambaut, A.; Bromham, L. (1998): Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* 15(4): 442-448.
- Roe, B. A.; Ma, D. P.; Wilson, R. K.; Wong, J. F. H. (1985): The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem.* 260: 9759-9774.
- Roy, M. (1997): Recent diversification in African bulbuls (Pycnonotidae: Andropadus) supports a montane speciation model. *Proc. R. Soc. Lond. Ser. B* 264:1337-1344.
- Rumjanek, F. D. (2001): *Introdução à biologia molecular*. Âmbito Cultural Edições Ltda., Rio de Janeiro, RJ.
- Saiki, R. K.; Scharf, S.; Faloona, F.; Mullis, K. B.; Horn, G. T.; Erlich, H. A.; Arnheim, N. (1985): Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230(4732):1350-1354.
- Sanders, H. L. (1968): Marine benthic diversity: a comparative study. *Am. Nat.* 102(925): 243-282.
- Sanders, H. L. (1969): Benthic marine diversity and the stability-time hypothesis. *Brookhaven Symp. Biol.* 22: 71-81.
- Sanders, H. L.; Hessler, R. H. (1969): Ecology of the deep-sea benthos. *Science* 163: 1419-1424.
- Sanderson, M. J. (1989): Confidence limits on phylogenies: the bootstrap revisited. *Cladistics* 5: 113-129.
- Sanderson, M. J. (1997): A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.*, 14, 1218-1231.
- Sanderson, M. J. (2002): Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19: 101-109.
- Sanderson, M. J.; Wojciechowski, M. F. (2000): Improved bootstrap confidence limits in large-scale phylogenies, with an example from Neo-Astragalus (Leguminosae). *Syst. Biol.* 49: 671-685.
- Sanger, F.; Nicklen, S.; Coulson, A. R. (1977): DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467.

- Sanson, G. F. O.; Kawashita, S. Y.; Brunstein, A.; Briones, M. R. S. (2002): Experimental phylogeny of neutrally evolving DNA sequences generated by a bifurcate series of nested polymerase chain reactions. *Mol. Biol. Evol.* 19(2): 170-178.
- Scharf, S. J.; Horn, G. T.; Erlich, H. A. (1986): Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science* 233(4768):1076-1078.
- Scheffler, I. E. (1999): *Mitochondria*. John Wiley & Sons, Inc., New York, NY.
- Short, L. L.; Horne, J. F. M. (2001): *Toucans, Barbets and Honeyguides*. Coleção "Bird Families of the World". Oxford University Press, New York, NY.
- Sibley, C. G.; Ahlquist, J. E. (1990): *Phylogeny and classification of birds: a study in molecular evolution*. Yale University Press, New Haven, CT.
- Sibley, C. G.; Monroe, B. L. (1990): *Distribution and taxonomy of birds of the world*. Yale University Press, New Haven, CT.
- Sick, H. (1967): Rios e enchentes na Amazônia como obstáculo para a avifauna. *Atas simp. sobre a biota Amazônica*, vol. 5 (Zoologia): 495-520.
- Sick, H. (1997): *Ornitologia Brasileira* (Tradução). Ed. Nova Fronteira, Rio de Janeiro, RJ.
- Siddall, M. E. (1998): Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris zone. *Cladistics* 14: 209-220.
- Simmons, M. P.; Pickett, K. M.; Miya, M. (2004): How meaningful are bayesian support values? *Mol. Biol. Evol.* 21(1): 188-199.
- Smith, C. H. (1989): Historical biogeography: geography as evolution, evolution as geography. *New Zealand Journal of Zoology* 16: 773-785.
- Smith, S. A.; de Oca, A. N. M.; Reeder, T. W.; Wiens, J. J. (2007): A phylogenetic perspective on elevational species richness patterns in middle american treefrogs - why so few species in lowland tropical rainforests? *Evolution* 61(5): 1188-1207.
- Smith, T. B.; Schneider, C. J.; Holder, K. (2001): Refugial isolation versus ecological gradients - testing alternative mechanisms of evolutionary divergence in four rainforest vertebrates. *Genetica* 112: 383-398.
- Sokal, R. R.; Rohlf, F. J. (1995): Biometry - *The principles and practice of statistics in biological research* - 3a edição. W. H. Freeman and Company, New York, NY.

- Sorenson, M. D.; Quinn, T. W. (1998): Numts: A challenge for avian systematics and population biology. *The Auk* 115: 214-221.
- Sorenson, M. D.; Ast, J. C.; Dimcheff, D. E.; Yuri, T.; Mindell, D. P. (1999): Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phyl. Evol.* 12(2): 105-114.
- Stephens, P. R.; Wiens, J. J. (2003): Explaining species richness from continents to communities - the time-for-speciation effect in emydid turtles. *Am. Nat.* 161:112-128.
- Stotz, D. F.; Fitzpatrick, J. W.; Parker III, T. A.; Moskovitz, D. K. (1996): *Neotropical birds: ecology and conservation*. University of Chicago Press, Chicago, IL.
- Suzuki, Y.; Glazko, G. V.; Nei, M. (2002): Overcredibility of molecular phylogenies obtained by bayesian phylogenetics. *Proc. Nat. Acad. Sci. USA* 99: 15138-16143.
- Swierczewski, E. V. & Raikow, R. J. (1981): Hind limb morphology, phylogeny and classification of the Piciformes. *The Auk* 98: 466-480.
- Swofford, D.L. (2002): PAUP* - Phylogenetic Analysis Using Parsimony (*and Other Methods). Versão 4.0b10. Sinauer Associates, Sunderland, MA
- Swofford, D. L. (2003): PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Takezaki, N.; Rzhetsky, A.; Nei, M. (1995): Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* 12: 823-833.
- Thompson, M. W.; McInnes, R. R.; Willard, H. F. (1991): *Genética médica - 5ª edição*. Editora Guanabara Koogan S. A., Rio de Janeiro, RJ.
- Thorne, J. L.; Kishino, H.; Painter, I. S. (1998): Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15(12): 1647-1657.
- Van Tuinen, M.; Sibley, C. G.; Hedges, S. B. (2000): The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. *Molecular Biology and Evolution* 17: 451-457.
- Vanzolini, P.; Williams, E. E. (1970): South American anoles: geographic differentiation and evolution of the *Anolis chrysolepis* species group (Sauria, Iguanidae). *Arq. Zool. São Paulo*, 19(1-4): 1-298.

- Wallace, A.R. (1852): On the monkeys of the Amazon. *Proceedings of the Zoological Society of London* 20: 107-110.
- Weckstein, J. D. (2005): Molecular phylogenetics of the *Ramphastos* toucans - implications for the evolution of morphology, vocalizations, and coloration. *Auk* 122(4): 1191-1209.
- Welch, J. J.; Bromham, L. (2005): Molecular dating when rates vary. *Trends Ecol. Evol.* 20(6): 320-327.
- Wiens, J. J.; Donoghue, M. J. (2004): Historical *biogeography, ecology and species richness*. *Trends Ecol. Evol.* 19(12): 639-644.
- Yang, Z. (1994b): Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39: 306-314.
- Yang, Z. (1996): Phylogenetic analysis using parsimony and likelihood methods. *J. Mol. Evol.* 42: 294-307.
- Yang, Z. (1997): How often do wrong models produce better phylogenies? *Mol. Biol. Evol.* 14: 105-108.
- Yang, Z.; Kumar, S. (1996): Approximate methods for estimating the pattern of nucleotide substitution rates among sites. *Mol. Biol. Evol.* 13: 650-659.
- Zuckerkandl, E.; Pauling, L. (1962): Molecular disease, evolution, and genetic heterogeneity. *In: Horizons in biochemistry* (Eds.: Kasha, M.; Pullman, B.). Academic Press, New York, NY.
- Zuckerkandl, E.; Pauling, L. (1965): Evolutionary divergence and convergence in proteins. *In: Evolving genes and proteins* (Eds.: Bryson, V.; Vogel, H. J.). Academic Press, New York, NY.

Capítulo 2 - Aspectos da evolução molecular do gene 12S nas espécies do gênero *Ramphastos*

Abstract

A description of the molecular characteristics of the mitochondrial ribosomal gene 12S is given for the species of the genus *Ramphastos* (Aves: Piciformes: Ramphastidae), and its utility for phylogenetic reconstruction is evaluated. The secondary structure was inferred by a comparison of the alignment obtained to a consensus published previously, which was in turn based on the analysis of several bird orders. The general structure is very conserved, showing the 40 intramolecular pairing stems and the 37 loops (internal or external) observed in other birds. The analysis of the corrected patristic distances showed some level of saturation, and also showed that loops (regions of single-stranded RNA) evolve at a higher relative rate than stems, in accordance with what is known for birds in general and for other animal groups. Three different approaches were carried out in order to avoid the effect that coevolution of pairing sites may exert over phylogenetic reconstructions: 1) topologies based on the analysis of only {loops + proximal portion of stems} or else {loops + distal portion of stems}; 2) maximum parsimony with differential weighting (weight=0.5) at sites pertaining to stems; 3) bayesian analysis using a mixed model, which assumed covariation of rates in stems, and independence in loops. There was a great topological similarity among the three different approaches when bootstrap levels $\geq 70\%$ and posterior probability estimates $\geq 95\%$ were considered; furthermore, such cladograms showed resemblance to the bootstrap tree obtained by both maximum likelihood and maximum parsimony analyses including all sites. Nevertheless, such similarity may be due to the effect of a polytomy near the base of the ingroup, perhaps because of a lack of resolution of the gene concerning the lineages studied. Even with such an absence of phylogenetic signal at the base of the ingroup, the cladograms corroborated by bootstrap and posterior probability estimates were congruent with a recently published tree, inferred from 3 mitochondrial coding genes. Such results indicate that the discrepancy of rates between stems and loops, as well as the use of a more adequate modelling regarding covariation in stems, were not decisive in terms of both precision and accuracy of the topological resolution, given the gene and taxa analysed. Two models assuming covariation at stems were further used, both indicating that single-substitution rates are higher than double-substitution rates, but with the latter occurring at an appreciable level.

Resumo

Uma descrição das características moleculares do gene ribossômico mitocondrial 12S para as espécies do gênero *Ramphastos* (Aves: Piciformes: Ramphastidae) é apresentada, sendo sua utilidade para reconstrução filogenética das espécies do grupo avaliada. A estrutura secundária foi inferida a partir de comparação do alinhamento com um consenso previamente publicado para diferentes ordens de aves. A estrutura geral é bastante conservada, tendo sido detectadas as 40 hastes de pareamento intra-molecular e as 37 alças (internas e externas) inferidas para outras aves. A análise com distâncias patrísticas corrigidas revelou a presença de saturação em algum grau, mostrando também que as alças (regiões de fita simples de DNA) evoluem a uma taxa relativa mais alta que as hastes, em concordância com o que é conhecido em aves em geral e em outros grupos de animais. Foram examinadas 3 diferentes abordagens para contornar o efeito que a co-evolução entre sítios parentes pode exercer nas reconstruções filogenéticas: 1) topologias baseadas em análises contendo apenas {alças + porção proximal das hastes} ou {alças + porção distal das hastes}; 2) máxima parcimônia com ponderação diferencial (peso=0,5) de sítios associados às hastes; 3) análise bayesiana com um modelo misto que assumiu covariação de taxas para as hastes, e independência para as regiões de alças. A semelhança entre as topologias para as três abordagens acima foi grande quando utilizados níveis de consenso de *bootstrap* $\geq 70\%$ e estimativas de probabilidade posterior $\geq 95\%$; tais cladogramas foram também semelhantes ao obtido pela análise com o número total de sítios. No entanto, tal similaridade pode ser simplesmente o efeito de uma politomia próxima à base do grupo interno, talvez devido a uma falta de resolução do gene para as linhagens estudadas. Mesmo mostrando essa ausência de sinal na base do grupo interno, os cladogramas corroborados pela análise de *bootstrap* se mostraram congruentes com uma árvore, publicada recentemente, inferida a partir de 3 genes codificadores mitocondriais para o gênero. Esses resultados indicam que a discrepância entre as taxas em alças e hastes, assim como o uso de uma modelização mais adequada para a questão da covariação nas hastes, não foram fatores decisivos tanto para a precisão quanto para a acurácia na resolução topológica com relação ao 12S e táxons estudados. Dois modelos foram aplicados para a análise das hastes, ambos indicando que as taxas de substituições simples são maiores que as duplas, mas com estas últimas ocorrendo em frequências não-desprezíveis.

2.1 Introdução

O gene 12S é um dentre dois genes mitocondriais envolvidos na formação de ribossomos, sendo o menor deles em tamanho. Bioquimicamente, ele é caracterizado pela presença de uma estrutura secundária relativamente conservada ao longo de um amplo espectro de grupos animais (Van de Peer *et al.*, 1994, 1997), sendo subdividido em quatro domínios (ou hélices) gerais (Neefs *et al.*, 1991). O 12S é comumente utilizado em estudos evolutivos, apresentando um grau de resolução filogenética útil tanto para divergências relativamente antigas, quanto para as relativamente recentes (Kocher *et al.*, 1989; Mindell & Honeycutt, 1990; Simon *et al.*, 1994).

A estrutura secundária do 12S obtida em aves apresenta poucas modificações em relação à estrutura geral para outros organismos (Espinosa de los Monteros, 2003), apresentando 40 hastes (regiões de pareamento intra-molecular) e 37 alças (regiões de fita simples) no total. Devido ao fato da estrutura secundária ser importante sob um ponto de vista bioquímico (e portanto evolutivo), mas não necessariamente as bases que compõem a estrutura primária, substituições ao longo do tempo podem mudar a composição dos sítios pareantes. Se uma determinada mutação ocorre em um nucleotídeo pertencente a um trecho de haste, ela pode reverter ao estado original, levar a uma substituição compensatória no outro nucleotídeo do par, ou formar um pareamento não-canônico relativamente estável (ex: A:C, U:G).

Estudos filogenéticos, tanto por máxima parcimônia (MP), máxima verossimilhança (MV), ou análise bayesiana (AB), que não se utilizam da estrutura secundária da molécula podem incorrer em dois tipos de erro:

I. *Ambigüidades no alinhamento* - regiões onde o alinhamento múltiplo é difícil, devido à existência de *indels* e/ou substituições nas alças e hastes, acabam sendo descartadas das análises; ou então mantidas e editadas por um processo manual (e portanto subjetivo) ou submetidas a algum tipo de otimização heurística que não reflete as pressões seletivas sofridas por essas regiões (Morrison & Ellis, 1997; Hickson *et al.*, 2000). Naturalmente, o uso da estrutura secundária como auxiliar no alinhamento dessas áreas pode não corrigir as ambigüidades completamente, mas minimiza o problema.

II. *Imprecisão topológica, por não ser considerada a covariação de taxas* - nas hastes, a evolução de sítios pareantes apresenta correlação. No entanto, MP, MV e AB, dentre outros métodos de reconstrução filogenética, assumem independência como padrão. Baseados nesse raciocínio, Jow e col. (2002) enfatizaram que modelos que consideram sítios pareantes como sendo independentes não devem ser utilizados no estudo das regiões como as das hastes nos RNAs. Galtier (2004) mostrou que desvios em relação a esse pressuposto podem deturpar a informação filogenética contida nos dados, podendo indicar fortemente a presença de ramos internos não-existentes na topologia real.

Algumas formas de se incorporar covariação foram aqui testadas, utilizando-se seqüências das espécies (e quase todas as subespécies) do gênero *Ramphastos* (Aves: Piciformes: Ramphastidae):

1) *MP com o conjunto de dados {alças + hastes 5'} e {alças + hastes 3'}, alternadamente* - uma maneira de se lidar com correlações entre taxas, para reconstrução filogenética, é simplesmente ignorar metade dos caracteres que ocorrem nas hastes; portanto, nesse caso, analisou-se separadamente a inclusão das regiões de alças com as hastes 5' (ou proximais), e hastes 3' (ou distais), e comparou-se os resultados; se a correlação=1, os dois conjuntos de dados devem resgatar a mesma topologia, e com valores de *bootstrap* semelhantes; no entanto, caso correlação<1 (indicando que há um certo grau de pareamentos não-canônicos entre as bases nas regiões 5' e 3' das hastes), a resolução topológica pode vir a ser ligeiramente diferente entre as duas partições de dados, devido ao resultado de assimetria na covariação de taxas.

2) *MP com ponderação diferencial dos sítios correlacionados* - uma outra forma plausível para contornar a correlação entre estados é atribuir um peso menor na análise de MP aos caracteres que já se sabe estarem correlacionados, como é o caso aqui (Wheeler & Honeycutt, 1988; Dixon & Hillis, 1993; Springer *et al.*, 1995). No presente estudo, um peso de 0,5 foi atribuído a todos os caracteres pertencentes às regiões de hastes; ou seja, assumiu-se correlação=1 entre caracteres pareados, para que a soma dos pesos entre os dois sítios de um mesmo par fosse igual a 1.

3) *AB com modelo de covariação entre posições* - assumiu-se um modelo híbrido, englobando ambos os tipos de regiões: um modelo variante de GTR (*General Time*

Reversible) para alças (4x4 estados), e um modelo de pareamento para as hastes, considerando pareamentos canônicos e não-canônicos.

2.2 Objetivos

- Descrever a estrutura secundária do 12S em tucanos (gênero *Ramphastos*);
- Realizar análises filogenéticas assumindo independência em todos os pontos da molécula (mesmo nas hastes), e comparar o resultado obtido com topologias geradas por métodos que levam em consideração a correlação de taxas nas hastes;
- Analisar quantitativamente as taxas de evolução molecular nas hastes do 12S no gênero estudado, através do uso de um modelo verossímil *a priori* para o grupo (RNA_8);

2.3 Material e Métodos

Foram utilizados 30 táxons para a inferência da estrutura secundária da maior parte da molécula e também para as análises filogenéticas, referenciados na tabela 1.1 (táxons 1-12, 16-33). O grupo interno foi constituído pelas espécies e subespécies do gênero *Ramphastos* (tucanos), num total de 25 seqüências analisadas; como grupo externo, foi utilizado um membro de cada gênero de arazaris, para um total de cinco seqüências incluídas. As amplificações e seqüenciamentos foram obtidos com os pares de *primers* Lphe1248-12SH1843 (porção inicial do gene) e 12SL1735-12SH2170 (porção mediana do gene), conforme descrito na tabela 1.2. A porção terminal do gene foi determinada apenas para os táxons da tabela 2.1; para os restantes, a leitura pelo seqüenciador apresentou picos múltiplos de bases a partir de um certo ponto, devido a uma região repetitiva mononucleotídica (>12 bases) localizada dentro dessa porção terminal, impedindo a correta atribuição de bases pelo leitor além desse limite.

Tabela 2.1: Táxons usados para determinação da estrutura secundária referente à região distal do gene 12S.

Espécie	Identificação (LGEMA)	Material	nº GenBank
<i>Ramphastos sulfuratus</i>	R-260	sangue	-
<i>Bailloni</i> <i>bailloni</i>	R-206	sangue	-
<i>Pteroglossus azara flavirostris</i>	-	-	NC 008549
<i>Selenidera maculirostris</i>	R-204	sangue	-
<i>Andigena laminirostris</i>	R-34	sangue	-
<i>Aulacorhynchus prasinus</i>	-	-	U89207

Tal região foi obtida utilizando-se o par de *primers* 12SL1858 (5'-CTA GAG GAG CCT GTT CTA TAA TCG) e H2294Val (5'-CTT TCA GGT GTA AGC TGA RTG C), ambos modificados de Sorenson e col. (1999), com ciclos de desnaturação = 94°C, hibridação de primers = 62°C, e extensão das cadeias recém-formadas = 72°C, por 40 ciclos consecutivos. As condições de PCR e seqüenciamento para as regiões inicial e mediana se deram conforme a seção 1.4 (Material e Métodos) do Capítulo 1.

Medidas foram tomadas para minimizar a chance de amplificação accidental de regiões não-ortólogas ao 12S: a grande maioria dos táxons teve seu DNA amplificado a partir de tecido rico em mitocôndrias; buscou-se indícios de mais de uma banda por amplificação, assim como sinais de substituições de qualquer tipo (seja pontuais ou *indels*) nas regiões de sobreposição entre as seqüências obtidas para cada par de *primers*, para um mesmo táxon; e amplificou-se, para alguns táxons, um fragmento único que sobrepôs as partes inicial e mediana do gene 12S, para se testar se haveria diferenças em relação aos PCRs parciais.

A estrutura secundária para o gênero *Ramphastos* foi estimada através de justaposição dos 2 alinhamentos prévios mencionados acima (porção inicial/mediana + distal), obtida através do programa ClustalW (Thompson *et al.*, 1994) com as opções-padrão, e posterior comparação deste alinhamento prévio com o consenso para aves publicado por Espinosa de los Monteros (2003); este autor, por sua vez, se utilizou de uma ampla base de dados para a estrutura secundária do 12S existente para diferentes grupos animais (Neefs *et al.*, 1991; Van de Peer *et al.*, 1994, 1997). As posições das hastes foram anotadas com ajuda do programa BioEdit 7.0.5.2 (Hall, 1999). Em regiões do alinhamento onde a definição das hastes não era tão clara, tais regiões relativas ao táxon *R. sulfuratus* R-260 foram submetidas ao site <http://rna.tbi.univie.ac.at/cgi-bin/RFold.cgi>, que executa o programa RNAfold *online* (ele calcula a estrutura secundária total, ou de apenas parte de uma seqüência fornecida, com base no menor valor de energia livre, a partir de um algoritmo que leva em consideração propriedades termodinâmicas do RNA), o que ajudou na inferência posicional dessas hastes. A representação gráfica da estrutura secundária do 12S foi realizada com auxílio do programa RnaViz 2.0 (De Rijk, 2003).

Todas as considerações deste ponto em diante se utilizam unicamente do alinhamento envolvendo a parte inicial + mediana do gene 12S mitocondrial de tucanos, que por simplificação será daqui em diante denominado *12S_Total*.

Posições ambíguas no alinhamento relativas às regiões de alças, ainda presentes depois de estabelecida a seqüência secundária, foram descartadas. As composições nucleotídicas para o *12S_Total*, apenas para a região de hastes, e apenas considerando-se a região de alças (incluindo os *bulges*, bases pertencentes a uma haste, mas que não pareiam), foram computadas usando-se o teste de χ^2 implementado no PAUP* 4b10 (Swofford, 2002), incluindo-se apenas um indivíduo por táxon, para minimização de efeitos de pseudo-replicação.

Gráficos de saturação foram calculados com auxílio do programa PATRISTIC (Fourment & Gibbs, 2006), colocando-se na abcissa os valores da distância patrística de máxima verossimilhança (calculada a partir do modelo que assume independência de todos os sítios, inclusive nas hastes) obtidos com o *12S_Total*, e na ordenada os valores de distância-p para alças e hastes.

Todas as análises filogenéticas por MP foram realizadas no PAUP* 4b10; as análises de MV, pelos programas PAUP* 4b10, TREEFINDER (Jobb, 2006) e HyPhy (Kosakovsky Pond *et al.*, 2005); e AB pelo programa MrBayes 3.1 (Ronquist & Huelsenbeck, 2003).

Árvores de MP foram calculadas para o *12S_Total*; para alças + hastes 5' (ou hastes proximais) apenas; para alças + hastes 3' (ou hastes distais) apenas; e para o *12S_Total* atribuindo-se um peso de 0,5 aos sítios que ocorrem em hastes. A busca pela melhor árvore de parcimônia se deu por busca exata do tipo *branch-and-bound*, e com *indels* sendo considerados um quinto caráter. As opções restantes não foram alteradas. Para cada réplica de *bootstrap* de MP, foi empregada uma busca heurística com 10 réplicas de adição aleatória de táxons e rearranjo de ramos por TBR, sendo o total de pseudo-réplicas de *bootstrap* igual a 10.000, para minimizar o efeito que variações estocásticas desse tipo de análise pudessem exercer sobre as estimativas de suporte de ramos.

No caso de MV, as três primeiras partições do conjunto de dados citadas acima foram usadas. O melhor modelo para cada partição foi aquele indicado pelo programa

Modeltest 3.7 (Posada & Crandall, 1998), empregando-se o AICc (este é o AIC corrigido para proporções relativamente pequenas entre o número amostral e o número de parâmetros livres do modelo) como critério de escolha. A busca heurística pela árvore de MV se deu através dos seguintes passos, para cada conjunto de dados: I. execução, no PAUP*, da linha de comando gerada pelo programa Modeltest; II. busca heurística utilizando uma árvore inicial obtida por NJ, com troca de ramos por SPR, utilizando os valores de parâmetros obtidos no passo anterior; III. otimização de todos os parâmetros para o modelo escolhido, mantendo-se a árvore de MV estimada pelo passo anterior fixa; IV. retorno ao passo II, e continuação do ciclo até que a topologia e as estimativas de MV para todos os parâmetros não mais se alterassem entre iterações sucessivas; V. busca heurística final, com 20 réplicas de adição aleatória de táxons, e troca de ramos por TBR. A análise de *bootstrap* para MV de cada partição foi realizada com o auxílio do programa TREEFINDER, sendo otimizados em cada réplica a topologia e os valores de todos os parâmetros do modelo. O número de réplicas foi sempre igual a 500.

Para a análise bayesiana, foi utilizado o programa MrBayes, com o emprego de um modelo misto (ABm): um sub-modelo (de acordo com o AICc) para alças, e um sub-modelo para hastes (modelo *doublet*). Este último sub-modelo considera os 16 estados de pares possíveis nas análises, não sendo possível a utilização de modelos para hastes com menor número de parâmetros nesse programa. O suporte foi obtido por estimativas de probabilidade posterior (PP), para cada ramo, a partir da fase estacionária das cadeias de *Markov Chain Monte Carlo* (MCMC). Quatro análises independentes, de 2.500.000 de gerações cada, foram implementadas, sendo a convergência de cada uma, assim como os respectivos períodos de *burn-in*, conferidos através do programa Tracer 1.3 (Rambaut & Drummond, 2006). Cada corrida apresentou 1 cadeia “fria” e 3 cadeias “quentes”.

As topologias obtidas por consenso de *bootstrap* $\geq 70\%$, ou por PP $\geq 95\%$ no caso de AB, foram comparadas com a árvore de MV (assumindo independência) para o *12S_Total*, e com a filogenia obtida por Weckstein (2005), mostrada na figura 2.1.

Figura 2.1 - Cladograma obtido por Weckstein (2005) para as espécies de *Ramphastos*, a partir de 3 genes codificantes mitocondriais (partes do COI e citocromo-b, e ND2 integral - total: 2468 bases). Legendas - R.: *Ramphastos*; a.: *ambiguus*; s.: *sulfuratus*; t.: *tucanus*; v.: *vitellinus*; SE: sudeste do Brasil; N: norte do Brasil. A análise filogenética original possui mais que um indivíduo para muitas das espécies e subespécies, mas por simplicidade apenas um indivíduo por táxon foi incluído nesta figura.

No programa HyPhy, foram empreendidos dois modelos para analisar algumas das propriedades relacionadas aos sítios pareantes: o modelo 6A (Savill *et al.*, 2001), e o modelo RNA_8, aqui proposto, ambos baseados apenas nos pareamentos de maior frequência (respectivamente 6 e 8 estados), para evitar um excesso de parametrização. Ambos tiveram seu ajuste aos dados calculado por AICc na topologia encontrada por Weckstein (2005). O modelo RNA_8 está descrito na tabela 2.2.

Tabela 2.2: Matriz de taxas do modelo RNA_8.

	AC	AU	CA	CG	GC	GU	UA	UG
AC	*	$\pi_{AT}^*S^*\mu$	0	0	$\pi_{GC}^*S^*\mu$	0	0	0
AU	$\pi_{AC}^*S^*\mu$	*	0	$\pi_{CG}^*\mu$	$\pi_{GC}^*D^*\mu$	$\pi_{GT}^*S^*\mu$	$\pi_{TA}^*\mu$	0
CA	0	0	*	$\pi_{CG}^*S^*\mu$	0	0	$\pi_{TA}^*S^*\mu$	0
CG	0	$\pi_{AT}^*\mu$	$\pi_{CA}^*S^*\mu$	*	$\pi_{GC}^*\mu$	0	$\pi_{TA}^*D^*\mu$	$\pi_{TG}^*S^*\mu$
GC	$\pi_{AC}^*S^*\mu$	$\pi_{AT}^*D^*\mu$	0	$\pi_{CG}^*\mu$	*	$\pi_{GT}^*S^*\mu$	$\pi_{TA}^*\mu$	0
GU	0	$\pi_{AT}^*S^*\mu$	0	0	$\pi_{GC}^*S^*\mu$	*	0	0
UA	0	$\pi_{AT}^*\mu$	$\pi_{CA}^*S^*\mu$	$\pi_{CG}^*D^*\mu$	$\pi_{GC}^*\mu$	0	*	$\pi_{TG}^*S^*\mu$
UG	0	0	0	$\pi_{CG}^*S^*\mu$	0	0	$\pi_{TA}^*S^*\mu$	*

Modelo RNA_8 (desenvolvido através do programa HyPhy) - características:

- **8 x 8 estados:** o estado AA, por ter frequência baixíssima (<0,1%), e os estados restantes, não-detectados, foram deixados de fora do modelo;
- **μ :** taxa de transversão dupla (mas apenas entre pares canônicos);
- **S:** taxa de transição simples (apenas um dos membros do par sofre uma transição);
- **D:** taxa de transição dupla (mas apenas entre pares canônicos);
- **π_{ij} :** frequência do estado I:J;
- nenhum outro tipo de passo é permitido em um intervalo infinitesimal de μt (onde μ é a taxa total de substituição, e t é o tempo);
- o valor de μ (taxa de transversão dupla) é otimizado para cada ramo (variável local); S e D são variáveis globais (todos os ramos compartilham o mesmo valor para ambos os parâmetros);
- todas as variáveis do modelo foram estimadas por MV, exceto pelas frequências das 8 classes (π_{ij}), que foram estimadas empiricamente a partir dos dados.

O modelo RNA_8 foi desenvolvido a fim de se evitar parametrização em excesso. Tal decisão faz sentido porque o número de sítios nas regiões de hastes é relativamente pequeno (196 pares), e assim o aumento da variância associada a um maior número de parâmetros poderia incorrer em diminuição excessiva na precisão de estimativas. Apesar de inédito, o RNA_8 foi baseado em extensão e adaptação do modelo 6B existente em Savill e col. (2001).

2.4 Resultados

2.4.1 - Evitando artefatos de amplificação

Uma preocupação pertinente, ligada à obtenção das seqüências, foi com a eventual amplificação (ou co-amplificação) de regiões não-ortólogas ao 12S mitocondrial (Bensasson *et al.*, 2001). As seguintes observações aumentam a verossimilhança de que uma mesma região ortóloga tenha sido amplificada e seqüenciada para todos os táxons (embora não neguem a possibilidade de que esta não represente o gene 12S - Collura & Stewart, 1995; Zhang & Hewitt, 1996; Sorenson & Quinn, 1998): a) a grande maioria dos táxons utilizados teve o DNA amplificado a partir de tecido rico em mitocôndrias; b) não foi detectada mais que uma banda por PCR; c) não houve sinais de mutações pontuais ou de *indels* nas regiões de sobreposição de seqüências a partir de amplificações de trechos diferentes do 12S; d) os seqüenciamentos não apresentaram picos duplos de leitura; e) um fragmento único englobando as porções inicial e mediana do gene foi amplificado e seqüenciado para alguns espécimes, e comparado com os fragmentos obtidos por amplificação em duas etapas, não tendo sido detectadas diferenças entre o fragmento único e os fragmentos parciais.

2.4.2 - Estrutura primária

A estrutura primária da molécula de 12S, para o táxon *R. sulfuratus* R-260, apresentou um total de 962 bases, situando-se dentro da faixa de tamanho observada para outras aves (Espinosa de los Monteros, 2003). O alinhamento obtido para todos os ranfastídeos baseou-se apenas nas regiões proximal e mediana do gene (*12S_Total*, 862 sítios no total), sendo o alinhamento relativo à porção terminal do gene descartada por ter sido seqüenciada apenas para poucos táxons (122 sítios) - tabela 2.1.

O alinhamento *12S_Total* apresentou, ao final da edição manual, 824 sítios (tendo sido descartados 33 sítios ambíguos do alinhamento, e 5 sítios da haste 24 que não apresentaram seus respectivos pares no *12S_Total*), com uma média de 817,53 bases por táxon. A composição geral do alinhamento baseado no *12S_Total* está descrita na tabela

2.3, que também indica o p-value obtido para o teste χ^2 de homogeneidade de composição de bases para cada partição.

Tabela 2.3: Frequências nucleotídicas por partição do 12S. O número de sítios, na última coluna, é o valor médio levando-se em conta todas as seqüências.

região	A	C	G	T	p-value	n° sítios
hastes	0.21	0.29	0.29	0.21	1.00	391.29
alças	0.40	0.31	0.13	0.16	1.00	426.24
12S_Total	0.31	0.30	0.20	0.19	1.00	817.53

Para o *12S_Total*, assim como para alças e hastes isoladamente, os números de caracteres constantes, não-informativos, e informativos para parcimônia foram calculados pelo programa PAUP. Os resultados estão mostrados na tabela 2.4.

Tabela 2.4: Números de caracteres constantes, não-informativos e informativos para o 12S, separados por partição.

Partição	Nº caracteres	constantes	não-informativos	informativos
hastes	392	340 (87%)	23 (6%)	29 (7%)
alças	432	302 (70%)	29 (7%)	101 (23%)
12S_Total	824	642 (78%)	52 (6%)	130 (16%)

2.4.3 - Estrutura secundária

A estrutura secundária do 12S para tucanos, baseada no espécime *R. sulfuratus* (LGEMA R-260), está mostrada na página seguinte, na figura 2.2.

Figura 2.2 - Estrutura secundária do 12S em tucanos (*em cinza* - número do sítio; *em vermelho* - número da haste; **pontos menores** - pareamentos canônicos; **pontos ressaltados** - pareamentos não-canônicos).

Figura 2.2 - continuação...

Na figura 2.3, está representado o histograma de frequências dos estados pareados mais comuns encontrados na estrutura secundária do 12S em tucanos, baseado no alinhamento do *12S_Total*. Os estados CC, UU, AG, GA e GG não foram observados.

Figura 2.3 - Histograma com as frequências dos estados pareados mais comuns nas regiões de hastes do *12s_Total*.

Os modelos escolhidos para cada partição, determinados pelo AICc no programa Modeltest, estão sumarizados na tabela 2.5. As reconstruções filogenéticas baseadas em MV, mostradas adiante nas figuras 2.5-2.7, se utilizaram dos modelos apropriados segundo essa tabela.

Tabela 2.5: Modelos markovianos de evolução molecular para diferentes partições do 12S, escolhidos pelo método AICc no programa Modeltest.

Partições (12S)	Modelo (AIC _c)
<i>12S_Total</i>	GTR + I + G
alças + hastes_5'	GTR + I + G
alças + hastes_3'	GTR + I + G
alças	TrN + I + G
hastes	K81uf + I

Legenda:

GTR - simétrico; matriz de taxas: **(abcdef)** - 6 taxas diferentes; frequências das bases estimadas por MV (Yang, 1994a).

I - proporção de sítios invariantes.

G - heterogeneidade de taxas ao longo da sequência; frequências de cada classe de taxa inferidas pela distribuição Gama discretizada, assumindo-se 4 classes.

TrN - simétrico; matriz de taxas: **(abaaca)** - 3 taxas diferentes; frequências das bases estimadas por MV (Tamura & Nei, 1993).

K81uf - simétrico; matriz de taxas: **(abccba)** - 3 taxas diferentes; frequências das bases estimadas por MV.

O gráfico de saturação, tanto para alças quanto para hastes (ainda assumindo independência entre os sítios parentais), pode ser observado na figura 2.4. No eixo x, estão os valores de distância patristica na árvore de MV usando-se o GTR+I+G para o *12S_Total* (tabela 2.5); e no eixo y, as distâncias-p para alças e hastes (não considerando-se covariação nestas últimas).

Figura 2.4 - Gráfico das distâncias-p de alças e hastes, em função das distâncias patristicas de MV para o *12S_Total* (modelo GTR+I+G).

Nas figuras 2.5-2.9 estão mostradas as reconstruções filogenéticas conforme explicado nas seções 2.1 e 2.3. Os valores de suporte para os ramos foram obtidos por *bootstrap* (no caso de MP e MV) ou por estimativas de probabilidade posterior (no caso de AB).

Figura 2.5 - Árvore de MV (*I2S_Total*), com valores de *bootstrap* de MP (acima dos ramos) e MV (abaixo dos ramos). Modelo de MV: GTR+I+G.

Figura 2.6 - Árvore de MV (alças + hastes 5'), com valores de *bootstrap* de MP (acima dos ramos) e MV (abaixo dos ramos). Modelo de MV: GTR+I+G.

Figura 2.7 - Árvore de MV (alças + hastes 3'), com valores de *bootstrap* de MP (acima dos ramos) e MV (abaixo dos ramos). Modelo de MV: GTR+I+G.

Figura 2.8 - Árvore de MP (*12S_Total*), com ponderação diferencial entre alças (peso=1.0) e hastes (peso=0.5), e valores de *bootstrap* para cada ramo.

Figura 2.9 - Árvore de AB (*12S_Total*), com modelo misto (ABm): alças - GTR+I+G; hastes - *doublet*-HKY+I. Este modelo considera os 16 estados possíveis de pares nas hastes, e assume que apenas 1 substituição simples é possível num intervalo infinitesimal de μt , sendo este passo proporcional à frequência do novo par (seja ele canônico ou não) e à taxa de substituição simples entre esses 2 estados (ex: $p[AA \rightarrow AG] = f(AG) \cdot ti$, onde ti é a taxa de transição simples, assumindo-se o modelo HKY) - Schöniger & von Haeseler, 1994.

Um modelo com apenas os seis estados mais comuns (“6A”) foi também empregado, para comparar com os resultados obtidos usando o RNA_8. A matriz de taxas instantâneas para o modelo 6A (Savill *et al.*, 2001) está mostrada na tabela 2.6, e a matriz

para o RNA_8 na tabela 2.7. Em ambos os casos, as taxas são dadas em referência à classe UA→UG, que foi fixada em 1,0 para facilitar as comparações.

Tabela 2.6: Matriz de taxas instantâneas (modelo 6A), para a árvore de Weckstein (2005). Em vermelho, transições duplas; em negrito, transições simples.

	AU	CG	GC	GU	UA	UG
AU	*	0.000	2.445	1.267	0.000	0.000
CG	0.000	*	0.000	0.000	2.069	0.000
GC	1.740	0.000	*	1.307	0.000	0.000
GU	12.167	0.000	17.633	*	0.000	0.000
UA	0.000	2.847	0.000	0.000	*	1.000
UG	0.000	0.000	0.000	0.000	34.694	*

Tabela 2.7: Matriz de taxas instantâneas (modelo RNA_8), para a árvore de Weckstein (2005). Em vermelho, transições duplas; em azul, transversões duplas; em negrito, transições simples.

	AC	AU	CA	CG	GC	GU	UA	UG
AC	*	27.40	0.00	0.00	38.50	0.00	0.00	0.00
AU	1.34	*	0.00	0.06	1.63	2.85	0.04	0.00
CA	0.00	0.00	*	47.74	0.00	0.00	34.69	0.00
CG	0.00	0.03	1.91	*	0.05	0.00	1.47	1.00
GC	1.34	1.16	0.00	0.06	*	2.85	0.04	0.00
GU	0.00	27.40	0.00	0.00	38.50	*	0.00	0.00
UA	0.00	0.03	1.91	2.02	0.05	0.00	*	1.00
UG	0.00	0.00	0.00	47.74	0.00	0.00	34.69	*

Os valores de AICc, para os três modelos usados nas hastes (K81uf+I, 6A e RNA_8), estão mostrados na tabela 2.8, onde K representa o número de parâmetros livres do modelo, e n o número amostral ^{*}.

^{*} A maneira aqui utilizada de se considerar o número amostral ($n = \text{nº sítios_independentes} \times \text{nº terminais}$) se baseou na discussão apresentada por Posada & Buckley (2004), e numa das sugestões de uso existentes no manual do Modeltest.

Tabela 2.8: Valores de AICc para os diferentes modelos usados para testar o ajuste a hastes. (topologia=Weckstein, 2005).

Modelo	Ln(L)	K	n	AICc
K81uf+I (independência)	-952.7088	63	11760	2032.11
6A	-356.7254	76	5160	867.75
RNA_8	-571.5792	66	5670	1276.74

$$AIC = -2 \ln L + 2 K$$

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1}$$

K(K81uf+I): 57 ramos + 3 freqs + 2 taxas + I = 63.

n(K81uf+I): 392 caracteres independentes * 30 táxons = 11.760.

K(6A): 57 ramos + 5 freqs + 14 taxas = 76.

n(6A): 172 caracteres independentes * 30 táxons = 5.160.

K(RNA_8): 57 ramos + 7 freqs + 2 taxas = 66.

n(RNA_8): 189 caracteres independentes * 30 táxons = 5.670.

Freqs - frequência de cada um dos estados (4 para sítios independentes; 6 ou 8 para pareados).

A figura 2.10 apresenta a correlação entre as distâncias patrísticas de alças e hastes, assumindo-se covariação de taxas nas últimas (através do modelo RNA_8).

Figura 2.10 - Gráfico comparando as distâncias patrísticas entre os modelos para alças e hastes, na topologia obtida por Weckstein (2005).

2.6 Discussão

2.6.1 - Estrutura primária do 12S

O teste do χ^2 no PAUP não detectou heterogeneidade na composição ($p = 1.00$), tanto para o *12S_Total*, quanto para alças e hastes quando estas foram examinadas isoladamente (tabela 2.3). As alças apresentaram composição de bases mais variável que as hastes. Houve equivalência de frequências entre C e G nas hastes, assim como entre A e U, evidenciando a seleção estabilizadora para pareamentos nessas regiões. Considerando-se o *12S_Total*, o nucleotídeo mais frequente foi A (31%), seguido de C (30%), G (20%) e U (19%). Essa sequência de estados é congruente com a obtida por Webb e Moore (2005) para pica-paus (que pertencem à ordem Piciformes, assim como a família Ramphastidae). A ordem obtida por Espinosa de los Monteros (2003) foi um pouco diferente, com U sendo ligeiramente mais frequente que G em média. No entanto, essa diferença observada entre G e U foi pequena (respectivamente, 20,2% e 20,6%), sendo que em 45 das 100 espécies estudadas pelo autor a frequência de G foi maior. Nas alças, a quantidade de A (40%) foi a maior, seguida de C (31%), U (16%) e G (13%); já nas hastes, C e G são os mais frequentes, com 29% cada, e A e U com 21% cada. Essas ordens relativas de frequências (tanto para alças quanto para hastes) também corroboram o estudo de Webb e Moore (2005). Quanto a outros grupos de vertebrados, tais resultados estão em acordo com estudos baseados em peixes (Wang & Lee, 2002), répteis (serpentes do grupo *Thamnophiini* - Alfaro & Arnold, 2001), e mamíferos (Springer & Douzery, 1996). Uma maior abundância de adenina nas alças foi hipotetizada por Gutell e col. (1985) como estando relacionada a interações hidrofóbicas com as proteínas ribossomais, já que adenina é o nucleotídeo menos polar; ou ainda isso pode ser devido a um enviesamento mutacional subjacente ao gene (Springer *et al.*, 1995). Por outro lado, uma maior proporção de GC nas hastes (58%) que nas alças (44%) pode ser explicada pelo fato de ligações GC possuírem um menor valor de energia livre do que as ligações AU (Turner *et al.*, 1988), aumentando assim a estabilidade nas hastes.

2.6.2 - Estrutura secundária do 12S

O número total de sítios pareantes foi 196 (392, se contados os sítios independentemente). Os estados mais freqüentes nas hastes foram, respectivamente, os pares canônicos CG (30,53%), GC (24,61%), UA (22,19%) e AU (17,51%), somando 94,84% das freqüências totais (figura 2.3). Foram observados, em proporção bem menor, os pares não-canônicos GU (1,81%), CA (1,19%), AC (0,82%), UG (0,63%), UC (0,51%), CU (0,14%); e o estado AA, que ocorreu com freqüência extremamente baixa (~0,05%). Não houve ocorrência dos demais estados no conjunto de dados analisado. Somente os 8 estados mais freqüentes foram considerados na modelagem do RNA_8 (99,3% da freqüência total). Os sítios que apresentaram os demais estados foram excluídos do alinhamento automaticamente pelo programa HyPhy (num total de 7 sítios excluídos). Já no caso do modelo 6A, foram desconsiderados das análises os dez estados menos freqüentes, para um total de 24 sítios removidos (sendo considerados nas análises 97% da freqüência total de sítios pareados).

Indels foram observados apenas em alças; sua presença em algumas regiões (em geral ricas em repetições de citosina) levou à determinação de 33 sítios ambíguos no alinhamento, que foram descartados das análises; além desses, 5 sítios pertencentes à haste 24 foram excluídos dos dados, pois a haste 3' associada não se mostrou presente no 12S_Total.

As 4 hélices (ou domínios) obtidas por Espinosa de los Monteros (2003) foram identificadas, num total de 40 hastes e 37 alças (internas e externas). A maioria dessas estruturas foi inferida da seguinte forma: I. busca, na seqüência relativa a *R. sulfuratus* R-260, das regiões conservadas descritas por Espinosa de los Monteros (2003); II. utilização de tais regiões como demarcadores de posição relativa, e a partir daí identificou-se sucessivamente as hastes esperadas, até que a simples observação e comparação com a estrutura proposta pelo autor acima já não pudesse mais esclarecer a posição de algumas hastes (por exemplo, devido a *indels* relativamente extensos em alças vizinhas à haste procurada); III. nesses casos onde a inspeção manual foi mais difícil, submeteu-se a região de interesse (do táxon *R. sulfuratus* R-260) à implementação online do programa RNAfold, que inferiu a melhor configuração da estrutura secundária baseada em um algoritmo que

considera como um de seus parâmetros a minimização da energia livre. IV. posteriormente, buscou-se corroborar a presença das hastes inferidas através da identificação de covariações nas seqüências (ex: num determinado sítio, a maioria dos táxons possui A:U; mas alguns deles apresentam C:G) - porém, para algumas dessas hastes, todos os pareamentos encontrados se apresentaram invariantes, como mencionado abaixo.

De forma geral, a inferência da estrutura secundária do 12S se baseou principalmente no uso dos passos I, II e IV; o passo III foi necessário apenas para o estabelecimento dos pareamentos e alças situados em algumas regiões das hélices II e III (figura 2.2). No caso da hélice II, para a haste 16, não houve covariação detectada (sítios invariáveis); a haste 17 foi a mais problemática da hélice: não há sinal claro de covariação, alguns pares não-canônicos são detectados, e na 6ª posição da haste 5' foram observados dois táxons (*S. reinwardtii* e *A. prasinus*) com o pareamento A:A. Isso pode ser um sinal de que a estimativa da haste 17 como um todo foi errônea, que a haste está corretamente identificada mas a 6ª posição (e suposto par na haste 3') é um *bulge* (sítio não-pareado localizado numa zona de haste), ou então que esta é uma haste legítima, mas que por algum motivo apresenta um relaxamento na pressão para pareamento naquela posição (seja em muitas das linhagens, ou apenas em algumas, dentre elas aquelas em que foi observado o estado A:A). Um outro caso de pareamento A:A foi observado na 6ª posição (porção 5') da haste 4 para o táxon *B. bailloni*, mas nesse caso foi detectada covariação na 1ª posição da mesma; entretanto, as explicações possíveis referentes à haste 17 também podem se aplicar aqui. A haste 18 apresentou sinais claros de covariação na 2ª posição (porção 5'). A haste 19 não mostrou covariação, sendo todas as quatro posições invariantes, exceto pelo táxon *B. bailloni*, que tem um C na 3ª posição (5') pareando com A, enquanto todos os outros táxons apresentam U:A. Espinosa de los Monteros (2000) também obteve dificuldade em localizar as hastes 17-21 em sua análise envolvendo diferentes grupos de aves, admitindo que não era possível estabelecer homologias posicionais claras para esse segmento em comparações com uma inferência prévia de estrutura secundária para mamíferos.

Na hélice III também há regiões cuja atribuição às classes corretas foi problemática (hastes 31-34). A haste 31 não mostra sinais claros de covariação, com alguns pareamentos não-canônicos presentes. A haste 32 possui covariação em algum grau em uma das

posições, também com alguns pareamentos não-canônicos, sendo as outras duas posições invariantes. A haste 33 tem um sítio covariante, os restantes sendo conservados, o que é um indício razoável de sua real existência. A haste 34 mostra apenas quatro pareamentos, de sítios invariantes, enquanto o consenso do autor acima apresentou seis; um eventual pareamento adicional pode ser invocado entre o sítio imediatamente anterior à haste 5' e o imediatamente posterior à haste 3', pois há covariação observável nos 30 táxons - porém, tais pareamentos eventuais são sempre na forma C:U ou U:C (ou C:A nos espécimes de *R. dicolorus*), que por não serem canônicos, e além disso localizados em um dos extremos da haste (onde os pareamentos supostamente devem ser mais fortes, portanto preferencialmente canônicos), levantam dúvidas quanto à sua real condição, tendo sido aqui considerados como sítios relacionados a alças. O fim da região 3' da hélice III, e toda a hélice IV (que juntos equivalem à porção distal do 12S, mencionada na seção 2.3) são regiões para as quais só seis táxons apresentaram sequência disponível (tabela 2.1), para um alinhamento total de 122 sítios de DNA (10 sítios no fim da hélice III, mais 112 sítios da hélice IV). Em tal alinhamento não foi observada covariação de pareamentos em quaisquer das hastes, mas no entanto a inferência de tais estruturas, baseada em comparações com a estrutura estimada pelo autor acima, foi relativamente simples.

É importante ressaltar que não foi detectada covariação em muitas das hastes inferidas, mesmo naquelas considerados relativamente acuradas (dado o grau de conservação entre *R. sulfuratus* R-260 e o consenso de Espinosa de los Monteros, 2003), devido, provavelmente, ao fato do nível de divergência estudado não ser muito abrangente num sentido temporal (o foco do estudo foram espécies dentro de um gênero).

2.6.3 - Estrutura secundária e reconstruções filogenéticas

Os RNAs ribossômicos desempenham um papel importante durante a síntese protéica (Dahlberg, 1989), sendo esta função fortemente dependente da estrutura molecular (Noller, 1984). Assim sendo, a consideração da estrutura secundária é muito importante para que o alinhamento final represente de forma mais acurada as relações de homologia entre as linhagens estudadas (Morrison & Ellis, 1997; Hickson *et al.*, 2000). A importância do uso de tal informação é maior em estudos onde os táxons sejam relativamente mais

distantes entre si, pois nesses casos é de se esperar que a quantidade de *indels* seja maior ao longo do alinhamento. No presente estudo, não houve diferenças quanto à topologia final por MP estimada usando-se ClustalW e a estimada a partir da estrutura secundária (dados não mostrados), em ambos os casos descartando-se as regiões ambíguas (35 e 33 sítios descartados, respectivamente). Isto é um sinal de que as ambigüidades observadas no alinhamento obtido ainda não são suficientes, num contexto evolutivo, para causar uma diferença nas estimativas topológicas, sejam elas baseadas ou não na estrutura secundária.

Três abordagens foram aqui empregadas para incluir a informação acerca da covariação entre taxas nas hastes, e testar o quanto esse fator poderia influir na resolução topológica: 1) MP e MV com utilização alternada de {alças + hastes 5'} (1A) e {alças + hastes 3'} (1B); 2) MP com pesagem menor dos sítios localizados em hastes (peso=0,5) em relação às alças (peso=1,0); 3) AB usando um modelo misto (ABm), que assumiu covariância na evolução dos sítios em hastes, e independência nos sítios em alças.

Como medida de comparação da acurácia entre os diferentes procedimentos, assumiu-se como hipótese filogenética mais próxima da realidade a árvore de Weckstein (2005) - figura 2.1. O esperado seria que a acurácia topológica das 3 diferentes implementações ocorresse na seguinte ordem: (1? >) 3 > (> 1?) > 2 (> 1?) > MV tratando os sítios independentemente > MP tratando os sítios independentemente. O resultado esperado acabou não ocorrendo, uma vez que todas as abordagens geraram topologias semelhantes quando utilizado um consenso de *bootstrap* $\geq 70\%$.

Poderia-se dizer que o método 1 serviu para testar se a covariação entre sítios de hastes seria igual a 1,0 - caso este não fosse o caso, a conclusão seria que existe assimetria entre as hastes 5' e as hastes 3'; no entanto, a própria análise fina do alinhamento, necessária para a identificação de alças e hastes, já mostrava que covariação < 1, evidenciando a presença de pareamentos não-canônicos. Desta feita, o método 1 só serviu para testar se a assimetria entre as porções 5' e 3' das hastes, seja ela estocástica ou não, seria responsável por uma resolução topológica diferencial entre 1A e 1B. Essa diferença realmente existiu (figuras 2.6 e 2.7); entretanto, para testar se tal assimetria é devida a eventuais fatores seletivos ou não, novos conjuntos de táxons deveriam ser analisados; caso em grande parte deles fosse observada uma maior taxa de evolução por parte de uma das

porções (ex: hastes 3'), isso seria um indício de pressão seletiva induzindo assimetria (talvez devido a algum tipo de necessidade de manutenção da estrutura terciária da molécula).

O método 2 permite que a questão da co-variação seja tratada de forma simples e relativamente rápida, atribuindo-se peso=0,5 aos sítios localizados em hastes. No entanto, há um pressuposto importante assumido para esse procedimento: o de que a covariação é igual a 1,0 em todos os sítios, e portanto simétrica em cada par. Dixon & Hillis (1993) sugeriram peso=0,8 para as hastes, tendo esta ponderação sido mais acurada que um peso=0,5 para o conjunto de dados analisado por eles. Já Springer e col. (1995) sugeriram peso=0.61 para as hastes. Ambos foram tentados com o *12S_Total* (dados não mostrados), gerando árvores de *bootstrap* $\geq 70\%$ idênticas à obtida com peso=0,5 (figura 2.8). Outros pesquisadores, por outro lado, advogam que a pesagem deveria ser feita da maneira oposta, atribuindo-se um peso maior às posições mais conservadas, por elas estarem menos sujeitas a eventos de homoplasia (Hillis *et al.*, 1993; Miyamoto *et al.*, 1994).

O método 3 implica em modelização explícita dos parâmetros para cada sub-modelo; implica, também, no uso de MCMC para aproximação da distribuição bayesiana posterior. Para a análise bayesiana no MrBayes, atribuiu-se um maior número de parâmetros para o sub-modelo relativo às alças, em relação a aqueles que foram escolhidos pelo AICc no Modeltest (tabela 2.5); procedeu-se assim devido a uma restrição da versão atual do programa, que não possui internamente muitos dos modelos existentes no Modeltest. Para as alças, o modelo escolhido pelo Modeltest foi o TrN+I+G; para hastes, K81uf+I para a região 5', e K80+I para a região 3' (dados não mostrados). A sugestão dos autores do MrBayes (Huelsenbeck & Ronquist, 2004) é a adoção de um modelo mais complexo que aquele sugerido pelo Modeltest, caso este último não esteja disponível no MrBayes. Assim, utilizou-se o GTR+I+G para as alças; para as hastes, o modelo seria então o GTR+I *doublet*, mas para evitar um excesso de parametrização devido a serem só 196 sítios pareados, preferiu-se o HKY+I *doublet*.

A comparação entre as topologias apresentadas, seja quando considerado um consenso de *bootstrap* $\geq 70\%$ (figuras 2.5 a 2.8), ou então PP $\geq 95\%$ (figura 2.9), mostrou resultados bastante similares entre si, sempre envolvendo uma politomia na base do grupo

interno com *R. toco*, *R. dicolorus*, *R. sulfuratus*, *R. tucanus ssp.*, *R. ambiguus ssp.* e (*R. brevis* + *R. vitellinus ssp.*); não foi observado suporte por *bootstrap* de parcimônia para o ramo separando *R. brevis* dos *R. vitellinus* da Amazônia (*vitellinus*, *culminatus* e *ariel N*), e nem sustentando os 2 espécimes de *R. v. vitellinus* como um clado monofilético. O primeiro caso, no entanto, apresenta valores altos de *bootstrap* de MV, sempre $\geq 90\%$. De forma geral, portanto, os resultados foram bastante congruentes, a despeito do fato de terem sido empregados métodos filogenéticos diferentes, e maneiras diversas de contornar a questão da estrutura secundária - sempre houve uma grande politomia na parte basal do grupo interno. As árvores de consenso obtidas são também muito semelhantes em relação à árvore de MV baseada no *12S_Total*. A árvore obtida por Weckstein (2005) mostra congruência com as árvores de consenso obtidas, embora a resolução destas seja bem menor. Alfaro & Arnold (2001) obtiveram um resultado semelhante, chegando à conclusão de que o 12S possui menor utilidade para reconstrução filogenética que quaisquer dos genes codificantes por eles analisados. Isso pode ser interpretado como evidência de que o uso do *12S_Total*, isoladamente, é insuficiente para resgatar relações filogenéticas em todo o espectro de divergências existentes na topologia real, seja por apresentar uma quantidade de bases no alinhamento relativamente pequena para inferências topológicas, e/ou então devido a fatores ligados à taxa de evolução do gene no grupo.

2.6.4 - Evolução molecular em alças e hastes

A tabela 2.4 mostra que a proporção de sítios informativos para parcimônia é bem maior nas alças (23%) que nas hastes (7%); estas últimas também apresentam a maior proporção de sítios constantes (87%, contra 70% nas alças). Tais observações, em conjunção com os gráficos das figuras 2.4 e 2.10, evidenciam que as hastes evoluem a uma taxa mais baixa que as alças.

O gráfico da figura 2.4 mostra a curva de saturação para alças e hastes. No eixo x foram consideradas as distâncias patrísticas de MV (que equivalem à soma dos tamanhos dos ramos ligando duas linhagens numa dada árvore filogenética, aqui no caso obtida por MV - Fourment & Gibbs, 2006) baseadas no modelo GTR+I+G para o *12S_Total*, já que estas distâncias correspondem apenas ao componente de herança evolutiva em comum

entre dois táxons, sendo desconsiderada a fração associada a eventos homoplásticos (Avise, 2004); e no eixo y foram consideradas as distâncias percentuais par-a-par não corrigidas, ou distâncias-p, para alças e hastes separadamente. É importante utilizar o modelo indicado por algum índice estatístico (no caso, o AICc) como estimativa mais adequada de correção das distâncias-p, pois em casos onde isso não é feito (ou seja, se um modelo sub-parametrizado é usado), a correção inferida para substituições múltiplas pode ser subestimada, em alguns casos deixando de detectar graus consideráveis de saturação, caso ela exista (Sullivan & Joyce, 2005). Para as alças, algum grau de saturação começa a ser detectado na faixa de 12% de divergência entre as seqüências. Em comparação, Espinosa de los Monteros, 2000, ao estudar um conjunto de dados incluindo diferentes grupos de aves, começou a detectar esse efeito a partir dos 15% de divergência, um valor um pouco acima do encontrado no presente estudo. No caso das hastes, esse efeito começa a ser observável a aproximadamente 5%. Tal observação evidencia que apesar da taxa relativa dessa partição ser menor que a das alças, homoplasias já estariam começando a ocorrer de maneira significativa. Esse fato, associado à observação de que a proporção de sítios constantes é bem maior nas hastes (87%) que nas alças (70%) (tabela 2.4), pode ser atribuído à seleção purificadora atuando na manutenção da estrutura secundária do 12S. Não obstante, o gráfico da figura 2.4 não deve ser interpretado como a melhor medida de saturação, já que todos os sítios de hastes (392 bases) foram incorporados na estimativa, ao invés de serem incluídos apenas os valores de distância-p referentes às classes de dinucleotídeos (196 no total).

Assim sendo, para estimar de forma mais acurada a proporcionalidade entre as taxas em alças e hastes, foram empregados os modelos 6A e RNA_8 para as hastes, comparando-se a correlação entre cada um deles, separadamente, com o modelo TrN+I+G para alças. Para esta finalidade, empregou-se inicialmente o AICc para testar se esses modelos que assumem correlação nas hastes são de fato mais satisfatórios que o modelo K81uf+I, que assume independência nessas regiões (tabela 2.8). Outros modelos mais complexos que o 6A ou o RNA_8, que também consideram correlação entre taxas, poderiam ter sido testados; no entanto, optou-se pela utilização de modelos com (relativamente) pouca parametrização/número de dimensões, visando equilibrar um nível de plausibilidade

biológico aceitável com o número de parâmetros a serem estimados, devido ao número relativamente baixo de pareamentos nas hastes. Apesar dessas considerações, é importante salientar que Savill e col. (2001) demonstraram, para um conjunto de 5 seqüências de Eubacteria, e um total de 296 sítios pareados, que uma maior parametrização aumentou de forma consistente o ajuste do modelo aos dados, e que além disso não houve evidência que modelos com menor número de dimensões teriam um ajuste mais adequado que modelos com mais dimensões (o estudo se baseou no uso de diferentes modelos com 6, 7 e 16 dimensões).

Observou-se que ambos os modelos que levam em conta covariação de taxas nas hastes foram superiores ao modelo K81uf+I (tabela 2.8). O mais indicado pelo AICc foi o 6A. No entanto, a correlação entre o modelo 6A e o modelo para alças (TrN+I+G) não foi clara, gerando dados relativamente dispersos e um R^2 relativamente baixo, mesmo quando outro tipo de curva foi tentado; a taxa de alças em relação a hastes, para o 6A, variou de 10-15x nesse gráfico (dados não mostrados), portanto essa regressão não deve ser considerada confiável. Por outro lado, com a utilização do modelo RNA_8 obteve-se uma relação aparentemente linear, com $R^2=0,9094$. A figura 2.10, com o gráfico de correlação entre hastes (modelo RNA_8) e alças (TrN+I+G), indica que a taxa nas alças é 4x maior que nas hastes. Essa razão entre as taxas de evolução equivale ao dobro da obtida por Wang & Lee (2002) para peixes. Uma provável explicação é que uma maior taxa de substituição estaria ocorrendo nas alças do 12S de tucanos devido a estes serem animais homeotérmicos (e portanto possuírem um metabolismo mais elevado), enquanto os peixes não o são; mas, por outro lado, a seleção para manutenção dos pareamentos nas hastes seria intensa em ambos os grupos - assim sendo, seria de se esperar um relativo aumento da razão entre as taxas em aves (e provavelmente também em mamíferos), em relação à detectada em peixes.

Tal razão, obtida analisando-se a figura 2.4, indica que quando é desconsiderada a covariação nas hastes, tende a ter seu valor decrescido (~2,5 - 3,2x). Isso faz sentido, já que quando se considera independência de todos os sítios nas hastes, o efeito é similar ao de uma parte dos caracteres tendo um peso maior no modelo (no caso, K81uf+I), portanto inflando o valor da taxa evolutiva associada às hastes, e diminuindo assim a razão entre taxas evolutivas nas alças e nas hastes.

Como já mencionado acima, o modelo com melhor ajuste aos dados, segundo o AICc, foi o 6A (tabela 2.8). Não obstante, essa conclusão deve ser considerada com cuidado: o emprego do AIC não é recomendado como método de escolha quando os modelos comparados possuem dimensionalidade diferentes (Savill *et al.*, 2001). As tabelas 2.6 (6A) e 2.7 (RNA_8) mostram aspectos em comum quanto às taxas relativas nas hastes. Em primeiro lugar, percebe-se que as transições simples para pares canônicos são as mais comuns; as mesmas mudanças, nos sentidos contrários, possuem taxa bem menor. Em seguida, no modelo 6A, tem-se as duplas transições; no RNA_8, essas taxas também vêm logo depois das transições simples, entretanto são intercaladas por taxas de substituições simples de pares canônicos para não-canônicos. Estas últimas, por sua vez, são as taxas mais baixas encontradas no 6A. Já no RNA_8, as transversões duplas (que só ocorrem entre pares canônicos) são as com frequência mais rara (tais tipos de substituição não são permitidos no 6A). É importante notar que uma das taxas que ocorreram com valor mais alto em outros modelos (Savill *et al.*, 2001; Jow *et al.*, 2002), assim como no RNA_8, não foi observada no 6A (taxa UG-CG), o que deve ser um efeito causado pela exclusão, no programa HyPhy, dos 24 sítios com estados não-permitidos no modelo 6A.

A detecção de substituições duplas (transições e, em menor escala, transversões) era esperada já que pareamentos canônicos devem sofrer seleção estabilizadora nas hastes. Tillier & Collins (1998) e Higgs (2000), utilizando modelos que permitem substituições duplas, detectaram taxas relativamente altas das mesmas, mostrando ser verossímil o uso de tais modelos. Savill e col. (2001) testaram modelos que permitem e que não permitem duplas substituições, demonstrando que modelos apoiados no pressuposto de que estas ocorrem se ajustaram melhor aos dados por eles analisados.

É importante salientar, no entanto, que a premissa de substituições duplas ocorrendo num único passo não implicam na idealização de uma dupla mutação ocorrendo num mesmo indivíduo, que é então fixada - a chance de um caso assim ocorrer é muito rara. O conceito de dupla substituição é apenas um recurso teórico, útil para modelagens markovianas; uma explicação para a observação de estados covariantes, aparentemente adquiridos de forma sincrônica, seria a manutenção em baixa frequência de um estado não-canônico na população, ligeiramente deletério, até que uma mutação posterior que leve a

um novo estado canônico surgisse e fosse fixada, por exemplo por deriva (Kimura, 1985; Stephan, 1996; Higgs, 1998).

Referências bibliográficas - Capítulo 2

- Alfaro, M. E.; Arnold, S. J. (2001): Molecular systematics and evolution of *Regina* and the Thamnophiine snakes. *Mol. Phyl. Evol.* 21(3): 408-423.
- Avise, J. C. (2004): *Molecular markers, natural history, and evolution* - 2ª edição. Sinauer Associates, Inc., Sunderland, MA.
- Bensasson, D.; Zhang, D. X.; Hartl, D. L.; Hewitt, G. M. (2001): Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. & Evol.* 16(6): 314-321.
- Collura, R. V.; Stewart, C. B. (1995): Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* 378: 485-489.
- Dahlberg, A. E. (1989): The functional role of ribosomal RNA in protein synthesis. *Cell* 57: 525-529.
- De Rijk, P.; Wuyts, J.; De Wachter, R. (2003): RnaViz 2 - an improved representation of RNA secondary structure. *Bioinformatics* 19(2): 299-300.
- Dixon, M. T.; Hillis, D. M. (1993): Ribosomal RNA secondary structure - compensatory mutations and implications for phylogenetic analysis. *Mol. Biol. Evol.* 10(1): 256-267.
- Espinosa de los Monteros, A. (2000): Higher-level phylogeny of Trogoniformes. *Mol. Phyl. Evol.* 14(1): 20-34.
- Espinosa de los Monteros, A. (2003): Models of the Primary and Secondary Structure of the 12S RNA of Birds: A Guideline for Sequence Aligning. *DNA Sequence* 14: 241-256.
- Fourment, M.; Gibbs, M. J. (2006): PATRISTIC - a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol. Biol.* 6(1): 5 págs.

- Galtier, N. (2004): Sampling properties of the bootstrap support in molecular phylogeny - influence of nonindependence among sites. *Syst. Biol.* 53(1): 38-46.
- Gutell, R. R.; Weiser, B.; Woese, C.; Noller, H. F. (1985): Comparative anatomy of 16S-like ribosomal RNA. *Nuc. Acid. Res.* 32: 155-215.
- Hall, T. A. (1999): BioEdit - a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41: 95-98.
- Hickson, R. E.; Simon, C.; Perrey, S. W. (2000): The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. *Mol. Biol. Evol.* 17(4): 530-539.
- Higgs, P. G. (1998): Compensatory neutral mutations and the evolution of RNA. *Genetica* 102/103: 91-101.
- Higgs, P. G. (2000): RNA secondary structure: physical and computational aspects. *Q. Rev. Biophys.* 30 (3).
- Hillis, D. M. ; Allard, M. W.; Miyamoto, M. M.; (1993): Analysis of DNA sequence data: phylogenetic inference. *Methods Enzymol.* 224: 456-87.
- Huelsenbeck, J. P.; Ronquist, F. (2004): Bayesian analysis of molecular evolution using MrBayes. In: *Statistical methods in molecular evolution* (Ed: R. Nielsen). Springer Science+Business Media, Inc. New York, NY.
- Jobb, G. (2006): *TREEFINDER version of May 2006*. Munich, Germany. Distributed by the author at www.treefinder.de.
- Jow, H.; Hudelot, C.; Rattray, M.; Higgs, P. G. (2002): Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. *Mol. Biol. Evol.* 19(9): 1591-1601.
- Kimura, M. (1985): The role of compensatory neutral mutations in molecular evolution. *J. Genet.* 64: 7-19.
- Kocher, T. D.; Thomas, W. K.; Meyer, A.; Edwards, S. V.; Pääbo, S.; Vallabhanca, F. X., Wilson, A. C. (1989): Dynamics of mitochondrial DNA evolution in animals - amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci., USA* 86: 6196-6200.

- Kosakovsky Pong, S. D.; Frost, W.; Muse, S. V. (2005): HyPhy - Hypothesis testing using phylogenies. *Bioinformatics* 21: 676-679.
- Mindell, D. P.; Honeycutt, R. L. (1990): Ribosomal RNA in vertebrates - evolution and phylogenetic applications. *Ann. Rev. Ecol. Syst.* 21: 541-566.
- Miyamoto, M. M.; Allard, M. W.; Adkins, R. M.; Janecek, L. L.; Honeycutt, R. L. (1994): A congruence test of reliability using linked mitochondrial DNA sequences. *Syst. Biol.* 43: 236-249.
- Morrison, D. A.; Ellis, J. T. (1997): Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of Apicomplexa. *Mol. Biol. Evol.* 14: 428-441.
- Neefs, J. M.; Van de Peer, Y.; De Rijk, P.; Goris, A.; De Wachter, R. (1991): Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res.* 19: 1987-2015.
- Noller, H. F. (1984): Structure of ribosomal RNA. *Annu. Rev. Biochem.* 53: 119- 162.
- Posada, D.; Buckley, T. R. (2004): Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53(5): 793-808.
- Posada, D.; Crandall, K. A. (1998): Modeltest - testing the model of DNA substitution. *Bioinformatics* 14 (9): 817-818.
- Rambaut, A.; Drummond, A. (2006) Tracer v1.3 - A program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes.
- Ronquist, F.; Huelsenbeck, J. P. (2003): MrBayes 3 - Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
- Savill, N. J.; Hoyle, D. C.; Higgs, P. G. (2001): RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics* 157: 399-411.
- Schöniger, M.; von Haeseler, A. (1994): A stochastic model and the evolution of autocorrelated DNA sequences. *Mol. Phyl. Evol.* 3: 240-247.
- Simon, C.; Frati, F.; Beckenbach, A.; Crespi, B.; Liu, H.; Flook, P. (1994): Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* 87(6): 651-701.

- Sorenson, M. D.; Quinn, T. W. (1998): Numts: A challenge for avian systematics and population biology. *The Auk* 115: 214-221.
- Sorenson, M. D.; Ast, J. C.; Dimcheff, D. E.; Yuri, T.; Mindell, D. P. (1999): Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phyl. Evol.* 12(2): 105-114.
- Springer, M. S.; Douzery, E. (1996): Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. *J. Mol. Evol.* 43: 357-373.
- Springer, M. S.; Hollar, L. J.; Burk, A. (1995): Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. *Mol. Biol. Evol.* 12(6): 1138-1150.
- Stephan, W. (1996): The rate of compensatory evolution. *Genetics* 144: 419-426.
- Sullivan, J.; Joyce, P. (2005): Model selection in phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 36: 445-466.
- Swofford, D.L. (2002): PAUP* - Phylogenetic Analysis Using Parsimony (*and Other Methods). Versão 4.0b10. Sinauer Associates, Sunderland, MA.
- Tamura, K.; Nei, M. (1993): Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10: 512-526.
- Thompson, J. D.; Higgins, D. G.; Gibson, T. J. (1994): CLUSTAL W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nuc. Acid. Res.*, 22: 4673-4680.
- Tillier, E. R. M.; Collins, R. A. (1998): High apparent rate of simultaneous compensatory base-pair substitutions in ribosomal RNA. *Genetics* 148: 1993-2002.
- Turner, D. H.; Sugimoto, N.; Freier, S. M. (1988): RNA structure prediction. *Annu. Rev. Biophys. Chem.* 17:167-192.
- Van de Peer, Y; Van den Broeck, I.; De Rijk, P.; De Wachter, R. (1994): Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res.* 22(17): 3488-3494.
- Van de Peer, Y; Jansen, J.; De Rijk, P.; De Wachter, R. (1997): Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res.* 25(1): 111-116.

- Wang, H.-Y.; Lee, S.-C. (2002): Secondary structure of mitochondrial 12S rRNA among fish and its phylogenetic applications. *Mol. Biol. Evol.* 19(2): 138-148.
- Webb, D. M.; Moore, W. S. (2005): A phylogenetic analysis of woodpeckers and their allies using 12S, Cyt b, and COI nucleotide sequences (class Aves; order Piciformes). *Mol. Phyl. Evol.* 36: 233-248.
- Weckstein, J. D. (2005): Molecular phylogenetics of the *Ramphastos* toucans - implications for the evolution of morphology, vocalizations, and coloration. *Auk* 122(4): 1191-1209.
- Wheeler, W. C.; Honeycutt, R. L. (1988): Paired sequence difference in ribosomal RNAs: evolution and phylogenetic implications. *Mol. Biol. Evol.* 5:90-96.
- Yang, Z. (1994a): Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* 39:105-111.
- Zhang, D.-X.; Hewitt, G. M. (1996): Nuclear integrations - challenges for mitochondrial DNA markers. *Trends Ecol. Evol.* 11: 247-251.

Capítulo 3 - Reavaliação filogenética das espécies e subespécies do gênero *Ramphastos* e considerações biogeográficas

Abstract

The genus *Ramphastos* (Piciformes: Ramphastidae) is composed by the toucans. These birds are distributed throughout the neotropical region, being characterized by a large, robust and colorful beak. Historically, the systematics of this group has been approached by morphological and molecular studies, which disagree in some aspects of taxonomic and evolutionary relevance. Morphology pointed to a total of seven species inside the genus, with most of them further divided into subspecies. A subsequent molecular study, based on analyses of DNA sequence data, corroborated six out of those seven species, but indicated that the monotypic species *R. brevis* is inside the *R. vitellinus* complex, turning the latter into a paraphyletic group. Furthermore, the latter study did not support the southeastern and northern brazilian populations of *R. vitellinus ariel* forming a single monophyletic entity, even though they are identical regarding general coloration. In order to further shed light on such discordances, the present study shows data relative to eight mitochondrial regions and one nuclear intron, which were analysed by maximum parsimony, maximum likelihood and bayesian methods. Additionally, samples of the subspecies *R. vitellinus citreolaemus*, not present in the former molecular study, were included. The analysis of the concatenated alignment showed a great resolution in terms of topological positioning of all the species (and also most of the subspecies), and a strong congruence among the different phylogenetic methods, with all branches having bootstrap support $\geq 70\%$ and posterior probability estimates $\geq 95\%$. The monophyletic status of the *R. vitellinus* complex was re-established, appearing here as the sister group to *R. brevis*. The results also confirm that the populations of southeastern and northern *R. vitellinus ariel* are genetically distinct. *R. v. citreolaemus* is the sister group of the amazon morphotypes of *R. vitellinus ssp.* Divergence dates were estimated by four different methods, all of them converging to similar periods in the past: the separation between genres may have taken place in the Upper Oligocene/Upper Miocene range; between species, from the Upper Miocene to the Lower Pleistocene; and among subspecies, from the Mid-Pliocene to the upper Pleistocene. Given those informations, possible correlations between the estimated divergences and paleogeographic events are discussed.

Resumo

O gênero *Ramphastos* (Piciformes: Ramphastidae) é composto pelos tucanos. Estas aves se apresentam distribuídas ao longo de toda a região neotropical, sendo caracterizadas pela presença de um bico grande, robusto e colorido. Historicamente, a sistemática do grupo já foi tratada com abordagens de natureza morfológica e molecular, que apresentaram algumas discordâncias de relevância taxonômica e evolutiva. A morfologia apontou para um total de sete espécies para o gênero, a maioria delas sendo subdivididas em subespécies. Um estudo molecular subsequente, baseado na análise de seqüências de DNA, corroborou seis destas sete espécies, mas sugeriram que a espécie monotípica *R. brevis* estaria incluída dentro do complexo *R. vitellinus*, que seria então parafilético. Além disso, tal estudo não indicou o monofiletismo das populações de *R. vitellinus ariel* das regiões sudeste e norte do Brasil, embora elas sejam idênticas quanto à coloração geral. No intuito de contribuir com a discussão em torno dessas discordâncias, o presente estudo reúne dados relativos a oito genes mitocondriais e um íntron nuclear, tratados sob os métodos de máxima parcimônia, máxima verossimilhança e análise bayesiana. Adicionalmente, incluiu-se nas análises amostras da subespécie *R. vitellinus citreolaemus*, não contemplada em estudos anteriores. O estudo do alinhamento total mostrou grande resolução quanto ao posicionamento topológico de todas as espécies (assim como para muitas das subespécies), e uma grande congruência entre as árvores dos três métodos de reconstrução aqui empregados, com praticamente todos os ramos apresentando *bootstrap* $\geq 70\%$ e estimativas de probabilidade posterior $\geq 95\%$. O monofiletismo do complexo *R. vitellinus* foi corroborado, aparecendo como grupo irmão de *R. brevis*. Os resultados também confirmam que as populações de *R. vitellinus ariel* do sudeste e do norte são geneticamente distintos. *R. vitellinus citreolaemus* é apontado como grupo irmão das subespécies do complexo *vitellinus* de distribuição amazônica. Datações das divergências foram estimadas através de quatro métodos distintos, todos eles convergindo para períodos semelhantes: separações entre gêneros teriam ocorrido na faixa Oligoceno Superior/Mioceno Superior; entre espécies, do Mioceno Superior ao Pleistoceno Inferior; e entre subespécies, do Plioceno Médio ao Pleistoceno Superior. A partir de tais informações, possíveis correlações entre as divergências observadas e eventos paleogeográficos são discutidas.

3.1 Introdução

Tucanos e araçaris (que juntos compõem a família Ramphastidae) têm em comum um bico relativamente longo e recurvado, presença de tómos nas laterais da parte superior do bico, arranjo único das vértebras caudais, postura de descanso típica, além de outras características relacionadas à morfologia (Haffer, 1974; Swierczewski & Raikow, 1981; Höfling, 1991, 1998; Prum, 1988; Short & Horne, 2001). Essa família está distribuída do México à Argentina, ocorrendo em toda a região neotropical. Quanto à monofilia do grupo, diferentes estudos evidenciaram a proximidade filogenética entre tucanos e araçaris (Lanyon & Zink, 1987; Prum, 1988; Höfling, 1991, 1998; Barker & Lanyon, 2000; Nahum *et al.*, 2003; Moyle, 2004).

Os tucanos do gênero *Ramphastos* são conspícuos, dentre outras características, pelo fato de possuírem um bico ainda maior que os araçaris, tamanho corporal mais avantajado, cobertura dorsal preta ou marrom-escuro, penas infracaudais vermelhas, peito com coloração branca ou em diferentes tons de amarelo, e tarso-metatarso azul (Short & Horne, 2001). Os calaus, da África (família Bucerotidae), ocupam nicho semelhante aos tucanos e possuem tamanho corporal e do bico similares, mas não são filogeneticamente relacionados (Cracraft, 1981; Feduccia, 1999; Braun & Kimball, 2002; García-Moreno *et al.*, 2003; Dyke & Van Tuinen, 2004; Fain & Houde, 2004).

Haffer (1974) foi o primeiro a postular relações filogenéticas entre as espécies de *Ramphastos*, sugerindo a existência de sete espécies. Tais táxons estariam divididos em dois grupos naturais segundo o autor, dados o padrão de vocalização e a morfologia apical do bico: os *croakers*, englobando as espécies *R. brevis*, *R. dicolorus*, *R. sulfuratus*, *R. toco* e *R. vitellinus*; e os *yelpers*, com *R. ambiguus* e *R. tucanus*. Dentre essas espécies, duas são monotípicas (*R. brevis* e *R. dicolorus*), sendo as restantes categorizadas em subespécies (Short & Horne, 2001).

Weckstein (2005) obteve uma filogenia molecular para as espécies do gênero a partir de três genes mitocondriais codificantes (parte do COI, Cyt-b quase completo, e ND2 completo), num total de 2468 sítios no alinhamento concatenado. Suas análises se basearam em máxima parcimônia (MP), máxima verossimilhança (MV) e análise bayesiana (AB). Todas as sete espécies acima foram analisadas pelo autor, e muitas das subespécies também

(*R. a. ambiguus*, *R. a. swainsonii*, *R. s. sulfuratus*, *R. s. brevicarinatus*, *R. t. tucanus*, *R. t. cuvieri*, *R. v. ariel*, *R. v. culminatus*, *R. v. vitellinus*). Os três métodos de inferência filogenética resgataram a mesma topologia. Dentre os ramos obtidos, apenas três não apresentaram índice de *bootstrap* $\geq 70\%$ concomitantemente para para MP e MV, embora todos tenham apresentado estimativa de probabilidade posterior (PP) $\geq 0,95$ (figura 3.1).

Figura 3.1 - Cladograma obtido por Weckstein (2005) para *Ramphastos*, a partir de 3 genes codificantes mitocondriais (partes do COI e Cyt-b, e ND2 integral - total: 2468 bases). Acima dos ramos, *bootstraps*-MP; abaixo, *bootstraps*-MV. Setas vermelhas apontam os nós com baixo valor de *bootstrap* de MP e/ou de MV.

A filogenia acima é congruente com a hipótese filogenética proposta por Haffer (1974) em muitos aspectos (ver também figura 1.2). Os *yelpers* são monofiléticos; com exceção de *R. toco* (sendo esta espécie basal dentro de *Ramphastos*), todos os outros *croakers* formam um componente monofilético; espécies simpátricas, de coloração geral idêntica, não são grupos irmãos; e seis dentre sete espécies propostas por Haffer são monofiléticas. A espécie *R. vitellinus*, no entanto, se mostrou parafilética, dada a inclusão do táxon *R. brevis* entre a subespécie *R. v. ariel* do sudeste do Brasil (*R. v. ariel* SE) e as subespécies de *vitellinus* restantes (figura 3.1). A subespécie *R. v. ariel* também se mostrou parafilética, surgindo como dois clados disjuntos dentro da topologia, com *R. v. ariel* SE numa posição basal dentro de *R. vitellinus* ssp., e *R. v. ariel* do norte (*R. v. ariel* N) próximo de *R. v. culminatus* (embora estas duas subespécies não apresentem monofiletismo recíproco - figura 3.1). Tal separação não era esperada *a priori*, já que a coloração geral de *R. v. ariel* SE e *R. v. ariel* N é a mesma. No entanto, esta não foi a primeira evidência de que há diferenças entre os membros desse táxon: dentre três tipos de medições morfológicas feitas por Haffer (tamanho de cauda, do bico, e da asa), o tamanho da asa mostrou uma grande diferença entre os dois morfótipos (Haffer, 1974), em acordo com o resultado obtido por Weckstein (2005). Um outro ponto interessante na figura 3.1 é que as subespécies estudadas de *R. ambiguus* (*R. a. ambiguus* e *R. a. swainsonii*) parecem ser linhagens evolutivas independentes; o mesmo também pode ser dito quanto a *R. v. vitellinus* em relação ao grupo [*R. v. culminatus* + *R. v. ariel* N].

As análises aqui realizadas visaram aumentar o número de regiões estudadas e a amostragem taxonômica, e comparar o resultado com aquele obtido por Weckstein (2005). Foram seqüenciados nove genes para o gênero *Ramphastos* e para um membro de cada gênero de araçari (*sensu* Haffer, 1974 - tais amostras foram utilizadas como grupo externo), num total de oito regiões mitocondriais (incluindo as três regiões estudadas por Weckstein, 2005) e um íntron de um gene nuclear. Indivíduos da subespécie *R. v. citreolaemus* foram acrescentados, assim como um novo exemplar de *R. v. ariel* SE (o estudo de Weckstein, de 2005, só havia incluído um espécime).

As análises filogenéticas empregadas foram de máxima parcimônia (MP), máxima verossimilhança (MV), e análise bayesiana (AB). Utilizou-se quatro métodos de datação de

divergências (relógio molecular, *penalized likelihood*-PL, estimativa bayesiana, e PATHd8), em conjunto com a informação de fósseis (um piciforme e uma coruja), e de um trabalho recente de datação em aves utilizando uma ampla gama de táxons, cinco regiões gênicas nucleares e uma grande quantidade de fósseis (Ericson *et al.*, 2006), para se estabelecer limites *a priori* para as divergências observadas. Em face dos resultados obtidos, procedeu-se a uma discussão biogeográfica, associando os períodos de divergência encontrados e a distribuição espacial dos táxons estudados a eventos paleogeográficos supostamente importantes para a diversificação do grupo.

3.2 Objetivos

- Reavaliar a filogenia das espécies do gênero *Ramphastos*;
- Estimar períodos de divergência entre as espécies;
- Associar datações com processos paleogeográficos;

3.3 Material e Métodos

Os indivíduos analisados das subespécies de *R. vitellinus* ssp., *R. tucanus* ssp. e *R. ambiguus* ssp. estão mostrados nas figuras 3.2-3.4 (ver também tabela 1.1).

Figura 3.2- Espécimes coletados de *R. vitellinus* ssp. Os números estão em acordo com a tabela 1.1. Amostras sem número correspondem a uma mesma localidade. A amostra 3 de *R. v. citreolaemus* é de cativeiro. Para detalhes das amostras, consultar a referida tabela.

Figura 3.3 - Espécimes coletados de *R. tucanus* ssp. Os números estão em acordo com a tabela 1.1. Amostras sem número correspondem a uma mesma localidade. Para detalhes das amostras, consultar a referida tabela.

Figura 3.4 - Espécimes coletados de *R. ambiguus* ssp. Os números estão em acordo com a tabela 1.1. Para detalhes das amostras, consultar a referida tabela. Em amarelo, *R. a. swainsonii* 1 (mais ao norte) e *R. a. swainsonii* 2 (ao sul); em vermelho, *R. a. ambiguus*.

A parte metodológica relativa à amplificação e seqüenciamento já foi descrita na seção 1.4, não sendo aqui repetida. As nove regiões gênicas seqüenciadas (e os respectivos tamanhos de alinhamento) para o presente estudo foram:

- *regiões mitocondriais*: parte dos genes ribossomais 12S (628 sítios) e 16S (546 sítios), parte da região controladora (699 sítios), ATPase 6 (674 sítios) e ATPase 8 (168 sítios) inteiros, parte da citocromo-oxidase I (COI - 379 sítios), citocromo-b (Cyt-b - 1048 sítios) quase completo, e a NADH-desidrogenase II (ND2 - 1041 sítios) completa;
- *região nuclear*: íntron 7 do β -fibrinogênio (Fib-7 - 380 sítios);

O alinhamento total, portanto, apresentou 5563 sítios. Para o 12S, incluiu-se todas as regiões de alças (exceto pela remoção de 33 sítios ambíguos, mesmo após a inferência da estrutura secundária - capítulo 2), mas apenas as regiões 5' (proximais) das hastes foram incluídas, para evitar o problema da correlação entre taxas nestas últimas. A opção pela região 5' foi arbitrária, dentre as duas possíveis (no capítulo 2, observou-se que qualquer umas das duas escolhas não altera a topologia obtida). Quanto ao 16S, devido à região

seqüenciada ser pequena em relação ao comprimento total do gene, não foi possível estabelecer as regiões de pareamento, sendo os sítios para este gene considerados independentes nas análises adiante. A parte da região controladora utilizada (de agora em diante referida como DLoop, por simplicidade) faz parte da terceira região, *sensu* Pereira e col., 2004 (domínio CSB). Esse gene apresentou problemas de amplificação para o grupo externo, tendo sido seqüenciado apenas para o táxon *Baillonius bailloni*. Para uma maior segurança de que tal produto não seria um pseudogene, comparou-se essa seqüência com uma amostra da mesma região de DNA para o espécime *Pteroglossus azara flavirostris*, baixada do GenBank (NC008549), espécime este que possui todo o seu genoma mitocondrial seqüenciado. O alinhamento dessa seqüência com *B. bailloni* foi relativamente simples, indicando que a porção seqüenciada deste táxon parece ser relativa à região ortóloga correta; assim, utilizou-se ambas as seqüências como grupo externo para as análises do DLoop. Para a análise dos alinhamentos concatenados, inseriu-se sucessivos “N” nos outros membros do grupo externo (*P. inscriptus*, *A. cucullata*, *S. reinwardtii* e *A. prasinus*) para o DLoop, e *P. azara flavirostris* foi removido.

O teste para homogeneidade de composição de bases foi realizado através do χ^2 do PAUP* para o alinhamento total, para cada um dos genes, e também para as três posições de codons no caso de genes que codificam proteínas, usando-se apenas um indivíduo por espécie (ou subespécie no caso dos táxons politípicos), e descartando-se os sítios invariáveis (Cunningham, 1997). Foram obtidos também os histogramas de freqüências nucleotídicas para cada gene. A relação de sítios invariáveis, não-informativos e informativos para parcimônia, em cada gene, foi apresentada. Curvas de saturação foram estimadas, comparando-se as distâncias patrísticas de MV com as distâncias-p para cada gene individualmente. Os genes foram analisados separadamente através do índice de *bootstrap* por MV para um total de 500 pseudo-réplicas, sendo o melhor modelo para cada gene estimado através do AICc no Modeltest.

O teste de homogeneidade entre partições implementado no PAUP* foi utilizado para testar se os genes possuem sinais filogenéticos congruentes ou não. Foram realizadas 200 permutações entre os conjuntos de dados individuais, sendo a cada rodada estimada a

soma do comprimento das árvores de MP para cada partição. Valores significativos sugerem a presença de genes com sinais incongruentes.

Para a sequência total concatenada, calculou-se o modelo de melhor ajuste segundo o AICc através do Modeltest (5563 sítios). A análise de MV e de *bootstrap* seguiu os mesmos passos apresentados na seção 1.4. Um total de 500 pseudo-réplicas de *bootstrap* foram utilizadas para MV. Pseudo-réplicas de *bootstrap* também foram calculadas para MP (num total de 2.000).

A análise bayesiana foi realizada implementando-se o modelo sugerido pelo AICc para MV. Um total de 5.000.000 de gerações foi empregado, com a execução em conjunto de uma cadeia “fria” e três “quentes”. O período de *burn-in* foi determinado através do programa Tracer v1.3 (Rambaut & Drummond, 2006). Três diferentes partições dos dados foram analisadas: 5 partições (DLoop + 12S + 16S + genes codificantes + Fib-7), 7 partições (as mesmas anteriores, com os genes codificantes sendo desmebrados nas três posições do codon), e 6 partições (as mesmas do caso anterior, mas sem o Fib-7).

Seja T_1 a melhor topologia possível dado que os *R. vitellinus* ssp. formem um grupo monofilético, e T_2 a melhor topologia possível onde os *R. vitellinus* ssp. formem um componente parafilético, um teste de *bootstrap* paramétrico foi empregado, a fim de se testar se existe uma diferença significativa quanto às hipóteses *R. vitellinus* ssp. monofilético x *R. vitellinus* ssp. parafilético. Um total de 200 alinhamentos, com o mesmo número de sítios que o original, foram simulados na árvore T_1 , com a topologia e valores dos parâmetros para a geração de cada um obtidos da árvore de MV para o conjunto total dos dados. A estatística de interesse, calculada para os dados originais e para cada um dos alinhamentos simulados, foi a diferença entre os logaritmos das verossimilhanças (δ) de T_1 (com valor L_1) e T_2 (com valor L_2). Seja k o valor de δ para os dados originais, tem-se abaixo as hipóteses nula e alternativa:

H₀: Não há diferença significativa entre T_1 e T_2 (ou seja, $L_1 - L_2 \geq k$);

H_A: T_2 é significativamente superior a T_1 (ou seja, $L_1 - L_2 < k$).

Caso o valor do k original se apresentasse dentro dos 5% inferiores da distribuição, o resultado seria considerado significativo, ou seja, a hipótese de parafiletismo ganharia mais força que a de monofiletismo para as formas de *R. vitellinus* ssp. Em caso negativo, não se poderia rejeitar que *R. vitellinus* ssp. é monofilético. A figura 3.5 mostra as hipóteses filogenéticas T_1 e T_2 ; em cada alinhamento simulado (e também para o original) foram permitidos rearranjos entre ramos tanto em T_1 quanto em T_2 , para que cada filogenia fosse otimizada seguindo apenas as restrições topológicas básicas acima estipuladas (figura 3.5). Ao todo, 200 alinhamentos foram simulados no programa Mesquite 1.12 (Maddison & Maddison, 2006), usando-se a topologia, modelo (GTR+I+G), e valores dos parâmetros obtidos da árvore de MV para os dados originais.

Figura 3.5 - Hipótese topológicas alternativas: **a)** *R. vitellinus* ssp. monofilético (T_1); **b)** *R. vitellinus* ssp. parafiletico (T_2).

As topologias obtidas por um conjunto de dados envolvendo apenas genes não codificantes de proteínas (Dloop + 12S + 16S), e por um conjunto de dados contendo apenas seqüências codificadoras de proteínas (ATPase 6 + ATPase 8 + COI + Cyt-b + ND2) foram comparadas entre si, a fim de se detectar eventuais diferenças de resolução entre os dois tipos de marcadores.

Um alinhamento alternativo foi obtido para o Cyt-b, incluindo-se três espécimes de *R. v. citreolaemus*, sendo esta a única região para a qual os referidos indivíduos foram sequenciados. Análises de MP e MV foram realizadas a partir desse conjunto de dados.

A datação das divergências foi estimada a partir de quatro métodos distintos: relógio molecular, PATHd8, PL e análise bayesiana de taxas. Para essas análises, o conjunto de dados original foi modificado, sendo mantidos apenas um espécime por táxon, e sendo descartadas das análises o Fib-7 (por ser nuclear e portanto apresentar características diferentes de evolução molecular) e o DLoop (por não ter sido possível o alinhamento desta região com os táxons escolhidos para calibrar alguns dos nós - ver abaixo). No caso do relógio molecular, o hLRT (*hierarchical Likelihood Ratio Test*) foi aplicado, comparando-se os valores de verossimilhança da árvore de MV com e sem a restrição de tamanhos de ramos iguais entre linhagens descendentes de um mesmo nó; caso o teste não apresentasse valor significativo, não haveria como rejeitar a hipótese de constância de taxas ao longo da árvore. O táxon *Dryocopus pileatus* (um piciforme) foi acrescentado ao alinhamento para servir de ponto de calibração para os quatro primeiros métodos, sendo sua idade mínima baseada na datação fóssil obtida para um espécime de piciforme do Oligoceno Inferior (Mayr, 2005). Duas estimativas de tempos de divergência foram realizadas para os três primeiros métodos, calibrando-se alternadamente a raiz a 30 m.a. (idade mínima de separação entre Picidae/Ramphastidae - Mayr, 2005) e 50 m.a. (limite superior e conservativo de datação, baseado no estudo de Ericson *et al.*, 2006).

A análise bayesiana de taxas foi feita pelo programa BEASTv1.4 (Drummond & Rambaut, 2003). O programa gera uma distribuição posterior de tempos de divergência e taxas evolutivas para os diferentes nós de uma dada filogenia, a partir da região estacionária da cadeia de MCMC. O programa não assume uma topologia *a priori* como é o caso dos métodos de datação anteriores; a busca pela melhor topologia é realizada simultaneamente com a estimação das datações - o próprio enraizamento das topologias é estimado ao longo da execução das cadeias de Markov. Ao final da análise, intervalos de confiança relacionados a tempos de divergência e taxas de evolução puderam ser obtidos. Para esta abordagem, um segundo táxon foi acrescentado ao alinhamento, *Ninox novaeseelandiae* (uma coruja, ordem Strigiformes) como ponto de calibração, para que os intervalos de

confiança obtidos ao final da execução da cadeia de MCMC fossem menores. Utilizou-se este táxon, ao invés de outro qualquer, por ele apresentar o melhor equilíbrio entre dois fatores: proximidade evolutiva relativa (Espinosa de los Monteros, 2000, 2003; Van Tuinen *et al.*, 2000; Ericson *et al.*, 2006), e existência de seqüências de DNA para todas as sete regiões mitocondriais analisadas (o espécime de *Ninox novaeseelandiae* utilizado, baixado do GenBank, teve todo seu genoma mitocondrial seqüenciado; o mesmo também é verdade para o táxon *Dryocopus pileatus*). Desta forma, dois pontos de calibração por fósseis foram utilizados, considerando-se um limite inferior de separação entre Ramphastidae/Picidae de 30 m.a. (Mayr, 2005), e um limite inferior para Strigiformes/Piciformes de 55 m.a. (Rich & Bohaska, 1976). Os limites superiores utilizados foram 50 m.a. (Ramphastidae/Picidae) e 80 m.a. (Strigiformes/Piciformes), baseados no estudo de Ericson e col. (2006), que analisaram uma ampla filogenia de aves empregando vários fósseis como limites inferiores de separações (os autores usaram PL e PATHd8 para o cálculo dos tempos de divergência). Os limites superiores aqui usados são mais remotos que as datações obtidas pelos autores acima, a fim de se minimizar a chance de enfiamentos no resultado final gerados pelo uso de intervalos muito restritivos. Uma distribuição LogNormal de taxas foi assumida *a priori*, sendo permitida a variação de taxas ao longo dos ramos para uma mesma topologia (média das taxas nos diferentes ramos entre 0 e 0,2 s/s/l/m.a.* por topologia proposta) e com um desvio padrão máximo por topologia (entre 0 e 0,5 s/s/l/m.a., um intervalo que abrange desde um relógio molecular, até variações bastante significativas de taxas ao longo da topologia). O modelo LogNormal foi escolhido por apresentar maior grau de acurácia em relação a dados simulados, e maior precisão em relação a topologias obtidas a partir de alguns conjuntos de dados reais e simulados (Drummond *et al.*, 2006). O intervalo para a média (0,0-0,2), relativamente liberal, visou englobar os valores máximos e mínimos para as taxas encontradas em estudos onde não houve extrapolações a partir de análises externas, e com pontos de calibração baseados em fósseis ou eventos paleogeográficos reconhecidos (as referências utilizadas para esses valores de taxas podem ser consultadas em García-

* s/s/l/m.a. = substituições por sítio por linhagem por milhão de anos. Em geral, os valores na literatura são reportados numa escala de s/s/m.a. (substituições por sítio por milhão de anos); para fazer a conversão, é só dividir esse último valor por 2.

Moreno, 2004; e Ho *et al.*, 2005). Para a análise, duas corridas independentes com 5.000.000 de gerações cada foram efetuadas, sendo descartadas as primeiras 200.000 gerações (*burn-in* determinado pelo programa Tracer 1.3).

3.4 Resultados

O teste do χ^2 para homogeneidade de frequências para cada um dos genes está mostrado na tabela 3.1, que mostra os valores obtidos para o conjunto total de genes, assim como para cada gene isoladamente, e também para as diferentes posições nos códons.

Tabela 3.1: *p*-values para o teste de homogeneidade de frequências implementado no PAUP*.

Gene	χ^2			
<i>Todos</i>	0.1470			
12S (loops + stems 5')	1.0000			
16S	0.9974			
Dloop	0.7932			
ATPase 6	0.9938			
ATPase 8	0.9886			
COI	1.0000			
Cyt-b	0.9618			
ND2	0.9993			
Fib-7	1.0000			
por posição no codon	pos. 1	pos. 2	pos. 3	
ATPase 6	0.9871	1.0000	0.7200	
ATPase 8	0.8799	0.9967	0.7306	
COI	1.0000	1.0000	1.0000	
Cyt-b	0.9999	1.0000	0.6730	
ND2	1.0000	1.0000	0.8417	

As frequências nucleotídicas para cada região (oito mitocondriais, e uma nuclear) estão sumarizadas na figura 3.6.

Figura 3.6 - Histogramas de frequências de bases, para cada gene analisado.

A tabela 3.2 indica a proporção de sítios invariáveis, não-informativos, e informativos para parcimônia, para cada um dos nove genes analisados no presente estudo.

Tabela 3.2: Valores de probabilidade para o teste de homogeneidade de frequências implementado no PAUP.

	Constantes	Não-informativos	Informativos	TOTAL
12S	473 (75,3%)	41 (6,5%)	114 (18,2%)	628
16S	374 (68,5%)	57 (10,4%)	115 (21,1%)	546
Dloop	498 (71,3%)	38 (5,4%)	163 (23,3%)	699
ATPase 6	456 (67,7%)	42 (6,2%)	176 (26,1%)	674
ATPase 8	98 (58,3%)	24 (14,3%)	46 (27,4%)	168
COI	280 (73,9%)	21 (5,5%)	78 (20,6%)	379
Cyt-b	697 (66,5%)	74 (7,1%)	277 (26,4%)	1048
ND2	640 (61,5%)	99 (9,5%)	302 (29%)	1041
Fib-7	350 (92,1%)	12 (3,2%)	18 (4,7%)	380

Os gráficos de saturação para cada gene (exceto para o Fib-7, que apresentou um gráfico com pontos muito dispersos) estão mostrados na figura 3.7. As distâncias patrísticas de MV foram calculadas a partir do modelo escolhido pelo AICc, no Modeltest, para cada gene (tabela 3.3).

Figura 3.7 - Distâncias patrísticas de MV (abcissa) x distâncias-p (ordenada).

Figura 3.7 - ... continuação.

Tabela 3.3: Modelo escolhido pelo AICc para cada região estudada.

Região	Modelo
TOTAL	GTR + I + G
12S	GTR + I + G
16S	GTR + I + G
Dloop	HKY + I
ATPase 6	TrN + I + G
ATPase 8	TrN + G
COI	HKY + I + G
Cyt-b	TrN + I + G
ND2	TrN + G

As topologias obtidas por MV, a partir do uso dos modelos acima para cada gene, estão mostradas nas figuras 3.8 a-i.

Figura 3.8 - Árvores de MV para cada gene individualmente. Acima de cada ramo, valores de *bootstrap* de MV. Legenda: R. - *Ramphastos*; a. - *ambiguus*; s. - *sulfuratus*; t. - *tucanus*; v. - *vitellinus*. O grupo externo é formado pelos espécimes *Andigena cucullata*, *Aulacorhynchus prasinus*, *Baillonius bailloni*, *Pteroglossus inscriptus* e *Selenidera reinwardtii* (ver tabela 1.1 para detalhes das amostras).

Figura 3.8 - ... continuação.

Figura 3.8 - ... continuação.

Figura 3.8 - ... continuação.

Figura 3.8 - ... continuação.

O resultado do teste de homogeneidade de partições implementado no PAUP apresentou um p-value não significativo ($p = 0.705$), não sendo possível rejeitar a hipótese de que o sinal filogenético das partições individuais seja semelhante.

A árvore de MV baseada no conjunto total de seqüências concatenadas, e com valores de *bootstrap* (tanto para MP quanto para MV) associados aos ramos, está apresentada na figura 3.9.

Figura 3.9 - Topologia de MV obtida para a seqüência total concatenada. Valores acima dos ramos: *bootstrap*_MP; abaixo dos ramos: *bootstrap*_MV.

A figura 3.10, na página seguinte, mostra os resultados das análises bayesianas utilizando o alinhamento total concatenado, com e sem o Fib-7.

■ **5 partições:** DLoop + 12S + 16S + genes de proteínas + Fib-7
 ■ **7 partições:** DLoop + 12S + 16S + 1ª pos. codon + 2ª pos. codon + 3ª pos. codon + Fib-7

■ **6 partições:** DLoop + 12S + 16S + 1ª pos. codon + 2ª pos. codon + 3ª pos. codon

Figura 3.10 - Análise bayesiana a) com o Fib-7; b) sem o Fib-7. As legendas respectivas indicam o valor de PP em cada ramo, para cada conjunto de partições.

O resultado do *bootstrap* paramétrico está mostrado na figura 3.11.

Figura 3.11 - Histograma com os valores de δ , obtidos por simulação de alinhamentos (*bootstrap* paramétrico), para um total de 200 réplicas. A seta amarela indica o valor de δ para os dados originais. Em rosa, região de rejeição para H_0 .

As topologias baseadas no DLoop + 12S + 16S, e no alinhamento conjunto dos genes codificantes de proteínas, estão mostradas na figura 3.12.

Figura 3.12 - Topologias de MV (e valores de *bootstrap* associados): a) DLoop+12S+16S; b) ATPase6+ATPase8+COI+Cyt-b+ND2.

A árvore de MV obtida da análise do Cyt-b com a adição de 3 espécimes do táxon *R. v. citreolaemus* está na figura 3.13 (apenas a porção relativa a *R. brevis* e *R. vitellinus* ssp. está mostrada).

Figura 3.13 - Porção da árvore de MV do Cyt-b, com o adicionamento de 3 espécimes de *R. v. citreolaemus*, e valores de *bootstrap* associados.

A topologia da figura 3.14a apresenta um enraizamento artificial do táxon *R. v. citreolaemus* na árvore obtida pela análise do conjunto total de dados (um espécime por táxon é utilizado nos casos de clados monofiléticos, para facilitar a interpretação), com as distribuições geográficas associadas a cada clado. A figura 3.14b mostra a seção da árvore 3.14a com alternância de distribuições entre a região andina e o sudeste da América do Sul.

Figura 3.14 - Árvore geral obtida no presente estudo: a) topologia com apenas um espécime por táxon (nos casos de monofiletismo recíproco entre duas espécies ou subespécies); b) relações filogenéticas dos táxons que apresentam alternância de distribuição entre os Andes e o sudeste da América do Sul.

A tabela 3.4 mostra as distâncias intra- e inter-taxonômicas entre as subespécies de *R. vitellinus* ssp., e a tabela 3.5 mostra as mesmas informações quanto às subespécies de *R. tucanus* ssp.

Tabela 3.4: Distâncias intra- e inter-taxonômicas em *R. vitellinus* ssp. (o asterisco em *R. v. citreolaemus* é para indicar que as distâncias intra-taxonômicas envolvendo esse táxon foram baseadas apenas no Cyt-b).

Distâncias intra-taxonômicas

<i>R. v. ariel</i> SE	0,263
<i>R. v. ariel</i> N	0,193
<i>R. v. vitellinus</i>	0,106
<i>R. v. culminatus</i>	0,143
<i>R. v. citreolaemus</i>	0,183 *

→ **Variação:**
0,106 – 0,263

Distâncias inter-taxonômicas

<i>R. v. ariel</i> SE x "outros <i>vitellinus</i> ssp."	2,203 - 2,454
<i>R. v. vitellinus</i> x <i>R. v. culminatus</i>	0,500 - 0,605
<i>R. v. vitellinus</i> x <i>R. v. ariel</i> N	0,596 - 0,642
<i>R. v. ariel</i> N x <i>R. v. culminatus</i>	0,123 - 0,215

Tabela 3.5: Distâncias intra- e inter-taxonômicas em *R. tucanus* ssp.

Distâncias intra-taxonômicas

<i>R. t. tucanus</i> 1 x <i>R. t. tucanus</i> 2	0,470
<i>R. t. cuvieri</i> 1 x <i>R. t. cuvieri</i> 2	0,145

Distâncias inter-taxonômicas

<i>R. t. cuvieri</i> 1 x <i>R. t. tucanus</i> 1	0,419
<i>R. t. cuvieri</i> 1 x <i>R. t. tucanus</i> 2	0,088
<i>R. t. cuvieri</i> 2 x <i>R. t. tucanus</i> 1	0,503
<i>R. t. cuvieri</i> 2 x <i>R. t. tucanus</i> 2	0,217

→ **Variação:**
0,088 – 0,503

Os tempos de divergência estimados a partir de quatro métodos diferentes estão mostrados na tabela 3.6.

Tabela 3.6: Datações obtidas por quatro abordagens diferentes. A faixa de valores para a análise bayesiana é referente a um intervalo de 95% dos valores da distribuição posterior. A amplitude total (penúltima coluna) e os intervalos em escala geológica (última coluna) correspondem à faixa entre os tempos de divergência mínimos e máximos estimados, levando-se em consideração os quatro métodos de datação. Épocas geológicas e tempos associados de acordo com a tabela estratigráfica internacional da *International Commission on Stratigraphy* (2006).

Divergência	rel. mol.	PATHd8	PL	Análise bayesiana	Amplitude total	Época geológica
Picidae / Ramphastidae	30 - 50*	30 - 50*	30 - 50*	32.01 - 49.79	30 - 50	Oligoceno Inferior - Eoceno Inferior
<i>Ramphastos</i> / <i>Araçaris</i>	13.56 - 22.60	12.20 - 20.33	12.91 - 21.51	14.61 - 25.00	12.20 - 25	Mioceno Médio - Oligoceno Superior
<i>R. toco</i> / outros <i>Ramphastos</i>	5.03 - 8.38	4.94 - 8.23	5.81 - 9.56	6.01 - 10.97	4.94 - 10.97	Plioceno Inferior - Mioceno Superior
Croakers / Yelpers	3.88 - 6.47	4.00 - 6.67	4.62 - 7.59	4.91 - 8.88	3.88 - 8.88	Plioceno Inferior - Mioceno Superior
<i>R. sulfuratus</i> / (dic. + brevis + vit. ssp.)	3.46 - 5.77	3.76 - 6.27	3.98 - 6.54	4.17 - 7.77	3.46 - 7.77	Plioceno Médio - Mioceno Superior
<i>R. dicolorus</i> / (brevis + vit. ssp.)	3.04 - 5.07	3.33 - 5.55	3.39 - 5.58	3.37 - 6.48	3.04 - 6.48	Plioceno Médio - Mioceno Superior
<i>R. brevis</i> / <i>R. vitellinus</i> ssp.	1.27 - 2.11	1.38 - 2.30	1.35 - 2.24	1.51 - 3.02	1.27 - 3.02	Pleistoceno Inferior - Plioceno Médio
<i>R. v. ariel</i> SE / vit. da Amazônia	1.03 - 1.71	1.14 - 1.90	1.07 - 1.77	1.28 - 2.93	1.03 - 2.93	Pleistoceno Inferior - Plioceno Médio
<i>R. v. vitellinus</i> / <i>culminatus</i> + <i>ariel</i> N)	0.21 - 0.35	0.23 - 0.39	0.21 - 0.35	0.21 - 0.60	0.21 - 0.60	Pleistoceno Médio
<i>R. v. culminatus</i> + <i>R. v. ariel</i> N [†]	0.04 - 0.07	0.05 - 0.08	0.04 - 0.07	0.02 - 0.164	0.02 - 0.164	Pleistoceno Superior - Pleistoceno Médio
<i>R. tucanus</i> ssp. / <i>R. ambiguus</i> ssp.	2.17 - 3.61	2.06 - 3.44	2.73 - 4.46	2.48 - 5.13	2.06 - 5.13	Plioceno Superior - Plioceno Inferior
<i>R. t. tucanus</i> + <i>R. t. cuvieri</i> [†]	0.17 - 0.28	0.16 - 0.26	0.21 - 0.34	0.13 - 0.49	0.13 - 0.49	Pleistoceno Médio
<i>R. a. ambiguus</i> / <i>R. a. swainsonii</i>	0.66 - 1.10	0.63 - 1.05	0.84 - 1.37	0.68 - 1.68	0.63 - 1.68	Pleistoceno Médio - Pleistoceno Inferior
(Ptero. + Bail.) / (Sel. + And. + Aul.)	12.08 - 20.13	12.20 - 20.33	10.52 - 17.57	12.07 - 21.30	10.52 - 21.30	Mioceno Superior - Mioceno Inferior
<i>Aulacorhynchus</i> / (Sel. + And.)	10.19 - 16.98	10.24 - 17.07	8.74 - 14.59	9.67 - 17.81	8.74 - 17.81	Mioceno Superior - Mioceno Inferior
<i>Andigena</i> / <i>Selenidera</i>	6.47 - 10.79	6.26 - 10.43	5.87 - 9.78	6.36 - 12.73	5.87 - 12.73	Mioceno Superior - Mioceno Médio
<i>Pteroglossus</i> + <i>Bailloniuss</i> [‡]	3.30 - 5.50	3.31 - 5.51	2.81 - 4.70	3.42 - 7.30	2.81 - 7.30	Plioceno Médio - Mioceno Superior

Legenda:

* - não corresponde a um intervalo de valores - nestes casos, apenas fixou-se a idade da raiz para os tempos mínimo e máximo de separação entre Picidae e Ramphastidae (mínimo: Mayr, 2005; máximo: Ericson *et al.*, 2006);

[†] - monofiletismo recíproco entre as duas linhagens não detectado;

[‡] - um estudo recente indica que *Bailloniuss* está dentro do gênero *Pteroglossus* (Kimura *et al.*, 2004).

Separações entre gêneros
Separações entre espécies
Separações entre subespécies

3.5 Conclusões

3.5.1 - Aspectos moleculares

A composição nucleotídica não se mostrou alterada de forma significativa nas diferentes partições analisadas (tabela 3.1), sejam elas os genes propriamente ditos, ou as posições do codon para os genes codificantes. O p-value para o conjunto total das seqüências foi baixo (0,1470), mas ainda acima dos 5% limítrofes. Os resultados como um todo mostraram que o pressuposto de freqüências estacionárias (assumido de forma geral nos diferentes modelos reversíveis, e para todos os modelos analisados pelo AICc no Modeltest) é bastante razoável para os presentes dados.

O nucleotídeo mais comum nos genes ribossomais (12S e 16S) foi a adenina (A) - figura 3.6. Sugeriu-se que isso é seletivamente mantido devido a interações hidrofóbicas necessárias para o correto funcionamento dos ribossomos, já que A é o nucleotídeo menos polar de todos (Gutell *et al.*, 1985). Depois de A, as freqüências em ordem decrescente foram: C, T e G. A relativa pobreza de G é condizente com o padrão encontrado em outras aves, répteis e roedores de uma forma geral (Kocher *et al.*, 1989). Já no caso do DLoop não há diferenciação perceptível entre A e C, podendo esta menor freqüência de A estar ligada à ausência de seleção para interações hidrofóbicas segundo o pressuposto acima. Nos genes que codificam proteínas, o nucleotídeo mais comum foi C, seguido de A, T e G (o Cyt-b mostrou-se uma exceção, com T sendo mais freqüente que A; e o COI não apresentou uma diferença clara). Não obstante, os diferentes tipos de genes mitocondriais apresentam um histograma de freqüências mais parecidos entre si, do que quando estes são comparados com o Fib-7; este gene apresentou, em ordem decrescente de freqüência, os nucleotídeos T, A, C e G. Isso pode estar associado às diferentes propriedades nas taxas de evolução molecular entre o núcleo e a mitocôndria (Graur & Li, 2000).

A porcentagem de sítios informativos para parcimônia foi maior no ND2, seguidos por ATPase 8, Cyt-b, ATPase 6, DLoop, 16S, COI, 12S e Fib-7. A quantidade de sítios invariáveis foi maior no Fib-7, com uma porcentagem significativamente menor que nas outras regiões gênicas (tabela 3.2). Isso faz sentido, já que as taxas de substituição no núcleo são mais baixas que na mitocôndria (Scheffler, 1999; Graur & Li, 2000; Avise, 2004).

Os gráficos de saturação (figura 3.7) indicaram duas peculiaridades nas regiões que não codificam proteínas, em relação às que codificam: 1. Níveis relativamente elevados de saturação para todas as regiões a partir de certo ponto (em média, ao redor de distâncias patrísticas da ordem de 0,2); 2. Taxas de evolução molecular para a terceira posição do codon maiores que nas regiões não-codificantes (como indicado pela inclinação dos gráficos até o ponto onde a saturação começa a ser detectada). Uma explicação para essas duas características seria que muitas das substituições possíveis na terceira posição do codon são silenciosas, tornando até certo ponto neutras as mudanças de bases; já nos genes ribossômicos, uma taxa equivalente de substituições poderia alterar a estrutura secundária do RNA gerado, acarretando em mal funcionamento da tradução nos ribossomos. No caso do DLoop, o esperado seria uma taxa mais elevada que no 12S e 16S, já que essa região não é codificante; no entanto, a região controladora é responsável pelo início da duplicação e transcrição mitocondriais (Scheffler, 1999), o que também é congruente com uma hipótese de seleção purificadora. As taxas nas terceiras posições dos genes que codificam proteínas, além de serem altas quando comparadas ao DLoop, 12S e 16S, são muito semelhantes de gene para gene^{*}. O mesmo não ocorreu nas primeiras e segundas posições dos codons: o ND2 mostrou uma taxa mais elevada que os outros genes para essas posições (figura 3.7); como o gene todo foi seqüenciado, problemas relativos ao tamanho amostral, ou então a algum tipo de enviesamento posicional do fragmento analisado foram descartados. Possíveis explicações são a existência de seleção direcional em algumas linhagens (e portanto seleção disruptiva entre linhagens com pressões diferenciadas); ou então um indício de que o gene possui muitos sítios relativamente neutros, nos quais diferentes bases seriam aceitas sem interferir significativamente na função do gene.

3.5.2 - *R. v. ariel SE* x *R. v. ariel N*

Weckstein (2005) obteve um resultado em que as formas de *R. v. ariel* do sudeste e do norte do Brasil não formaram uma unidade monofilética. O autor, no entanto, analisou apenas um espécime representante da distribuição encontrada no sudeste, o que não

^{*} com exceção da ATPase 8 - estranhamente, distâncias patrísticas de MV obtidas entre táxons para este gene foram maiores que 1,0 para muitas comparações, tendo este resultado se mantido após re-análises no Modeltest e PAUP; isso pode ser o efeito de um pequeno número amostral gerando estimativas paramétricas imprecisas (Felsenstein, 2003).

descarta a hipótese de um artefato filogenético relacionado a eventuais contaminações e/ou troca de amostras. Para o presente estudo, duas amostras pertencentes à distribuição do sudeste foram analisadas. Pode-se observar, nas filogenias individuais apresentadas na figura 3.8, que há dois clados filogeneticamente distintos de *R. v. ariel*, um do sudeste da América do Sul (*R. v. ariel* SE) que é basal aos *R. vitellinus* ssp. restantes, e outro da Amazônia (*R. v. ariel* N), próximo a *R. v. vitellinus* e a *R. v. culminatus*, corroborando o estudo de Weckstein (2005). Todas as regiões seqüenciadas indicaram o mesmo resultado. Quando alinhamentos envolvendo mais de um gene foram analisados (figuras 3.9, 3.10 e 3.12), o mesmo resultado se manteve, como seria de se esperar. Assim sendo, constatou-se que o táxon não constitui um clado natural. A tabela 3.4 corrobora ainda mais essas observações, mostrando que as distâncias genéticas não corrigidas entre *R. v. ariel* SE e os outros membros do grupo *vitellinus* é mais alta que as existentes entre outras subespécies do táxon.

3.5.3 - O clado *R. vitellinus* ssp. é monofilético ou parafilético?

Quanto à questão do monofiletismo ou parafiletismo de *R. vitellinus* ssp., pode-se dizer que há evidência em favor tanto de uma quanto de outra hipótese, quando examinados os genes mitocondriais individualmente: o 12S e o ND2 indicaram que o clado é parafilético (*bootstrap* $\geq 70\%$); por outro lado, Dloop e Cyt-b sugeriram que o clado deve ser monofilético (figura 3.8). Os outros genes não mostraram um sinal claro. Entretanto, o íntron nuclear Fib-7 indica também que o clado é monofilético; por não fazer parte do mesmo grupo de ligação que os genes mitocondriais, e por apresentar características de evolução molecular diferentes em relação aos genes mencionados anteriormente, pode-se considerar que a evidência em favor do monofiletismo da espécie é maior que a hipótese contrária, quando os genes são estudados individualmente. Quando partições compostas foram analisadas o resultado foi sempre favorável à hipótese de monofiletismo, com os índices de *bootstrap* de MP e MV acima dos 70%, e PP altas (figuras 3.9, 3.10, 3.12a; e em menor escala, figura 3.12b).

Para testar de forma mais concisa a questão do monofiletismo ou não de *R. vitellinus* ssp., um teste de *bootstrap* paramétrico foi implementado. O valor de δ para os dados originais ($\delta=13,8$) foi então comparado com o histograma de distribuição de valores

dos δ para os dados simulados. O resultado não foi considerado significativo (figura 3.11), portanto não se pôde rejeitar a hipótese de monofiletismo, um resultado congruente com as análises filogenéticas do conjunto total de dados, mostradas acima.

3.5.4 - O posicionamento dos táxons *R. dicolorus* e *R. sulfuratus*

Com relação a *R. dicolorus*, dois genes (ATPase 6 e ND2) indicaram o posicionamento topológico conforme o obtido por Weckstein (2005), sendo que os outros genes não evidenciaram uma posição bem sustentada (*bootstrap* $\geq 70\%$) do táxon em qualquer outro ponto da árvore (figura 3.8). Os alinhamentos baseados no conjunto de genes codificantes de proteínas, assim como no conjunto total de dados, corroboraram o posicionamento do táxon como grupo irmão do componente [*brevis* + *vitellinus*] (figuras 3.9, 3.10a e 3.12b), embora a análise bayesiana com 6 partições, sem a inclusão do Fib-7, tenha apresentado um valor de PP abaixo de 95% (figura 3.10b).

Quanto à espécie *R. sulfuratus*, uma posição basal em relação a *R. dicolorus* foi indicada pelo gene ATPase 6, e de forma menos robusta pelo ND2 (figura 3.8). Além disso, o DLoop mostrou uma inserção relativamente mais basal deste táxon em comparação a *R. dicolorus*, com *bootstrap* $\geq 70\%$ (para esse marcador, *R. sulfuratus* foi basal em relação a todas as outras espécies de *Ramphastos*). O alinhamento envolvendo [DLoop + 12S + 16S] também indicou esse padrão relativo, embora com um valor de *bootstrap* baixo (52% - figura 3.12a). As análises da sequência total concatenada, assim como do conjunto dos genes codificantes de proteínas, corroboraram os resultados obtidos para ATPase 6 e ND2, situando *R. sulfuratus* como grupo irmão do clado [*R. dicolorus* + *R. brevis* + *R. vitellinus* ssp.] (figuras 3.9, 3.10 e 3.12b).

3.5.5 - *R. v. citreolaemus*

Um alinhamento para o Cyt-b formado por inclusão dos espécimes *R. v. citreolaemus* 1, 2 e 3 foi realizado, a fim de se averiguar o monofiletismo e a posição do táxon na topologia. O resultado filogenético obtido (figura 3.13) mostrou monofilia do táxon, sendo ele o grupo-irmão das formas amazônicas de *R. vitellinus* ssp. A figura 3.14a sintetiza as informações topológicas contidas nas figuras 3.9 e 3.13.

3.5.6 - Ausência de monofiletismo recíproco em subespécies de *R. vitellinus* ssp. e *R. tucanus* ssp.

Os clados *R. v. culminatus* e *R. v. ariel* N não se mostraram reciprocamente monofiléticos, o mesmo sendo verdade para as subespécies *R. t. tucanus* e *R. t. cuvieri*, em acordo com os resultados obtidos por Weckstein (2005). As tabelas 3.4 e 3.5 corroboram essa observação: as distâncias intra-taxonômicas sobrepõem as distâncias inter-taxonômicas. Todas as amostras de *R. v. culminatus*/*R. v. ariel* N, assim como as amostras de *R. t. tucanus*/*R. t. cuvieri*, foram obtidas das áreas de morfótipo puro (ou seja, fora da área de intergradação entre as respectivas subespécies), o mesmo sendo verdade para as demais subespécies de *R. vitellinus* ssp. (e também para as de *R. ambiguus* ssp.). Assim sendo, se mesmo nessas áreas já não é observado um sinal indicativo de monofiletismo recíproco e de distâncias genéticas inter-taxonômicas sobrepujando as intra-taxonômicas, então estes são indícios de que *R. v. culminatus* / *R. v. ariel* N talvez não sejam linhagens evolutivas independentes, o mesmo sendo verdade para as subespécies *R. t. tucanus* / *R. t. cuvieri*. No entanto, aqui só foram utilizados dois indivíduos por subespécie; apenas um estudo com maior número de espécimes e pontos de coleta é que poderá estimar mais claramente se é este ou não o caso.

3.5.7 - Digressões biogeográficas

A tabela 3.6 indica que as divergências entre categorias taxonômicas acima de gênero teriam ocorrido na faixa Eoceno Inferior/Plioceno Inferior; entre gêneros, Oligoceno Superior/Mioceno Superior; entre espécies, Mioceno Superior/Pleistoceno Inferior; e entre subespécies, Plioceno Médio/Pleistoceno Superior. Pode-se perceber que há um certo padrão temporal para os níveis de divergência na família Ramphastidae, com as separações entre gêneros em geral ocorrendo ao longo do Mioceno, entre espécies no Plioceno, e entre subespécies no Pleistoceno (tabela 3.6). Isso parece indicar uma correlação entre o nível taxonômico e a distância genética, o que é compatível com a hipótese de relógio molecular. Tal associação faz sentido, pois o teste de constância de taxas não rejeitou a hipótese de relógio molecular (apesar do p-value para o teste de hLRT obtido ter sido de 0,057, um valor marginal a um $\alpha=0,05$).

Nahum e col. (2003) estimaram divergências dentro dos Picidae e Ramphastidae, detectando uma divergência entre estes de ~82,6m.a.. Este valor é bastante remoto, sendo tal divergência localizada temporalmente no Cretáceo Superior. Como os próprios autores citam em relação aos ramfastídeos, isso seria evidência em favor do surgimento de algumas das linhagens atuais de aves antes da extinção do Cretáceo/Terciário (Cooper & Penny, 1997; Cracraft, 2001; Van Tuinen & Hedges, 2001). O método utilizado naquele trabalho, entretanto, foi a adoção de uma árvore linearizada e exclusão dos táxons que diferiram significativamente da hipótese de constância de taxas, o que pode não ser a maneira mais recomendada de se estimar datações (seção 1.2.6); além disso, a topologia foi enraizada com *Gallus gallus*, pertencente à ordem Galliformes, um grupo filogeneticamente distante dos Piciformes (Ericson *et al.*, 2006; Livezey & Zusi, 2007), o que pode ter comprometido as datações estimadas devido a graus elevados de saturação, dado o longo tempo de separação entre essas duas linhagens. Os dados presentes não mostram divergências tão remotas, situando a separação Picidae/Ramphastidae entre o Eoceno Inferior e o Oligoceno Inferior.

A separação entre *Ramphastos* e aracarís, segundo nossos dados, teria se dado entre o Oligoceno Superior/Mioceno Médio; o surgimento de gêneros em outras famílias de aves também foi relacionado ao Mioceno (Psittacidae - Tavares *et al.*, 2006; Cracidae - Pereira & Baker, 2004). Durante esse período, houve atividade geológica intensa, mudando drasticamente a geografia e o clima globais (Putzer, 1984; Burton *et al.*, 1997; Lundberg *et al.*, 1998; Burnham & Graham, 1999). Aqui no hemisfério sul, estava ocorrendo um grande pulso no soerguimento das cadeias ocidental e central dos Andes, que mudou o clima, as bacias de drenagem e a posição de vários rios, assim como transgressões marinhas de médio e grande porte (Lundberg *et al.*, 1998). Tais fatores podem ter sido importantes para a diversificação de linhagens em larga escala durante o Mioceno.

As especiações dentro de *Ramphastos* teriam se iniciado entre o Mioceno Superior e o Pleistoceno Inferior, e continuado até bem próximo do presente (tabela 3.6). A espécie mais basal dentro do gênero, *R. toco*, teria se originado entre o Mioceno Superior e o Plioceno Inferior. Os eventos mais notáveis durante o fim do Mioceno e início do Plioceno são uma aceleração no soerguimento dos Andes (Lundberg *et al.*, 1998; Burnham & Graham, 1999) e um aumento gradual de áreas de vegetação aberta, conforme extrapolado

a partir da expansão de habitats ocupados por gramíneas do tipo C₄ nessa época (Latorre *et al.*, 1997). Este deve ter sido um fator importante na evolução de *R. toco*, pois dentre todos os *Ramphastos sp.*, esta é a única espécie adaptada a áreas abertas (Haffer, 1974).

A separação entre *R. tucanus ssp./R. ambiguus ssp.* tem datação no período Plioceno. Alguns dos fatores paleogeográficos mais importantes nesse período foram um pulso de soerguimento da cadeia oriental dos Andes, e o gradual fechamento do istmo do Panamá ao final dessa época (Lundberg *et al.*, 1998; Burnham & Graham, 1999), o que teria afetado o clima na América do Sul. Essas duas hipóteses podem ajudar a explicar a evolução dos *yelpers* - embora o fator mais verossímil seja a primeira (soerguimento), já que as duas espécies são separadas parcialmente pela cadeia oriental dos Andes; no entanto, *R. a. ambiguus* ocorre nas escarpas orientais dessa cadeia, o que não anula a possibilidade de que a espécie *R. ambiguus ssp.* tenha evoluído a partir da Amazônia. Datações do mesmo período sugerindo diversificação intra-genérica nos neotrópicos foram encontrados em pelo menos outro grupo de aves (Psittacidae - Ribas *et al.*, 2005).

Já as diversificações entre os *R. vitellinus ssp.* e os *R. ambiguus ssp.* se deram respectivamente no Plioceno/Pleistoceno e no Pleistoceno. Nessa época, os ciclos paleoclimáticos de Milankovitch já estariam causando a fragmentação de habitats previamente contínuos (Haffer, 1997). Com relação aos *R. ambiguus ssp.*, nota-se que há uma clara separação filogenética entre as subespécies (embora tenha-se analisado apenas um indivíduo de *R. a. ambiguus*, e dois de *R. a. swainsonii*). Uma das cadeias dos Andes isola completamente as distribuições de *R. a. ambiguus* e *R. a. swainsonii*, um fator que deve estar associado à separação, apesar desta ser considerada relativamente recente, tanto pelo presente estudo, quanto por Haffer (1974). Entretanto, não há como dizer se os Andes seriam uma barreira primária ou secundária entre as duas subespécies, ou seja, se ela foi responsável pela vicariância, ou se o pulso de soerguimento do Plioceno teria se dado após o isolamento e evolução em separado dessas duas populações em refúgios, que nos períodos glaciais teriam se expandido, já diferenciadas, até a base das cordilheiras.

Com relação ao complexo *vitellinus*, o padrão de diversificação é mais complexo (figuras 3.9, 3.10 e 3.12a), sugerindo *R. v. ariel* SE como grupo basal dos demais *R. vitellinus ssp.* da (Amazônia+Andes) no Plioceno/Pleistoceno, com todas as diversificações posteriores ocorrendo no Pleistoceno. Uma sequência plausível de eventos levando ao

padrão de cladogênese observado em *R. vitellinus ssp.* (figuras 3.14a e b) seria, inicialmente, o isolamento biogeográfico entre o ancestral de *R. v. ariel* do sudeste da América do Sul e os táxons mais ao norte. Este isolamento no sudeste seria decorrente da ausência de ligação entre a Mata Atlântica e a Amazônia observada atualmente, ligação esta que provavelmente existiu no passado, tendo se partido quando um aumento de aridez no Terciário provocou a formação de um cinturão de formações xeromórficas entre elas (Bigarella *et al.*, 1975; Ledru, 1993; Prado & Gibbs, 1993). Posteriormente, houve uma estruturação de *R. v. citreolaemus* entre as cordilheiras dos Andes, talvez devido ao crescente soerguimento da cordilheira oriental nessa época (Putzer, 1984; Lundberg *et al.*, 1998; Burnham & Graham, 1999), através também de alopatria; pode-se especular que a zona de intergradação existente hoje entre essa subespécie e *R. v. culminatus*, existente na região mais ao norte dos Andes onde termina a elevação da cordilheira (Haffer, 1974), possa ser um efeito de contato secundário, já que a monofilia de *R. v. citreolaemus* é bem estabelecida pelo índice de *bootstrap*, e *R. v. citreolaemus* não é o grupo-irmão de *R. v. culminatus* (figuras 3.13 e 3.14). O evento seguinte teria sido a evolução de *R. v. vitellinus* ao norte do rio Amazonas, sugerindo este rio como mecanismo de diversificação importante (esta hipótese torna-se mais consistente quando examinadas as localidades das amostras do presente estudo: ambos os indivíduos de *R. v. vitellinus* estão mais próximos de *R. v. ariel* N do que estes últimos estão de *R. v. culminatus*, desfavorecendo um cenário de isolamento por distância). Aqui é mais provável que esta tenha sido uma barreira primária, já que o rio Amazonas teria se formado de 8 a 11 m.a. (Hoorn, 1994; Lundberg *et al.*, 1998), muito antes da separação estimada entre essas duas subespécies.

Eventuais pressões seletivas podem ter permitido a grande variação morfológica encontrada entre subespécies *R. v. ariel* N e *R. v. culminatus*, encontradas abaixo do rio Amazonas. Nossos dados indicam que tal variação morfológica não possui um paralelo em nível molecular, já que não foi detectado monofiletismo recíproco entre elas (figuras 3.9, 3.10, 3.12b e 3.13), e além disso há sobreposição entre as distâncias intra-taxonômicas e as inter-taxonômicas (tabela 3.4). Haffer (1974) já havia descrito uma ampla região de intergradação entre *R. v. vitellinus* e *R. v. culminatus* na margem norte do rio Amazonas, assim como entre este último e *R. v. ariel* N na margem sul do rio. Segundo nossos resultados filogenéticos (figuras 3.9, 3.10, 3.12b e 3.13), então, o primeiro caso seria um

evento de re-expansão e contato secundário, e o segundo caso uma instância de divergência incipiente (fluxo gênico e/ou divergência recente) ou ainda uma clina estável. Apenas um estudo filogeográfico, mais pormenorizado, poderá distinguir entre essas diferentes alternativas. Um fator condizente com a diferenciação morfológica entre *R. v. culminatus* e *R. v. ariel* N seria o mimetismo agressivo (Sick, 1997, Short & Horne, 2001) de *R. v. culminatus* em relação a *R. t. cuvieri*, já que eles são simpátricos em grande parte de suas respectivas distribuições (figuras 3.2 e 3.3), e ambos apresentam exatamente a mesma coloração corporal e do bico - essa mesma explicação serve para justificar a simpatria entre *R. brevis* e *R. a. swainsonii*, na região trans-andina da América do Sul, já que em ambos os casos similaridade morfológica devido a um parentesco evolutivo está descartada (figuras 3.9, 3.10 e 3.12). Segundo essa série de eventos, ainda poderia existir uma pressão seletiva no sentido de manter o morfótipo de *R. v. ariel* N igual ao de *R. v. ariel* SE, ou por outro lado poderia não ter surgido pressão seletiva suficiente para modificações de coloração na parte norte da distribuição de *R. v. ariel*.

Quanto às subespécies de *R. tucanus* ssp., não houve diferenciação molecular perceptível, embora haja diferenças morfológicas, espelhando o que foi encontrado para *R. v. culminatus* / *ariel* N; no primeiro caso, no entanto, essas diferenças fenotípicas não são exacerbadas como no último, conforme reconhecido por Haffer (1974) e Short e Horne (2001), a principal diferença sendo o quão escura é a faixa escarlate na lateral do bico (em *R. t. cuvieri*, a oeste da Amazônia, esta é quase totalmente preta). Aqui, portanto, a separação em subespécies parece ser ainda menos justificada do que no caso das subespécies de *R. vitellinus* ssp. mencionadas acima.

A sequência cladogenética observada nas figuras 3.14a e 3.14b pode indicar sucessivas colonizações a partir de eventuais conexões entre os Andes e o sudeste da América do Sul. A existência de tais conexões foi inferida por Willis (1992), baseando-se num grande número de distribuição de aves. Uma provável explicação seria a presença de ciclos climáticos, que propiciariam o isolamento das proto-espécies em fragmentos de mata no sudeste e no noroeste da América do Sul nos períodos glaciais, com posterior expansão das espécies/subespécies recém-evoluídas nos períodos interglaciais. Durante estes últimos, teriam existido conexões entre o noroeste e sudeste da América, via oeste flanqueando as bordas dos andes, ou mais ao norte, ou pela Amazônia (seja por uma conexões mais

centrais na América do Sul ou então descendo pela costa leste brasileira), que seriam novamente interrompidas durante os ciclos de aridez; isso permitiria a divergência sucessiva de *R. sulfuratus*., *R. dicolorus*, *R. brevis* e *R. ariel* NE. Inicialmente, Haffer (1974) sugeriu que tais ciclos só tivessem ocorrido durante o Pleistoceno, mas mais recentemente ele passou a admitir a existência deles em épocas anteriores (Haffer, 1997, Haffer & Prance 2001). Vale ressaltar que esse modelo de constrição drástica de florestas durante as épocas glaciais, no entanto, não é universalmente aceito (Bush, 1994; Colinvaux *et al.*, 2001; Cowling *et al.*, 2001).

3.5.8 - *R. v. ariel*: considerações taxonômicas

Conforme já mencionado, *R. v. ariel* SE e *R. v. ariel* N são reciprocamente monofiléticos; além disso, não são grupos irmãos, sendo o primeiro basal aos demais membros do complexo *vitellinus*, e o segundo próximo a *R. v. culminatus*. Por fim, as distâncias genéticas indicam que se trata de duas linhagens muito divergentes.

Poderia-se propôr que *R. v. ariel* do sudeste fosse elevado à condição de espécie após o presente estudo, mas isto ainda não é aconselhável. Ocorre que Vigors, em 1826, fez a descrição de *R. v. ariel* baseado no morfótipo do sudeste, tendo esta linhagem a prioridade nomenclatural; assim, *R. v. ariel* SE deveria se tornar simplesmente *R. ariel*. O táxon do norte do Brasil passaria a ter então outro nome;

Não obstante, como mostrado acima, os táxons *R. v. culminatus* e *R. v. ariel* N não apresentam diferenciação genética detectável, ao menos considerando-se os espécimes estudados. Caso houvesse uma separação clara, poderia-se adotar um novo nome para *R. v. ariel* N, e a questão estaria resolvida. Caso tal separação evolutiva entre esses dois clados realmente não exista, então ambas as linhagens poderiam ser englobadas em *R. v. culminatus*. Porém, como no presente estudo foram incluídos apenas dois indivíduos por subespécie, não se tem ainda definição quanto à independência evolutiva ou não entre essas duas linhagens - assim, o mais indicado é adotar uma posição conservativa e aguardar o resultado baseado em um trabalho filogeográfico com maior número de espécimes de *R. v. ariel* N e *R. v. culminatus*, com mais localidades de coleta.

Referências bibliográficas - Capítulo 3

- Avice, J.C. (2004): *Molecular markers, natural history, and evolution* - 2ª edição. Sinauer Associates, Inc., Sunderland, MA.
- Barker F. K.; Lanyon, S. M. (2000): The Impact of Parsimony Weighting Schemes on Inferred Relationships among Toucans and Neotropical Barbets (Aves: Piciformes). *Mol. Phyl. Evol.* 15(2): 215–234.
- Bigarella, J. J.; Andrade-Lima, D.; Riehs, P. J. (1975): Considerações a respeito das mudanças paleoambientais na distribuição de algumas espécies vegetais e animais no Brasil. *Anais da Academia Brasileira de Ciências* 47(Supl.): 411-464.
- Braun, E. L.; Kimball, R. T. (2002): Examining basal avian divergences with mitochondrial sequences: model complexity, taxon sampling, and sequence length. *Syst. Biol.* 51: 614-625.
- Burnham, R. J.; Graham, A. (1999): The History of Neotropical vegetation: new developments and status. *Annals of the Missouri Botanical Garden* 86:546-589.
- Burton, K. W.; Ling, H.-F.; O'Nions, R. K. (1997): Closure of the Central American Isthmus and its effect on deep-water formation in the north Atlantic, *Nature* 386: 382-385.
- Bush, M. B. (1994): Amazonian speciation - a necessarily complex model. *J. Biogeogr.* 21: 5-17.
- Colinvaux, P. A.; Irion, G.; Räsänen, M. E.; Bush, M. B.; Nunes de Mello, J. A. S. (2001): A paradigm to be discarded - geological and paleoecological data falsify the Haffer & Prance refuge hypothesis of Amazonian speciation. *Amazoniana*, 16, 609-646.
- Cooper, A.; Penny, D. (1997): Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. *Science* 275: 1109-1113.
- Cowling, S. A.; Maslin, M. A.; Sykes, M. T. (2001): Paleovegetation simulations of lowland Amazonia and implications for neotropical allopatry and speciation. *Quaternary Res.* 55: 140-149.
- Cracraft, J. (1981): Toward a phylogenetic classification of birds of the world (class Aves). *Auk* 98: 681-714.

- Cracraft, J. (2001): Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. *Proc. R. Soc. B* 268: 459-469.
- Cunningham, C. W. (1997): Can three incongruence tests predict when data should be combined? *Mol. Biol. Evol.* 14(7): 733-740.
- Drummond, A. J.; Rambaut, A. (2003): BEAST v1.0 - Available from <http://evolve.zoo.ox.ac.uk/beast/>.
- Drummond, A. J.; Ho, S. Y. W.; Phillips, M. J.; Rambaut, A. (2006): Relaxed phylogenetics and dating with confidence. *PLOS Biology* 4(5): 699-710.
- Dyke G.J.; Van Tuinen, M. (2004): The evolutionary radiation of modern birds (Neornithes): reconciling molecules, morphology and the fossil record. *Zool. J. Linn. Soc.* 141: 153-177.
- Ericson, P. G. P.; Anderson, C. L.; Britton, T.; Elzanowski, A.; Johansson, U. S.; Källersjö, M.; Ohlson, J. I.; Parsons, T. J.; Zuccon, D.; Mayr, G. (2006): Diversification of Neoaves – integration of molecular sequences data and fossils. *Biol. Lett.* 2(4): 543-547.
- Espinosa de los Monteros, A. (2000): Higher-level phylogeny of Trogoniformes. *Mol. Phyl. Evol.* 14(1): 20-34.
- Espinosa de los Monteros, A. (2003): Models of the Primary and Secondary Structure of the 12S RNA of Birds: A Guideline for Sequence Aligning. *DNA Sequence* 14: 241-256.
- Fain, M. G.; Houde, P. (2004): Parallel radiations in the primary clades of birds. *Evolution* 58(11): 2558-2573.
- Feduccia, A. (1999): *The Origin and Evolution of Birds* - 2^a edição. Yale University Press, New Haven, CT.
- García-Moreno, J. (2004): Is there a universal mtDNA clock for birds?. *J. Avian Biol.* 35: 465-468.
- García-Moreno, J.; Sorenson, M. D.; Mindell, D. P. (2003): Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. *Journal of Molecular Evolution* 57: 27-37.
- Graur, D.; Li, W.-H. (2000): *Fundamentals of molecular evolution* - 2^a edição. Sinauer Associates, Inc., Sunderland, MA.

- Gutell, R. R.; Weiser, B.; Woese, C.; Noller, H. F. (1985): Comparative anatomy of 16S-like ribosomal RNA. *Nuc. Acid. Res.* 32: 155-215.
- Haffer, J. (1974): *Avian speciation in tropical South America*. Coleção “Publications of the Nuttall Ornithological Club”, n° 14, Cambridge, MA.
- Haffer, J. (1997): Alternative models of vertebrate speciation in Amazonia: An overview. *Biodiversity and Conservation* 6: 451-476.
- Haffer, J.; Prance, G. T. (2001): Climatic forcing of evolution in Amazonia during the Cenozoic: on the refuge theory of biotic differentiation. *Amazoniana* 16: 579–607.
- Ho, S. Y. W.; Phillips, M. J.; Cooper, A.; Drumond, A. J. (2005): Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* 22(7): 1561-1568.
- Höfling, E. (1991): Etude comparative du crane chez des Ramphastidae (Aves, Piciformes). *Bonner Zoologische Beiträge* 42(1): 55-65.
- Höfling, E. (1998): Comparative cranial anatomy of Ramphastidae and Capitonidae. *Ostrich* 69: 389-389.
- Hoorn, C. (1994): An environmental reconstruction of the palaeo-Amazon River system (Middle-Late Miocene, NW Amazonia). *Palaeogeography, Palaeoclimatology, Palaeoecology* 112: 187–238.
- Kocher, T. D.; Thomas, W. K.; Meyer, A.; Edwards, S. V.; Pääbo, S.; Vallabhan, F. X., Wilson, A. C. (1989): Dynamics of mitochondrial DNA evolution in animals - amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci., USA* 86: 6196-6200.
- Lanyon, S. M.; Zink, R. M. (1987): Genetic variation in piciform birds: monophyly and generic and familial relationships. *The Auk* 104: 724-732.
- Latorre, C.; Quade, J.; McIntosh, W. C. (1997): The expansion of C4 grasses and global change in the late Miocene: stable isotope evidence from the Americas. *Earth Planet Sci. Lett.* 146: 83-96.
- Ledru, M.-P. (1993): Late quaternary environmental and climatic changes in central Brazil. *Quaternary Research*, 39: 90-98.

- Livezey, B. C.; Zusi, R. L. (2007): Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool. J. Linn. Soc.* 149: 1-95.
- Lundberg, J. G.; Marshall, L. G.; Guerrero, J.; Horton, B.; Malabarba, M. C.; Wesselingh, F. (1998): *The Stage for Neotropical Fish Diversification: A History of Tropical South American Rivers*. In: Phylogeny and Classification of Neotropical Fishes. Eds: L.R. Malabarba, R.E.Reis, R.P.Vari, C.A.S.Lucena & Z.M.S.Lucena eds.. Museu de Ciências e Tecnologia, PUCRS. Porto Alegre, Brazil.
- Maddison, W. P.; Maddison, D. R. (2006): Mesquite - a modular system for evolutionary analysis. Version 1.12 <http://mesquiteproject.org>.
- Mayr, G. 2005. A tiny barbet-like bird from the lower Oligocene of Germany: the smallest species and earliest substantial fossil record of the Pici (woodpeckers and allies). *Auk* 122, 1055-1063.
- Moyle, R.G. (2004): Phylogenetics of barbets (Aves: Piciformes) based on nuclear and mitochondrial DNA sequence data. *Mol. Phyl. Evol.* 30: 187-200.
- Nahum, L. A.; Pereira, S. L.; Fernandes, F. M. C.; Matioli, S. R.; Wajntal, A. (2003): Diversification of Ramphastinae (Aves, Ramphastidae) prior to the Cretaceous/Tertiary boundary as shown by molecular clock of mtDNA sequences. *Genetics and Molecular Biology*, 26(4): 411-418.
- Pereira, S. L.; Grau, E. T.; Wajntal, A. (2004): Molecular architecture and rates of DNA substitutions of the mitochondrial control region of cracid birds. *Genome* 47: 535-545.
- Pereira, S. L.; Baker, A. J. (2004): Vicariant speciation of curassows (Aves, Cracidae) – a hypothesis based on mitochondrial DNA phylogeny. *Auk* 121(3): 682-694.
- Prado, D. E.; Gibbs, P. E. (1993): Patterns of species distributions in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden* 80: 902-927.
- Prum, R. O. (1988): Phylogenetic interrelationships of the barbets (Aves: Capitonidae) and toucans (Aves: Ramphastidae) based on morphology with comparisons to DNA-DNA hybridization. *Zool. J. Linn. Soc.* 92: 313-343.
- Putzer, H. (1984). *The geological evolution of the Amazon Basin and its mineral resources*. In: The Amazon. Ed: Sioli, H.. W. Junk Publishers, Dordrecht, The Netherlands. P. 15-46.

- Rambaut, A.; Drummond, A. (2006) Tracer v1.3 - A program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes.
- Ribas, C. C.; Gaban-Lima, R.; Miyaki, C. Y.; Cracraft, J. (2005): Historical biogeography and diversification within the neotropical parrot genus *Pionopsitta* (Aves: Psittacidae). *J. Biogeogr.* 32: 1409-1427.
- Rich, P. V.; Bohaska, D. J. (1976): The World's Oldest Owl: A new strigiform from the Paleocene of southwestern Colorado. *Smithsonian Contributions to Paleobiology* 27, 87-93.
- Scheffler, I. E. (1999): *Mitochondria*. John Wiley & Sons, Inc., New York, NY.
- Short, L. L.; Horne, J. F. M. (2001): *Toucans, Barbets and Honeyguides*. Coleção "Bird Families of the World". Oxford University Press, New York, NY.
- Sick, H. (1997): *Ornitologia Brasileira* (Tradução). Ed. Nova Fronteira, Rio de Janeiro, RJ.
- Swierczewski, E. V. & Raikow, R. J. (1981): Hind limb morphology, phylogeny and classification of the Piciformes. *The Auk* 98: 466-480.
- Tavares, E. S.; Baker, A. J.; Pereira, S. L.; Miyaki, C. Y. (2006): Phylogenetic relationships and historical biogeography of neotropical parrots (Psittaciformes: Psittacidae: Arini) inferred from mitochondrial and nuclear DNA sequences. *Syst. Biol.* 55(3): 454-470.
- Van Tuinen, M.; Hedges, S. B. (2001): Calibration of avian molecular clocks. *Mol. Biol. Evol.* 18: 206-213.
- Van Tuinen, M.; Sibley, C. G.; Hedges, S. B. (2000): The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. *Molecular Biology and Evolution* 17: 451-457.
- Weckstein, J. D. (2005): Molecular phylogenetics of the *Ramphastos* toucans - implications for the evolution of morphology, vocalizations, and coloration. *Auk* 122(4): 1191-1209.
- Willis, E. O. (1992): Zoogeographical origins of eastern brazilian birds. *Ornitologia Neotropical* 3: 1-15.

Capítulo 4 - Conclusões

O presente trabalho buscou analisar aspectos tanto de micro-evolução quanto de macro-evolução do gênero *Ramphastos*. O capítulo 2, totalmente focado na dinâmica evolutivo-molecular do gene mitocondrial ribossomal 12S no grupo, visou estimar apropriadamente a estrutura secundária da molécula; estudos onde esta etapa não é realizada podem trazer enviezamentos nas análises, devido a erros no alinhamento que podem ser gerados, e também à presença de eventuais ramos com suporte excessivo nas reconstruções filogenéticas. No presente estudo não houve diferenças quanto à resolução topológica usando-se o 12S inteiro ou incluindo-se alternadamente [alças + hastes 5'] ou [alças + hastes 3']. O nível taxonômico do estudo talvez tenha sido o responsável, já que foram analisadas amostras dentro de um mesmo gênero; é de se esperar que quanto maior o nível de divergências estudado, maior tende a ser o erro no alinhamento e durante a etapa de reconstrução filogenética. O trabalho de Espinosa de los Monteros (2003) mostra isso: o alinhamento baseado na estrutura secundária foi fundamental para uma maior precisão em relação a topologias aceitas na comunidade ornitológica (mas o autor usou sempre o 12S inteiro).

O capítulo traz também a descrição de um modelo de Markov inédito na literatura, o RNA_8, voltado para padrões de evolução molecular nas regiões de hastes. O desenvolvimento de modelos markovianos verossímeis é importante para uma apreciação embasada dos processos moleculares em nível de DNA. O RNA_8 possui, como o próprio nome sugere, oito classes de taxas relativas. Modelos de correlação entre taxas descritos na literatura possuem 6, 7 ou 16 estados, nos dois primeiros casos deixando de levar em consideração estados relativamente freqüentes no conjunto de dados, e no último caso incluindo muitas dimensões, o que pode não ser indicado dado que muitos graus de liberdade são perdidos no processo, diminuindo a precisão das estimativas de taxas, e portanto seu poder explanatório (vale lembrar que as regiões de hastes, pelo menos no 12S, não passam de ~300 pareamentos de bases no total). O RNA_8 tem dois atrativos: 1. número balanceado de dimensões, conforme mencionado; 2. número balanceado de parâmetros (além das freqüências de cada par, tem-se apenas três parâmetros, um para

substituições simples, outro para duplas transições, e outro para duplas transversões). A tabela 2.7, obtida da análise com o RNA_8, consegue captar informações importantes acerca da evolução molecular da região estudada, não deixando de incluir estados relativamente freqüentes (como seria o caso com modelos de 6 e 7 dimensões), e ao mesmo tempo sendo mais econômico que modelos com 16 dimensões, sem perder precisão. Sua utilidade ou não para estudos futuros é ainda uma incógnita; suas propriedades devem ser testadas quanto à acurácia para reconstruções filogenéticas em dados simulados, precisão para reconstruções filogenéticas em dados reais, e testes estatísticos em relação a modelos alternativos de oito dimensões para uma mesma topologia e um mesmo conjunto de dados. Seguramente, sua aplicabilidade estará longe de ser universal; de qualquer forma, o RNA_8 é uma opção a mais para estudos evolutivos envolvendo correlação entre sítios nucleotídicos.

O capítulo 3, apesar de também apresentar análises e digressões em nível de evolução molecular, teve seu enfoque principal na história natural das espécies do gênero *Ramphastos*. Os tucanos são emblemáticos para estudos ornitológicos tanto pela peculiaridade na forma do bico, quanto por terem sido um dentre dois grupos estudados por Jürgen Haffer em sua monografia de 1974, intitulada “*Avian speciation in tropical South America*”, que é sob muitos aspectos um marco (e uma fonte) em estudos evolutivos e biogeográficos em aves, além de ter sido o primeiro trabalho a propôr uma hipótese de parentesco entre todas as espécies do grupo a partir de medições morfológicas e padrões de vocalização.

Weckstein (2005) estudou o grupo a partir de seqüências de DNA mitocondrial (~2.500 bases), obtendo uma filogenia que em alguns pontos foi congruente com a hipótese de Haffer, e em outros incongruentes. Os principais resultados obtidos por ele foram o parafiletismo dos *croakers*, o parafiletismo de *R. vitellinus* ssp., com a inserção da espécie *R. brevis* dentro desse táxon, e uma aparente falta de diferenciação genética entre as subespécies *R. v. culminatus/ ariel* N e *R. t. tucanus/ cuvieri*.

O presente estudo apresentou uma maior inclusão taxonômica e gênica dentro do grupo, a fim de buscar resolver as questões ainda em aberto. Obteve-se um resultado muito parecido com o de Weckstein; as novidades foram um aparente retorno do complexo *vitellinus* ao status de espécie monofilética (e grupo irmão de *R. brevis*), e o

posicionamento de *R. v. citreolaemus* como grupo irmão das subespécies do complexo *vitellinus* que ocorrem na Amazônia. Uma das regiões gênicas usadas nesse estudo foi o 12S; entretanto, não se usou a sequência inteira, mas sim apenas as regiões de alça e as hastes proximais - justamente para não cometer os enviezamentos analíticos mencionados anteriormente (alinhamento inadequado e imprecisão topológica). Nesse sentido, o estudo do capítulo 2 foi fundamental para que os estudos realizados no capítulo 3 fossem feitos de forma correta, já que o alinhamento total incluiu também esse subconjunto do 12S.

Datações de divergências são ainda um campo em desenvolvimento; muitas técnicas utilizadas no passado já são consideradas obsoletas em face de alternativas mais atuais (ex: métodos que assumem uma única taxa evolutiva ao longo da topologia); possivelmente, muitos dos métodos correntes na literatura também o serão no futuro. Aqui foram utilizados quatro métodos, todos assumindo premissas diferentes, dois deles amplamente utilizados na literatura (relógio molecular e PL), e dois ainda pouco utilizados (PATHd8 e análise bayesiana pelo programa BEAST). Como o objetivo principal aqui foi o de buscar tempos de divergência assumindo o menor número de restrições possível, todos foram utilizados, e optou-se por gerar intervalos a partir dos tempos mínimo e máximo obtidos, mesmo que o mínimo e o máximo não viessem da mesma metodologia - assim, buscou-se errar o menos possível, optando-se por um intervalo mais abrangente, mas também mais conservativo.

Do ponto de vista biogeográfico, pôde-se: 1. associar a evolução de *R. toco* com a expansão de áreas abertas na América do Sul durante o fim do Mioceno; 2. identificar um padrão de alternância de cladogêneses com relação a táxons que ocupam o sudeste e o noroeste da América do Sul; 3. inferir que categorias taxonômicas semelhantes teriam evoluído mais ou menos nas mesmas épocas (gêneros, no Mioceno; espécies, no Plioceno; subespécies, no Pleistoceno).

Referências bibliográficas - Capítulo 4

- Espinosa de los Monteros, A. (2003): Models of the Primary and Secondary Structure of the 12S RNA of Birds: A Guideline for Sequence Aligning. *DNA Sequence* 14: 241-256.
- Haffer, J. (1974): *Avian speciation in tropical South America*. Coleção “Publications of the Nuttall Ornithological Club”, nº 14, Cambridge, MA.
- Weckstein, J. D. (2005): Molecular phylogenetics of the *Ramphastos* toucans - implications for the evolution of morphology, vocalizations, and coloration. *Auk* 122(4): 1191-1209.