

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS HUMANAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS**

**Padrão de variação intra e intergenérica na estrutura
proteica do Citocromo b em oito gêneros de Oryzomyini
(Rodentia)**

Eduardo Loyola Muhl

**Vitória, ES
Fevereiro, 2017**

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS HUMANAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS**

**Padrão de variação intra e intergenérica na estrutura
proteica do Citocromo b em oito gêneros de Oryzomyini
(Rodentia)**

Eduardo Loyola Muhl

Orientadora: Valéria Fagundes

**Dissertação submetida ao Programa de Pós-Graduação
em Ciências Biológicas (Biologia Animal) da
Universidade Federal do Espírito Santo como requisito
parcial para a obtenção do grau de Mestre em Biologia
Animal**

**Vitória, ES
Fevereiro, 2017**

AGRADECIMENTOS

Expresso meus sinceros agradecimentos a todos que ajudaram na conclusão desta jornada. Agradeço a Prof^ª. Dr^ª. Valéria Fagundes, que me abriu as portas do Laboratório de Genética Animal ainda na graduação me acolhendo novamente como aluno de mestrado. Sua postura, amizade, companherismo e conselhos vem me mostrando ao longo desses anos como me tornar um profissional melhor, e acima de tudo, uma pessoa melhor, estando sempre presente nos momentos de maiores dificuldades em minha carreira. Muito obrigado.

Ao Prof. Dr. Arlan da Silva Gonçalves (IFES/Vila Velha), que compartilhou seu vasto conhecimento em metodologias de modelagem proteica, que foram cruciais para a realização do presente estudo.

Ao Prof. Dr. Rinaldo Wander Montalvão (PUC/RS) por ter me recebido em seu laboratório e me auxiliado em diferentes análises e discussões dos dados, contribuindo em muito para a finalização deste trabalho.

Em especial, ao doutorando Antonio Marinho (USP/São Carlos), meu amigo Tonho, por toda a ajuda para a realização deste trabalho, estando sempre presente seja para discutir dados e metodologias ou para entregar palavras de incentivo nos momentos de maiores angústias. Muito obrigado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Universidade Federal do Espírito Santo (UFES) pelo financiamento do projeto, da bolsa de mestrado e suporte.

Aos professores do PPGBAN que se dedicaram para contribuir com a minha formação profissional e pessoal, através da ajuda na elaboração do projeto, discussões em disciplinas e auxílio no meu entendimento da metodologia científica. Muito obrigado.

Aos membros titulares e suplentes da banca por aceitarem participar da avaliação da dissertação.

Aos amigos do Laboratório de Genética Animal: Arturo Martinelli, Ana Heloisa de Carvalho, Cristina Massariol, Débora Dalvi, Fernanda Zaidan, Gabriel Dalbem, Marina Monjardim, Marianna Xavier, Rosana Nunes, Victor Colombi, e por todos que lá passaram ao longo desses anos. Levarei comigo todas as boas lembranças, alegrias, agústias e superações que compartilhamos. Em especial à Thaís Volpi, pela amizade e conselhos ao longo deste curso de Mestrado. Muito obrigado a todos vocês.

A todos os colegas de turma do Mestrado, pelo companherismo e amizade nesses últimos dois anos (e meio). Levo a certeza que, juntamente as disciplinas que cursamos juntos, a convivência e experiências compartilhadas contribuíram para minha formação. Muito obrigado.

Aos meus pais Carmen e Masir, minha avó Anésia Ferraz, e todos os meus demais familiares por me apoiarem ao longo de todo o caminho, sendo referência de amor e respeito. Graças a vocês mais uma etapa é cumprida, e devido a tudo isso vocês possuem minha eterna gratidão e amor.

Muito obrigado.

SUMÁRIO

Resumo.....	6
Introdução.....	8
Resultados e Discussão	12
Variação na sequência de DNA	12
O gene Citb mostrou-se com alto poder de agrupamento de espécies e gêneros.....	12
Geração dos modelos tridimensionais da proteína do Citb	15
Análise da estrutura da proteína do Citb	18
Variação estrutural da proteína na posição 14.....	20
Avaliando a possibilidade de modelagem do Citb a partir da metodologia de fragment-based assembly	23
Mapeamento das mutações na estrutura da proteína Citb	25
Mutações sinônimas e não-sinônimas.....	25
Porções hipervariáveis e sua importância	29
Correlação entre posição do códon e tipo de mutação	31
Variação nas cadeias cd1 e cd2.....	33
Padrão de mutação do Citb: seleção x deriva genética.....	35
Métodos.....	37
Matriz de dados	37
Modelagem do Citb.....	38
Mapeamento das mutações	38
Comparação da estrutura tridimensional do Citb.....	39
Análise da Entropia de Shannon	39
Análises estatísticas.....	40
Análises genéticas dos dados	40
REFERÊNCIAS BIBLIOGRÁFICAS.....	41
Anexo 1	48
Anexo 2:	49
Anexo 3	50
Anexo 4:	52
Anexo 5:	53

Padrão de variação intra e intergenérica na estrutura proteica do Citocromo b em oito gêneros de *Oryzomyini* (Rodentia)

(A ser submetido no periódico *BMC Genetics*, *Qualis B2*)

RESUMO

Background: O gene Citocromo b (Citb) é utilizado para elucidar as relações filogenéticas entre famílias, gêneros e espécies de eucariotos, destacando-se na identificação de diversas linhagens crípticas do gênero de roedor *Oryzomys* (Coues, 1890) na década de 2000. Em 2006, *Oryzomys* havia sido subdividido em onze gêneros, baseado em estudos de morfologia, e as análises do gene Citb (anteriores e posteriores à subdivisão) mostraram-se congruentes com a nova organização. Por esse motivo, esse estudo se tornou clássico em evidenciar a importância do Citb como um marcador gênero- e espécie-específico em roedores. Diferentes trabalhos demonstram que, ainda que o Citb seja considerado um gene de evolução rápida, é possível observar porções mais variáveis do que outras ao longo da estrutura terciária da proteína. Assim, no presente estudo, nosso objetivo foi a) investigar se as mutações que sustentam a subdivisão dos gêneros e espécies derivadas do antigo gênero *Oryzomys* alteram as sequências primárias e estruturas tridimensionais da proteína Citb; b) mapear se as mutações ocorrem em regiões específicas do gene, e se podem ser associadas à uma variação intra e/ou intergenérica da proteína Citb; e c) se há zonas quentes de variação do gene para gênero ou espécies. Dessa forma, foram tomadas 75 sequências do gene Citb obtidas disponíveis no GenBank de 24 espécies de oito dos 10 gêneros derivados de *Oryzomys*: *Aegialomys*, *Cerradomys*, *Euryoryzomys*, *Hylaeamys*, *Nephelomys*, *Oreoryzomys*, *Sooretamys* e *Transandinomys*. Foi realizada uma análise filogenética a fim de identificar as mutações que diferenciavam gêneros (análise intergenérica) e espécies de um mesmo gênero (análise intragenérica), sendo essas mutações mapeadas na cadeia proteica do Citb. A modelagem da proteína do Citb foi realizada pelas metodologias de modelagem por homologia e pela *fragment-based assembly*.

Resultados: As análises filogenéticas recuperaram o monofiletismo de todos os oito gêneros analisados, revelando 58 haplótipos exclusivos, tanto para espécies como gêneros. Embora variável na estrutura primária dos 1143 pb do gene, a estrutura tridimensional da proteína com 380 resíduos mostrou-se conservada, exceto por uma variação estrutural na posição 14, uma região de *loop* próxima ao N-Terminal da proteína, com 28% das sequências contendo Valina (V14) e 72% Isoleucina (I14), ambos aminoácidos alifáticos e

de propriedades semelhantes. A mutação V14 foi observada somente em *Euryoryzomys* e *Hylaeamys*. O mapeamento das mutações ao longo da estrutura do Citb apontou que as mutações corresponderam majoritariamente a mutações sinônimas (MS), tanto para a análise inter quanto para a intragenérica (81 e 88%, respectivamente), distribuídas ao longo de toda a extensão da proteína. Porém, as mutações ocorrem com maior frequência em regiões de α -Hélice, seguidas por mutações em região de *loop* e Folha- β , tanto nas análises intra como intergenéricas, sendo observado um maior grau de conservação nos primeiros 267 resíduos da proteína. Quanto às mutações não sinônimas (MNS), observa-se que aquelas capazes de diferenciar os gêneros ocorrem em porções distintas daquelas que diferenciam espécies. Há uma região gênero-específica (intergenérica) nos primeiros 80 aminoácidos, enquanto que uma região espécie-específica (intragenérica) ocorre entre as posições 110 e 123. Além disso, observa-se que tanto para a análise intragenérica quanto para a intergenérica, as MNS ocorreram entre aminoácidos de propriedade semelhantes quanto à classificação, carga, polaridade e tamanho da cadeia lateral. Ainda, as MNS registradas no presente estudo estão distantes dos principais sítios ativos do Citb, como os sítios Qi e Qo, demonstrando a sensibilidade desses locais à mutação.

Conclusões: Nosso estudo demonstra que, apesar da estrutura terciária do Citb mostrar-se altamente conservada, é possível identificar a presença de MNS em porções distintas quando comparadas homologias entre gêneros e entre espécies. A estrutura primária do Citb é mais variável do que a estrutura terciária, sendo a sequência de DNA informativa para as análises filogenéticas. Isso se explica pois a maioria das mutações registradas correspondem à MS e porque a maioria das MNS ocorrerem entre aminoácidos com propriedades semelhantes de tamanho e carga da cadeia lateral. Além disso, a distribuição das mutações ao longo das estruturas secundárias do Citb segue a proporcionalidade com que essas ocorrem na proteína, sendo observada uma maior quantidade de mutações em regiões de α -Hélice, seguido por mutações em *loop* e Folha β , tanto para as análises intergenéricas quanto para as intragnéricas. Ainda, foi possível observar que a ação da seleção natural mantém os principais sítios ativos do Citb inalterados, uma vez que mutações nesses locais corresponderam majoritariamente à MS ou a MNS entre aminoácidos de propriedades semelhantes. Isso explica o maior grau de conservação observado nos primeiros 267 resíduos da proteína, uma vez que essa região abriga os principais sítios ativos do Citb.

Palavras-chave: Citb; Modelagem proteica; *Oryzomys*; Oryzomyini; Homologia; Filogenia

INTRODUÇÃO

A abordagem da modelagem proteica comparativa ou modelagem por homologia é uma metodologia na qual a estrutura tridimensional de uma proteína pode ser construída a partir do alinhamento entre sequências de aminoácidos da proteína alvo e de proteínas cujas estruturas são conhecidas e que atuam como molde para a construção do modelo tridimensional desejado (Martí-Remon *et al.*, 2000). Assim, a modelagem por homologia prevê uma busca prévia por moldes com alto valor de similaridade com a sequência de aminoácidos alvo, uma vez que um alinhamento preciso se faz necessário para a inferência de um modelo proteico de alta precisão.

A correlação entre a similaridade das sequências de aminoácidos e a estrutura proteica foi originalmente identificada por Chothia & Lesk (1986) e quantificada por Sander & Schneider (1991). Em um trabalho posterior, Rost (1999) inferiu o limite preciso da aplicação dessa regra, determinando uma equação que descreve o limite entre a *safe zone* e a *twilight zone*. De forma geral, sequências nas quais o comprimento e a porcentagem de resíduos similares entre si encontram-se nos limites da *safe zone* possuem uma alta probabilidade de adotarem estruturas tridimensionais semelhantes.

As metodologias utilizadas pelos programas de modelagem por homologia inferem a estrutura da proteína alvo a partir de informações derivadas de moldes, podendo essas metodologias serem agrupadas em dois grandes grupos: o primeiro tem início com uma grande quantidade de restrições derivada de diferentes fontes, tais como proteínas homólogas e dados experimentais por exemplo (*satisfaction of spatial restrains*). A proteína alvo é então arranjada de forma a satisfazer tais restrições (Greer, 1980; Sali & Blundell, 1993). O segundo grupo (*fragment-based assembly*) visa identificar partes da proteína com estrutura conhecida que são mais similares àquelas da proteína alvo, identificando regiões estruturalmente conservadas (REC) e regiões estruturalmente variáveis (REV). Os fragmentos são então ranqueados de acordo com o nível de similaridade com a sequência alvo, sendo o modelo inferido a partir dos melhores fragmentos (Johnson *et al.*, 1994).

Conforme novos modelos proteicos foram gerados, observou-se que o nível de variação de aminoácidos flutua ao longo da estrutura proteica, observando-se uma maior quantidade de resíduos conservados naquelas posições próximas à sítios ativos das proteínas (Chelliah *et al.*, 2004). Uma vez que a estrutura terciária de uma dada proteína

está finamente adaptada para exercer uma função específica, mutações que alterem significativamente essa funcionalidade tenderão a ser expurgadas pela ação da seleção natural negativa. Worth *et al.* (2009) argumentam que mutações na sequência de DNA podem influenciar a funcionalidade proteica de diferentes formas, modificando diretamente a estrutura terciária ou interferindo nas suas interações quaternárias com outras macromoléculas, substratos, ligantes ou reguladores alostéricos, por exemplo.

Em eucariotos, o Citocromo b (Citb) é um dos 11 componentes do Ubiquinol citocromo c redutase, conhecido também como complexo bc1. Esse complexo protéico transmembranar está envolvido na geração de elétrons e transferência de prótons H⁺ contra o gradiente eletroquímico, que serão utilizados para produção de ATP na membrana interna da mitocôndria (Anugolu *et al.*, 2011). O Citb é a principal subunidade catalítica do complexo bc1, uma vez que possui os sítios nos quais ocorrem a oxidação da molécula de ubiquinol (QH₂) e redução da ubiquinona (Q), gerando assim os elétrons que serão utilizados ao longo da cadeia respiratória (Esposti *et al.*, 1993).

Por ser a única subunidade de bc1 codificada pelo genoma mitocondrial (mtDNA), o gene Citb é rotineiramente utilizado em análises filogenéticas para recuperar tanto as relações ao nível genérico quanto específico (Chen *et al.*, 2009). Isso por que, sendo codificado pelo mtDNA, o citocromo b está em um grande número de cópias em cada célula, facilitando metodologias de amplificação de DNA *in vitro*, como a PCR. Adicionalmente, trata-se de um gene altamente variável e herdado de forma assexuada pela linhagem materna, o que permite uma inferência filogenética de forma mais refinada, além do fato de uma única árvore ser representativa da história evolutiva da molécula como um todo (Birky, 1995; Ballard & Whitlock, 2004; Lynch, 2010).

A utilização do Citb em análises filogenéticas é observada em diversos grupos de vertebrados, como anfíbios anuros (Read *et al.*, 2001; Cunningham, 2002; Faivovich *et al.*, 2005), serpentes (Burbrink, 2005), mamíferos (Redondo *et al.*, 2008; de Carvalho *et al.*, 2011) e aves (Thomas *et al.*, 2004; Batalha-Filho *et al.*, 2014). Entretanto, tem-se demonstrado que o nível de divergência genética do Citb flutua nas diferentes ordens de mamíferos. Há quase duas décadas, Johns & Avise (1998) já haviam observado que espécies irmãs de mamíferos, répteis e peixes apresentam uma taxa de divergência genética maior (até 25%) do que aquelas apresentadas entre espécies por aves (até 16%). Especificamente em mamíferos, Nabholz *et al.* (2008) apontam que espécies de Rodentia apresentam os maiores níveis de divergência interespecífica (2 a 10%), enquanto as de Perissodatyia possuem os menores, com valores inferiores a 1%.

Apesar da sequência do gene *Citb* se mostrar variável, pouco se conhece sobre a variação da estrutura tridimensional da proteína em relação aos eventos de cladogênese nos níveis de gênero e espécie (Esposti *et al.*, 1993; Martin & Palambi, 1993; Anugolu *et al.*, 2011). Na verdade, observa-se que os trabalhos com essa proteína limitam-se à descrição da estrutura proteica em si (Chen *et al.*, 2009) ou na análise das interações da proteína *Citb* com outras macromoléculas (Esser *et al.*, 2006), não sendo realizada uma análise da variação proteica ao nível estrutural em relação a diferenciação entre espécies e gêneros. Assim, o uso do *Citb* para a realização de análises filogenéticas, limita-se a investigação ao nível de DNA, não se levando em conta a localização das mutações na proteína que diferenciam os diferentes táxons.

O *Citb* tem sido frequentemente utilizado para resolver relações evolutivas em grupos complexos, como no caso de *Oryzomys* (Coues, 1890), um gênero de roedor semiaquático da família Cricetidae que possui ampla distribuição geográfica, ocorrendo do sudeste da América do Norte à Patagônia na América do Sul. O gênero possui espécies extremamente semelhantes morfologicamente e um histórico de grande variação do número de espécies a depender do autor que o estuda (Gardner & Patton, 1976; Patton & Hafner, 1983; Dickerman & Yates, 1995; Myers *et al.*, 1995; Patton & da Silva, 1995; Weksler, 2003). Os estudos envolvendo *Citb* nesse grupo mostravam que haviam agrupamentos com alto suporte, que representam táxons com valores de divergência de 13 a 25%, levando o questionamento sobre a existência de mais de um gênero no grupo. Ainda, não havia uma congruência quanto à organização dos grupos sob o enfoque morfológico e citogenético, uma vez que os dados citogenéticos mostravam mais de um cariótipo (e muito distintos) por espécie em alguns casos, inviabilizando a possibilidade de se tratar de variações intraespecíficas e sustentando a hipótese de *Oryzomys* representar mais de um gênero (Patton & da Silva, 1995; Myers *et al.*, 1995; Bonvicino *et al.*, 2001; Andrade & Bonvicino, 2003; Gómez-Laverde *et al.*, 2004).

Em um estudo filogenético marcante entre roedores, Weksler *et al.* (2006) sumarizaram as variações de 99 caracteres morfológicos e craniométricos de todas as espécies de *Oryzomys* descritas até então, recuperando 11 grupos monofiléticos com alto suporte, e propuseram que cada clado corresponderia a um gênero, sendo um deles *Oryzomys* e os demais, dez novos gêneros. Nesse estudo, os novos gêneros foram descritos e nomeados: *Aegialomys*; *Cerradomys*; *Eremoryzomys*; *Euryoryzomys*; *Hylaeamys*; *Mindomys*; *Nephelomys*; *Oreoryzomys*; *Sooretamys* e *Transandinomys*. Ainda, nesse estudo, Weksler e colaboradores observaram que os agrupamentos monofiléticos obtidos

por dados morfológicos corroboravam os grupos obtidos por dados de sequência do Citb em estudos prévios (eg. Myers *et al.*, 1995; Bonvicino *et al.*, 2001), destacando que esse achado foi um dos critérios que os levou à manter a proposta de subdivisão de *Oryzomys*.

Trabalhos posteriores com o gene Citb corroboraram a divisão dos novos gêneros de oryzomíinos, refinaram os relacionamentos filogenéticos e determinaram as taxas de divergência entre gêneros e espécies (Miranda *et al.*, 2007; Bonvicino *et al.*, 2014; Chiquito & Percequillo, 2014; Tavares *et al.*, 2014). Porém, esses estudos limitaram-se à análise de um único gênero, com o objetivo de refinar o posicionamento de suas espécies, não sendo realizada uma revisão dos posicionamentos filogenéticos dos diferentes gêneros de orizomíinos.

Nesse sentido, o presente estudo pretendeu utilizar como grupo modelo os gêneros de roedores descritos por Weksler *et al.* (2006), nos quais as variações do gene Citb sustentaram a subdivisão de um gênero (*Oryzomys*) em 11, sendo 10 novos. Assim, a partir da utilização da abordagem de modelagem por homologia, investigamos se há variações nas estruturas da proteína Citb entre os gêneros e espécies e se essas estão correlacionadas às variações intra e intergêneros. Para tanto, realizou-se um mapeamento das mutações que sustentam os diferentes gêneros e espécies ao longo da estrutura do Citb, levando-se em consideração a localização na cadeia da proteína, assim como a estrutura secundária associada.

Para composição da amostra foram utilizadas sequências do Citb disponíveis no Genbank provenientes de todos os trabalhos que utilizaram representantes dos gêneros descritos por Weksler *et al.* (2006), e sob uma abordagem filogenética, recuperadas as relações dos gêneros e espécies e verificado se as variações na sequência do Citb são espécie e/ou gênero-específicas. Assim, investigamos o monofiletismo dos diferentes gêneros e espécies de orizomíinos, além de realizar uma análise filogenética envolvendo o maior número possível de gêneros a fim de determinar o relacionamento entre os gêneros, uma vez que trabalhos posteriores à Weksler *et al.* (2006) realizaram somente análises intragenéricas. Também, foi feito o mapeamento das mutações que sustentam os diferentes gêneros e espécies na estrutura tridimensional do Citb, afim de verificar se as mutações que levaram aos diferentes eventos de cladogênese em orizomíinos ocorreram em locais de importância funcional para a proteína do Citb e em locais distintos daqueles de gêneros e espécies. Tais análises se mostram relevantes pois a relação filogenética, em um contexto molecular, é determinada exclusivamente pela sequência de DNA. Assim, a questão central do presente estudo é, sendo a variação na sequência de DNA do Citb informativa para

sustentar a separação de *Oryzomys* em dez novos gêneros, haveria alteração equivalente na estrutura terciária da proteína, e se sim, se essas ocorrem em locais específicos e distintos no *Citb* entre os diferentes gêneros e espécies.

RESULTADOS E DISCUSSÃO

Variação na sequência de DNA

Foi gerada uma matriz geral com 75 sequências disponíveis no Genbank para oito dos 11 gêneros de orizomíinos: *Aegialomys*, *Cerradomys*, *Euryoryzomys*, *Hylaeamys*, *Nephelomys*, *Oreoryzomys*, *Sooretamys* e *Transandinomys*, totalizando 24 espécies (Anexo 1). Das 75 sequências, 32 (42%) possuem o gene completo, com 1143 pb. As posições ausentes nas sequências parciais (801 - 1103 pb) foram substituídas por “N” a fim de se realizar as análises filogenéticas.

Foram identificados 58 haplótipos (Anexo 1), exclusivos e espécie-específicos, exceto pelo haplótipo H17, que foi compartilhado entre duas espécies (ver discussão abaixo). Em 16 das 24 espécies (67%) foram observados mais de um haplótipo por espécie.

As mutações ocorreram majoritariamente na terceira posição do códon, seguido por mutações na primeira e segunda posições. Nas análises entre espécies dentro de um mesmo gênero (intra-genéricas), verificou-se que dentre as 214 mutações observadas, 180 (84,1%) ocorreram na terceira posição, 28 (13,1%) na primeira posição e seis (2,8%) na segunda posição. Nas análises intergenéricas, verificou-se que dentre as 934 mutações, 705 (75,6%) ocorreram na terceira posição, 179 (19,2%) na primeira posição e 48 (5,1%) na segunda posição.

O gene Citb mostrou-se com alto poder de agrupamento de espécies e gêneros

Para a realização da análise de Máxima Verossimilhança (MV), o modelo evolutivo que mais se adequou aos conjuntos de dados foi o GTR+I+G, com uma proporção de sítios invariáveis de 0,5340. A divergência genética entre gêneros variou de 13,7 a 23,0%, enquanto que entre espécies variou de 0,06 a 18% (Anexo 2).

O monofiletismo de todos os oitos gêneros analisados foi recuperado, com valores de *bootstrap* acima de 70% (Fig. 1). No entanto, não foi possível recuperar as relações filogenéticas entre os gêneros, sendo observado um polifiletismo basal, exceto pelo clado

formado por *Euryoryzomys* + *Transandinomys* + *Hylaeamys* + *Nephelomys*, identificado como clado B por Weksler *et al.* (2006) e Percequillo *et al.* (2011).

No nível específico, foi recuperado o monofiletismo em 21 espécies. Em dois casos observou-se que a inclusão de duas sequências na análise (uma sequência associada a *Aegialomys xantheolus* - EU579479 e outra a *Hylaeamys yunganus* - DQ224408) tornaram as duas espécies parafiléticas. Se excluídas da análise, estas duas espécies são recuperadas como monofiléticas.

A sequência EU579479, de um exemplar do Peru identificado como *A. xantheolus*, mostrou divergência de 5,1% da sequência EU340015, obtida de um exemplar do Equador também identificado como *A. xantheolus*, enquanto exibe divergência de 1,5% em relação à sequência EU579478, um exemplar de Galápagos, identificado como *A. galapagoensis*. Em 2012, do Prado (2012) realizou uma revisão taxonômica para o gênero *Aegialomys* baseada exclusivamente em dados morfológicos e identificou três grupos, sugerindo haver uma terceira espécie no gênero, com localidade no Peru, podendo, segundo os autores, se tratar de uma nova espécie. Weksler e colaboradores (2006) já haviam apontado que nenhum estudo esclareceu as relações de parentesco das espécies de *Aegialomys*, e que é possível a existência de, no mínimo, uma espécie não descrita para o gênero. Nossas análises de Citb revelaram o agrupamento ((Galápagos+Peru)Equador), no qual a linhagem proveniente do Peru é distinta daquela do Equador, podendo corresponder a um novo táxon, corroborando assim com os dados da literatura. Nesse caso, se explicaria o não monofiletismo da espécie *A. xantheolus* em nosso estudo, podendo, portanto, a sequência EU579479 corresponder a um novo táxon de *Aegialomys*, mais relacionado a *A. galapagoensis* do que a *A. xantheolus*.

A sequência DQ224408 do Peru, associada a *H. yunganus*, mostrou divergência de 13-21% com as demais espécies do gênero, 16,4 - 18,2% de divergência com as demais sequências de *H. yunganus* (OYU58380, AF251520 e EU579500), contrastante com nossas análises que apontam uma divergência intraespecífica de 3-3,2% para *H. yunganus*. Assim, acredita-se que essa sequência do Peru pode estar associada a outra espécie do gênero, distinta de *H. yunganus*. Nos registros do *Genbank*, a sequência DQ224408 foi originalmente depositada como *H. yunganus*. Entretanto, o artigo original (Amman *et al.*, 2006) associa essa sequência a *O. perenensis*, atualmente *H. perenensis*, sendo possível um erro de registro da sequência no *Genbank*. Se confirmado, se explica a grande divergência observada entre essas sequências de *H. yunganus*, uma vez que a sequência DQ224408

corresponderia a *H. perenensis*, apresentando assim uma divergência interespecífica congruente às espécies de *Hylaeamys*.

Outro caso específico foi verificado com a sequência EU579481 (haplótipo H17, Anexo 2), associada a *C. subflavus*, que mostrou-se parafilética às demais sequências da espécie, com até 2% de divergência, e agrupou-se com KP122249, associada a *C. vivoi*, com divergência de 0% (Fig. 1). Isso sugere um novo erro de identificação da sequência EU57948 e justifica o compartilhamento do haplótipo H17 entre *C. subflavus* e *C. vivoi*.

A problemática para erros de identificação de *C. subflavus* já foi levantada na literatura. Sua distribuição abrange os estados de Goiás, São Paulo e Minas Gerais no Brasil, possuindo populações na Bolívia e Paraguai (Percequillo *et al.*, 2008). Diferentes pesquisadores apontaram a possibilidade de *C. subflavus* corresponder à um grupo de espécies crípticas, uma vez que ocorre em biomas tão distintos como Cerrado e Mata Atlântica, além do fato que ser observada uma grande variação quanto ao cariótipo (Bonvicino *et al.*, 2001). No passado, quatro cariótipos distintos foram associados ao grupo *C. subflavus* ($2n=48-50$, $2n=54-56$, $2n=50$, e $2n=58$), cada qual associados a diferentes regiões geográficas (Bonvicino *et al.*, 1999).

Estudos moleculares e morfológicos subsequentes (e.g. Bonvicino *et al.*, 2003; Weksler *et al.*, 2006, Tavares *et al.*, 2011) levaram à separação das espécies nas atualmente reconhecidas: *C. maracajuensis*; *C. marinhui*; *C. scotti*; *C. subflavus*; *C. langguthi*; *C. vivoi* e *C. goytaca*. Assim, nossos dados sugerem que a sequência EU579481, identificada como *C. subflavus*, seja associada a *C. vivoi*, uma vez que essas sequências compartilham o mesmo haplótipo, e formam um clado distinto de todas as demais sequências de *C. subflavus* (*bootstrap*=96).

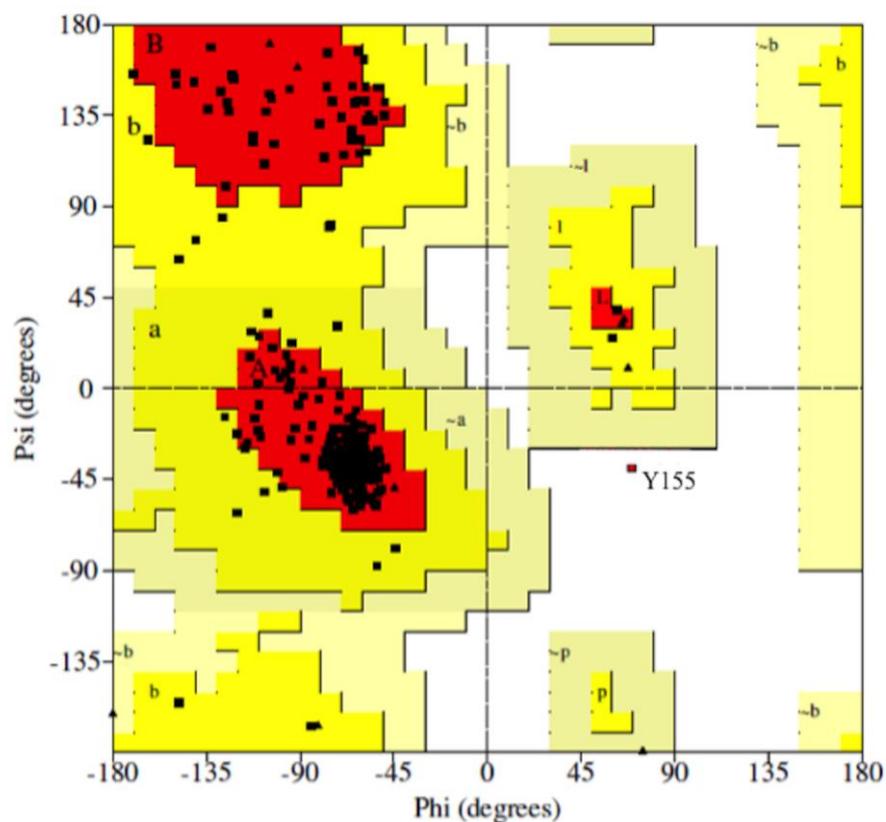
Em suma, nossos resultados demonstram que o Citb não somente é capaz de agrupar táxons, como também auxilia na detecção de erros de identificação, principalmente ao se comparar níveis de diversidade intra e interespecíficos. Adicionalmente, trabalhos anteriores que objetivaram investigar as relações filogenéticas das espécies de orizomíinos, utilizaram o Citb juntamente com genes nucleares, em uma análise concatenada. No presente estudo, demonstra-se que o uso do Citb sozinho é capaz de identificar as diferentes linhagens dessas espécies.

Geração dos modelos tridimensionais da proteína do Citb

Todas as 75 sequências de DNA foram utilizadas para gerar os modelos tridimensionais, utilizando como modelo a estrutura 2A06 de bovino (*Bovine cytochrome bc1 complex*). As sequências alinharam com a porção correspondente do Citb do modelo com similaridade de 79% a 84%. Não foram observados *gaps* entre as sequências de aminoácidos alvo e do modelo. As sequências parciais apresentaram os maiores valores de similaridade com o modelo 2A06. Os 75 modelos gerados não apresentaram erros esteroquímicos.

A verificação da confiabilidade dos modelos a partir da análise das torções dos ângulos de Ramachandran apontaram que 93% dos resíduos estão localizados nas regiões “mais favoráveis” e 6,5% estão nas regiões “permitidas” (Fig. 2). Somente um resíduo (0,3%) de Tirosina (Y) na posição 155 (Y155) está localizado em região “não permitida”, localizada no *loop* curto entre as cadeias cd1 e cd2, na superfície da estrutura do Citb.

A Tirosina possui uma cadeia lateral aromática composta por seis carbonos, sendo assim considerada uma cadeia volumosa. Uma possível explicação para a sua localização em região não permitida nos gráficos de Ramachandran é a de que a Y155 assume uma torção que evita choques com os aminoácidos vizinhos, uma vez que na posição 154 observamos uma Prolina (P154), também de cadeia aromática (Fig. 3). Assumimos que ao evitar o choque com a P154, que se encontra mais internamente (“enterrada”), a Y155 assumiu uma torção que a plotou em região “não-permitida” nos gráficos de Ramachandran. Entretanto, a ocorrência da torção Y155 evitou perda de estabilidade da cadeia proteica do Citb.



Plot statistics		
Residues in most favoured regions [A,B,L]	299	93.1%
Residues in additional allowed regions [a,b,l,p]	21	6.5%
Residues in generously allowed regions [-a,-b,-l,-p]	0	0.0%
Residues in disallowed regions	1	0.3%

Number of non-glycine and non-proline residues	321	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	20	

Total number of residues	369	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig. 2. Gráfico de Ramachandran para um dos modelos do Citb gerados, mostrando sua qualidade em termos estereoquímicos. As regiões mais favoráveis estão em vermelho, e as regiões permitidas e não permitidas estão mostradas em amarelo e branco, respectivamente. Os pontos escuros representam cada aminoácido na cadeia do Citb. Destaca-se que somente o aminoácido Tirosina na posição 155 (Y155), como um ponto vermelho, está localizado em região não permitida.

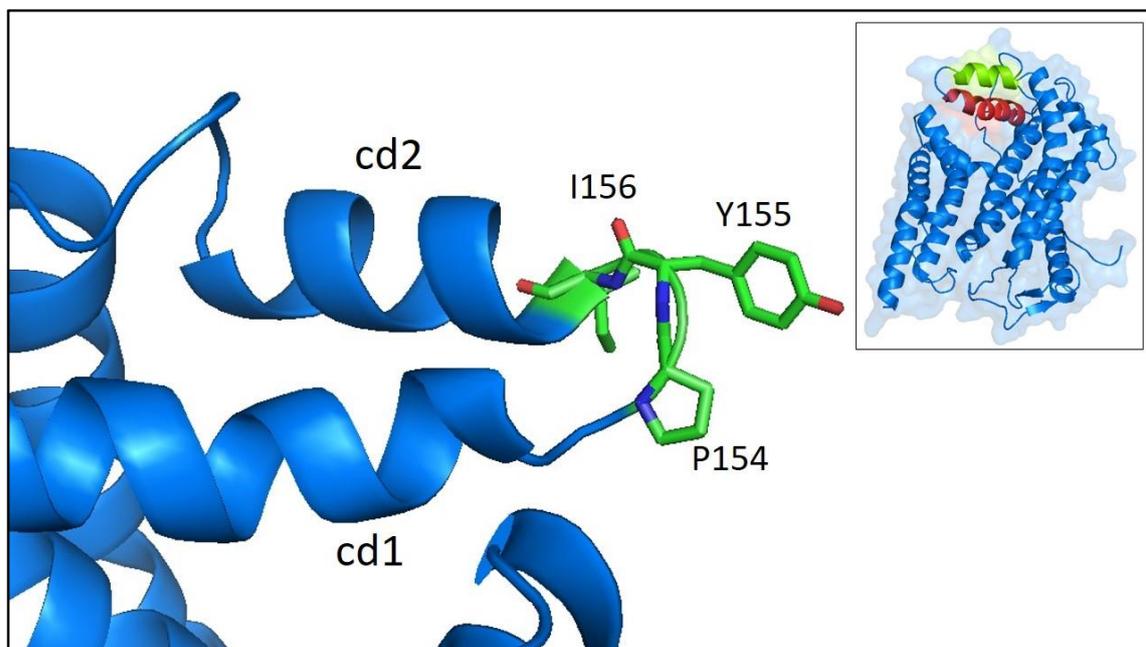


Fig. 3. Detalhe da estrutura tridimensional do Citb evidenciando a Tirosina na posição 155 (Y155), em região não permitida segundo das análises dos gráficos de Ramachandran, posicionando a cadeia lateral externa à cadeia principal da proteína. A figura exibe também os aminoácidos vizinhos à Y155. O quadro mostra a estrutura completa do Citb e destaca as regiões cd1 (vermelha) e cd2 (verde).

Pal & Chakrabarti (2002) demonstraram que aminoácidos localizados na superfície da proteína, e aminoácidos de *loops* curtos, tendem a ser registrados em regiões “não permitidas” do gráfico de Ramachandran. Os autores argumentam que isso ocorre devido ao fato de que aminoácidos na superfície apresentam uma menor quantidade de ligações e interações com aminoácidos vizinhos, permitindo assim uma maior plasticidade de suas conformações. Ainda, reforçando nossos achados, Guo *et al.* (2004), por sua vez, argumentaram que mutações em resíduos localizados em regiões internas da proteína, requerem que ocorram mutações concomitantes em aminoácidos próximos para que se mantenha um ótimo dobramento (*folding*) na estrutura proteica. No caso de orizomino, a mutação Y155, é superficial e única, corroborando as proposições anteriores.

Análise da estrutura da proteína do Citb

A estrutura geral da proteína dos 75 modelos gerados é similar àquelas encontradas na literatura para mamíferos, sendo composta por oito α -Hélices maiores (A a H), três α -Hélices menores (cd1, cd2 e ef) e um par de Folhas- β com direções opostas (Fig. 4), semelhante à estrutura observada para o molde 2A06 (Huang *et al.*, 2005).

As hélices são organizadas aos pares, conectadas por sete *loops*, sendo *loops* longos com até 14 aminoácidos entre os pares de hélices AB, CD, DE e EF e *loops* curtos de até seis aminoácidos nos pares de hélices BC; FG; GH. Todas as estruturas apresentaram 24 aminoácidos que estão envolvidos na ligação do grupo Heme (*Protoporphyrin IX Containing Fe*).

A sobreposição de todos os modelos de Citb gerados mostra que as áreas de α -Hélice localizam-se de modo agrupado, centralmente na estrutura da proteína, enquanto que as extremidades C-Terminal e N-terminal e porções de Folha- β e *loops* agrupam-se superficialmente à porção central de α -Hélice (Fig. 5A). Esta configuração geral está de acordo com alguns estudos (eg. da Fonseca *et al.*, 2008; Chen, 2009) que demonstram que o Citb distribui-se de forma transmembranar na mitocôndria. Assim, a porção da proteína em orizomino que estaria embebida pela membrana interna da mitocôndria seria constituída principalmente de regiões de α -Hélices, compondo a maior parte da proteína, abrangendo 71,81% dos aminoácidos.

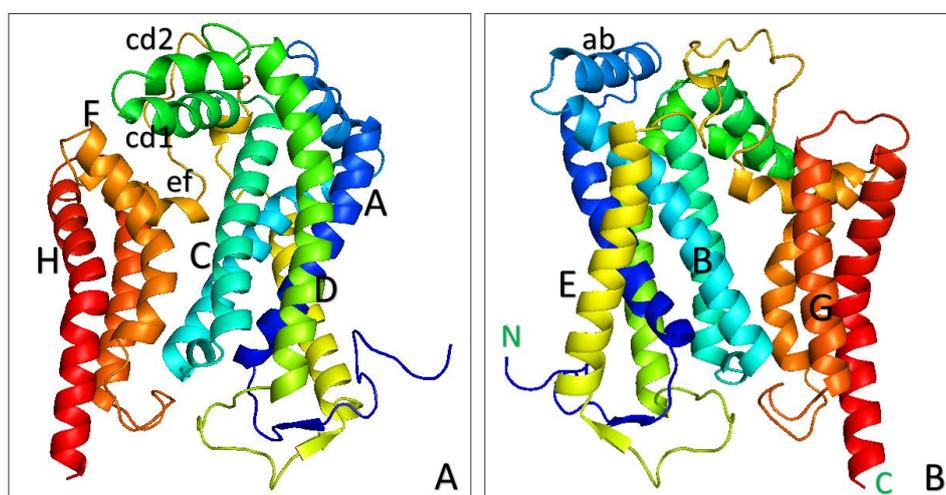


Fig. 4. (A, B) Representação geral no formato *cartoon* da estrutura do Citb para as espécies de oryzominos do presente estudo. As letras em negrito (A-H; cd1; cd2 e ef) identificam as diferentes cadeias de α -Hélices recuperadas a partir da modelagem por homologia. Em B, vista em 180° da mesma estrutura para melhor identificação das cadeias. As letras em verde identificam as extremidades C-Terminal e N-Terminal.

A análise par-a-par das estruturas tridimensionais, tanto entre espécies do mesmo gênero quanto espécies de gêneros distintos, mostraram que há alta semelhança estrutural entre os modelos, com valor de RMSD máximo de 0,22, não revelando um padrão espécie-

ou gênero-específico, ao contrário da variação observada na sequência de DNA, que se mostrou distinta entre gêneros e entre espécies de um mesmo gênero.

Esse resultado não é surpreendente, já que as sequências de aminoácidos são altamente conservadas. Sequências nas quais o comprimento e a porcentagem de aminoácidos semelhantes estão localizadas nos limites da *safe zone* tendem a assumir estruturas tridimensionais semelhantes. Em nosso conjunto de dados, as sequências de aminoácidos do Citb apresentaram um alto grau de conservação, além de não ter sido observada a presença de *gaps* no alinhamento. Dessa forma, a alta similaridade entre as estruturas tridimensionais do Citb é uma consequência do fato de suas sequências de aminoácidos estarem localizadas no limite da *safe zone*.

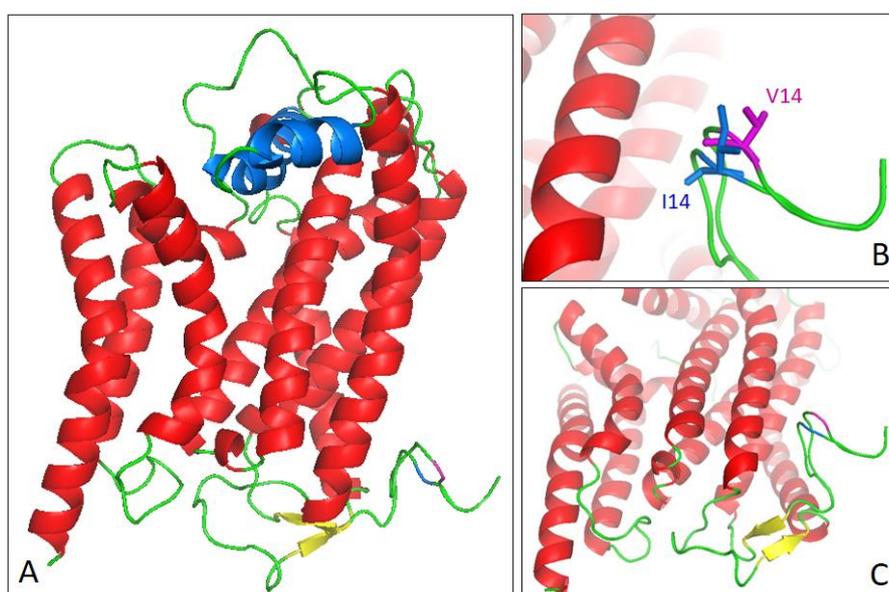


Fig. 5. (A) Sobreposição de todos os modelos de Citb gerados no presente trabalho. As áreas de α -hélice (vermelho) correspondem às porções majoritariamente embebidas na membrana interna da mitocôndria (exceto cd1 e cd2, destacadas em azul), as áreas em amarelo correspondem às estruturas em folha- β e em verde aos *Loops*. (B) e (C) Detalhes da região N-terminal do Citb, mostrando a variação entre as estruturas na posição 14, com variação entre valina (V14, rosa) e isoleucina (I14, azul).

Varição estrutural da proteína na posição 14

A partir da sobreposição de *backbone* de todos os modelos tridimensionais gerados para o Citb pela modelagem por homologia, observou-se somente uma variação estrutural da proteína, localizada na posição 14, próximo à região do N-terminal, com 28% das sequências contendo Valina (V14) e 72% Isoleucina (I14), ambos aminoácidos alifáticos e

de propriedades semelhantes, com distância de 2,1Å naquela posição quando os diferentes modelos são sobrepostos (Fig. 5B e 5C).

As trincas do DNA em orizominos que codificam a I14 foram ATT e ATC e a para V14 foi GTA. Considerando-se que o grupo externo, representado *Thomasomys* e *Rhipidomys*, apresentou exclusivamente a trinca ATT, é possível sugerir a Isoleucina na posição 14 é um caráter plesiomórfico e a mutação V14 seria a forma mutante. Assim, para gerar a forma mutante em orizominos seria necessária a ocorrência de duas mutações, uma transição de Adenina para Guanina na primeira posição e uma transversão de Timina para Adenina na terceira posição. Ainda, uma transição de Timina para Citosina na terceira posição gerou uma mutação sinônima de Isoleucina de ATT para ATC em orizominos.

Dos oito gêneros de orizominos, seis apresentam somente I14, enquanto *Hylaeamys* (13 sequências de três espécies) apresenta somente V14 e em *Euryoryzomys* as duas formas, V14 e I14 são observadas (Fig. 1). Das 17 sequências de cinco espécies de *Euryoryzomys*, três espécies (*E. macconnelli*, *E. russatus* e *E. lamia*) apresentam somente V14; e duas espécies (*E. macconnelli* e *E. emmonsae*) apresentam somente I14. *E. macconnelli* é a única espécie analisada que apresenta ambas as formas, sendo quatro sequências com I14 (Peru e Amazonas/Brasil) e uma com V14 (Amazonas/Brasil).

O mapeamento das mutações V14 na árvore filogenética de orizominos (Fig.1) sugere que a mutação V14 tenha evoluído mais de uma vez e independente nos gêneros *Euryoryzomys* e *Hylaeamys*.

Apesar do presente estudo não ter recuperado as relações entre os diferentes gêneros de orizominos, trabalhos anteriores utilizando o Citb de forma concatenada à genes nucleares (IRBP) foram capazes de demonstrar a relação entre esses gêneros. Percequillo *et al.* (2011) e Weksler (2006) demonstraram que *Euryoryzomys* e *Hylaeamys* constituem o clado B, juntamente com espécies de *Nephelomys* e *Transandinomys*. Entretanto, ao analisarmos a relação proposta por aqueles autores observamos a seguinte configuração: (((*Euryoryzomys*+*Transandinomys*)*Hylaeamys*)*Nephelomys*). Tal disposição filogenética aponta a origem independente da mutação V14 em *Euryoryzomys* e *Hylaeamys*, uma vez que nem *Transandinomys* nem *Nephelomys* a possui.

Vale destacar que a mutação V14 parece apontar um cenário complexo em relação à *E. macconnelli*. Das cinco sequências analisadas no presente estudo, três são provenientes do Amazonia brasileira (AF251528, AF251527e OMU58379) e duas são provenientes de Loreto no Peru (EU579484 e GU126538). Dessas cinco, somente a AF251528, da

Amazonia, possui a mutação V14, fazendo com que essa sequência esteja localizada basalmente a todas as demais da espécie em nossas análises (*bootstrap*=100).

Uma das questões intrigantes no processo de formação de uma proteína é como a variação de códons influencia na síntese proteica. Observa-se no código genético que um dado aminoácido pode ser codificado por até seis códons distintos, chamados códons sinônimos (Fredrick & Ibba, 2010). Sabe-se que a frequência de utilização de diferentes códons varia em diferentes organismos, um fenômeno conhecido como uso de códon (*codon usage*). Uma das primeiras questões em relação ao uso de códon, é como sua variação influencia na taxa de transcrição do RNA mensageiro (RNAm).

Para correlacionar a taxa de transcrição do RNAm e o uso de códon, Tuller *et al.* (2010) investigaram a eficiência na transcrição de diferentes códons e a disponibilidade de RNA transportador (RNAt) correspondentes no RNAm. Esses autores demonstraram que códons comuns tendem a ser transcritos de forma mais rápida e eficiente, uma vez que há uma maior disponibilidade de RNAt durante esse evento. Entretanto, diferentes trabalhos apontam que a transcrição é influenciada por outros fatores celulares além da disponibilidade de RNAt. Elf *et al.* (2003) demonstraram que a velocidade de transcrição nos primeiros 30-50 códons (região conhecida como *ramp*) tende a ser menor se comparada ao restante do gene, indiferente dos códons presentes nesta região inicial. Esses autores argumentam que uma velocidade menor de transcrição no início do gene permite a criação de um espaçamento entre os ribossomos mais avançados na sequência de RNAm, evitando assim um congestionamento dos ribossomos e aumentando a eficiência da biossíntese proteica.

Nossos dados de modelagem por homologia para o Citb apontam a existência de duas formas proteicas para os orizomios, com variação na posição 14, com uma contendo uma Valina (V14) e outra uma Isoleucina (I14). Observa-se que esses aminoácidos diferem entre si quanto à quantidade de códons que os codificam, sendo a Valina codificada por um códon (GTA) enquanto a Isoleucina é por dois (ATT e ATC). Se por um lado o conceito do uso do códon favoreceria no sentido de uma maior eficiência biossintética da Isoleucina, há de se considerar que a velocidade de síntese na região inicial da proteína (região de *ramp* do Citb) seria baixa, independente do aminoácido existente. Nesse sentido, a mutação nessa posição, embora sendo mutação não sinômima, não necessariamente conferiria uma vantagem evolutiva. Dessa forma, apesar das mutações I14 e V14 gerarem uma variação ao nível de estrutura tridimensional do Citb, é provável que haja pouca alteração na funcionalidade e biossíntese entre essas duas formas.

Avaliando a possibilidade de modelagem do Citb a partir da metodologia de fragment-based assembly

Com o objetivo de recuperar os melhores modelos para a realização da metodologia de *fragment-based assembly*, foi realizada uma busca no servidor online FUGUE (Shi *et al.*, 2011), utilizando todas as 75 sequências de orizominos. Para todas as sequências de aminoácidos, foram identificados como os melhores modelos as sequências 1bcc, associado à *Gallus gallus*, e 1bgyc, associado à *Bos taurus*, sendo observado um *Z-score* variando de 55-81.

Após a recuperação dos melhores modelos, realizamos então a identificação das regiões estruturalmente conservadas (REC) e regiões estruturalmente variadas (REV) a partir da implementação do CHORAL (Montalvao *et al.*, 2005). Para tanto, optou-se pela realização de uma modelagem inicial para a sequência EU579478, associada à *A. galapagoensis*. Como a sequência alvo do Citb está completa, foi possível verificar o nível de variação entre a sequência alvo e os moldes selecionados, assim como entre os dois moldes entre si. A estrutura 1bcc apresentou uma curvatura máxima de 1,83 e uma torção máxima de 1,64, enquanto a 1bgyc apresentou uma curvatura máxima de 2,15 e uma torção máxima de 1,69. Os moldes 1bgyc e 1bcc possuíram um PID com a sequência EU579478 de 78,2 e 69, com um *score* de 2319 e 2083 respectivamente.

Ao total, foram gerados cinco modelos para a sequência de *A. galapagoensis*, com uma porcentagem de resíduos modelados variando de 99,5 a 97,1%. Por fim, foram observadas duas REV, uma localizada no *loop* curto entre as cadeias B e C e uma na região do N-terminal, sendo o Citb composto majoritariamente por REC em nossas análises (Fig. 6). Assim, é possível observar que os moldes de *Gallus* e *Bos* são altamente semelhantes entre si, além de apresentarem uma alta similaridade com a sequência alvo de *Aegilomys*. O alinhamento gerado para um dos modelos pode ser observado no anexo 5.

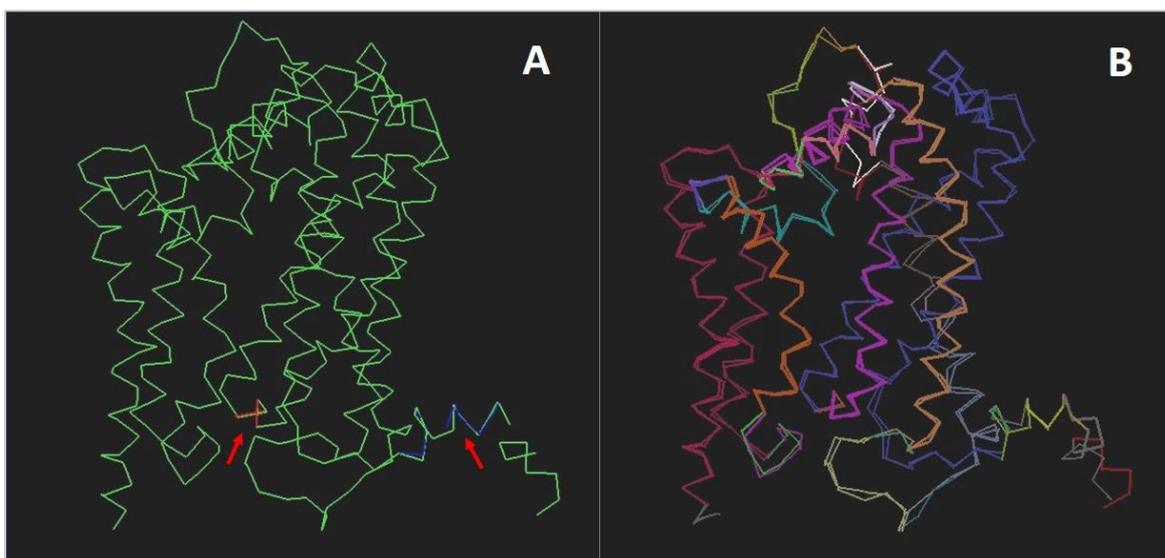


Fig. 6: (A) Sobreposição dos *backbones* para os cinco modelos gerados a partir do CHORAL para a sequência EU579478 de *Aegilomys*. As setas vermelhas evidenciam os pontos de variação na estrutura do Citb. (B) Sobreposição dos moldes de *Gallus* (1bcc) e de *Bos* (1bgyc) evidenciando REC e REV para o Citb entre essas estruturas.

Como mencionado, o principal aspecto da metodologia de *fragment-based assembly* consiste na identificação de fragmentos considerados similares entre os moldes selecionados e a sequência alvo, a partir da identificação das REC e REV (Greer, 1980). As REC são então divididas em fragmentos que são classificados de acordo com o nível de similaridade (curvatura e torção) com a sequência alvo, sendo que os modelos são então inferidos a partir dos melhores fragmentos (Johnson *et al.*, 1994).

Uma das principais vantagens da metodologia *fragment-based assembly* quando comparada com aquelas baseadas em *spatial constraints* é o potencial para lidar com famílias proteicas que apresentam uma baixa porcentagem de similaridade (o que não é o caso de orizomins). Isso por que em tais casos, a grande quantidade de *indels* e variação estrutural gera um sistema complexo, dificultando a modelagem comparativa.

A definição de REC como aplicada pelo CHORAL foi elaborada por Montalvao *et al.* (2005), que define uma REC como uma forma geométrica (*shape*) que pode ser única para qualquer membro do alinhamento estrutural ou comum para qualquer combinação de estruturas. Assim, no CHORAL, resíduos alinhados de estruturas sobrepostas são considerados agregados, formando uma REC, quando todas as sepações de C α -C α da região são menores do que 3.6 Å. Assim, uma região de duas ou mais proteínas, com pelos menos três resíduos de comprimento, é considerada conservada quando todos os seus resíduos são agrupados e as diferenças para todas as suas características geométricas

(curvatura e torção) estão abaixo de um limite específico. Como curvatura e torção caracterizam uma curva regular em um espaço tridimensional euclidiano, regiões com valores similares nessas características tenderão a ser similares.

Nossos dados de modelagem gerados a partir do CHORAL apontam que a sequência EU579478 possui uma alta similaridade estrutural com os moldes selecionados, uma vez que tanto os valores de curvatura máxima quanto torção estão abaixo dos limites mínimos estabelecidos para metodologia de *fragment-based assembly*. Tal resultado era esperado, uma vez que a sequência alvo está localizada na *safe zone* como postulado por Rost (1999). Assim, como os moldes 1bcc e 1bgyc são estruturalmente similares entre si, acrescido ao fato das sequências de orizominos possuírem um alto valor de similaridade, é possível afirmar que as sequências recuperadas para o presente estudo possuem estruturas similares, havendo uma mínima variação. Dessa forma, a partir da análise de *fragment-based assembly*, nota-se que mesmo que a variação no DNA observada no Citb tenha sido suficiente para separar os diferentes gêneros e espécies de orizominos, essa variação não se reflete em uma alteração na estrutura tridimensional da proteína, uma vez que as estruturas observadas se mostraram virtualmente idênticas.

Mapeamento das mutações na estrutura da proteína Citb

Mutações sinônimas e não-sinônimas

A análise comparativa da estrutura proteína entre os grupos recuperados pela análise filogenética do Citb permitiu identificar as mutações que diferenciam as espécies dentro de um gênero (abordagem intragenérica) e os gêneros entre si (abordagem intergenérica).

Na análise intragenérica, identificou-se 214 mutações (Anexo 2), sendo 88% correspondente à mutações sinônimas (MS) e 12% a mutações não sinônimas (MNS). De todas as mutações, ~75% mutações ocorreram em regiões de α -Hélices, ~21% em regiões de *loop* e somente 1,4% ocorreram nas Folha- β (Tabela 1). MNS não foram observadas nas regiões de Folha β . Não houve correlação entre o tipo de mutação (MS e MNS) e a estrutura secundária da posição mutada (χ^2 ; p-valor=0,2613).

Na análise par-a-par entre os gêneros, registrou-se 934 mutações (Anexo 3), das quais 81% corresponderam a MS e 19% MNS. Observou-se que ~71% estão em α -Hélices, ~26% em regiões de *loop*, e somente ~1% em região de Folha β (Tabela 1, Fig.

7). O teste de significância mostrou que há correlação entre o tipo de mutação e a estrutura secundária associada para mutações intergenéricas (X^2 ; p-valor= 0,04784).

Dessa forma, observa-se que tanto para a análise intragenérica quanto para a intergenérica, as mutações ocorreram com maior frequência na região de α -Hélices, seguidas por mutações em região de *loop* e Folha- β , nessa ordem. De forma semelhante, a distribuição dos tipos de mutações (MS e MNS) segue a mesma ordem. Porém, somente há significância para a correlação entre a estrutura secundária e o tipo de mutação nas mutações intergêneros.

Tabela 1: Frequência (em porcentagem) das mutações sinômimas (MS) e não sinônimas (MNS) e respectivas estruturas secundárias ao longo dos 380 resíduos na proteína do citocromo b.

Análise	Total de mutações	Alfa Hélice		Loop		Folha Beta		ND ¹
		MS	MNS	MS	MNS	MS	MNS	
Intergenérica	934	58,45	12,42	19,59	5,99	0,96	-	2,35
Intragenérica	214	64,02	10,75	20,03	1,40	1,40	-	3,40

¹ND=estrutura não determinada para a respectiva mutação.

A partir da análise da Figura 7, observamos que tanto MS quanto MNS, de forma geral, estão distribuídas ao longo de toda a estrutura do Citb, tanto na abordagem intragenérica quanto intergenérica. Apesar disso, observa-se que as MNS que distinguem gêneros estão localizadas em posições distintas daquelas que distinguem espécies. Para a abordagem intragenérica registramos 13 posições onde observou-se exclusivamente MNS, sendo elas distribuídas nas diferentes espécies da seguinte forma: *Aegialomys galapagoensis* (posições 142 e 266), *Cerradomys goytaca* (posição 164), *Cerradomys maracajuensis* (posição 117), *Cerradomys marinhuis* (posição 236), *Hylaeamys acritus* (posições 66 e 375), *Hylaeamys laticeps* (posição 118), *Nephelomys levipes* (posições 60 e 88) e *Transandinomys boliviensis* (posições 42, 115, 315 e 349).

Por outro lado, para a abordagem intergenérica, foram registradas 13 posições onde ocorreram exclusivamente MNS, estando elas distribuídas da seguinte forma: *Aegialomys* (posições 43, 107, 204 e 370), *Hylaeamys* (posições 14 e 324), *Cerradomys* (posições 107 e 234), *Euryoryzomys* (posições 107 e 324), *Nephelomys* (posição 323), *Oreoryzomys* (posições 29 e 241), *Sooretamys* (posições 107, 323; 345 e 371) e *Transandinomys* (posição 214).

As MNS que separam as diferentes espécies em um mesmo gênero ocorrem nas posições 14, 42, 60, 66, 88, 115, 117, 118, 142, 164, 236, 266, 315 e 349. Por outro lado, as MNS que separam os diferentes gêneros são distintas (exceto pela posição 14) e ocorreram nas posições 14, 29, 43, 107, 214, 234, 241, 324, 323, 345, 370 e 371. Dessa forma, é possível notar que há uma distinção entre a localização das MNS entre as análises intergenéricas e intragenéricas.

A partir da análise da Figura 7, observa-se que para a abordagem intergenérica, há uma faixa de MNS que estende-se nas primeiras 80 posições, enquanto na intragenérica, essa faixa estende-se entre as posições 110 e 126. Dessa forma, é possível notar faixas de MNS distintas em nossas análises intergenéricas e intragenéricas.

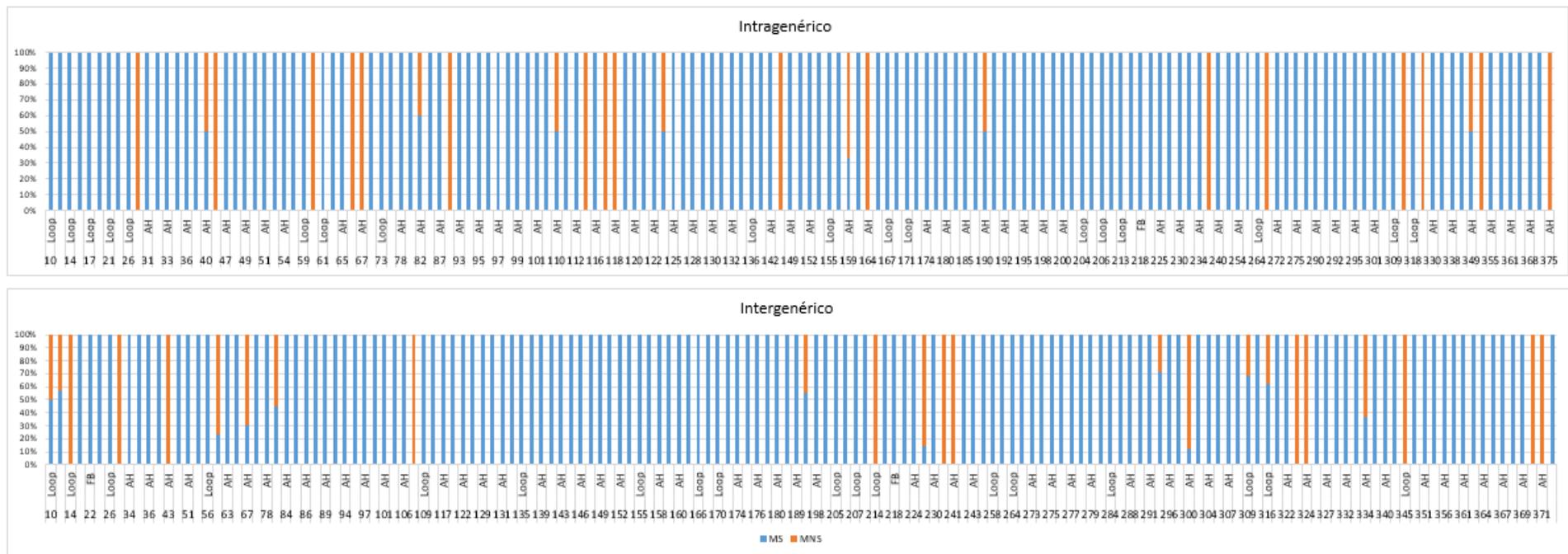


Fig.7: Proporções entre mutações sinônimas (MS) e não sinônimas (MNS) em relação as estruturas secundárias do Citb tanto para a abordagem intragenérica quanto intergenérica. A imagem exibe somente posições onde registrou-se eventos de mutação, não sendo mostrado posições conservadas. AH= α -Hélice; FB= Folha β .

Porções hipervariáveis e sua importância

A análise da entropia de Shannon é importante para diferentes finalidades, sendo utilizada no presente estudo com o objetivo de identificar as posições mais variáveis no conjunto de dados de orizominos. A análise revelou cinco posições (43, 82, 108, 234 e 323) com valores de entropia acima de 1.0 (Fig. 8), sendo observado no máximo cinco aminoácidos distintos para essas posições, com os aminoácidos observados apresentando propriedades de carga e polaridade semelhantes, com exceção da Treonina (T).

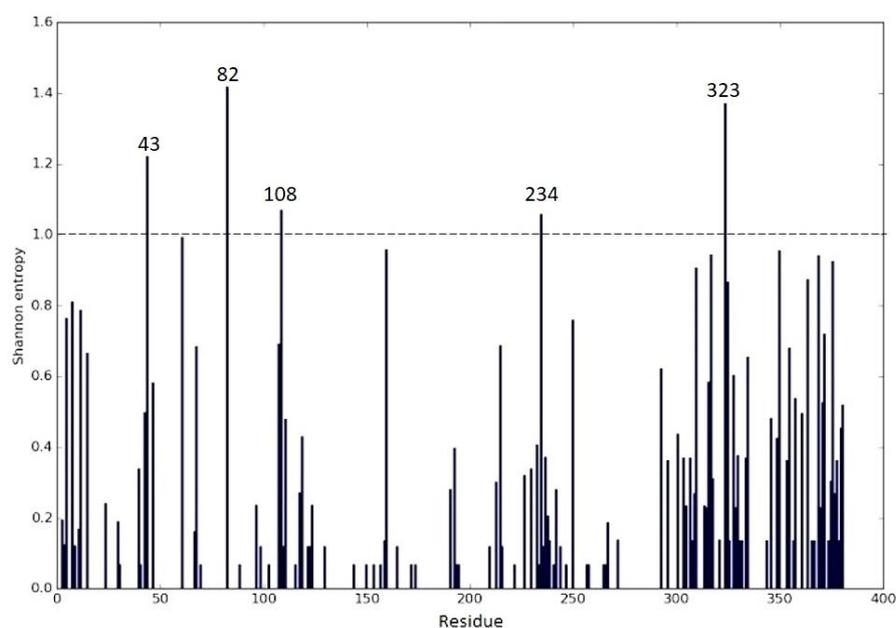


Fig. 8: Gráfico da entropia de Shannon para a matriz composta por 77 sequências. A linha pontilhada identifica aquelas posições nas quais foram observados valores de entropia acima de 1,0. Posições conservadas são identificadas pela ausência de colunas.

Todas essas posições localizam-se em regiões do Citb distribuídas pela membrana interna mitocondrial, estando todas elas em regiões de α -Hélice, com exceção da posição 108, que localiza-se em uma região de *loop* (Tabela 2). Adicionalmente, as cinco posições estão localizadas na superfície do Citb, com as cadeias laterais dos aminoácidos posicionando-se externamente ao Citb, não estando próximo a nenhum sítio ativo ou alostérico da proteína.

Tabela 2: Posições nas quais foram observados valores da entropia de Shannon acima de 1,0. A quantidade de sequências por aminoácido é mostrada entre parênteses, assim como a respectiva cadeia para a posição.

Posição	Aminoácidos	Cadeia
43	V(36); M(26); I(8); T(2); L(4)	A
82	T(4); M(27); L(24); I(15); A(7)	B
108	M(37); V(29); T(10); L(1)	<i>Loop</i>
234	L(36); I(30); F(11)	E
323	F(5); L(1); T(10); I(13); M(1)	G

É sabido que as diferentes regiões de uma dada proteína podem apresentar diferentes níveis de variação, sendo a análise da entropia de Shannon (Cover & Thomas, 1991) uma importante ferramenta para identificação de regiões variáveis ou conservadas em sua cadeia. A entropia de Shannon pressupõe que se considerarmos uma série estacionária $S(i)$, $i=1, 2, \dots, N$ (onde N corresponde ao comprimento da série), sendo observado para $S(i)$ diferentes valores ou estados (M), a entropia de Shannon calcula então a incerteza da série (Mazaheri *et al.*, 2010). A tolerância da posição i para mudança de aminoácidos é dada por $S(i) = -\sum_{a_i=1}^{20} P(a_i) \log P(a_i)$, onde $P(a_i)$ é a probabilidade da ocorrência do aminoácido do tipo a na i th posição da cadeia proteica (Liu & Bahar, 2012). $S(i)$ possui uma variação de 0-3, com o valor mínimo e máximo correspondendo a uma posição conservada e completamente randômica respectivamente (chances iguais da ocorrência de quaisquer dos 20 aminoácidos na i th posição), dessa forma, uma posição supervariável.

Nossas análises sobre a entropia de Shannon apontam que as posições mais variáveis do Citb estão localizadas em regiões distantes dos principais sítios, sugerindo assim que mutações nessas posições geram pouco impacto para a funcionalidade da proteína como um todo. Tal resultado contradiz os achados de da Fonseca *et al.* (2008) para as diferentes ordens de mamíferos, uma vez que esses autores sugerem que posições hipervariáveis do Citb possuem uma relevância adaptativa elevada em muitas espécies.

Entretanto, apesar das posições identificadas pela entropia de Shannon para as espécies de orizomíinos do presente estudo estarem distantes dos principais sítios ativos do Citb, não se pode descartar a hipótese de essas mutações não serem neutras, principalmente em relação à estabilidade e mobilidade da estrutura, uma vez que aminoácidos dessas posições estão em contato com a membrana interna mitocondrial.

Um importante conceito no estudo da dinâmica estrutural de proteína é a relação entre estabilidade e mobilidade. Em alguns casos, a dinâmica é manifestada por um grande movimento de estruturas intactas, conhecido como “movimentos globais”. Muitas enzimas ou proteínas membranares envolvidas em transporte apresentam esse tipo de movimento (Tama & Brooks, 2006; Yang *et al.*, 2009). Por outro lado, “movimentos locais” tendem a ser mais restritivos, como rearranjos de *loops*. Wright & Dyson (2009) argumentam que “movimentos locais” facilitam o reconhecimento de diferentes substratos e otimizam ligações. Liu & Bahar (2012) demonstraram que o aumento da variabilidade da sequência, ou seja, maiores valores de entropia, é acompanhada pelo aumento da mobilidade proteica.

Nossos dados identificam que as posições mais variáveis localizam-se em regiões embebidas na membrana interna mitocondrial, dessa forma, não participando de ligações com nenhum substrato. Entretanto, essas posições hipervariáveis poderiam afetar a interação entre o Citb e a membrana mitocondrial, principalmente em relação à mobilidade e estabilidade da proteína. Assim, se faz necessário um estudo futuro de dinâmica molecular com o objetivo de investigar a influência dessas posições em relação a interação do Citb e a membrana mitocondrial.

Correlação entre posição do códon e tipo de mutação

Nas análises intragenéricas, observou-se que em todas as mutações na segunda posição (n=6) tratava-se de MNS, enquanto que mutações nas demais posições, ocorreram MS e MNS, sendo observada uma maior quantidade de MNS na primeira posição do que na terceira (Tabela 2). A partir da análise de χ^2 observou-se uma correlação significativa entre o tipo de mutação e a posição no códon (p-valor<0,005)

Nas análises intergenéricas, de forma semelhante à abordagem intragenérica, observou-se que todas as mutações na segunda posição (n=48) corresponderam a MNS, enquanto para as demais posições registrou-se tanto MS quanto MNS, sendo observada uma maior quantidade de MNS na primeira do que na terceira posição (Tabela 2). Da mesma forma que em nossas análises intragenéricas, na intergenérica encontramos valores significativos de χ^2 entre posição do códon e tipo de mutação, com *p-valor* muito menor do que 0,05.

Tabela 3: Frequência (em porcentagem) das mutações sinômimas (MS) e não-sinônimas (MNS) e posição do códon.

Análise	Total de mutações	1a		2a		3a	
		MS	MNS	MS	MNS	MS	MNS
Intergenérica	934	6,96	12,2	-	5,1	73,2	2,0
Intragenérica	214	6,54	6,54	-	2,8	81,3	2,8

Assim, verifica-se que a maior proporção de MS ocorrem na 3ª posição do códon, enquanto que alterações na 1ª e 2ª posições geraram prioritariamente MNS tanto para as análises inter quanto para intragenéricas.

Sabe-se que as quatro diferentes bases nitrogenadas podem ocorrer em proporções distintas conforme a posição no códon em que elas ocupam. Majumdar *et al.* (1999) trabalhando com genomas procarióticos demonstraram que há uma maior frequência de purinas, i.e, A e G, na primeira posição, sendo observado para a segunda posição uma maior frequência de A, enquanto para a terceira posição não há uma predominância clara de bases. Adicionalmente, aqueles autores apontam que quando observado o desvio padrão em relação a predominância das bases nas diferentes posições do códon, é possível observar que a terceira posição possui os maiores níveis de variação enquanto a segunda posição possui os menores. Assim, as primeiras duas posições do códon estariam sob forte seleção purificadora, enquanto a maior parte da variação na terceira posição seria sinônima e estaria sob fraca seleção purificadora. Em suma, Majumbar *et al.* (1999) apontam que as primeiras duas posições do códon são importantes para determinar a estrutura proteica enquanto a terceira posição é importante para a diferenciação de espécies.

Nossos dados para as espécies de orizomíinos corroboram com os achados de Majumbar *et al.* (1999) ao demonstrar que a maior parte das mutações na terceira posição são representadas por MS. Além disso, é possível afirmar a seleção purificadora atua de forma mais ríspida nas duas primeiras posições, uma vez que observamos uma menor quantidade de mutações nas duas primeiras posições, tanto para MS quanto para MNS, do que para a terceira posição.

Variação nas cadeias cd1 e cd2

Duas regiões de grande importância funcional para o Citb compreendem as regiões α -Hélices cd1 e cd2, que possuem diferença de quantidade de aminoácidos (aa) em suas cadeias, tendo 16 aa em cd1 e 10 aa em cd2.

Na abordagem intergenérica, para cd1, foram observadas nove das 16 posições mutadas (139, 142, 143, 144, 146, 147, 149, 150 e 152), com MNS somente em duas delas (posições 143 e 149). Para a posição 143, observa-se a presença de uma Guanina (G) exclusiva de *Aegialomys* (*A. galapagoensis*), com Alanina (A) nos demais gêneros. Em relação a posição 149, verificou-se uma Prolina (P) em somente uma das três sequências na única espécie de *Oreoryzomys* (*O. balneator*), enquanto todas as demais sequências apresentando uma Leucina (L) para aquela posição.

Para cd2, foram registrados seis posições mutadas (157, 158, 159, 160, 161 e 164), sendo registradas MNS somente nas posições 159 e 164. Para a posição 159, *Cerradomys* e *Hylaeamys* são os únicos gêneros que apresentam S (Serina), sendo que *Hylaeamys* possui adicionalmente uma Treonina (T). Todos os demais seis gêneros, assim como o grupo externo, apresentam T159. Na posição 164, *Cerradomys* é o único gênero que possui uma Valina (V164) em somente duas sequências de *C. goytaca*. Para todos os demais sete gêneros de orizomíinos analisados, assim como as demais espécies de *Cerradomys*, registramos a presença de uma Isoleucina (I164, Anexo 4).

Em relação à abordagem intragenérica para a região cd1, foram registradas seis posições mutadas, sendo cinco (139, 142, 149, 151 e 152) com somente MS. A posição 143 foi a única na qual foi registrada uma MNS, correspondendo a uma Guanina (G) na única sequência de *A. galapagoensis*, sendo que *A. xantheolus* e todas as demais espécies possuem uma Alanina (A) para essa posição. Em relação à cd2, verificou-se quatro posições mutadas (157, 159, 160 e 164), sendo MS em três delas (157, 159 e 160) e MNS em duas delas (159 e 164). Na posição 159 verificou-se uma Asparagina (N159) em *Cerradomys goytaca*, *C. languthi*, *C. scotti*, *C. subflavus* e *C. vivoi* e uma Serina (S159) para *C. marinhui* e *C. maracajuensis*. *Hylaeamys acritus* é a única espécie do gênero que possui uma S159, enquanto *H. laticeps* e *H. magacephalus* possuem uma tirosina (T159). *H. yunganus* foi a única espécie do gênero a apresentar variação intragenérica para na posição 159, sendo que das quatro sequências analisadas, três possuem N159 e somente uma T159. Já para a posição 164, foi registrado somente um evento de MNS, que compreende a presença de uma V164 nas duas sequências de *C.*

goytaca analisadas, sendo que as demais espécies de *Cerradomys* possuem I164 (Anexo 4). Assim, observa-se que tanto para a abordagem intragenérica quanto para a intergenérica, as MNS nas regiões de cd1 e cd2 ocorrem entre aminoácidos de propriedades semelhantes, seja quanto a polaridade, carga ou comprimento da cadeia lateral.

Nossos achados, tanto para a abordagem intragenérica quanto para a intergenérica, mostram um maior número de posições mutáveis em cd2 (20% resíduos variáveis) do que cd1 (12,5% resíduos variáveis), mesmo cd1 possuindo uma maior quantidade absoluta de posições mutáveis, nota-se que a proporção de MNS e MS em cd1 é menor do que em cd2. Esses dados são congruentes aos achados de Esser *et al.* (2006) que demonstraram que devido cd1 possuir uma importância funcional maior, essa cadeia possui uma menor proporção de mutações (~93% de resíduos conservados) do que cd2 por exemplo (~69% de resíduos conservados), mesmo cd1 possuindo uma maior quantidade de aminoácidos em sua sequência.

De acordo com o mecanismo de funcionamento do complexo bc1 aceito atualmente para a geração dos elétrons que serão transferidos ao longo da cadeia respiratória na mitocôndria (Esser *et al.* 2006), o sítio no qual a ubiquinona (Q) é reduzido localiza-se próximo ao lado negativo da membrana mitocondrial, conhecido como *Qi site*. O *Qi site* conecta-se via dois grupos protéticos (*high-potential b heme* (bH); *low-potential b heme* (bL)) com o sítio de redução do ubiquinol (Qo), localizado próximo ao lado positivo da membrana. Em Qo, um elétron é transferido do Citb para o Citc conduzido pela *iron-sulfur protein* (ISP), que também constitui uma das subunidades do complexo bc1, enquanto um segundo direciona-se via bH e bL para o sítio Qi. Dessa forma, a ISP é um componente essencial para a conexão entre Qo e Qi, sendo assim fundamental para o funcionamento do Citb como um todo. A ISP possui diferentes pontos de ancoragem ao longo do Citb, sendo estas realizadas por pontes de hidrogênio com aminoácidos dos *loops* CD e EF, porém tais pontes entre ISP e o Citb são majoritariamente realizadas com aminoácidos localizados na cadeia de cd1 (Esser *et al.*, 2006).

Sabe-se que MNS tendem a possuir um maior impacto na proteína, pois, uma vez que a sequência primária de uma proteína está finamente associada a função que esta exerce, MNS podem desviar a proteína de sua função, seja por alterações na estabilidade geral da estrutura proteica ou padrões de interações com ligantes. Dessa forma, devido à essa característica, MNS que alteram negativamente de alguma forma a proteína, tendem

a ser expurgadas por seleção, mantendo a longo prazo somente MS e MNS benéficas (Lercher *et al.*, 2003; Pál *et al.*, 2006).

Nossos dados estão de acordo com a premissa da seleção natural atuar de forma mais rígida em sítios de maior importância funcional para proteína, uma vez que foi observada uma menor proporção de resíduos variáveis em cd1 do que em cd2. Além disso, tais sequências foram geradas a partir de organismos silvestres, sendo assim possível concluir que mesmo havendo a existência de MNS, tais mutações não acarretam uma problemática maior na molécula do Citb, uma vez que a seleção positiva atua de forma mais rígida em proteínas de funções essenciais (Nei & Kumar, 2000). Adicionalmente, observou-se que em todos os registros de MNS nas cadeias cd1 e cd2, os aminoácidos registrados para uma dada posição correspondiam a aminoácidos de propriedades semelhantes. Assim, é possível inferir que as características gerais nas posições mutáveis se mantiveram inalteradas, gerando assim pouco ou nenhum impacto nas funções das cadeias cd1 e cd2.

Padrão de mutação do Citb: seleção x deriva genética

Os resultados encontrados no presente estudo demonstram que as mutações no Citb em orizomíinos ocorreram majoritariamente em regiões de α -Hélices, seguida por mutações em *loop* e Folha β , tanto para a abordagem intra quanto para a intergenérica, sendo observada significância estatística somente nas análises intergenéricas. Em nossas análises intragenéricas, registrou-se 75% das mutações em α -Hélice, 21% em *loop* e 1% em Folha β e esses dados estão proporcionais ao tamanho de cada região na proteína: 71,81% dos resíduos correspondem a α -Hélices; 26,56% a *loops*; e somente 1,63% contituem Folhas β . O mesmo padrão é verificado nas análises intergenéricas: 71% das mutações foram registradas em α -Hélice, 26% em regiões de *loop* e somente 1% em Folha β . Dessa forma, a distribuição das mutações inter e intragenéricas registradas no presente estudo, apresentam uma proporcionalidade similar à observada para as estruturas secundárias do citocromo b e a frequência das mutações ao longo das estruturas secundárias do Citb seria explicada pela proporção do número de resíduos observados nas α -Hélices, *loops* e Folhas β .

Esses achados do DNA mitocondrial em orizomíinos (Rodentia) são contrários à regra proposta genes nucleares de mamíferos. Chiusano *et al.* (1999) estudaram 34 proteínas nucleares em quatro ordens distintas de mamíferos (Primates, Artiodactyla,

Lagomorpha e Rodentia (Muridae) e apontaram que tanto mutações sinônimas quanto não-sinônimas tendem a ocorrer com maior frequência em regiões de *loops*, seguida por α -Hélices e Folhas- β . Outros trabalhos com proteínas nucleares em mamíferos concordam com a correlação entre mutações sinônimas e a estrutura secundária (Adzhubei *et al.*, 1996), enquanto outros com invertebrados não (Tao & Dafu, 1998). Assim, essa correlação mutação e a estrutura secundária parece depender do tipo de proteína e do organismo analisado.

A ideia de investigar a distribuição de variações ao longo da estrutura proteica não é nova. O aumento da disponibilidade de modelos proteicos na literatura nos últimos anos permitiu a realização de comparações entre diferentes organismos a fim de evidenciar se há uma regra geral acerca da distribuição das mutações nas estruturas proteicas.

Wilbur (1985) realizou um dos primeiros trabalhos investigando a distribuição de diferentes códons ao longo da estrutura terciária de diversas proteínas, não incluindo o Citb em suas análises, e demonstrou que em regiões de grande importância funcional há um maior grau de conservação, sugerindo a atuação da seleção natural ao nível de códons nesse processo. Além disso, tem-se demonstrado que não somente a funcionalidade, mas também outros fatores, como a manutenção da estabilidade proteica (Depristo *et al.*, 2005), podem limitar a ocorrência de mutações em determinadas regiões da proteína.

Nossos achados dão suporte aos estudos de Wilbur (1985), sendo as regiões de *loop* do Citb as que apresentam maior sensibilidade às mutações, e portanto com maior grau de conservação em comparação às α -Hélice. Sabe-se que no Citb, os *loops* longos AB, CD, DE e EF são responsáveis pela geração das cavidades que contém os sítios ativos da proteína, enquanto as cadeias de α -Hélices desempenham principalmente função estrutural (Chen *et al.*, 2009). Ainda, todas as sequências parciais com 801 pb abrangem os 267 primeiros resíduos do Citb, que apresentaram maior similaridade com estrutura molde de bovino, e é a região mais conservada quando comparada com as regiões finais da proteína, que possui 380 resíduos. Nos 267 resíduos ocorrem as primeiras cinco cadeias do Citb (A, B, C, D e E) com seus respectivos *loops*, não cobrindo as cadeias F, G e H. Os principais sítios ativos (Qi e Qo), assim como pontos de ancoragem de substratos e ligantes (ISP e grupo Heme) localizam-se nas cadeias A-E. Assim, podemos destacar que o grupo de cadeias A-E possuem uma maior sensibilidade a mutações do que o grupo formado pelas F, G e H.

Sabe-se que MS, causam um menor impacto na estrutura e funcionalidade proteica como um todo, uma vez que mesmo havendo alteração da base nitrogenada, mantém-se o

aminoácido codificado. As MNS, uma vez que alteram o aminoácido codificado, possuem uma maior probabilidade de alterar o ótimo adaptativo da proteína, sendo selecionada negativamente pela seleção natural (Smith & Hurst, 1999). Tal aspecto é respeitado quando observamos a discrepância do tipo de mutação (sinônima e não-sinônima) ao longo da proteína. A maioria das mutações são sinônimas, tanto para as análises inter quanto intragenérica (81 a 88% respectivamente) demonstrando que MNS tendem a ser eliminadas pela ação da seleção natural, e por isso, são observadas em menor frequência. Além disso, mesmo as mutações não-sinônimas registradas ocorreram entre aminoácidos com propriedades físicas similares na maior parte dos casos e não afetaram a estrutura da proteína.

Assim, o padrão observado respeita o postulado na teoria neutra de Kimura (1983) na qual a maioria das mutações ao nível molecular são deletérias, uma vez que ocasionam ao organismo algum efeito negativo, como capacidade de reprodução e sobrevivência. Assim, as mutações neutras seriam fixadas na população e em seguida na espécie por deriva genética.

Concluindo, o Citb mostrou-se uma proteína altamente conservada do ponto de vista estrutural, devido à maioria das mutações serem sinônimas ou não-sinônimas silenciosas, muito embora tenha havido um grande número de mudança na sequência do DNA. As mutações no Citb em orizomíneos ocorreram majoritariamente em regiões de α -Hélices, seguida por mutações em *loop* e Folha β . Podemos sugerir que há um processo de seleção atuando mais fortemente nos primeiros 267 resíduos do Citb, associados ao grupo de cadeias α -Hélices e *loops* nas regiões A-E. O mapeamento dessas alterações mostrou padrões distintos de distribuição das mutações não-sinônimas dentro e entre gêneros, mostrando que as mutações que levaram à separação de gêneros foram distintas daquelas que levaram à separação de espécies.

MÉTODOS

Matriz de dados

Foi construída uma matriz de dados composta por 75 sequências de Citb obtidas do GenBank (www.ncbi.nlm.nih.gov/genbank/) de oito dos 10 gêneros descritos por *Weksler et al.* (2006), totalizando 24 espécies (Anexo 1). Das sequências, 21 estavam completas (1143 pb) e 54 estavam incompletas, variando de 786 a 1140 pb. As sequências foram

alinhas no MEGA 7 (<http://www.megasoftware.net/>), gerando uma matriz de 1143 nucleotídeos e 380 aminoácidos. Os nucleotídeos ausentes nas sequências incompletas foram completados com *n*. Foram utilizados como grupo externo *Thomasomys aureus* (código *genbank* U03540) e *Rhipidomys leucodactylus* (código *Genbank* RLU03550).

Modelagem do Citb

Os modelos comparativos foram construídos baseados no *template* 2A06 (*Bovine cytochrome bc1 complex*) obtido do Protein Data Bank (PDB), com similaridade de 79 a 84%, pelo método de modelagem comparativa realizada pela plataforma online Swiss-Model (Biasini *et al.* 2014). A estrutura conhecida da proteína foi utilizada como *template* para obtenção do modelo da sequência *query* (Baker & Sali, 2001). Os modelos gerados para cada espécie foram armazenados como *pdb output*, que foram visualizados no programa PyMol (DeLano, 2002). A confiabilidade e qualidade dos modelos foram verificadas pela distribuição dos ângulos torcionais Φ (Phi) e Ψ (Psi), examinados através do gráfico de Ramachandran (Ramachandran *et al.* 1968, que descrevem a rotação do *backbone* do polipeptídeo ao redor do N-C α (Φ) e C α -C (Ψ), analisando-se o enovelamento e qualidade estereoquímica, por meio da plataforma online PROCHECK (Laskowski *et al.* 1993). Para cada modelo foi criado o gráfico de Ramachandran (Ramachandran *et al.* 1968) e foram verificadas as regiões favoráveis em relação às torções, assim como a sua respectiva estrutura secundária. Os resíduos de aminoácidos com problemas estereoquímicos foram identificados em regiões desfavoráveis do gráfico. Assim, para a análise dos modelos inferidos neste estudo, foram gerados os respectivos gráficos de Ramachandran para cada modelo/espécie.

Mapeamento das mutações

Com o objetivo de identificar as mutações que sustentam cada clado foi realizada uma análise de Parcimônia, através de uma busca heurística utilizando-se o *software* Winclada (Nixon, 1999) com implementação do NONA (Golobof, 1999), com cálculo dos índices de consistência (CI) e retenção (RI). Em seguida, as variações nas sequências de DNA para cada espécie de um gênero foram mapeadas, levando-se em consideração a ocorrência de mutações exclusivas de cada espécie (variação intragenérica), baseada em uma matriz contendo somente as sequências das espécies e do grupo externo. Foram registradas as mutações exclusivas que sustentavam cada espécie. De forma semelhante, as

mutações intergenéricas foram registradas comparando-se par-a-par as sequências de cada gênero de modo que todos os gêneros foram comparados entre si. Foram mapeados os aminoácidos associados a cada mutação (sinônima - MS ou não-sinônima – MNS), a posição do códon da referida mutação, a posição da mutação na estrutura primária no Citb no *software* Geneious (<http://www.geneious.com/>). No caso de MNS nas variações intragenéricas registrou-se os aminoácidos presentes nas demais espécies do gênero, enquanto nas variações intergenéricas registrou-se os aminoácidos presentes nos demais gêneros. Ainda, verificou-se a potencial alteração na polaridade (Polar; Não-Polar; Basic-Polar; Acidic-Polar) e na carga (Positiva; Negativa; Neutra) da cadeia lateral do aminoácido alterado, assim como em qual estrutura secundária (α -Hélice, Folha β , ou *loop*) cada mutação ocorreu, sua cadeia de alfa-hélice correspondente na proteína (A-H; cd1 e cd2) por meio da visualização *Cartoon* do arquivo PDB no Pymol.

Comparação da estrutura tridimensional do Citb

A similaridade par-a-par entre os modelos tridimensionais do Citb foi calculada pela distância média, em ângstrons, entre os carbonos alfa de todos os aminoácidos das cadeias (Desvio da Raiz Média Quadrática - DRMQ) no programa SPDVIEWER (<http://spdbv.vital-it.ch/>), sendo que quanto menor a distância, maior a similaridade entre os dois modelos. Somente foram utilizadas nessa análise sequências completas, 15 das 75, do Citb com 1143 pb e 380 resíduos (Anexo 2) foram utilizadas nessa análise, totalizando assim 15 espécies analisadas, uma vez que somente espécies com sequências completas foram analisadas e uma sequência por espécie foi modelada. Não foram considerados nos modelos os hidrogênios e ligantes (grupo heme), sendo utilizados nos modelos somente a cadeia principal da proteína. Posteriormente para identificar as regiões de divergência foram realizadas as comparações de estruturas tridimensionais dos modelos através da sobreposição de *backbone* o programa PyMol.

Análise da Entropia de Shannon

A fim de identificar as posições mais variáveis no Citb dos Orizomíneos analisados, foi realizada uma análise da entropia de Shannon (Cover & Thomas, 1991), utilizando a matriz de aminoácidos das 77 sequências. Para tanto, foi utilizado o pacote gratuito *ProDy* (disponível em <http://prody.csb.pitt.edu/>), assim como o pacote *Evol*

(<http://prody.csb.pitt.edu/evol/>), ambos utilizados através do *Python*. No ambiente *Python*, a matriz de dados, ou *multiple sequence alignment* (MSA), foi refinada com o objetivo de minimizar a presença de *gaps*, com *row occupancy*=0,8. O cálculo de entropia foi então realizado com a matriz já refinada. Para a identificação das posições mais variáveis da proteína foi utilizado como *threshold* o valor de 1,0.

Análises estatísticas

A fim de verificar a correlação par-a-par entre os tipos de mutação (sinônima e não-sinônima) e a posição no códon (1^a, 2^a ou 3^a posição), o tipo de estrutura secundária onde a mutação ocorreu (α -Hélice, Folha β , ou *loop*), o tipo de cadeia onde a mutação ocorreu (Cadeia A-H; cd1 e cd2) e a polaridade da cadeia lateral do aminoácido associado à cada mutação (polar ou apolar) foi realizada a análise de qui-quadrado através do uso da plataforma R 3.2.4. Foram considerados significativos aqueles resultados que mostraram um p-valor igual ou menor que 0,05.

Análises genéticas dos dados

A relação hierárquica entre os gêneros e espécies foi determinada a partir da aplicação da análise de Verossimilhança (ML) através do *software* MEGA 7 (<http://www.megasoftware.net/>). Os nucleotídeos ausentes nas sequências incompletas foram completados com *n*. Foram utilizados como grupo externo *Thomasomys aureus* (código genbank U03540) e *Rhipidomys leucodactylus* (código genbank RLU03550). O modelo de evolução molecular que melhor se ajusta aos dados foi calculado pelo uso do critério de informação de Akaike, implementando no *software* Jmodeltest v0.1 (Posada, 2008), selecionando-se o modelo GTR+I+G. Os valores para suporte dos clados da ML foi gerado por meio de 1000 replicações de *bootstrap*. As distâncias genéticas par-a-par entre as 75 sequências analisadas foram calculadas pelo MEGA 7, utilizando o modelo de Kimura-2-parâmetros (K2P)

REFERÊNCIAS BIBLIOGRÁFICAS

- ADZHUBEI, I. A; ADZHUBEI, A. A; NEIDLE, S. 1998. An Integrated Sequence-Structure Database incorporating matching mRNA sequence, amino acid sequence and protein three-dimensional structure data. **Nucleic Acids Research** 26(1): 327-331
- ANDRADE, A.F.B., & BONVICINO, C.R. 2003. A new karyological variant of *Oecomys* (Rodentia: Sigmodontinae) and its phylogenetic relationship based on molecular data. **Genome** 46: 195–203
- AMMAN, B.R., HANSON, J.D., LONGHOFER, L.K. AND BRADLEY, R.D. 2006. Intron 2 (ADH1-I2) of the alcohol dehydrogenase gene: a potential nuclear DNA phylogenetic marker for mammals. **Occasional Papers Museum of Texas Tech University**. 256.
- ANUGOLU, R., GUNDA, S., KAKARLA, R. 2011. Homology modeling and ligand interaction of Cytochrome b protein. **Journal of Medical and Allied Sciences**, 1(2): 79–83.
- BAKER, D. & SALI, A. 2001. Protein structure prediction and structural genomics. **Science**, 294: 93–96.
- BALLARD, J.W.O. & WHITLOCK M. C. 2004. The incomplete natural history of mitochondria. **Molecular Ecology**, 13: 729–744.
- BATALHA-FILHO, H.; PESSOA, R. O.; FABRE, P.H.; FJELDSÅ, J.; IRESTEDT, M. ERICSON, G. P.; SILVEIRA, L. F.; MIYAKI, C. 2014. Phylogeny and historical biogeography of gnateaters (Passeriformes, Conopophagidae) in the South America forests. **Molecular Phylogenetics and Evolution**, 79(1), 422–432.
- BIRKY C.W 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. **Proceedings of the National Academy of Sciences of the United States of America**, 92, 11331–11338.
- BIASINI, M.; BIENERT, S.; WATERHOUSE, A.; ARNOLD, K.; STUDER, G.; SCHMIDT, T.; KIEFER, F.; GALLO C.; BERTONI, M.; BORDOLI, L.; SCHWEDE, T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic Acids Research**, 42(1): 252–258
- BONVICINO, C. R., OTAZU, I. B., BORODIN, P. M. 1999. Chromosome variation in *Oryzomys subflavus* species group (Sigmodontinae, Rodentia) and its taxonomic implication. **Cytologia** 64: 327– 332.
- BONVICINO, C. R., & MOREIRA, M. A. M. 2001. Molecular Phylogeny of the Genus *Oryzomys* (Rodentia: Sigmodontinae) Based on Cytochrome b DNA Sequences. **Molecular Phylogenetics and Evolution**, 18(2): 282–292.
- BONVICINO, C. R. 2003. A new species of *Oryzomys* (Rodentia, Sigmodontinae) of the subflavus group from the Cerrado of Central Brazil. **Mammalian Biology**, 68: 78–90.

- BONVICINO, C. R., CASADO, F., WEKSLER, M. 2014. A new species of *Cerradomys* (Mammalia: Rodentia: Cricetidae) from Central Brazil, with remarks on the taxonomy of the genus. **Zoologia**, 31(6): 525–540.
- BURBRINK, F. T. 2005. Inferring the phylogenetic position of *Boa constrictor* among the Boinae. **Molecular Phylogenetics and Evolution**, 34(1), 167–80.
- CHELLIAH, V., CHEN, L., BLUNDELL, T. L., LOVELL, S. C. 2004. Distinguishing structural and functional restraints in evolution in order to identify interaction sites. **Journal of Molecular Biology**, 342(5): 1487–1504.
- CHEN, Q. L., TANG, X. S., YAO, W. J., LU, S. Q. 2009. Bioinformatics analysis of the complete sequences of cytochrome b of *Takydromus sylvaticus* and modeling the tertiary structure of encoded protein. **International Journal of Biological Sciences**, 5(6): 596–602.
- CHIUSANO, M. L., D'ONOFRIO, G., ALVAREZ-VALIN, F., JABBARI, K., COLONNA, G., & BERNARDI, G. 1999. Correlations of nucleotide substitution rates and base composition of mammalian coding sequences with protein structure. **Gene**, 238(1), 23–31.
- CHIQUITO, E. D., D'ELIA, G., PERCEQUILLO, A. R. 2014. Taxonomic review of genus *Sooretamys* Weksler, Percequillo & Voss (Rodentia: Cricetidae: Sigmodontinae): an integrative approach. **Zoological Journal of the Linnean Society**, 171(4): 842–877.
- CHOTHIA, C. & LESK, A. M. 1986. The relation between the divergence of sequence and structure in proteins. **EMBO J.** 5, 823–826.
- COUES, E. 1890. *Oryzomys* (verbete). In “The Century Dictionary: An Encyclopedic Lexicon of the English Language” (W. D. Whitney, Org.), Vol. 1, p. 4164. Century, New York.
- COVER T, THOMAS J. 1991. Elements of information theory. New York: Wiley-Interscience.
- CUNNINGHAM, M. 2002. Identification and evolution of Australian torrent treefrogs (Anura: Hylidae: *Litoria nannotis* group). **Memoirs of the Queensland Museum** 48: 93–102.
- DELANO W. L. The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA 2002; <http://www.pymol.org>.
- DEPRISTO, M. A, WEINREICH, D. M., & HARTL, D. L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. **Nature Reviews. Genetics**, 6(9), 678–687.

- de CARVALHO, A. B., OLIVEIRA, L. F. B., LANGGUTH, A., FREYGANG, C. C., FERRAZ, R. S., MATTEVI, M. S. 2011. Phylogenetic relationships and phylogeographic patterns in *Monodelphis* (Didelphimorphia: Didelphidae). **Journal of Mammalogy**, 92(1): 121–133.
- da FONSECA, R. R., JOHNSON, W. E., O'BRIEN, S. J., RAMOS, M. J., ANTUNES, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. **BMC Genomics**, 9, 119.
- do PRADO, J. R. 2006. Revisão taxonômica de *Aegialomys* (Weksler, Percequillo & Voss, 2006) (Cricetidae: Sigmodontinae). Dissertação (Mestrado em Ecologia Aplicada) – Escola Superior de Agricultura “Luiz Queiroz”, **Universidade de São Paulo**, São Paulo.
- ELF, J., NILSSON, D., TENSON, T., EHRENBERG, M. 2003. Selective Charging of tRNA Isoacceptors Explains Patterns of Codon Usage **Science** 300(5626), 1718–1722.
- ESPOSTI, M. D., DE VRIES, S., CRIMI, M., GHELLI, A., PATARNELLO, T., MEYER, A. 1993. Mitochondrial cytochrome b: evolution and structure of the protein. **Biochimica et Biophysica Acta**, 1143(3): 243–271.
- ESWAR, N. 2003. Tools for comparative protein structure modeling and analysis. **Nucleic Acids Research**, 31(13): 3375–3380.
- ESSER, L., GONG, X., YANG, S., Y. U., L., YU, C. A., XIA, D. 2006. Surface-modulated motion switch: Capture and release of iron-sulfur protein in the cytochrome bc1 complex. **Proceedings of the National Academy of Sciences of the United States of America**, 103(35): 13045–13050.
- FAIVOVICH, J., HADDAD, C.F.B., GARCIA, P.C.A., FROST, D.R., CAMPBELL, J. A., WHEELER, W.C. 2005. Systematic review of the frog family Hylidae, with special reference to hylinae: phylogenetic analysis and taxonomic revision. **Bulletin of the American Museum of Natural History**, 294: 6–228
- FREDRICK, K., & IBBA, M. 2010. How the sequence of a gene can tune its translation. **Cell**, 141(2), 227–229.
- GARDNER, A. L., & PATTON, J. L. 1976. Karyotypic variation in oryzomyine rodents (Cricetinae) with comments on chromosomal evolution in the Neotropical cricetine complex. **Occasional Papers of the Museum of Zoology Louisiana State University** 49: 1–48.
- GOLOBOFF, P. 1999. NONA (NO NAME) ver. 2 Published by the author, Tucumán, Argentina.
- GÓMEZ-LAVERDE, M., R.P. ANDERSON, GARCIA, L.F..2004. Integrated systematic reevaluation of the Amazonian genus *Scolomys* (Rodentia: Sigmodontinae). **Mammalian Biology** 69: 119–139.

- GREER, J. 1980. Model for haptoglobin heavy chain based upon structural homology. **Proc. Natl. Acad. Sci. USA.** 77: 3393-3397.
- GUO, H. H., CHOE, J., LOEB, L. A. 2004. Protein tolerance to random amino acid change. **Proceedings of the National Academy of Sciences of the United States of America**, 101(25): 9205–9210.
- GU, W., ZHOU, T., WILKE, C. O. 2010. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. **PloS Computational Biology**, 6(2).
- HILBERT, M. BOHM, G., JAENICKE, R. 1993. Structural relationships of homologous protein as a fundamental principle in homology modeling. **Proteins.** 17: 138-151.
- HUANG, L. S., COBESSI, D., TUNG, E. Y, BERRY, E. A. 2005. Binding of the Respiratory Chain Inhibitor Antimycin to the Mitochondrial bc(1) Complex: A New Crystal Structure Reveals an Altered Intramolecular Hydrogen-bonding Pattern. **Journal of Molecular Biology**, 351: 573-597.
- JHONSON, M. S. SRINIVASAN, N. SOWDHAMINI, R. BLUNDELL, T. L. 1994. Knowledge-Based Protein Modeling. **Critical Reviews In Biochemistry and Molecular Biology.** 29: 1-68.
- JOHNS, G. C., & AVISE, J. C. 1998. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. **Molecular Biology and Evolution**, 15(11), 1481–1490.
- KIMURA, M. 1983. The neutral theory of molecular evolution. **Cambridge**
- LASKOWSKI R.A., ANTOON, J. C., MACARTHUR, M. W., KAPTEIN, R., THORNTON J. M. 1996 . AQUA and PROCHECK: programs for checking the quality of protein structures solved by NMR. **Journal of Biomolecular NMR.** 8:477-486.
- LERCHER, M. J., WILLIAMS, E. J. B., HURST, L. D. 2001. Local similarity in evolutionary rates extends over whole chromosomes in human–rodent and mouse–rat comparisons: Implications for understanding the mechanistic basis of the male mutation bias. **Molecular Biology and Evolution.** 18, 2032–2039
- LIU, Y., & BAHAR, I. 2012. Sequence evolution correlates with structural dynamics. **Molecular Biology and Evolution**, 29(9), 2253–2263.
- MARTIN, A P.; PALUMBI, S. R. 1993. Protein evolution in different cellular environments: cytochrome b in sharks and mammals. **Molecular Biology and Evolution**, 10(4): 873–891.
- MAJUMDAR, S.; GUPTA, S. K.; SUNDARARAJAN, V. S.; GHOSH, T. C. 1999. Compositional correlation studies among the three different codon positions in 12 bacterial genomes. **Biochemical and Biophysical Research Communications**, 266(1): 66–71.

- MARTÍ-RENOM, M. A., STUART, A. C., FISER, A., SÁNCHEZ, R., MELO, F., SALI, A. 2000. Comparative protein structure modeling of genes and genomes. **Annual Review of Biophysics and Biomolecular Structure**, 29: 291–325.
- MAZAHERI, P., SHIRAZI, A. H., SAEEDI, N., REZA JAFARI, G., & SAHIMI, M. 2010. Differentiating the Protein Coding and Noncoding Rna Segments of Dna Using Shannon Entropy. **International Journal of Modern Physics C**, 21(01): 1–9.
- MIRANDA, G. B., ANDRADES-MIRANDA, J., OLIVEIRA, L. F. B., LANGGUTH, A., MATTEVI, M. S. 2007. Geographic patterns of genetic variation and conservation consequences in three South American rodents. **Biochemical Genetics**, 45(11-12): 839–856.
- MONTALVAO, R. W., SMITH, R. E., LOVELL, S. C., BLUNDELL, T. L. 2005. CHORAL: a differential geometry approach to the prediction of the cores of protein structures. **Bioinformatics**, bti595.
- MUSSER, G. G., CARLETON, M. D., BROTHERS, E. M., GARDNER, A. L. 1998. Systematic studies of Oryzomyine rodents (Muridae, Sigmodontinae): Diagnoses and distributions of species formally assigned to *Oryzomys capito*. **Bulletin of the American Museum of Natural History**. 236: 1–376.
- MYERS, P., B. LUNDRIGAN, P.K. TUCKER. 1995. Molecular phylogenetics of oryzomyine rodents: the genus *Oligoryzomys*. **Molecular Phylogenetics and Evolution**. 4: 372–382.
- NABHOLZ, B., MAUFFREY, J. F., BAZIN, E., GALTIER, N., GLEMIN, S. 2008. Determination of mitochondrial genetic diversity in mammals. **Genetics**, 178(1), 351–361.
- NEI, M. & KUMAR, S. **Molecular Evolution and Phylogenetics** (Oxford Univ. Press, Oxford, 2000).
- NIXON, K. C. 1999-2002. WinClada ver. 1.0000 Published by the author, Ithaca, NY, USA.
- PATTON, J.L., & HAFNER M.S. 1983. Biosystematics of the native rodents of the Galapagos Archipelago, Ecuador. In R.I. Bowman, M. Berson, and A.E. Leviton (editors), Patterns of evolution in Galapagos organisms. 539–568. **San Francisco: Pacific Division AAAS**.
- PATTON, J.L. & da SILVA M.N.F. 1995. A review of the spiny mouse genus *Scolomys* (Rodentia: Muridae: Sigmodontinae) with the description of a new species from the western Amazon of Brazil. **Proceedings of the Biological Society of Washington** 108: 319–337
- PÁL, D., & CHAKRABARTI, P. 2002. On residues in the disallowed region of the Ramachandran map. **Biopolymers**, 63(3):195–206.

- PÁL, C., PAPP, B., & LERCHER, M. J. 2006. An integrated view of protein evolution. **Nature Reviews. Genetics**, 7(5), 337–348.
- POSADA, D. & CRANDALL, K.A. 1998 Modeltest: Testing the model of DNA substitution. **Bioinformatics**. 14:817-818.
- RAMACHANDRAN, G. N.; SASISEKHARAN, V. 1968. **Advances in Protein Chemistry**, 23: 283-437.
- READ, K., J. S. KEOGH, I. A.W. SCOTT, J. D. ROBERTS, P. DOUGHTY. 2001. Molecular phylogeny of the Australian frog genera *Crinia*, *Geocrinia*, and allied taxa (Anura, Myobatrachidae). **Molecular Phylogenetics and Evolution**. 21: 294–308.
- REDONDO, R. A F., BRINA, L. P. S., SILVA, R. F., DITCHFIELD, A. D., SANTOS, F. R. 2008. Molecular systematics of the genus *Artibeus* (Chiroptera: Phyllostomidae). **Molecular Phylogenetics and Evolution**, 49(1): 44–58.
- ROST, B. 1999. Twilight zone of protein sequence alignment. **Protein Eng.** 12: 85-94.
- SANDER, C; SCHENEIDER, R. 2001. Database of homology-derived protein structures and the structural meaning of sequence alignment. **Proteins**. 9: 56-68.
- SHI, J., BLUNDELL, T. L. MIZUGUCHI, K. 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. **Journal of Molecular Biology**. 310(1): 243-257.
- SMITH, N. G. C., & HURST, L. D. 1999. The effect of tandem substitutions on the correlation between synonymous and nonsynonymous rates in rodents. **Genetics**, 153(3), 1395–1402.
- TAO & DAFU. 1998. The relationship between synonymous codon usage and protein structure. **FEBS Letters**. 433:93-96.
- TAMA F, BROOKS C.L. 2006. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. **Annual Review of Biophysics and Biomolecular Structure**. 35:115–133.
- TAVARES, W. C., PESSÔA, L. M., GONÇALVES, P. R. 2011. New species of *Cerradomys* from coastal sandy plains of southeastern Brazil (Cricetidae: Sigmodontinae). **Journal of Mammalogy**, 92(3), 645–658.
- THOMAS, G. H., WILLS, M. A., SZÉKELY, T. 2004. Phylogeny of shorebirds, gulls, and alcids (Aves: Charadrii) from the cytochrome-b gene: Parsimony, Bayesian inference, minimum evolution, and quartet puzzling. **Molecular Phylogenetics and Evolution**, 30(3): 516–526.
- TULLER, T., CARMI, A., VESTSIGIAN, K., NAVON, S., DORFAN, Y., ZABORSKE, J., PAN, T., DAHAN, O., FURMAN, I., PILPEL, Y. 2010. An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. **Cell**. 141: 344-354.

- WEKSLER, M. 2003. Phylogeny of Neotropical oryzomyine rodents (Muridae: Sigmodontinae) based on the nuclear IRBP exon. **Molecular Phylogenetics and Evolution** 29: 331–349
- WEKSLER, M.; PERCEQUILLO, A. R.; VOSS, R. S. 2006. Ten New Genera of Oryzomyine Rodents (Cricetidae: Sigmodontinae). **American Museum Novitates**. 3537 (3537): 1.
- WORTH, C. L., GONG, S., BLUNDELL, T. L. 2009. Structural and functional constraints in the evolution of protein families. **Nature Reviews. Molecular Cell Biology**, 10(10): 709–20.
- WRIGHT P.E, DYSON H.J. 2009. Linking folding and binding. **Current Opinion in Structural Biology** . 19:31–38.
- YANG Z, MAJEK P, BAHAR I. 2009. Allosteric transitions of supramolecular systems explored by network models: application to chaperonin GroEL. **PloS Comput. Biol.**

ANEXO 1: Relação dos espécimes com sequências do gene *Citb* obtidas do GenBank assim como seus respectivos vouchers.

Espécie	n	nH	Haplótipo	Voucher¹, Número do GenBank (negrito), tamanho em nucleotídeos (grifado)
<i>Aegialomys galapagoensis</i>	1	1	H1	ASK4105, EU579478 , <u>1143</u>
<i>A. xantheolus</i>	2	2	H2 e H3	MVZ145533, EU579479 , <u>1143</u> ; TK134912, EU340015 , <u>1143</u> .
<i>Cerradomys goytaca</i>	2	1	H4	LBCE-JCM11, KP122222 , <u>801</u> ; LBCE-JCM10, KP122221 , <u>801</u>
<i>C. langguthi</i>	1	1	H5	LBCE15905, KP122223 , <u>801</u> .
<i>C. maracajuensis</i>	3	2	H6 e H7	LBCE-SVS883, KP122226 , <u>801</u> ; LBCE-SVS849, KP122225 , <u>801</u> ; MN44178, AF181278 , <u>801</u> .
<i>C. scotti</i>	10	6	H10; H11; H12; H13; H14; H15	TK61881, EU579482 , <u>1143</u> ; LBCE4884, KP122236 , <u>801</u> ; LBCE5713, KP122237 , <u>801</u> ; LBCE5724, KP122238 , <u>801</u> ; LBCE5725, KP122239 , <u>801</u> ; LBCE5848, KP122240 , <u>801</u> ; LBCE7546, KP122241 , <u>801</u> ; LBCE8749, KP122242 , <u>801</u> ; LBCE9446, KP122243 , <u>801</u> ; MN61672, KP122244 , <u>801</u> .
<i>C. subflavus</i>	6	5	H16; H17; H18; H19 e H20	AF181274 , <u>801</u> ; MNRJ61885, EU579481 , <u>1065</u> ; MNRJ61665, GU126543 , <u>801</u> ; LBCE15485, KP122245 , <u>801</u> ; LBCE17255, KP122247 , <u>801</u> ; LBCE17293, KP122248 , <u>801</u> .
<i>C. vivoi</i>	1	1	H17	MN61666, KP122249 , <u>786</u> ;
<i>C. marinhos</i>	5	2	H8 e H9	LBCE17294, KP122228 , <u>801</u> ; LBCE18191, KP122230 , <u>801</u> ; LBCE17786, KP122229 , <u>801</u> ; LBCE17286, KP122227 , <u>801</u> ; CRB1560, AF181279 , <u>801</u> .
<i>Euryoryzomys emmonsae</i>	2	2	H21 e H22	MZUSP 27150, AF251526 , <u>801</u> ; CS37, AF251525 , <u>801</u> .
<i>E. lamia</i>	2	1	H23	CRB983, GU126537 , <u>801</u> ; CRB968, AF181273 , <u>801</u> .
<i>E. macconnelli</i>	5	5	H24, H25, H26, H27 e H28	AMNH272678, EU579484 , <u>1143</u> ; AMNH272669, GU126538 , <u>1140</u> ; MNFS 156, OMU58379 , <u>801</u> ; CCM44-2, AF251528 , <u>801</u> ; MNFS747, AF251527 , <u>801</u> .
<i>E. nitidus</i>	3	3	H29, H30 e H31	TK14571, EU579485 , <u>1143</u> ; MNFS 1419, ONU58383 , <u>801</u> ; MSB166027, AF251529 , <u>801</u> .
<i>E. russatus</i>	5	3	H31, H32 e H33	ORG067, EU579486 , <u>1143</u> ; MNRJ50230, GU126542 , <u>801</u> ; ORG67, AF181272 , <u>801</u> ; AL3466, AF181271 , <u>801</u> ; UFPB381, EF455025 , <u>801</u> .
<i>Hylaeamys acritus</i>	3	3	H35, H36 e H37	USNM597566, AY940625 , <u>1140</u> ; USNM597572, AY940624 , <u>1140</u> ; USNM597565, AY940623 , <u>1140</u> .
<i>H. laticeps</i>	3	3	H38, H39 e H40	MVZ198262, EU579498 , <u>1143</u> ; ML101, AF251522 , <u>801</u> ; EDH22, AF251521 , <u>801</u> .
<i>H. megacephalus</i>	3	3	H41, H42 e H43	MHNLS8061, EU579499 , <u>1143</u> ; FSF149r, KR190461 , <u>801</u> ; USNM549548, AF108695 , <u>1144</u> .
<i>H. yunganus</i>	4	4	H44, H45, H46 e H47	CM76926, EU579500 , <u>1143</u> ; MNFS 1101, OYU58380 , <u>801</u> ; TK73011, DQ224408 , <u>1143</u> ; JLP15535, AF251520 , <u>801</u> .
<i>Nephelomys albigularis</i>	2	2	H 48 e H49	AMNH268125, EU579505 , <u>1143</u> ; TEL1447, DQ224407 , <u>1143</u>
<i>N. levipes</i>	1	1	H50	MVZ171468, JF693875 , <u>801</u>
<i>Oreoryzomys balneator</i>	3	2	H51 e H52	AMNH268144, EU579510 , <u>1143</u> ; AMNH268144, GU126535 , <u>1039</u> ; TEL1854, EU258534 , <u>1143</u> .
<i>Sooretamys angouya</i>	4	3	H53, H54 e H55	MN78781, KF815444 , <u>1137</u> ; TK61763, EU579512 , <u>1143</u> ; MNRJ50234, EU579511 , <u>1143</u> ; MNRJ50234, GU126534 , <u>801</u> ;
<i>Transandinomys bolivaris</i>	1	1	H56	TK135687, EU579513 , <u>1143</u> .
<i>T. talamancae</i>	3	2	H57 e H58	USNM449894, EU579515 , <u>1143</u> ; TK135289, EU579514 , <u>1143</u> ; USNM449894, GU126544 , <u>1103</u> .
Total	75	58		

¹ O número do voucher foi obtido do Genbank ou do artigo original.

ANEXO 2: Relação das mutações exclusivas para a análise intragenérica, por espécie, com o número total de sítios mutáveis (N), número total (n) mutações sinônimas (MS) e não sinônimas (MNS) e posição onde ocorrem as mutações nas cadeias α -Hélices e pares de cadeias (vide Figura 2) .Os pares de cadeias conectados pelos *loops* são peradados por “/”.

Espécie	N	MS		MNS	
		n	Posição	n	Posição
<i>Aegialomys galapagoensis</i>	5	3	cd1, cd2, ef, F	2	cd2; E/ef
<i>Cerradomys goytaca</i>	2	0	-	2	C, cd2
<i>C. langguthi</i>	6	5	A, B, C, D	1	A/N/Terminal
<i>C. maracajuensis</i>	15	14	A, B, C, D, E, cd1, cd2, E/ef	1	C/cd1
<i>C. marinhui</i>	12	10	A/ab, A/N-Terminal, C, D, E	2	E, cd2
<i>C. scotti</i>	7	7	A/N-Terminal, A, B, C, cd1/cd2	0	-
<i>Euryoryzomys emmonsae</i>	10	10	A, B, C, D, cd1, cd2	0	-
<i>E. macconnelli</i>	2	2	B, D/E	0	-
<i>E. nitidus</i>	4	4	A/ab, A/N-Terminal, B	0	-
<i>E. russatus</i>	7	7	A/N-Terminal, A, ab, B, C, D	0	-
<i>Euryoryzoys lamia</i>	9	7	A/ab, A/N-Terminal, A, B	2	ab, C
<i>Hylaeomys acritus</i>	21	18	A, B, C, D, ef, F, cd1, D/E, E/ef, F/G	3	ab, cd2, H
<i>H. laticeps</i>	4	2	B	2	C, D
<i>H. megacephalus</i>	8	8	A/N-Terminal, ab/B, B/C, C, E, F, H, D/E	0	-
<i>H. yunganus</i>	10	10	A, B, C, D, cd1/cd2, E/ef	0	-
<i>Nephelomys levipes</i>	21	17	A/N-Terminal, A, B, C, D, cd2/D, D/E	4	A/ab, A, B
<i>N.ambigularis</i>	14	14	A/N-Terminal, A, ab, B, D, cd1, cd2, cd2/D, E/ef	0	-
<i>Oreoryzomys balneator</i>	5	4	C, E	1	A
<i>Sooretamys angouya</i>	3	2	F	1	G
<i>Transandinomys boliviarius</i>	36	31	A/ab, A/N-Terminal, A, ab, B, C, D, E, ef, F, G, H, cd1, cd1/cd2	5	A, C, H, F/G
<i>T. talamancae</i>	13	13	B, C, D, E, ef, G, H, cd2/D, D/E, F/G	0	-
Total	214	188		26	

ANEXO 3: Relação das mutações exclusivas para a análise intergnérica, por espécie, com o número total de sítios mutáveis (N), número total (n) mutações sinônimas (MS) e não sinônimas (MNS) e posição onde ocorrem as mutações nas cadeias α -Hélices e pares de cadeias (vide Figura 2). Os pares de cadeias conectados pelos *loops* são separados por “/”.

Par de Espécies	N	MS		MNS	
		n	Localização	n	Localização
<i>Aegialomys/Cerradomys</i>	33	27	A/N-Terminal; A; B; B/C; C; cd1; cd2; D; E/ef; F; F/G; G; H	6	ab; B; D/E; F; F/G; H
<i>Aegialomys/Euryoryzomys</i>	21	17	A; B; B/C; C; cd1; cd2; E/ef; ef; ef/F; F; F/G; G; H	4	B; B/C; F/G; H
<i>Aegialomys/Hylaeamys</i>	13	11	A; B; B/C; C; cd1; cd2; ef; F; G; H	2	B; H
<i>Aegialomys/Nephelomys</i>	29	26	A; B; B/C; C; cd1; cd1/cd2; cd2; E; E/ef; ef; F; F/G; G; H	3	B/C; F; H
<i>Aegialomys/Oreoryzomys</i>	31	28	A; B; B/C; C; cd2; cd2/D; D; E/ef; F; F/G; G; H	3	B/C; F/G; H
<i>Aegialomys/Sooretamys</i>	41	34	A; B; B/C; C; cd1; cd2; D; D/E; E/ef; ef; ef/F; F; F/G; G; H	7	A; ab; D/E; F; F/G; H
<i>Aegialomys/Transandinomys</i>	28	21	A/N-Terminal; A; B; B/C; C; cd1; D; E/ef; ef; F; G; H	7	A; B; B/C; F; F/G; H
<i>Cerradomys/Aegialomys</i>	7	6	A/N-Terminal; ab; B; C/cd1; cd2/D; D	1	E
<i>Cerradomys/Euryoryzomys</i>	4	3	A/N-Terminal; C-cd1	1	B/C
<i>Cerradomys/Hylaeamys</i>	9	9	A/N-Terminal; A; ab; B; C/cd1; cd2/D; D	0	-
<i>Cerradomys/Nephelomys</i>	13	11	A/N-Terminal; ab; B; C/cd1; cd2/D; D; E; E/ef	2	B/C; E
<i>Cerradomys/Oreoryzomys</i>	4	2	A/N-Terminal; C/cd1	2	B/C; E
<i>Cerradomys/Sooretamys</i>	11	10	A; ab; B; B/C; C; C/cd1; cd2/D; D/E	1	E
<i>Cerradomys/Transandinomys</i>	8	6	A/N-Terminal; B; C; C/cd1; D	2	B/C; E
<i>Euryoryzomys/Aegialomys</i>	5	2	G; H	3	A/ab; F/G; G
<i>Euryoryzomys/Cerradomys</i>	4	2	A; D	2	A/ab; ab
<i>Euryoryzomys/Nephelomys</i>	5	4	cd2/D; E; ef	1	F/G
<i>Euryoryzomys/Oreoryzomys</i>	7	4	A; cd2; cd2/D; D	3	A/ab; F/G; G
<i>Euryoryzomys/Sooretamys</i>	6	5	A; ab; D	1	A/ab
<i>Hylaeamys/Cerradomys</i>	12	7	cd1; D; ef; F/G	5	A/N-Terminal; ab; G
<i>Hylaeamys/Euryoryzomys</i>	6	5	B; E/ef; ef; F/G; H	1	G
<i>Hylaeamys/Nephelomys</i>	7	4	cd1; E	3	A/N-Terminal; G
<i>Hylaeamys/Oreoryzomys</i>	8	4	A/ab; cd1; D	4	A/N-Terminal; G
<i>Hylaeamys/Sooretamys</i>	11	6	A/ab; cd1; D; H	5	A/N-Terminal; ab; G
<i>Hylaeamys/Transandinomys</i>	5	3	B; E/ef; ef	2	A/N-Terminal; G
<i>Hylaeamys/Aegialomys</i>	8	4	A/ab; cd1; H	4	A/N-Terminal; G

Par de Espécies	N	MS		MNS	
		n	Localização	n	Localização
<i>Nephelomys/Aegialomys</i>	10	7	A/ab; ab; cd1; cd2; D; D/E	3	N-Terminal
<i>Nephelomys/Cerradomys</i>	7	5	A; cd1; D	2	A/N-Terminal; D/E
<i>Nephelomys/Euryoryzomys</i>	10	8	D; E; ef; F; F/G	2	G
<i>Nephelomys/Hylaeamys</i>	3	2	cd1; D	1	A/N-Terminal
<i>Nephelomys/Oreoryzomys</i>	10	8	A; A/ab; ab; C; cd1; cd2; D/E	2	A/N-Terminal
<i>Nephelomys/Sooretamys</i>	14	10	A; A/ab; ab; C; cd1; D/E	4	A/N-Terminal
<i>Nephelomys/Transandinomys</i>	5	4	C; cd2; D	1	A/N-Terminal
<i>Oreoryzomys/Aegialomys</i>	36	32	A/ab; ab; B; C; cd1; cd2; cd2/D; D/E; E; E/ef; ef; F; F/G; H	4	A; E
<i>Oreoryzomys/Cerradomys</i>	23	18	A/N-Terminal; A/ab; B; C; cd1; D; E; ef; F; F/G; G; H	5	A; ab; B; D/E; E
<i>Oreoryzomys/Euryoryzomys</i>	23	19	A/N-Terminal; B; C; D/E; E; E/ef; ef; F; F/G; G; H	4	A; B; E
<i>Oreoryzomys/Hylaeamys</i>	15	12	A/N-Terminal; ab; B; C; cd2/D; E; E/ef; ef; G	3	A; B; E
<i>Oreoryzomys/Nephelomys</i>	31	29	A/ab; ab; B; C; cd2/D; D; E; E/ef; ef; F; F/G; H; cd1	2	A; E
<i>Oreoryzomys/Sooretamys</i>	39	32	A; ab; B; B/C; C; cd1; D; D/E; E; E/ef; ef; ef/F; F; G; H; cd2/D	5	A; ab; D/E; E
<i>Oreoryzomys/Transandinomys</i>	21	18	A/N-Terminal; B; C; D; E/ef; ef; F; H	3	A; B; E
<i>Sooretamys/Aegialomys</i>	41	34	A/N-Terminal; A; A/ab; B; C; cd1; cd1/cd2; cd2; cd2/D; D; D/E; E; ef; F; F/G; G; G/H; H	7	E; F; G; G/H; H
<i>Sooretamys/Cerradomys</i>	28	21	A/N-Terminal; A/ab; B/C; cd1; cd1; D; D/E; ef; F; F/G; G; G/H	7	E; F; G; G/H; H
<i>Sooretamys/Euryoryzomys</i>	24	20	A/N-Terminal; B; B/C; cd1; D; D/E; ef; F; G; G/H	4	B/C; f; G; H
<i>Sooretamys/Hylaeamys</i>	20	15	A/N-Terminal; B/C; D; D/E; ef; F; F/G; G; G/H	5	E; F; G; G/H; H
<i>Sooretamys/Nephelomys</i>	48	41	A/N-Terminal; A; A/ab; B; B/C; cd1; cd1/cd2; cd2; D; E; E/ef; ef; F; F/G; G; G/H; H	7	B/C; E; F; G; G/H; H
<i>Sooretamys/Oreoryzomys</i>	40	35	A/N-Terminal; A; B; B/C; C; cd1; cd2; cd2/D; D/E; F; F/G; G; G/H; H	5	B/C; E; F; G/H; H
<i>Sooretamys/Transandinomys</i>	34	28	A/N-Terminal; A; B; B/C; C; cd1; cd2; D; D/E; ef; F; G; G/H; H	6	B/C; E; F; G; G/H; H
<i>Transandinomys/Aegialomys</i>	20	15	A/ab; ab; B; cd1; cd2; cd2/D; D/E; E; ef; F/G; G; H	5	A/N-Terminal; A/ab; D; G
<i>Transandinomys/Cerradomys</i>	23	17	A; ab; B; cd1; cd2; D; ef; F; F/G	6	A/ab; ab; D; D/E; G
<i>Transandinomys/Euryoryzomys</i>	10	9	B; cd1; cd2; D/E; E/ef; ef; F; F/G	1	A/N-Terminal
<i>Transandinomys/Hylaeamys</i>	11	8	ab; B; cd2; cd2/D; ef; F; F/G	3	A/N-Terminal; A/ab; cd2/D
<i>Transandinomys/Nephelomys</i>	6	5	ab; B; C	1	A/N-Terminal
<i>Transandinomys/Oreoryzomys</i>	20	16	A; A/ab; ab; B; C; cd1; cd2; D; D/E; F; F/G; G; H;	3	A/N-Terminal; A/ab; G
<i>Transandinomys/Sooretamys</i>	25	20	A; A/ab; ab; B; C; cd1; cd2; cd2/D; D; ef; F; F/G	5	A/N-Terminal; ab; D; D/E; G
Total	933	749		181	

ANEXO 4: Variação dos aminoácidos presentes nas cadeias cd1 e cd2, assim como o respectivo *loop* que as conectam. Os “.” são utilizados naqueles aminoácidos semelhantes aos encontrados no grupo externo.

Espécie	cd1															Loop			cd2											
	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	
Grupo Externo	Q	M	S	F	W	G	A	T	V	I	T	N	L	L	S	A	I	P	Y	I	G	N	T	L	V	E	W	I	W	
<i>A. xantheolus</i>
<i>A. galapagoensis</i>	G
<i>C. goytaca</i>	N	V	.	.
<i>C. langguthi</i>	N
<i>C. maracajuensis</i>	S
<i>C. marinhus</i>	S
<i>C. scotti</i>	N
<i>C. subflavus</i>	N
<i>C. subflavus_KP122245</i>	V	N
<i>C. vivoi</i>	N
<i>E. emmonsae</i>
<i>E. lamia</i>
<i>E. macconnelli</i>
<i>E. nitidus</i>
<i>E. russatus</i>
<i>H. acritus</i>	S
<i>H. laticeps</i>
<i>H. megacephalus</i>
<i>H. yunganus</i>	N
<i>H. yunganus_DQ224408</i>
<i>H. albigularis</i>
<i>H. levipes</i>
<i>O. balneator_EU258534</i>	P
<i>O. balneator</i>
<i>S. angouya</i>
<i>T. bolivaris</i>
<i>T. talamancae</i>

ANEXO 5: Alinhamento da sequência EU579478 (TARGET) com os *templates* 1bcc e 1bgyc de um dos cinco modelos gerados através do uso da metodologia implementada pelo CHORAL.

		10	20	30	40	50
1	1bcc	A- PNI RKSHP	LLKMI NNSLI	DL PAPSNI	S AWWNFGS	LLAVCLMTQI LTGL
2	1bgyc	MTNI RKSHP	LMKI VNNA- FI	DL PAPSNI	S S WWNFGS	LLGI CLI LQI LTGL
seq	TARGET	MTI MRKKH-	PLLKI I NHS FI	DL PTPSNI	S S WWNFGS	LLGI CLMVQI I TGL
		60	70	80	90	100
1	1bcc	LLAMHYTADTS	LAFSS VAHTCRNVQY	GWLI RNLHANGAS	FFFI CI FLHI	G
2	1bgyc	FLAMHYTSDTTT	AFSS VTHI CRD	VNYGW I RYMHANGAS	MFFI CLYMHV	G
seq	TARGET	FLAMHYTSDTTT	AFSS VTHI CRD	VNYGWLI RYTHANGAS	MFFI CLFI	HVG
		110	120	130	140	150
1	1bcc	RGLYYGS YLYKET	WNTGVI LLLTLMATAF	VG YVLPWGQMS	FWGATVI TNL	
2	1bgyc	RGLYYGS YTFLET	WNI GVI LLLTVMATAF	MG YVLPWGQMS	FWGATVI TNL	
seq	TARGET	RGI YYGS FMLNET	WNI GI I LLLTTMATAF	VG YVLPWGQMS	FWGGTVI TNL	
		160	170	180	190	200
1	1bcc	FS AI PYI GHTL	VEWA WGGFS VDNPTL	TRFFALHFLLPFAI	AGI TI I HLT	F
2	1bgyc	LS AI PYI GTNL	VEW WGGFS VDKATL	TRFFAFHFI LPFI	I MAI AMVHLL	F
seq	TARGET	LS AI PYI GTTL	VEW WGGFS VDKATL	TRFFAFHFI LPFI	I TALVLVHLL	F
		210	220	230	240	250
1	1bcc	LHESGS NNPLGI	SSDS DKIPFHP	YYSFKDI LGLTLM	TPFLTALFSPNL	
2	1bgyc	LHETGS NNPTGI	SSDVKIPFHP	YYTKDI LGALLI	LALMLLVLFAPDL	
seq	TARGET	LHETGS NNPS	GLNSDS DKIPFHP	YYTKDLLGI LLLLMVLM	LVLFFPDV	
		260	270	280	290	300
1	1bcc	LGD PENFTPAN	PLVTPPHI KPEWYFL	FAYAI LRSI PNKL	GGVLALAASVL	
2	1bgyc	LGD PDNYTPAN	PLNTPPHI KPEWYFL	FAYAI LRSI PNKL	GGVLALAFSIL	
seq	TARGET	LGD PDNYTPAN	PLNTPPHI KPEWYFL	FAYAI LRSI PNKL	GGVMALLLSIL	
		310	320	330	340	350
1	1bcc	I LFLI PFLHKS	KQRTMFRPLS	QTLFWLLVANLLI	LTWIGSQPVEHPFI	I
2	1bgyc	I LALI PLLHTS	KQRSMMFRPLS	QCLFWALVADLLT	LWIGSQPVEHPYI	T
seq	TARGET	I LAAFPLLNYS	KQHGLI YRPI	TQFLYWFI ANLLI	LTWIGSQPVEYPTT	
		360	370	380	390	400
1	1bcc	I GQMASLSYFTI	LLI LFPTI GTLEN	KMLNY-		
2	1bgyc	I GQLASVLYFLLI	LVLMP	TAGTI ENKLLK	W-	
seq	TARGET	I GQISSI LYFTI	I VI LMP	MATMI ENDI	LKLH	

