

**Rodrigo Aparecido Fernandes Redondo**

**Estudos evolutivos em um segmento do  
gene da  $\alpha$ -amilase em morcegos**

Dissertação apresentada ao  
programa de Pós Graduação  
em Genética do Departamento  
de Biologia Geral do Instituto  
de Ciências Biológicas da  
Universidade Federal de Minas  
Gerais, como requisito parcial  
para a obtenção do título de  
Mestre em Genética

Orientador: Prof. Dr. Fabrício R. Santos

Belo Horizonte  
Departamento de Biologia Geral  
Instituto de Ciências Biológicas  
2003

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

043 Redondo, Rodrigo Aparecido Fernandes  
R319e Estudos evolutivos em um segmento do gene da  $\alpha$ -  
amilase em morcegos / Rodrigo Aparecido Fernandes  
Redondo. – Belo Horizonte : UFMG/ ICB, 2003.  
85p. : il.; tab.

Orientador: Fabrício R. Santos

Dissertação (mestrado) – Universidade Federal de  
Minas Gerais, Departamento de Biologia Geral.

1. Morcego – Evolução – Teses. 2. Amilase – Teses.  
3. Evolução molecular – Teses. 4. Filogenia – Aspectos  
moleculares – Teses. 5. Seleção natural – Teses. I.  
Santos, Fabrício Rodrigues dos. II. Universidade Federal  
de Minas Gerais. Departamento de Biologia Geral. III.  
Título.

# Agradecimentos

Agradeço primeiramente ao meu orientador Fabrício Santos, por ter me dado a oportunidade de trabalhar no Lbem e me orientar nesta dissertação. Talvez ele não faça idéia o quão importante ele foi para meu crescimento profissional e pessoal;

À minha família, sempre presentes e importantes na minha formação;

Ao pessoal do Lbem (e membros honorários), Leandro, Simone (ôxe), Gisele, Raquel, Luciana, Beth, Renata, Juliana, Bia, Daniela, Pedro e Cláudia. À Cláudia, Jú e ao Pedro, um especial pelo tempo de convivência;

Aos professores do curso de Mestrado em Genética, em especial à Mônica, à Dolores e ao Romeu pelos bate-papos e discussões;

Aos marmotas, Marcelão, Ricardo, Ted, Marcelo Fred Mercury, Michel, Ewerton, Éder e Fred, pela amizade e presença. Especialmente pro Marcelão;

Aos amigos Luiz, Cris, Flávio e Érika, pela amizade e companheirismo;

Ao pessoal da Zoo, (incluindo os já citados) à Maria Olimpia (MO) e principalmente ao Fernando Perini, pela ajuda nas coletas e discussões;

Ao Ted Baxter (*in memoriam*) e banda Leo, Nata e Lydston;

Ao Lydston devo também muito da minha Iniciação Científica. Valeu por tudo. E não me esqueci da Vanessa, grande e querida amiga;

À turma feliz, Flávia, Renata, Jarina, Lud, Carol e Marina. À Flávia ainda agradeço ao carinho, paciência e bom humor;

À Maja, Lucíola, Xavier, Cibele e Maurício também pela amizade;

Aos amigos do Mestrado Lino, Daniela, Cris, Patrícia, Míriam e Letícia, e também à Alice (OK), que é da Bioquímica, mas não poderia faltar;

Aos meus eternos amigos "outsiders", Pedro, Tarley, Daniel, João, Tércio, Mul, Bocão, X-man, Kaliu e Cláudio. "...e nada mais será como antes..."

Aos laboratórios de Mastozoologia e Citogenética e Escola de Veterinária da UFMG e Departamento de Zoologia da UFAL pelas amostras;

À CAPES, PELD-UFMG e Fundação O Boticário de Proteção à natureza pela bolsa de estudos e financiamento do projeto e ao IBAMA-MG pela concessão da licença;

E, a todos aqueles que de uma forma ou de outra contribuíram para esta dissertação.

# Índice

|  |           |
|--|-----------|
| <b>AGRADECIMENTOS</b>  | <b>2</b>  |
| <b>LISTA DE TABELAS</b>  | <b>5</b>  |
| <b>LISTA DE FIGURAS</b>  | <b>6</b>  |
| <b>LISTA DE ABREVIATURAS</b>   | <b>7</b>  |
| <b>RESUMO</b>  | <b>8</b>  |
| <b>ABSTRACT</b>  | <b>9</b>  |
| <b>INTRODUÇÃO</b>  | <b>10</b> |
| <b>EVOLUÇÃO MOLECULAR</b>  | <b>10</b> |
| FILOGENIA MOLECULAR  | 10        |
| MÉTODOS DE FILOGENIA MOLECULAR   | 11        |
| PROBLEMAS NA INFERÊNCIA FILOGENÉTICA MOLECULAR                           | 14        |
| <b>SELEÇÃO NATURAL E ADAPTAÇÃO MOLECULAR</b>                             | <b>16</b> |
| MÉTODOS PARA A DETECÇÃO DE ADAPTAÇÃO MOLECULAR                           | 17        |
| <b>A ENZIMA <math>\alpha</math>-AMILASE</b>                              | <b>22</b> |
| <b>QUIRÓPTEROS</b>   | <b>24</b> |
| <b>OBJETIVOS</b>   | <b>28</b> |
| <b>MATERIAIS E MÉTODOS</b>   | <b>29</b> |
| <b>OBTENÇÃO DAS AMOSTRAS</b>   | <b>29</b> |
| <b>PREPARAÇÃO DAS AMOSTRAS E EXTRAÇÃO DE DNA</b>                         | <b>30</b> |
| <b>AMPLIFICAÇÃO DOS SEGMENTOS DO GENE DA <math>\alpha</math>-AMILASE</b> | <b>31</b> |
| DESENHO DOS INICIADORES  | 31        |
| AMPLIFICAÇÃO PELA PCR  | 33        |
| PURIFICAÇÃO DOS PRODUTOS DA PCR  | 34        |
| <b>SEQÜENCIAMENTO DOS FRAGMENTOS GERADOS PELA PCR</b>                    | <b>34</b> |

|   |                  |
|---|------------------|
| <b>MONTAGEM E VALIDAÇÃO DAS SEQÜÊNCIAS</b>                              | <b>34</b>        |
| <b>ANÁLISES DOS DADOS</b>   | <b>35</b>        |
| CONJUNTOS DE DADOS  | 35               |
| ALINHAMENTO DAS SEQÜÊNCIAS  | 36               |
| RECONSTRUÇÕES FILOGENÉTICAS   | 36               |
| CARACTERÍSTICAS DOS CONJUNTOS DE DADOS                                  | 37               |
| DETECÇÃO DE SELEÇÃO NATURAL E ADAPTAÇÕES MOLECULARES                    | 37               |
| <b><u>RESULTADOS E DISCUSSÃO</u></b>                                    | <b><u>41</u></b> |
| <b>AMPLIFICAÇÃO POR PCR</b>   | <b>41</b>        |
| <b>OBTENÇÃO E ANÁLISE DAS SEQÜÊNCIAS DE <math>\alpha</math>-AMILASE</b> | <b>43</b>        |
| <b>CARACTERÍSTICAS BÁSICAS DOS CONJUNTOS DE DADOS ANALISADOS</b>        | <b>46</b>        |
| <b>ANÁLISES FILOGENÉTICAS</b>   | <b>48</b>        |
| <b>DETECÇÃO DE ADAPTAÇÕES MOLECULARES</b>                               | <b>56</b>        |
| <b><u>CONCLUSÕES</u></b>  | <b><u>74</u></b> |
| <b><u>REFERÊNCIAS BIBLIOGRÁFICAS</u></b>                                | <b><u>75</u></b> |

## Lista de tabelas

|                |   |    |
|----------------|---|----|
| <b>Tab. 1</b>  | Lista das espécies inicialmente utilizadas neste estudo_____          | 30 |
| <b>Tab. 2</b>  | Pares de iniciadores utilizados nas reações de PCR_____               | 33 |
| <b>Tab. 3</b>  | Sequências obtidas nos bancos de dados do GenBank do NCBI_____        | 35 |
| <b>Tab. 4</b>  | Seqüências de Citocromo B e RAG-2 baixadas do GenBank_____            | 36 |
| <b>Tab. 5</b>  | Resultados das ampliações dos 5 pares de iniciadores selecionados_    | 42 |
| <b>Tab. 6</b>  | Resultados do seqüenciamento_____                                     | 45 |
| <b>Tab. 7</b>  | Estatísticas gerais descritivas para os conjuntos de dados PMR e VERT | 47 |
| <b>Tab. 8</b>  | Teste de Neutralidade NG_____   | 57 |
| <b>Tab. 9</b>  | Teste de Seleção Positiva NG_____                                     | 58 |
| <b>Tab. 10</b> | Teste de Seleção Purificadora NG_____                                 | 59 |
| <b>Tab. 11</b> | Teste de Neutralidade NG (modificado) _____                           | 60 |
| <b>Tab. 12</b> | Teste de Seleção Positiva NG (modificado) _____                       | 61 |
| <b>Tab. 13</b> | Teste de Seleção Purificadora NG (modificado)_____                    | 62 |
| <b>Tab. 14</b> | Resultado do teste de Seleção para as médias dos grupos_____          | 63 |
| <b>Tab. 15</b> | Resultados do teste exato de Fisher_____                              | 64 |
| <b>Tab. 16</b> | Resultado do $r^3t$ do método de Creevey- McInerney_____              | 65 |
| <b>Tab. 17</b> | Resultado dos testes com diversos modelos de substituição de códon_   | 66 |

## Lista de figuras

- Figura 1:** Nova filogenia proposta para as diversas ordens de mamíferos\_\_\_\_\_25
- Figura 2:** Relações filogenéticas entre *Icaronycteris*, *Archeonycteris*, *Hassianycteris* e *Palaeochiropteryx* e as Sub-ordens de morcegos viventes\_\_\_\_\_26
- Figura 3:** Alinhamento das seqüências de cDNA de  $\alpha$ -amilase encontrados no GenBank do NCBI\_\_\_\_\_32
- Figura 4:** Desenho esquemático do gene da  $\alpha$ -amilase\_\_\_\_\_32
- Figura 5:** Cromatogramas das seqüências de  $\alpha$ -amilase\_\_\_\_\_44
- Figura 6:** Árvore de NJ utilizando todas as posições do éxon 4 da  $\alpha$ -amilase para o conjunto PMR de dados\_\_\_\_\_49
- Figura 7:** Consenso das 6 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com todas as posições do éxon 4 de  $\alpha$ -amilase no conjunto de dados PMR\_\_\_\_\_49
- Figura 8:** Consenso das 20 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com apenas as duas primeiras posições dos códons do éxon 4 da  $\alpha$ -amilase no conjunto de dados PMR\_\_\_\_\_50
- Figura 9:** Consenso das 9 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com a seqüência deduzida de aminoácidos do éxon 4 da  $\alpha$ -amilase no conjunto de dados PMR\_\_\_\_\_50
- Figura 10:** Consenso das 5 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com um segmento de 330 pb do Citocromo B\_\_\_\_\_51
- Figura 11:** Reconstrução filogenética utilizando MP com o segmento correspondente ao íntron C da  $\alpha$ -amilase\_\_\_\_\_52
- Figura 12:** Reconstrução filogenética utilizando MP com o segmento correspondente ao íntron C e éxon 4 da  $\alpha$ -amilase\_\_\_\_\_52
- Figura 13:** Consenso das 8 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com todas as posições do éxon 4 de  $\alpha$ -amilase no conjunto de dados VERT\_\_\_\_\_54
- Figura 14:** Consenso das 3 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com todas as posições do éxon 4 de  $\alpha$ -amilase no conjunto de dados PMR mais grupo externo\_\_\_\_\_55
- Figura 15:** Consenso das 6 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com aminoácidos deduzidos no conjunto de dados PMR mais grupo externo\_\_\_\_\_55
- Figura 16:** Numeração dos ramos da filogenia para o teste CM2002\_\_\_\_\_65
- Figura 17:** Alinhamento das seqüências de aminoácidos deduzidas para o éxon 4 da  $\alpha$ -amilase no conjunto de dados PMR\_\_\_\_\_69
- Figura 18:** Alinhamento das sequências de  $\alpha$ -amilase de *Asterias ruber* (estrela-do-mar) e *Sus scrofa* (porco domestico) por BLAST 2 e Dot-Plot de identidade mostrando regiões de similaridade das sequências\_\_\_\_\_71

# Lista de abreviaturas

$\omega$ : Razão das taxas de substituições não-sinônimas e sinônimas,  $d_N/d_S$

$\kappa$ : taxa da razão de transições e transversões

**aa**: aminoácido(s).

**CM2002**: Método para detecção de seleção natural de Creevey & McInerney, 2002

**df**: (degrees of freedom): graus de liberdade

**$d_N$** : taxa de substituições não-sinônimas  $n/N$

**DNA** (Deoxiribonucleic Acid ): Ácido desoxiribonucleico

**$d_S$** : taxa de substituições sinônimas,  $s/S$

**$K_N$** : número de mutações não-sinônimas no modelo de Li et al., 1985

**$K_S$** : número de mutações sinônimas no modelo de Li et al., 1985

**LRT**: (Likelihood Ratio Test): Teste de razão de verossimilhança

**ML** (Maximum Likelihood): Máxima verossimilhança

**mNG**: Método para detecção de seleção natural de Nei & Gojobori, modificado por Zhang et al., 1998

**MP**: (Maximum Parsimony): Parcimônia máxima

**n**: número de mutações não-sinônimas

**N**: numero de sitio não-sinônimos

**NG**: Método para detecção de seleção natural de Nei & Gojobori, 1987

**NJ**: (Neighbor Joining): Agrupamentos Vizinhos

**pb**: pares de bases

**PCR** (Polymerase Chain Reaction): Reação em cadeia da polimerase

**R**: razão de transições sobre transversões,  $ts/tv$

**$r^3t$** : (relative rate ratio test): teste de razão de taxas relativas

**s**: número de mutações sinônimas

**S**: numero de sitios sinônimos

**t**: divergência entre seqüências homólogas de DNA ou proteína

**ts**: transição

**tv**: transversão

**BI** (Bayesian Inference): Inferência Bayesiana

**MCMC** (Markov Chain Monte Carlo): Cadeias de Markov Monte Carlo

**RI** (Replacement Invariable): mutação não-sinônima que é uma sinapomorfia

**RV** (Replacement Variable): mutação não-sinônima que não é uma sinapomorfia

**SI** (Silent Invariable): mutação sinônima que é uma sinapomorfia

**SV** (Silent Variable): mutação sinônima que não é uma sinapomorfia

## Resumo

A Ordem Chiroptera é a segunda mais numerosa ordem de mamíferos e a primeira em diversidade de hábitos alimentares. Estudos comparativos de glândulas salivares em morcegos sugeriram que a evolução destas estruturas está relacionada com a radiação adaptativa destes animais. Parece plausível supor que a diversidade de recursos alimentares explorados pelos morcegos pode estar relacionada à diversificação de suas enzimas digestivas.

Este estudo teve por objetivo principal verificar se existe relação entre a diversidade de hábitos alimentares encontrados em morcegos e modificações adaptativas num segmento do gene codificador da enzima  $\alpha$ -amilase. E também testar algumas hipóteses de relação filogenética entre morcegos e outros mamíferos.

Um segmento de 663 pb correspondendo à totalidade do éxon 4 e uma grande parte do íntron C da  $\alpha$ -amilase foi seqüenciado em nove espécies de morcegos, o éxon 4 foi também seqüenciado em outras nove espécies de mamíferos para comparações e estes dados foram suplementados com seqüências do GenBank. Reconstruções filogenéticas e diversos testes de detecção de seleção natural foram utilizados para a análise.

As árvores filogenéticas geradas para morcegos, primatas e roedores foram bastante congruentes em relação a diferentes métodos de reconstrução filogenética. Quando analisados separadamente, o íntron C e o éxon 4 apresentam filogenias conflitantes para a espécie *Sturnira lilium*. Ambas as reconstruções são diferentes das morfológicas tradicionais e de outras reconstruções moleculares (Citocromo B e Proteína Ativadora de Recombinação (RAG-2)). As árvores para todos os mamíferos e alguns outros vertebrados gerou politomias em demasia para uma análise detalhada. Quando analisados apenas morcegos, roedores e primatas, juntamente com outros vertebrados não mamíferos, a filogenia gerada é congruente com as hipóteses atuais sobre as relações entre estes grupos, indicando roedores e primatas mais relacionados entre si do que com morcegos.

Os testes de seleção revelaram que o segmento codificador analisado está sob forte influência de seleção purificadora, provavelmente devido às restrições relacionadas à função das  $\alpha$ -amilases, mas também revelaram a presença de sítios (códon) sob seleção Darwiniana não direcional. As conflitantes filogenias apresentadas para morcegos não podem ser explicadas por convergência evolutiva, já que as espécies envolvidas não possuem o mesmo hábito alimentar, por isso é provável que a causa desta discrepância seja a retenção de caracteres plesiomórficos ou a aquisição de características semelhantes ao acaso por paralelismo evolutivo.

## Abstract

The Order Chiroptera is the second most diverse order of mammals concerning the number of species and the first one in number of feeding habits. Comparative studies of bats' salivary glands showed that their evolution was related to adaptative radiation of these animals. So it is possible that the diversity of feeding resources used by bats could be correlated with a similar diversification of their digestive enzymes.

In this study we have been searching for a likely relationship between the different feeding habits found in bats, and possible adaptative changes in a coding segment of the  $\alpha$ -amilase enzyme. Besides we have tested some hypothesis about the phylogenetic relationship of bats and other mammals.

A 663 bp segment of the  $\alpha$ -amilase gene, corresponding to the exon 4 and part of the intron C was sequenced in nine species of bats. The exon 4 was also sequenced in further nine mammalian species and we have also used some sequence data retrieved from GenBank. Phylogenetic reconstructions and tests to detect molecular selection were applied to the data analysis.

The phylogenetic trees generated for bats, primates and rodents with different methods produced the same results. When the intron C and the exon 4 were independently analyzed, they showed different phylogenies for the bat species *Sturnira lilium*. Both reconstructions were different from the trees using morphological and other molecular characters (Cytochrome B and Recombination Activator Protein (RAG-2)). The trees obtained with the data set for all mammals and some other vertebrates generated to many polytomies to allow a detailed analysis. When analyzing only bats, primates, rodents and other non-mammal vertebrates, the generated phylogeny was congruent with some modern hypothesis about the relationship between these groups, showing rodents and primates more related between them than with bats.

The tests of Selection showed that the segment of  $\alpha$ -amilase studied is under strong purifying selection, probably because of functional constraints, but the tests also showed the existence of some sites (codons) under non-directional selection. The conflicting bat phylogenies showed could not be explained by evolutionary convergence, because the species involved have not the same feeding habit. Then, the differences in the phylogenies found here probably could be better explained by retention of plesiomorphic characters or the independent acquisition by evolutionary paralelism.

# Introdução

## Evolução Molecular

### Filogenia Molecular

O sonho de reconstruir a história da vida na Terra remonta desde o tempo de Darwin (Li, 1997) e intriga a imaginação de inúmeros cientistas. E tentar reconstruir as relações evolutivas entre os organismos é justamente o papel dos sistematas.

As primeiras metodologias em reconstrução filogenética apareceram com o desenvolvimento de duas escolas de sistemática, a taxonomia numérica (Sneath & Sokal, 1973) e a sistemática filogenética (Hennig, 1966).

A taxonomia numérica baseia-se no cálculo de similaridades entre diversos caracteres existentes nos organismos, os quais são posteriormente agrupados por distintos métodos (ex. UPGMA, *unweighted pair grouping with mean averages*; Sneath & Sokal, 1973). Na realidade os métodos de agrupamento apresentados por Sneath e Sokal não tinham a intenção de apresentar hipóteses sobre as relações evolutivas dos organismos e sim apresentar uma classificação biológica sem premissas evolutivas.

O método de Hennig é um pouco mais complexo e utiliza uma classificação de caracteres em ancestrais (plesiomorfias) e derivados (apomorfias). Esta mudança de estado, ancestral e derivado, bem como o compartilhamento dos caracteres derivados são utilizados para a representação da evolução dos táxons num tipo de dendrograma chamado de cladograma (por isso o método é conhecido como método cladístico). Para a construção do cladograma é utilizado o princípio da parcimônia (se existem duas hipóteses para explicar um fato escolha a mais simples, ex. menor número de mudanças num cladograma) é utilizado como critério para a escolha do "melhor" caminho evolutivo. Este método é utilizado para a classificação, e também mostra as relações de descendência e ancestralidade entre os organismos (Hennig, 1966; revisão em Amorin, 1997).

Os dendrogramas (e cladogramas) passaram a ser conhecidos como árvores filogenéticas, onde as relações evolutivas entre os organismos podiam ser representadas. Posteriormente os métodos acima se expandiram e passaram a utilizar, além de caracteres morfológicos, caracteres fisiológicos, ecológicos e até mesmo comportamentais.

Os estudos moleculares em filogenia se desenvolveram junto com as técnicas de biologia molecular e foram incorporados em ambas as metodologias acima que passaram a utilizar padrões eletroforéticos de DNA e proteínas, sítios de enzimas de restrição, hibridização DNA-DNA, e mais tarde, seqüências de aminoácidos e nucleotídeos (Li, 1997).

Com o aumento do volume de dados gerados (tanto morfológicos quanto moleculares) sua interpretação começou a se tornar mais complexa e novas discussões a respeito dos métodos começaram a se formar, e de uma forma geral, chegou-se à conclusão de que a inferência filogenética é um problema também estatístico e deveria ser abordada de forma semelhante (Li, 1997; Nei & Kumar, 2000; Yang, 2000). Deste modo a filogenia molecular hoje pode ser interpretada como sendo um conjunto de técnicas de biologia molecular e inferência estatística que visa estudar as relações evolutivas entre os organismos (ou genes).

### **Métodos de filogenia molecular**

De uma forma geral os métodos atuais para estudos filogenéticos com seqüências de DNA ou proteína, podem ser resumidos em três tipos de métodos (Nei & Kumar, 2000): Métodos de distância; métodos de parcimônia e métodos probabilísticos. Cada método tem suas vantagens e desvantagens e podem ser testados estatisticamente de diversas maneiras tanto para a topologia das árvores criadas, quanto os tamanhos de seus ramos (Li, 1997; Nei & Kumar, 2000).

Nos métodos de distância, uma matriz de diferenças (ou semelhanças) é construída com base nos dados obtidos (seqüências ou outros dados moleculares) que são então reduzidos a um valor (distância) entre os pares de táxons e um algoritmo de agrupamento é utilizado para montar a árvore filogenética (Li, 1997; Nei & Kumar, 2000; Russo *et al.*, 2001).

Apesar de aparentemente simples, os métodos de distância podem ser bastante complexos. A "distância" entre as seqüências de DNA é calculada com base em modelos evolutivos, que podem ser simples, como o de Jukes e Cantor de 1969, onde taxas de transições e transversões são iguais (Jukes & Cantor, 1969 *apud* Li, 1997; Nei & Kumar, 2000; Russo *et al.*, 2001), um pouco mais complexos como o Tajima 2-parâmetros, onde se diferencia as taxas de transições e transversões (Tajima, 1980; Li, 1997; Nei & Kumar, 2000; Russo *et al.*, 2001) até modelos muito complexos como o Tamura e Nei, que considera taxas de substituição diferenciadas para purinas e pirimidinas em transições e transversões, conteúdo GC das seqüências e presença de substituições múltiplas (Tamura & Nei, 1993; Li, 1997; Nei & Kumar, 2000; Russo *et*

*al.*, 2001). Existem vários outros modelos, inclusive para seqüências de aminoácidos (revisões em Nei & Kumar, 2000; Russo *et al.*, 2001).

Com isso os métodos de distância passaram a ter um importante critério evolutivo, baseado nos modelos de evolução de seqüências (Nei & Kumar, 2000). No entanto alguns autores os consideram aproximados e por isso sujeito a erros (Hasegawa & Fugiwara, 1993).

Diversos métodos de agrupamento para as distâncias genéticas foram criados (ex. UPGMA, Quadrados Mínimos, Agrupamentos Vizinhos, Evolução Mínima, etc. revisões em Li, 1997; Nei & Kumar, 2000; Russo *et al.*, 2001), mas efetivamente apenas alguns são utilizados atualmente para reconstruções filogenéticas.

O mais utilizado é o método de Saitou e Nei (1987) de Agrupamentos Vizinhos (NJ, *Neighbor-joining*), que é um método computacionalmente rápido de construção de árvores sem raiz pré-definida (não existe direção evolutiva pré-definida, pode ser enraizada depois para a análise) e que facilmente incorpora todos os modelos evolutivos para seqüências de DNA e proteínas. Além disto é possível utilizar diversos métodos estatísticos para testar tanto a topologia da árvore quanto o tamanho de cada ramo (Nei & Kumar, 2000).

Os métodos de parcimônia para reconstrução de filogenias moleculares são grandemente embasados no método original de Hennig (1966; Miyaki *et al.*, 2000), porém com modificações para atender as necessidades dos modelos evolutivos moleculares, principalmente a teoria neutra das mutações de Kimura (Li, 1997; Nei & Kumar, 2000; Miyaki *et al.*, 2000), este método e suas variantes passaram a receber o nome de Máxima Parcimônia.

A Máxima Parcimônia (MP) tem a grande vantagem de ser simples em sua concepção, onde se uma mudança basta para explicar um evento não existe a necessidade de explicar este evento de forma mais complexa (ex. com 2 mudanças) (Miyaki *et al.*, 2000). Por ser um método de otimização (utiliza um critério de decisão), a MP é computacionalmente trabalhosa, isso porque temos que examinar todas as árvores e passos evolutivos possíveis e escolher o de menor número de mudanças (Li, 1997; Nei & Kumar, 2000; Miyaki *et al.*, 2000).

O número de árvores possíveis é relacionado ao número de táxons (organismos, genes) que se está estudando, se o número for muito alto a busca por todas as árvores possíveis (busca exaustiva) passa a ser proibitiva computacionalmente, por isso foram desenvolvidos vários métodos de busca de conjuntos mais prováveis de árvores (busca heurística, ex. troca de ramos e decomposição estelar) (revisões em Li, 1997; Nei & Kumar, 2000; Miyaki *et al.*, 2000). A busca pelas árvores pode ser

tanto feita com árvores enraizadas (como no método original de Hennig, 1966), como pode ser feita com árvores sem raiz, o que facilita o trabalho computacional, e se mostra mais adequado ao modelo de evolução neutra (Nei & Kumar, 2000; Miyaki *et al.*, 2000).

Muitas vezes há várias árvores mais parcimoniosas na MP, por isso técnicas estatísticas para a obtenção de uma árvore de consenso foram criadas, como o "Bootstrap", que é um método de reamostragem capaz de gerar valores de confiança para cada ramo na topologia da árvore e pode ser utilizado para testar árvores inclusive dos métodos de distância e probabilidade (Russo, 1997; Nei & Kumar, 2000).

A principal crítica à MP é a consideração de uma taxa constante de mutações por sítio, ignorando a possibilidade de heterogeneidade de substituições, apesar de não estar explícito na sua concepção é uma característica que acompanha o método e pode diminuir a sua eficiência, segundo alguns autores (Yang, 1996; Miyaki *et al.*, 2000; Nei & Kumar, 2000).

Métodos probabilísticos incluem métodos de otimização que utilizam a teoria da probabilidade como critério de escolha da melhor árvore filogenética (Pereira *et al.*, 2000). O método mais utilizado atualmente é o da Máxima Verossimilhança (ML, *Maximum Likelihood*; Felsenstein, 1981).

Reconstruções de árvores de ML são feitas utilizando modelos evolutivos explícitos, tanto pra evolução de seqüências de DNA e proteínas (os mesmos utilizados para os métodos de distância e alguns outros exclusivos e muito complexos) e também modelos de substituição de códons (Goldman & Yang, 1994). Ao contrário dos métodos de distância, a ML utiliza os modelos de evolução molecular para calcular a árvore de maior verossimilhança (probabilidade) (Felsenstein, 1981; Li, 1997; Nei & Kumar, 2000; Pereira *et al.*, 2000). Parâmetros como a razão das taxas de transições/transversões ( $\kappa$ ), o conteúdo GC das seqüências e muitos outros, são estimados diretamente dos dados utilizando também o critério de ML, diminuindo as pressuposições sobre características do modelo evolutivo (Yang, 1996; Pereira *et al.*, 2000).

A busca pela árvore de maior verossimilhança também pode ser feita por busca exaustiva ou heurística, como na MP, porém o gasto computacional da ML é muito superior ao da MP, tanto que buscas exaustivas são proibitivas com um número igual a nove ou dez táxons (organismos, populações ou genes) (Nei & Kumar, 2000).

A reconstrução filogenética por ML tem ganhado grande impulso nos últimos anos, com o aumento do conhecimento sobre os mecanismos de evolução de

seqüências e a melhoria dos métodos computacionais (Yang, 1996), e se considerarmos a inferência filogenética como um problema realmente estatístico, a ML deveria ser o método de escolha para inferências, dado seu papel central na teoria de probabilidade atual (Yang, 2000). Todavia os métodos de ML sofrem diversas críticas, tanto do ponto de vista prático (tempo computacional), como também do teórico, já que os parâmetros são estimados pelos dados, onde erros de amostragem (ex. erros de seqüenciamento ou baixa amostragem populacional) podem ser extremamente prejudiciais (Nei & Kumar, 2000).

Um outro método probabilístico para a estimativa de filogenias que vem sendo utilizado atualmente é o de inferência Bayesiana (IB) (Rannala & Yang, 1996; Yang & Rannala, 1997; Huelsenbeck & Ronquist, 2001; Huelsenbeck *et al.*, 2001). Métodos de inferência Bayesiana se baseiam no cálculo de uma quantificação denominada de probabilidade posterior. Numa reconstrução filogenética, essa probabilidade posterior é, de forma simplificada, a probabilidade da existência de uma árvore condicionada às observações (dados que temos disponíveis) (Huelsenbeck *et al.*, 2001) e o condicionamento para a existência desta probabilidade é calculada utilizando o Teorema de Bayes.

A probabilidade posterior é impossível de ser calculada analiticamente, pois envolve a integração de todas as possíveis combinações de tamanhos de ramos e parâmetros de modelos evolutivos (os mesmos de ML) em todas as árvores (Rannala & Yang, 1996; Yang & Rannala, 1997; Huelsenbeck & Ronquist, 2001; Huelsenbeck *et al.*, 2001), por isso existe a necessidade da utilização de modelos de cálculo numérico conhecidos como MCMC (Markov Chain-Monte Carlo) para a integração do cálculo (Yang & Rannala, 1997).

Apesar de muito semelhante em concepção à ML, a IB é muito mais veloz computacionalmente, mas a sua eficiência em reconstruções filogenéticas ainda precisa ser checada, pois além de ser um método pouco utilizado e recente, sofre das mesmas críticas da ML.

## **Problemas na inferência filogenética molecular**

Na reconstrução filogenética é importante se ter um cuidado especial na escolha dos caracteres (ou seqüências) que se está utilizando para estudar os organismos (ou genes), isso porque apenas os caracteres homólogos (mesma origem) podem ser naturalmente comparáveis (Hennig, 1966).

Quando trabalhamos com seqüências de DNA e proteínas, é preciso saber distinguir não só se duas moléculas são homólogas, mas também de que forma elas são homólogas (Li, 1997; Gogarten & Olendzenski, 1999).

Em moléculas de DNA existem três formas de homologia: ortologia, paralogia e xenologia (Li, 1997, Gogarten & Olendzenski, 1999). Ortologia é a homologia propriamente dita, é quando uma seqüência de DNA é duplicada por um evento de especiação; Paralogia é quando duas cópias são homólogas por eventos de duplicação gênica na mesma espécie e Xenologia é quando uma cópia de um gene é transferida lateralmente para outra espécie, não por descendência (não será discutido aqui) (Li, 1997; Gogarten & Olendzenski, 1999).

Quando há uma duplicação gênica, existe a possibilidade de que uma das cópias comece a se divergir, acumulando mutações e eventualmente se torne um novo gene, e é dessa forma que provavelmente evoluíram inúmeras famílias gênicas (ex. globinas) (Li, 1997).

Em geral quando precisamos fazer inferências filogenéticas devemos utilizar apenas seqüências ortólogas, e se quisermos estudar relações entre famílias gênicas ou o tempo do evento de duplicação devemos utilizar as parálogas (Li, 1997).

O uso incorreto de seqüências parálogas para inferência filogenética pode ter conseqüências como erros na estimativa do tempo de divergência, uma vez que o evento de duplicação gênica pode ser mais antigo do que o evento de divergência entre as espécies (Li, 1997; Gogarten & Olendzenski, 1999; Nei & Kumar, 2000).

Quando trabalhamos com genes nucleares devemos ter consciência deste problema, pois o aumento do tamanho dos genomas está muito relacionado a eventos de duplicação gênica, e a divergência entre genes duplicados é em geral proporcional ao tempo de ocorrência da duplicação, dificultando ainda mais sua correta identificação (Li, 1997). O problema pode se agravar ainda mais, pois pode acontecer por acaso, de duas linhagens terem cópias não ortólogas deletadas e, neste caso, cada linhagem pode possuir apenas uma cópia dos parálogos, um caso de paralogia muito difícil de ser detectado (Li, 1997; Gogarten & Olendzenski, 1999).

Outro fato que dificulta a identificação de cópias ortólogas e parálogas de genes é a possibilidade de ocorrência de evolução em concerto, que pode homogeneizar duas ou mais cópias de genes (ou segmentos) e neste caso, a evolução das cópias gênicas duplicadas não é independente (Li, 1997; Gogarten & Olendzenski, 1999).

Os principais mecanismos conhecidos de evolução em concerto são o "crossing-over" desigual e a conversão gênica, sendo que o "crossing-over" desigual é também um grande responsável pelo aumento do número de cópias em genes, como no caso dos genes ribossomais (Li, 1997). Genes ribossomais estão presentes em grande quantidade (cerca de 400-600 cópias em rãs do gênero *Xenopus*), e a diferença entre as cópias mais divergentes é apenas de 10% (Li, 1997).

A conversão gênica é ainda mais problemática, pois ela não aumenta o número de cópias do gene (Li, 1997; Gogarten & Olendzenski, 1999), e sim homogeneiza segmentos (ou a totalidade) de duas cópias gênicas. Isso a torna particularmente difícil de ser detectada e pode trazer problemas de inferência filogenética de famílias multigênicas (ex. em Li, 1997, Gogarten & Olendzenski, 1999).

Por outro lado a evolução em concerto também pode minimizar problemas de tempos de divergência, se as cópias parálogas são mantidas sob conversão gênica, a divergência entre elas é pequena e o uso destes parálogos, nesse caso inclui um erro menor na estimativa filogenética (Gogarten & Olendzenski, 1999).

Outras complicações que podem interferir na análise filogenética é a ocorrência de eventos de convergência e paralelismos decorrentes de seleção natural e adaptação (Li, 1997; Zhang & Kumar 1997).

## **Seleção natural e adaptação molecular**

Desde a publicação de "Sobre a Origem das Espécies por meio da Seleção Natural" por Darwin em 1859, a seleção natural tem sido parte do pensamento evolucionista tradicional e muitas vezes é até confundida diretamente com o termo evolução (Li, 1997). As mutações que surgissem e conferissem alguma vantagem ou uma melhoria no valor adaptativo ("fitness") de um indivíduo seriam selecionadas e estariam representadas nas próximas gerações daquela população pelos descendentes deste indivíduo, enquanto que o contrário, mutações prejudiciais, seriam subseqüentemente eliminadas do conjunto genético ("*gene pool*") da população (Li, 1997).

Biólogos evolucionistas que trabalham principalmente com morfologia, comportamento e fisiologia, praticamente não encontram dificuldades em comprovar seleção Darwiniana em seus estudos, porém, para os atuais dados genéticos e moleculares (seqüências de DNA e aminoácidos), a existência e, principalmente, a força da seleção natural são ainda motivos de debate (Kimura, 1983; Li, 1997).

Segundo a "Teoria Neutra da Evolução Molecular" de Kimura, a maior parte das mudanças (mutações) em ácidos nucléicos e proteínas são seletivamente equivalentes, sendo mantidas por deriva genética e outros processos estocásticos. Todavia na própria Teoria Neutra, é ressaltado que uma pequena parte da variação não está relacionada ao acaso, mas sim relacionada à fixação por seleção natural (positiva) de alguma mutação vantajosa ou eliminação por seleção natural "purificadora" (negativa) de alguma desvantagem (Kimura, 1983).

A disputa da importância entre a deriva genética e a seleção natural para a evolução, levou à criação de diversos testes estatísticos para a neutralidade de mutações em populações (Li, 1997; Nei & Kumar, 2000; Skibinski, 2000; Fay & Wu, 2001; Nielsen, 2001). A maioria destes testes são utilizados para verificar vantagem seletiva de alelos e polimorfismos dentro da mesma espécie (Fay & Wu, 2001; Nielsen, 2001), sendo que sua interpretação depende de premissas sobre a dinâmica populacional e detalhamento de modelos de seleção. A seleção e a adaptação molecular, que correram de forma episódica durante a diferenciação de espécies ou duplicação gênica, são muito mais difíceis de serem detectadas nestes testes (Yang & Bielawski, 2000; Yang, 2002). Na verdade a maioria dos testes nesse caso consegue rejeitar a hipótese de neutralidade estrita, mas não são poderosos o suficiente para evidenciar de forma inequívoca a presença de seleção Darwiniana (Yang & Bielawski, 2000).

De fato, são raros os casos bem estabelecidos onde existem provas consistentes de seleção natural episódica em nível molecular (Endo *et al.*, 1996; Yang & Nielsen, 1998; Yang & Bielawski, 2000). A maior parte destes casos de adaptação molecular foram detectados em toxinas de origem animal (fosfolipase A<sub>2</sub>, conotoxinas, serina-proteases, revisão em Kordis & Gubensek 2000), proteínas relacionadas aos sistemas de defesa e/ou imunidade do organismo (imunoglobulinas: Tanaka & Nei 1989; Zhang & Nei, 2000; Su *et al.*, 2002), genes de invasão do sistema imune em vírus (Zanotto *et al.*, 1999), genes relacionados à reprodução (espermolisina em baleias : Vacquier *et al.*, 1997; Yang *et al.*, 2000; protamina P1 em esperma de primatas: Rooney & Zhang, 1999) e principalmente genes envolvidos em digestão (lisozimas em ruminantes: Irwin, 1995a, 1995b; lisozimas em primatas: Messier & Stuart, 1997; Zhang *et al.*, 1997; Yang, 1998; ribonucleases em artiodátilos: Jermann *et al.*, 1995).

A maioria dos casos citados acima foram principalmente comprovados através da comparação das taxas de substituições sinônimas ( $d_S$ ) e não-sinônimas ( $d_N$ ) na seqüência de DNA codificante destas proteínas.

## **Métodos para a detecção de adaptação molecular**

Reconstruções da história evolutiva de genes que supostamente estariam sob seleção poderiam ajudar na identificação de evolução adaptativa em nível molecular (Zhang & Kumar 1997). Estas reconstruções poderiam apontar tanto convergências evolutivas (seletivas) quanto evolução paralela (ao acaso) de algumas seqüências (Zhang & Kumar 1997).

O método mais simples para se detectar convergências e paralelismos é o de comparação direta da reconstrução filogenética utilizando o gene alvo contra outra

reconstrução feita com um gene sabidamente neutro, ou com múltiplos caracteres morfológicos (Li, 1997). Porém este método, por ser observacional, não pode ser estatisticamente comprovado, portanto as relações encontradas podem estar simplesmente demonstrando obras do acaso (Zhang & Kumar, 1997).

Outros métodos mais avançados utilizam reconstruções filogenéticas para traçar uma possível seqüência ancestral para cada nó em um cladograma. Esta seqüência pode ser então testada contra as seqüências estudadas, e submetidas a testes estatísticos para cada posição (aminoácido ou nucleotídeo) considerada paralela ou convergente, permitindo verificar se as mudanças observadas são realmente adaptativas ou se ocorreram simplesmente ao acaso (Zhang & Kumar 1997). Todavia estes métodos não são muito bem apoiados por embasamento teórico adequado (Li, 1997; Nei & Kumar, 2000).

Os mais convincentes casos de detecção de seleção natural e adaptação molecular utilizaram a razão ( $\omega$ ) das taxas  $d_N/d_S$  (ver mais adiante) que nos dá uma estimativa a qual pode ser interpretada como sendo neutra se for  $\omega=1$ ; seleção "purificadora" se  $\omega<1$  e seleção positiva se  $\omega>1$  (Yang & Bielawski 2000, Yang, 2002). Este critério é muito mais rigoroso e robusto do que os testes tradicionais (Yang, 2002).

Historicamente Li e colaboradores (1985) foram os primeiros a propor um tipo de teste semelhante ao descrito acima. No método de Li, era feita uma comparação par a par das seqüências, baseada em uma distância que era calculada com base nas possibilidades de degeneração dos códons estudados e mudança nas características físico-químicas dos aminoácidos por eles codificados. A estimativa do número de mutações sinônimas ( $K_S$ ) e não-sinônimas ( $K_N$ ) era calculado separadamente e a diferença entre elas lhe dava uma medida de seleção.

Este método tem diversos problemas: o poder de predição do teste é baixo e a comparação feita apenas por pares não permite discussão de quando o evento aconteceu e não leva em consideração a filogenia das espécies (Nei & Gojobori, 1986; Creevey & McInerney, 2002; Yang & Bielawski, 2000).

Outro método baseado em distâncias par a par foi desenvolvido por Nei e Gojobori (1986). Neste método os conceitos das taxas  $d_S$  como sendo a razão do número de substituições sinônimas (**s**) por sítios sinônimos (**S**) e  $d_N$ , como sendo o número de substituições não-sinônimas (**n**) por sítios não-sinônimos (**N**) foram apresentados. Estes valores são calculados diretamente da comparação entre seqüências utilizando modelos de evolução de códons (ex. Nei & Gojobori, 1986).

Os cálculos de **S** e **N**, são realizados, de forma simplificada (Nei & Kumar, 2000), da seguinte maneira: para cada nucleotídeo de um códon é calculado o número de possíveis substituições sinônimas e não-sinônimas, por exemplo, CGA, CGT, CGC, CGG, AGA e AGG codificam Arginina; se tivermos um CGA em uma seqüência, veremos que a terceira posição é totalmente degenerada, logo esta posição é 100% sinônima (1), se observarmos a segunda posição veremos que esta é totalmente não degenerada logo ela é 100% não-sinônima (1), mas na primeira posição temos duas possíveis mudanças não-sinônimas T e G e uma sinônima A, logo este sítio é 1/3 sinônimo e 2/3 não-sinônimo, se fossemos contar **S** e **N** neste códon teríamos:  $S=4/3= 1,33$  e  $N=5/3= 1,66$ . Para uma dada seqüência, **S** e **N** são o somatório de todos os possíveis sítios sinônimos e não-sinônimos respectivamente. No entanto, na comparação entre duas seqüências as médias de **S** e **N** são utilizadas.

Os cálculos dos valores de **s** e **n** só podem ser realizados em comparações entre seqüências, e são realizados, simplificada, da seguinte maneira: imagine que em uma comparação de duas seqüências homólogas tenhamos TTT em uma e GTA na outra. Entre estas duas seqüências existem apenas dois caminhos parcimoniosos: 1: TTT (Phe) ↔ GTT (Val) ↔ GTA (Val) e 2: TTT (Phe) ↔ TTA (Leu) ↔ GTA (Val). O primeiro caminho utiliza uma mutação não-sinônima e uma sinônima, e o segundo utiliza duas mutações não-sinônimas, logo os valores de **n** e **s** são respectivamente 3/2 (1,5) e 1/2 (0,5). O somatório destes cálculos em todos os códons é utilizado na comparação (Nei & Kumar, 2000). Logicamente existem diversos modelos matemáticos para a evolução de seqüências que são utilizados para realizar estes cálculos levando em consideração aspectos evolutivos.

Os resultados da diferença  $d_N - d_S$ , calculados a partir dos valores **S**, **s**, **N** e **n** entre cada par de seqüências podem ser avaliados com testes estatísticos tradicionais, como um teste Z, onde pode-se testar: a neutralidade se  $d_N \neq d_S$ ; seleção positiva se  $d_N > d_S$  e seleção purificadora se  $d_N < d_S$ . Este método pode ser utilizado para um grande número de seqüências ao mesmo tempo se estimarmos as médias de  $d_N$  e  $d_S$ .

Messier e Stewart (1997) incorporaram algumas modificações no método acima, entre elas a possibilidade de estudar as mudanças num contexto filogenético. A metodologia consiste na construção de uma árvore filogenética representando a história evolutiva das seqüências estudadas, e, para cada nó interno da árvore, seqüências ancestrais hipotéticas são inferidas por parcimônia, e então todas as distâncias (taxas)  $d_N$  e  $d_S$  par a par entre todas as seqüências reais e hipotéticas são calculadas e testadas. Esta metodologia aumentou o poder de predição do teste, e

ainda colocou-o num contexto filogenético discriminando o tempo em que um evento evolucionário pode ter ocorrido.

Vários outros métodos baseados em distâncias foram desenvolvidos (revisão em Yang & Bielawski, 2000; Nei & Kumar, 2000) e vários autores apresentaram testes estatísticos mais apropriados para determinadas situações utilizando os modelos de distância, entre eles a substituição do teste Z (considerado ineficaz para amostras pequenas) por teste t ou por um teste exato de Fisher (Zhang *et al.*, 1997, Nei & Kumar, 2000).

Os métodos até agora descritos são ditos como métodos aproximativos. Críticas a estes métodos são feitas principalmente porque características da evolução das seqüências, como razão (R) e taxa ( $\kappa$ ) de transições e transversões e tendências no uso de códon não eram incorporadas nos modelos (Li, 1993 e 1997, Nei & Kumar, 2000 Yang & Bielawski, 2000). Considerar as taxas de transições e transversões como sendo iguais pode ter um efeito desastroso nas estimativas de  $d_N$  e  $d_S$ , uma vez que as mutações na terceira posição ("wobble") dos códons são mais prováveis de serem transições do que transversões, o que causa uma subestimativa de **S** (número de sítios sinônimos) e com isso uma superestimativa de  $d_S$  (Yang & Bielawski, 2000). Recentemente estas características de evolução de seqüências foram incorporadas nos métodos de distância (Li, 1993; Zhang *et al.*, 1998; Yang & Nielsen, 2000; Nei & Kumar, 2000).

Métodos mais rigorosos e avançados para a estimativa de  $d_N$  e  $d_S$  vem sendo desenvolvidos mais recentemente (Yang, 1998). Estes métodos utilizam Máxima Verossimilhança (ML, *Maximum Likelihood*; Felsenstein, 1981) para estimar diversos parâmetros como divergência entre as seqüências (t), razão das taxas de transições/transversões ( $\kappa$ ) e a razão  $\omega$  ( $d_N/d_S$ ) diretamente dos dados. Estes parâmetros são utilizados em modelos explícitos de substituição de códons (Goldman & Yang, 1994).

O teste estatístico nestes casos é um Teste de Razão de Verossimilhança ou LRT (*Likelihood Ratio Test*), que é utilizado para testar qual entre dois modelos melhor se enquadra no caso estudado. Por exemplo podemos utilizar um modelo com o parâmetro  $\omega=1$  e outro com  $\omega$  como um parâmetro livre, o primeiro serve como hipótese nula. Assim para compará-los o LRT é calculado como duas vezes a diferença entre o logaritmo natural de cada verossimilhança (ou  $2\Delta\log_e(L)$ ), então este valor pode ser comparado a uma distribuição  $\chi^2$  com o número de graus de liberdade (df) igual à diferença no número de parâmetros utilizados nos modelos.

Os parâmetros dos modelos de ML podem ser otimizados para calcular  $d_N$  e  $d_S$  para os dados como um todo, para testar a presença de seleção em ramos específicos

em uma árvore filogenética (Yang 1998; Yang & Nielsen, 1998), para detecção de sítios individuais sob seleção (Yang *et al.*, 2000) e ainda para detectar seleção em sítios individuais em linhagens específicas (Yang & Nielsen, 2002).

No caso dos sítios individuais, após os LRTs são utilizadas análises Bayesianas para apontar quais destes sítios estão sob seleção, permitindo o cálculo de probabilidades posteriores que são muito utilizadas em diversos métodos de estatística avançada (Yang *et al.*, 2000; Anisimova *et al.*, 2002).

Os métodos de ML são bastante rigorosos, e simulações computacionais mostraram que o poder de predição e a exatidão dos LRTs são bastante elevados (Anisimova *et al.*, 2001; 2002), mas alguns autores parecem discordar disto, afirmando que a aproximação de  $\chi^2$  não é confiável para amostras pequenas (Nei & Kumar, 2000).

Um último método (CM2002) a ser comentado foi desenvolvido recentemente por Creevey e McInerney em 2002, e pode ser considerado um teste de razão de taxas relativas ( $r^3t$  - *relative rate ratio test*). Estes autores se basearam no teste descrito por MacDonald e Kreitman (1991) onde substituições sinônimas são utilizadas como estimativas de um padrão de substituição neutralista e se observa como as substituições não-sinônimas desviam deste padrão (MacDonald & Kreitman, 1991; Creevey & McInerney, 2002).

A diferença básica entre os dois modelos é que o de MacDonald e Kreitman era um modelo principalmente utilizado para diferenciação entre o limiar de espécies, necessitando de um grande número de seqüências da mesma espécie, enquanto que o modelo de Creevey e McInerney foca em testes de árvores filogenéticas, podendo ser utilizado tanto dentro quanto entre espécies (Creevey & McInerney, 2002).

O método CM2002 consiste em estimar todas as distâncias par a par de  $d_N$  e  $d_S$ , utilizando o método de Li (1993), depois, utilizando uma árvore filogenética, todas as seqüências hipotéticas ancestrais são inferidas por parcimônia (como no método de Messier e Stewart, 1997) então todas as seqüências (hipotéticas e reais) são comparadas, e as mutações são classificadas em RI (*Replacement Invariable* ou mutação não-sinônima que é uma sinapomorfia), RV (*Replacement Variable* ou mutação não-sinônima que não é uma sinapomorfia), SI (*Silent Invariable* ou mutação sinônima que é uma sinapomorfia) e SV (*Silent Variable* ou mutação sinônima que não é uma sinapomorfia). As razões de RI/RV e SI/SV são comparadas em um teste G com distribuição  $\chi^2$  ou teste exato de Fisher, com a hipótese nula sendo RI/RV = SI/SV como neutralidade, em caso de diferença significativa, RI e RV podem ser comparados, e um número maior de RI é considerado como seleção direcional e o contrário como não direcional. Em caso de falha do primeiro teste, é

possível testar o contrário e verificar seleção purificadora (Creevey & McInerney, 2002).

Na prática e na realidade, não é possível saber se as mudanças detectadas em qualquer dos métodos acima são realmente adaptativas, pois normalmente não se sabe o seu efeito real na proteína. Seria interessante observar a estrutura tridimensional das proteínas em estudo. Proteínas podem possuir uma estrutura muito semelhante e diferir na seqüência linear de aminoácidos, mantendo a mesma função. Portanto é plausível que as alterações na estrutura ocorram onde a seleção natural mais provavelmente atua (Golding & Dean, 1998). Algumas substituições em resíduos de aminoácidos podem alterar conformações na proteína, ou alterar sítios catalíticos, permitindo a proteína a atuar com maior eficiência ou em condições antes impossíveis, como em pHs diferentes ou em altas temperaturas (Irwin, 1995a, 1995b, Golding & Dean 1998, Kordis & Gubensek 2000, Todd *et al.* 2001).

Outro método interessante seria investigar no caso de enzimas, como elas se diferenciam na forma de atuação real com o substrato, produzindo artificialmente as seqüências ancestrais por mutagênese sítio-dirigida, como no caso das ribonucleases de artiodátilos (Jermann *et al.*, 1995).

Na impossibilidade de se reconstruir e observar as proteínas atuais e ancestrais, nada parece comprovar melhores adaptações em nível molecular do que se utilizar os vários métodos estatísticos, reconstruções filogenéticas, e informações sobre a proteína estudada em conjunto (Golding & Dean, 1998). Se a seleção de um caráter adaptativo puder ser realmente comprovada por mais de um método independente, a possibilidade da ocorrência deste na realidade passa a ser extremamente plausível.

## **A enzima $\alpha$ -Amilase**

A  $\alpha$ -amilase ( $\alpha$ -1,4-glucano-4-glucanohidrolase E.C. 3.2.1.1) é uma hidrolase glicolítica que catalisa a hidrólise de ligações glicosídicas  $\alpha$ -(1-4) de amido e outros poli- e oligossacarídeos relacionados (Janecek, 1997) e é extremamente distribuída em todos os três Domínios da vida: *Bacteria*, *Eukarya* e *Archaea* (Janecek 1994, Janecek, 1997).

A enzima faz parte de um grupo de  $\alpha$ -glicosilases denominadas de família da  $\alpha$ -amilase, justamente por esta ser a mais bem estudada do ponto de vista estrutural (Janecek, 1994, 1997). Sua estrutura é formada por barris  $(\alpha/\beta)_8$ , e parece ser muito mais bem conservado evolutivamente do que sua seqüência primária (Janecek, 1994, 1997). Em todos os mamíferos em que a estrutura primária da  $\alpha$ -amilase foi descrita até agora (humanos, ratos, camundongos e porcos), a enzima parece ter 496

resíduos de aminoácidos com um sítio ativo formado por uma tríade catalítica de Asp, Glu e Asp (Nishide *et al.*, 1986a, 1986b; Schibler *et al.*, 1982; MacDonald *et al.*, 1980; Daris *et al.*, 1999).

A proteína é formada por 10 éxons, sendo o éxon 4, o mais importante adaptativamente por ser a localização de dois dos aminoácidos da tríade catalítica (Glu e Asp) e conter sítios de ligação de cálcio (Nishide *et al.*, 1986a; Nishide *et al.*, 1986b).

Dois *loci* de  $\alpha$ -amilase, um da amilase salivar (Amy1) e um da pancreática (Amy2) parecem co-existir nos genomas de mamíferos (Groot *et al.*, 1989), porém no banco de dados do NCBI ([www.ncbi.nlm.nih.gov/LocusLink/](http://www.ncbi.nlm.nih.gov/LocusLink/)) apenas estão descritos e mapeados os dois *loci* de humanos, camundongos e ratos. Para a espécie do porco apenas o *locus* da  $\alpha$ -amilase 2 (Amy2) está depositado no GenBank e na espécie do boi há apenas o da  $\alpha$ -amilase 1 (Amy1).

Em humanos, várias cópias dos dois genes existem, sendo o conjunto mais comum o de 5 genes: Amy1A, 1B e 1C (1C é um pseudogene) e Amy2A e Amy2B (Groot *et al.*, 1989). Uma duplicação *de novo* semelhante existe na Amy2 em uma linhagem de camundongos de laboratório (Strahler & Meisler, 1982; Wiebauer *et al.*, 1985). As cópias de  $\alpha$ -amilase em humanos são extremamente semelhantes, com bordas éxon/íntron idênticas e 98% de similaridade em seqüência de nucleotídeos da região codificadora (Horii *et al.*, 1987). As cópias parecem ser mantidas semelhantes por processos de evolução em concerto (Gumucio *et al.*, 1988). Samuelson e colaboradores (1988, 1990 e 1996) chegaram à conclusão que todas as cópias dos genes de  $\alpha$ -amilase em humanos são eventos de duplicações recentes que ocorreram durante a evolução dos primatas. Aparentemente os promotores dos genes das enzimas salivares de primatas e roedores têm origem independente sugerindo eventos de duplicações também independentes (Samuelson *et al.*, 1988, 1990 e 1996). Neste contexto, a cópia humana mais antiga parece ser a Amy2B e a mais nova Amy1C e o promotor para as Amy1 não existe em primatas do novo mundo (Samuelson *et al.*, 1988, 1990 e 1996). Estes dados corroboram os de Junqueira e colaboradores (1973) sobre a expressão da amilase salivar (Amy1). Neste trabalho, os autores só detectaram amilase na saliva de algumas espécies de roedores, lagomorfos, primatas e morcegos.

Desta forma Meisler e Gumucio em 1986 (*apud* Samuelson *et al.*, 1996) criaram dois modelos para a evolução das amilases em mamíferos, um no qual o ancestral dos mamíferos possuía tanto o *locus* da amilase pancreática, quanto o da salivar, sendo esta extinta em várias linhagens, e outro no qual o ancestral possuía apenas a amilase pancreática, e sendo a salivar adquirida de forma independente em algumas

linhagens. A aquisição independente de amilase salivar em diferentes linhagens parece ser mais provável devido à origem independente dos promotores em roedores e primatas (Samuelson *et al.*, 1988, 1990 e 1996). Além disso, é sugerido que a aquisição de amilase salivar ajude a identificar por paladar, fontes melhores de nutrientes, sendo isso favorecido por seleção natural (Ting *et al.*, 1992).

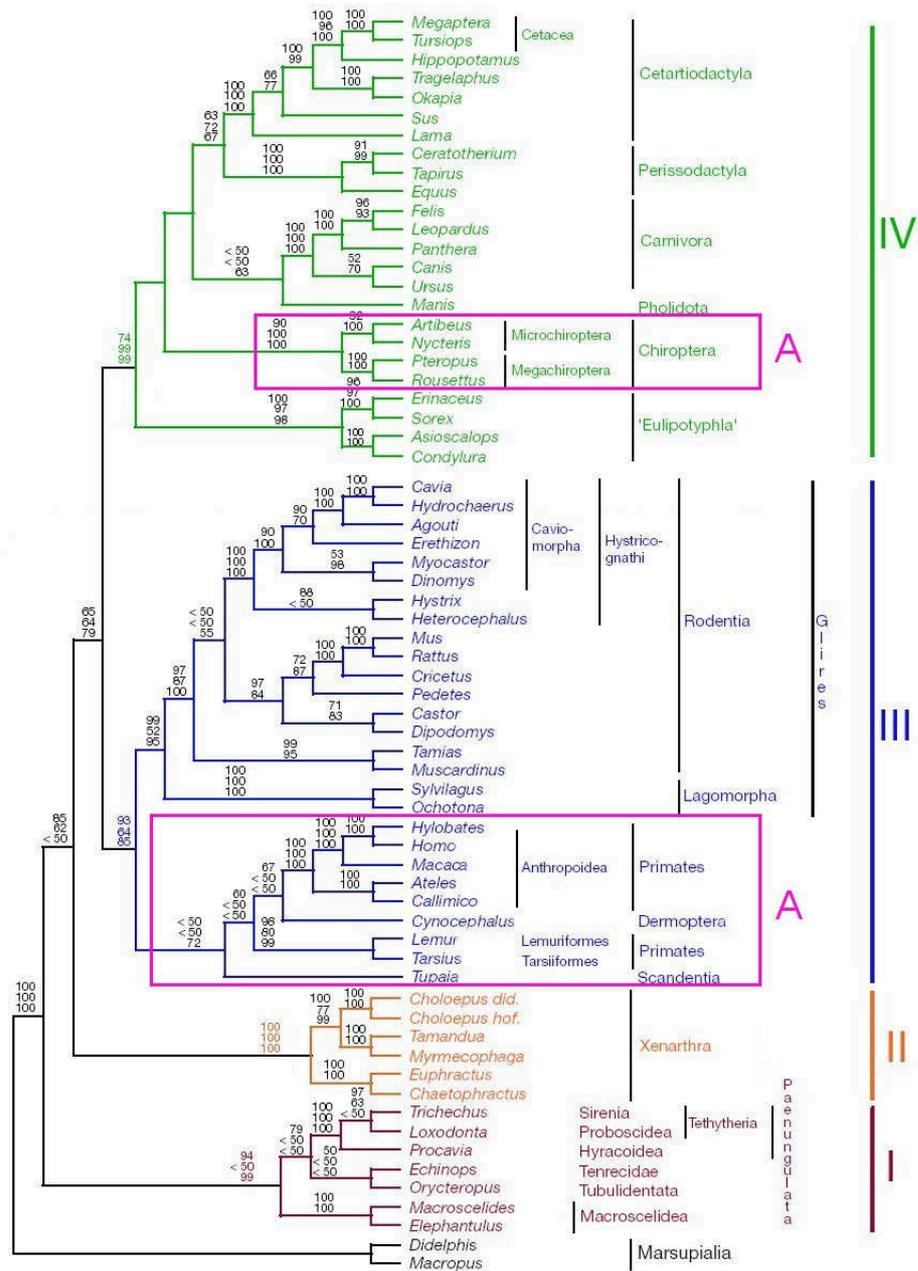
## Quirópteros

A ordem Chiroptera é a segunda mais numerosa ordem de mamíferos, possuindo cerca de 925 espécies (Koopman, 1993; Nowak, 1997). A ordem que compreende os morcegos é dividida em duas sub-ordens, uma restrita a algumas regiões da África, Índia, Austrália, sudeste da Ásia e algumas ilhas do Oceano Pacífico: Megachiroptera; e outra possuindo uma ampla distribuição mundial, só não ocorrendo nas regiões polares: Microchiroptera (Koopman, 1993; Nowak, 1997).

Cerca de 150 das 925 espécies de morcegos conhecidas (aproximadamente 16%) habitam os Neotrópicos, destas, 141 ocorrem no Brasil, quase 30% do total de mamíferos continentais com distribuição registrada para o território brasileiro (Fonseca *et al.*, 1996; Nowak, 1997).

Classicamente os quirópteros são agrupados com as ordens Dermoptera (lêmures voadores), Primata (macacos, micos, lêmures, humanos, etc.) e Scandentia (musaranhos arborícolas) na superordem Archonta (Novacek, 1992) (Figura 1). Esta classificação tem recebido suporte de algumas análises realizadas com dados moleculares também (Adkins & Honeycutt, 1991; Allard *et al.*, 1996).

Entretanto, trabalhos recentes (de Jong, 1998; Nikaido *et al.*, 2000; Madsen *et al.*, 2001; Murphy *et al.*, 2001; Arnason *et al.*, 2002) baseados em seqüências completas de mitocôndrias e concatenação de genes nucleares não dão apoio à existência de Archonta, e sim a uma outra classificação, onde Primata, Scandentia e Dermoptera são agrupados com Lagomorpha (coelhos) e Rodentia (roedores), formando o grupo Euarchontoglires (Madsen *et al.*, 2001; Murphy *et al.*, 2001; Arnason *et al.*, 2002; Deusuc *et al.*, 2002), enquanto os Chiroptera são classificados no grupo Laurasiatheria, junto com Carnivora (gatos, cães, ursos, focas, etc.), Pholidota (pangolin), Perissodactyla (cavalos, rinocerontes, etc.), Cetartiodactyla (Artiodactyla: porco, boi, hipopótamo, carneiro, mais Cetacea: baleias, golfinhos) e Soricomorpha (toupeiras) (Nikaido *et al.*, 2000; Madsen *et al.*, 2001; Murphy *et al.*, 2001; Arnason *et al.*, 2002; Deusuc *et al.*, 2002) (Figura 1).



**Figura 1:** Nova filogenia proposta para as diversas ordens de mamíferos. **I** Afrotheria; **II** Xenarthra; **III** Euarchontoglires; **IV** Laurasiatheria. Destacado nos retângulos assinalados com **A**: Archonta. Modificado de Murphy *et al.*, 2001.

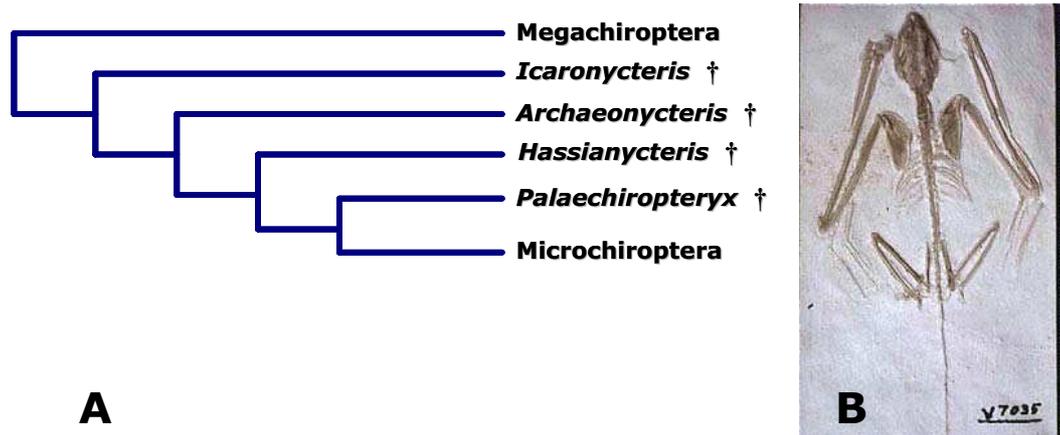
Além da posição da ordem dentro dos grupos de mamíferos, a monofilia das sub-ordens de Chiroptera também é controversa (Pettigrew, 1986; Mindel *et al.*, 1991; Ammerman & Hillis, 1992; Bailey *et al.*, 1992; Kirsch, 1996). Pettigrew (1986) estudou a morfologia cerebral e sugeriu que os megaquirópteros são mais relacionados aos primatas do que aos microquirópteros, porém esta hipótese não é apoiada nem por dados moleculares (Bailey *et al.*, 1992) e nem por outros caracteres morfológicos (Springer *et al.*, 2001). Os estudos recentes (Madsen *et al.*, 2001;

Murphy *et al.*, 2001; Arnason *et al.*, 2002; Deusuc *et al.*, 2002) entre as ordens de mamíferos citados acima, também suportam a monofilia de Chiroptera.

A controversa evolução dos morcegos deve-se principalmente à escassez de seu registro fóssil e sua rápida radiação adaptativa (Jepsen, 1977; Simmons & Geisler, 1998).

Os fósseis de morcegos são datados desde o Eoceno e foram encontrados em todos os continentes com exceção da Antártida (Simmons & Geisler, 1998). *Icaronycteris*, *Archeonycteris*, *Hassianycteris* e *Palaeochiropteryx* são os únicos fósseis encontrados até hoje que não pertencem a uma família vivente de morcegos (Simmons & Geisler, 1998).

A posição destes fósseis em Chiroptera já foi muito controversa (ver revisão em Simmons & Geisler, 1998), mas hoje já existe um consenso de que todos eles são mais relacionados aos microquirópteros do que aos megaquirópteros, na verdade eles são considerados uma série de táxons irmãos de Microchiroptera (Simmons and Geisler, 1998) (Figura 2).



**Figura 2:** **A:** Relações filogenéticas entre *Icaronycteris*, *Archeonycteris*, *Hassianycteris* e *Palaeochiropteryx* e as Sub-ordens de morcegos viventes (Simmons and Geisler, 1998). **B:** Fóssil de *Icaronycteris index*.

O mais antigo fóssil de morcego é o *Icaronycteris index*, que é datado do início do Eoceno, e, apesar de possuir algumas características consideradas ancestrais em morcegos (e algumas semelhantes a megaquirópteros), o *Icaronycteris* possui todas as características necessárias ao vôo batido, e análises cranianas sugerem fortemente a capacidade de ecolocação igual ao dos microquirópteros atuais (Jepsen, 1977; Simmons and Geisler, 1998). Estas características avançadas sugerem que os morcegos tenham surgido em algum ponto do final do Paleoceno e se diversificaram muito rapidamente (Jepsen, 1977; Simmons and Geisler, 1998).

A mobilidade dada pelo vôo, e seus hábitos noturnos permitiram a estes animais a exploração de nichos relativamente sem competição durante sua evolução e diversificação, e parece ter sido responsável pela grande riqueza de espécies e diversidade de hábitos alimentares existentes nestes animais hoje em dia, que incluem: frugivoria, nectarivoria, carnivoria, piscivoria, insetivoria, onivoria, e, únicos entre os mamíferos, sanguivoria (hematofagia) (Ferarezzi & Gimenez, 1996; Nowak, 1997).

Estudos realizados comparando glândulas salivares em mamíferos sugeriram que a evolução destas estruturas está relacionada com a radiação adaptativa destes animais (Phillips & Tandler, 1996). Os mesmos autores já apontavam para diferenciações únicas em glândulas salivares de morcegos (Phillips *et al.*, 1987; Tandler *et al.*, 1996; Phillips *et al.*, 1998), e algumas destas diferenciações estão atribuídas mais a guildas alimentares do que a grupos filogeneticamente próximos, indicando convergência adaptativa ou paralelismo (Phillips *et al.*, 1987; Tandler *et al.*, 1996). Além disso, estudos de imuno-histoquímica e níveis de produção de lisozima já foram realizados em morcegos e chegaram a conclusões semelhantes (Phillips *et al.*, 1998). No entanto, nenhum destes estudos realizou uma investigação molecular de possíveis adaptações evolutivas nos genes de morcegos relacionados à digestão. Parece muito plausível supor que a diversidade de recursos alimentares explorados pelos morcegos pode estar relacionada a uma possível diversidade das enzimas digestivas destes animais.

No Brasil, apesar de muito estudados do ponto de vista ecológico, como dispersores de sementes, polinizadores e em outros processos (Sazima & Sazima, 1975; Uieda, 1992; Aguiar, 1994; Faria & Sazima, 1995; Tavares & Anciães, 1998; Tavares, 1999a e 1999b), são poucos os estudos genéticos e moleculares com morcegos. A maioria dos estudos são de genética populacional utilizando o DNA mitocondrial (Ditchfield & Burns, 1998; Ditchfield, 2000) e citogenéticos (Silva, *et al.*, 1997 e 1998; Noronha *et al.*, 1998).

Até o momento não existem trabalhos publicados de evolução molecular relacionados à busca por seleção natural e adaptação ao nicho alimentar em genes de morcegos.

# Objetivos

Visto a inexistência de estudos de seleção e adaptação molecular em morcegos este estudo tem como objetivo principal:

- Verificar a existência de seleção natural em um segmento do gene da  $\alpha$ -amilase de morcegos em relação aos seus hábitos alimentares.

Como objetivos específicos pretende-se:

- Desenvolver protocolos de amplificação por PCR e seqüenciamento de um segmento do gene da  $\alpha$ -amilase em morcegos e outros mamíferos;
- Reconstruir a filogenia molecular de um segmento do gene da  $\alpha$ -amilase, verificando se existem convergências evolutivas relacionadas à evolução de hábitos alimentares de quirópteros;
- Verificar se este segmento sofre influência de seleção natural e de que forma esta seleção estaria agindo;
- Verificar a existência de modificações adaptativas neste segmento do gene;
- Verificar se este segmento pode ser útil para o estudo da filogenia das ordens de mamíferos, com ênfase na posição dos grupos Chiroptera, Primata e Rodentia.

# **Materiais e Métodos**

## **Obtenção das Amostras**

As amostras utilizadas neste trabalho foram provenientes principalmente da captura de animais em campo, empréstimos de animais pertencentes a Coleções Taxonômicas institucionais e coleta de animais em cativeiro. Estas amostras estão listadas na Tabela 1.

As coletas em campo foram realizadas no Parque Estadual do Rio Doce (município de Marliéria, MG), Estação de Desenvolvimento Ambiental de Peti (Santa Bárbara, MG) e Estação Ecológica da Universidade Federal de Minas Gerais (Belo Horizonte, MG) e foram feitas com o objetivo de capturar morcegos de diferentes espécies e distintos hábitos alimentares.

A captura é feita com o auxílio de Redes de Neblina de 2 x 6 m e 3 x 12 m. Os animais de interesse capturados foram sacrificados em campo utilizando câmara mortífera com éter, já descrita em literatura (Tavares, 1999). Quando possível, os animais foram congelados para o transporte ao laboratório, alternativamente, um fragmento de tecido (fígado, baço e/ou músculo) era retirado do animal e colocado em álcool 95%.

Outros mamíferos silvestres foram gentilmente cedidos pelo Laboratório de Mastozoologia do Departamento de Zoologia do ICB-UFMG, pelo Laboratório de Citogenética do Departamento de Biologia Geral do ICB-UFMG, pelo Departamento de Zoologia da UFAL ou se encontravam depositados no Banco de DNA do Laboratório de Biodiversidade e Evolução Molecular do Departamento de Biologia Geral do ICB-UFMG (Santos *et al.*, 2002; [www.icb.ufmg.br/~lbem](http://www.icb.ufmg.br/~lbem) ; Tabela 1).

As amostras de animais domésticos utilizadas neste estudo foram coletadas em cativeiro na Escola de Veterinária da UFMG. Doações de morcegos capturados acidentalmente na região metropolitana de Belo Horizonte também foram utilizadas.

Todos os animais silvestres capturados neste estudo, e doações não institucionais, serão doados à Coleção de Mastozoologia do Departamento de Zoologia do ICB-UFMG, após a conclusão do estudo.

**Tabela 1:** Lista das espécies inicialmente utilizadas neste estudo e local de procedência

| Espécie                        | procedência                    | hábito alimentar     |
|--------------------------------|--------------------------------|----------------------|
| <i>Artibeus lituratus</i>      | PERD (3); EEUFMG (2); EDAP (2) | Frugívoro (onívoro)  |
| <i>Carollia perspicillata</i>  | PERD (2); EDAP (2)             | Frugívoro            |
| <i>Chiroderma villosum</i>     | PERD (2)                       | Frugívoro            |
| <i>Chrotopterus auritus</i>    | EDAP (1)                       | Carnívoro            |
| <i>Desmodus rotundus</i>       | PERD (4); EDAP (3)             | Sanguívoro           |
| <i>Diaemus youngi</i>          | Lmasto (1)                     | Sanguívoro           |
| <i>Eptesicus sp.</i>           | Lbem (1)                       | Insetívoro           |
| <i>Molossus molossus</i>       | BH (2)                         | Insetívoro           |
| <i>Myotis nigricans</i>        | PERD (1); EEUFMG (1); EDAP (1) | Insetívoro           |
| <i>Noctilio leporinus</i>      | UFAL (1)                       | Piscívoro            |
| <i>Phyllostomus discolor</i>   | EEUFMG (3)                     | Insetívoro (onívoro) |
| <i>Phyllostomus hastatus</i>   | Lbem (1)                       | Carnívoro (onívoro)  |
| <i>Platyrrhinus lineatus</i>   | PERD (2); EEUFMG (4); EDAP (1) | Frugívoro            |
| <i>Pygoderma bilabiatum</i>    | EDAP (2)                       | Frugívoro            |
| <i>Sturnira lilium</i>         | PERD (2); EDAP (1)             | Frugívoro            |
| <i>Vampyressa pusilla</i>      | EDAP (2)                       | Frugívoro            |
| <i>Bos taurus</i>              | VetUFMG (1)                    | Herbívoro            |
| <i>Equus caballus</i>          | VetUFMG (1)                    | Herbívoro            |
| <i>Canis familiaris</i>        | VetUFMG (1)                    | Carnívoro (onívoro)  |
| <i>Pan troglodytes</i>         | Lbem (1)                       | Onívoro              |
| <i>Homo sapiens</i>            | Lbem (1)                       | Onívoro              |
| <i>Bradypus variegatus</i>     | Lcito (1)                      | Herbívoro            |
| <i>Cerdocyon thous</i>         | Lmasto (1)                     | Carnívoro (onívoro)  |
| <i>Conepatus semistriatus</i>  | Lbem (1)                       | Carnívoro (onívoro)  |
| <i>Dugong dugon</i>            | Lbem (2)                       | Herbívoro            |
| <i>Eira barbara</i>            | Lbem (1)                       | Piscívoro/ Carnívoro |
| <i>Sylvilagus brasiliensis</i> | Lmasto (1)                     | Herbívoro            |
| <i>Oryzomys sp.</i>            | Lmasto (1)                     | Onívoro              |
| <i>Trinomys dimidiatus</i>     | Lmasto (1)                     | Onívoro              |

*legendas*

EDPAP: Estação de Desenvolvimento Ambiental de Peti; EEUFMG: Estação Ecológica da UFMG; PERD: Parque Estadual do Rio Doce; BH: Região Metropolitana de Belo Horizonte; Lbem: Laboratório de Biodiversidade e Evolução Molecular ICB-UFMG; Lcito: Laboratório de Citogenética ICB-UFMG; Lmasto: Laboratório de Mastozoologia ICB-UFMG; VetUFMG: Escola de Veterinária da UFMG; UFAL: Departamento de Zoologia da UFAL.

## Preparação das Amostras e Extração de DNA

No laboratório, os animais capturados, ou cedidos, foram dissecados, sendo removidos o fígado, baço e/ ou um fragmento da musculatura, dependendo do estado de conservação do animal. O tecido foi recortado em pequenos pedaços e estocado em álcool 95-100% geralmente em *freezer* a  $-20^{\circ}\text{C}$  até o momento da extração de DNA.

As extrações de DNA foram realizadas pelo método de Fenol/Clorofórmio (Sambrook *et al.*, 2001) com modificações, ou com a utilização de Kits de extração DNeasy® (Qiagen) utilizando as recomendações do fabricante.

O DNA extraído foi visualizado em gel de agarose 0,8% (Sambrook *et al.*, 2001), principalmente para checar a sua integridade e pureza, assim como estimar sua quantidade.

Uma parte do DNA extraído foi aliquoteado para a utilização neste projeto, e o restante foi depositado no Banco de DNA do Laboratório de Biodiversidade e Evolução Molecular, ICB-UFMG (Santos *et al.*, 2002; [www.icb.ufmg.br/~lbem/ddb](http://www.icb.ufmg.br/~lbem/ddb)).

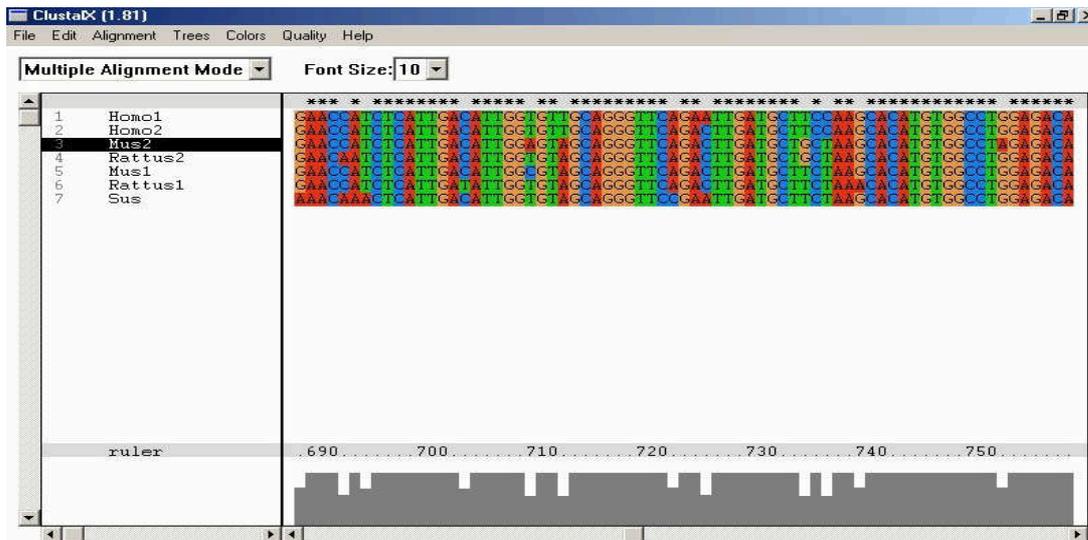
## **Amplificação dos segmentos do gene da $\alpha$ -amilase**

### **Desenho dos iniciadores**

Os iniciadores utilizados foram desenhados com base em seqüências do gene da  $\alpha$ -amilase1 existentes no GenBank do NCBI ([www.ncbi.nih.gov](http://www.ncbi.nih.gov)).

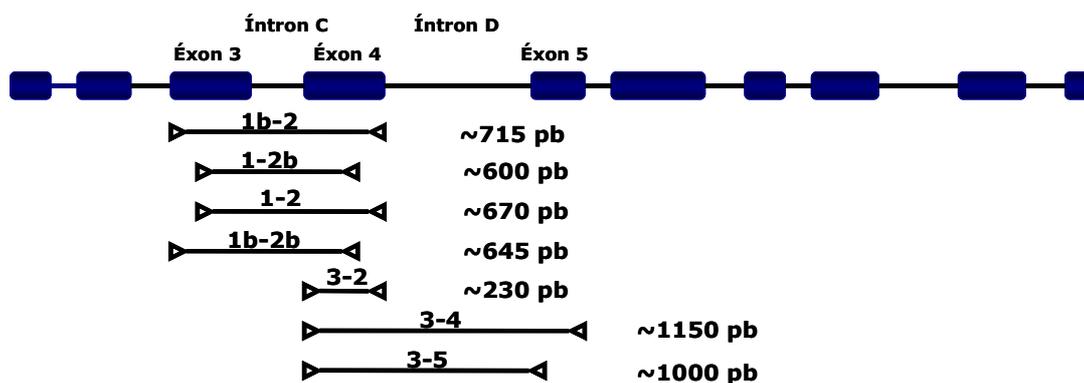
As seqüências de cDNA de *Mus musculus* (camundongo), *Rattus norvegicus* (rato comum), *Homo sapiens* (homem) e *Sus scrofa* (porco doméstico) foram alinhadas no programa ClustalX 1.81 (Thompson *et al.*, 1994 e 1997; Jeanmougin *et al.*, 1998), utilizando os parâmetros *default* do programa (Figura 3). Este alinhamento permitiu a inferência de uma seqüência consenso com o programa SeqPup 9 (Gilbert, 1997). Todas as ambigüidades foram solucionadas com base na seqüência humana, visto que os primatas estão filogeneticamente mais próximos dos morcegos, do que os outros animais apresentados anteriormente (Novacek, 1992).

Esta seqüência foi depois alinhada manualmente no programa SeqPup 9 contra a seqüência total do gene da proteína humana (única disponível nos bancos de dados até aquela data), para a localização dos éxons e íntrons.



**Figura 3:** Alinhamento das seqüências de cDNA de  $\alpha$ -amilase encontrados no GenBank do NCBI.

Com base na localização dos sítios ativo e de ligação de cálcio, os iniciadores foram desenhados para amplificarem principalmente a região do éxon 4, que contém dois resíduos catalíticos (Asp, Glu) que fazem parte do sítio ativo da enzima e dois sítios de ligação de cálcio (Asp, His, necessários para a ativação da enzima). Os iniciadores desenhados também devem amplificar o Éxon 3, Íntron C, Íntron D, e o Éxon 5 (Figura 4).



**Figura 4:** Desenho esquemático do gene da  $\alpha$ -amilase, mostrando a posição dos iniciadores desenhados e utilizados neste trabalho, com o tamanho esperado dos segmentos.

Os iniciadores (Tabela 2) foram testados em simulações de computador nos programas: Amplify 1.2 para MacIntosh (Engel, 1992) e Oligos 8.72 para Windows (Kalendar, 2001) utilizando a seqüência completa do gene da  $\alpha$ -Amilase 1 humana.

Os fragmentos resultantes gerados por diferentes combinações dos iniciadores foram entre 250 e 1150 pb nas simulações (Figura 4).

## Amplificação pela PCR

Após diversas padronizações, as reações de amplificação pela PCR foram feitas em um volume final de 15  $\mu$ L, contendo aproximadamente 40-80 ng de DNA genômico, 1X tampão da reação 1B (Phoneutria® - 1,5 mM de  $MgCl_2$ , 10 mM Tris-HCl (pH 8,4), 50 mM KCl, 0,1% Triton X-100), 200  $\mu$ M de dNTPs (Amersham-Biosciences®), 0,25-0,6  $\mu$ M de iniciadores (Tabela 2) e 1,25 U de Taq DNA polimerase (Phoneutria®).

**Tabela 2:** Pares de iniciadores utilizados nas reações de PCR, com a concentração utilizada nas reações, temperatura de anelamento e tamanho do segmento.

| Pares de Iniciadores 5' → 3'                  | Concentração na PCR | Temperatura de anelamento | Tamanho  |
|---|---------------------|---------------------------|----------|
| <b>BatAmy1</b><br>CGAGAACTATAATGATGCTACTCAGG  | 0,3 $\mu$ M         | 45-55°C                   | ~670* pb |
| <b>BatAmy2</b><br>CTCCTGGTAAATGAAAGGTTTACTACC |                     |                           |          |
| <b>BatAmy1b</b><br>GTTTCGTATTTATGTGGATGCTGTAA | **                  | **                        | **       |
| <b>BatAmy2b</b><br>GACCAGACAGACGACAATCTCTGAC  |                     |                           |          |
| <b>BatAmy1</b><br>CGAGAACTATAATGATGCTACTCAGG  | **                  | **                        | **       |
| <b>BatAmy2b</b><br>GACCAGACAGACGACAATCTCTGAC  |                     |                           |          |
| <b>BatAmy1b</b><br>GTTTCGTATTTATGTGGATGCTGTAA | 0,25 $\mu$ M        | 47-55°C                   | ~ 715*pb |
| <b>BatAmy2</b><br>CTCCTGGTAAATGAAAGGTTTACTACC |                     |                           |          |
| <b>BatAmy3</b><br>GTCAGAGATTGTCGTCTGTCTGGTC   | 0,25 $\mu$ M        | 50-53°C                   | ~250 pb  |
| <b>BatAmy2</b><br>CTCCTGGTAAATGAAAGGTTTACTACC |                     |                           |          |
| <b>BatAmy3</b><br>GTCAGAGATTGTCGTCTGTCTGGTC   | 0,5 $\mu$ M         | 55°C                      | ~1150 pb |
| <b>BatAmy4</b><br>AAGTAAGACATCTTCTCTCCATTCC   |                     |                           |          |
| <b>BatAmy3</b><br>GTCAGAGATTGTCGTCTGTCTGGTC   | 0,6 $\mu$ M         | 60° C                     | ~950 pb  |
| <b>BatAmy5</b><br>GGCTCACCACCCAGATCAATTAC     |                     |                           |          |

\* tamanho variável em algumas espécies

\*\* iniciadores que não funcionaram adequadamente

Os produtos da PCR foram visualizados em gel de agarose 0,8% como descrito anteriormente.

## **Purificação dos produtos da PCR**

Antes do seqüenciamento, os produtos da PCR foram purificados para ser eliminado o excesso de iniciadores e dNTPs. Este procedimento é realizado com as enzimas Exonuclease I (EXO I) e SAP (Shrimp Alkaline Phosphatase).

A EXO I é uma exonuclease que quebra ácidos nucléicos de fita simples e visa eliminar os iniciadores ainda presentes na PCR, e a SAP é uma fosfatase alcalina que cliva os fosfatos dos dNTPs.

Uma mistura de 1 U de SAP e 10 U de EXO I é colocada na proporção de 1:1 com o produto da PCR e a reação é incubada a 37°C durante 45'. Existe a necessidade de que as enzimas sejam inativadas antes da reação de seqüenciamento, incubando-as a 80°C durante 15'.

## **Seqüenciamento dos fragmentos gerados pela PCR**

O seqüenciamento dos segmentos amplificados foi realizado no equipamento a *laser* MegaBACE 1000® da Amersham-Biosciences, utilizando *kits* para seqüenciamento DYEnamic ET® Terminator da mesma companhia e protocolo sugerido pelo fabricante.

Os iniciadores utilizados no seqüenciamento foram os mesmos utilizados na PCR e todas as amostras foram seqüenciadas no sentido direto e reverso usando apenas um dos respectivos iniciadores por reação.

## **Montagem e validação das seqüências**

A qualidade das seqüências foi verificada inicialmente no programa Sequence Analyzer 2.4 que acompanha o seqüenciador de DNA e posteriormente utilizamos o programa Phred v.0.20425 (Ewing *et al.*, 1998; Ewing & Green, 1998), utilizando o parâmetro phred=20 (probabilidade de 1 possível erro em 100) para a validação das bases.

Os fragmentos seqüenciados foram montados e alinhados no programa Phrap v.0.990319 (Green, 1994), e visualizados e editados no programa Consed 12.0 (Gordon *et al.*, 1998). Todas as seqüências pertencentes a indivíduos da mesma espécie foram colocadas juntas para a determinação de uma seqüência consenso de alta qualidade para a espécie.

A determinação da presença de possíveis variantes alélicas foi feita no programa Polyphred 4.0 (Nickerson *et al.*, 1997) e visualizadas no Consed 12.0.

# Análises dos dados

## Conjuntos de Dados

As seqüências de  $\alpha$ -amilase obtidas para os morcegos e para os outros animais foram somadas a outras seqüências encontradas nos bancos de dados do GenBank do NCBI (Tabela 3).

**Tabela 3:** Sequências obtidas nos bancos de dados do GenBank do NCBI com número de acesso.

| <b>Espécie</b>                | <b>Molécula (nº de acesso)</b>   |
|-------------------------------|--|
| <i>Homo sapiens</i>           | Amy1A(gi537511); Amy2B(gi20070311); Contig genômico do cromossomo 1 (gi27477789)             |
| <i>Sus scrofa</i>             | Amy2 (gi6056337)   |
| <i>Mus musculus</i>           | Amy1 (gi6996908); Amy2 (gi6753051); Contig genômico do cromossomo 3 (gi20875733; gi20971184) |
| <i>Rattus norvegicus</i>      | Amy1 (gi13928683); Amy2 (gi11528628); Contig genômico do cromossomo 2 (gi26007434)           |
| <i>Gallus gallus</i>          | Amy2 (gi1498467)   |
| <i>Xenopus laevis</i>         | Amy 2 (gi19387129)   |
| <i>Tetraodon nigroviridis</i> | Amy (gi20502193)   |
| <i>Anguila japonica</i>       | Amy (gi1903229)  |

*Legendas:* Amy: mRNA de  $\alpha$ -amilase; Amy1: mRNA de  $\alpha$ -amilase de glândula salivar; Amy2: mRNA de  $\alpha$ -amilase de pâncreas;

A localização dos éxons e íntrons foi feita com o auxílio dos programas Blast (Altschul *et al.*, 1997) e Blast 2 sequences (Tatusova & Madden, 1999) e a utilização de ferramentas de bioinformática encontradas na Página do Núcleo de Bioinformática da UFMG ([www.icb.ufmg.br/~infobio](http://www.icb.ufmg.br/~infobio)).

Como a maioria das seqüências presentes no GenBank provém de cDNA, e nem todas as seqüências geradas foram de excelente qualidade (Ver resultados) não foi possível utilizar comparações de íntrons e éxons para todas as espécies, assim, dois conjuntos de dados foram analisados:

1. Conjunto **PMR**, formado pelas seqüências do éxon 4 e íntron C de primatas, morcegos e roedores seqüenciados e obtidos no GenBank. *Vampyressa pussilla* só foi utilizada nas análises de éxon (ver resultados). Este conjunto de dados foi utilizado nas análises filogenéticas e nos testes de seleção e adaptação moleculares.

2. Conjunto **VERT**, formado pelas seqüências do éxon 4 de todos os animais seqüenciados e obtidos no GenBank. Este conjunto de dados foi utilizado apenas nas análises filogenéticas.

Os conjuntos de dados tiveram sua seqüência traduzida em aminoácidos, utilizando o programa "on-line" *Sequence Manipulation Suit* disponível no Núcleo de Bioinformática da UFMG ([www.icb.ufmg.br/~infobio](http://www.icb.ufmg.br/~infobio)).

Seqüências adicionais de Cyt-b (Citocromo-b oxidase) mitocondrial e de RAG-2 (*Recombination activator protein G-2*) nuclear também foram baixadas dos bancos de dados para serem utilizados em comparações com o conjunto PMR (Tabela 4).

**Tabela 4:** Seqüências de Citocromo B e RAG-2 baixadas do GenBank e utilizadas neste estudo

| <b>Espécie</b>                 | <b>Molécula (nº de acesso)</b>       |
|--------------------------------|--------------------------------------|
| <i>Homo sapiens</i>            | Cyt-b (gi13273284); RAG-2 (20557254) |
| <i>Artibeus lituratus</i>      | Cyt-b (gi1513279)                    |
| <i>Artibeus jamaicensis</i> *  | RAG-2 (gi12700342)                   |
| <i>Chiroderma vilosum</i>      | RAG-2 (gi12247574)                   |
| <i>Chiroderma trinitatum</i> * | Cyt-b (gi304317)                     |
| <i>Platirrhinus helleri</i> *  | Cyt-b (gi309897); RAG-2 (gi12247658) |
| <i>Sturnira lilium</i>         | Cyt-b (gi310655); RAG-2 (gi12247672) |
| <i>Phyllostomus hastatus</i>   | RAG-2 (gi12247654)                   |
| <i>Phyllostomus discolor</i>   | Cyt-b (Lbem, ainda não depositado)   |
| <i>Myotis nigricans</i>        | Cyt-b (gi14248972)                   |
| <i>Molossus molossus</i>       | Cyt-b (gi309031)                     |

\* seqüências congênicas utilizadas na ausência das específicas.

### **Alinhamento das seqüências**

As seqüências foram alinhadas com o programa ClustalX 1.81 (Thompson *et al.* 1994, 1997), utilizando o padrão *default* do programa, com uma penalidade para a abertura de *gaps* de 15 pontos no *Score* final, e penalidade de 0,66 para prolongamento do *gap*.

### **Reconstruções Filogenéticas**

Os conjuntos de dados foram analisados com os métodos de Máxima Parcimônia e Agrupamentos Vizinhos. O éxon 4 ainda foi analisado de três formas: com todas as posições de nucleotídeos, apenas as duas primeiras posições de nucleotídeos e através da seqüência deduzida de aminoácidos.

### *Máxima Parcimônia - MP*

As reconstruções por máxima parcimônia foram feitas para todos os conjuntos de dados, utilizando o programa PAUP\* 4.08b (Swofford, 1998) e o programa MEGA 2.1 (Kumar, *et al.*, 2001). A busca pela árvore mais parcimoniosa foi feita utilizando busca heurística com o algoritmo TBR (*Tree Bisection Reconnection* no PAUP) e CNI (*Closest Neighbor Interexchange* no MEGA). Um teste estatístico de consistência da topologia foi feito por *bootstrap* com 1.000 replicações e 10 turnos de adições aleatórias de seqüências.

### *Agrupamentos Vizinhos (Neighbor-Joining) - NJ*

As reconstruções por agrupamentos vizinhos utilizaram o modelo de Tamura-Nei 1993, também com teste de topologia de 1.000 replicações de *bootstrap* no programa MEGA 2.1 (Kumar, *et al.*, 2001).

A distância, no caso dos aminoácidos (aa) foi medida com a distância  $p$  de proporção de aa compartilhados (Nei & Kumar, 2000).

## **Características dos conjuntos de dados**

Características estatísticas dos conjuntos de dados foram inferidas utilizando o programa MEGA 2.1 (Kumar *et al.*, 2001).

Razão das transições sobre transversões (R), calculada utilizando o modelo de Tamura-Nei (Tamura & Nei, 1993);

Distância (d) de divergência entre as seqüências calculada com o número de substituições por sítio utilizando o número de transições mais transversões com o modelo de Tamura-Nei (Tamura & Nei, 1993);

Sítios degenerados: não degenerado (0X), é aquele em que qualquer mudança é uma mutação não-sinônima; duas vezes degenerado (2X) é aquele em que uma (ou duas) das três possíveis mudanças são sinônimas e quatro vezes degenerado (4X) é aquele em que qualquer mudança é sinônima;

Sítios sinônimos (**S**) e não-sinônimos (**N**) calculados utilizando a versão modificada do método de Nei-Gojobori (Zhang *et al.*, 1998).

## **Detecção de seleção natural e adaptações moleculares**

A detecção de possíveis episódios de adaptação molecular nos genes de  $\alpha$ -amilase, foi testado por três metodologias:

- 1** Teste de distância par a par de Nei-Gojobori (1986) e sua versão modificada (Zhang *et al.*, 1998).

O conjunto de dados PMR foi analisado de duas formas:  $d_N$  e  $d_S$  calculados par a par entre todas as seqüências e utilizando a média de  $d_N$  e  $d_S$  dos grupos (primatas, morcegos e roedores), em ambos os casos um teste Z foi realizado de modo que:

$$Z = \frac{d_N - d_S}{\sqrt{\text{Var}(d_S) + \text{Var}(d_N)}}$$

Onde  $\text{Var}(d_N)$  e  $\text{Var}(d_S)$  são as variâncias de  $d_N$  e  $d_S$ , calculadas pelo método de *bootstrap* com 1000 replicações.

O valor de Z pode ser testado para três hipóteses alternativas:

$$\begin{array}{lll} H_0: d_N = d_S & H_0: d_N = d_S & H_0: d_N = d_S \\ H_1: d_N \neq d_S & H_1: d_N > d_S & H_1: d_N < d_S \end{array}$$

Sendo cada uma destas um teste de neutralidade, um de seleção positiva e um de seleção purificadora respectivamente.

Além disso, foi realizado um teste exato de Fisher como no quadro de contingência abaixo:

|                     | não-sinônimas | sinônimas | Soma          |
|---------------------|---------------|-----------|---------------|
| substituições       | n             | s         | n+s           |
| nº de substituições | N-n           | S-s       | (N-n) + (S-s) |
| Totais              | N             | S         | N+S           |

Onde o número de sítios sinônimos (**S**) e não-sinônimos (**N**) e o número de substituições sinônimas (**s**) e não-sinônimas (**n**) foram estimados utilizando o teste modificado de Nei-Gojobori (Zhang *et al.*, 1998) em cada par de seqüências ancestrais reconstruídas para os ramos 1, 2, 3, 4, 5 e P na árvore filogenética da Figura 6.

As seqüências ancestrais foram inferidas pelo método Bayesiano (Yang *et al.*, 1995) utilizando o programa PAML (Yang, 1997).

- 2** Teste CM2002 de Creevey e McInerney (2002), utilizando o programa Crann (Creevey, 2002).

Foram utilizados os parâmetros *default* do programa e dois testes, um utilizando o Teste CM2002 propriamente dito que é um teste de razão de taxas relativas ( $r^3t$ ) para detecção de seleção, e um outro teste para detecção de seleção purificadora.

A árvore filogenética gerada pelo programa tem topologia idêntica à árvore da Figura 6.

**3** Testes de LRT desenvolvidos para modelos de evolução de códons (Goldman & Yang, 1994; Yang, 1998; Yang & Nielsen, 1998; Yang *et al.*, 2000; Yang & Nielsen, 2002).

Uma série de LRTs foram utilizados para testar diversos modelos de substituição de códons, onde a verossimilhança de cada modelo pode ser testada com:

$$2x [\text{Log}(\ell_{H_0}) - \text{Log}(\ell_{H_1})]$$

Onde  $\ell_{H_0}$  e  $\ell_{H_1}$  são a verossimilhança da hipótese nula e alternativa respectivamente. O valor obtido é verificado numa distribuição de  $\chi^2$  com graus de liberdade igual à diferença entre o número de parâmetros no modelo.

Os primeiros LRTs foram testes de ramos para se detectar a presença de seleção em linhagens específicas. As linhagens utilizadas foram as mesmas utilizadas do teste de Nei-Gojobori (Figura 6), e foram utilizados dois modelos: R0, R1 e R2.

R0 utiliza um único  $\omega$  ( $d_N/d_S$ ) para todas as linhagens e serve como hipótese nula, e R1 utiliza o parâmetro  $\omega$  livre para todos os ramos da árvore, de modo que  $\omega$  possa ser estimado em cada linhagem a partir dos dados. O LRT neste caso utiliza um grau de liberdade.

Caso o modelo R1 seja mais adequado do que o R0, isso apenas significa que  $\omega$  é variável, mas não é indício de seleção então o modelo R2 é testado contra R0. R2 fixa  $\omega_1$  para uma linhagem e estima  $\omega_0$  para as outras, no nosso caso  $\omega_1=1$  (R2 $\omega_1$ ) para se testar seleção. Este teste utiliza um grau de liberdade e deve ser testado fixando  $\omega_1$  para cada linhagem de interesse.

A segunda série de LRTs são testes para verificar se existe diferença na taxa de  $d_N$  e  $d_S$  ( $\omega$ ) para diferentes sítios nas seqüências. Três LRTs foram utilizados para isso.

O primeiro modelo M0 utiliza  $\omega$  constante para todos os sítios e é o modelo nulo para comparar com M3. M3 divide os sítios em classes discretas (ex.  $0 < \omega < 0,25$ ;  $0,25 < \omega < 0,5$ , etc), com a proporção ( $p$ ) de sítios pertencentes a estas classes estimada a partir dos dados. M0 e M3 podem ser comparados utilizando 4 graus de liberdade.

O teste acima também é apenas um indício de que existam sítios com  $\omega$  variáveis e pode apontar sítios com  $\omega$  elevado, mas o teste de seleção é utilizado com os modelos M1 e M2. Onde M1 é um modelo de seleção neutra, com duas classes fixas de  $\omega$ , uma com  $\omega=1$  e outra com  $\omega=0$ , e as proporções de cada uma estimada dos dados, e é hipótese nula para M2 que utiliza uma terceira classe com  $\omega$  estimado dos dados, para se verificar se existem sítios com  $\omega>1$ , portanto alvo de seleção. Este LRT é feito com dois graus de liberdade.

Como o teste acima é muito rigoroso, um novo modelo pode ser utilizado, com distribuição de  $\omega$  mais homogênea. Este teste utiliza os modelos M7 $\beta$  e M8 $\beta\omega$ . A hipótese nula neste caso é M7 $\beta$ , que é um modelo discreto onde as diferentes classes de  $\omega$  utilizam uma distribuição  $\beta$ , que varia entre 0 e 1. M8 $\beta\omega$  coloca mais uma classe de  $\omega$  variável, estimada a partir dos dados. LRT com dois graus de liberdade.

Em todos os modelos de sítios, aqueles com maior probabilidade de estarem sujeitos a seleção são inferidos pelo teorema de Bayes (Yang, 1997). Os últimos LRTs são modelos de ramo e sítio, onde é possível verificar se existe seleção em sítios de linhagens específicas.

Dois LRTs são utilizados nestes testes. O primeiro utiliza como hipótese nula o modelo M1, e a alternativa é o modelo de ramo e sítio MA. MA estima  $\omega_1=1$  para a linhagem de interesse e  $\omega_0=0$  para as outras linhagens e só permite variação nos  $\omega$  de sítios para a linhagem de interesse. Este LRT é realizado com dois graus de liberdade.

O segundo utiliza M3 como hipótese nula e MB como alternativa. MB é semelhante ao modelo MA pois estima diferentes  $\omega$  para a linhagem de interesse e as outras, porém a variação de  $\omega$  nos sítios da linhagem de interesse é discreta (como no modelo M3). Também utiliza dois graus de liberdade no LRT.

# Resultados e Discussão

## Amplificação por PCR

A partir dos alinhamentos de seqüências depositadas no GenBank foram desenhados e sintetizados iniciadores para distintas regiões do gene da  $\alpha$ -amilase (Tabela 2). A maioria dos pares de iniciadores funcionaram bem, gerando fragmentos únicos com o tamanho esperado pelas simulações (Tabela 2; Figura 4). Alguns pares de iniciadores não produziram um produto de PCR com a qualidade esperada. Os conjuntos BatAmy1-2b e BatAmy1b-2b geravam sempre múltiplos fragmentos de peso molecular variável, por isso foram rapidamente descartados.

No entanto, não foi possível gerar fragmentos para todas as espécies com todos os pares de iniciadores (Tabela 5). Em algumas espécies nenhum dos iniciadores funcionaram. Vários motivos podem ter sido a causa dessa observação, por exemplo, os morcegos *Eptesicus sp.*, *Diaemus young* e *Noctilio leporinus* foram animais vindos de coleções taxonômicas e estavam conservados em formol. Sabe-se que formol pode criar compostos oxidantes que reduzem a eficiência das reações de PCR (Sambrook *et al.*, 2001) além de degradar bastante o DNA genômico. Nos animais citados acima, as extrações de DNA resultaram na obtenção de um DNA bastante degradado.

Em *Desmodus rotundus* e *Chrotopterus auritus* não houve sinal de amplificação com nenhum dos pares em que um dos iniciadores se acoplam na região do éxon 3 (BatAmy1 e BatAmy 1b), semelhantemente, em *Carollia perspicillata* não ocorreu amplificação com estes mesmos iniciadores e nem com o BatAmy2 que se liga na porção 3' do éxon 4 e em *Myotis nigricans* os iniciadores que se ligam na região do éxon 5 também não funcionaram (BatAmy4 e BatAmy5; Tabela 5, Figura 4). Isso pode indicar ou que mutações existam nos sítios de ligação dos iniciadores indicando taxas de evolução diferentes para estes dois fragmentos ou que o íntron entre estes iniciadores seja muito grande impedindo a elongação da fita dupla de DNA na PCR. Mutações nos sítios parecem ser prováveis, mas seriam necessárias várias mutações em pelo menos duas regiões de cada éxon. A questão do tamanho do íntron parece mais plausível uma vez que a variação no tamanho do segmento foi observada em *Mollossus mollossus* onde os pares de iniciadores BatAmy1-2 e BatAmy1b-2 geraram fragmentos muito maiores do que o esperado (cerca de 1200 pb; Tabela 2), estes mesmos iniciadores também geraram fragmentos maiores do que o esperado nos carnívoros *Conepatus semistriatus* e *Eira barbara* (cerca de 900 pb), mostrando uma variação no tamanho do íntron C (Figura 4).

**Tabela 5:** Resultados das ampliações dos 5 pares de iniciadores selecionados em todas as espécies estudadas

| Espécie                        | Família<br>(subfamília) <sup>#</sup><br>Apenas nos morcegos | Pares de iniciadores BatAmy |      |     |     |     |
|--------------------------------|---|-----------------------------|------|-----|-----|-----|
|                                |   | 1-2                         | 1b-2 | 3-2 | 3-4 | 3-5 |
| <i>Artibeus lituratus</i>      | Phyllostomidae (S)  | X                           | X    | X   | X   | X   |
| <i>Carollia perspicillata</i>  | Phyllostomidae (C)  | *                           | *    | *   | X   | X   |
| <i>Chiroderma villosum</i>     | Phyllostomidae (S)  | X                           | X    | X   | X   | X   |
| <i>Chrotopterus auritus</i>    | Phyllostomidae (P)  | *                           | *    | X   | X   | X   |
| <i>Desmodus rotundus</i>       | Phyllostomidae (D)  | *                           | *    | X   | X   | X   |
| <i>Diaemus youngi</i>          | Phyllostomidae (D)  | *                           | *    | *   | *   | *   |
| <i>Eptesicus sp.</i>           | Vespertilionidae  | *                           | *    | *   | *   | *   |
| <i>Molossus molossus</i>       | Molossidae  | X**                         | X**  | X   | X   | X   |
| <i>Myotis nigricans</i>        | Vespertilionidae  | X                           | X    | X   | *   | *   |
| <i>Noctilio leporinus</i>      | Noctilionidae   | *                           | *    | *   | *   | *   |
| <i>Phyllostomus discolor</i>   | Phyllostomidae (P)  | X                           | X    | X   | X   | X   |
| <i>Phyllostomus hastatus</i>   | Phyllostomidae (P)  | X                           | X    | X   | X   | X   |
| <i>Platyrrhinus lineatus</i>   | Phyllostomidae (S)  | X                           | X    | X   | X   | X   |
| <i>Pygoderma bilabiatum</i>    | Phyllostomidae (S)  | X                           | X    | X   | X   | X   |
| <i>Sturnira lilium</i>         | Phyllostomidae (S)  | X                           | X    | X   | X   | X   |
| <i>Vampyressa pussila</i>      | Phyllostomidae (S)  | X                           | X    | X   | X   | X   |
| <hr/>                          |   |                             |      |     |     |     |
| <i>Bos taurus</i>              |   | X                           | X    | X   | X   | X   |
| <i>Equus caballus</i>          |   | X                           | X    | X   | X   | X   |
| <i>Canis familiaris</i>        |   | X                           | X    | X   | X   | X   |
| <i>Pan troglodytes</i>         |   | X                           | X    | X   | X   | X   |
| <i>Homo sapiens</i>            |   | X                           | X    | X   | ?   | ?   |
| <hr/>                          |   |                             |      |     |     |     |
| <i>Cerdocyon thous</i>         |   | X                           | X    | X   | X   | X   |
| <i>Conepatus semistriatus</i>  |   | X**                         | X**  | X   | X   | X   |
| <i>Eira barbara</i>            |   | X**                         | X**  | X   | X   | X   |
| <i>Dugong dugon</i>            |   | X                           | X    | ?   | ?   | ?   |
| <i>Sylvilagus brasiliensis</i> |   | X                           | X    | X   | X   | X   |
| <i>Oryzomys sp.</i>            |   | X                           | X    | X   | X   | X   |
| <i>Trinomys dimidiatus</i>     |   | X                           | X    | X   | X   | X   |

**Legendas** X amplificação funcionou; X\*\* amplificação funcionou, porém com tamanho diferente do esperado; \* sem amplificação; ? não testado. <sup>#</sup> (S) Stenodermatinae, (C) Caroliinae, (D) Desmodontinae, (P) Phyllostominae.

## Obtenção e análise das seqüências de $\alpha$ -amilase

Todos os produtos de PCR obtidos a partir do gene  $\alpha$ -amilase de diferentes indivíduos foram submetidos ao seqüenciamento no equipamento MegaBACE que faz uma leitura a *laser* de fragmentos da reação de seqüenciamento, marcados por distintas fluorescências. Espera-se que quando mais de um segmento distinto de DNA é utilizado numa mesma reação de seqüenciamento no equipamento, o sinal fica confuso e extremamente difícil de ser lido (Manual do Megabace 1000, Amersham-Biosciences). Quando são utilizados dois segmentos diferentes, porém de tamanho e composição de bases muito semelhantes, existe a possibilidade de que a seqüência seja gerada com uma qualidade elevada. Isto é o caso dos estudos com seqüências parálogas sob evolução em concerto, tais como os genes de RNA16S, 12S, etc freqüentemente utilizados em filogenia (Li *et al.*, 1997). Neste caso só é possível verificar o co-seqüenciamento de seqüências distintas, que podem ser parálogos ou alelos de um mesmo gene, se há alguma variação entre os *loci*, que se revela através da análise direta do fluorograma. Neste caso, apenas substituições nucleotídicas podem ser detectadas pois pequenas deleções ou inserções iriam impedir a leitura correta pelo seqüenciador.

As seqüências foram conferidas nos cromatogramas para o caso da existência de dois segmentos distintos, possíveis parálogos ou formas alélicas do gene da  $\alpha$ -amilase estarem sendo seqüenciados. Não foram encontrados cromatogramas que revelassem seqüências diferentes em nenhum dos indivíduos analisados de distintas espécies (ex. Figura 5).

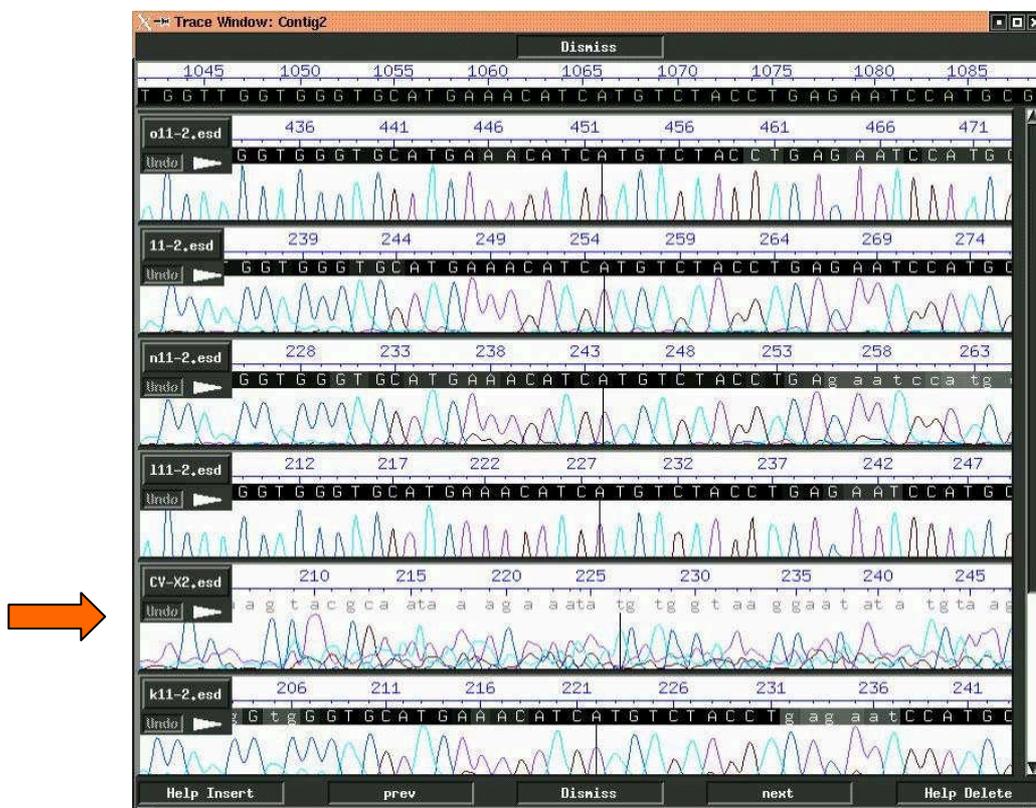
Foi utilizado também o programa Polyphred, que foi desenhado principalmente para verificar a presença de possíveis alelos entre seqüências de distintos clones, mas que também utilizamos no caso acima, onde não se encontrou variações.

Apesar de não se encontrar diferenças na análise dos cromatogramas, isso não descarta em 100% a possibilidade de, mesmo sem variação aparente, estarmos trabalhando com parálogos da  $\alpha$ -amilase. A existência de fenômenos como evolução em concerto pode dificultar muito esta análise e dependendo do nível de correção entre dois genes num evento de conversão gênica, seria impossível distinguir um gene de outro. Além disso pode existir amplificação preferencial do segmento de uma cópia ou outra.

Entretanto no caso de um nível dessa magnitude de convergência gênica vir a ocorrer, os resultados das reconstruções filogenéticas provavelmente não seriam alterados (Li, 1997; Gogarten & Olendzenski, 1999).

Inicialmente as seqüências foram selecionadas utilizando o critério de qualidade do programa Phred como >20 (probabilidade de 1 possível erro em 100 nucleotídeos) na análise do fluorograma. Algumas seqüências foram assim excluídas das análises e outras simplesmente não foram geradas a partir do produto de PCR de qualidade duvidosa. Com isso só foi possível analisar algumas das espécies, e mesmo assim apenas para alguns dos segmentos do gene da  $\alpha$ -amilase (Tabela 6).

A não obtenção de seqüências de boa qualidade pode ser também um indício da existência de múltiplas cópias da  $\alpha$ -amilase em alguns indivíduos, semelhante ao que acontece em humanos, onde algumas populações apresentam um maior número de cópias do que o normal (Groot *et al.*, 1989).



**Figura 5:** Cromatogramas das seqüências de  $\alpha$ -amilase de *Chiroderma doria* mostrados no programa Consed. A seta indica uma seqüência de baixa qualidade.

No entanto, as seqüências escolhidas para análise foram de excelente qualidade (Phred >20; Figura 5) e não houve diferença significativa para o seqüenciamento entre indivíduos da mesma espécie.

Foram analisadas apenas as espécies de morcegos em que se conseguiu seqüenciar a totalidade do éxon 4 (231 pb) e quase a totalidade do íntron C (432 pb). A única exceção foi *Vampiroessa pussila* que apresentou uma seqüência do éxon 4 de

excelente qualidade, mas apenas 153 pb confiáveis para o íntron C, por isso foi incluída no conjunto de dados PMR, apenas para as análises de éxon.

**Tabela 6:** Resultados do seqüenciamento nas espécies estudadas

| Espécie                        | Região do gene da $\alpha$ -amilase |         |       |         |       |
|--------------------------------|-------------------------------------|---------|-------|---------|-------|
|                                | éxon3                               | íntronC | éxon4 | íntronD | éxon5 |
| <i>Artibeus lituratus</i>      | P                                   | X       | X     |         | P     |
| <i>Carollia perspicillata</i>  |                                     |         | P     |         |       |
| <i>Chiroderma villosum</i>     |                                     | P       | X     |         | P     |
| <i>Chrotopterus auritus</i>    |                                     |         | P     |         | P     |
| <i>Desmodus rotundus</i>       |                                     |         | P     |         | P     |
| <i>Molossus molossus</i>       |                                     | P       | X     |         | P     |
| <i>Myotis nigricans</i>        | P                                   | X       | X     |         |       |
| <i>Phyllostomus discolor</i>   |                                     | P       | X     |         | P     |
| <i>Phyllostomus hastatus</i>   |                                     | P       | X     |         |       |
| <i>Platyrrhinus lineatus</i>   | P                                   | X       | X     | X       | P     |
| <i>Pygoderma bilabiatum</i>    |                                     |         | P     |         |       |
| <i>Sturnira lilium</i>         |                                     | P       | X     |         |       |
| <i>Vampyressa pussila</i>      |                                     | P**     | X     |         | P     |
| <i>Bos taurus</i>              |                                     | P       | X     | X       | P     |
| <i>Eqqus caballus</i>          |                                     | P       | X     |         | P     |
| <i>Canis familiaris</i>        |                                     | X       | X     |         | P     |
| <i>Pan troglodytes</i>         | P                                   | X       | X     |         | P     |
| <i>Homo sapiens</i>            | X                                   | X       | X     | ?       | ?     |
| <i>Bradypus variegatus</i>     |                                     |         |       |         |       |
| <i>Cerdocyon thous</i>         |                                     | X       | X     |         | P     |
| <i>Conepatus semistriatus</i>  |                                     | P       | X     |         |       |
| <i>Eira barbara</i>            |                                     | P       | X     |         |       |
| <i>Dugong dugon</i>            |                                     |         | P     |         |       |
| <i>Sylvilagus brasiliensis</i> |                                     | P       | X     |         |       |
| <i>Oryzomys sp.</i>            |                                     |         | P     |         |       |
| <i>Trinomys dimidiatus</i>     |                                     |         | P     |         |       |

**Legendas** **X** seqüenciamento completo da região correspondente; **P** seqüenciamento parcial; **P\*\*** seqüenciamento parcial (<50%); **?** não realizado.

Então foram analisados os dados para as espécies *Artibeus lituratus*, *Chiroderma villosum*, *Sturnira lilium*, *Platyrrhinus lineatus*, *Vampyressa pussila*, *Phyllostomus hastatus*, *Phyllostomus discolor*, *Myotis nigricans* e *Molossus molossus*.

Apenas o segmento do éxon 4 foi utilizado nas análises comparativas com as demais espécies de mamíferos: *Bos taurus*, *Equus caballus*, *Canis familiaris*, *Cerdocyon thous*, *Conepatus semistriatus*, *Eira barbara*, *Sylvilagus brasiliensis*, *Pan troglodytes* e *Homo sapiens*.

Além destas foram também utilizadas seqüências obtidas no GenBank para  $\alpha$ -amilases de *Mus musculus*, *Rattus norvegicus*, *Sus scrofa*, *Gallus gallus*, *Xenopus laevis*, *Tetraodon nigroviridis*, *Anguila japonica* e *Homo sapiens* (Tabela 3).

## **Características básicas dos conjuntos de dados analisados**

Foram utilizados então, para o conjunto de dados PMR, 663 pb correspondendo à totalidade do éxon 4 e uma grande parte do íntron C com 231 e 432 pb respectivamente. Destes, 455 pb são variáveis (100 pb no éxon 4) e 197 pb são conservados (131 pb no éxon 4) (Tabela 8).

No conjunto VERT, contendo diversos mamíferos e alguns outros vertebrados, foram analisados 231 pb (éxon 4) sendo que dentre estes 147 pb são variáveis e 84 conservados (Tabela 7).

Não foram encontradas inserções nem deleções na região do éxon 4, demonstrando um grande grau de conservação desta região (Figura 7).

A razão de transições sobre transversões na região codificante é de  $R= 1,65$  e  $1,60$  para os conjuntos PMR e VERT respectivamente, e a composição nucleotídica é de  $A=25,4\%$ ,  $T=25,4\%$ ,  $C=23,6\%$  e  $G=25,6\%$  para PMR e de  $A=25,5\%$ ,  $T=26,4\%$ ,  $C= 22,5\%$  e  $G=25,6\%$  para VERT (Tabela 7). O valor de  $R$  está dentro do esperado para genes nucleares ( $0,5 < R < 2,0$ ; Nei & Kumar, 2001), e a composição de nucleotídeos pode ser considerada praticamente como igual em frequência, com um conteúdo GC cerca de 49%, também dentro da média para regiões gênicas em vertebrados (40-60%; Nei & Kumar, 2001).

**Tabela 7:** Características gerais descritivas para os conjuntos de dados PMR e VERT

| <b>Características</b>                     | <b>PMR</b>                     |           |           | <b>VERT</b> |           |           |           |           |
|--|--------------------------------|-----------|-----------|-------------|-----------|-----------|-----------|-----------|
| <b>n° de seqüências</b>                    | 17 (15 nos íntrons)            |           |           | 29          |           |           |           |           |
| <b>n° total de sítios</b>                  | 663, 432 íntron C e 231 éxon 4 |           |           | 231 éxon 4  |           |           |           |           |
| <b>d= ts+tv</b>                            | d=0.20                         |           |           | d=0,24      |           |           |           |           |
| <b>sítios variáveis</b>                    | 455 (100 no éxon 4)            |           |           | 147         |           |           |           |           |
| <b>sítios conservados</b>                  | 197 (131 no éxon 4)            |           |           | 84          |           |           |           |           |
| <b>sítios degenerados</b>                  | <b>0x</b>                      | <b>2X</b> | <b>4X</b> | <b>0x</b>   | <b>2X</b> | <b>4X</b> |           |           |
| <b>S</b>                                   | 141                            | 32        | 26        | 133         | 20        | 14        |           |           |
| <b>N</b>                                   | 63                             |           |           | 64          |           |           |           |           |
| <b>R= ts/tv</b>                            | R= 1,65                        |           |           | R= 1,60     |           |           |           |           |
| <b>Frequência dos nucleotídeos</b>         | <b>A</b>                       | <b>T</b>  | <b>C</b>  | <b>G</b>    | <b>A</b>  | <b>T</b>  | <b>C</b>  | <b>G</b>  |
| <b>3a posição dos códons</b>               | 25,4%                          | 25,4%     | 23,6%     | 25,6%       | 25,5%     | 26,4%     | 22,5%     | 25,6%     |
| <b>Frequência na 3a posição dos códons</b> | <b>A3</b>                      | <b>T3</b> | <b>C3</b> | <b>G3</b>   | <b>A3</b> | <b>T3</b> | <b>C3</b> | <b>G3</b> |
|  | 15,3%                          | 28,6%     | 30,3%     | 25,8%       | 16,2%     | 31,0%     | 27,5%     | 25,3%     |

Dos 231 nucleotídeos do éxon 4, 141 são considerados não-degenerados para o conjunto PMR e 133 para VERT (Tabela 7). Isso significa que cerca de 60% dos sítios estão sempre sujeitos a mutações não-sinônimas, se somarmos isso aos sítios 2 vezes e 4 vezes degenerados, teremos que cerca de 86% das mutações serão não-sinônimas, uma frequência maior do que a esperada para frequências iguais de nucleotídeos (75%, Nei & Kumar, 2001). Isso pode ser devido às frequências de nucleotídeos GC na 3ª posição dos códons, A3=15,3%, T3=28,6%, C3=30,3% e G3=25,8% (16,2%, 31,0%, 27,5%, 25,3% em VERT), serem muito altas indicando um elevado conteúdo GC para a terceira posição (Nei & Kumar, 2001).

Isso pode estar demonstrando uma saturação dos sítios sinônimos, causado por restrição funcional ou de conformação do produto gênico (Li, 1997; Golding & Dean, 1998; Nei & Kumar, 2001), esse assunto será discutido mais adiante.

## Análises filogenéticas

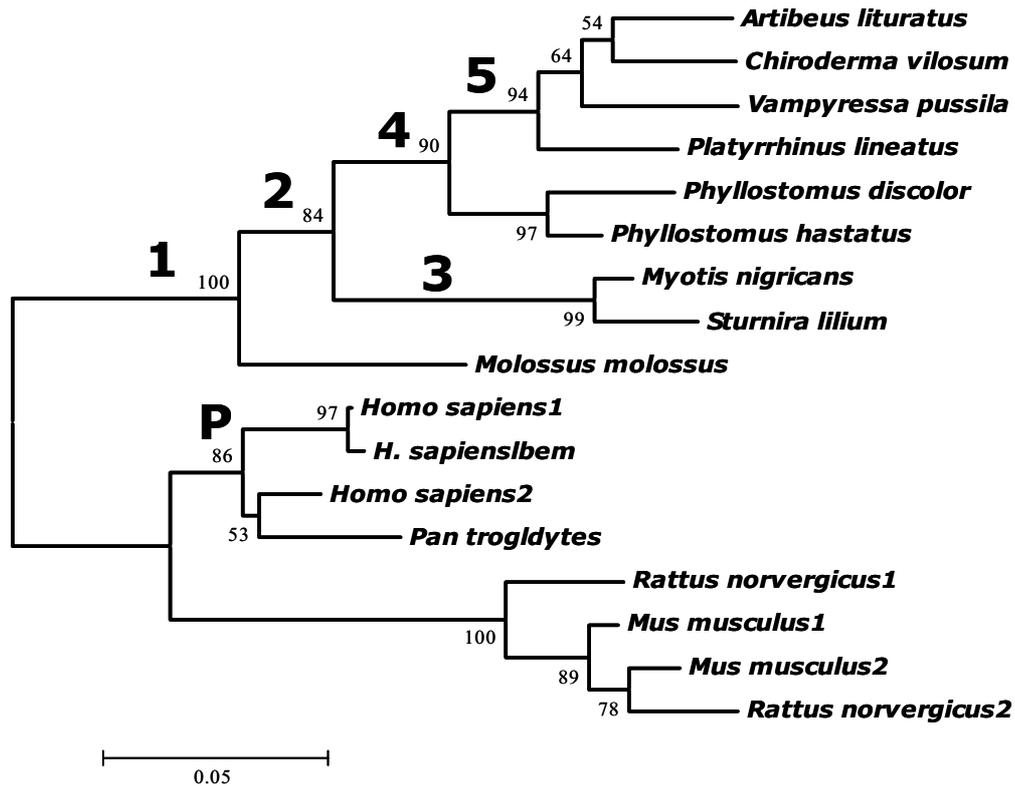
As reconstruções filogenéticas tiveram principalmente o objetivo de verificar a existência de convergência evolutiva guiada por uma possível seleção na região estudada do gene da  $\alpha$ -amilase, bem como verificar se o gene da  $\alpha$ -amilase poderia ser útil para resolver problemas de sistemática, como a posição relativa de morcegos e primatas em relação aos outros mamíferos.

O conjunto de dados PRM foi analisado primeiramente em sua parte codificante (éxon 4, 231 pb), as árvores de NJ e de MP foram praticamente idênticas em topologia (Figuras 6 e 7).

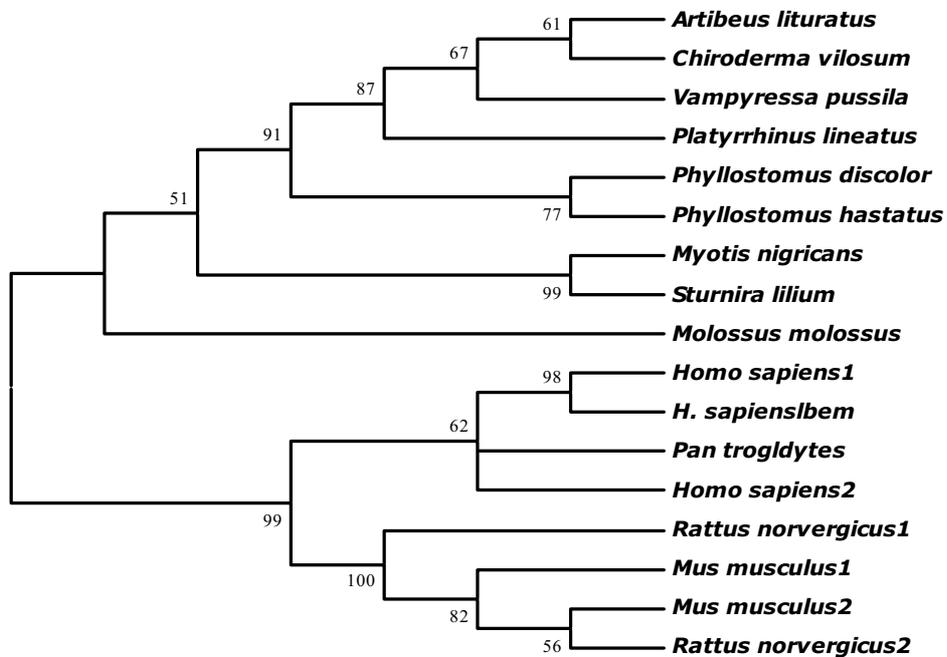
O número de árvores igualmente parcimoniosas quando se utilizava todas as posições, apenas as duas primeiras posições dos códons e a seqüência deduzida de aminoácidos foi de 6, 20 e 9 respectivamente, apenas o consenso é mostrado (Figuras 7-9).

Apesar de praticamente idênticas entre os métodos (MP e NJ), as árvores não apresentaram exatamente a mesma topologia, quando diferentes posições de nucleotídeos (todos ou somente as duas primeiras posições dos códons) eram analisadas e também quando as seqüências deduzidas de aminoácidos foram analisadas (Figuras 6-9).

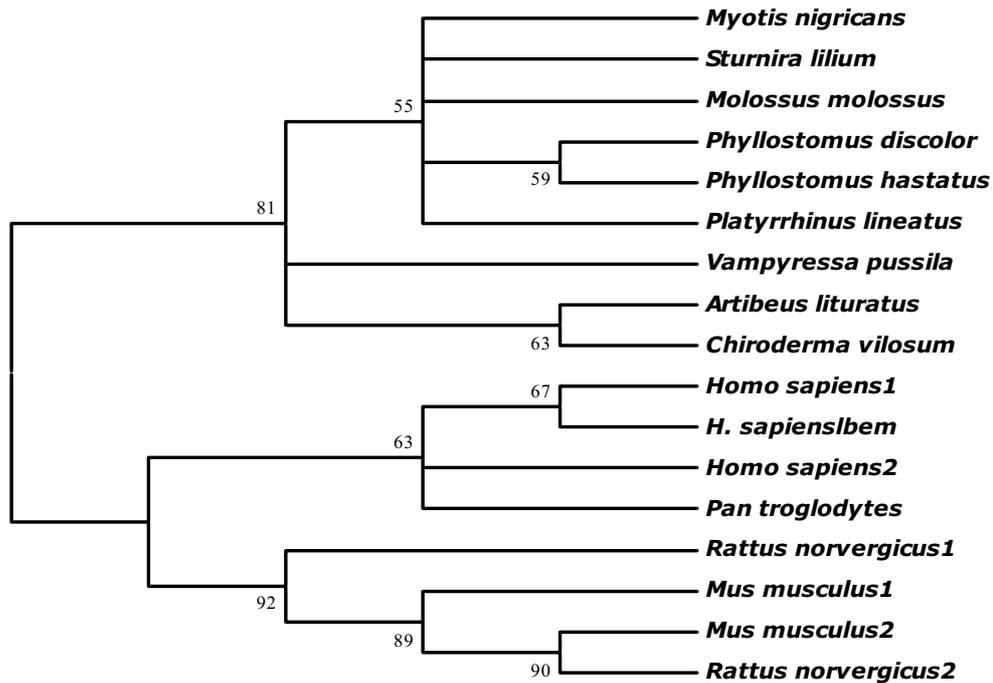
As principais diferenças estão em sua resolução quanto ao grupo dos morcegos. A utilização de todos os sítios de nucleotídeos aparentemente é a que melhor resolve as politomias nos ramos dos morcegos, e estas estão condizentes com filogenias tradicionais morfológicas (Simmon & Geisler, 1998) e filogenias moleculares utilizando Citocromo B e RAG-2 (idênticas em topologia, Figura 10). A única e marcante discrepância contra as filogenias tradicionais é o agrupamento de *Sturnira lilium* e *Myotis nigricans* em quase todas as árvores analisadas. Este é um agrupamento bastante incomum, pois estas espécies pertencem claramente a famílias diferentes de morcegos, Phyllostomidae e Vespertilionidae respectivamente (Wilson & La Val, 1974; Gannon *et al.*, 1989; Novak, 1997; Simmons & Geisler, 1998).



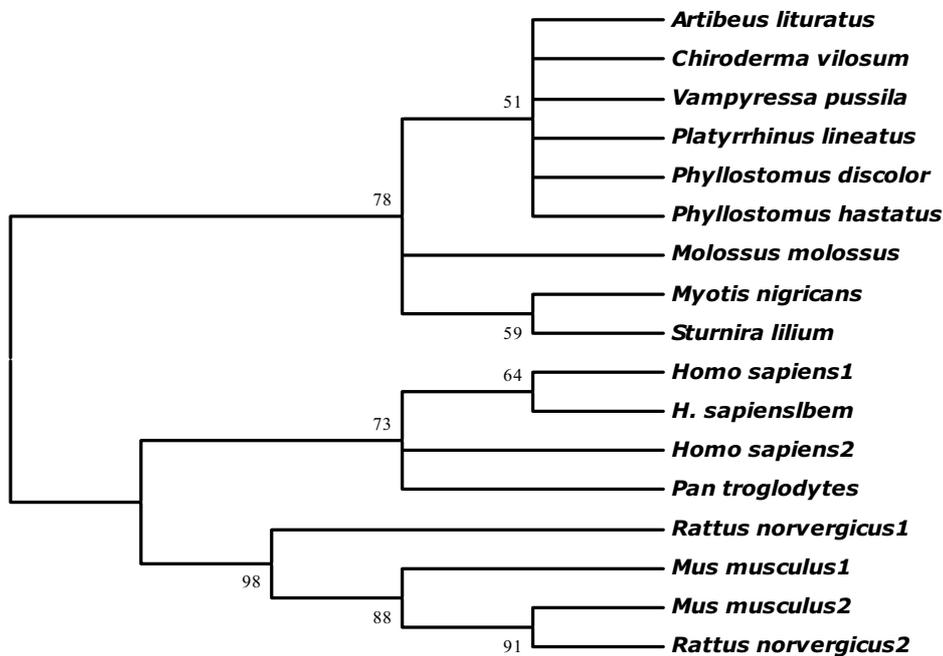
**Figura 6:** Árvore de NJ utilizando todas as posições (231 bp) do éxon 4 da  $\alpha$ -amilase para o conjunto PMR de dados. Os números pequenos são porcentagens de *bootstrap* após 1.000 replicações. Os números grandes e a letra P referem-se aos ramos utilizados para as análises de seleção.



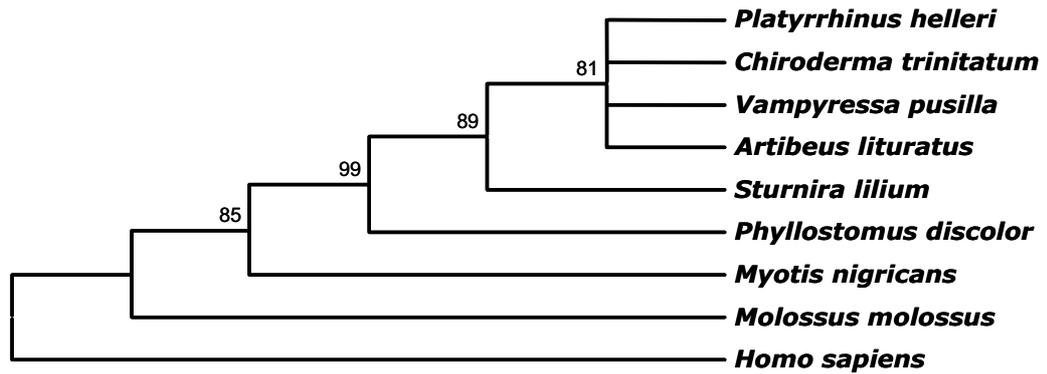
**Figura 7:** Consenso das 6 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com todas as posições (231 bp) do éxon 4 de  $\alpha$ -amilase no conjunto de dados PMR. Valores nos ramos indicam porcentagens de *bootstrap* após 1.000 replicações.



**Figura 8:** Consenso das 20 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com apenas as duas primeiras posições dos códons do éxon 4 da  $\alpha$ -amilase no conjunto de dados PMR. Valores nos ramos indicam porcentagens de *bootstrap* após 1.000 replicações.



**Figura 9:** Consenso das 9 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com a seqüência deduzida de aminoácidos (77 aa) do éxon 4 da  $\alpha$ -amilase no conjunto de dados PMR. Valores nos ramos indicam porcentagens de *bootstrap* após 1.000 replicações.

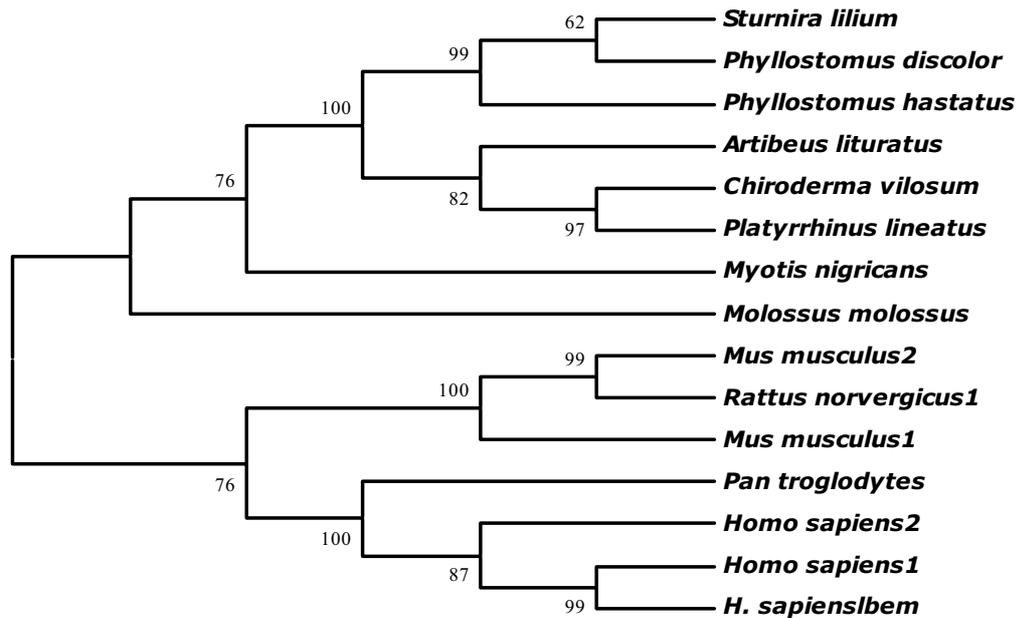


**Figura 10:** Consenso das 5 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com um segmento de 330 pb do Citocromo B. Valores nos ramos indicam porcentagens de *bootstrap* após 1.000 replicações.

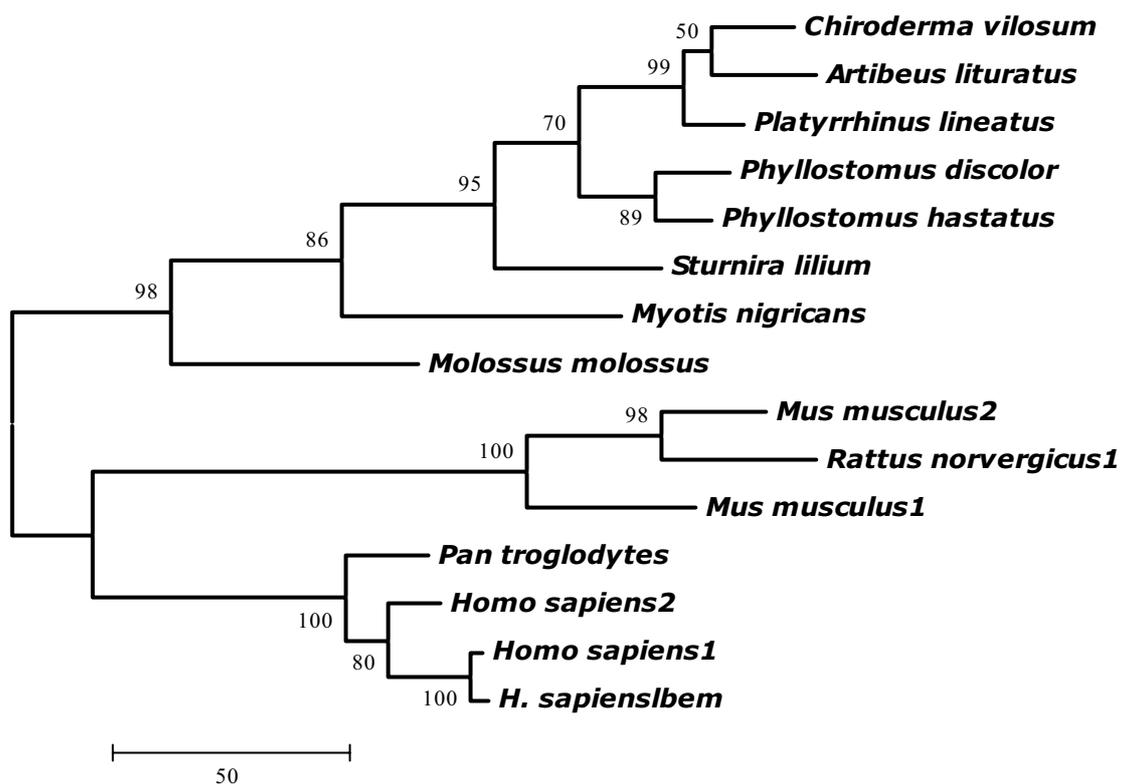
Neste mesmo conjunto de dados (PMR), foram feitas análises filogenéticas com os íntrons (íntron C, 432 pb) e com o conjunto inteiro de nucleotídeos (663 pb) (Figuras 11 e 12).

A topologia das árvores utilizando os íntrons entre os métodos (NJ e MP) também foi bastante semelhante. Mas a posição de *S. lilium* e *M. nigricans* mudou consideravelmente. *S. lilium* nestas reconstruções se agrupa com os outros Phyllostomidae, porém junto com *Phyllostomus hastatus* e *Phyllostomus discolor* que pertencem à sub-família Phyllostominae, enquanto que *S. lilium* pertence à sub-família Stenodermatinae.

Utilizando o conjunto inteiro (éxon 4 e íntron C) também não há diferença entre os métodos, ressaltando que métodos diferentes, importam menos que as suposições quando o conjunto de dados estudado obedece às mesmas propriedades (Russo *et al.*, 1996). *Sturnira lilium* nesta análise ficou como grupo irmão de Phyllostomidae, assumindo basicamente o ponto médio entre as duas outras análises.



**Figura 11:** Reconstrução filogenética utilizando MP com o segmento correspondente ao íntron C da  $\alpha$ -amilase (432 bp). Valores nos ramos indicam porcentagens de *bootstrap* após 1.000 replicações.



**Figura 12:** Reconstrução filogenética utilizando MP com o segmento correspondente ao íntron C e éxon 4 da  $\alpha$ -amilase (432 bp). Valores nos ramos indicam porcentagens de *bootstrap* após 1.000 replicações. Tamanho dos ramos relativos à número de mudanças.

Árvores gênicas são inferências das árvores verdadeiras das espécies que podem nem sempre serem idênticas (Li, 1997; Nei & Kumar, 2001). Genes e segmentos de DNA podem possuir taxas de evolução e contingências evolutivas diferentes, dependendo de inúmeros fatores, incluindo vários fatores estocásticos ou processos de seleção natural, tais como os que levam à convergência evolutiva em distintas linhagens (Li, 1997; Nei & Kumar, 2001).

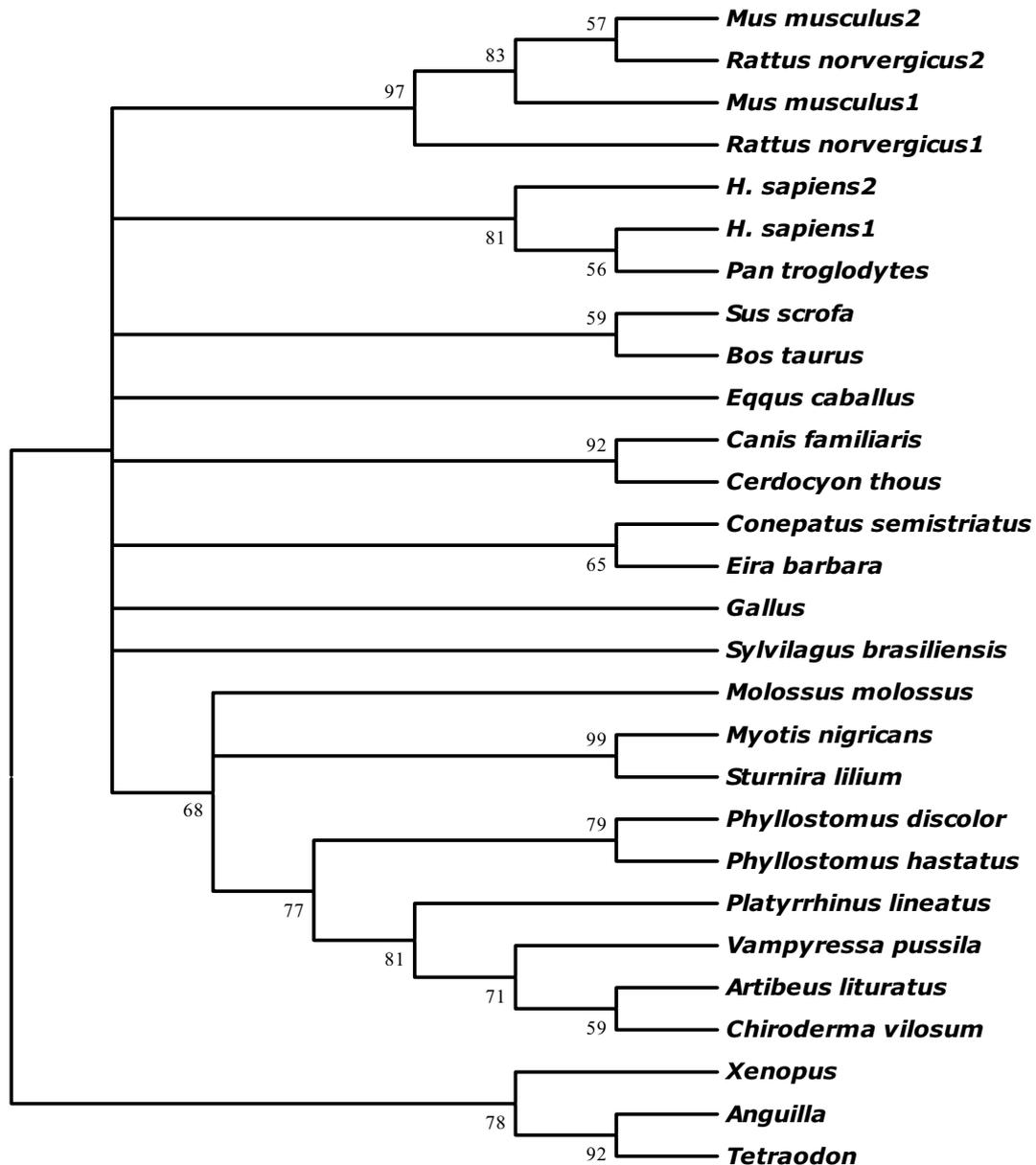
A posição dúbia de *S. liliium* pode estar associada a um destes fatores, e será mais discutida adiante. Nenhuma outra possível distorção na topologia das árvores foi encontrada entre morcegos com diferentes hábitos alimentares.

As reconstruções filogenéticas utilizando o conjunto VERT, também foram congruentes entre os métodos (Figura 13) e produziu topologias semelhantes às outras árvores para os morcegos. No entanto este conjunto causou diversas politomias nos diferentes grupos de mamíferos, portanto não é possível assumir uma topologia confiável.

A maioria das árvores do conjunto PMR aponta para um agrupamento entre primatas e roedores, ao invés de morcegos e primatas, dando suporte aos grupos Laurasiatheria e Euarchontoglires sugeridos recentemente por vários autores (Madsen *et al.*, 2001; Murphy *et al.*, 2001; Arnason *et al.*, 2002; Deusuc *et al.*, 2002), ao invés do tradicional Archonta (Novacek, 1992).

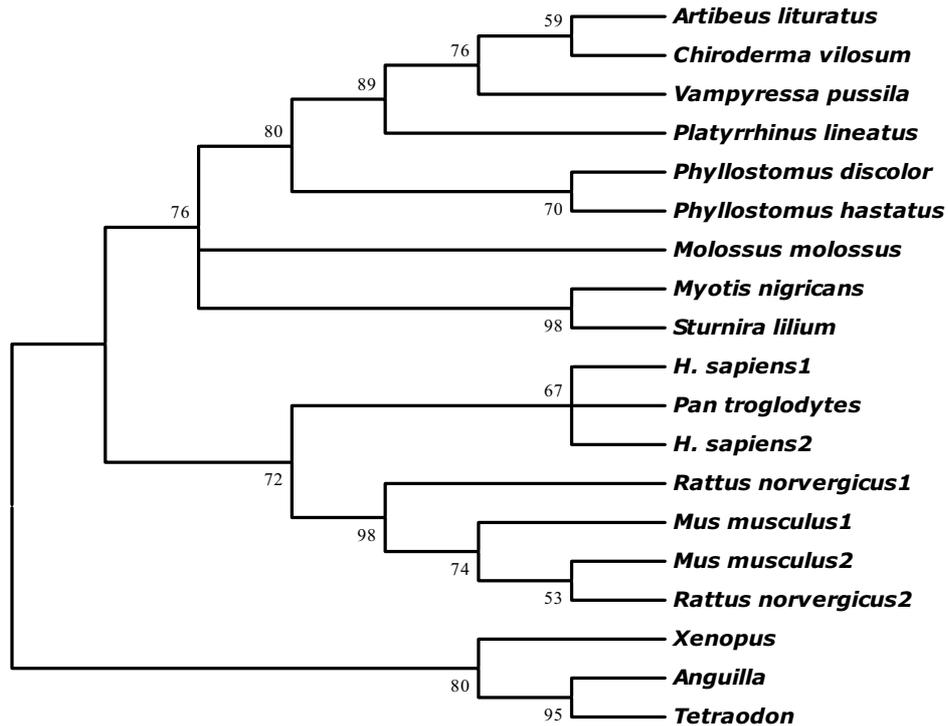
No entanto estas topologias são obtidas com a técnica de enraizamento no ponto médio (Nei & Kumar, 2001), e recentemente alguns autores demonstraram que a eficiência deste método é de 50% dos casos (comunicação pessoal, Cláudia Russo 2003 em publicação). Como a topologia da árvore gerada com o conjunto todo de mamíferos não pode ser analisada, uma nova comparação precisou ser feita.

Essa nova reconstrução foi feita utilizando o conjunto PMR com a adição de seqüências de *Tetraodon nigroviridis* e *Anguila japonica* (Piscis) e *Xenopus laevis* (Anphibia) (Figura 14). Estas seqüências foram escolhidas porque foram as únicas com a certeza de serem divergentes o suficiente para, pelo menos, agrupar mamíferos (*Gallus gallus* - Aves - agrupou junto com os mamíferos na análise de VERT).

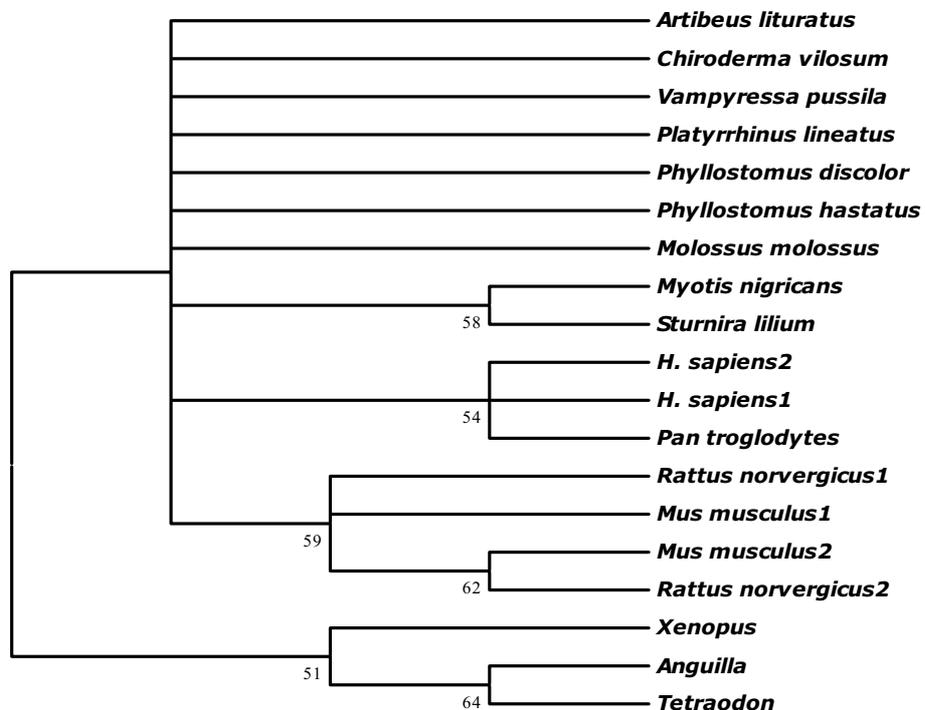


**Figura 13:** Consenso das 8 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com todas as posições do éxon 4 de  $\alpha$ -amilase no conjunto de dados VERT. Valores nos ramos indicam porcentagens de *bootstrap* após 1.000 replicações.

Mais uma vez ambos os métodos sugeriram árvores semelhantes, entretanto, a utilização de nucleotídeos dá suporte às filogenias recentemente propostas (Laurasiatheria e Euarchontoglires; Madsen *et al.*, 2001; Murphy *et al.*, 2001; Arnason *et al.*, 2002; Deusuc *et al.*, 2002), enquanto a utilização de aminoácidos deduzidos não esclarece adequadamente as relações (Figuras 14 e 15)



**Figura 14:** Consenso das 3 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com todas as posições do éxon 4 de  $\alpha$ -amilase no conjunto de dados PMR mais grupo externo. Valores nos ramos indicam porcentagens de *bootstrap* após 1.000 replicações.



**Figura 15:** Consenso das 6 árvores mais parcimoniosas da reconstrução filogenética utilizando MP com aminoácidos deduzidos no conjunto de dados PMR mais grupo externo. Valores nos ramos indicam porcentagens de *bootstrap* após 1.000 replicações.

## Detecção de adaptações moleculares

A análise deste segmento da  $\alpha$ -amilase não mostra qualquer evidência de seleção natural relacionada a diferentes hábitos alimentares nas análises filogenéticas. Isso não significa que esta não ocorra, uma vez que na maioria dos trabalhos onde esta investigação ocorreu, as filogenias obtidas eram idênticas (ou semelhantes) às tradicionais (Messier & Stewart, 1997; Vacquier *et al.*, 1997; Rooney & Zhang, 1999; Yang *et al.*, 2000). Portanto os testes de seleção a seguir foram utilizados para verificar sua possível presença.

O teste de seleção de Nei-Gojobori tradicional apontou um grande desvio da hipótese de neutralidade (Tabela 8), mas não detectou seleção em nenhum dos pares de seqüência analisados, utilizando o teste Z (Tabela 9), isto aponta para a hipótese de que seleção purificadora (negativa) estaria agindo neste segmento do gene e isto foi indicado com um terceiro teste para esta hipótese (Tabela 10). Porém sabe-se que o teste de Nei-Gojobori não leva em consideração a razão de transições e transversões, utilizando como parâmetro  $R = 0,5$ , isso pode afetar muito os resultados, já que este leva a uma estimativa errada de **N** e **S** e conseqüentemente de  $d_N$  e  $d_S$ .

A razão  $R$  média no conjunto de dados analisados é de  $R = 1,65$  (Tabela 7), por isso realizamos mais uma vez o teste de seleção, em sua versão modificada por Zhang e colaboradores (1998). Os resultados continuam mostrando desvio da neutralidade e seleção negativa com a utilização do teste Z (Tabela 8-13).

**Tabela 8: Teste de Neutralidade NG** - Resultado do teste de desvio da neutralidade utilizando o método de Nei-Gojobori para o conjunto de dados PMR. Em vermelho valores com significância  $p < 0,05$ ; sublinhados valores com significância marginal; n/c impossível de calcular

|                       | 1     | 2            | 3     | 4     | 5     | 6     | 7     | 8     | 9            | 10           | 11    | 12    | 13    | 14    | 15    | 16    | 17 |
|-----------------------|-------|--------------|-------|-------|-------|-------|-------|-------|--------------|--------------|-------|-------|-------|-------|-------|-------|----|
| Homo_sapiens1         |       |              |       |       |       |       |       |       |              |              |       |       |       |       |       |       |    |
| Homo_sapiens_2        | 0.074 |              |       |       |       |       |       |       |              |              |       |       |       |       |       |       |    |
| H._sapienslbem        | 0.320 | <u>0.058</u> |       |       |       |       |       |       |              |              |       |       |       |       |       |       |    |
| Pan_troglodytes       | 0.255 | 0.215        | 0.207 |       |       |       |       |       |              |              |       |       |       |       |       |       |    |
| Molopus_molossus      | 0.001 | 0.002        | 0.001 | 0.004 |       |       |       |       |              |              |       |       |       |       |       |       |    |
| Myotis_nigricans      | 0.001 | 0.000        | 0.001 | 0.001 | 0.001 |       |       |       |              |              |       |       |       |       |       |       |    |
| Artibeus_lituratus    | 0.000 | 0.001        | 0.000 | 0.001 | 0.001 | 0.001 |       |       |              |              |       |       |       |       |       |       |    |
| Chiroderma_vilosum    | 0.002 | 0.002        | 0.003 | 0.002 | 0.001 | 0.000 | 0.010 |       |              |              |       |       |       |       |       |       |    |
| Platyrrhinus_lineatus | 0.001 | 0.001        | 0.001 | 0.001 | 0.001 | 0.001 | 0.070 | 0.033 |              |              |       |       |       |       |       |       |    |
| Vampyressa_pussila    | 0.001 | 0.001        | 0.001 | 0.001 | 0.001 | 0.001 | 0.007 | 0.008 | 0.023        |              |       |       |       |       |       |       |    |
| Phyllostomus_discolor | 0.001 | 0.001        | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.005        | 0.004        |       |       |       |       |       |       |    |
| Phyllostomus_hastatus | 0.001 | 0.001        | 0.001 | 0.001 | 0.001 | 0.001 | 0.004 | 0.003 | 0.007        | 0.005        | 0.072 |       |       |       |       |       |    |
| Sturnira_lilium       | 0.001 | 0.001        | 0.001 | 0.001 | 0.001 | 0.744 | 0.002 | 0.001 | 0.001        | 0.001        | 0.000 | 0.001 |       |       |       |       |    |
| Mus_musculus1         | 0.002 | 0.005        | 0.002 | 0.009 | 0.001 | 0.006 | 0.020 | 0.021 | 0.006        | 0.007        | 0.017 | 0.006 | 0.008 |       |       |       |    |
| Mus_musculus2         | 0.005 | 0.012        | 0.004 | 0.027 | 0.001 | 0.006 | 0.035 | 0.100 | 0.009        | 0.025        | 0.018 | 0.015 | 0.009 | 0.582 |       |       |    |
| Rattus_norvergicus1   | 0.001 | 0.002        | 0.001 | 0.003 | 0.001 | 0.071 | 0.249 | n/c   | <u>0.045</u> | n/c          | 0.343 | 0.027 | 0.108 | 0.013 | 0.035 |       |    |
| Rattus_norvergicus2   | 0.005 | 0.017        | 0.004 | 0.033 | 0.001 | 0.011 | 0.027 | 0.221 | 0.040        | <u>0.049</u> | 0.097 | 0.028 | 0.016 | 0.228 | 0.076 | 0.074 |    |



**Tabela 10: Teste de Seleção Purificadora NG** - Resultado do teste de seleção purificadora (negativa) utilizando o método de Nei-Gojobori para o conjunto de dados PMR. Em vermelho valores com significância  $p < 0,05$ ; sublinhados valores com significância marginal; n/c impossível de calcular

|                       | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8            | 9     | 10    | 11           | 12    | 13           | 14    | 15    | 16    | 17 |
|-----------------------|-------|-------|-------|-------|-------|-------|-------|--------------|-------|-------|--------------|-------|--------------|-------|-------|-------|----|
| Homo_sapiens1         |       |       |       |       |       |       |       |              |       |       |              |       |              |       |       |       |    |
| Homo_sapiens_2        | 0.037 |       |       |       |       |       |       |              |       |       |              |       |              |       |       |       |    |
| H._sapienslbem        | 1.000 | 0.029 |       |       |       |       |       |              |       |       |              |       |              |       |       |       |    |
| Pan_troglodytes       | 0.128 | 1.000 | 0.104 |       |       |       |       |              |       |       |              |       |              |       |       |       |    |
| Molopus_molossus      | 0.001 | 0.001 | 0.001 | 0.002 |       |       |       |              |       |       |              |       |              |       |       |       |    |
| Myotis_nigricans      | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |       |       |              |       |       |              |       |              |       |       |       |    |
| Artibeus_lituratus    | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 |       |              |       |       |              |       |              |       |       |       |    |
| Chiroderma_vilosum    | 0.001 | 0.001 | 0.001 | 0.001 | 0.000 | 0.000 | 0.005 |              |       |       |              |       |              |       |       |       |    |
| Platyrrhinus_lineatus | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.035 | 0.016        |       |       |              |       |              |       |       |       |    |
| Vampyressa_pussila    | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.003 | 0.004        | 0.011 |       |              |       |              |       |       |       |    |
| Phyllostomus_discolor | 0.001 | 0.000 | 0.001 | 0.001 | 0.001 | 0.000 | 0.001 | 0.001        | 0.003 | 0.002 |              |       |              |       |       |       |    |
| Phyllostomus_hastatus | 0.001 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.002 | 0.002        | 0.004 | 0.002 | 0.036        |       |              |       |       |       |    |
| Sturnira_lilium       | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | 0.372 | 0.001 | 0.000        | 0.000 | 0.000 | 0.000        | 0.000 |              |       |       |       |    |
| Mus_musculus1         | 0.001 | 0.002 | 0.001 | 0.004 | 0.000 | 0.003 | 0.010 | 0.010        | 0.003 | 0.003 | 0.009        | 0.003 | 0.004        |       |       |       |    |
| Mus_musculus2         | 0.003 | 0.006 | 0.002 | 0.013 | 0.001 | 0.003 | 0.018 | <u>0.050</u> | 0.005 | 0.012 | 0.009        | 0.007 | 0.005        | 0.291 |       |       |    |
| Rattus_norvergicus1   | 0.000 | 0.001 | 0.000 | 0.001 | 0.000 | 0.036 | 0.124 | n/c          | 0.022 | n/c   | 0.172        | 0.014 | <u>0.054</u> | 0.007 | 0.017 |       |    |
| Rattus_norvergicus2   | 0.002 | 0.008 | 0.002 | 0.017 | 0.001 | 0.006 | 0.014 | 0.111        | 0.020 | 0.024 | <u>0.049</u> | 0.014 | 0.008        | 0.114 | 0.038 | 0.037 |    |

**Tabela 11: Teste de Neutralidade NG (modificado)** - Resultado do teste de desvio da neutralidade utilizando o método modificado de Nei-Gojobori, com  $R=1.65$ , para o conjunto de dados PMR. Em vermelho valores com significância  $p < 0,05$ ; sublinhados valores com significância marginal; n/c impossível de calcular

|                       | 1            | 2     | 3     | 4            | 5     | 6     | 7     | 8     | 9     | 10    | 11           | 12    | 13    | 14    | 15    | 16           | 17    |
|-----------------------|--------------|-------|-------|--------------|-------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|--------------|-------|
| Homo_sapiens1         |              |       |       |              |       |       |       |       |       |       |              |       |       |       |       |              |       |
| Homo_sapiens_2        | <u>0.058</u> |       |       |              |       |       |       |       |       |       |              |       |       |       |       |              |       |
| H._sapienslbem        | 1.000        | 0.042 |       |              |       |       |       |       |       |       |              |       |       |       |       |              |       |
| Pan_troglodytes       | 0.222        | 1.000 | 0.176 |              |       |       |       |       |       |       |              |       |       |       |       |              |       |
| Molossus_molossus     | 0.001        | 0.001 | 0.001 | 0.005        |       |       |       |       |       |       |              |       |       |       |       |              |       |
| Myotis_nigricans      | 0.000        | 0.000 | 0.000 | 0.000        | 0.000 |       |       |       |       |       |              |       |       |       |       |              |       |
| Artibeus_lituratus    | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.001 |       |       |       |       |              |       |       |       |       |              |       |
| Chiroderma_vilosum    | 0.000        | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.007 |       |       |       |              |       |       |       |       |              |       |
| Platyrrhinus_lineatus | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.000 | 0.065 | 0.029 |       |       |              |       |       |       |       |              |       |
| Vampyressa_pussila    | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.000 | 0.005 | 0.006 | 0.020 |       |              |       |       |       |       |              |       |
| Phyllostomus_discolor | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.000 | 0.001 | 0.001 | 0.004 | 0.003 |              |       |       |       |       |              |       |
| Phyllostomus_hastatus | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.000 | 0.003 | 0.003 | 0.006 | 0.004 | <u>0.050</u> |       |       |       |       |              |       |
| Sturnira_lilium       | 0.000        | 0.000 | 0.000 | 0.001        | 0.001 | 0.490 | 0.002 | 0.000 | 0.001 | 0.001 | 0.000        | 0.001 |       |       |       |              |       |
| Mus_musculus1         | 0.002        | 0.005 | 0.001 | 0.012        | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 |
| Mus_musculus2         | 0.007        | 0.015 | 0.005 | 0.042        | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.000 | 0.385        |       |
| Rattus_norvergicus1   | 0.001        | 0.001 | 0.000 | 0.003        | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.009 | 0.029        |       |
| Rattus_norvergicus2   | 0.006        | 0.025 | 0.005 | <u>0.055</u> | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.170 | <u>0.051</u> | 0.065 |



**Tabela 13: Teste de Seleção Purificadora NG (modificado)** - Resultado do teste de seleção positiva utilizando o método modificado de Nei-Gojobori, com R=1.65, para o conjunto de dados PMR. Em vermelho valores com significância  $p < 0,05$ ; sublinhados valores com significância marginal; n/c impossível de calcular.

|                       | 1            | 2     | 3     | 4            | 5     | 6     | 7            | 8     | 9     | 10    | 11           | 12    | 13    | 14    | 15    | 16           | 17    |
|-----------------------|--------------|-------|-------|--------------|-------|-------|--------------|-------|-------|-------|--------------|-------|-------|-------|-------|--------------|-------|
| Homo_sapiens1         |              |       |       |              |       |       |              |       |       |       |              |       |       |       |       |              |       |
| Homo_sapiens_2        | <u>0.058</u> |       |       |              |       |       |              |       |       |       |              |       |       |       |       |              |       |
| H._sapienslbem        | 1.000        | 0.042 |       |              |       |       |              |       |       |       |              |       |       |       |       |              |       |
| Pan_troglodytes       | 0.222        | 1.000 | 0.176 |              |       |       |              |       |       |       |              |       |       |       |       |              |       |
| Molossus_molossus     | 0.001        | 0.001 | 0.001 | 0.005        |       |       |              |       |       |       |              |       |       |       |       |              |       |
| Myotis_nigricans      | 0.000        | 0.000 | 0.000 | 0.000        | 0.000 |       |              |       |       |       |              |       |       |       |       |              |       |
| Artibeus_lituratus    | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.001 |              |       |       |       |              |       |       |       |       |              |       |
| Chiroderma_vilosum    | 0.000        | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.007        |       |       |       |              |       |       |       |       |              |       |
| Platyrrhinus_lineatus | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.000 | 0.065        | 0.029 |       |       |              |       |       |       |       |              |       |
| Vampyressa_pussila    | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.000 | 0.005        | 0.006 | 0.020 |       |              |       |       |       |       |              |       |
| Phyllostomus_discolor | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.000 | 0.001        | 0.001 | 0.004 | 0.003 |              |       |       |       |       |              |       |
| Phyllostomus_hastatus | 0.000        | 0.000 | 0.000 | 0.000        | 0.001 | 0.000 | 0.003        | 0.003 | 0.006 | 0.004 | <u>0.050</u> |       |       |       |       |              |       |
| Sturnira_lilium       | 0.000        | 0.000 | 0.000 | 0.001        | 0.001 | 0.490 | <u>0.002</u> | 0.000 | 0.001 | 0.001 | 0.000        | 0.001 |       |       |       |              |       |
| Mus_musculus1         | 0.002        | 0.005 | 0.001 | 0.012        | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 |
| Mus_musculus2         | 0.007        | 0.015 | 0.005 | 0.042        | 0.001 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.000 | 0.385        |       |
| Rattus_norvergicus1   | 0.001        | 0.001 | 0.000 | 0.003        | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.009 | 0.029        |       |
| Rattus_norvergicus2   | 0.006        | 0.025 | 0.005 | <u>0.055</u> | 0.001 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.000        | 0.000 | 0.000 | 0.000 | 0.170 | <u>0.051</u> | 0.065 |

Mesmo utilizando a média dos grupos (Tabela 14), para o teste Z, este não foi significativo para seleção, significando que em média não há diferença significativa entre  $d_N$  e  $d_S$ .

**Tabela 14:** Resultado do teste de Seleção para as médias dos grupos utilizando os métodos de Nei-Gojobori e Nei-Gojobori modificado

|                 | <b>Nei-Gojobori</b> | <b>Nei-gojobori modificado</b> |
|-----------------|---------------------|--------------------------------|
| <b>Primatas</b> | 0,414               | 0,44                           |
| <b>Morcegos</b> | 0,117               | 0,159                          |
| <b>Roedores</b> | 0,337               | 0,36                           |

O teste Z é considerado um teste para grandes amostras, isto é, o número de mutações sinônimas e não-sinônimas deve ser grande. Mas esse número varia de autor para autor, para alguns  $d_S \geq 0,5$  (Yang, 2000) é o suficiente, para outros a distância geral (não a de mutações sinônimas) deve ser acima de 0,25, (Nei & Kumar, 2000) ou ainda o número de substituições sinônimas (**s**) e não-sinônimas (**n**) deve ser 10 ou mais. Como os nossos dados podem ser muito pequenos para o teste Z ( $d_S = 0,45$  e  $d = 0,2$ ;  $s \sim 23$ ,  $n \sim 13$ ), utilizamos o teste exato de Fisher, indicado para pequenas amostras (Zhang *et al.*, 1997 e 1998).

Para o teste de Fisher foram utilizadas as seqüências ancestrais inferidas para cada ramo indicado na árvore da Figura 6.

Os resultados do teste de Fisher foram todos significativos, isso indica que existe desvio da neutralidade (Zhang *et al.*, 1997 e 1998) (Tabela 15). O teste é bidirecional, por isso devemos comparar o número de **n** e **s** para as seqüências reconstruídas. Neste caso veremos que o número de **s** ( $\sim 27$ ) é, em geral muito maior do que de **n** ( $\sim 11$ ) se considerarmos que temos em média 168 sítios não-sinônimos e 63 sinônimos (Tabela 7) isso implica que  $d_N (=0,06)$  é muito menor do que  $d_S (0,4)$  logo a seleção neste caso é do tipo purificadora com  $\omega = 0,15$ .

Yang (2002) critica nitidamente o uso do teste exato de Fisher, pois este depende de reconstruções de seqüências ancestrais, que pode conter erros, além disso os métodos utilizados para esta comparação não levam em consideração o conteúdo GC da 3ª posição dos códons, o que pode causar erros na contagem do número de sítios sinônimos **S** e não-sinônimos **N** e na estimativa do número de mutações sinônimas (**s**) e não-sinônimas (**n**) (Yang 2002).

**Tabela 15:** Resultados do teste exato de Fisher para as seqüências ancestrais dos ramos 1, 2, 3, 4, 5 e P da árvore da Figura 6.

|   | P      | 1      | 2      | 3      | 4      | 5 |
|---|--------|--------|--------|--------|--------|---|
| P |        |        |        |        |        |   |
| 1 | 0,0000 |        |        |        |        |   |
| 2 | 0,0001 | 0,0000 |        |        |        |   |
| 3 | 0,0000 | 0,0000 | 0,0000 |        |        |   |
| 4 | 0,0000 | 0,0000 | 0,0015 | 0,0000 |        |   |
| 5 | 0,0000 | 0,0017 | 0,0000 | 0,0000 | 0,0002 |   |

O teste de Creevey & McInerney (CM2002) é um teste principalmente das linhagens utilizando  $r^3t$  (teste de razão das taxas relativas). Apenas em três ramos foi detectado desvio de neutralidade e um aumento significativo no número de substituições não-sinônimas não direcionais ( $RV > RI$ ), um no ramo 3 dos morcegos e em dois ramos dos primatas (Tabela 16, Figura 16).

Este teste ainda apontou as outras linhagens como evoluindo sob seleção negativa e apenas dois ramos foram considerados estritamente neutros. Um dos ramos já era esperado, pois agrupa apenas o éxon 4 da  $\alpha$ -amilase 1 humana do GenBank e a seqüência humana gerada no Lbem. O outro ramo engloba as  $\alpha$ -amilases 2 de ratos e camundongos (Tabela 16, Figura 16).

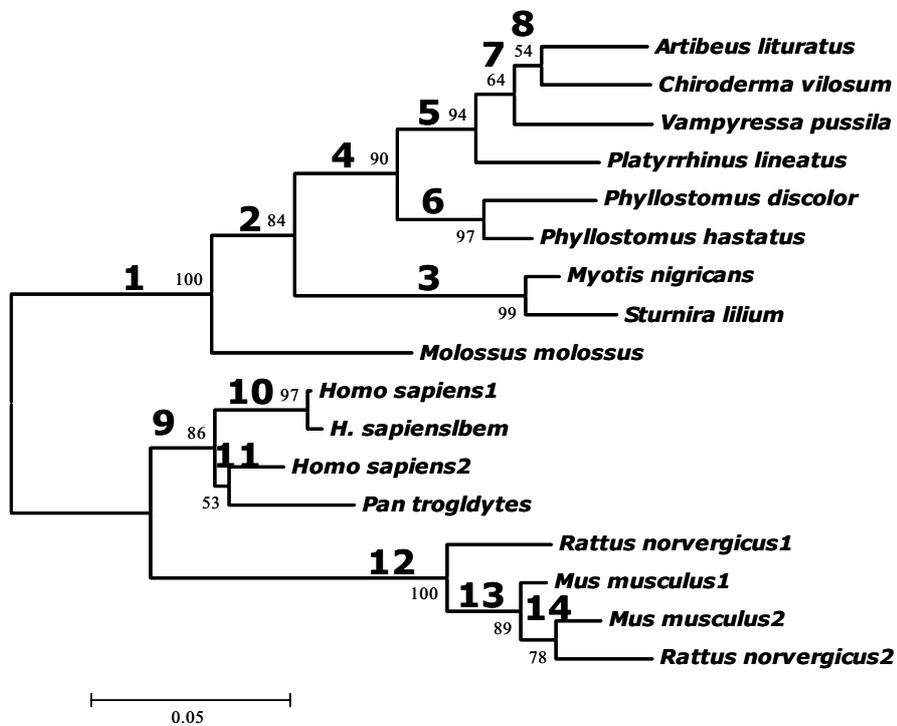
Outro evento interessante é que no ramo 3, tanto o  $r^3t$ , quanto o teste de seleção negativa foram significativos ( $p < 0,05$ ). Isso indica que pode estar ocorrendo seleção em apenas alguns poucos sítios e o restante estar sob seleção purificadora por restrição funcional (Creevey & McInerney, 2002). O ramo detectado é também o único ramo nos morcegos, que desvia da filogenia tradicional quando analisado apenas o éxon 4 (ramo 3, Figura 16).

Quando o conjunto VERT é utilizado detecta-se também o mesmo padrão acima. Além disso, no ramo dos carnívoros mustelídeos *Conepatus semistriatus* e *Eira barbara* acontece um padrão de substituição misto de seleção e purificação semelhante ao de *Myotis nigricans* e *Sturnira liliium* (ramo 3). Este padrão pode ser a explicação de porquê estes carnívoros (mustelídeos) não se agrupam com os outros carnívoros (canídeos) nas reconstruções filogenéticas (Figura 13). Estes dois mustelídeos também apresentaram variações idênticas de tamanho do segmento gerado pelo par de iniciadores Batamy 1b-2 e 1-2, demonstrando taxas de evolução diferentes entre estes e as demais linhagens de carnívoros.

**Tabela 16:** Resultado do  $r^3t$  do método de Creevey-McInerney.

| Ramo      | Substituições |    |    |    | seleção positiva |        | seleção purificadora |        |
|-----------|---------------|----|----|----|------------------|--------|----------------------|--------|
|           | RI            | RV | SI | SV | Teste G          | Fisher | Teste G              | Fisher |
| <b>1</b>  | 13            | 37 | 29 | 48 | 0,2<p<0,1        | Inc    | 0,005<p<0,000        | Inc    |
| <b>2</b>  | 7             | 28 | 14 | 37 | 0,5<p<0,2        | Inc    | 0,005<p<0,000        | Inc    |
| <b>3</b>  | 1             | 5  | 9  | 3  | 0,025<p<0,01     | 0,03   | 0,005<p<0,000        | Inc    |
| <b>4</b>  | 5             | 23 | 12 | 28 | 0,5<p<0,2        | Inc    | 0,005<p<0,000        | Inc    |
| <b>5</b>  | 1             | 16 | 5  | 19 | 0,2<p<0,1        | Inc    | 0,005<p<0,000        | Inc    |
| <b>6</b>  | 0             | 4  | 5  | 5  | 0,1<p<0,05       | 0,12   | 0,025<p<0,01         | 0,01   |
| <b>7</b>  | 0             | 7  | 3  | 15 | 0,2<p<0,1        | 0,35   | 0,005<p<0,000        | Inc    |
| <b>8</b>  | 0             | 4  | 2  | 11 | 0,5<p<0,2        | 0,57   | 0,005<p<0,000        | 0,001  |
| <b>9</b>  | 1             | 10 | 5  | 1  | 0,005<p<0,000    | 0,005  | 0,5<p<0,2            | 0,26   |
| <b>10</b> | 1             | 1  | 5  | 0  | 0,2<p<0,1        | 0,28   | 0,5<p<0,2            | 0,28   |
| <b>11</b> | 1             | 7  | 5  | 0  | 0,005<p<0,000    | 0,004  | 0,1<p<0,05           | 0,087  |
| <b>12</b> | 6             | 7  | 11 | 13 | 0,99<p<0,95      | Inc    | 0,005<p<0,000        | Inc    |
| <b>13</b> | 4             | 4  | 2  | 8  | 0,2<p<0,1        | 0,22   | 0,05<p<0,025         | Inc    |
| <b>14</b> | 2             | 3  | 0  | 5  | 0,2<p<0,1        | 0,20   | 0,5<p<0,2            | 0,2    |

*Legendas* RI e RV: mutação não-sinônima sinapomórfica e não-sinapomórfica respectivamente. SI e SV: mutação sinônima sinapomórfica e não-sinapomórfica respectivamente. Resultados significativos ( $p<0,05$ ) em vermelho. Inc= incalculável, utiliza-se o resultado do teste G.



**Figura 16:** Numeração dos ramos da filogenia para o teste CM2002.

As últimas análises foram feitas com os LRTs para se testar vários modelos de verossimilhança (Tabela 17).

**Tabela 17:** Testes com diversos modelos de substituição de códon.

| Modelos        |                  | Log( $\ell$ ) | Parâmetros                       | Sítios detectados     |
|----------------|------------------|---------------|----------------------------------|-----------------------|
|                | R0 e M0          | -1186,08      | $\omega=1639$                    |                       |
|                | R1               | -1161,05      | *                                | não aplicável         |
| Ramo 1         | R2               | -1185,77      | $\omega=0,1715$                  | não aplicável         |
| Ramo 2         | R2               | -1186,58      | $\omega=0,1715$                  | não aplicável         |
| Ramo 3         | R2               | -1184,33      | $\omega=0,1796$                  | não aplicável         |
| Ramo 4         | R2               | -1184,93      | $\omega=0,1750$                  | não aplicável         |
| Ramo 5         | R2               | -1185,28      | $\omega=0,1709$                  | não aplicável         |
| Ramo P         | R2               | -1185,72      | $\omega=0,1593$                  | não aplicável         |
| Sítios         | M1               | -1192,75      |                                  | não permitido         |
|                | M2               | -1192,70      | $p_2=0,03$ ( $\omega=4,07$ )     |                       |
|                | M3               | -1171,47      | $p_2=0,01$ ( $\omega=2,38$ )     | 7 (98%)               |
|                | M7 $\beta$       | -1172,85      | **                               |                       |
|                | M8 $\beta\omega$ | -1171,89      | $p_1=0,01$ ( $\omega=2,37$ );    | 7 (90%)               |
| Ramo e sítio 1 | MA               | -1189,05      |                                  |                       |
|                | MB               | -1171,12      |                                  |                       |
| Ramo e sítio 2 | MA               | -1187,54      |                                  |                       |
|                | MB               | -1171,13      |                                  |                       |
| Ramo e sítio 3 | MA               | -1185,01      |                                  |                       |
|                | MB               | -1171,43      |                                  |                       |
| Ramo e sítio 4 | MA               | -1188,48      |                                  |                       |
|                | MB               | -1172,27      |                                  |                       |
| Ramo e sítio 5 | MA               | -1189,97      |                                  |                       |
|                | MB               | -1172,61      |                                  |                       |
| Ramo e sítio P | MA               | -1188,20      | $P_2=0,018$<br>( $\omega=0,99$ ) | 63 (99%)              |
|                | MB               | -1168,95      | $P_2=0,024$<br>( $\omega=0,99$ ) | 23 (77%);<br>63 (99%) |

Os valores de  $\omega$  obtidos para os ramos são em sua maioria muito baixos, o que é um indicador de seleção purificadora (Yang & Bielawski, 2000; Yang, 2002). Entretanto uma parcela muito pequena dos sítios apresenta valores de  $\omega$  elevados (Tabela 17), o que pode indicar uma espécie de mosaico com alguns sítios sob

pressão seletiva e a grande maioria sob restrição funcional (Yang, 2002). A utilização dos LRTs a seguir mostra quais modelos são mais verossímeis para se tirar conclusões.

O primeiro é um teste de neutralidade feito para se verificar se existe diferença nos valores de  $\omega$  dos ramos da árvore. Neste teste, um modelo com um único  $\omega$  para todas as linhagens (R0) é testado contra um outro que estima os diferentes  $\omega$  para cada uma das linhagens (R1). Este teste demonstrou que existe diferença ( $p < 0,05$ ) entre os  $\omega$  de diferentes linhagens.

Então cada ramo de interesse na árvore foi testado para verificar se  $\omega > 1$ . Desta forma o teste é realizado utilizando o modelo (R0) contra um modelo com  $\omega_1 = 1$  para a linhagem de interesse e  $\omega_0$  para todas as outras linhagens (R2 $\omega_1$ ). Este modelo foi testado nos ramos de 1 a 5 (Figura 6) dos morcegos e no ramo dos primatas (P). Nenhum dos LRTs foi significativo, indicando que em nenhuma das linhagens o parâmetro  $\omega$  é igual ou maior do que 1. Mais uma vez no ramo 3, observa-se uma probabilidade marginal ( $p = 0,06$ ). Yang e colaboradores (2002) citam a possibilidade de probabilidades marginais como indicativa de um possível aumento no valor do  $\omega$  sem atingir o valor de 1, indicando possível seleção balanceada ou seleção fraca, neste caso os LRTs podem não ter o poder suficiente para verificar esta seleção (Yang, 2002).

Outro teste foi utilizado para a verificação de sítios com diferentes taxas de  $\omega$ . No modelo acima,  $\omega$  é considerado constante nas linhagens, isso significa que  $\omega$  é constante nos sítios e apenas a média de  $\omega$  é testada.

Vários LRTs foram realizados para testar diferentes modelos. O primeiro LRT é entre um modelo que fixa um valor de  $\omega$  igual para todos os sítios (M0) e um que institui classes diferentes de valores para  $\omega$  (ex.  $0 < \omega < 0,25$ ;  $0,25 < \omega < 0,5$ , etc; M3). O modelo discreto (M3) se adequou muito melhor aos dados do que o modelo fixo ( $p < 0,05$ ). Mais uma vez isto indica que existem diferentes pressões sobre os sítios. Este teste ainda identificou um sítio provável sob seleção forte ( $\omega \geq 1$ ) no códon 7, com probabilidade de 97%.

O segundo teste é o mais restritivo, utilizando um modelo neutro com duas classes de  $\omega$  para os sítios, uma com  $\omega = 0$  e outra com  $\omega = 1$  (M1). Este modelo é testado contra um outro que adiciona uma terceira classe de sítios com  $\omega$  estimada a partir dos dados, para se verificar se existem sítios com  $\omega > 1$  (M2). Este teste não foi significativo ( $P = 0,95$ ). Este alto valor mostra que, para o nosso conjunto de dados que  $\omega$  pode ser variável, mas não é maior do que 1.

O terceiro LRT utiliza como modelo nulo uma distribuição  $\beta$  para descrever as classes de  $\omega$  para os sítios ( $\beta$  varia entre 0 e 1). Este modelo (M7 $\beta$ ) é testado contra outro que coloca uma única classe a mais com  $\omega$  estimado dos dados (M8 $\beta\omega$ ). Este teste também não foi significativo, mas também aponta para a existência de um sítio no códon 7 como alvo de seleção com uma probabilidade de 90%.

O último teste de LRTs foi feito com os modelos de ramo e sítio (Yang *et al.*, 2002). Nestes modelos, tanto os ramos das árvores quanto os sítios podem ser variáveis. O primeiro LRT é feito com um modelo que utiliza um  $\omega_1$  para a linhagem de interesse e um outro  $\omega_0$  para todas as outras linhagens, e ainda fixa estes  $\omega$  não permitindo a existência de sítios sob seleção em todas as linhagens (MA). Este modelo pode ser testado contra um modelo neutro (M1). O modelo MA se adequou muito melhor aos dados nos ramos 1 a 4 e no ramo dos primatas (Figura 6). Isso indica que cada um destes ramos possui um  $\omega$  diferente e sítios distintos possuem diferentes classes de  $\omega$ . Apenas no ramo dos primatas foi identificado um sítio sob possível seleção no códon 63 com 99% de probabilidade. Este modelo, no entanto, não é considerado realístico (Yang *et al.*, 2002), e serve mais para prover evidência de sítios sob seleção em determinados ramos que possam ser testados com o LRT a seguir.

O modelo MB é exatamente igual ao MA, porém a distribuição de classes para os sítios é igual ao do modelo M3 com distribuição discreta das classes de  $\omega$  para os sítios, por isso ele pode ser testado diretamente contra o modelo M3 num LRT. Este modelo foi significativo apenas no ramo dos primatas, onde identificou o mesmo sítio do modelo anterior com 99% de probabilidade e um outro no códon 23 com 71% de probabilidade.

Uma mistura de alguns sítios sob seleção positiva e muitos outros sob negativa poderia explicar os dados acima, mas até que ponto isso pode estar correto?

Se observarmos a Figura 17, podemos constatar que em nenhum dos sítios indicadas como seleção (códon 7, códon 23 e códon 63) o aminoácido resultante muda radicalmente suas propriedades físico-químicas. Apenas no códon 7 existe alguma alteração de polaridade, mas esta é de um aminoácido pequeno polar para um aminoácido pequeno e neutro, uma mudança considerada branda (Creighton, 1993). Por isso é pouco provável que estes aminoácidos possam estar modificando a ação desta proteína (Goldman & Dean, 1998; Tourasse & Li, 2000). Além disso, a maioria das outras alterações observáveis também não alteram significativamente as propriedades físico-químicas do aminoácido codificado.



Na realidade, apenas algumas alterações na seqüência deduzida de aminoácidos são radicais e dignas de nota em morcegos, no entanto, as relações sinérgicas entre os aminoácidos, quando em proteína, são ainda pouco compreendidas, por isso mesmo sem alterações drásticas em propriedades físico-químicas individuais dos aminoácidos, não podemos ter certeza do seu efeito na proteína.

A primeira é uma alteração principalmente de tamanho ocorrendo apenas em Stenodermatinae na posição 57 onde se encontra um aminoácido pequeno, polar de baixa carga Glutamina (**Q**), e nos outros organismos é encontrado um aminoácido grande polar Lisina (**K**) (Arginina (**R**) em Phyllostominae).

Outra mudança semelhante é encontrada no códon 59, onde apenas *S. liliium* e *M. nigricans* possuem **Q** (pequeno polar de baixa carga) enquanto os outros organismos possuem Histidina (**H**) (grande polar).

No códon 60 primatas e roedores possuem Asparagina (**N**), um aminoácido polar, pequeno de baixa carga, *M. molossus* possui **Q**, um aminoácido semelhante, porém em *S. liliium* e *M. nigricans* encontramos Prolina (**P**), que é pequeno e neutro e nos outros morcegos encontramos **H** que é grande e polar uma mudança considerável (Creighton, 1993).

Modificação interessante também ocorre no códon 64, onde temos primatas possuindo **N** (polar de baixa carga e pequeno), roedores com **K** (polar positivo e grande) e morcegos com aminoácidos neutros e pequenos Serina (**S**) em *P. discolor* e *S. liliium* Treonina (**T**) e nos outros.

A última é na posição 77 que é um dos aminoácidos catalíticos (**N** em humanos) da  $\alpha$ -amilase. Nessa posição, apenas *S. liliium* e *P. discolor*, possuem um aminoácido neutro e pequeno (Glicina - **G**) enquanto em todos os outros organismos deste trabalho este aminoácido é pequeno de carga negativa (Ác. Glutâmico - **E**). Variações neste sítio da  $\alpha$ -amilase já foram registradas em literatura para outros organismos, indicando que este é o menos importante sítio da tríade catalítica (Janecek, 1994 e 1997), além disso alguns autores consideram glicina como possuindo carga ligeiramente negativa (Creighton, 1993).

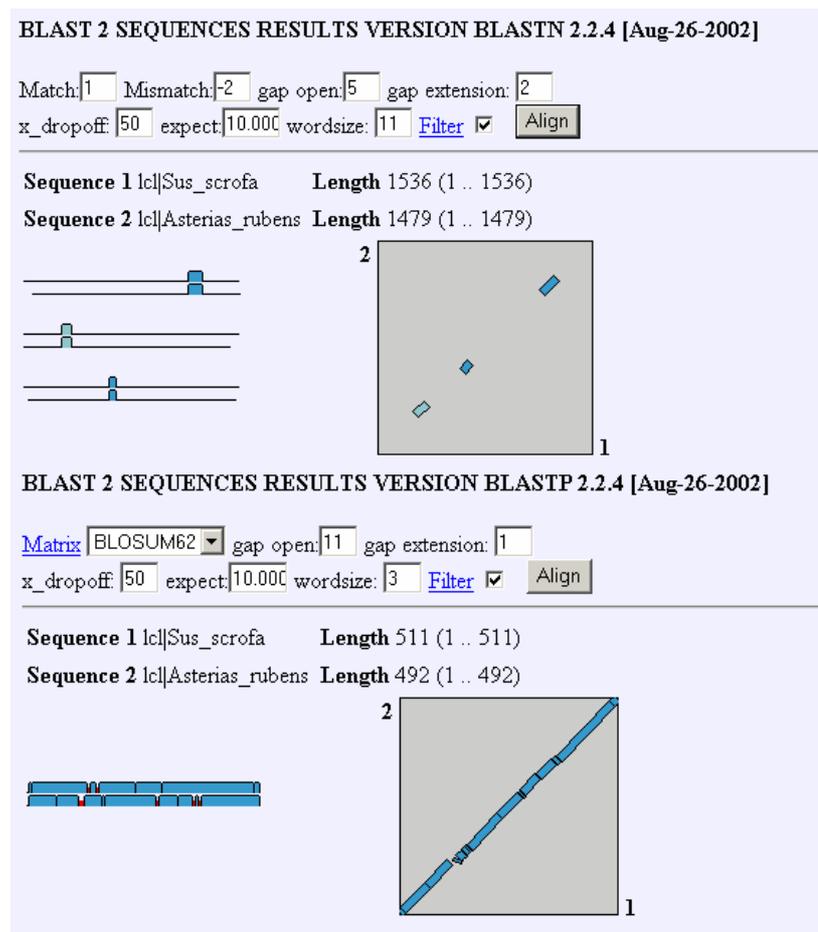
Nenhuma alteração foi detectada no outro sítio catalítico 41 **D**, nem nos sítios de ligação de cálcio 11 **N** e 45 **H**, que são essenciais para o funcionamento da  $\alpha$ -amilase (Janecek, 1994 e 1997).

Graur (1985) achou uma grande correlação entre a composição de aminoácidos de uma proteína e sua taxa de evolução. Esta correlação, no entanto, foi muito criticada e diversos trabalhos provaram exatamente o contrário (Golding & Dean, 1998; Tourasse & Li, 2000), que esta relação não tinha importância, e que a taxa de

evolução de uma proteína está muito mais relacionada à sua estrutura tridimensional (Golding & Dean, 1998; Tourasse & Li, 2000). A seleção em proteínas estaria limitada pela restrição funcional imposta pela necessidade de se manter a estrutura para exercer determinada função (Golding & Dean, 1998; Tourasse & Li, 2000).

A  $\alpha$ -amilase parece então estar sob fortes restrições estruturais, e sob forte seleção purificadora, como atestada pela maioria dos métodos utilizados. Sendo a principal enzima para a digestão de amido e sacarose (Janecek, 1994 e 1997), seu funcionamento deve ser crucial para o metabolismo primário.

A grande restrição estrutural da amilase pode ser atestada na Figura 18, onde mostramos o alinhamento de nucleotídeos e aminoácidos de *Asterias rubens*, uma estrela do mar, e de *Sus scrofa*, o porco doméstico, o mamífero com a estrutura molecular de  $\alpha$ -amilase mais bem estudada (Janecek, 1994 e 1997). As sequências de DNA são muito divergentes, mas as de aminoácidos são extremamente semelhantes, mesmo em espécies extremamente distantes filogeneticamente.



**Figura 18:** Alinhamento das sequências de  $\alpha$ -amilase de *Asterias ruber* (estrela-do-mar) e *Sus scrofa* (porco doméstico) por BLAST 2 e Dot-Plot de identidade mostrando regiões de similaridade das sequências. O gráfico superior mostra a sequência de nucleotídeos e o gráfico inferior mostra a de aminoácidos.

Os métodos utilizados para a detecção de seleção parecem responder de forma ambígua quanto ao tipo de seleção que está ocorrendo neste segmento da  $\alpha$ -amilase, indicando um mosaico com alguns poucos sítios sob seleção positiva e uma enorme quantidade de sítios sob estritas restrições funcionais, levando à seleção purificadora. Fica claro então que este segmento (e possivelmente o gene) não está evoluindo neutramente.

O tamanho do segmento analisado também pode ser o limitante destes métodos e é possível que nenhum deles tenha poder de predição bom o suficiente para uma amostra deste tamanho. O sequenciamento de uma região maior do gene (ou do gene inteiro) poderia esclarecer esta questão e ainda encontrar um maior número de sítios sob pressão seletiva, para se tentar uma busca por adaptação molecular utilizando mutagênese sítio-dirigida, como no caso das ribonucleases de artiodátilos (Jermann *et al.*, 1995), então poderíamos comparar realmente as enzimas para verificar mudanças no modo de atuação ou resistência à diferentes pHs, etc.

As discrepâncias filogenéticas encontradas (ramo 3, Figura 6) não podem ser explicadas utilizando seleção direcional ligada a adaptações ao hábito alimentar, uma vez que os dois animais envolvidos tem hábitos completamente diferentes (*S. lillium* é um frugívoro e *M. nigricans* é insetívoro). O mais provável seria ou que as substituições fossem independentemente adquiridas por *M. nigricans* e *S. lillium* por paralelismo ou que *S. lillium* exibisse retenção de alguns caracteres ancestrais (plesiomorfias).

Apesar de *S. lillium* compartilhar inclusive algumas mutações não-sinônimas radicais com *M. nigricans*, o número de diferenças em aminoácidos entre as duas espécies é o mesmo do que entre *M. nigricans* e *M. molossus* (Figura 17). As reconstruções filogenéticas do éxon 4 retirando a 3ª posição dos códons (Figura 8) gera uma politomia incluindo os três morcegos citados acima e ainda *Platyrrhinus lineatus*, *Phyllostomus hastatus* e *P. discolor*, mostrando que a maioria das diferenças que agrupam o ramo 3 da Figura 6 estão na 3ª posição do códon, que é a mais sujeita a alterações (Li, 1997; Nei & Kumar, 2000), logo a possibilidade de evolução paralela ao acaso para este segmento do gene nas duas espécies é aceitável.

Entretanto sabe-se que *S. lillium* é considerado basal na filogenia de Stenodermatinae (Gannon *et al.*, 1989; Simmons & Geisler, 1998), este inclusive possui características cranianas muito diferentes dos outros membros da sua subfamília, e sua relação caixa craniana - cérebro é muito diferente até mesmo comparando com outros Phyllostomidae (Gannon *et al.*, 1989). A retenção de

características plesiomórficas é sempre uma possibilidade, mas praticamente impossível de ser comprovada a não ser em uma filogenia bem estabelecida.

Para verificar se os dados observados para *S. lillium* e/ou *M. nigricans* estão relacionados à uma possível retenção de caracteres, o sequenciamento de um maior número de espécies de Phyllostomidae e Vespertilionidae poderia ser esclarecedor, assim como o sequenciamento de uma região maior do gene. Outra medida seria a utilização de espécies de grupos irmãos de Phyllostomidae, como Noctilionidae e Mystacinidae e também de Phyllostomidae basais, como *Macrotus* spp. Caso se verifique a ocorrência das características observadas em outros animais, a retenção de caracteres passa a ser uma situação mais provável.

Uma outra possibilidade é que duplicações independentes e posterior deleção de cópias parálogas dos genes da  $\alpha$ -amilase poderiam estar desviando as análises da filogenia correta. Duplicações independentes dos genes da  $\alpha$ -amilase já foram descritas (Samuelson *et al.*, 1988, 1990 e 1996), e é uma possibilidade, apesar de não se encontrar evidência direta da presença de dois genes de  $\alpha$ -amilase em morcegos neste trabalho.

As diferenças encontradas nas reconstruções filogenéticas utilizando o éxon 4, o íntron C e ambos segmentos (Figura 6, 9, 13 e 14) podem estar indicando taxas de evolução diferente para os segmentos deste gene em *S. lillium*. Outro exemplo pode ser observado indiretamente nos resultados das ampliações de alguns dos outros animais, como por exemplo *Desmodus rotundus*. A não amplificação destes animais quando se utiliza iniciadores que se acoplam ao éxon 3 pode ser um indício de um grande acúmulo de mutações.

A região estudada da  $\alpha$ -amilase para mamíferos não é de utilidade para a resolução de todos grandes grupos (Figura 13). No entanto esta se mostrou útil na definição das relações entre morcegos, roedores e primatas, dando suporte às classificações recentes do grupo Euarchontoglires, contra a existência de Archonta (Figuras 14 e 15).

O seqüenciamento de um segmento maior do gene da  $\alpha$ -amilase compreendendo regiões codificadoras e não codificadoras poderia dar uma maior resolução às reconstruções filogenéticas, tanto para o grupo dos morcegos, como para as relações entre mamíferos.

## Conclusões

1. Foram padronizados protocolos de amplificação e seqüenciamento da  $\alpha$ -amilase para várias espécies de morcegos e de outros mamíferos.
2. A região estudada do gene da  $\alpha$ -amilase está sob grande influência de seleção purificadora, provavelmente relacionada a limitações estruturais da sua importância funcional e catalítica, mesmo assim em alguns sítios parece estar ocorrendo seleção natural não direcional que pode ser devido provavelmente a um relaxamento das restrições funcionais naqueles sítios.
3. A seleção natural detectada não parece relacionada à diversidade de hábitos alimentares nos morcegos estudados, por isso pode não ser adaptativa.
4. As reconstruções filogenéticas utilizando o segmento codificador do gene da  $\alpha$ -amilase diferem das filogenias tradicionais de morcegos quanto ao posicionamento da espécie *Sturnira lilium*, que se agrupa com animais de famílias distintas e de diferentes hábitos alimentares. Além disto as taxas evolutivas entre o íntron C e o éxon 4 da  $\alpha$ -amilase são diferentes também nesta espécie.
5. As possíveis causas para as discrepâncias nas reconstruções filogenéticas envolvem retenção de caracteres plesiomórficos, paralelismo evolutivo ou erro de amostragem por utilização de cópia paráloga do gene.
6. A região estudada da  $\alpha$ -amilase porém pode ser útil na resolução das relações filogenéticas entre as ordens Primata, Rodentia e Chiroptera, quando utilizado um grupo externo filogeneticamente distante. Sugerindo uma relação mais próxima entre primatas e roedores do que primatas e morcegos, dando apoio à classificação dos grupos Euarchontoglires e Laurasiatheria em mamíferos

## Referências Bibliográficas

- Adkins, R.M.; & Honeycutt, R.L. 1991. Molecular phylogeny of the superorder Archonta. **Proceedings of the National Academy of Science U.S.A.** **88**: 10317-10321.
- Aguiar, L. M. S. 1994. **Comunidades de Chiroptera em três áreas de Mata Atlântica em diferentes estádios de sucessão - Estação Biológica de Caratinga, Minas Gerais**. Dissertação de Mestrado, Universidade Federal de Minas Gerais, Belo Horizonte. Brazil
- Allard, M.W.; McNiff, B.E. & Miyamoto, M.M. 1996. Support for interordinal eutherian relationships, with an emphasis on Primates and their archontan relatives. **Molecular Phylogenetics and Evolution** **5**: 78-88.
- Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research** **25**:3389-3402.
- Ammerman, L.K., & Hillis, D.M. A molecular test of bat relationships: monophyly or diphyly? **Systematic Biology** **41**: 222-232.
- Amorin, D.S. 1997. **Elementos básicos de sistemática filogenética**. Holos editora e Sociedade Brasileira de Entomologia. Ribeirão Preto BR.
- Anisimova, M.; Bielawski, J.P. & Yang, Z. 2001. The accuracy and power of likelihood ratio tests to detect positive selection at amino acid sites. **Molecular Biology and Evolution** **18**: 1585-1592.
- Anisimova, M.; Bielawski, J.P. & Yang, Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. **Molecular Biology and Evolution** **19**: 950-958.
- Arnason, U.; Adegoke, A.J.; Bodin, K.; Born, E.W.; Esa, Y.B.; Gullberg, A.; Nilsson, M.; Short, R.V.; Xu, X & Janke, A. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. **Proceedings of the Natural Academy of Sciences U.S.A.** **99**(12): 8151-8156.
- Bailey, W.J.; Slightom, J.L. & Goodman, M. 1992. Rejection of the "flying primate" hypothesis by phylogenetic evidence from the e-globin gene. **Science** **256**: 86-89.
- Creevey, C. and McInerney, J. O. 2002. An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. **Gene** **300**: 43-51.

- Creighton, T.E. 1993. **Proteins: Structures and Molecular Properties**. 2.ed. W.H. Freeman publisher, New York, USA.
- Darnis, S.; Juge, N.; Guo, X.; Marchis-Mouren, G.; Puigserver, A. & Chaix, J. 1999. Molecular cloning and primary structure analysis of porcine pancreatic  $\alpha$ -amylase. **Biochimica et Biophysica Acta** **1430**: 281-289.
- de Jong, W.W. 1998. Molecules remodel the mammalian tree. **Trends in Ecology and Evolution** **13**(7): 270-275.
- Deusuc, F.; Scally, M.; Madsen, O.; Stanhope, M.J.; de Jong, W.W.; Catzeflis, F.M.; Springer, M.S. & Douzery, E.J.P. 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. **Molecular Biology and Evolution** **19**(10): 1656-1671
- Ditchfield, A.D. & Burns, K. 1998. DNA sequences reveal phylogeographic similarities of neotropical bats and birds. **Journal of Comparative Biology** **3**(2):165-170.
- Ditchfield, A.D. 2000. The comparative phylogeography of Neotropical mammals: patterns of intraspecific mitochondrial DNA variation among bats contrasted to nonvolant small mammals. **Molecular Ecology** **9**: 1307-1318.
- Endo, T.; Ikeo, K. & Gojobory, T. 1996. Large scale search for genes on which positive selection may operate. **Molecular Biology and evolution** **13**: 685-690.
- Engel, W. 1992. **Amplify 1.2**. Department of Genetics, University of Wisconsin.
- Ewing, B. & Green, P. 1998. Basecalling of automated sequencer traces using Phred II. Error probabilities. **Genome Research** **8**: 186-194.
- Ewing, B.; Hillier, L.; Wendl, M. & Green, P. 1998. Basecalling of automated sequencer traces using Phred I: Accuracy assesment. **Genome Research** **8**: 175-185.
- Faria, D.M. & Sazima, I. 1995. Food resources utilization by a frugivorous-nectarivorous bat guild in Southeastern Brazil. **Chiroptera Neotropical** **1**(2): 31.
- Fay, J.C.; Wu, C.-I. 2001. The neutral theory in the genomic era. **Current Opinion in Genetics and Development** **11**: 642-646.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. **Journal of Molecular Evolution** **17**: 368-376.
- Ferarezzi, H. & Gimenez, E.A. 1996. Systematic patterns and the evolution of feeding habits in Chiroptera (Archonta: Mammalia). **Journal of Comparative Biology** **1**: (3-4): 75-94.

- Fonseca, G.A.B.; Herrmann, G.; Leite, Y.L.R.; Mittermeier, R.A.; Rylands, A.B. & Patton, J.L. 1996. Lista Anotada dos Mamíferos do Brasil. **Occasional Papers in Conservation Biology n4**. Conservation International.
- Gannon, M.R.; Willig, M.R. & Knox-Jones, J. *Sturnira lilium*. **Mammalian Species 333**: 1-5.
- Gilbert, D.G. 1997. **SeqPup 9**: Sequence editor and analyser. Biology Department, Indiana University.
- Gogarten, J.P. & Olendzenski, L. 1999. Orthologs, paralogs and genome comparisons. **Current Opinion in Genetics and Development 9**: 630-636.
- Golding, G.B. & Dean, A.M. 1998. The structural basis of molecular adaptation. **Molecular Biology and Evolution 15**(4): 355-369.
- Goldman, N. & Yang, Z. 1994. A codon-based model of nucleotide substitution for protein coding DNA sequences. **Molecular Biology and Evolution 11**(5): 725-736.
- Gordon, D.; Abajian, C. & Green, P. 1998. Consed: A graphical tool for sequence finishing. **Genome Research 8**: 195-202.
- Graur, D. 1985. Amino acid composition and the evolutionary rates of protein-coding genes. **Journal of Molecular Evolution 22**: 53-62.
- Green, P. 1994. **Phrap**. [www.genome.washington.edu/UWGC/analysis\\_tools/phrap.htm](http://www.genome.washington.edu/UWGC/analysis_tools/phrap.htm)
- Groot, P.C.; Mager, W.H.; Henriquez, N.V.; Pronk, J.C.; Arwert, F.; Planta, R.J.; Eriksson, A.W. & Frants, R.R. 1990. Evolution of the human alpha-amylase multigene family through unequal, homologous, and inter- and intrachromosomal crossovers. **Genomics 8**: 97-105.
- Gumucio, D.L.; Wiebauer, K.; Caldwell, R.M.; Samuelson, L.C.; Meisler, M.H. Concerted evolution of human amylase genes. **Molecular and Cellular Biology 8**: 1197-1205.
- Horii, A.; Emi, M.; Tomita, N.; Nishide, T.; Ogawa, M.; Mori, T. Matsubara, M. 1987. Primary structure of human pancreatic  $\alpha$ -amylase gene: its comparison with the human salivary  $\alpha$ -amylase gene. **Gene 60**: 57-64.
- Hasegawa, M. & Fujiwara, M. 1993. Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogeny. **Molecular Phylogenetics and Evolution 2**: 1-5.
- Hennig, W. 1966. Phylogenetic systematics. **Annual Reviews of Entomology 10**: 97-116.

- Huelsenbeck, J.P. & Ronquist, F. 2001. MrBayes: Bayesian inference of phylogenetic trees. **Bioinformatics** **17**: 754-755.
- Huelsenbeck, J.P.; Ronquist, F.; Nielsen, R. & Bollback, J.P. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. **Science** **294**: 2310-2314.
- Irwin, D.M. 1995a. Evolution of bovine lysozyme gene family: Changes in gene expression and reversion of function. **Journal of Molecular Evolution** **41**: 299-312.
- Irwin, D.M. 1995b. Molecular Evolution of ruminant lysosymes. In: Jollès, P (Ed). **Lysozymes, Model Enzymes in Biochemistry and Biology**. Birkhäuser - Verlag, Basel.
- Janecek, S. 1994. Sequence similarities and evolutionary relationships of microbial, plant and animal  $\alpha$ -amylases. **European Journal of Biochemistry**. **224**: 519-524.
- Janecek, S. 1997.  $\alpha$ -amylase family: molecular biology and evolution. **Programs in Biophysics and Molecular Biology** **67**(1): 67-97.
- Jeanmougin, F.; Thompson, J.D.; Gouy, M.; Higgins, D.G. & Gibson, T.J. 1998. Multiple sequence alignment with Clustal X. **Trends In Biochemical Science** **23**: 403-405.
- Jepsen, G.L. 1977. Bat origins and Evolution. In. **Biology of the bats** v.I. W.A. Wimsatt (Ed). Academic Press: New York, London.
- Jermann, T.M.; Opitz, J.S. & Benner, S.A. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. **Nature** **374**: 57-59.
- Junqueira, L.C.U., Toledo, A. M. S. & Doine, A. I. 1973. Digestive enzymes in the parotid and submandibular glands of mammals. **Anais da Academia Brasileira Ciências** **45**(3/4): 629-643.
- Kalendar, R. 2001. **Oligos 8.72**. Institute of Biotechnology, University of Helsinki.
- Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. **Journal of Molecular Evolution** **16**: 111-120.
- Kimura, M. 1983. **The neutral theory of molecular evolution**. Cambridge University Press. Cambridge UK.
- Kirsch, J.A.W. 1996. Bats are monophyletic; megabats are monophyletic; but are microbats also? **Bat Research News** **36**: 78.

- Koopman, K. F. 1993. Chiroptera. In **Mammal Species of the World**. D.E. Wilson and D.M. Reeder (Eds.) Smithsonian Institution Press, Washington USA.
- Kordis, D. & Gubensek, F. 2000. Adaptive evolution of animal toxin multigene families. **Gene** **261**: 43-52.
- Kumar, S.; Tamura, K.; Jakobsen, I.B. & Nei, M. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. **Bioinformatics** **17**: 12:1244-1245.
- Li, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. **Journal of Molecular Evolution** **36**: 96-99.
- Li, W.-H. 1997. **Molecular Evolution**. Sinauer Associates Inc., Massachusetts USA.
- Li, W.-H.; Wu, C.I. & Luo, C.C. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likelihood of nucleotide and codon changes. **Molecular Biology and Evolution** **2**(2): 150-174.
- MacDonald, R.J.; Crerar, M.M.; Swain, W.F.; Pictet, R.L.; Thomas, G. & Rutter, W.J. 1980. Structure of a family of rat amylase genes. **Nature** **287**: 117-122
- Madsen, O.; Scally, M.; Douardy, C.J.; Kao, D.J.; DeBry, R.W.; Adkins, R.M.; Amrine, H.M.; Stanhope, M.J.; de Jong, W.W. & Springer, S. 2001. Parallel adaptive radiations in two major clades of placental mammals. **Nature** **409**: 610-614.
- McDonald, J. H. & Kreitman, M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. **Nature** **351**: 652-654.
- Messier, W. & Stewart, C. B. 1997. Episodic adaptive evolution of primate lysozymes. **Nature** **385**: 151-154.
- Mindell, D.P.; Dick, C.W. & Baker, R.J. 1991. Phylogenetic relationships among megabats, microbats, and primates. **Proceedings of the National Academy of Sciences U.S.A.** **88**(22): 10322-10326.
- Murphy, W.J.; Eizirik, E.; Johnson, W.E.; Zhang, Y.P.; Ryder, O.A. & O'Brien, S.J. 2001. Molecular phylogenetics and the origin of placental mammals. **Nature** **409**:614-618.
- Myiaki, C.Y; Russo, C.A.M & Pereira, S.L. 2001. Reconstrução filogenética: Introdução e o método da máxima parcimônia. In **Biologia molecular e evolução**. S.R. Mاتيoli (Ed) Holos editora. Ribeirão Preto BR.
- Nei, M. & Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. **Molecular Biology and Evolution** **3**:418-426.

- Nei, M. & Kumar, S. 2000. **Molecular Evolution and Phylogenetics**. Oxford University Press, New York USA.
- Nickerson, D.A. Tobe, V.O. & Taylor, S.L. 1997. Polyphred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. **Nucleic Acids Research** **25**: 2745-2751.
- Nielsen, R. 2001. Statistical tests of selective neutrality in the age of genomics. **Heredity** **86**: 641-647.
- Nikaido, M.; Harada, M.; Cao, Y.; Hasegawa, M. & Okada, N. 2000. Monophyletic origin of the Order Chiroptera and its phylogenetic position among Mammalia, as inferred from the complete sequence of the mitochondrial DNA of a Japanese megabat, the Ryukyu flying fox (*Pteropus dasymallus*). **Journal of Molecular Evolution** **51**: 318-328.
- Nishide, T.; Emi, M.; Nakamura, Y. & Matsubara, K. 1986b. Corrected sequences of cDNAs for human salivary and pancreatic alpha-amylases. **Gene** **50**(2): 371-372
- Nishide, T.; Nakamura, Y.; Emi, M.; Yamamoto, T.; Ogawa, M.; Mori, T. & Matsubara, K. 1986a. Primary structure of human salivary alpha-amylase gene. **Gene** **41**(2-3): 299-304.
- Noronha, R.C.R.; Rodrigues, L.R.R.; SILVA, A. M.; Marques-Aguiar, S.A.; Nagamachi, C.Y.; Pieczarka, J.C. & Barros, R.M.S. 1998. Meiose e cromossomos sexuais de duas espécies de Quirópteros da Amazônia (Phyllostomidae). **Genetics and Molecular Biology**. **21** (3): 74.
- Novacek, M.J. 1992. Mammalian phylogeny: shaking the tree. **Nature** **256**: 121-125.
- Nowak, R.N. 1997. **Walker's Mammals of the World on line 5.1**. Johns Hopkins University Press, London. [Http://www.jhu.edu/books/walker](http://www.jhu.edu/books/walker).
- Pereira, S.L.; Myiaki, C.Y. & Russo, C.A.M. 2001. Reconstruções filogenéticas: Métodos probabilísticos. In. **Biologia molecular e evolução**. S.R. Matioli (Ed) Holos editora. Ribeirão Preto BR.
- Pettigrew, J.D. 1986. Flying primates? Megabats have the advanced pathway from eye to midbrain. **Science** **231**:1304-1306.
- Phillips, C.J. & Tandler, B. 1996. Salivary glands, cellular evolution, and adaptive radiation in mammals. **European Journal of Morphology** **34**(3):155-161.
- Phillips, C.J.; Tandler, B. & Pinkstaff, C.A. 1987. Unique salivary glands in two genera of tropical microchiropteran bats: an example of evolutionary convergence in histology and histochemistry. **Journal of Mammalogy** **68**(2):235-242.

- Phillips, C.J.; Weiss, A. & Tandler, B. 1998. Plasticity and patterns of evolution in mammalian salivary glands: comparative immunohistochemistry of lysozyme in bats. **European Journal of Morphology** **36**(supplement):19-26.
- Rannala, B. & Yang, Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. **Journal of Molecular Evolution** **43**: 304-311.
- Rooney, A.P. & Zhang, J. 1999. Rapid evolution of a primate sperm protein: relaxation of functional constraints or positive selection?. **Molecular Biology and Evolution** **16**: 706-710.
- Russo, C.A.M. 1997. Efficiencies of different statistical tests in supporting a known vertebrate phylogeny. **Molecular Biology and Evolution** **14**(10): 1078-1080.
- Russo, C.A.M.; Takezaki, N. & Nei, M. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. **Molecular Biology and Evolution** **13**(3): 525-536.
- Russo, C.A.M.; Myiaki, C.Y. & Pereira, S.L. 2001. Reconstrução filogenética: Métodos geométricos. In. **Biologia molecular e evolução**. S.R. Mاتيoli (Ed) Holos editora. Ribeirão Preto BR.
- Saitou, N. & Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution** **4**:406-425.
- Sambrook, J.; Russell, D.W. & Sambrook, J. 2001. **Molecular Cloning: A Laboratory Manual**. 3.Ed. CSHL press, New York USA.
- Samuelson, L.C.; Phillips, R.S. & Swanberg, L.J. Amylase gene structures in Primates: retroposon insertions and promoter evolution. **Molecular Biology and Evolution** **13**(6): 767-779.
- Samuelson, L.C.; Wiebauer, K.; Gumucio, D.L. & Meisler, M.H. 1998. Expression of the human amylase genes: recent origin of a salivary amylase promoter from an actin pseudogene. **Nucleic Acid Research** **16**: 8261-8275.
- Samuelson, L.C.; Wiebauer, K.; Snow, C.M. & Meisler, M.H. 1990. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. **Molecular and Cellular Biology** **10**: 2513-2520.
- Santos, F.R.; Guimarães, P.E.M. & Redondo, R.A.F. 2002. Bancos de DNA: coleções estratégicas para estudos da biodiversidade. **Lundiana: International Journal Of Biodiversity** **3**(2): 93-98.

- Sazima, M. & Sazima, I. 1975. Quiropterofilia em *Lafoensia pacari* St. Hil. (Lythraceae), na Serra do Cipó, Minas Gerais. **Ciência e Cultura** **27**(4): 405-416.
- Schibler, U.; Pittet, A.C.; Young, R.A.; Hagenbuchle, O.; Tosi, M.; Gellman, S. & Wellauer, P.K. 1982. The mouse alpha-amylase multigene family. Sequence organization of members expressed in the pancreas, salivary gland and liver. **Journal of Molecular Biology** **155**(3): 247-266.
- Silva, A.M.; Rodrigues, L.R.R.; Barros, R.M.S.; Assis, M.F.L.; Marques-Aguiar, S.; Pieczarka, J.C. & Nagamachi, C.Y. 1997. Estudo citogenético em *Artibeus jamaicensis* (Phyllostomidae: Chiroptera) e comparação com *A. lituratus*. **Revista Brasileira de Genética** **20** (3):74.
- Silva, A.M.; Rodrigues, L.R.R.; Barros, R.M.S.; Marques-Aguiar, S.A.; Nagamachi, C.Y. & Pieczarka, J.C. 1998. Descrição cariotípica da espécie *Uroderma magnirostrum* (Chiroptera: Phyllostomidae). **Genetics and Molecular Biology**. **21** (3): 75.
- Simmons, N.B. & Geisler, J.H. 1998. Phylogenetic relationships of *Icaronycteris*, *Archeonycteris*, *Hassianycteris*, and *Palaeochiropteryx* to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in microchiroptera. **Bulletin of the American Museum of Natural History**. **235**:1-182.
- Skibinski, D.O.F. 2000. DNA tests of neutral theory: application in marine genetics. **Hydrobiologia** **420**: 137-152.
- Sneath, P.H.A. & Sokal, R.R. 1973. **Numerical Taxonomy**. W.H. Freeman publisher, San Francisco USA.
- Springer, M.S.; Teeling, E. & Stanhope, M.J. 2001. External Nasal Cartilages in Bats: Evidence for Microchiropteran Monophyly?. **Journal of mammalian evolution** **8**(3): 231-236.
- Strahler, J.R & Meisler, M.H. 1982. Two distinct pancreatic amylase genes are active in YBR mice. **Genetics** **101**(1):91-102.
- Su, C.; Nguyen, V.K. & Nei, M. Adaptative evolution of variable region genes encoding an unusual type of immunoglobulin in camelids. **Molecular Biology and Evolution** **19**(3): 205-215.
- Swofford, D.L. 1998. **PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods)**. Version 4.08b. Sinauer Associates Inc, Massachusetts USA.

- Tamura, K. and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. **Molecular Biology and Evolution** **10**:512-526.
- Tanaka, T. & Nei, M. 1989. Positive darwinian selection observed at the variable region genes of immunoglobulins. **Molecular Biology and Evolution** **6**(3): 447-459.
- Tandler, B.; Phillips, C.J. & Nagato, T. 1996. Histological convergent evolution of the accessory submandibular glands in four species of frog-eating bats. **European Journal of Morphology** **34**(3): 163-168.
- Tatusova, T.A. & Madden, T.L. 1999. Blast 2 sequences - a new tool for comparing protein and nucleotide sequences. **F.E.M.S. Microbiology Letters** **174**: 247-250
- Tavares, V.C. & Anciães, M. 1998. Artificial Roosts and Diet of Some Insectivorous Bats in the Parque Estadual do Rio Doce, Brazil. **Bat Research News** **39**(3):142.
- Tavares, V.C. 1999a. **Ecomorfologia do vôo, dieta das espécies e composição de uma taxocenose de morcegos (Mammalia: Chiroptera) do Parque Estadual do Rio Doce, Leste de Minas Gerais, Sudeste do Brasil.** Dissertação de Mestrado. Belo Horizonte, Universidade Federal de Minas Gerais. 111 pp.
- Tavares, V.C. 1999b. Flight Morphology, diet, and composition of a bat assemblage (Mammalia: Chiroptera) at the Rio Doce state Park, Southeast Brazil. **Chiroptera Neotropical** **5**(1-2): 117-118.
- Thompson, J.D.; Higgins, D.G. & Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. **Nucleic Acids Research** **22**:4673-4680.
- Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F. & Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Research** **24**:4876-4882.
- Tourasse, N.J. & Li, W-H. 2000. Selective constraints, amino acid composition and the rate of protein evolution. **Molecular Biology and Evolution** **17**(4): 656-664.
- Uieda, W. 1992. Período de atividade alimentar e tipos de presas dos morcegos hematófagos (Phyllostomidae) no sudeste do Brasil. **Revista Brasileira de Biologia** **52**(4): 563-573.

- Vacquier, V.D.; Swanson, W.J. & Lee, W.-H. 1997. Positive darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence?. **Journal of Molecular Evolution** **44**(supplement): 15-22.
- Wiebauer, K; Gumucio, D.L.; Jones, J.M.; Caldwell, R.M.; Hartle, H.T.; Meisler M.H., 1985. A 78-kilobase region of mouse chromosome 3 contains salivary and pancreatic amylase genes and a pseudogene. **Proceedings of the Natural Academy of Sciences U.S.A.** **82**(16):5446-5449.
- Wilson, D.E. & LaVal, R.K. 1974. *Myotis nigricans*. **Mammalian Species** **39**: 1-3
- Yang, Z. & Bielawski, J.P. 2000. Statical methods for detecting molecular adaptation. **Trends in Ecology and Evolution** **15**(12): 496-503.
- Yang, Z. & Nielsen, R. 1997. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. **Journal of Molecular Evolution** **46**: 409-418.
- Yang, Z. & Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. **Molecular Biology and Evolution** **17**: 32-43.
- Yang, Z. & Rannala, B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. **Molecular Biology and Evolution** **14**(7): 714-724.
- Yang, Z. 1996. Phylogenetic analysis using parsimony and likelihood methods. **Journal of Molecular Evolution** **42**: 294-307.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood **CABIOS** **13**:555-556.
- Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to Primate lysozyme evolution. **Molecular Biology and Evolution** **15**(5): 568-573.
- Yang, Z. 2000. Complexity of the simplest phylogenetic estimation problem. **Proceedings of the Royal Society of London B.** **267**: 109-116.
- Yang, Z. 2002. Inference of selection from multiple species alignments. **Current Opinion in Genetics and Development** **12**: 688-694.
- Yang, Z.; Swanson, W.J. & Vacquier, V.D. 2000. Maximum likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. **Molecular Biology and Evolution** **17**: 1446-1455.
- Yang, Z; Kumar, S. & Nei, M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. **Genetics** **141**: 1641-1650.

- Zanotto, P.M.; Kallas, E.G.; de Souza, R.F. & Holmes, E.C. 1999. Genealogical evidence of positive selection in the *nef* gene of HIV-1. **Genetics** **153**: 1077-1089.
- Zhang, J. & Kumar, S. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. **Molecular Biology and Evolution** **14**(5): 527-536.
- Zhang, J. & Nei, M. 2000. Positive selection in the evolution of mammalian interleukin-2 genes. **Molecular Biology and Evolution** **17**(9): 1413-1416.
- Zhang, J.; Kumar, S. & Nei, M. 1997. Small-sample tests of episodic adaptive evolution: a case of Primate lysozymes. **Molecular Biology and Evolution** **14**(12): 1335-1338.
- Zhang, J.; Rosenberg, H.F. & Nei, M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. **Proceedings of the National Academy of Sciences USA** **95**:3708-3713.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)