Genetica (2006) 126:199–213
DOI 10.1007/s10709-005-1449-9

Evolutionary studies on an α-amylase gene segment in bats and other mammals

Rodrigo A.F. Redondo* & Fabrício R. Santos
*Laboratório de Biodiversidade e Evolução Molecular, Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, C.P. 486, 31.270-010, Belo Horizonte, Minas Gerais, Brazil; *Author for correspondence (Phone: +55-31-3499-2581; Fax: +55-31-3499-2570; E-mail: redondo@icb.ufmg.br)*

Key words: α-amylase, bat diversity, mammal order phylogeny, natural selection tests

Abstract

Comparative studies of salivary glands showed that they maybe related to the adaptive radiation of bats, especially in the family Phylostomidae. In this study we have been searching for a likely relationship between different feeding habits found in bats and possible adaptive changes in a coding segment of the α-amylase enzyme. We have also tested some hypothesis about the phylogenetic relationship of bats and other mammals. A 663 bp segment of the α-amylase gene, corresponding to the exon 4 and part of the intron c, was sequenced in nine bat species. The exon 4 was also sequenced in further ten mammalian species. The phylogenetic trees generated with different methods produced the same results. When the intron c and the exon 4 were independently analyzed, they showed distinct topologies involving the bat species *Sturnira lilium*, different from the traditional bat phylogeny. Phylogenetic analysis of bats, primates and rodents supports the Euarchontoglires-Laurasiatheria hypothesis about the relationship among these groups. Selection tests showed that the α-amylase exon 4 is under strong purifying selection, probably caused by functional constraints. The conflicting bat phylogenies could not be explained by evolutionary convergence due to adaptive forces, and the different topologies may be likely due to the retention of plesiomorphic characters or the independent acquisition by evolutionary parallelism.

Introduction

The Chiroptera is the second most diverse order of mammals in number of species, encompassing more than 925 species occurring worldwide (Nowak, 1999). Classically the Chiroptera belongs to the Superorder Archonta, that includes Primates, Dermoptera (flying lemurs) and Scandentia (tree shrews) (Novacek, 1992), but the traditional view of mammalian phylogeny has been challenged by recent molecular evidence from nuclear and mitochondrial data (Nikaido et al., 2000; Madsen et al., 2001; Murphy et al., 2001; Arnason et al., 2002; Deusuc et al., 2002; Springer et al., 2004).

These recent studies give support to a very different and more geographically correlated classification, besides it is less clear at first look. In this new mammalian tree, the bats are grouped within the Laurasiatheria together Carnivora, Pholidota, Perissodactyla, Cetartiodactyla (Artiodactyla plus Cetacea) and Soricomorpha. Primates, Dermoptera and Scandentia still cluster together making up a sister group with Rodentia and Lagomorpha within the Euarchontoglires (Springer et al., 2004).

Although the most recent molecular data support the above views of the mammal tree, there are still some molecular (and a bulk of morphological) data supporting the traditional classification (Adkins & Honeycutt, 1991; Allard, McNiff & Miyamoto, 1996).

About 270 bat species live in the Neotropics, most of these species (143) belongs to the

Phyllostomidae family (Nowak, 1999). Phyllostomid bats exhibit amazing adaptations and accounts for the most radiated mammal group in terms of feeding habits, including omnivory, frugivory, nectarivory, insectivory/carnivory (including piscivorous and frog eating species) and the famous blood eating vampire bats (Ferarezzi & Gimenez, 1996; Nowak, 1999).

Comparative studies using salivary glands in mammals have previously suggested that their evolution was correlated with the adaptive radiation of these animals (Phillips & Tandler, 1996). The same authors also found unique modifications on bat's salivary glands and showed that these features were more correlated to their feeding habits than to their phylogenetic relationships, suggesting convergent evolution (Phillips, Tandler & Pinkstaff, 1987; Tandler, Phillips & Nagato, 1996). Immunohistochemistry assays and analysis of lysozyme production in bats showed the same interesting pattern (Phillips, Weiss & Tandler, 1998). Some previous studies have also suggested that in frugivorous mammalian species, the hydrolysis of starch (and other polysaccharides) in the mouth helps the identification of more nutritive food resources and a salivary α-amylase would be favored by natural selection (Ting et al., 1992).

However, molecular evolutionary studies of digestive enzymes in bats have not been found in literature, and it seems possible that the diversity of feeding resources used by bats could be correlated with a similar diversification of their digestive enzymes.

The α-amylase (1,4-α-D-glucan glucanohydrolase E.C. 3.2.1.1) acts on the hydrolysis of 1,4-α-glucosidic linkages in oligosaccharides and polysaccharides (as starch and glycogen) in a random manner, reducing the size of these molecules (Janecek, 1997).

In mammals, the primary structure of the α-amylase have 496 amino acids and the catalytic domain is composed by a triad of small and polar amino acids, Aspartic acid, Glutamic acid and Aspartic acid, and it is dependent of calcium binding sites (Janecek, 1994, 1997). Most of these features are found in the exon 4 (two catalytic sites and two calcium binding sites).

Duplication of the α-amylase gene have been found in some mammalian species (as primates and rodents) but are thought to be independent duplication events since their promoters are unrelated (Samuelson, Phillips & Swanberg, 1988; Samuelson et al., 1990). The duplicated copies are usually maintained extremely similar by concerted evolution and gene conversion (Gumucio et al., 1988).

In this study we have been searching for possible adaptive changes in a coding segment of the α-amylase enzyme that is likely correlated to feeding habits. Besides, we have tested some hypothesis about the phylogenetic relationship between bats and other mammals.

Materials and methods

Samples

Nine bat species were collected with mist-nets in distinct areas of the Atlantic Rain forest in Southeast Brazil (Table 1). DNA was extracted from liver tissue by standard phenol-chloroform-isoamyl-alcohol method (Sambrook, Russell & Sambrook, 2001).

Samples from 10 other mammalian species (Table 1) used in this study were obtained in scientific collections. All DNA samples were kept at −20°C in the DNA repository of our laboratory (Santos, Guimarães & Redondo, 2002).

DNA amplification

PCR reactions were carried out in 15 μl reactions containing ~40 ng of genomic DNA, 1X reaction buffer 1B (Phoneutria® −1.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.4), 50 mM KCl, 0.1% Triton X-100), 200 μM dNTP mix (Amersham-Biosciences®), 0.5 μM of each primer (Table 2) and 1.25 U of Taq DNA polymerase (Phoneutria®). PCR cycling was conducted with a first denaturing step at 94°C for 5 min, 35 cycles at 94°C for 30 s, annealing temperature (Table 2) for 30s and extension at 72°C for 80 s, and a final extension step at 72°C for 5 min. PCR products were checked on ethidium-bromide stained agarose gels.

The primers (Table 2, Figure 1) were designed using sequences of α-amylase found in GenBank: *Homo sapiens* (gi537511), *Sus scrofa* (gi6056337), *Mus musculus* (gi6996908) and *Rattus norvergicus*

Table 1. Species studied and source of the sequences

Species	Order	Sequence
Equus caballus (Linnaeus, 1758)	Perissodactyla	This study
Canis familiaris (Linnaeus, 1758)	Carnivora	This study
Cerdocyon thous (Linnaeus, 1766)	Carnivora	This study
Conepatus semistriatus (Boddaert, 1785)	Carnivora	This study
Eira barbara (Linnaeus, 1758)	Carnivora	This study
Bos taurus (Linnaeus, 1758)	Cetartiodactyla	This study
Sus scrofa (Linnaeus, 1758)	Cetartiodactyla	GenBank (AF064742.1)
Artibeus lituratus (Olfers, 1818)	Chiroptera*	This study
Chiroderma villosum (Peters, 1860)	Chiroptera*	This study
Molossus molossus (Pallas, 1766)	Chiroptera***	This study
Myotis nigricans (Schinz, 1821)	Chiroptera****	This study
Phyllostomus discolor (Wagner, 1843)	Chiroptera**	This study
Phyllostomus hastatus (Pallas, 1767)	Chiroptera**	This study
Platyrrhinus lineatus (E. Geoffroy, 1810)	Chiroptera*	This study
Sturnira lilium (E. Geoffroy, 1810)	Chiroptera*	This study
Vampyressa pusilla (Wagner, 1843)	Chiroptera*	This study
Sylvilagus brasiliensis (Linnaeus, 1758)	Lagomorpha	This study
Homo sapiens (Linnaeus, 1758)	Primates	GenBank (NT_021977, NM_020978.1, M24895.1) and this study
Pan troglodytes (Blumenbach, 1775)	Primates	This study
Mus musculus (Linnaeus, 1758)	Rodentia	GenBank (NW_000201.1, NW_021792.1, NM_007446.1, NM_009669.1)
Rattus norvegicus (Berkenhout, 1769)	Rodentia	GenBank (NM_031502.1, J00703.2, NW_043548.1)
Dugong dugon (Müller, 1776)	Sirenia	This study
Gallus gallus (Linnaeus, 1758)	Outgroup (Aves)	GenBank (U63411.1)
Xenopus laevis (Daudin, 1802)	Outgroup (Amphybia)	GenBank (AF468647)
Tetraodon nigroviridis (de Procé 1822)	Outgroup (Teleostei)	GenBank (AJ427289)
Anguilla japonica (Temminck & Schlegel, 1846)	Outgroup (Teleostei)	GenBank (BAB85635.1)

*Phyllostomidae/Stenodermatinae.
**Phlyllostomidae/Phylostominae.
***Molossidae.
****Vespertillionidae.

(gi13928683). The designed primers were previously tested on the simulation software Amplify 1.2 (Engel, 1992).

Sequencing

PCR products were purified with a 1:1 mix of 1 U of Exonuclease I (EXO I) and 10 U of Shrimp Alkaline Phosphatase (SAP). The sequencing reactions were carried out with ET-DYE® kit (Amersham-Biosciences) following manufacturer's protocol, with the each one of the primers used in the PCR amplification. The sequencing runs were performed in a MegaBACE 1000® capillary sequencer (Amersham-Biosciences).

Base calling and consensus sequence generation

Sequence quality was checked on Phred v.0.20425 (Ewing & Green, 1998) and single individual consensus sequences were assembled on Phrap v.0.990319 (Green, 1994), and visualized and edited on Consed 12.0 (Gordon, Abajian & Green, 1998). Only sequences and positions with Phred values

Table 2. Primer pairs developed and used in this study

Primers pairs	Anealing temperature	Expected size
BatAmy1 CGAGAACTATAATGATGCTACTCAGG BatAmy2 CTCCTGGTAAATGAAAGGTTTACTACC	53°C	~670 bp
BatAmy1b GTTCGTATTTATGTGGATGCTGTAA BatAmy2 CTCCTGGTAAATGAAAGGTTTACTACC	53°C	~715 bp
BatAmy3 GTCAGAGATTGTCGTCTGTCTGGTC BatAmy2 CTCCTGGTAAATGAAAGGTTTACTACC	51°C	~250 bp

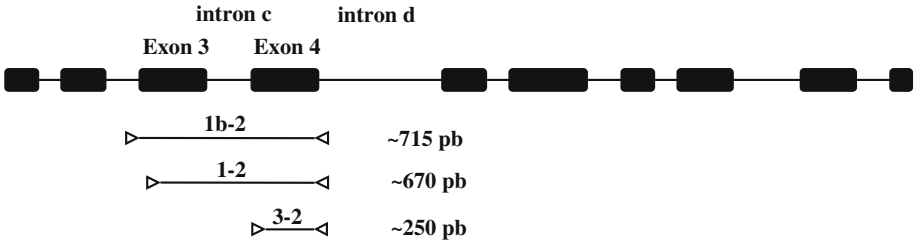

Figure 1. Schematic representation of the α-amylase gene showing the location of the segments studied, the annealing primer sites and the expected PCR fragment size. Numbers refer to primers listed in Table 2

over 20 were used in the analysis. Possible allelic variants were checked on Polyphred software (Nickerson, Tobe & Taylor, 1997). Some α-amylase DNA sequences from other mammals and Vertebrates were retrieved from GenBank (Table 1).

Data analyses

A consensus sequence was generated for each species. Because few sequences showed low quality data in positions particularly in exon 3, we limited our investigation on intron c and exon 4 (Figure 1) totalizing 663 bp positions. *Vampyressa pussilla* was only included on exon 4 analyses, because only a small segment of intron c matched our quality cut off. We have used two datasets in the analyses (Table 1), the first one included the entire segment (intron c + exon 4) and was analyzed in species of Chiroptera, Primates and Rod entia (named CPR dataset) and the second one include only the exon 4 sequences in all available data from distinct mammalian orders and other Vertebrates (named ALL dataset).

Phylogenetic analysis

Phylogenetic analysis were conducted both on PAUP* 4.08b (Swofford, 1998) and MEGA 2.1 (Kumar et al., 2001), with identical results. Maximum parsimony trees were searched with heuristic procedure by Tree Bisection Reconnection algorithm (TBR) on PAUP and Closest Neighbor Inter-exchange (CNI) on MEGA, both tested with 1000 bootstrap replicates and 10 rounds of random addiction sequences. Neighbor Joining trees were searched on MEGA 2.1, using Tamura-Nei model (Tamura & Nei, 1993). Topology was tested with a 1000 bootstrap replicates.

Parameters such as transition/transversion ratio (R), GC content, nucleotide composition, number of synonymous (S) and nonsynonymous (N) sites and synonymous (s) and nonsynonymous

(n) changes, and other sequence features were calculated in MEGA 2.1.

Tests of natural selection and molecular adaptation

Only the CPR dataset was tested. We used three different methodologies to test the influence of natural selection in an α-amylase gene segment and a possible episode of molecular adaptation in the lineages.

First method: It is a pairwise distance test described by Nei and Gojobory (1986) and its modified version (Zhang, Rosemberg & Nei, 1998), both implemented on MEGA 2.1 . The ratios d_N ($=n/N$) and d_S ($=s/S$) of nonsynonymous and synonymous substitutions were also calculated on MEGA 2.1.

A Z test was conducted for each sequence pair with:

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

where $Var(d_N)$ and $Var(d_S)$ are the variances of d_N and d_S as calculated by 1000 bootstrap replications.

The Z values were tested under three alternative hypotheses:

(a) a neutrality test; (b) a positive (darwinian) selection test; and (c) a negative (purifying) selection test.

(a) $H_0: d_N = d_S$ $H1: d_N \neq d_S$ (b) $H_0: d_N = d_S$ $H_1: d_N > d_S$ (c) $H_0: d_N = d_S$ $H_1: d_N < d_S$

We have also inferred the ancestral sequences for the 1, 2, 3, 4, 5 and P branches of the phylogenetic tree on Figure 2a using Bayesian methods implemented on PAML (Yang, Kumar & Nei, 1995; Yang, 1997; Nielsen & Yang, 1998). These sequences were used to calculate a Fisher exact test for each branch and test selection among the lineages as described in Zhang, Rosemberg and Nei (1998).

Second method: The lineage test described by Creevey and McInerney (2002) uses the Crann software. In our study we tested both diversifying and purifying selection and we have used the tree shown in Figure 2a to perform this test.

Third method: The Likelihood Ratio Tests (LRTs) were developed to detect selection using many codon substitution models (Goldman & Yang, 1994; Yang, 1998; Yang & Bielawski, 2000; Yang & Nielsen, 2000, 2002; Yang, 2002).

The LRTs were tested as $2\times$ [Log(ℓH_0) – Log (ℓH_1)] where ℓH_0 and ℓH_1 are the likelihood functions of null and alternative hypothesis (models) respectively and this value is checked on a χ^2 distribution with degrees of freedom equals to the difference between the number of parameters on each model. In all site-models the sites with probability to be targets of selection were inferred by Bayesian methods (Nielsen & Yang, 1998). All calculations were made in the PAML program (Yang, 1997).

Three kinds of models were tested:

(1) Branch models to detect selection on specific lineages. The lineages studied were the same used in the Nei-Gojobory test (Figure 2a);
(2) Site specific models that assume variable selective pressure among sites;
(3) Branch-sites models.

Results

Sequence data

A total of 663 bp, corresponding to the entire exon 4 (231 bp) and most of the intron c (432 bp) of the α-amylase gene were analyzed on the CPR dataset consisting of nine at species, a human and a chimpanzee sequences generated in our lab (Table 1) and the sequences of human, rat and mouse α-amylases 1 and 2. From the 663 bp, 455 sites were variable (100 bp in exon 4) and 197 bp were invariable sites (131 in exon 4). In the ALL dataset, several mammals were included as well as other Vertebrates sequenced in our lab or retrieved from the GenBank database (Table 1). The 231 bp α-amylase exon 4 in the ALL dataset presented 147 variable and 84 invariable sites.

No indels were found in the coding region studied and the transition/transversion ratio was $R = 1.65$ (CPR) and $R = 1.60$ (ALL), and nucleotide composition A = 25.4%, T = 25.4%, C = 23.6% and G = 25.6% for CPR dataset and A = 25.5%, T = 26.4%, C = 22.5% and G = 25.6% for ALL dataset. The GC content is about 49% in the CPR dataset and 48% in the ALL dataset. These values are according to the range observed

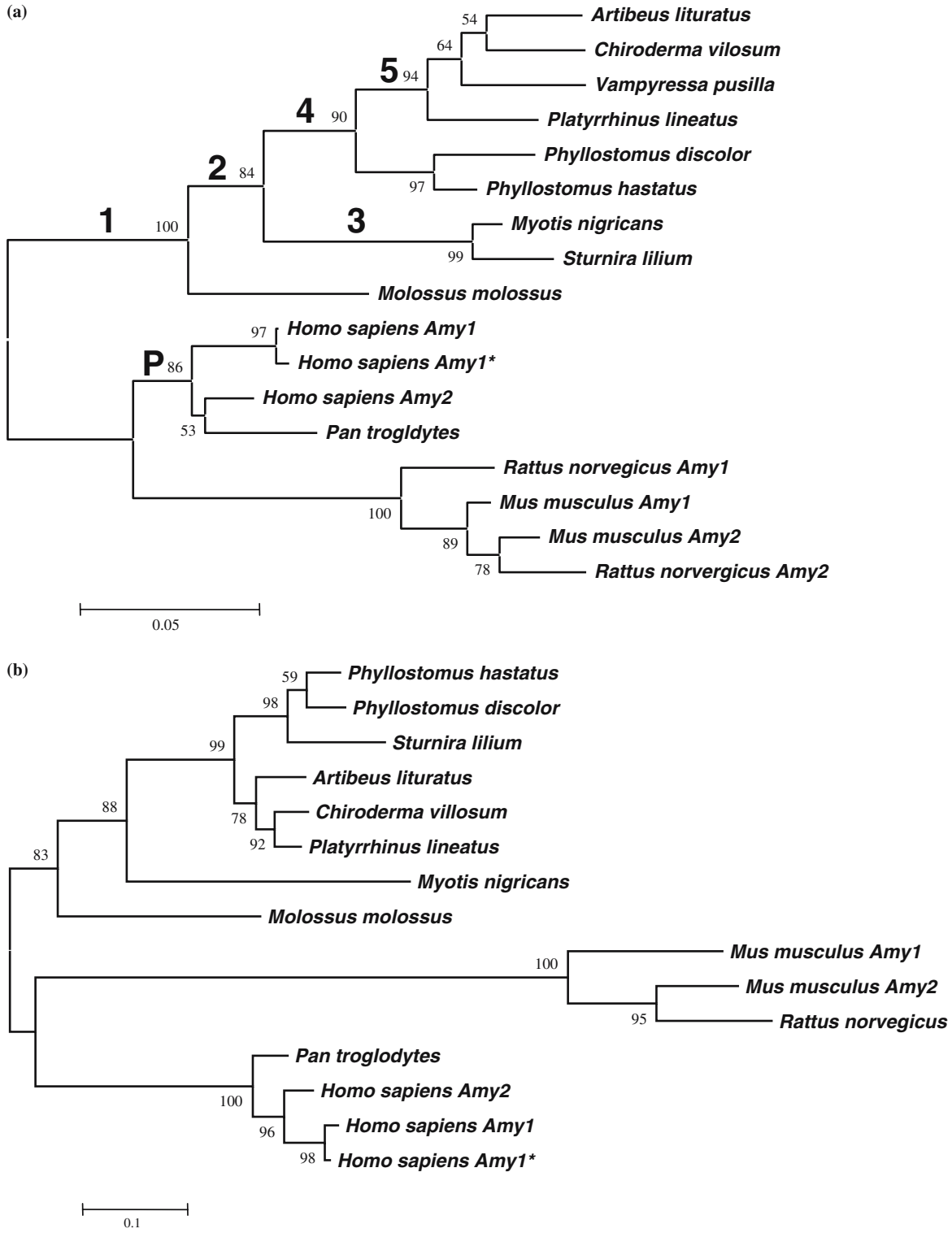

Figure 2. Neighbor Joining phylogenetic reconstructions for the CPR dataset with different segments of the α-amylase gene. The numbers near the branches refer to percentage of 1000 bootstrap replications. (a) exon 4 tree (231 bp), the numbers 1–5 and letter 'P' refer to the branches used in the selection tests (see material and methods). (b) intron c tree (432 bp).

in gene regions from Vertebrates (Nei & Kumar, 2000).

From the 231 positions in the α-amylase exon 4, 141 are zerofold degenerated for the CPR dataset and 133 for the ALL dataset. This estimate suggests that at least 60% of the sites are subjected to nonsynonymous mutations. If we consider the two- and fourfold degenerated sites we will have about 86% of potential mutations to be nonsynonymous, a number much greater than expected in gene regions with equal nucleotide frequencies (75%, Nei & Kumar, 2001). Those values could be explained by the high GC content at 3rd codon positions found in this exon (A3 = 15.3%, T3 = 28.6%, C3 = 30.3%, G3 = 25.8% in CPR and 16.2, 31.0, 27.5, 25.3% in ALL) and could be showing saturation in synonymous positions caused by restriction of function or spatial conformation of the gene product (Li, 1997; Golding & Dean, 1998; Nei & Kumar, 2001).

Phylogenetic analyses

Our main interest in the phylogenetic reconstructions was to search for a pattern of molecular convergence between the bats and their feeding habits and to test if the α-amylase gene could be useful to solve systematic problems as the relationship of mammalian orders including bats, primates and rodents.

The first analysis was in the coding segment of the CPR dataset (exon 4). Both neighbor joining (NJ) and maximum parsimony consensus (MP, 6 trees with CI = 0.742, RI = 0.850, RCI = 0.630) methods showed trees with identical topology (Figure 2a), even when the deduced amino acids sequences were used in the reconstruction (data not shown). The trees revealed a group including the vespertilionid bat *Myotis nigricans* and the phyllostomid *Sturnira lilium*, in both nucleotide and amino acid reconstructions. This grouping was totally unexpected since the placement of these bat species in their respective families were well established (Gannon, Willig & Knox-Jones, 1989; Lim, 1993; Simmon & Geisler, 1998; Nowak, 1999) and their feeding habits are quite different (*M. nigricans* is a strict insectivorous and *S. lilium* is a frugivorous bat) rejecting an adaptive convergence hypothesis (see discussion). The positions of other bat species in the tree were according to the traditional view of bat phylogeny using molecular data (Cytochrome-*b*, RAG-2, Van den Bussche et al., 1993; Baker et al., 2000) and morphology (Lim, 1993; Simmon & Geisler, 1998).

No resolution of the phyllostomid bat phylogeny could be observed at the amino acid level. However both trees (data not shown) grouped together the rodents and primates, supporting the Euarchontoglires hypothesis (Madsen et al., 2001; Murphy et al., 2001; Arnason et al., 2002; Deusuc et al., 2002; Springer et al., 2004).

When the entire segment (intron c + exon 4, 663 bp) was analyzed, the tree topology changed considerably (Figure 3a) and again the unique discrepancy was *S. lilium* that became a basal branch of the phyllostomid bats when the traditional placement of *S. lilium* is at the base of Stenodermatinae subfamily (Lim, 1993; Simmon & Geisler, 1998). When only the intron c was analyzed (Figure 2b), *S. lilium* grouped with the Phyllostominae bats *Phyllostomus discolor* and *P. hastatus*.

The phylogenetic reconstructions using the ALL dataset generated a tree with lots of polytomies (Figure 4), presenting no resolution in the relationship of mammal orders; even the avian *Gallus gallus* sequence was placed in the middle of the mammal polytomy.

As a final test to the Euarchontoglires versus Archonta hypotheses, we constructed two new trees. The CPR dataset was analyzed together either Teleostei/Amphybia (not shown) or *Dugong dugon* (Figure 3b), an Afrotheria/Panungulata mammal, as outgroups. Both trees again gave support to the Euarchontoglires grouping of rodents and primates.

Selection tests

The traditional Nei-Gojobory test showed that most of comparisons of sequence pairs revealed non neutral evolution, suggesting purifying selection on *Z*-tests. This test does not account for different transition/transversion ratios (use $R = 0.5$), with could dramatically affect the d_N and d_S estimations (Zhang, Rosemberg & Nei, 1998). We have estimated the ratio $R = 1.65$ in the CPR dataset, a value at least three times higher. Thus a modified version of this test (Zhang, Rosemberg & Nei, 1998) that accounts for these features was applied, but the same results were obtained.

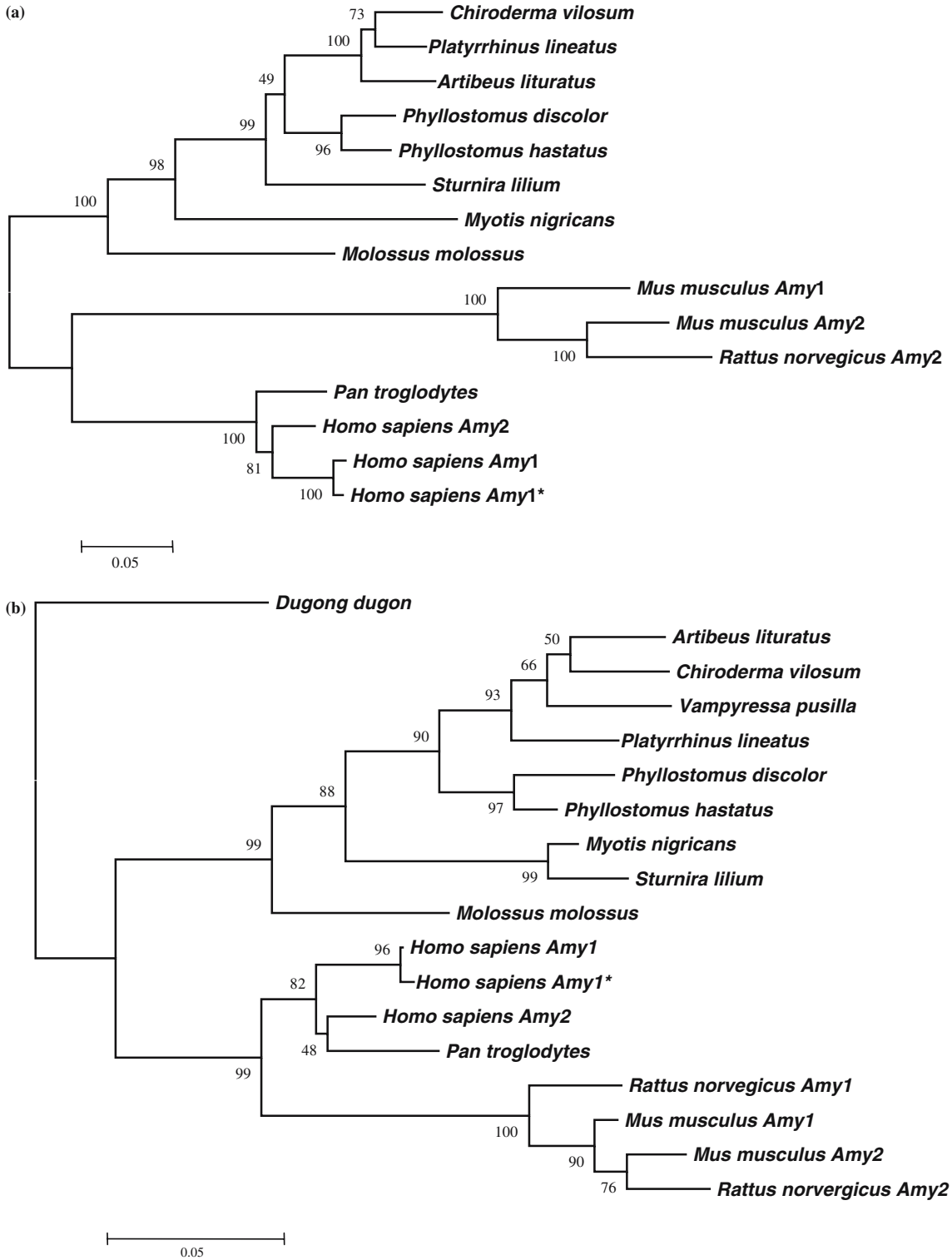

Figure 3. (a) Neighbor Joining phylogenetic reconstruction for the CPR dataset with the entire fragment (663 bp) of the α-amylase gene studied. The numbers near the branches refer to percentage of 1000 bootstrap replications. (b) Neighbor Joining phylogenetic reconstruction for the CPR dataset (exon 4 only) using *Dugong dugon* (Sirenia: Afrotheria) as an outgroup. The numbers near the branches refer to percentage of 1000 bootstrap replications.

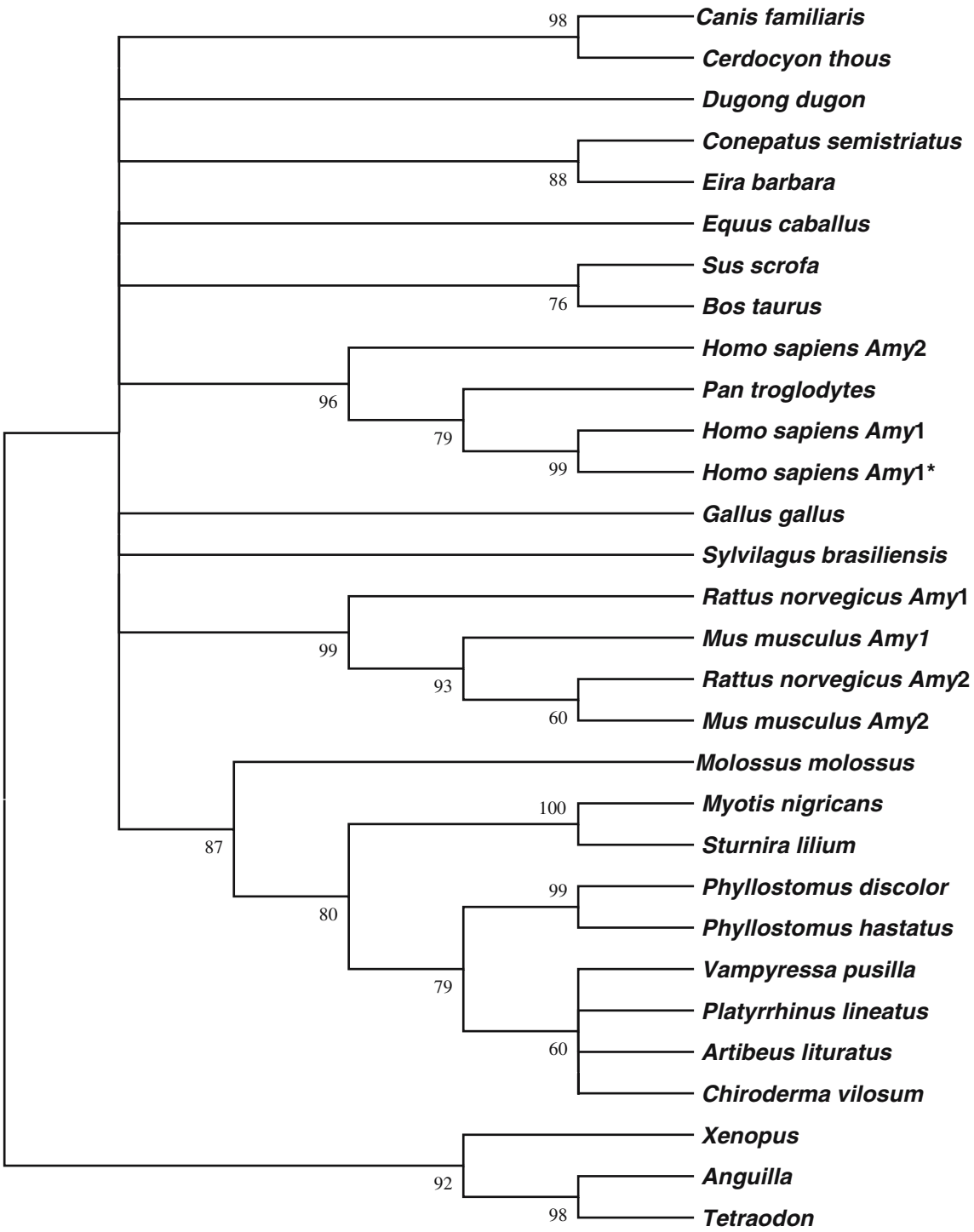

Figure 4. Phylogenetic reconstruction for the ALL dataset using the exon 4 segment. The numbers near the branches refer to percentage of 1000 bootstrap replications. Branches with less than 50% bootstrap values were collapsed.

Because the Z test is designed for large samples, the number of synonymous and nonsynonymous positions must be usually large. Thus we have also used a Fisher exact test on the inferred ancestral sequences, which is indicated for small samples (Zhang, Kumar & Nei, 1997; Zhang, Rosemberg & Nei, 1998). The selection tests using the Fisher exact test were all significant at $p < 0.05$. However this is a bidirectional test and the values of d_N and d_S must be compared. With 168 nonsynonymous sites and $s \sim 27$ and 63 synonymous sites with $n \sim 11$ we have $d_N = 0.06$ and $d_S = 0.4$, implying a massive presence of purifying selection with $\omega = 0.15$.

Yang (2002) made several critiques on the use of the Fisher exact tests to search for selection in gene lineages. First, they are based on ancestral reconstruction that many times can contain errors or underestimate the substitutions. Second, this kind of test does not account for GC content on 3rd codon positions, so the estimates of d_S and d_N are inaccurate.

The Creevey–MacInerley test detected non neutral evolution occurring in all but two branches. It also detected an increase in nonsynonymous non-directional changes in three branches, one in the branch 3 in Figure 2a and two others in the Primates lineage (the basal and *Pan troglodytes-Homo sapiens* α-amylase two branches). The two neutral ones included the branch with sequences of α-amylase 1 from GenBank and the human sequence from our lab (this result was expected), and the branch of α-amylase 2 in mouse and rat. The most interesting result in this test is in the branch 3 in Figure 2a where both the purifying and diversifying selection tests were shown to be significant ($p < 0.05$). This may indicate that some sites could be under purifying selection by functional restrictions while others may be released from those restrictions (Creevey & McInerney, 2002). This is the only branch that contradicts the traditional phylogeny of bats.

Table 3. Likelihood results and parameter estimates in the codon models

Models		Log(ℓ)	Estimates of parameters	Detected sites
	M0	−1186.08	$\omega = 0.1639$	
Branch 1	B1	−1185.77	$\omega = 0.1715$	NA
Branch 2	B1	−1186.58	$\omega = 0.1715$	NA
Branch 3	B1	−1184.33	$\omega = 0.1796$	NA
Branch 4	B1	−1184.93	$\omega = 0.1750$	NA
Branch 5	B1	−1185.28	$\omega = 0.1709$	NA
Branch P	B1	−1185.72	$\omega = 0.1593$	NA
Sites	M1	−1192.75		NA
	M2	−1192.70	$p_2 = 0.03$ ($\omega = 4.07$)	
	M3	−1171.47	$p_2 = 0.01$ ($\omega = 2.38$)	7 (98%)
	M7β	−1172.85	$p = 0.4$	
	M8βω	−1171.89	$p_1 = 0.01$ ($\omega = 2.37$);	7 (90%)
Branch-sites 1	MA	−1189.05		
	MB	−1171.12		
Branch-sites 2	MA	−1187.54		
	MB	−1171.13		
Branch-sites 3	MA	−1185.01		
	MB	−1171.43		
Branch-sites 4	MA	−1188.48		
	MB	−1172.27		
Branch-sites 5	MA	−1189.97		
	MB	−1172.61		
Branch-sites P	MA	−1188.20	$P_2 = 0.018$ ($\omega = 0.99$)	63 (99%)
	MB	−1168.95	$P_2 = 0.024$ ($\omega = 0.99$)	23 (77%); 63 (99%)

NA: not allowed.

In the codon-substitution models tested, all ω ratios estimated for the lineages presented very low values (Table 3), which suggest purifying selection (Yang & Bielawski, 2000; Yang, 2002). Because some high values of ω were observed in the sites analysis, this could indicate the presence of a mosaic sequence with some sites that are under functional constraints and others that are not (Yang & Nielsen, 2002).

The first LRT analyzed was between models M0 (one ω ratio) and B1 (two ω ratios), which was intended to verify if the ω values were different between the lineage of interest and others. Each branch labeled in Figure 2a was tested for positive selection. The B1 model did not fit the data better than the M0 (Table 3). Only the branch 3 presented a 'marginal' statistic significance ($p = 0.06$). Marginal probabilities have been used as an indicative of a probable increase in ω (Yang & Nielsen, 2002). However our estimates for ω in this model was very low ($\omega = 0.1796$) and this hypothesis seems unlikely.

In the tests using the sites models, the first one compares the one ratio model M0 with the M3 model, where ω is divided in discrete classes, in our case $K = 3$ classes, with the proportion (p) of sites belonging to each class estimated from data. The discrete model fits the data much better than the one ratio model, with $p = 0.000$ and a proportion of sites $p_2 = 0.01$ with $\omega_2 = 2.38$. Bayesian analysis also indicates a probable site under strong selective pressure at codon 7, with 97% of probability (Table 3). Again it seems that different selective pressures occur among the sites. The second test on sites models was performed with the M1 (neutral) and M2 (selection) models. M1 sets two ω classes, with $\omega_1 = 1$ and $\omega_0 = 0$, and M2 adds a third class with ω_2 estimated from the data to verify sites with ω values higher than 1. Again this test was not significant ($p = 0.95$).

The above test could be very conservative (Yang & Nielsen, 2002) and a new comparison with a more homogeneous distribution of ω was used. This third test on sites models uses the M7β model as a null hypothesis. This model make use of a β distribution of ω classes (β varies from 0 to 1) that is a flexible null hypothesis for testing positive selection. The alternative model M8βω adds a new site class with ω estimated from data, and despite it does not fit the data better than the M7β model ($p = 0.38$), it estimates a $\omega = 2.37$ in 1% of the sites and the bayesian analysis also indicates the codon number 7 as a possible target of selection with 90% of probability.

The last selection test uses the branch-sites models. The MA model only allows sites under selective pressure in the interest (foreground) branches, and it was applied in the labeled branches of Figure 2a. In this LRT, the M1 model is the null hypothesis against the MA. The MA model fits the data better in all but one branch (Table 3). Only the branch leading to the Stenodermatinae bats (branch 5) presented a 'marginal' p value (0.06). The model also indicates a site under selective pressure at codon 63 with 99% probability in the primate branch. The final LRT was between the neutral M1 model and the branch-site model MB that is the branch-site model version of the M3 model with a discrete class distribution for ω among the sites, so it can be tested against the M3 model. This model only fits the data better in the primate branch (P) and it identified the same codon under selection as the MA model (63; 99% probability) and a new one at codon 23 with 71% probability.

Discussion

The incoherence found in the bat phylogeny using α-amylase exon 4 could not be explained by adaptive convergence since both animals involved have very distinct feeding habits and belong to different lineages (Gannon, Willig & Knox-Jones, 1989; Lim, 1993; Nowak, 1999). It is most likely that the substitutions found to be shared between *S. lilium* and *M. nigricans* could be either an evidence of ancestral character states (plesiomorphies) retained in *S. lilium* or the independent acquisition of these characters by both taxa.

Although *S. lilium* and *M. nigricans* share some radical nonsynonymous substitutions (Figure 5), the raw number of differences between them is the same as found between *M. nigricans* and *Molossus molossus*. A phylogenetic tree constructed without the 3rd codon positions (data not shown) groups all three bats above and *Platyrrhinus lineatus*, *Phyllostomus hastatus* and *Phyllostomus discolor*. It confirms that most of differences among them are at 3rd codon positions, normally the most degenerated position in coding regions (Li, 1997;

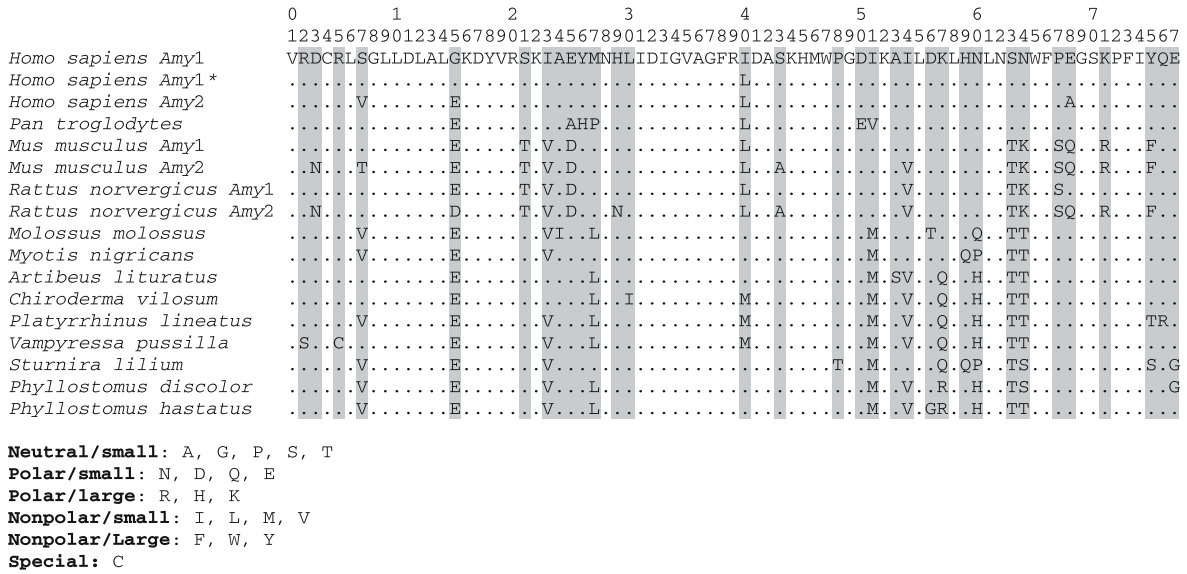

Figure 5. Alignment of the deduced amino acid sequences for the exon 4 of the α-amylase gene in the CPR dataset, showing the physicochemical properties classification of amino acids according to Zhang (2000). Dots (.) in the alignment mean identical amino acid to first sequence.

Nei & Kumar, 2000). Thus, the possibility of independent acquisition of shared mutations between *S. lilium* and *M. nigricans* lineages seems likely particularly because of the presence of saturation at 3rd codon positions.

S. lilium have been traditionally considered a basal member of Stenodermatinae subfamily (Lim, 1993; Simmon & Geisler, 1998) and it displays remarkable cerebral morphological differences as well as other features, when compared to other Stenodermatinae bats (Gannon, Willig & Knox-Jones, 1989).

To further investigate the place of *S. lilium* in the Phyllostomidae family, it would be needed an increment in the taxon sampling, with a higher number of bat species, especially basal phyllostomids as *Macrotus* and other species of the *Sturnira* genus. If a similar pattern related to *S. lilium* were found in other bats, then the occurrence of independent acquisition would become more likely.

Another possible explanation is that an independent duplication of the α-amylase gene may have occurred, followed by a (likely) deletion of distinct paralogous copies, misguiding our phylogenetic reconstructions. Independent duplications of α-amylase genes have been described in literature for other mammals (Samuelson, Phillips & Swanberg, 1988; Samuelson et al., 1990). However we have not found any evidence of the presence of distinct α-amylase copies in bats. All the sequences have been checked with Polyphred (Nickerson, Tobe & Taylor, 1997) to search for differences that could be attributed to paralogous gene sequences.

The exon 4 was not useful to discriminate mammal orders because of the presence of many polytomies (Figure 4). Maybe the analysis of a larger segment of the α-amylase gene, with many coding and noncoding regions, would increase its resolution. However, exon 4 was used to investigate the relationships among the orders Chiroptera, Primates and Rodentia, giving support to the Euarchontoglires hypothesis against the existence of Archonta.

Most of selection tests showed strong evidence of purifying selection. However some methods imply the possible existence of sites under diversifying selection in a 'sea' of functionally restricted sites. No radical amino acid change in size or polarity could be verified in any of the sites proposed to be under selection in the codon models (codons 7, 23, 63; Table 3, Figure 5). It is unlikely that the amino acid changes in the proposed sites affect significantly the function of this protein (Golding & Dean, 1998; Tourasse & Li, 2000). In fact, only codon 7 displays a slight change in polarity from a nonpolar to a neutral amino acid

(Creighton, 1993; Zhang, 2000) and it seems to be nondirectional, uncorrelated with lineage relationship or feeding habits (Figure 5).

Only a few amino acid alterations found in this study could be considered radical changes and may affect the enzyme function; however there is no indication that they are under selection. The first one in codon 57 is a size change that occurs only in the Stenodermatinae lineage (*Sturnira lilium* included). The small and polar amino acid Glutamine (Q) is found instead of the large and polar Arginine (R in Phyllostomidae) and Lysine (K in all other organisms studied) (Figure 5).

Another similar size change is found in codon 59, where *S. lilium* and *M. nigricans* have again a Glutamine and the other organisms presented the large and polar Histidine (H) (Figure 5).

In codon 60, another radical change shared by *S. lilium* and *M. nigricans* that carry the small and neutral Proline (P) where the phyllostomid bats have Histidine (H, large and Polar), and *M. molossus* and the others have small polar amino acids (Glutamine and Asparagine, respectively) (Figure 5). This change in size and polarity found in the phyllostomid bats (except *S. lilium*) could be taking to an important change in protein structure and function (Creighton, 1993).

An interesting radical change appears in codon 64 and it is lineage related. Primates have the small and polar Asparagine (N), Rodentia present the large and polar Lysine (K) and the Chiroptera displays the neutral and small amino acids Threonine (T) and Serine (S) (Figure 5).

The last radical change happens in one of the catalytic amino acids in the 77 codon position, where we found *Sturnira lilium* and *Phyllostomus discolor* with the small and neutral Glycine (G) and the other organisms studied have small and polar (negatively charged) amino acids. Besides the apparent importance of this site (codon 77) as a part of the catalytic triad of the α-amylase protein, variations have been reported in other organisms in literature, and it is considered the least important of the triad (Janecek, 1994, 1997). However, Glycine has been also reported to show a weak negative charge (Creighton, 1993).

No alterations in the other catalytic site (codon 41) or on the calcium binding sites (codon 11 and 45) were detected and these sites are the most important for the correct functioning of α-amylase (Janecek, 1994, 1997).

Graur (1985) showed a great correlation between the amino acid composition and the evolution rate in proteins, but this correlation was extremely criticized and several works confirm that the evolutionary rate of a protein depends mostly on the three-dimensional structure than on its primary sequence (Golding & Dean, 1998; Tourasse & Li, 2000).

In fact, the importance of α-amylase functional restrictions could be seen when the nucleotide (mRNA) and the amino acid (protein) sequences of *Asterias ruber*, a starfish, and *Sus scrofa*, the domestic pig, (the best known α-amylase structure; Janecek, 1994, 1997). are aligned (data not shown). The nucleotide sequences are very divergent but at the amino acid level they are much more similar, showing a conserved amino acid sequence even between such phylogenetically distant groups.

The α-amylase gene seems to be under strong structural and functional restrictions with purifying selection affecting almost all sites as attested by our results. Because it is the main enzyme related to the digestion of starch and sucrose (Janecek, 1994, 1997), its correct functioning is essential for the primary metabolism. However our results may indicate it is not related to the distinct feeding habits observed in bats.

The present work represents a detailed analysis of gene segments that could be likely influenced by Natural Selection, and related to species diversification and adaptation. Through the use of distinct selection tests, we could identify several evidences of purifying selection, indicating functional constraints affecting in the same way distinct lineages. Besides, the comparison with distinct mammalian orders supported some recent findings of the Euarchontoglires grouping. The same strategy could be further applied to larger gene segments and several other genes that could be under selection influence.

Acknowledgments

We would like to thank to FBPN (Fundação o Boticário de Proteção à Natureza), FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais) and PELD-CNPq for the financial support to the project, and CAPES and CNPq from Brazil for the fellowships to RAFR and FRS. Native

animal samples were collected under licenses 011/2000, 053/2001 and 070/2002 issued by IBAMA, and this work was allowed by the special authorization of access to genetic resource # 03/2004 emitted by the Ministry of Environment of Brazil (CGEN/MMA).

References

Adkins, R.M. & R.L. Honeycutt, 1991. Molecular phylogeny of the superorder Archonta. Proc. Natl. Acad. Sci. USA 88: 10317–10321.

Allard, M.W., B.E. McNiff & M.M. Miyamoto, 1996. Support for interordinal eutherian relationships, with an emphasis on Primates and their archontan relatives. Mol. Phylogenet. Evol. 5: 78–88.

Arnason, U., A.J. Adegoke, K. Bodin, E.W. Born, Y.B. Esa, A. Gullberg, M. Nilsson, R.V. Short, X. Xu & A. Janke, 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. USA 99(12): 8151–8156.

Baker, R.J., C.A. Porter, J.C. Patton & R.A. Van Den Bussche, 2000. Systematics of Bats of the Family Phyllostomidae based on RAG2 DNA Sequences. Occasional Papers The Museum Texas Technological University 202: 1–16.

Creevey, C. & J.O. McInerney, 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. 2nd edn. W.H. Freeman publisher, New York, USA.

Deusuc, F., M. Scally, O. Madsen, M.J. Stanhope, W.W. de Jong, F.M. Catzeflis, M.S. Springer & E.J.P. Douzery, 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Mol. Biol. Evol. 19(10): 1656–1671.

Engel, W., 1992. Amplify 1.2. Departament of Genetics, University of Wisconsin.

Ewing, B. & P. Green, 1998. Basecalling of automated sequencer traces using Phred II Error probabilities. Genome Res. 8: 186–194.

Ferarezzi, H. & E.A. Gimenez, 1996. Systematic patterns and the evolution of feeding habits in Chiroptera (Archonta: Mammalia). J. Compar. Biol. 1(3–4): 75–94.

Gannon, M.R., M.R. Willig & J. Knox-Jones, 1989. *Sturnira lilium*. Mammalian Species 333: 1–5.

Golding, G.B. & A.M. Dean, 1998. The structural basis of molecular adaptation. Mol. Biol. Evol. 15(4): 355–369.

Goldman, N. & Z. Yang, 1994. A codon-based model of nucleotide substitution for protein coding DNA sequences. Mol. Biol. Evol. 11(5): 725–736.

Gordon, D., C. Abajian & P. Green, 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8: 195–202.

Graur, D., 1985. Amino acid composition and the evolutionary rates of protein-coding genes. J. Mol. Evol. 22: 53–62.

Green, P. 1994. Phrap.www.genome.washington.edu/UWGC/analysistools/phrap.htm.

Gumucio, D.L., Wiebauer, K., Caldwell, R.M.; Samuelson, L.C. & Meisler, M.H. Concerted evolution of human amylase genes. Mol. Cell. Biol. 8: 1197–1205.

Janecek, S., 1994. Sequence similarities and evolutionary relationships of microbial, plant and animal α-amylases. Eur. J. Biochem. 224: 519–524.

Janecek, S., 1994. Sequence similarities and evolutionary relationships of microbial, plant and animal α-amylases. Eur. J. Biochem. 224: 519–524.

Janecek, S., 1997. α-amylase family: molecular biology and evolution. Prog. Biophys. Mol. Biol. 67(1): 67–97.

Kumar, S., K. Tamura, I.B. Jakobsen & M. Nei, 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17(12): 1244–1245.

Li, W.-H., 1997. Molecular Evolution. Sinauer Associates Inc, Massachusetts USA.

Lim, B.K., 1993. Cladistic reapraisal of neotropical Stenodermatinae bat phylogeny. Cladistics 9: 147–165.

Madsen, O., M. Scally, C.J. Douardy, D.J. Kao, R.W. DeBry, R.M. Adkins, H.M. Amrine, M.J. Stanhope, W.W. de Jong & S. Springer, 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409: 610–614.

Murphy, W.J., E. Eizirik, W.E. Jonhson, Y.P. Zhang, O.A. Ryder & S.J. O'brien, 2001. Molecular phylogenetics and the origin of placental mammals. Nature 409: 614–618.

Nei, M. & T. Gojobori, 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3: 418–426.

Nei, M. & S. Kumar, 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York USA.

Nickerson, D.A, V.O. Tobe & S.L. Taylor, 1997. Polyphred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucl. Acids Res. 25: 2745–2751.

Nielsen, R. & Z. Yang, 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148: 929–936.

Nikaido, M., M. Harada, Y. Cao, M. Hasegawa & N. Okada, 2000. Monophyletic origin of the Order Chiroptera and its phylogenetic position among Mammalia, as infered from the complete sequence of the mitocondrial DNA of a japanese megabat, the Ryukyu flying fox (*Pteropus dasymallus*). J.Mol. Evol. 51: 318–328.

Novacek, M.J., 1992. Mammalian phylogeny: shaking the tree. Nature 256: 121–125.

Nowak, R.M., 1999. Walker's Mammals of the World. 6 edn., Vol. I. The Johns Hopkins University Press.

Phillips, C.J. & B. Tandler, 1996. Salivary glands, cellular evolution, and adaptive radiation in mammals. Eur. J. Morphol. 34(3): 155–161.

Phillips, C.J., B. Tandler & C.A. Pinkstaff, 1987. Unique salivary glands in two genera of tropical microchiropteran bats: an example of evolutionary convergence in histology and hystochemistry. J. Mammal. 68(2): 235–242.

Phillips, C.J., A. Weiss & B. Tandler, 1998. Plasticity and patterns of evolution in mammalias sallivary glands: comparative immunohistochemistry of lysozyme in bats. Eur. J. Morphol. 36(suplement): 19–26.

Sambrook, J., D.W. Russell & J. Sambrook, 2001. Molecular Cloning: A Laboratory Manual. 3CSHL press, New York USA.

Samuelson, L.C., R.S. Phillips & L.J. Swanberg, 1988. Amylase gene structures in Primates: retroposon insertions and promoter evolution. Mol. Biol. Evol. 13(6): 767–779.

Samuelson, L.C., K Wiebauer, C.M. Snow & M.H. Meisler, 1990. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. Mol. Cell. Biol. 10: 2513–2520.

Santos, F.R., P.E.M. Guimarães & R.A.F. Redondo, 2002. Bancos de DNA: coleções estratégicas para estudos da biodiversidade. Lundiana: Int. J. Biodivers. 3(2): 93–98.

Simmons, N.B. & J.H. Geisler, 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. Bull. Am. Museum Nat. History. 235: 1–182.

Springer, M.S., M.J. Stanhop, O. Madsen & W.W. de Jong, 2004. Molecules consolidates the placental mammal tree. Trends Ecol. Evol. 19(8): 430–438.

Swofford, D.L. 1998. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.08b. Sinauer Associates Inc, Massachusetts USA.

Tamura, K. & M. Nei, 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512–526.

Tandler, B., C.J. Phillips & T. Nagato, 1996. Histological convergent evolution of the accessory submandibular glands in four species of frog-eating bats. Eur. J. Morphol. 34(3): 163–168.

Ting, C.-N., M.P. Rosemberg, C.M. Snow, L.C. Samuelson & H.M. Meisler, 1992. Parotid-specific expression of a human salivary amylase gene in transgenic mice. Genes Develop. 6: 1457–1465.

Tourasse, N.J. & W.-H. Li, 2000. Selective constraints, amino acid composition and the rate of protein evolution. Mol. Biol. Evol. 17(4): 656–664.

Van den Bussche, R.A., R.J. Baker, H.A. Wichman & M.J. Hamilton, 1993. Molecular phylogenetics of Stenodermatini bat genera: congruence of data from nuclear and mitochondrial DNA. Mol. Biol. Evol. 10(5): 944–959.

Yang, Z. & J.P. Bielawski, 2000. Statical methods for detecting molecular adaptation. Trends Ecol. Evol. 15(12): 496–503.

Yang, Z. & R. Nielsen, 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17: 32–43.

Yang, Z. & R. Nielsen, 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19(6): 908–917.

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. Mol. Biol. Evol. 15(5): 568–573.

Yang, Z., 2002. Inference of selection from multiple species alignments. Curr. Opin. Genet. Develop. 12: 688–694.

Yang Z, Kumar S., Nei M. (1995) A new method of inference of ancestral nucleotide and amino acid sequences Genetics 141:1641–1650.

Zhang, J., S. Kumar & M. Nei, 1997. Small-sample tests of episodic adaptive evolution: a case of Primate lysozymes. Mol. Biol. Evol. 14(12): 1335–1338.

Zhang, J., H.F. Rosenberg & M. Nei, 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA 95: 3708–3713.

Zhang, J., 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J. Mol. Evol. 50: 56–68.