

# Isolation and Characterization of the *CYP2D6* Gene in Felidae with Comparison to Other Mammals

Tamara Schenekar · Kathrin A. Winkler ·  
Jennifer L. Troyer · Steven Weiss

Received: 17 June 2010 / Accepted: 29 November 2010 / Published online: 28 December 2010  
© Springer Science+Business Media, LLC 2010

**Abstract** The highly polymorphic CYP2D6 protein metabolizes about 25% of commonly used drugs and underlies a broad spectrum of drug responses among individuals. In contrast to extensive knowledge on the human *CYP2D6* gene, little is known about the gene in non-human mammals. CYP2D6 mRNA from 23 cats (Felidae) spanning seven species were compared to available *CYPD6* sequences in ten additional mammals and multiple allelic variants in humans. A relatively high mean *dN/dS* ratio (0.565) was observed, especially within Felidae. Pairwise *dN/dS* ratios were non-monotonically distributed with respect to evolutionary distance suggesting either positive selection or retention of slightly deleterious mutations. Positive selection on specific codons, most notably in regions involved in substrate recognition and membrane anchoring is supported and the possible influence of diet on specific amino acid changes in substrate binding sites is discussed.

**Keywords** Cytochrome P450 · Cats · Selection · *dN/dS* · Slightly deleterious mutations

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-010-9424-1) contains supplementary material, which is available to authorized users.

T. Schenekar · K. A. Winkler · S. Weiss (✉)  
Karl-Franzens Universität Graz, Institut für Zoologie,  
Universitätsplatz 2, A-8010 Graz, Austria  
e-mail: steven.weiss@uni-graz.at

J. L. Troyer  
Laboratory of Genomic Diversity, SAIC-Frederic, Inc., National  
Cancer Institute-Frederick, Frederick, MD 21702, USA

## Introduction

Cytochrome P450s (P450s) are a superfamily of hemoproteins known to metabolize about 75% of marketed human drugs (Guengerich 2008). P450 genes additionally play important roles in the biosynthesis of steroid hormones, oxidation of unsaturated fatty acids and the metabolism of fat-soluble proteins; they are also involved in the activation of procarcinogens like polycyclic aromatic hydrocarbons (PAHs) in tobacco smoke (Hasler 1999). Over 3000 P450 genes have been described in animals, whereas humans possess 57 functional genes (Guengerich 2008). Allelic variants of several human P450 genes underlie the high among individual variation in metabolic conversion and excretion rates of drugs and therefore drug response (Weide and Steijns 1999). As up to 100,000 deaths per year in the USA alone are attributed to adverse responses to drugs, understanding the variable expression of P450 genes in drug response has become a major goal in the burgeoning field of pharmacogenetics (Lazarou et al. 1998; Pirmohamed et al. 2004). The best studied P450 gene is the highly polymorphic *CYP2D6*, accounting for 25% of the metabolism of commonly used drugs (Ingelman-Sundberg 2005). Described mutations resulting in various changes in enzyme function include whole-gene deletion, mutations leading to altered splicing or frame shifts, single nucleotide polymorphisms and gene duplication (Gaedigk et al. 1991; Johansson et al. 1993; Oscarson et al. 1997; Ingelman-Sundberg 2004). Numerous approaches for genotyping and phenotyping the CYP2D6 enzyme have been developed (Frank et al. 2007; Leandro-Garcia et al. 2009) leading to the description of 78 allelic variants (<http://www.cypalleles.ki.se/cyp2d6.htm>). In contrast to such detailed studies in humans, little is known about the sequence variation or function of this gene in non-human animals.

Orthologs of the *CYP2D6* gene have been characterized in several animals of interest to veterinary medicine, whereby the main focus of such research is expression and substrate specificity (Garfinkel 1958; Ioannides 2006; Trepanier 2006; Seliskar and Rozman 2007). The role of *CYP2D6* in domestic cats has attracted attention due to their low levels of high-affinity acetaminophen-UDP-glucuronosyltransferase, which underlies a high risk of acetaminophen intoxication (Court and Greenblatt 1997; Dong et al. 2000). While there is a great deficit in knowledge of non-human P450s, intraspecific polymorphisms have been found (Court et al. 1999; Paulson et al. 1999; Bull et al. 2003; Ioannides 2006), suggesting that the high polymorphism and general importance of this gene family is a prevalent phenomenon among mammals.

If variation in mammalian P450s, and specifically in the *CYP2D6* gene, is adaptive, there should be signals of selection across a broader range of taxa than those thus far characterized. To gain perspective on the evolutionary significance of mammalian *CYP2D6*, we assessed the distribution of synonymous and non-synonymous mutations within and among taxa as well as changes in amino acid (AA) characteristics in selected felids and compared our data to that available from other mammals, including allelic variants from humans.

## Materials and Methods

Liver samples of three domestic cats (*Felis catus*), four cheetahs (*Acinonyx jubatus*), three bobcats (*Lynx rufus*), three leopard cats (*Prionailurus bengalensis*), three cougars (*Puma concolor*), three lions (*Panthera leo*), and three Siberian tigers (*Panthera tigris altaica*), were stored at  $-70^{\circ}\text{C}$  or in liquid nitrogen. One additional liver sample from a domestic cat was received from a local veterinarian and stored at  $4^{\circ}\text{C}$  in RNAlater (Ambion, Austin, TX; Table S1).

### Primer Design

Primers for Polymerase Chain Reactions (PCR) were designed against the predicted feline *CYP2D6*-mRNA sequence retrieved from the “Genome Annotation Resource Fields” (GARFIELD) online browser (Pontius and O’Brien 2007) and placed in conserved regions based on an alignment of mRNA sequences from *Homo sapiens*, *Pan troglodytes*, *Macaca mulatta*, *Bos taurus*, *Canis lupus familiaris*, and *Felis catus*, resulting in selected primers FcCYP2D6-190fb (5'-AGCCCAGCACGCACTGAG-3'), FcCYP2D6-4191r (5'-CAGGAGCACAGGGATGGAGTTCAGG-3'), FcCYP2D6-6125f (5'-ATAAAGCTGTGAGCAACG-3') and FcCYP2D6-8179r (5'-ACTGGTTTATTGCATATCAGG-3').

### RNA Extraction and Reverse Transcription

Total RNA was extracted using RNeasy Mini Kit (Qiagen) or SV Total RNA Isolation System (Promega) according to manufacturer’s guidelines and stored at  $-60^{\circ}\text{C}$ . cDNA was prepared using the ImProm-II<sup>TM</sup> Reverse Transcription System (Promega) and reverse transcription reactions were carried out with both random and oligodT primers using 1–2  $\mu\text{g}$  total RNA per reaction and in a final reaction volume of 40  $\mu\text{l}$ .

### Amplification and Sequencing of *CYP2D6*

A 719-bp long 5' fragment of the *CYP2D6* gene was amplified using primer pair FcCYP2D6-190fb/FcCYP2D6-4191r and a 1020-bp long 3' fragment was amplified using primer pair FcCYP2D6-6125f/FcCYP2D6-8179r. Each 25  $\mu\text{l}$  reaction mix consisted of 4  $\mu\text{l}$  of cDNA, 25 pmol of both forward and reverse primer, 5  $\mu\text{l}$  GC buffer (5 $\times$ ), 0.75  $\mu\text{l}$  DMSO, 0.25 mM dNTP mix, 0.5 U Phusion<sup>TM</sup> DNA Polymerase (New England Biolabs) and H<sub>2</sub>O or 2  $\mu\text{l}$  AmpliTaq<sup>®</sup> 360 Buffer, 10 $\times$ , 1 mM MgCl<sub>2</sub>, 1 mM dNTP mix, 0.75 U AmpliTaq<sup>®</sup> 360 DNA Polymerase (Applied Biosystems), and H<sub>2</sub>O. PCR conditions were as follows: denaturation of  $98^{\circ}\text{C}$  for 30 s, followed by 35 cycles at  $98^{\circ}\text{C}$  for 10 s,  $60^{\circ}\text{C}$  for 30 s and  $72^{\circ}\text{C}$  for 35 s, and final extension at  $72^{\circ}\text{C}$  for 10 min for the 5' fragment and identical conditions with an annealing temperature of  $55^{\circ}\text{C}$  for the 3' fragment. If necessary, PCR products were excised—from an agarose gel and purified using the Wizard<sup>®</sup> SV Gel and PCR Clean-Up System (Promega).

PCR products (20–50 ng) were sequenced in both directions (Big Dye, ABI) on an ABI 3130xl Genetic Analyzer and manually checked for heterozygote positions. Homology of the 5' and 3' PCR product was verified by a 100% sequence match of the 133-bp long overlapping region.

### *CYP2D6* Sequences Obtained from Public Databases

Complete mRNA sequences of putative *CYP2D6* orthologs of 11 mammals, including 33 cDNA sequences of functional human *CYP2D6* alleles were downloaded from GenBank. An mRNA sequence of the putative feline *CYP2D6* ortholog was obtained by searching the GARFIELD web browser generated by the Laboratory of Genomic Diversity and Advanced Biomedical Computing Center (Pontius and O’Brien 2007, Table 1).

### ND5, RAG2, GNAZ, and ATP7A Sequences

Sequences of three nuclear protein coding genes, recombination activating gene 2 (RAG2), guanine nucleotide

**Table 1** GenBank accession numbers of gene sequences used in this study

Species	CYP2D6 <sup>a</sup>	ATP7A	GNAZ	NADH-5	RAG2
<i>Acinonyx jubatus</i> (Cheetah)	HQ263387; HQ263388; HQ263389; HQ263390 <sup>b</sup>	DQ082604	DQ082128	n.s. <sup>c</sup>	DQ082344
<i>Bos taurus</i> (Cattle)	NM_174529.2	XM_615430.4	n.a.	NC_006853.1	NM_001082469.1
<i>Callithrix jacchus</i> (White-tufted- ear marmoset)	D29822.1	n.a. <sup>d</sup>	n.a.	n.a.	n.a.
<i>Canis lupus familiaris</i> (Dog)	NM_001003333.1	AY011436.1	XM_843512.1	NC_002008.4	XM_540537.2
<i>Equus caballus</i> (Horse)	NM_001111306.1	AY011430.1	XM_001488441.2	NC_001640.1	XM_001487973.2
<i>Felis catus</i> (Domestic cat)	HQ263391; HQ263392; HQ263393; HQ263394 <sup>b</sup> ChrB4:154039538..154047755 <sup>e</sup>	DQ082590	DQ082114	n.s.	DQ082330
<i>Homo sapiens</i> (Human)	AY545216.1; BC067432.1; BC075023.2; BC075024.2; BC106758.1; BC126858.1; DD182331.1-DD182334.1; DQ282144.1- DQ282147.1; DQ282151.1-DQ282159.1; EU530602.1; EU530603.1; GQ374179.1; M33388.1; NG_008376.1; BD318197.1- BD318200.1; NM_000106.4	NM_000052.4	NM_002073.2	NC_012920.1	NM_000536.2
<i>Lynx rufus</i> (Bobcat)	HQ263395; HQ263396; HQ263397 <sup>b</sup>	DQ082608	DQ082132	n.s.	DQ082348
<i>Macaca fascicularis</i> (Crab-eating macaque)	U38218.1	n.a.	n.a.	n.a.	n.a.
<i>Macaca mulatta</i> (Rhesus monkey)	XM_001107251.1	AY011415.1	XM_001092302.1	NC_005943.1	XM_001114748.1
<i>Mus musculus</i> (House mouse)	NM_019823.2	AY011397.1	NM_010311.3	NC_010339.1	NM_009020.3
<i>Pan paniscus</i> (Pygmy chimpanzee)	DQ282163.1	n.a.	n.a.	n.a.	n.a.
<i>Pan troglodytes</i> (Common chimpanzee)	NM_001040622.1	AF535046.1	XM_515016.2	NC_001643.1	XM_001155150.1
<i>Panthera leo</i> (Lion)	HQ263404; HQ263405; HQ263406 <sup>b</sup>	DQ082622	DQ082146	n.s.	DQ082362
<i>Panthera tigris altaica</i> (Siberian Tiger)	HQ263407; HQ263408; HQ263409 <sup>b</sup>	DQ082624	DQ082148	n.s.	DQ082364
<i>Prionailurus bengalensis</i> (Leopard cat)	HQ263398; HQ263399; HQ263400 <sup>b</sup>	DQ082599	DQ082123	n.s.	DQ082339
<i>Puma concolor</i> (Cougar)	HQ263401; HQ263402; HQ263403 <sup>b</sup>	DQ082602	DQ082126	n.s.	DQ082342
<i>Rattus norvegicus</i> (Norway rat)	NM_138515.1	AY011398.1	NM_013189.1	NC_001665.2	XM_230374.4

<sup>a</sup> Or putative ortholog<sup>b</sup> Sequences obtained from this study<sup>c</sup> Sequence data not submitted to Genbank (data from Johnson et al. 2006)<sup>d</sup> Sequence data not available. Species was excluded from analysis involving this gene<sup>e</sup> Positional information of the putative CYP2D6-gene. Data from GARFIELD-Web Browser (<http://lgs.abcc.ncifcrf.gov>)

binding protein, alpha z subunit (GNAZ) and ATPase, copper-transporting, alpha polypeptide (ATP7A), as well as of the mitochondrial NADH dehydrogenase subunit 5 (ND5) gene were either taken from Johnson et al. (2006) or downloaded from GenBank (Table 1). To the best of our knowledge these genes are either neutrally evolving or at least do not show signs of being under strong positive selection.

The nuclear genes were chosen based on sequence availability for the taxa under study and to provide a baseline for among taxa comparisons of dN/dS ratios. The dN/dS ratio is the ratio between the rate of nonsynonymous (dN) substitutions and the rate of synonymous (dS) substitutions in a coding sequence (Nei and Gojobori 1986; Yang and Nielsen 2000). Alignments were done using the ClustalW algorithm in MEGA 4.1 software (Tamura et al. 2007).

### Data Analysis

Pairwise *p* distances were calculated in MEGA 4.1 software and Maximum Likelihood (ML) distances in PAUP 4.0 (Swofford 2002) using substitution models chosen using MODELTEST 3.7 (Posada and Crandall 1998) via a Hierarchical Likelihood Ratio Test. Models were calculated separately for human alleles and the remaining sequences (including one human reference sequence). Sliding windows analysis was accomplished using DnaSP 5.10 (Librado and Rozas 2009) with 100 bp windows and 25 bp steps. The number of segregating sites (*s* values) were calculated for the global data set as well as subsets defined as “felids”, “human alleles”, and “primates without humans”. Putative secondary structure for all the sequences was predicted using the online tool GOR IV (Garnier et al. 1996) and a “consensus-secondary structure” was created by selecting the most frequent predicted secondary structure at each amino acid position. For each secondary structure type (alpha-helix, extended strand, and random coil) mean *s* values per site (standardized by dividing by the standard deviation of each subset after subtracting the arithmetic mean) were calculated. The numbers of synonymous and non-synonymous substitutions were calculated in 99-bp step windows. For both felids and humans, for which intraspecific data was available, synonymous and nonsynonymous substitutions were identified and mapped on a phylogenetic tree (from Johnson et al. 2006). These substitutions were then classified as either polymorphic (within) or fixed (between) species and placed in a 2 × 2 contingency table. The data were then subject to a McDonald–Kreitman test with a William’s correction (McDonald and Kreitman 1991).

Pairwise dN/dS ratios of nuclear genes were calculated in DnaSP following Nei and Gojobori (1986) with a Jukes

Cantor correction and plotted against pairwise distances between taxa drawn from the ND5 sequence data.

For the Felidae subset as well as for a “large dataset” (whole dataset but including only one reference sequence for humans), codon-specific dN/dS ratios were calculated with a ML approach using multiple models (M0: one ratio, M3: discrete, M7: beta and M8: beta and  $\omega$ ) implemented in the program CODEML of the PAML software package (Yang 2007). The tree topology for the PAML input file was based on combining the felid phylogeny from Johnson et al. (2006) with our own UPGMA-tree topology for the remaining taxa. Likelihood ratio tests (LRTs) were conducted by comparing likelihood values of two site-model pairs: M0 versus M3 and M8a versus M8. M3 and M8 are more general models than M0 and M8a, both allowing the presence of positive selection. Twice the log-likelihood differences of the model pairs have a  $\chi^2$  distribution and significant *P* values indicate the presence of positive selection. The M0 versus M3 comparison can also be considered as a test for a varying dN/dS ratios among codon sites (Yang et al. 2000a, b; Swanson et al. 2003). In addition, a LRT based on branch models (one ratio vs. free ratio model, Yang 1998), testing for varying dN/dS ratios among lineages, as well as the branch-site test of positive selection (Zhang et al. 2005), testing for positive selection on specified “foreground” branches within the tree (in our case, felids) were conducted. The degree of changes in amino acid properties indicating positive selection was calculated using the software TreeSAAP (Woolley et al. 2003). Out of the 31 amino acid properties implemented in this program, those putatively indicating positive selection were selected and a sliding window analysis involving the selected properties was conducted to detect regions under positive selection. Magnitudes of property changes were divided into eight categories and, to reduce false positives, only highly significant (*P* < 0.001) changes within categories 6–8 were considered as radical (McClellan et al. 2005).

### Results

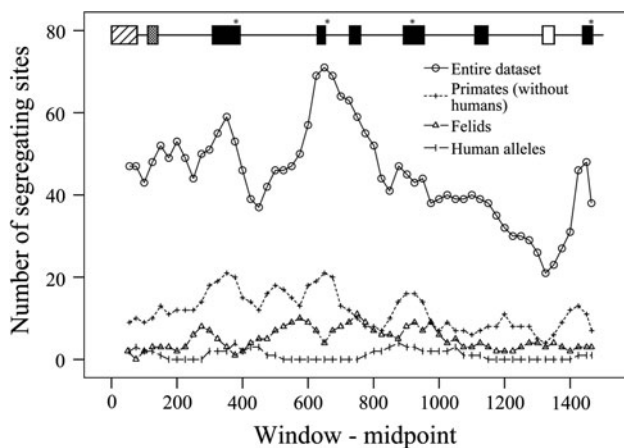
A total of 1506 base pairs were aligned in 67 sequences (Fig. S1; GenBank Acc. No.: Table 1). There were 676 segregating sites, 487 of which were parsimony informative resulting from 931 substitutions and 15 insertion/deletion polymorphisms. All primates carried a 9-bp deletion at position 10 resulting in three amino acids less than other species. Two Siberian tigers (PTI-72, 73) showed a 3 bp insertion (TAC) at positions 331–333 and therefore an additional tyrosine in the protein. A 3 bp deletion in the human allele *CYP2D6*\*9 occurred at

position 853–855. Overall mean pairwise  $p$  distances were 0.108 ( $\pm 0.087$ ) and 0.002 ( $\pm 0.001$ ) for inter- and intra-specific comparisons, respectively (Table S2).

Predicted amino acids were also aligned (Fig. S2); 257 of the 501 sites were variable and 187 of these were parsimony informative. The P450 signature motif was located at sites 440–449. Of the four amino acid residues involved in substrate binding (Rowland et al. 2006), only one was conserved. Little saturation was seen at first and third codon sites, whereas no saturation was observed with respect to transversions and substitutions at second codon positions.

### Sliding Window Analysis

Sliding window analysis revealed a high heterogeneity in sequence variation across the *CYP2D6* gene. The highest variability was found between positions 526 and 875 with a peak around position 650 (Fig. 1), whereas felid sequences displayed a low anomaly at this position, as well as some increase in variation near the 3' end (i.e., pos. 1251–1450). Primate sequences showed an additional signal of high variation around position 352. The most conserved region was between positions 1251 and 1450 encompassing the cysteine pocket; however, felids showed additional variation in this region. Relative to interspecific comparisons, human alleles showed little variation throughout the gene, but nonetheless showed their regions of highest  $s$  values between positions 250–500 and 750–1100.



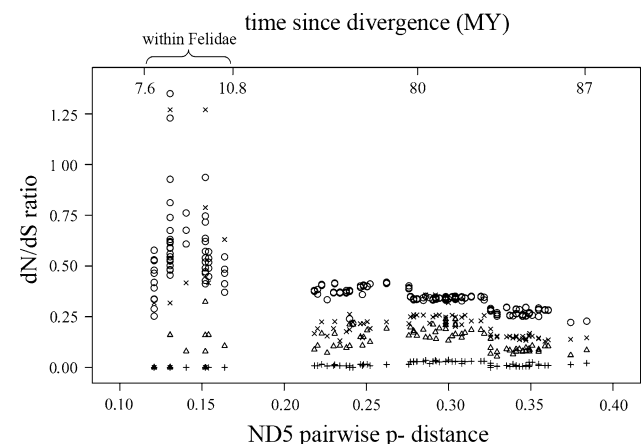
**Fig. 1** Results of sliding window analysis for various data subsets (window size: 100 bp; step size 25 bp): *Black bars on top* indicate substrate recognition sites described by Gotoh (1992). The membrane anchor (*striped*) and the proline-rich hinge region (*checked*) described by Yamazaki et al. (1993) as well as the cysteine pocket (*white bar*; Modi et al. 1996) are also shown. *Asterisks* indicate amino acids involved in substrate binding (Rowland et al. 2006)

### dN/dS Ratios

The mean dN/dS ratio of *CYP2D6* ( $0.565 \pm 0.484$ ) was higher than the other nuclear genes analyzed (i.e., RAG2,  $0.127 \pm 0.085$ ; GNAZ,  $0.008 \pm 0.112$ ; ATP7A,  $0.317 \pm 0.298$ ). Intraspecific *CYP2D6* dN/dS ratios were significantly higher ( $0.730 \pm 0.663$ ) than interspecific ratios ( $0.440 \pm 0.253$ ) (Mann–Whitney  $U$ ;  $P < 0.0001$ ). The distribution of dN/dS ratios showed a marked peak within Felidae, followed by a sharp reduction and gradual negative correlation with increasing evolutionary distance (Fig. 2). The highest interspecific dN/dS ratios were all within Felidae. Ratios  $>1$  involved tigers and lions, as well as puma and leopard cats, reflecting evolutionary distances of 3.72 and 6.7 Ma, respectively (Johnson et al. 2006).

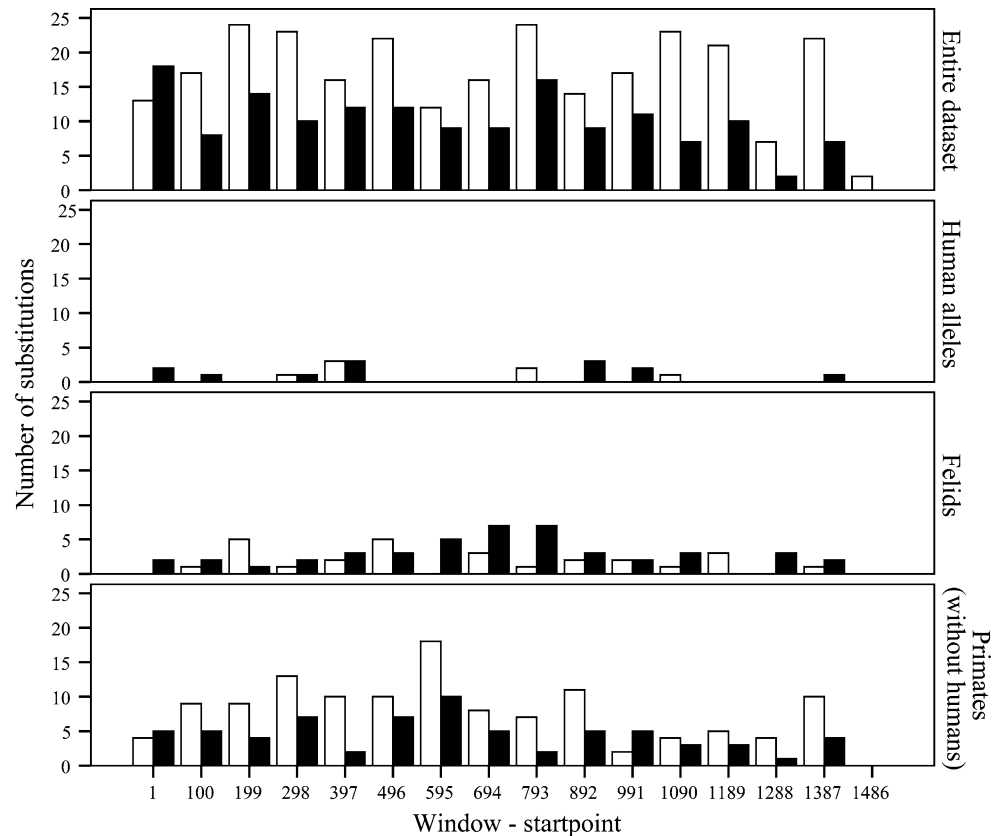
The number of synonymous substitutions exceeded the number of nonsynonymous substitutions for the gene as a whole. Across gene regions, only the first 99-bp window had more nonsynonymous (18) than synonymous substitutions (13) (Fig. 3). Primates showed the same pattern with five nonsynonymous and four synonymous substitutions in the first window. Within felids, 11 out of the 16 windows displayed more nonsynonymous than synonymous substitutions. Because of the low sequence variation in human alleles, no conclusion about the distribution of synonymous and nonsynonymous substitutions throughout the gene could be drawn. There were no significant differences in the number of segregating sites with respect to the three specified secondary structures of the gene.

The McDonald–Kreitman test revealed no statistically significant differences between the ratios of synonymous and nonsynonymous substitutions within- and between species (data not shown). PAML Likelihood ratio tests indicated both varying dN/dS ratios among sites and positive



**Fig. 2** Bi-variate plot of pairwise dN/dS ratios of *CYP2D6* (*open circle*), RAG2 (*open triangle*), GNAZ (*plus*) and ATP7A (*cross*) against pairwise  $p$  distances of the mtDNA NADH-5 gene. Divergence times are from Johnson et al. (2006) and Springer et al. (2003)

**Fig. 3** Distribution of synonymous and nonsynonymous substitutions throughout *CYP2D6* in various data subsets. *Open bars* represent synonymous substitutions, *filled bars* represent nonsynonymous substitutions



selection within felids (M0 vs. M3:  $P < 0.001$ ; M8 vs. M8a:  $P < 0.001$ ), as well as across the entire dataset (M0 vs. M3:  $P < 0.001$ ; M8a vs. M8:  $P = 0.0076$ ). Substrate recognition sites and the transmembrane anchor included multiple codons with high  $dN/dS$  ratios, whereas the conserved cysteine pocket and the proline-rich hinge region revealed only one ratio  $>1$  (Fig. 4a). Interestingly, the codon just upstream of the completely conserved cysteine showed a high number of nonsynonymous substitutions. Posterior probabilities, based on Bayes empirical Bayes (BEB) procedure, PAML detected positive selection for 16 codons within Felidae, and 15 overall (Table 2). A LRT of the two different branch models supported varying  $dN/dS$  ratios among lineages (Fig. S3). Based on LRT of the branch-site models, there was no statistical significant support for sites under positive selection within Felidae.

Over the whole dataset, three of 31 analyzed amino acid properties revealed signals of positive selection ( $P < 0.001$ ): alpha-helical tendencies, equilibrium constant and power to be at the C-terminal. Regions of significant positive selection in each of the three properties were identified by sliding window analysis (Fig. 4b). Codon sites with high  $dN/dS$  ratios showed high numbers of radical amino acid property changes (Table 2), compared to the mean across all sites ( $0.46 \pm 0.94$ ). For the Felidae subset, no amino acid property was detected to be under positive destabilizing selection.

## Discussion

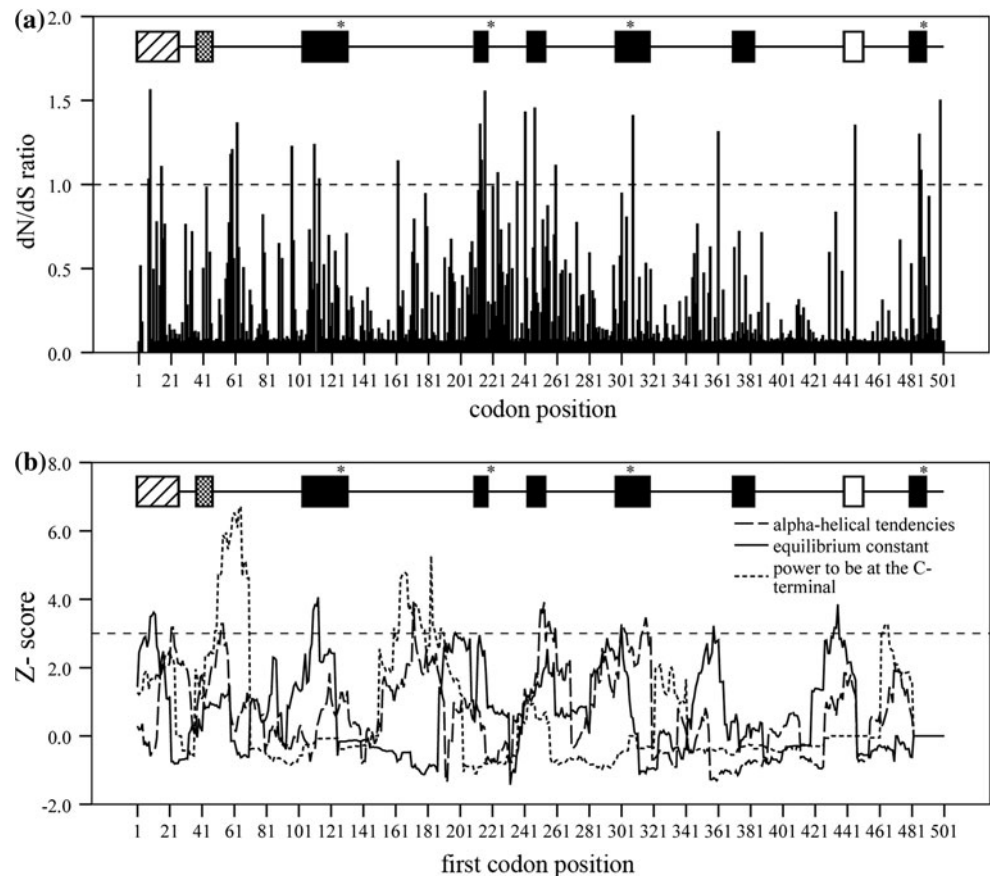
Our analysis of the *CYP2D6* gene revealed higher than average levels of variability, various signs of positive selection and distinct patterns of sequence variation with respect to specific functional domains.

Sliding window analysis clearly showed that the region of the highest sequence variability coincides with predicted substrate recognition sites. These sites have been shown to exhibit the highest sequence variability within P450s in general (Gotoh 1992; Werck-Reichhart and Feyereisen 2000). The region of lowest sequence variability is located around the heme-binding cysteine (Rowland et al. 2006). The conservation of this region and especially the cysteine AA itself is of great importance because it provides the thiolate ligand to bind the prosthetic group (Gotoh et al. 1983).

More detailed support for positive selection is gained through the  $dN/dS$  ratio analysis.

The mean  $dN/dS$  ratio across *CYP2D6* (0.565) was substantially higher than the four other nuclear genes tested, and considerably higher than averages reported in multiple studies of mammalian genes. Wang et al. (2007) reported a mean  $dN/dS$  ratio between human and mouse of 0.138 (12,052 genes) and between human and old world monkeys of 0.177 (14,204 genes). Zhang (2000) calculated

**Fig. 4** Results of PAML and TreeSAAP analyses for “large dataset”: **a** Mean  $dN/dS$  ratio for each codon. *Dotted line* indicates a ratio of 1. **b** Sliding window analysis of amino acid properties detected to be under positive selection. *Dotted line* represents cut-off values for significant ( $P < 0.001$ ) positive selection. *Black bars on top* indicate substrate recognition sites described by Gotoh (1992). The membrane anchor (*striped*) and the proline-rich hinge region (*checkered*) described by Yamazaki et al. (1993) as well as the cysteine pocket (*white bar*; Modi et al. 1996) are shown. *Asterisks* indicate amino acids involved in substrate binding



mean  $dN/dS$  ratios of 47 nuclear genes among primates (0.282), artiodactyls (0.274) and rodents (0.191). A study on 240 human–mouse orthologs, expressed specifically in liver, yielded an average  $dN/dS$  ratio of 0.233 (Zhang and Li 2004). *CYP2D6*, also a liver-expressed gene, exhibited a  $dN/dS$  ratio of 0.257 for the human–mouse comparison.

While some of our analytical results may suggest or support positive selection, the non-monotonic distribution of  $dN/dS$  values with respect to evolutionary distance (Fig. 2) might also be a sign of initial retention and eventual loss of slightly deleterious mutations (Balbi and Feil 2007). However, such patterns across evolutionary time frames of several million years do not correspond with literature examples of this phenomena, which involve comparisons within or among closely related species (Yang et al. 2000a, b; Rocha et al. 2006). While some effect from small sample bias cannot be excluded, an alternate explanation may relate to the fact that slightly deleterious mutations are not removed as effectively in species with small populations (Hughes and Friedman 2009). Three of the studied felids are listed as vulnerable or endangered by IUCN and felids in general have relatively low effective population sizes. But our comparisons are interspecific so it is not clear how large an effect small effective population sizes have on our data.

The distribution of codon-specific  $dN/dS$  ratios across the gene was not homogeneous as higher ratios were primarily located near substrate recognition sites or the membrane anchor. Interestingly, the distribution of radical changes in amino acid properties did not coincide with this pattern as such changes were not found at substrate recognition sites, nor were relevant within Felidae. We speculate that AA changes (or variability in general) at substrate recognition or member anchor sites could be of adaptive importance, but most likely AA changes do not involve “radical” property changes because such sites must maintain their basic properties to be functional. Thus, the approach developed with the program TreeSAAP, aimed at revealing radical AA changes that might be associated with adaptive significance may not have the same application potential across all types of genes. Such changes may signal major functional shifts in a gene, but not minor adjustments in substrate specificity.

Analysis within Felidae revealed a different pattern than that seen for the global data set in that the region around the heme pocket was not as conserved as within humans or primates, perhaps indicating relaxed selection pressure on this region, relative to other mammalian orders. In addition, the variability peak associated with substrate recognition sites was not as pronounced as for the global data

**Table 2** Results of PAML and TreeSAAP analyses

PAML						TreeSAAP	
Felid subset			Large subset			Large subset	
Codon	dN/dS	post. prob.	Codon	dN/dS	post. prob.	Codon	# rad. changes
8	6.845	0.992**	8	1.561	0.945	8	2
165	3.987	0.545	58	1.177	0.530	58	5
191	6.847	0.992**	59	1.206	0.552	59	3
212	6.517	0.936	62	1.364	0.721	62	2
224	5.414	0.771	96	1.226	0.575	96	0
236	5.263	0.749	110	1.236	0.587	110	1
241	6.514	0.936	213	1.357	0.708	213	2
256	3.790	0.517	216	1.552	0.933	216	5
273	6.525	0.937	241	1.430	0.786	241	2
312	6.167	0.881	247	1.452	0.811	247	2
346	5.525	0.787	308	1.407	0.761	308	3
438	5.051	0.718	361	1.313	0.663	361	4
446	6.255	0.896	446	1.349	0.700	446	4
460	4.678	0.664	486	1.297	0.646	486	3
493	4.880	0.692	499	1.499	0.866	499	2
499	6.704	0.969*					

Codon suggested codons under positive selection, dN/dS mean dN/dS ratio, post. prob. BEB posterior probability, and # rad. changes number of radical changes in selected amino acid properties

\* post. prob. > 0.95; \*\* post. prob. > 0.99

set. This result may relate to the fact that felids are strict carnivores and thus may be less exposed to plant toxins, long thought to be the most common P450 substrates.

Surprisingly, sequence variation of human alleles was very low despite the gene's high expression variability. While we only analyzed putatively functional alleles, several of them are described to result in decreased enzyme function. Moreover, a number of reported allelic variants are not functional. Thus, some sequence variability results in reduced or disrupted function, whereby variable copy number is most likely responsible for increased functionality. The prevalence of variable copy number or enzyme dysfunction of *CYP2D6* in non-humans is not known.

Another very important factor influencing the *CYP2D6* enzyme and its gene might be diet. So far, no clear pattern of P450 expression between species of particular dietary type, such as herbivores or carnivores has been found (Fink-Gremmels 2008), but several authors report dietary effects on P450 enzyme expression. For example, rabbits showed an increase in mRNA expression of certain P450s corresponding to weaning time (Pineau et al. 1991). Variation of quality and quantity of particular micro- and macro-nutrients in the diet has also been reported to influence P450 expression (Ioannides 1999 and references therein). However, it should be noted that those studies

only focused on the expression levels. Our data encompassed sequences of species representing different dietary types (e.g., carnivorous felids, herbivorous horse and cattle and omnivorous rat, mouse, and humans). Out of the four identified substrate binding sites of *CYP2D6*, only one was conserved and two revealed AA changes correlated with diet. Position 124 revealed a common phenylalanine across all carnivores and primates, whereas herbivores had either a valine or isoleucine; position 220 revealed a common aspartic acid across all carnivores whereas herbivores and omnivores had either glutamine or glutamine acid.

We conclude that the *CYP2D6* gene in non-human mammals exhibits considerable variation that may reflect some adaptive significance and should be considered a candidate gene in understanding dietary adaptation. Such studies may be particularly relevant in domesticated animals. Domesticated species tend to have elevated dN/dS ratios as compared to their wild relatives due to reduced selection pressure, altered breeding patterns and reduced effective population sizes (Cruz et al. 2008; MacEachern et al. 2009). Enzymatic studies on domestic cats report variable *CYP2D* activities, at least between males and females (Shah et al. 2007). Thus, there is some potential to pursue our general overview of felids with a more extensive intra-specific study of domestic cats. Whether based on copy number or sequence variation, a more thorough consideration of *CYP2D6* function in domestic or veterinary animals may uncover breed-specific differences that can support more effective drug therapy.

**Acknowledgments** We thank Stephen O'Brien for his generous support, provision of tissue samples and access to facilities. We also thank Natalie Baggett for cDNA preparation. The work was partially funded by the Faculty of Natural Sciences of the Karl-Franzens University Graz. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and has also been funded in part with federal funds from the National Cancer Institute and National Institutes of Health, under contract HHSN26120080001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

## References

- Balbi KJ, Feil EJ (2007) The rise and fall of deleterious mutation. Res Microbiol 158:779–786
- Bull S, Nijmeier S, Werners A, Fink-Gremmels J (2003) Cytochrome P450 2D6 polymorphism in horses and ponies. J Vet Pharmacol Therap 26:85–86
- Court MH, Greenblatt DJ (1997) Molecular basis for deficient acetaminophen glucuronidation in cats. An interspecies comparison of enzyme kinetics in liver microsomes. Biochem Pharmacol 53:1041–1047



- Court MH, Hay-Kraus BL, Hill DW, Kind AJ, Greenblatt DJ (1999) Propofol hydroxylation by dog liver microsomes: assay development and dog breed differences. *Drug Metab Dispos* 27:1293–1299
- Cruz F, Vila C, Webster MT (2008) The legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol Biol Evol* 25:2331–2336
- Dong H, Haining RL, Thummel KE, Rettie AE, Nelson SD (2000) Involvement of human cytochrome P450 2D6 in the bioactivation of acetaminophen. *Drug Metab Dispos* 28:1397–1400
- Fink-Gremmels J (2008) Implications of hepatic cytochrome P450-related biotransformation processes in veterinary sciences. *Eur J Pharmacol* 585:502–509
- Frank D, Jaehde U, Fuhr U (2007) Evaluation of probe drugs and pharmacokinetic metrics for CYP2D6 phenotyping. *Eur J Clin Pharmacol* 63:321–333
- Gaedigk A, Blum M, Gaedigk R, Eichelbaum M, Meyer UA (1991) Deletion of the entire cytochrome P450 CYP2D6 gene as a cause of impaired drug metabolism in poor metabolizers of the debrisoquine/sparteine polymorphism. *Am J Hum Genet* 48:943–950
- Garfinkel D (1958) Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions. *Arch Biochem Biophys* 77:493–509
- Garnier J, Gibrat J-F, Robson B (1996) GOR secondary structure prediction method version IV. *Methods Enzymol* 266:14
- Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* 267:83–90
- Gotoh O, Tagashira Y, Iizuka T, Fujii-Kuriyama Y (1983) Structural characteristics of cytochrome P-450. Possible location of the heme-binding cysteine in determined amino-acid sequences. *J Biochem* 93:807–817
- Guengerich FP (2008) Cytochrome p450 and chemical toxicology. *Chem Res Toxicol* 21:70–83
- Hasler JA (1999) Pharmacogenetics of cytochromes P450. *Mol Aspects Med* 20(12–24):25–137
- Hughes AL, Friedman R (2009) More radical AMINO ACID replacements in primates than in rodents: support for the evolutionary role of effective population size. *Gene* 440:50–56
- Ingelman-Sundberg M (2004) Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms. *Naunyn-Schmiedeberg Arch Pharmacol* 369:89–104
- Ingelman-Sundberg M (2005) Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. *Pharmacogenomics* 5:6–13
- Ioannides C (1999) Effect of diet and nutrition on the expression of cytochromes P450. *Xenobiotica* 29:109–154
- Ioannides C (2006) Cytochrome p450 expression in the liver of food-producing animals. *Curr Drug Metab* 7:335–348
- Johansson I, Lundqvist E, Bertilsson L, Dahl ML, Sjoqvist F, Ingelman-Sundberg M (1993) Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. *Proc Natl Acad Sci USA* 90:11825–11829
- Johnson WE, Eizirik E, Pecon-Slattery J, Murphy WJ, Antunes A, Teeling E, O'Brien SJ (2006) The late Miocene radiation of modern Felidae: a genetic assessment. *Science* 311:73–77
- Lazarou J, Pomeranz BH, Corey PN (1998) Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 279:1200–1205
- Leandro-Garcia LJ, Leskela S, Montero-Conde C, Landa I, Lopez-Jimenez E, Leton R, Seeringer A, Kirchheiner J, Cascon A, Robledo M, Rodriguez-Antona C (2009) Determination of CYP2D6 gene copy number by multiplex polymerase chain reaction analysis. *Anal Biochem* 389:74–76
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
- MacEachern S, McEwan J, McCulloch A, Mather A, Savin K, Goddard M (2009) Molecular evolution of the Bovini tribe (Bovidae, Bovinae): is there evidence of rapid evolution or reduced selective constraint in domestic cattle? *BMC Genomics* 10:179
- McClellan DA, Palfreyman EJ, Smith MJ, Moss JL, Christensen RG, Sailsbery JK (2005) Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. *Mol Biol Evol* 22:437–455
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654
- Modi S, Paine MJ, Sutcliffe MJ, Lian LY, Primrose WU, Wolf CR, Roberts GC (1996) A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. *Biochemistry* 35:4540–4550
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Oscarson M, Hidestrand M, Johansson I, Ingelman-Sundberg M (1997) A combination of mutations in the CYP2D6\*17 (CYP2D6Z) allele causes alterations in enzyme function. *Mol Pharmacol* 52:1034–1040
- Paulson SK, Engel L, Reitz B, Bolten S, Burton EG, Maziasz TJ, Yan B, Schoenhard GL (1999) Evidence for polymorphism in the canine metabolism of the cyclooxygenase 2 inhibitor, celecoxib. *Drug Metab Dispos* 27:1133–1142
- Pineau T, Daujat M, Pichard L, Girard F, Angevain J, Bonfils C, Maurel P (1991) Developmental expression of rabbit cytochrome P450 CYP1A1, CYP1A2 and CYP3A6 genes. Effect of weaning and rifampicin. *Eur J Biochem* 197:145–153
- Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, Farrar K, Park BK, Breckenridge AM (2004) Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ* 329:15–19
- Pontius JU, O'Brien SJ (2007) Genome Annotation Resource Fields-GARFIELD: a genome browser for *Felis catus*. *J Hered* 98:386–389
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Rocha EP, Smith JM, Hurst LD, Holden MT, Cooper JE, Smith NH, Feil EJ (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol* 239:226–235
- Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, Oxbrow AK, Lewis CJ, Tennant MG, Modi S, Eggleston DS, Chenery RJ, Bridges AM (2006) Crystal structure of human cytochrome P450 2D6. *J Biol Chem* 281:7614–7622
- Seliskar M, Rozman D (2007) Mammalian cytochromes P450—importance of tissue specificity. *Biochim Biophys Acta* 1770:458–466
- Shah SS, Sanda S, Regmi NL, Sasaki K, Shimoda M (2007) Characterization of cytochrome P450-mediated drug metabolism in cats. *J Vet Pharmacol Ther* 30:422–428
- Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci USA* 100:1056–1061
- Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 20:18–20
- Swofford DL (2002) PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods), Version 4. Sinauer Associates, Sunderland, Massachusetts

- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Trepanier LA (2006) Cytochrome P450 and its role in veterinary drug interactions. *Vet Clin North Am Small Anim Pract* 36:975–985
- Wang HY, Chien HC, Osada N, Hashimoto K, Sugano S, Gojobori T, Chou CK, Tsai SF, Wu CI, Shen CK (2007) Rate of evolution in brain-expressed genes in humans and other primates. *PLoS Biol* 5:0335–0342
- Weide vd,J, Steijns LS (1999) Cytochrome P450 enzyme system: genetic polymorphisms and impact on clinical pharmacology. *Ann Clin Biochem* 36(Pt 6):722–729
- Werck-Reichhart, D., and R. Feyereisen. 2000. Cytochromes P450: a success story. *Genome Biol* 1:REVIEWS3003
- Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19:671–672
- Yamazaki S, Sato K, Suhara K, Sakaguchi M, Mihara K, Omura T (1993) Importance of the proline-rich region following signal-anchor sequence in the formation of correct conformation of microsomal cytochrome P-450s. *J Biochem* 114:652–657
- Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568–573
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 17:32–43
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000a) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449
- Yang Z, Swanson WJ, Vacquier VD (2000b) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol Biol Evol* 17:1446–1455
- Zhang J (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol* 50:56–68
- Zhang L, Li WH (2004) Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol Biol Evol* 21:236–239
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479