

Preliminary analysis of the mitochondrial genome evolutionary pattern in primates

Liang ZHAO^{1,2}, Xingtao ZHANG¹, Xingkui TAO¹, Weiwei WANG¹, Ming LI^{2,*}

1. Faculty of Biology, Suzhou University, Suzhou Anhui 234000, China;

2. Institute of Zoology, the Chinese Academy of Sciences, Beijing 100101, China

Abstract: Since the birth of molecular evolutionary analysis, primates have been a central focus of study and mitochondrial DNA is well suited to these endeavors because of its unique features. Surprisingly, to date no comprehensive evaluation of the nucleotide substitution patterns has been conducted on the mitochondrial genome of primates. Here, we analyzed the evolutionary patterns and evaluated selection and recombination in the mitochondrial genomes of 44 Primates species downloaded from GenBank. The results revealed that a strong rate heterogeneity occurred among sites and genes in all comparisons. Likewise, an obvious decline in primate nucleotide diversity was noted in the subunit rRNAs and tRNAs as compared to the protein-coding genes. Within 13 protein-coding genes, the pattern of nonsynonymous divergence was similar to that of overall nucleotide divergence, while synonymous changes differed only for individual genes, indicating that the rate heterogeneity may result from the rate of change at nonsynonymous sites. Codon usage analysis revealed that there was intermediate codon usage bias in primate protein-coding genes, and supported the idea that GC mutation pressure might determine codon usage and that positive selection is not the driving force for the codon usage bias. Neutrality tests using site-specific positive selection from a Bayesian framework indicated no sites were under positive selection for any gene, consistent with near neutrality. Recombination tests based on the pairwise homoplasy test statistic supported complete linkage even for much older divergent primate species. Thus, with the exception of rate heterogeneity among mitochondrial genes, evaluating the validity assumed complete linkage and selective neutrality in primates prior to phylogenetic or phylogeographic analysis seems unnecessary.

Keywords: Mitochondrial genome; Evolutionary pattern; Codon usage bias; Complete linkage; Evolution neutrality; Primates

The mitochondrial genome of vertebrates exhibits several peculiar features including maternal inheritance, the presence of single-copy orthologous genes, a lack of recombination, evolutionary neutrality, and a high mutation rate—characteristics that make it well suited for evolutionary studies (Pesole et al, 1999; Saccone et al, 2000). Among all the characteristics and assumptions of the mitochondrial genome, the two most important aspects for explaining the evolutionary process are the presumed freedom from positive Darwinian (adaptive) natural selection and lack of recombination. Mitochondrial proteins are central to the cellular oxidative phosphorylation pathway and are functionally conserved across metazoan phyla (Gray et al, 1999). The importance of these mtDNA protein products supports the hypothesis that some variation in the mtDNA

molecule is adaptive. Nonetheless, of the numerous single mtDNA gene studies and several complete mitochondrial genome comparisons published over the past two decades, very few have been able to reject neutrality in favor of positive Darwinian selection (García-Martínez et al, 1998; Rand et al, 1994). Indeed, previous studies on vertebrates indicated that different levels of variability are attributable to functional constraints and/or slightly deleterious polymorphisms, a

Received: 19 January 2012; Accepted: 27 June 2012

Foundation items: This project was supported by the National Basic Research Program of China (973 Program: 2007CB411600), the Natural Science Foundation of China (30630016; 30570292)

* Corresponding author, E-mail: lim@ioz.ac.cn

pattern consistent with near neutrality (Nachman *et al.*, 1996; Templeton, 1996).

Recombination breaks down the correlation in genealogical history between different regions of a genome, which may be the incorrect inference of evolutionary history (Schierup & Hein, 2000). The issue of whether recombination occurs in the human mitochondrial genome remains controversial as the necessary enzymes for recombination are in fact present in the mitochondria, and a few paternal mitochondria do penetrate the egg during fertilization (Thyagarajan *et al.*, 1996), providing evidence that suggests recombination is possible, at least in humans. Recent broad surveys of animal mitochondrial genomes have also concluded that recombination is widespread (Piganeau *et al.*, 2004; Tsousis *et al.*, 2005). Another interesting issue about evolutionary patterns of the mitochondrial genome is the codon usage bias in its protein-coding genes. Many studies have analyzed the differences in codon usage between different organisms and between proteins within the same organism (Bulmer, 1991; Kanaya *et al.*, 2001; Sharp *et al.*, 1995) and one of their main conclusions is that there are strong correlations between codon usage and genomic GC content. Moreover, human codon usage may be determined solely by GC content and its isochores composition (Kanaya *et al.*, 2001). Despite suggestive evidence and hypothesis, the relationships between codon usage and the underlying evolutionary constraints are still not fully understood (Prat *et al.*, 2009).

Since the birth of molecular evolutionary analysis, the evolutionary relationships of our own order, primates, have been of central interest. Strangely, no comprehensive and accurate evaluation for the nucleotide substitution pattern(s) of the various mtDNA genes (*i.e.*, tRNA and rRNA genes, synonymous and nonsynonymous positions of protein-coding genes) have been conducted to date. Most primate studies demonstrated mtDNA evolutionary neutrality and no recombination, and generally reported mtDNA still evolves about 5- to 10-fold more rapidly than single-copy nuclear DNA (nDNA) (Brown *et al.*, 1982) at an overall nucleotide substitution rate of 1% per million years (Brown *et al.*, 1979; Wilson *et al.*, 1985).

As large libraries of whole-mtDNA genome sequences from primates accumulate, we are able to conduct more sophisticated genome-wide analyses to obtain solid information on the pattern of whole-mtDNA genome evolution in primates while also determining the relative evolutionary rate of each mitochondrial component. Here, we report on the rates and patterns of mtDNA sequence variation and the codon usage in primates and describe the results of tests conducted to detect recombination and selection. These results will not only be of interest in their own right but will also have implications for the use of mtDNA in evolutionary studies of primates.

Materials and Methods

Sequences

54 Mitogenomic sequences from 44 primate species were downloaded from GenBank (Table 1). All 54 sequences were aligned, edited, and compared using the Sequencher 4.5 (Gene Codes Corporation, Ann Arbor, MI). We excluded all gaps and ambiguous alignment sites, resulting in 14,764 bp of sequence from the 13 protein-coding, 2 rRNA, and 21 of the 22 tRNA genes. Protein-coding sequences from the L-strand-encoded genes (ND6 and 8 tRNA genes) were converted into complementary strand sequences. The tRNA-Glu gene and the Control Region were not included because the alignments of those sequences are difficult due to a large number of highly variable sites, insertions, and deletions.

Phylogenetic analyses and sequence diversity

The genealogical relationships between the mitochondrial genomes were analyzed by parsimony using PAUP* (Swofford, 2002). Bootstrapping (Felsenstein, 1985) was used to test monophyly. For this study, 1,000 pseudosamples were generated to estimate the bootstrap proportions.

Nucleotide diversity (π) was calculated for all 13 protein-coding, 2 rRNA, and 21 concatenated tRNA genes. Measures of sequence diversity were also calculated by the sliding window method using the DnaSP 4.10 (Rozas *et al.*, 2003). A window (500 bp in length) was moved along the sequences in steps of 50 bp. Patterns of nucleotide diversity for all positions as well as patterns of synonymous and nonsynonymous nucleotide variation were calculated in each window, and the value was assigned to the nucleotide at the midpoint of the window. The gamma parameter alpha (α), which represents the extent of rate heterogeneity among sites, was estimated from the individual data sets with TREE-PUZZLE 5.2 (Schmidt *et al.*, 2002) using the 8-categories option.

Codon usage bias

Several parameters related to codon usage bias, such as the codon bias index (Morton, 1993), the effective number of codons (Wright, 1990), and G + C content at second and third positions as well as overall were estimated for each primate mitochondrial protein-coding genes using DnaSP 4.10 (Rozas *et al.*, 2003). Furthermore, to determine whether the compositional changes of the nucleotide content in the primate mitochondrial protein-coding genes are caused by directional mutation pressure or a result of positive selection, we performed a correlation analysis. If the nucleotide bias affects both the synonymous and nonsynonymous sites in protein-coding genes, positive selection could not solely explain this phenomenon, because positive selection should not affect silent

Table 1 Primate taxa and GenBank accession numbers for mitochondria genome genes.

Family	Taxon	GenBank Accession No.	Family	Taxon	GenBank Accession No.
Galagidae	<i>Galago senegalensis</i>	AB371092		<i>Presbytis melalophos</i>	DQ355299
	<i>Otolemur crassicaudatus</i>	AB371093		<i>Pygathrix nemaeus</i>	DQ355302
Lorisidae	<i>Nycticebus coucang</i>	AJ309867		<i>Pygathrix roxellana</i>	DQ355300
	<i>Loris tardigradus</i>	AB371094		<i>Nasalis larvatus</i>	DQ355298
	<i>Perodicticus potto</i>	AB371095		<i>Chlorocebus pygerythrus 1</i>	EF597501
Daubentonidae	<i>Daubentonia madagascariensis 1</i>	AM905039		<i>Chlorocebus pygerythrus 2</i>	EF597500
	<i>Daubentonia madagascariensis 2</i>	AB371085		<i>Cercopithecus aethiops sabaues</i>	DQ069713
Indriidae	<i>Propithecus verreauxi</i>	AB286049		<i>Papio hamadryas</i>	Y18001
Lemuridae	<i>Varecia variegata</i>	AB371089		<i>Theropithecus gelada</i>	FJ785426
	<i>Lemur catta</i>	AJ421451		<i>Macaca sylvanus</i>	AJ309865
	<i>Eulemur macaco</i>	AB371088		<i>Macaca thibetana</i>	EU294187
	<i>Eulemur mongoz</i>	AM905040		<i>Macaca mulatta</i>	AY612638
	<i>Eulemur fulvus mayottensis</i>	AB371087		<i>Macaca fascicularis</i>	FJ906803
	<i>Eulemur fulvus fulvus</i>	AB371086	Hylobatidae	<i>Hylobates lar</i>	X99256
Tarsiidae	<i>Tarsius syrichta</i>	AB371090		<i>Pongo pygmaeus</i>	D38115
	<i>Tarsius bancanus</i>	AF348159		<i>Pongo abelii</i>	X97707
Cebidae	<i>Saguinus oedipus</i>	FJ785424		<i>Gorilla gorilla 1</i>	D38114
	<i>Cebus albifrons</i>	AJ309866		<i>Gorilla gorilla 2</i>	X93347
Aotidae	<i>Aotus lemurinus</i>	FJ785421		<i>Pan troglodytes 1</i>	EU095335
Atelidae	<i>Ateles belzebuth</i>	FJ785422		<i>Pan troglodytes 2</i>	D38113
Pitheciidae	<i>Callicebus donacophilus</i>	FJ785423		<i>Pan troglodytes 3</i>	X93335
Cebidae	<i>Saimiri sciureus 1</i>	AB371091		<i>Pan paniscus</i>	D38116
	<i>Saimiri sciureus 2</i>	FJ785425		<i>Homo sapiens 1</i>	AM948965
Cercopithecidae	<i>Procolobus badius</i>	DQ355301		<i>Homo sapiens 2</i>	X93334
	<i>Colobus guereza</i>	AY863427		<i>Homo sapiens 3</i>	D38112
	<i>Semnopithecus entellus</i>	DQ355297		<i>Homo sapiens 4</i>	AM711903
	<i>Trachypithecus obscurus</i>	AY863425		<i>Homo sapiens 5</i>	AF346992

nucleotide positions. The correlation analysis of GC content at the second codon position (nonsynonymous mutations related, GC₂) and GC content at third position (synonymous mutations related, GC_{s3}) with CG content of all protein codon genes (GCC) were implemented using R 2.6.2 (Ihaka & Gentleman, 1996).

Tests of neutrality

To investigate protein evolution, we first calculated d_N/d_S ratios among the set of 54 primate data and tested their significance using Z-tests (Nei & Kumar, 2000) for each of the 13 protein-coding genes as implemented by MEGA 4.0 (Tamura et al, 2007). We also tested for site-specific positive selection with Bayesian posterior probabilities under the Ny98 fitness regime (0 < ω₁ < 1, ω₂ = 1, and ω₃ > 1) (Nielsen & Yang, 1998) as described and implemented by MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003).

Tests of recombination

We calculated the Kelly (1997) measure of linkage

disequilibrium (ZnS); ZnS being the average of R² (Hill & Robertson, 1968) over all pairwise comparisons. We also employed the pairwise homoplasy test (Φ_W), a relatively new and robust statistical estimator of recombination (Buen et al, 2006) that is a coalescent-based estimator of the genealogical correlation or compatibility among sites, which is negatively correlated to recombination. The estimator was calculated, and its significance was estimated via a permutation test using Splitstree 4.10 (Huson & Bryant, 2006).

Results

Phylogenetic analyses and sequence diversity

The parsimony tree for primates, based on the mitochondrial genome, is shown in Figure 1. Each internal node was supported by more than 80% bootstrap (data not shown). The mtDNA polymorphism data and estimates of nucleotide divergence are summarized in Table 2, and the distributions of sequence diversity across the mitochondrial genome are presented in Figure



Figure 1 Parsimony tree of 54 primates based on mitochondrial genome ($H_1 \sim H_6$ were selected nodes for recombination tests)

2. An obvious decline in primate nucleotide diversity was noted in the subunit rRNAs and tRNAs as compared to the protein-coding genes. Among the 13 protein-coding genes, ATP8-ATP6 and the NADH dehydrogenase (ND) complex (with the exception of ND1) showed higher rates of sequence diversity, while the cytochrome oxidase (COX) complex exhibited lower diversity within Primates. Notably, the nonsynonymous divergence plots were similar to those obtained from the total nucleotide divergence, but the trend was not observed in the sliding-window plot of synonymous divergence. Although the COX complex exhibited an obvious decline in total nucleotide divergence and nonsynonymous divergence, these genes had higher rates of synonymous diversity. Based on the estimated gamma parameter α (α), the rate heterogeneity among sites was not even, but the pattern of heterogeneity among sites was distributed

across the primate mitochondrial genome.

Codon usage

Several parameters related to codon usage bias were estimated to check whether synonymous mutations are selectively neutral. The codon bias index (CBI) is a measure of the deviation from the equal use of synonymous codons with values ranging from 0 (uniform use of synonymous codons) to 1 (maximum codon bias). CBI values were intermediate in our study, ranging from 0.354 to 0.504 (Table 3). The effective number of codons (ENC), which may range from 20 (only one codon is used for each amino acid; i.e., the codon bias is maximum) to 61 (all synonymous codons for each amino acid are equally used; i.e., no codon bias), were likewise intermediate in the primates mitochondrial protein condoning genes (Table 3). GC content at the second

Table 2 Characteristics for each gene data set in the 54 Mitogenomic sequences from 44 primate species

Gene data set	<i>N</i>	π_{total}	$\pi_{syn.}$	$\pi_{nonsyn.}$	α
Protein genes					
ND1	954	0.22594	0.56062	0.10797	0.6400
ND2	1032	0.28488	0.55369	0.19488	0.6829
COX1	1533	0.19501	0.65372	0.04446	0.6678
COX2	678	0.24309	0.62540	0.11977	0.6908
ATP8- ATP6	834	0.28099	0.59298	0.17451	0.8017
COX3	783	0.21378	0.60615	0.08494	0.7019
ND3	342	0.26373	0.58340	0.15246	0.9531
ND4L	294	0.25770	0.61033	0.13563	0.6112
ND4	1359	0.25191	0.56225	0.14085	0.6478
ND5	1788	0.25525	0.56755	0.14766	0.7309
ND6	486	0.25715	0.50608	0.17487	0.6829
Cyt b	1134	0.22837	0.57817	0.11319	0.7729
13 concatenated protein genes	11217	0.24287	0.58072	0.13166	0.7410
RNA genes					
12s rRNA	864	0.15497			0.6910
16s rRNA	1424	0.16676			0.6700
21 concatenated tRNA genes (excluding lysine tRNA)	1259	0.13533			0.5689
All genes	14764	0.22121			0.7229

N: Number of sites; π_{total} : nucleotide diversity for all sites; $\pi_{syn.}$: nucleotide diversity for synonymous sites; $\pi_{nonsyn.}$: nucleotide diversity for nonsynonymous sites; α : rate heterogeneity index / Gamma parameter.

position (GC₂) ranged from 38.0% to 40.4%, GC content at third codon position (GC_{3s}) ranged from 30.7% to 50.4%, and overall CG content of the protein codon genes (GCc) ranged from 37.4% to 46.5%. Correlation analysis indicates that the nucleotide bias affects both the synonymous and nonsynonymous sites in protein-coding genes. Likewise, there is a highly significant correlation between the GC_{3s} and GCc ($R^2 = 0.9932$, $P < 0.001$; Figure 3 A). The correlation between GC₂ and GCc is also highly significant ($R^2 = 0.7772$, $P < 0.001$; Figure 3B).

Neutrality Tests

As a conservative measure of selection, Z-tests were conducted for each protein-coding gene in all species. In all cases, d_N/d_S was highly significant ($P = 0$), consistent with strong purifying selection of mitochondrial genes in these comparisons. The d_N/d_S ratios differed among genes (Figure 4) and these differences may potentially be attributable to positive selection or variation in levels of functional constraint among genes. The ATP8, ND2, and ND6 genes were characterized by a d_N/d_S ratio well above the average, while the COX1 gene ratios were the lowest among the species. Site-specific positive selection inferred from Bayesian posterior probabilities indicated that no sites were under positive selection for any gene of any of the species.

Table 3 Summary of codon usage index of 13 concatenated protein genes from 44 Primates species Mitogenomic sequences

Organism	ENC	CBI	GC ₂	GC _{3s}	GCc
<i>Galago senegalensis</i>	44.147	0.406	0.393	0.390	0.412
<i>Otolemur crassicaudatus</i>	43.363	0.428	0.399	0.449	0.441
<i>Nycticebus coucang</i>	41.565	0.470	0.397	0.400	0.423
<i>Loris tardigradus</i>	46.337	0.363	0.394	0.386	0.413
<i>Perodicticus potto</i>	42.444	0.461	0.390	0.348	0.394
<i>Daubentonia madagascariensis 1</i>	44.075	0.399	0.383	0.363	0.402
<i>Propithecus verreauxi</i>	43.910	0.406	0.399	0.417	0.420
<i>Varecia variegata</i>	44.723	0.389	0.389	0.327	0.388
<i>Lemur catta</i>	45.347	0.383	0.394	0.373	0.407
<i>Eulemur macaco</i>	44.925	0.383	0.393	0.381	0.410
<i>Eulemur mongoz</i>	45.215	0.380	0.394	0.382	0.410
<i>Eulemur fulvus mayottensis</i>	45.045	0.396	0.391	0.331	0.388
<i>Eulemur fulvus fulvus</i>	43.987	0.406	0.390	0.346	0.392
<i>Tarsius syrichta</i>	45.321	0.395	0.392	0.348	0.399
<i>Tarsius bancanus</i>	45.541	0.364	0.388	0.338	0.393
<i>Saguinus oedipus</i>	42.664	0.409	0.383	0.339	0.389
<i>Cebus albifrons</i>	42.905	0.398	0.385	0.307	0.379
<i>Aotus lemurinus</i>	44.815	0.367	0.385	0.334	0.390
<i>Ateles belzebuth</i>	44.227	0.385	0.385	0.325	0.388
<i>Callicebus donacophilus</i>	44.608	0.367	0.387	0.331	0.391
<i>Saimiri sciureus 1</i>	44.552	0.368	0.387	0.331	0.390
<i>Procolobus badius</i>	45.041	0.420	0.401	0.458	0.441
<i>Colobus guereza</i>	44.081	0.427	0.397	0.447	0.434
<i>Semnopithecus entellus</i>	44.262	0.415	0.403	0.428	0.429
<i>Trachypithecus obscurus</i>	46.006	0.394	0.382	0.328	0.388
<i>Presbytis melalophos</i>	46.065	0.357	0.389	0.379	0.407
<i>Pygathrix nemaeus</i>	46.553	0.373	0.387	0.321	0.384
<i>Pygathrix roxellana</i>	46.949	0.354	0.387	0.333	0.387
<i>Nasalis larvatus</i>	45.514	0.394	0.384	0.311	0.380
<i>Chlorocebus pygerythrus 1</i>	44.783	0.415	0.400	0.434	0.431
<i>Cercopithecus aethiops sabaeus</i>	44.474	0.426	0.399	0.439	0.436
<i>Papio hamadryas</i>	44.504	0.425	0.396	0.440	0.434
<i>Theropithecus gelada</i>	44.556	0.426	0.400	0.442	0.436
<i>Macaca sylvanus</i>	44.741	0.434	0.400	0.438	0.434
<i>Macaca thibetana</i>	45.325	0.375	0.387	0.338	0.389
<i>Macaca mulatta</i>	45.586	0.377	0.380	0.307	0.374
<i>Macaca fascicularis</i>	46.576	0.374	0.387	0.324	0.385
<i>Hylobates lar</i>	43.281	0.457	0.397	0.498	0.461
<i>Pongo pygmaeus</i>	43.111	0.436	0.397	0.455	0.442
<i>Pongo abelii</i>	43.123	0.435	0.398	0.454	0.443
<i>Gorilla gorilla 1</i>	43.177	0.435	0.398	0.454	0.442
<i>Pan troglodytes 1</i>	44.396	0.427	0.394	0.456	0.443
<i>Pan paniscus</i>	41.489	0.504	0.404	0.504	0.465
<i>Homo sapiens 1</i>	42.622	0.457	0.397	0.474	0.451

ENC: Effective number of codons (Wright, 1990); CBI: Codon bias index (Morton, 1993); GC₂: GC content at the second position; GC_{3s}: GC content at third codon position; GCc: overall CG content.

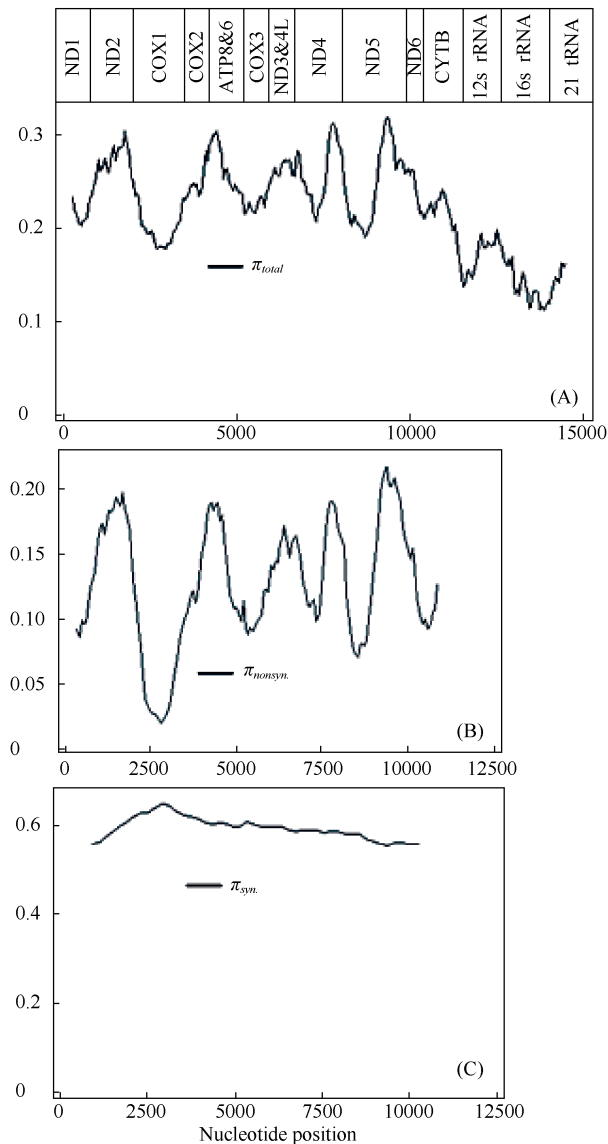


Figure 2 Sequence diversity of the mitochondrial genome by nucleotide position
Diversity measure was calculated for a sliding window of 500 bp with a step size of 50 bp. (A) Patterns of total nucleotide diversity (π_{total}) for all positions; (B) patterns of nonsynonymous nucleotide diversity ($\pi_{nonsyn.}$) for concatenated protein genes; (C) patterns of synonymous nucleotide diversity ($\pi_{syn.}$) for concatenated protein genes.

Tests of Recombination

The measures of linkage disequilibrium are summarized in Table 3. The average R^2 (Hill & Robertson 1968) across all pairwise comparisons (ZnS) for each individual data set and the pooled data provided evidence of recombination between relatively distant divergent species but not for more recent divergent species. The pairwise homoplasmy test statistic (Φ_W) for the pooled data also provided similar results to the ZnS-

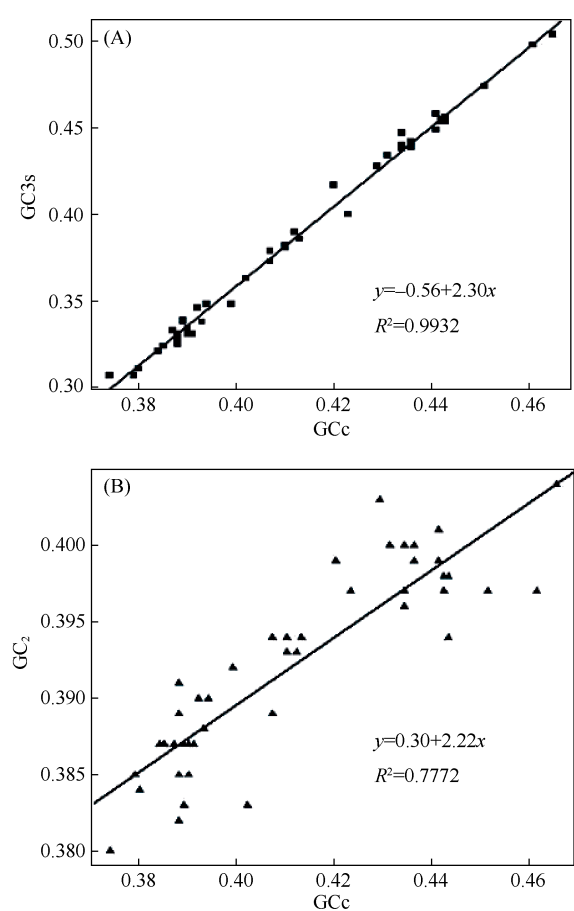


Figure 3 Correlation of GC content at the third codon position (GC_{s3}), and GC content at the second position (GC_2) with overall CG content (GC_c)
(A) Correlation of GC_{s3} and GC_c of coding sequences; (B) Correlation of GC_2 and GC_c of coding sequences.

test, but the Φ_W test statistic provided no evidence of recombination for each individual data set. Based on our data sets, the assumption of complete linkage of the mtDNA genome appears to be justified only between more recently divergent species.

Discussion

Primate mitochondrial genomes exhibit striking patterns of both polymorphism and substitution rate heterogeneity among sites and genes, as evidenced by sliding-window estimates of sequence diversity, nonsynonymous and synonymous divergence, and the estimated gamma parameter alpha (α) (Table 2, Figure 2). Less primate nucleotide diversity was observed in the subunit rRNAs and tRNAs than that in the protein-coding genes. Among the 13 coding genes, rates were higher in the ND genes and the ATP8- ATP6 gene and lower in the COX complex genes, consistent with the

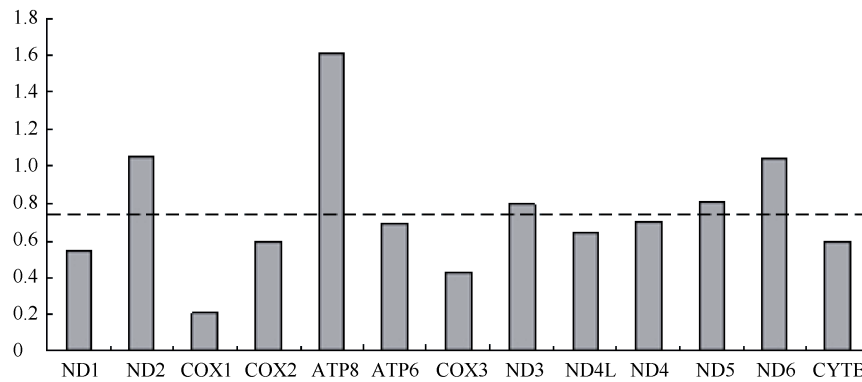


Figure 4 Ratios of nonsynonymous/synonymous rate per gene in the mitochondrial genomes of 54 primates. Dotted line indicates the average ratio.

Table 4 Recombination test for primate mitochondrial genomes.

Gene	Recombination test	Internal subclades (node)					
		H ₁	H ₂	H ₃	H ₄	H ₅	H ₆
ND1	ZnS	0.0405**	0.0458**	0.0588**	0.0770**	0.2884	0.7266
	Φ _W	1.2488	1.0927	0.8219	0.6476	0.1496	0.0659
ND2	ZnS	0.0489**	0.0566**	0.0699**	0.0810*	0.2874	0.7251
	Φ _W	1.4217	1.2213	0.9230	0.7143	0.1779	0.0579
COX1	ZaS	0.0424**	0.0487**	0.0611**	0.0784**	0.2823	0.7544
	Φ _W	1.6117	1.4236	1.1128	0.8265	0.1662	0.0345
COX2	ZnS	0.0795**	0.0874**	0.1147	0.0947*	0.2966	0.7880
	Φ _W	1.3825	1.1986	0.9053	0.8271	0.1300	0.0000
ATP8-ATP6	ZnS	0.0479**	0.0558**	0.0696**	0.0845**	0.3012	0.6755
	Φ _W	1.4795	1.2841	0.9917	0.8155	0.1424	0.0772
COX3	ZnS	0.0407**	0.0553**	0.0575**	0.0730**	0.2273	0.7828
	Φ _W	1.4295	1.2397	0.9946	0.7820	0.2722	0.0029
ND3	ZnS	0.0375**	0.0493**	0.0610**	0.0790*	0.2893	0.7436
	Φ _W	1.5716	1.3769	0.9752	0.7419	0.2004	0.0751
ND4L	ZnS	0.0406**	0.0489**	0.0628**	0.0676**	0.2877	0.7621
	Φ _W	1.3633	1.2171	0.9726	0.7554	0.1937	0.0811
ND4	ZnS	0.0445**	0.0513**	0.0579**	0.0787*	0.2811	0.7561
	Φ _W	1.4206	1.2099	0.9199	0.7240	0.1476	0.0101
ND5	ZnS	0.0401**	0.0458**	0.0608**	0.0770**	0.2812	0.7088
	Φ _W	1.3903	1.2037	0.9311	0.7230*	0.1596	0.0472
ND6	ZnS	0.0644*	0.0726*	0.0815*	0.0750**	0.2754	0.7926
	Φ _W	1.264	1.1078	0.9061	0.7237	0.1910	0.0163
CYT B	ZnS	0.0381**	0.0430**	0.0545**	0.0683*	0.2680	0.7586
	Φ _W	1.3633	1.2238	0.9792	0.7067	0.2383	0.0398*
All-Pro.	ZnS	0.0442**	0.0504**	0.0640**	0.0770**	0.2835	0.7394
	Φ _W	1.2291**	1.0311**	0.9620**	0.7483	0.1691	0.0411

* indicates $p < 0.05$; ** indicates $p < 0.01$

results from various mammals (Kumazawa et al, 2004). Most notably, the nonsynonymous divergence plots were similar to those obtained from the pooled data nucleotide divergence, while the synonymous changes were only slightly different between individual genes of mtDNA (Figure 2). Accordingly, rate heterogeneity may result from the rate of change at nonsynonymous sites.

The observed pattern of heterogeneity across the mitochondrial genome was similar to the pattern of

mitochondrial genome evolution at the intraspecific level in humans (Elson et al, 2004) and both the inter- and intraspecific levels in *Drosophila lard* (Ballard, 2000), where the synonymous distribution was more homogeneous and the nonsynonymous distribution was more heterogeneous. Additionally, the results obtained here differed from the inter- and intraspecific studies based on whole genome in *Gadus morhua* and its relative species (Marshall et al, 2009). Interestingly, all

mitochondrial genome comparisons mentioned above showed significant rate heterogeneity and site specific hyper-mutability. Possible explanations for rate heterogeneity in the mitochondrial genome include selection pressure, variation in the mutation rate of each gene (involving base composition bias and properties of the L-strand during mtDNA replication) (Marshall *et al.*, 2009), and the consequence of neutral mutational effects (Galtier *et al.*, 2006). The detected evolutionary patterns of the mitochondrial genome will influence the choice of mtDNA regions used for phylogeographic studies, since the more variable regions of the genome harbor excess homoplasy relative to the more slowly evolving regions (Galtier *et al.*, 2006). Marshall *et al.* (2009) proposed that the more variable regions of the mitochondrial genome harbor excess homoplasy relative to the more slowly evolving regions, so choosing to examine more slowly evolving regions for evolutionary analysis among species, and even within species, with an increase in the sequence length may be wiser. Based on the primate mitochondrial genome comparison in this study, we suggest that the subunit rRNAs genes, tRNAs genes, and the COX complex genes may be promising sites for future investigations of primate phylogeography.

CBIs estimated in this study revealed that there was intermediate codon usage bias in primate protein-coding genes (Table.3). Several evolutionary processes have been postulated as the major factors that determine codon usage. For a number of different organisms, it was suggested that codon usage was best explained by selection for tRNA abundance, gene expression levels, and translational optimization (Duret, 2000). In other cases, the codon usage was explained by mutation rate, mutation preference (Powell & Moriyama, 1997), environmental conditions (Goodarzi *et al.*, 2008), generation time (Subramanian, 2008), and recombination rates (Meunier & Duret, 2004). Although the relationships between codon usage and the underlying evolutionary constraints are still not fully understood, two major paradigms of codon usage have been found in most species (Prat *et al.*, 2009). Natural selection is expected to yield a correlation with codon bias in more highly expressed genes. Alternatively, codon bias could be generated by strictly neutral processes, such as mutational biases, where some mutations occur more than others across the genome by local variations in the base composition (Duret, 2002). Correlation analysis indicates that the nucleotide bias affects both the synonymous and nonsynonymous sites in primate protein-coding genes. There is a highly significant correlation between the GC3s and GCc ($R^2 = 0.9932$,

Figure 3A) and the correlation between GC₂ and GCc is also highly significant ($R^2 = 0.7772$, Figure 3B). Because positive selection should not affect silent nucleotide positions, our results support the idea that adaptive evolution is not the driving force for the codon usage bias, indicating that GC mutation pressure may determine codon usage in mitochondrial genomes of different primate species.

Using a conservative measure of selection, Z-tests indicated that $d_N < d_S$ was highly significant ($P = 0$) in all cases; this result is consistent with purifying selection of mitochondrial genes. A similar result was first found during analysis of the mitochondrial genomes of humans, chimpanzees, and gorillas (Hasegawa *et al.*, 1998) and also reported in additional studies on fruit flies, humans and Atlantic cod (Ballard, 2000; Elson *et al.*, 2004; Marshall *et al.*, 2009). However, site-specific positive selection from a Bayesian perspective indicates that no sites are under positive selection for any gene of any of the species, indicating that most mutations are nearly neutral or slightly deleterious, a pattern consistent with the nearly neutral theory of molecular evolution proposed by Ohta (1973). Our results therefore indicate that the evolutionary pattern of primate mitochondrial genomes and each single mitochondrial gene did not severely deviate from evolutionary neutrality.

A presumed absence of recombination is another key feature of mtDNA that makes it a useful phylogeographic molecule (Marshall *et al.*, 2009). The uniparental inheritance of mtDNA indicates that recombination should be rare or absent, but a broad survey of 267 published animal mtDNA data sets by Piganeau *et al.* (2004) provided indirect evidence of recombination in more than 14% of cases. Based on the pairwise comparisons (ZnS) and the pairwise homoplasy test statistic (Φ_W) for the 13 concatenated protein genes, the assumption of complete linkage of the mtDNA genome appears to be justified only between more recently divergent species. However, no evidence of recombination among the individual genes was detected with the pairwise homoplasy test statistic (Φ_W) even among the more divergent species, indicating that recombination is perhaps a result of substitution rate heterogeneity in the primate mtDNA genomes. Additionally, as the lineage divergence increases, the divergence of each individual gene may produce artificial significant ZnS values based on all pairwise comparisons from the pooled data and may lead to false detection of recombination. Thus, it appears unnecessary to consider the effects of recombination on the evolution of mtDNA in primates.

References

Ballard JWO. 2000. Comparative genomics of mitochondrial DNA in

Drosophila simulans [J]. *J Mol Evol*, **51**(1): 64-75.

- Brown WM, George M Jr, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA [J]. *Proc Natl Acad Sci USA*, **76**(4): 1967-1971.
- Brown WM, Prager EM, Wang A, Wilson AC. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution [J]. *J Mol Evol*, **18**(4): 225-239.
- Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination [J]. *Genetics*, **172**(4): 2665-2681.
- Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage [J]. *Genetics*, **129**(3): 897-907.
- Duret L. 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes [J]. *Trends Genet*, **16**(7): 287-289.
- Duret L. 2002. Evolution of synonymous codon usage in metazoans [J]. *Curr Opin Genet Dev*, **12**(6): 640-649.
- Elson JL, Turnbull DM, Howell N. 2004. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection [J]. *Am J Hum Genet*, **74**(2): 229-238.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap [J]. *Evolution*, **39**(4): 783-791.
- Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K. 2006. Mutation hot spots in mammalian mitochondrial DNA [J]. *Genome Res*, **16**(2): 215-222.
- García-Martínez J, Castro JA, Ramón MR, Latorre A, Moya A. 1998. Mitochondrial DNA haplotype frequencies in natural and experimental populations of *Drosophila subobscura* [J]. *Genetics*, **149**(3): 1377-1382.
- Goodarzi H, Torabi N, Najafabadi HS, Archetti M. 2008. Amino acid and codon usage profiles: adaptive changes in the frequency of amino acids and codons [J]. *Gene*, **407**(1-2): 30-41.
- Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution [J]. *Science*, **283**(5407): 1476-1481.
- Hasegawa M, Cao Y, Yang Z. 1998. Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species [J]. *Mol Biol Evol*, **15**(11): 1499-1505.
- Hill WG, Robertson A. 1968. Linkage disequilibrium in finite populations [J]. *Theor Appl Genet*, **38**(6): 226-231.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees [J]. *Bioinformatics*, **17**(8): 754-755.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies [J]. *Mol Biol Evol*, **23**(2): 254-267.
- Ihaka R, Gentleman R. 1996. R: A language for data analysis and graphics [J]. *J Comput Graph Stat*, **5**(3): 299-314.
- Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. 2001. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis [J]. *J Mol Evol*, **53**(4-5): 290-298.
- Kelly JK. 1997. A test of neutrality based on interlocus associations [J]. *Genetics*, **146**(3): 1197-1206.
- Kumazawa Y, Azuma Y, Nishida M. 2004. Tempo of mitochondrial gene evolution: Can mitochondrial DNA be used to date old divergences? [J]. *Endocytobiosis Cell Res*, **15**(1): 136-142.
- Marshall HD, Coulson MW, Carr SM. 2009. Near neutrality, rate heterogeneity, and linkage govern mitochondrial genome evolution in Atlantic Cod (*Gadus morhua*) and other gadine fish [J]. *Mol Biol Evol*, **26**(3): 579-589.
- Meunier J, Duret L. 2004. Recombination drives the evolution of GC-content in the human genome [J]. *Mol Biol Evol*, **21**(6): 984-990.
- Morton BR. 1993. Chloroplast DNA codon use: evidence for selection at the *psb A* locus based on tRNA availability [J]. *J Mol Evol*, **37**(3): 273-280.
- Nachman MW, Brown WM, Stoneking M, Aquadro CF. 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees [J]. *Genetics*, **142**(3): 953-963.
- Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics [M]. New York: Oxford University Press.
- Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene [J]. *Genetics*, **148**(3): 929-936.
- Ohta T. 1973. Slightly deleterious mutant substitutions in evolution [J]. *Nature*, **246**(5428): 96-98.
- Pesole G, Gissi C, De Chirico A, Saccone C. 1999. Nucleotide substitution rate of mammalian mitochondrial genomes [J]. *J Mol Evol*, **48**(4): 427-434.
- Piganeau G, Gardner M, Eyre-Walker A. 2004. A broad survey of recombination in animal mitochondria [J]. *Mol Biol Evol*, **21**(12): 2319-2325.
- Powell JR, Moriyama EN. 1997. Evolution of codon usage bias in *Drosophila* [J]. *Proc Natl Acad Sci USA*, **94**(15): 7784-7790.
- Prat Y, Fromer M, Linial N, Linial M. 2009. Codon usage is associated with the evolutionary age of genes in metazoan genomes [J]. *BMC Evol Biol*, **9**(1): 285.
- Rand DM, Dorfsman M, Kann LM. 1994. Neutral and non-neutral evolution of drosophila mitochondrial DNA [J]. *Genetics*, **138**(3): 741-756.
- Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models [J]. *Bioinformatics*, **19**(12): 1572-1574.
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods [J]. *Bioinformatics*, **19**(18): 2496-2497.
- Saccone C, Gissi C, Lanave C, Larizza A, Pesole G, Reyes A. 2000. Evolution of the mitochondrial genetic system: an overview [J]. *Gene*, **261**(1): 153-159.
- Schierup MH, Hein J. 2000. Consequences of recombination on traditional phylogenetic analysis [J]. *Genetics*, **156**(2): 879-891.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing [J]. *Bioinformatics*, **18**(3): 502-504.
- Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF. 1995. DNA

- sequence evolution: the sounds of silence [J]. *Philos Trans R Soc Lond B Biol Sci*, **349**(1329): 241-247.
- Subramanian S. 2008. Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes [J]. *Genetics*, **178**(4): 2429-2432.
- Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony. (*and other methods), Version 4.0b10. Sunderland, Massachusetts: Sinauer Associates.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0 [J]. *Mol Biol Evol*, **24**(8): 1596-1599.
- Templeton AR. 1996. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates [J]. *Genetics*, **144**(3): 1263-1270.
- Thyagarajan B, Padua RA, Campbell C. 1996. Mammalian mitochondria possess homologous DNA recombination activity [J]. *J Biol Chem*, **271**(44): 27536-27543.
- Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E. 2005. Widespread recombination in published animal mtDNA sequences [J]. *Mol Biol Evol*, **22**(4): 925-933.
- Wilson AC, Cann RL, Carr SM, George M, Gyllenstein UB, Helms-Bychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics [J]. *Biol Linn Soc*, **26**(4): 375-400.
- Wright F. 1990. The 'effective number of codons' used in a gene [J]. *Gene*, **87**(1): 23-29.