

Molecular Evolution of Cytochrome *c* Oxidase Subunit IV: Evidence for Positive Selection in Simian Primates

Wei Wu,¹ Morris Goodman,^{1,2} Margaret I. Lomax,³ Lawrence I. Grossman¹

¹ Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, 540 E. Canfield Avenue, Detroit, MI 48201, USA

² Department of Anatomy and Cell Biology, Wayne State University School of Medicine, 540 E. Canfield Avenue, Detroit, MI 48201, USA

³ Department of Anatomy and Cell Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

Received: 22 May 1996 / Accepted: 24 November 1996

Abstract. Cytochrome *c* oxidase (COX) is a multi-subunit enzyme complex that catalyzes the final step of electron transfer through the respiratory chain on the mitochondrial inner membrane. Up to 13 subunits encoded by both the mitochondrial (subunits I, II, and III) and nuclear genomes occur in eukaryotic organisms ranging from yeast to human. Previously, we observed a high number of amino acid replacements in the human COX IV subunit compared to mouse, rat, and cow orthologues. Here we examined COX IV evolution in the two groups of anthropoid primates, the catarrhines (hominoids, cercopithecoids) and platyrrhines (ceboids), as well as one prosimian primate (lorisiform), by sequencing PCR-amplified portions of functional *COX4* genes from genomic DNAs. Phylogenetic analysis of the *COX4* sequence data revealed that accelerated nonsynonymous substitution rates were evident in the early evolution of both catarrhines and, to a lesser extent, platyrrhines. These accelerated rates were followed later by decelerated rates, suggesting that positive selection for adaptive amino acid replacement became purifying selection, preserving replacements that had occurred. The evidence for positive selection was especially pronounced along the catarrhine lineage to hominoids in which the nonsynonymous rate was first faster than the synonymous rate, then later much slower. The rates of

three types of “neutral DNA” nucleotide substitutions (synonymous substitutions, pseudogene nucleotide substitutions, and intron nucleotide substitutions) are similar and are consistent with previous observations of a slower rate of such substitutions in the nuclear genomes of hominoids than in the nuclear genomes of other primate and mammalian lineages.

Key words: COX — Primates — Nonsynonymous substitutions — Synonymous substitutions — Rates — Positive selection

Introduction

Cytochrome *c* oxidase (COX) is the terminal enzyme complex of the mitochondrial electron transport chain. It plays a critical role in energy metabolism by catalyzing the transfer of electrons from reduced cytochrome *c* to oxygen, forming water. It also acts as a proton pump, forming an ion gradient across the mitochondrial membrane that drives the synthesis of ATP (Kadenbach et al. 1987; Capaldi 1990). Crystallographic analysis has confirmed that COX is a 13-subunit enzyme in mammals (Tsukihara et al. 1995); the three largest subunits are encoded by mitochondrial DNA (mtDNA) and carry out the catalytic functions, whereas the functions of the ten nuclear DNA encoded subunits are largely unknown but likely to be concerned with regulation and assembly of the holoenzyme. COX is one of four multi-subunit en-

Abbreviations: CSY, changes/site/year; η , nucleotide substitutions; S and N, synonymous and nonsynonymous substitutions

Correspondence to: L.I. Grossman; e-mail: lg@cmb.biosci.wayne.edu

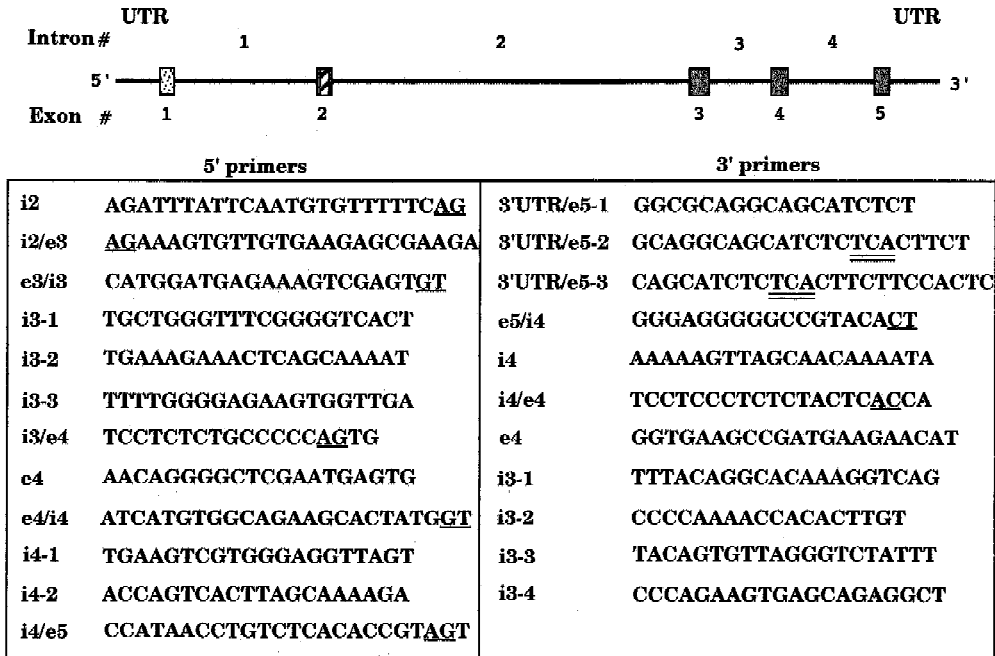


Fig. 1. PCR primers utilized. Primers and their locations are shown with respect to a diagram of the human *COX4* gene. Exons are boxed (stippled box, untranslated; stripes, leader sequence; filled, mature protein coding region) and both exons (e) and introns (i) are numbered. Single underline, intron-exon junction; double underline, stop codon.

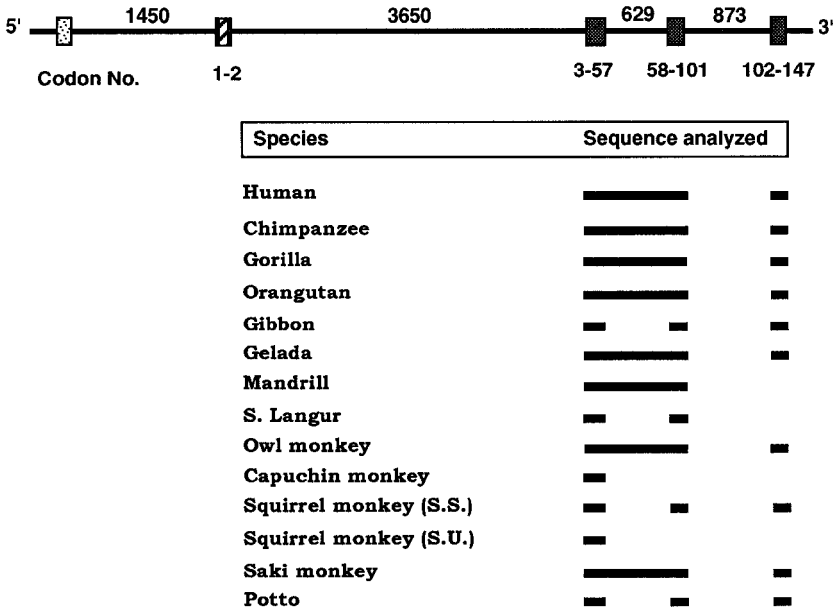


Fig. 2. Primate *COX4* sequences obtained in this study. The sequences generated by PCR, utilizing the primers shown in Fig. 1, are indicated with respect to a diagram of the human *COX4* gene (Fig. 1, legend). Intron sizes are shown in bp.

zyme complexes (Chomyn and Attardi 1987; Kadenbach et al. 1987) that are found in mitochondria and have subunits of mixed genetic origin: At least one subunit in each complex is encoded by the organelle genome, whereas the others are encoded by nuclear DNA. Thus, these multi-subunit complexes offer opportunities to investigate the effects of their mixed genetic origin on their molecular evolution.

Early studies showed that the mitochondrial genome in primates has accumulated mutations at a higher rate than single-copy nuclear DNA (Brown et al. 1982).

However, the rate of sequence evolution in the nuclear genome, which varies in different phylogenetic groups (Wu and Li 1985), decreases in the primate lineage leading to humans (Goodman 1985; Britten 1986; Li and Tanimura 1987).

In contrast to the slow rate of nonsynonymous substitutions (N) in most nuclear genes, sequence comparisons of rat (Gopalan et al. 1989), mouse (Grossman and Akamatsu 1990), cow (Lomax et al. 1984; Bachman 1995), and human (Zeviani et al. 1987) COX IV cDNAs indicated that a large number of N occurred in the pri-

		10										20									
		*										*									
HUMAN	↓	NAA	AGT	GTT	GTG	AAG	AGC	GAA	GAC	TTT	TCG	CTC	CCA	GCT	TAT	ATG	GAT	CGG	CGT	GAC	CAC
CHIMPANZEE	
GORILLA		T..
ORANGUTAN		T..	G..
GIBBON	C	G..T
GELADA		A..	G..	T.T
MANDRILL		A..	G..	T.T
S. LANGUR	GC	G..	T.T
macaq-1		G..CC	G..	T.T
macaq-2		G..CC	G..	T.T
macaq-3		G..	-.	..AT	...	AA.	G..	G..	A..	..A	T..	T.T
CAPUCHIN		.G.TA	G..	AG.	...	G..	T.T
SQ.MON(S.U)		.G.A	G..	..G	...	AG.	...	G..	T.T
SQ.MON(S.S)		.G.A	G..A	..G	...	AG.	..C	G..	T.T
OWL MONKEY		.G.A	G..	AG.	...	G..	T.T
SAKI MONKEY		.G.A	A..A	AG.	..C	G..	T.T
human		G..	-.TA	G..	G..A	T.T
chimp		G..	-.TA	G..	G..A	T.T
orang		G..	-.TAG	G..	G..A	T.T
sqmon-1		A..	CA.C	A.T	C..	..A	G..	G..	...	T..	..A	T.T
sqmon-2		GG.A	..T	G..	..C	T..	T..	T.T
sqmon-3		GG.A	..T	G..	..C	T..	T..	T.T
POTTO		???	???	???	???	???	???	???	??.	..A	G..TG	AG.	..C	G..T	T..
RAT		GGGTA	G..TG	T..	...	G..T	T..
MOUSE		GGGTA	G..T	T..	..C	A..	..C	GCT	T..
COW		GG.ATT	..A	G..TG	AG.	...	G..	..C	T..
bovine		GG.ATT	..A	G..TG	AG.	...	G..	..C	T..

		30										40									
		*										*									
HUMAN		CCC	TTG	CCG	GAG	GTG	GCC	CAT	GTC	AAG	CAC	CTG	TCT	GCC	AGC	CAG	AAG	GCA	CTG	AAG	GAG
CHIMPANZEE	C	T..
GORILLA	C	T..
ORANGUTAN	C	T..
GIBBON	C	T..
GELADA	CC	T..
MANDRILL	A	..CC	T..
S. LANGUR	TCT	T..
macaq-1	TCC	T..
macaq-2	TCC	T..
macaq-3	A	..T	..TTCA	T..	C..	..C	T..
CAPUCHIN	T	..CGAC	T..
SQ.MON(S.U)	T	..CGAC	T..
SQ.MON(S.S)	T	..CGAC	T..
OWL MONKEY	C	..CGGC	T..
SAKI MONKEY	T	..TGGC	T..
human		...	G..A	..C	..TAA	..A	..C	G..A
chimp		...	G..A	..C	..TAA	..A	..C	T..A
orang	A	..C	..TAAC	T..
sqmon-1	C	A.TCA	..A	..CAC	T..
sqmon-2	T	..T	-.ATGC	T..	...	---	---	---
sqmon-3	T	..T	-.ATGC	T..	...	---	---	---
POTTO	T	..TCT	T..	..A
RAT	T	..TCTGC
MOUSE	T	..TC	ATGG
COW	C	..CAC	T..
bovine	C	..CCAC	T..C

Fig. 3. Aligned *COX4* coding region (exons 3–5) sequences from primates, rodents, and cow. Identity to the human sequence is indicated by dots. Codon numbers start with the mature protein's amino terminus. Exon/exon junctions are indicated by arrows. Uppercase names signify functional genes; lowercase names, pseudogenes. Gaps are in-

dicated by *hyphens* (-); gaps placed in pseudogenes are needed to accommodate deletions in these pseudogenes, whereas gaps placed in all functional genes are needed to accommodate insertions in the pseudogenes. Blank spaces designate unsequenced positions; ?, uncertain nucleotides at position of PCR primers.

mate lineage to the human functional gene, with most of these changes occurring after the *COX4P1* anthropoid pseudogene locus (Lomax et al. 1990) arose approximately 41 million years ago (MYA) (Lomax et al. 1992). To further define when the change from a slow to rapid N rate occurred, we have now extended our analysis for functional *COX4* sequences to a total of 13 primates

from the infraorders Catarrhini and Platyrrhini of the suborder Anthropoidea and Lorisoformes of the suborder Strepsirhini. The catarrhines examined consist of five hominoid species (human, chimpanzee, gorilla, orangutan, and gibbon) and three cercopithecoid or Old World monkey species of subfamilies Cercopithecinae (gelada and mandrill) and Colobinae (silvery langur). The plat-

	50														60					
	*														*					
HUMAN	AAG	GAG	AAG	GCC	TCC	TGG	AGC	AGC	CTC	TCC	ATG	GAT	GAG	AAA	GTC	GAG	TTG	TAT	CGC	ATT
CHIMPANZEE
GORILLA
ORANGUTAN
GIBBON	G..
GELADAG
MANDRILLG
S. LANGURG
macaq-1G
macaq-2G
macaq-3	G..GC.C	...
CAPUCHIN
SQ.MON(S.U)
SQ.MON(S.S)
OWL MONKEYT	...
SAKI MONKEYT	...
humanA.	.G.C.G.A.	...
chimpA.	.G.C.G.A.	...
orangA.	.GTC.G.
sqmon-1A	CA.	T.AA.	...
sqmon-2T.AG.	A.A	C.A	.T.
sqmon-3T.AG.	A.A	C.A	.T.
POTTOTA.C
RAT	GA.TG.	C.ACCC
MOUSE	GA.TG.T	C.CCC
COWTTT	.A	C.	..CCC
bovineTC	.G.T	.A	C.	..C	..C	GA.	..C	..C

	70														80					
	*														*					
HUMAN	AAG	TTC	AAG	GAG	AGC	TTT	GCT	GAG	ATG	AAC	AGG	GGC	TCG	AAC	GAG	TGG	AAG	ACG	GTT	GTG
CHIMPANZEE
GORILLA
ORANGUTAN
GIBBONC
GELADA	C.T
MANDRILL	C.T
S. LANGURC.T
macaq-1	C.GT
macaq-2	C.T
macaq-3G.T
CAPUCHINT
SQ.MON(S.U)
SQ.MON(S.S)
OWL MONKEY	C.C	..T
SAKI MONKEY	C.C	..T
human	C.	GT.
chimp	C.	GT.T
orang	C.	G.A
sqmon-1	C.	C.	..C	A.	A.	..TG.	..A	A.	..-
sqmon-2	CC.-	AT.
sqmon-3	CC.-	AT.
POTTO	C.	A.C	...
RAT	C.	..T	..CCA.	...	A.C	..TG
MOUSE	C.	..T	..CC	..C	A.C	..T	..AA
COWC	..C	A.	A.A	..TG
bovineC	CG.	...	A.	A.A	..-	AG.	...	C.	..A	..G

Fig. 3. Continued.

Yrrhines examined consist of five ceboid or New World monkey species (capuchin monkey, two species of squirrel monkey, owl monkey, and saki monkey). The loriform consists of one species (potto). Our results show that rates of N accelerated initially in both catarrhine and platyrrhine lineages and then decelerated in these lineages. The acceleration was most pronounced in the catarrhine lineage to the earlier hominoids. By contrast, the rates of synonymous substitutions (S) were slowest in

hominoids and fastest in New World monkeys. Pseudogene and intron nucleotide substitution (η) rates were also slowest in hominoids.

Materials and Methods

DNA Samples. Human genomic DNA was extracted from the human cell line 293. Capuchin monkey (*Cebus apella*), squirrel monkey

	90												100											
	*												*											
HUMAN	GGC	GGT	GCC	ATG	TTC	TTC	ATC	GGT	TTC	ACC	GCG	CTC	---	---	GTT	ATC	ATG	TGG	CAG	AAG				
CHIMPANZEE	---	---				
GORILLA	---	---				
ORANGUTANC	---	---	A..				
GIBBONC	---	---	A..C				
GELADA	...	AC.	T..	..C	A..	---	---	A..	G..	G..	...				
MANDRILL	...	AC.TC	A..	---	---	G..	...				
S. LANGUR	...	AC.CC	---	---	A..				
macaq-1	...	AC.CG	---	---	---	G..	...				
macaq-2	...	AC.CG	---	---	---	G..	...				
macaq-3	..T.	A..C.	C..	...	T..C	---	---	---	A..				
CAPUCHIN																								
SQ.MON(S.U)																								
SQ.MON(S.S)	..T	..CGCA	...	A..T	---	---	C..	G..	..C	CT.	G..	...					
OWL MONKEY	..T	..C.CA	..A	A..T	---	---	C..C	..T.	G..	...					
SAKI MONKEY	..T	..C.CA	..A	A..T	---	---	C..C	..T.	G..	...					
humanC.	C..T	..CG	...	T..	---	---	A..C	...	G..	...				
chimpC.	C..T	..CG	...	T..	---	---	A..C	...	G..	...				
orangC.A	C..T	..CTG	...	T..	---	---	A..C	..A	G..	...				
sqmon-1C.CT	..C	...	GGG	..A	T..	---	---	A..C	...	G..	...					
sqmon-2	..T	..C.CTG	..A	...	ATT	ATC	A..C	...	G..	...					
sqmon-3	..T	..C.CTG	..A	...	ATT	ATC	A..C	...	G..	...					
POTTO	...	ACC	...	C..T	..CA	---	---	A..	G..	..C	...	G..	...					
RAT	...	CTGCTT	---	---	..G	C..G	..C	...	G..	...					
MOUSE	...	ATGT	..CT	---	---	...	C..G	..T	...	G..	...					
COWCGC	---	---	C..C	C..	..C	...	G..	...					
bovineCGC.C	---	---	C..C	---	..CC	...	G..	...					

	110												120											
	*												*											
HUMAN	CAC	TAT	GTG	TAC	GGC	CCC	CTC	CCG	CAA	AGC	TTT	GAC	AAA	GAG	TGG	GTG	GCC	AAG	CAG	ACC				
CHIMPANZEE				
GORILLA				
ORANGUTAN	..G.	C..				
GIBBONC.				
GELADA	..G.C.T.				
MANDRILL	..T.				
S. LANGUR	..G.				
macaq-1	..T.C.T.				
macaq-2	..G.	..G.	T..	A..	..C	..T.	T..T.				
macaq-3	AG.T.	G..	..C	..CTA				
CAPUCHIN																								
SQ.MON(S.U)																								
SQ.MON(S.S)	..G.C	..CT.				
OWL MONKEY	..G.C	..CT.				
SAKI MONKEY	..GTC	..CT.				
human	..G.	..G.	A..	..C	..C	C..	..T.				
chimp	..G.	..G.	A..	..C	..C	C..	..T.				
orang	GG.	..G.	A..	..T	..C	..CT.				
sqmon-1	...	CGA	...	G..T	A..	..TT	..C	..CT.				
sqmon-2	..T.T	...	---	---	---	..CT.				
sqmon-3	..T.T	...	---	---	---	..CT.				
POTTO	..G.	..C	A..	..A	..C	..CT	CCT	..C				
RAT	AG.	..CT	A..	..T	..T	..CT	CGT	..CT.				
MOUSE	AG.T	A..	..T	..T	..CT	..C	..T	CGT	..CT.				
COW	A..	..C	..CA	G..G				
bovine	..G.	..G.	A..	..T	..C	..CA	G..G				

Fig. 3. Continued.

(*Saimiri ustus* and *S. sciureus*), saki monkey (*Pithecia pithecia*), and owl monkey (*Aotus azarac*) genomic DNAs were generously provided by Dr. Horacio Schneider, Federal University of Para, Balem, Brazil. Genomic DNAs of chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), gibbon (*Hylobates agilis*), mandrill (*Papio sphinx*), gelada (*Theropithecus gelada*), and silvery langur (*Presbytis cristata*) were generously donated by Dr. Kathy Neiswanger, University of Pittsburgh. Potto (*Perodicticus potto*) DNA was provided by Dr. Calvin Porter using tissue from the Duke University Primate Center.

Polymerase Chain Reaction. PCR primers (Fig. 1) were based on the human *COX4* sequence. Intron-based primers used to amplify exons

terminated with the conserved nucleotides at the intron-exon junction. For exon 5, the 3' primer is located at the 3' untranslated region, with the last nucleotide representing the third base of the stop codon. Since introns are divergent and some species have identical coding sequences, primers flanking an entire intron or part were also used to generate PCR products containing both exon and intron sequences (Fig. 2). All primers were synthesized by Dr. Michael Hagen, Wayne State University, using an Applied Biosystems DNA synthesizer. PCR reactions were carried out in 50 μ l containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM MgCl₂, 10% DMSO, 0.2 mM each dNTP, 0.35 μ M each primer, 500 ng of DNA template, and 1.25 units of DNA polymerase (Perkin-Elmer). Reactions were performed in a Perkin-Elmer/

	130												140							
	*																			
HUMAN	AAG	AGG	ATG	CTG	GAC	ATG	AAG	GTG	AAC	CCC	ATC	CAG	GGC	TTA	GCC-	TCC	AAG	TGG	GAC	TAC
CHIMPANZEE
GORILLA
ORANGUTAN
GIBBON
GELADA
MANDRILL
S. LANGUR
macaq-1
macaq-2	G..	G.G	..AT	...	G..	..AT
macaq-3	C..
CAPUCHIN
SQ. MON (S.U)
SQ. MON (S.S)	C..	...	A..	G.C	...	C.C
OWL MONKEY	C..T	G..	G.T	...	C.C
SAKI MONKEY	C..	G.T	G.T	...	C.CT	..T	..T
humanG	..-C	..A	G..T	..T
chimpG	..-C	..A	G..T	..T
orangAG	..-	C..-	..C	..A	G..	..AAT
sqmon-1G	GGGC	A..-	A..	G.T	..T
sqmon-2AA	..-T	..C	A..-	G..T	..T
sqmon-3AA	..-T	..C	A..-	G..T	..T
POTTOATCC	...	G..
RAT	...	C..CTC	T..-	G..
MOUSE	...	C..ACCTC	T..-	G..T	..T
COWC	GC.C	T..G-	G..
bovine	..G.C	...	G..	...	GC.C	T..A-	G..	...	C..

HUMAN	GAA	AAG	AAC	GAG	TGG	AAG	AAG	TGA
CHIMPANZEE
GORILLA
ORANGUTANT
GIBBON
GELADA
MANDRILL
S. LANGUR
macaq-1
macaq-2	..C
macaq-3
CAPUCHIN
SQ. MON (S.U)
SQ. MON (S.S)	..GG
OWL MONKEY	..CG
SAKI MONKEYG
human	..C
chimp	..C
orang	..C
sqmon-1	..C
sqmon-2	..CT
sqmon-3	..CT
POTTO	..G	...	C..A
RAT	A..CT	..A
MOUSE	..CT
COW	..CA
bovine	..CA

Fig. 3. Continued.

Cetus thermal cycler 480, utilizing a modified touch-down protocol (Roux 1994) with the following conditions: 94°C, 40 s; 60°C, 45 s; 72°C, 1 min, 5 cycles; 94°C, 40 s; 55°C, 45 s; 72°C, 1 min, 5 cycles; 94°C, 40 s; 50°C, 45 s; 72°C, 1 min, 5 cycles; 94°C, 40 s; 45°C, 45 s; 72°C, 1 min, 10 cycles; 94°C, 40 s; 40°C, 45 s; 72°C, 1 min, 10 cycles.

Purification, Cloning, and Sequencing of PCR Products. PCR products were isolated using agarose gel electrophoresis (1.2%) and purified with either Ultra-free MC purification units (Millipore) or QiaQuick DNA purification columns (QIAGEN). Purified products were either cloned into pCR-Script vectors at the *SrfI* site, using the

pCR-Script SK(+) cloning kit (Stratagene), or cloned into pGEM-T vectors utilizing a pGEM-T cloning kit (Promega) and transformed to XL1-Blue (Stratagene) and DH5 α (GIBCO BRL) competent cells, respectively. Sequencing reactions were performed using the [γ ³²P] ATP (NEN Dupont) end-labeling method (Maxam and Gilbert 1977). Each of 20 cycles with CircumVent Thermal Cycle Dideoxy DNA Sequencing kit (New England BioLabs) consisted of 95°C, 20 s; 55°C, 20 s; 72°C, 20 s. For each exon, more than three independent clones, representing at least two PCR reactions, were sequenced on both strands. All manipulations followed manufacturers' protocols.

Human, orangutan, pigtail macaque (*Macaca nemestrina*), and squirrel monkey *COX4* pseudogenes were previously isolated by PCR

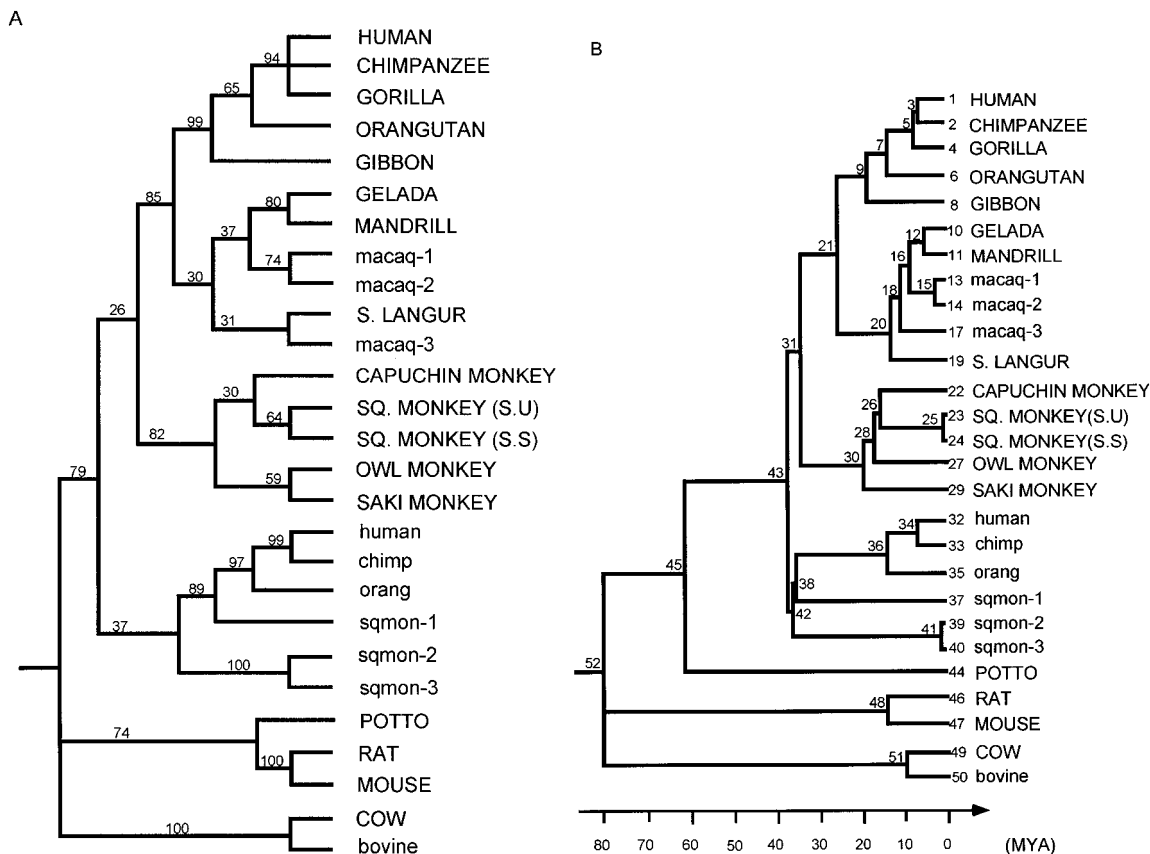


Fig. 4. Maximum parsimony analysis of the aligned *COX4* exonic sequences shown in Fig. 3. **A** One of 30 most parsimonious trees, each requiring the lowest number of η (406). Although this tree does not consistently depict orthologous relations among the functional *COX4* sequences, it has fewer inconsistencies than any of the other 29 trees. *Numbers above branches* represent bootstrap values (as percentages) obtained in 500 replications. **B** The near-most parsimonious tree (410 steps) that is fully consistent with strongly supported evidence on primate phylogeny (Bailey et al. 1992; Goodman et al. 1994; Harada et al. 1995; Schneider et al. 1996). *Numbers at each branch* represent branch-points (interior nodes) and the terminal exterior nodes of the tree. Branches were drawn to the time scale shown below the tree. This time scale is consistent with paleontological evidence on the ages of branch-points during early eutherian phylogeny (Romer 1966; Novacek 1992) and during successive stages of primate phylogeny (Gingerich 1984; Fleagle 1988, 1990; Kay 1990; MacFadden 1990; Rosenberger et al. 1990; Delson 1992; Gingerich and Uhen 1994). The ages for branch-points within Hominoidea and Ceboidea (Bailey et al. 1992;

and sequenced (Lomax et al. 1992). A chimpanzee *COX4* pseudogene was previously isolated and sequenced from a genomic library (Lomax et al. 1990). Like gelada and mandrill, pigtail macaque is a member of the subfamily Cercopithecinae.

Sequence Data Analysis. Nucleotide sequences were unambiguously aligned by eye. For the dataset restricted to only functional *COX4* coding sequences, no gaps were required. However, in the full dataset of functional and pseudogene *COX4* exonic sequences (Fig. 3), some gaps were required to accommodate the insertions and deletions that had occurred in the evolution of the pseudogene sequences. The gaps used markedly increased sequence identity between pseudogene and functional gene sequences. Similarly, the gaps used in the dataset of

Schneider et al. 1993; Goodman 1996) are based on use of a local molecular clock approach to date branch-points. This approach focuses on localized regions of the phylogenetic tree and for each region uses one or more well-established paleontological time points at or near the root of that region as reference dates for the clock calculations. Dated branch-points with their *ages in parentheses (MYA)* are as follows: 52 (80), 45 (63) 31 (35), 21 (25), 9 (18), 7 (14), 5 (7), 3 (6), 30 (20), 28 (17) 26 (16), 38 (35), 36 (14), 34 (6). An increase of the N rate is evident on the anthropoid stem using the parsimony ancestral reconstructions; thus, the common ancestor for the anthropoid *COX4* functional genes and *COX4P1* pseudogenes (i.e., node 43) was used to demarcate the time when the switch occurred from a slow to rapid N rate. The procedure that we utilized to calculate the age of this early anthropoid node was based on the assumption that after the *COX4P1* pseudogene lineage arose from node 43 the rapid N rate on the *COX4* functional gene lineage was equivalent to the ensuing N rate in the platyrrhine and catarrhine stems (i.e., on the 31–30 and 31–21 branches). With this assumption, we calculate a time of 40 MYA for node 43.

aligned *COX4* intron 3 sequences markedly increased overall sequence identity.

A maximum parsimony analysis, implemented with PAUP (Swofford 1990) on the aligned nucleotide sequences, found the trees that accounted for the descent of these sequences from their common ancestor by the smallest number of η . By providing bootstrap values for the nodes in these trees, this analysis also indicated how much confidence to have in each of the nodes. Because some of the weakly supported nodes for the dataset of all exonic (functional and pseudogene) sequences were inconsistent with strongly supported evidence on primate phylogeny (Bailey et al. 1992; Goodman et al. 1994; Harada et al. 1995; Schneider et al. 1996), the lowest η tree length that was most consistent with the evidence on primate phylogeny was modified to be completely consistent. This completely consistent phylogeny tree had

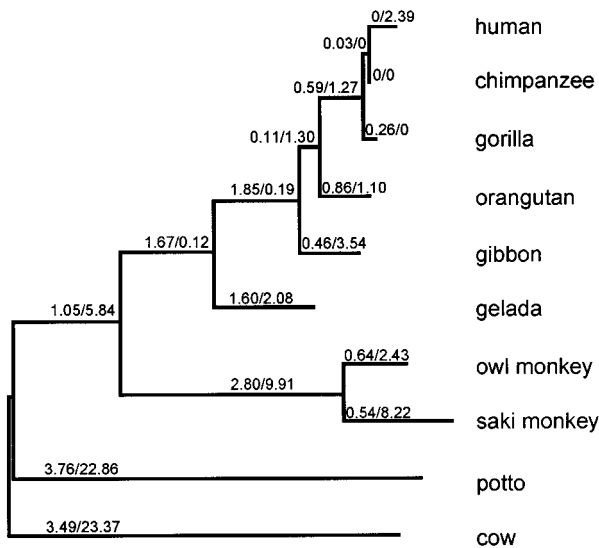


Fig. 5. Phylogenetic tree of complete coding region sequences of *COX4* functional genes. The tree was generated by FITCH, using the distance matrix for all η , and branches are drawn to reflect their lengths. Numbers above each link (N/S) designate substitutions per 100 sites.

an η length only slightly larger than the lowest η length (see Fig. 4, legend).

The maximum parsimony ancestral sequences at the interior nodes of the phylogenetic tree for each dataset of aligned sequences were reconstructed by the A-solution procedure of Goodman et al. (1974). When there are alternative maximum parsimony choices for ancestral nucleotides at an aligned sequence position, the A-solution choice places each required nucleotide substitution on a lineage to fewer contemporary species than would any alternative choice. The A-solution more than any other maximum parsimony solution counteracts the tendency for nucleotide change to be underestimated in the regions of the phylogenetic tree sparsely represented by contemporary sequence data and, in general, insures that the best use is made of the contemporary sequence data in reconstructing ancestral sequences.

Three types of distance matrices were generated with MEGA (Kumar et al. 1993) using a correction (Jukes and Cantor 1969) for (1) all η , (2) only N, and (3) only S. This was done for the complete set of exonic sequences. In order to have accurate pairwise comparisons between pseudogenes and functional gene sequences, the alignment excluded from the pseudogenes their mutation-induced stop codons and their codons with partial deletions. Internodal distances were calculated using two different approaches. For the parsimony A-solution, the reconstructed ancestral sequences at the interior nodes were included with the contemporary sequences to generate N and S distance matrices. Thus, included in the pairwise comparisons were those that yielded the internodal distances—that is, the distances on the links connecting ancestral to descendant nodes. For the distance approach, the internodal distances were calculated by the algorithm of Fitch and Margoliash (1967), as implemented by FITCH in PHYLIP (Felsenstein 1991) with the user-defined tree option. When this option is employed with the given phylogenetic tree, FITCH apportions the pairwise distances among the contemporary sequences to the links of the tree, thereby providing the internodal distances. Another option of FITCH allows the program to find a tree that yields the closest fit between the original input matrix of pairwise distances and the output matrix on tallying the apportioned internodal distances for each pair of contemporary sequences. When pseudogenes, incomplete sequences, and the rodent genes were removed from the set of contemporary sequences, the tree found by FITCH (Fig. 5) using a distance matrix of all η agreed with

the accepted primate phylogeny. The internodal N and S distances were calculated using the same tree. Substitution rates were derived based on dates in units of MYA for interior nodes as inferred from paleontological evidence on the ages of the common ancestors of the contemporary species (Fig. 4B, legend).

Results

The COX4 Exon Sequences Examined

The human *COX4* gene is a 7.4-kb region that consists of five exons and four introns (Bachman et al. 1997). It encodes a 169-aa (amino acid) polypeptide, including a 22-residue amino-terminal leader. We limited our analysis to the mature protein since different evolutionary pressures are likely to apply to the presequence. The first two residues of the mature protein are encoded by exon 2 and the remainder by exons 3, 4, and 5. Prior to the availability of the human *COX4* gene sequence, we utilized a two-step approach for obtaining coding sequences by PCR. First we amplified introns with primers based on the human cDNA sequence (Zeviani et al. 1987) and on the positions of intron–exon boundaries found in rodent *COX4* genes (Yamada et al. 1990; Carter and Avadhani 1991). Next, we sequenced portions of the amplified introns. Then primers based on the intron sequences, or combinations of adjoining introns and exons, were used to amplify the desired fragments. After sequences of *COX4* functional genes in several species of catarrhine primates became available, primers were designed based on the most conserved regions of the available aligned sequences and were utilized to amplify the functional genes of phylogenetically more distant primate species. Figure 1 presents the sequences and locations of these primers. Figure 2 depicts the stretches of nucleotide sequences determined for the primate *COX4* functional genes. These primate *COX4* sequences, with the rodent and cow sequences, and the primate and bovine pseudogene sequences, were aligned against the human sequence (Fig. 3).

Phylogenetic Branching Pattern of COX4 Exon Sequences

Maximum parsimony analysis of the aligned *COX4* gene and pseudogene exon nucleotide sequences (Fig. 3) generated 30 trees, each requiring 406 η . These trees varied in the order by which the functional platyrrhine *COX4* sequences grouped together and in the position of the macaque and squirrel monkey pseudogenes. We selected the tree (Fig. 4A) closest to the phylogenetic tree that is strongly supported by previous molecular evidence on primate phylogeny (Bailey et al. 1992; Goodman et al. 1994; Harada et al. 1995; Schneider et al. 1996). Three small changes in the branching pattern of the tree, re-

Table 1. Evolution of *COX4* gene: rates and ratios^a

(A) <i>COX4</i> functional gene coding sequences					
Phylogenetic lineages	Accumulated time (MY)	Distance approach		Parsimony approach	
		Rate (N/S)	Ratio	Rate (N/S)	Ratio
52-Cow (49)	80	0.33/2.49	0.13	0.49/2.57	0.19
52-Rodents (46,47)	80 ⁺	0.84/3.97	0.21	0.67/4.21	0.16
52-Potto (44)	80	0.27/2.41	0.11	0.61/3.59	0.17
52-43	40	0.02/2.80	0.005	0.15/2.04	0.07
43-Platyrrhine ancestor (30)	20	0.52/2.67	0.20	*1.93/3.50	0.55
30-Platyrrhines (24,27,29)	57	0.42/3.73	0.11	0.55/3.73	0.15
43-Catarrhine ancestor (21)	15	*1.64/0.53	*3.08	*2.36/0.76	*3.11
21-Gelada (10)	25	0.68/1.26	0.54	1.04/1.11	0.93
43-Hominoid ancestor (9)	22	*2.04/0.40	*5.04	*2.14/1.04	*2.06
9-Gibbon (8)	18	0.31/1.87	0.16	0.32/1.96	0.16
9-Orangutan (6)	18	0.50/1.28	0.39	0.48/1.29	0.37
9-Hominin ancestor (5)	11	0.49/2.59	0.19	1.05/2.12	0.50
5-Hominins (1,2,4)	20	0.15/1.24	0.12	0.15/1.19	0.12

(B) <i>COX4</i> pseudogene sequences					
Phylogenetic lineages	Accumulated time (MY)	Distance approach		Parsimony approach	
		Rate (N/S)	Ratio	Rate (N/S)	Ratio
38-Sqmon-1 (37)	35	2.44/3.60	0.68	2.49/3.25	0.77
38-Orang (35)	35	0.65/1.65	0.39	0.88/1.68	0.52
38-34	29	0.85/1.53	0.55	1.14/1.60	0.71
34-Hominans (32,33)	12	0.50/0.00	—	0.50/0.00	—

(C) <i>COX4</i> intron 3 sequences				
Phylogenetic lineages	Accumulated time (MY)	Distance rate	Parsimony rate	
Between catarrhine ancestor (11) and platyrrhine ancestor (14)	25	2.50	2.46	
11-Cercopithecines (8,9)	25 ⁺	1.62	1.53	
11-Orangutan (6)	25	1.37	1.31	
11-Hominin ancestor (5)	18	1.47	1.81	
5-Hominins (1,2,4)	20	0.84	0.88	
14-Platyrrhines (12,13)	40	1.20	1.25	

^a Phylogenetic lineage is defined as the internodal link(s) from an ancestral node to one or more descendant nodes in the phylogenetic tree shown in Fig. 4B (**A and B**) or to the tree shown in Fig. 8 (**C**). When a descendent node represents a single present-day species or an ancestor of a phylogenetic group, the name of either the species or the phylogenetic group is given, along with the node number. When the descendent nodes represent a group of present-day species, the name of the group is given, along with the node numbers for these species. Accumulated time (MY) on the internodal link(s) representing a phylogenetic lineage is determined from the times assigned to the relevant ancestral and descendant nodes in Figs. 4B and 8. The rate of a phy-

logenetic lineage is calculated from the accumulated distance (not shown) and accumulated time in units of 10^{-9} CSY. ⁺As an exception to giving total accumulated time, the time given is that of the age of the ancestral node; thus, the average of the accumulated distances to the present-day descendants is used in calculating these particular rates. Significance of an accelerated N rate was evaluated with a one-tailed t-test; * $P < 0.005$. Significance of deviation of ratios >1 from 1 itself was tested by the K-means cluster analysis module of STATISTICA (Statsoft, Tulsa, OK). N, nonsynonymous substitutions; S, synonymous substitutions.

quiring only four additional η , made the modified tree (Fig. 4B) completely consistent with the evidence on primate phylogeny. These changes involved (1) moving the loriform primate potto from the stem of the rodents to the stem of the primates, which added 2 η ; (2) moving the owl monkey from the saki monkey to the stem of the branch of capuchin and squirrel monkeys, which added 1 η ; and (3) moving the macaque pseudogene macaq-3

from the colobine langur to the stem of the cercopithecine sequences, which added another 1 η .

Evolution of COX4 Exon Sequences of the Functional Genes

Our analysis of the N and S distances found on the links of the phylogenetic tree (Fig. 4B) for the complete set of

	10	20	30	40	50	60	70	80
ANCESTOR	GSVVKSE Y DA	LPTYVDRRDY	PLP D V A H V KH	<u>LSASQKALKE</u>	<u>KEKASWSLS</u>	<u>MDEKVELYRI</u>	QFKESFAEMN	RGTNEWKTVV
HUMAN	X.....FS	..A.M...H	..E.....	K.....SS
CHIMPANZEE	X.....FS	..A.M...H	..E.....	K.....SS
GORILLA	X.....FS	..A.M...H	..E.....	K.....SS
ORANGUTAN	X.....FS	..SA...H	..E.....	K.....SS
GIBBON	X.....FS	..A...H	..E.....	A.....	K.....SS
GELADA	X.....FT	..A.....	K.....RSS
MANDRILL	X.....FT	..A.....	K.....RSS
S. LANGUR	X...R..FS	..A.....	K...T...SS
CAPUCHIN	X.....	..S.....	..R.....
SQ.MON (S.U)	X.....	..R.S.....	..R.....
SQ.MON (S.S)	X.....	..R.S.....	..R.....	X.....
OWL MONKEY	X.....	..S.....	..R.....S
SAKI MONKEY	X.....T	..QS.....	..R.....S
POTTO	XXXXXXXX	..S.....	N.....
RATS.....L	D.....	R...Q...NK
MOUSEF..A.....	..TM.....	D.....	R...Q...N
COWS.....	..N.....	I...L...KS
	90	100	110	120	130	140		
ANCESTOR	<u>GAAMFFIGFT</u>	<u>ALLIWEKRY</u>	VYGP I PH T FD	HEWVAM O TKR	MLDMKVNP I Q	GFSAKWDYEK	NEWKK*	
HUMAN	.G.....	..VIM.Q.H.	...L.QS..	K...K...LAS.....	
CHIMPANZEE	.G.....	..VIM.Q.H.	...L.QS..	K...K...LAS.....	
GORILLA	.G.....	..VIM.Q.H.	...L.QS..	K...K...LAS.....	
ORANGUTAN	.G.....	..IM.Q..H	...L.QS..	K...K...LAS.....	
GIBBON	.G.....	..IM.Q.H.	...L.Q...	K...K...LAS.....	
GELADA	.T...F.I.	..VM.....	...L.Q...	K.....LAS.....	
MANDRILL	.T.....I.	..VIM...L	
S. LANGUR	.T.T.....	..IM.Q..	
CAPUCHIN	
SQ.MON (S.U)	
SQ.MON (S.S)ILV.L...	...L.....	K.....	..L.M...D	.LAS.....	K.....	
OWL MONKEYILI.L...	...L.....	K.....	..L...VD	.LAS...D.	K.....	
SAKI MONKEYILIML...	...L.....	K.....	..L...VD	.LAS.....	K.....	
POTTO	.T..L....	..V.....	PD...K...KA.....Q	
RAT	.L.....	..V...S.	RD.....N.	
MOUSE	.M.....	..V...S.	RD.....	..A.....D.	
COWL...H.	...E.E...A.....D.	

Fig. 6. Aligned COX IV amino acid sequences. The sequence alignment is numbered from the beginning of the mature protein and starts at residue 3. Only amino acids that differ from the mammalian ancestral sequence are shown. This ancestor sequence was obtained by translating the A-solution maximum parsimony ancestral nucleotide sequence

for node 52 of Fig. 4B into the corresponding amino acid sequence. Arrows indicate positions of introns. In the ancestor sequence, the transmembrane domain is designated by the underline, and the large conserved region is designated by the double underline.

exonic sequences in Fig. 3 revealed a remarkable upsurge of the N rate on the anthropoid lineage that separated into platyrrhines and catarrhines (Table 1A). This upsurge is especially evident on the lineage descending first to the catarrhine ancestor and then to the hominoid ancestor (from node 43 to node 9). As estimated by either the distance or parsimony approach, the N rate on the nonanthropoid lineages was always slower than 1×10^{-9} changes/site/year (CSY) (Table 1A) and at 0.02×10^{-9} CSY (distance approach) and 0.15×10^{-9} CSY (parsimony approach) was very slow on the primate lineage descending from the cow-rodent-primate ancestor to the anthropoid COX4 functional-pseudogene ancestor (from node 52 to node 43). In contrast, by either approach the N rate increased to over 2×10^{-9} CSY on the anthropoid COX4 functional gene lineage descending to the catarrhine (node 21) and then hominoid (node 9) ancestor. From the hominoid ancestor to the present the N rate

slowed down and became very slow (0.15×10^{-9} CSY) in descent from the hominin (gorilla-chimpanzee-human) ancestor (node 5) to the present.

This pattern of an accelerated N rate followed by a decelerated rate suggests that positive selection for adaptive amino acid replacements became purifying selection, preserving replacements that had occurred. Due to purifying selection, N accumulate more slowly than S and the ratio of N to S rates (N/S ratio) in the evolution of functional coding sequences is normally much less than 1, as is evident (Table 1A) for all nonanthropoid COX4 functional gene lineages (those to cow, rodents, and potto), as well as the primate lineage to the anthropoid COX4 functional-pseudogene ancestor and the hominin and later platyrrhine lineages. However, in the anthropoid COX4 functional gene lineage descending to the catarrhine and then hominoid ancestor, the N/S ratio ranging between 2 and 5 was significantly elevated

					50			
					*			
HUMAN	GTGGGTATTG	AAGGGACCCA	CAGGCGCGCC	CAGCAGCTCT	CGGAAGCGTG	TGTGTG	-----	-----
CHIMPANZEEA...T...	-----	-----
GORILLAT..T	-----	-----
ORANGUTAN	T...T...TGTG	TGTGTGTGTG	-----
GELADAG..A..T..	A...T...	-----	-----
MANDRILLG..A..T--A.....	G...T...	-----	-----
OWL MONKEYAA.....T...	-----	-----
SAKI MONKEYA.....C	T..G..T...	-----	-----
			100					
			*					
HUMAN	-----ACAGA	GCCTCTGCCTC	ACTTCTGGGC	CTACTGTCTA	GAGAGCAGTC	TTGCACAGGA	GGTTTGCCTCT	-----
CHIMPANZEE	-----	-----
GORILLA	-----	-----
ORANGUTAN	TGTGTGCG--CG.	-----
GELADA	-----A.....C.....	-----
MANDRILL	-----G..A.....G...C.....	-----
OWL MONKEY	-----CGG.C	--...C...	TG..A...TGC.....	-----
SAKI MONKEY	-----CG...	---TGC.....	-----
		150				200		
		*				*		
HUMAN	GCTGGGTTTC	GGGGTCACIG	TGCCAGGGCC	CCAGTTTATG	<u>TGCTCACCAG</u>	<u>TCACTTAGCT</u>	<u>CTGCCAGCTG</u>	-----
CHIMPANZEEC.....	.T.....	-----
GORILLAC.....	-----
ORANGUTANC.....	-----
GELADA	..C.....C.....T.	-----
MANDRILL	..C.....C.....T.	-----
OWL MONKEY	..C.....	T.....	.C..T.....GC.	-----
SAKI MONKEY	..C.....	T.....	.C.....G.	-----
			250					
			*					
HUMAN	ACAGGATCTT	TTGCTAGGCC	CCCTTCTCTG	TGCTGAGTGG	AGGTAGCCTC	TCAGCATATC	TGCTGGGTAA	-----
CHIMPANZEEC.....	-----
GORILLAA	-----
ORANGUTAN	..A.....TA.....	-----
GELADA	.TG.....	-----
MANDRILL	.TG.....	-----
OWL MONKEY	---T.....G.....	..TT.....	..G.....G	-----
SAKI MONKEY	...T.....G.....	..TT.....G	-----
		300				350		
		*				*		
HUMAN	GACATAGTGA	ACTGTAAAIT	ATTGAAAGAA	ACTCAGCAA	ATGCATAGTG	TTTGGTATGA	AAGGGGCAGA	-----
CHIMPANZEET.....	-----
GORILLAT.....	-----
ORANGUTANT.....C.....A.....T.	-----
GELADAT..TC.....TT	-----
MANDRILLT..TC.....TT	-----
OWL MONKEYT..TG.....T..C..	..C.....	G.....TG	-----
SAKI MONKEYT..TG.....T..CC.	G..A...TG	-----

Fig. 7. Aligned *COX4* intron 3 sequences. Identify to the human sequence indicated by dots. Dashes indicate gaps; underline, short tandem repeat.

above 1, providing clear evidence of strong positive selection for adaptive amino acid replacements. Very similar results were observed in our analysis of the dataset containing only complete functional *COX4* gene coding sequences (Fig. 5).

The N rate also accelerated on the platyrrhine stem and then decelerated in the later platyrrhine lineages (Table 1A, Fig. 5). However, due to high S rates in the platyrrhines the N/S ratio both in the stem and later platyrrhines remained less than 1. Thus, the evidence for adaptive replacements on the platyrrhine stem is not as

striking as the evidence for these replacements on the catarrhine and hominoid stems.

Evolution of *COX IV* Protein

A number of amino acid replacements separate the contemporary sequences from their common ancestor (Fig. 6). None of these replacements alters the hydrophobicity of the *COX IV* protein. Several portions of the sequence alignment have a high amino acid replacement density in

*

HUMAN	AAAATAACAA	GATTAAATAG	ACCCTAATAC	TGTAATTCAA	GTAAGAAATA	ATTTTGCAGT	TTTAAATTGC
CHIMPANZEE
GORILLAC.....
ORANGUTANG.C.
GELADATG.C.C.C.
MANDRILLTG.C.C.C.C.
OWL MONKEY--C.C.C.C.	A.....C.C.
SAKI MONKEYG.C.G.C.C.C.	A.....CG.C.

450

*

HUMAN	ACCTGAAGCG	AACTGTATGC	ATTTTCTTCC	TTCC-TTGCC	CTGTCACATG	CCTGCGTGGG	CACGTGTGTG
CHIMPANZEE
GORILLA
ORANGUTAN	...A...AT...TC...
GELADA	...A...A	...CC..ATT...-C...C.
MANDRILL	...A...A	...CC..TT...-C.....C.....C.....
OWL MONKEY	...A...A	...CC...	...C.....	C..T...-T.CA.C	...A.....
SAKI MONKEY	...A...A	..G.CC...	...C.....	C..T.....	..A.....ACA.C	...A.....

500

*

550

*

HUMAN	---CACGTAT	GTGCGTGAAC	ATGATGTGGC	CTGGGTTGGT	GTATCCTTCA	GCTCTGTGTT	TCCTCCTTCA
CHIMPANZEEC	...A.....
GORILLAC	...A.....
ORANGUTANA.C	...A.....C.....
GELADAA..CC.C.T.
MANDRILLA..CC.	G.....C.....G...G.C.
OWL MONKEY	ACG..AC..C.....	..CA..C.C.	..G.....	C.....G...
SAKI MONKEY	ACA...C..C.....	..CA..C.C.	..GC.....	C.....G...

600

*

HUMAN	CAAGTGTGGT	TTTGGGGAGA	AGTGGTTGAA	TGTTGCAGAG	GAGGGAGCTG	CTGAC-----	-CTTTGTGCC
CHIMPANZEE
GORILLA
ORANGUTANA..
GELADAG.	CA.....
MANDRILL	CA.....
OWL MONKEY	CC.....C.CACT	T.....
SAKI MONKEYC.....	CC.....CCGCT	T.....

650

*

HUMAN	TGTAATGGC	TGTCTCTCT	GCCCCCAG
CHIMPANZEE
GORILLA
ORANGUTAN
GELADAC.....
MANDRILLC.....T..
OWL MONKEYT.
SAKI MONKEYTT.

Fig. 7. Continued.

primates compared to the ancestral sequence, as well as several portions that are absolutely conserved. The portion (11–32) that contains the greatest number of changes in hominoids compared to the other species is sandwiched between two highly conserved portions. On one side are seven residues (positions 4–10; Fig. 6) that, with only one exception, are invariant in all species examined. These residues have been correlated in partial proteolysis experiments (Capitanio et al. 1994) with the ability of COX to carry out proton pumping. On the other side nine of 20 residues (positions 33–52) are invariant in all mammals thus far examined. We can speculate that these two

sets of conserved residues circumscribe a region important for subunit function and that only the central portion can be acted upon successfully by evolutionary selection to modulate function.

Evolution of COX4 Pseudogenes

Nine retropositioned primate pseudogenes were utilized in our analysis. The oldest clade (*COX4PI*), one represented by six of the retropositioned sequences, arose in the stem anthropoid lineage about 40 MYA before this

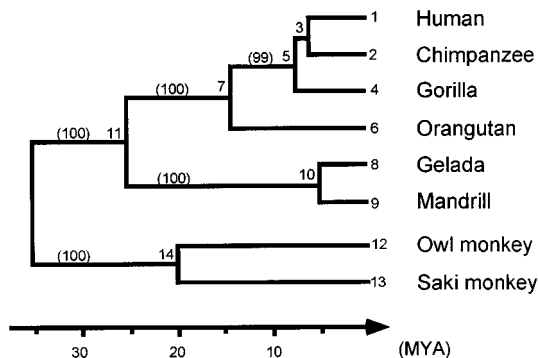


Fig. 8. Phylogenetic tree of *COX4* intron 3 sequences found by maximum parsimony analysis. Branches were drawn to the time scale previously defined (Fig. 4B, legend). Numbers at branches represent branch points (interior nodes) and the terminal exterior node of the tree. Numbers in parentheses represent bootstrap values (as percentages) obtained in 500 replications.

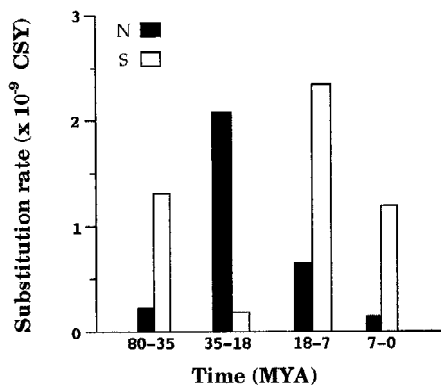


Fig. 9. Rates of N and S in *COX4* during different evolutionary periods in descent from the mammal ancestor to present-day hominins (gorilla, chimpanzee, human), as calculated from the distances of the relevant branches in Fig. 5. Filled bar, N rate; empty bar, S rate. For the terminal descent of hominins from their last common ancestor at 7 MYA, the accumulated time on the three lineages is approximately 20 MY.

lineage split at about 35 MYA into platyrrhines and catarrhines (Fig. 4). From the platyrrhine–catarrhine ancestor (node 38) to the present, this clade’s four orthologously related pseudogene lineages accumulated about two-thirds as many N as S. Both the N and S rates were much higher in the squirrel monkey (i.e., platyrrhine) lineage than in the hominoid lineages; the slowest rate was from the human–chimpanzee ancestor to the present (Table 1B). By calculating N/S on each link of the *COX4P1* clade, we found the pseudogene was functional right after it diverged from the main stem ($N/S < 1$) and eventually lost activity ($N/S \approx 1$). The observation that $N/S \approx 1$ is consistent with the neutral theory that, when an expressed gene loses function, selective pressure on N sites is relaxed and the rates of N and S will converge.

Evolution of *COX4* Noncoding (Intron) Sequences

To compare *COX4* noncoding DNA rates to silent rates of the coding sequences, we amplified and sequenced

COX4 intron 3 in eight of the primate species (Fig. 7). A dinucleotide repeat of (GT) n was found in all eight species, with $n = 15$ in orangutan and $n = 4–6$ in each of the seven primates. A 16-nt insertion was also found in the hominoids, flanked by a 4-nt direct repeat (GCTC) (Fig. 7, underlined). Phylogenetic analysis by the maximum parsimony method generated only one tree (Fig. 8). The branching pattern of this tree is consistent with the previous molecular evidence on primate phylogeny (Bailey et al. 1992; Goodman et al. 1994). Internodal distances on the branches of this *COX4* intron 3 tree were determined by both distance and parsimony approaches, and these distances over known periods of time were converted into rates (Table 1C), which could then be compared to rates of other selectively neutral substitutions (S in the *COX4* functional genes and both S and N in the *COX4P1* pseudogenes). Interestingly, the slowest rates for these three types of selective neutral η were observed in the hominoids. This result is consistent with previous observations of the slowdown in substitution rates in higher primates (Goodman 1985; Wu and Li 1985; Britten 1986; Li and Tanimura 1987).

Discussion

We have extended our previous observation of an elevated rate of N in the lineage to the human *COX4* gene (Lomax et al. 1992) by examining other anthropoid primates. An upsurge of N rates followed later by a downsurge occurred in both catarrhines and platyrrhines. However, during the upsurge, the proportion of substitutions that were N was higher in catarrhines than in platyrrhines. In the catarrhines, as emphasized in Fig. 9, pronounced positive selection for adaptive amino acid replacements was evident by higher N than S rates on the lineage encompassing catarrhine and hominoid stems. The markedly decelerated N rate that occurred in the terminal lineages to the human, chimpanzee, and gorilla *COX4* genes (Fig. 9) further suggests that the replacements were indeed selected for and were then preserved by purifying selection.

The initially high rate of COX IV protein evolution in anthropoid primates engenders the question, “What is the nature of the evolutionary pressure responsible?” However, it is difficult to pose testable hypotheses because a specific role for subunit IV is not yet available. Since COX is the terminal and potentially rate-limiting component of the respiratory chain, it is plausible to consider any major functional change in primate evolution that coincided with the period of rapid change in subunit IV. One such striking change was the development of a larger neocortex, in fact among the most aerobic tissues. Although not yet examined systematically in primates, a second change in COX is the loss in humans of the heart of isoform of subunit VIII (Van Kuilenburg

et al. 1988; Rizzuto et al. 1989). Both of these changes pose the dual challenges of deducing the evolutionary pressure(s) that led to them and the biochemical and physiological consequences that resulted.

Acknowledgments. We thank Dr. Kathy Neiswanger, University of Pittsburgh, and Dr. Horacio Schneider, Federal University of Para, Belem, Brazil, for generously supplying primate DNAs. We also thank Dr. Timothy Schmidt (of this laboratory) for help with the computer programs utilized in data analysis. This work was supported by NIH grant GM 48800.

References

- Bachman NJ (1995) Isolation and characterization of the functional gene encoding bovine cytochrome *c* oxidase subunit IV. *Gene* 162: 313–318
- Bachman NJ, Wu W, Grossman LI, Lomax MI (1997) *COX4*, the gene for cytochrome *c* oxidase subunit IV, and *COX4AL*, a closely linked gene, are controlled by a directional promoter. (submitted for publication)
- Bailey W, Hayasaka K, Skinner C, Kehoe S, Sieu L, Slightom J, Goodman M (1992) Reexamination of the African hominoid trichotomy with additional sequences from the primate beta-globin gene cluster. *Mol Phylogenet Evol* 1:97–135
- Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. *Science* 231:1393–1398
- Brown WM, Prager EM, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol* 18:225–239
- Capaldi RA (1990) Structure and function of cytochrome-*c* oxidase. *Annu Rev. Biochem* 59:569–596
- Capitanio N, Peccarisi R, Capitanio G, Villani G, De Nitto E, Scacco S, Papa S (1994) Role of nuclear-encoded subunits of mitochondrial cytochrome *c* oxidase in proton pumping revealed by limited enzymatic digestion. *Biochemistry* 33:12521–12526
- Carter RS, Avadhani NG (1991) Cloning and characterization of the mouse cytochrome-*c* oxidase subunit-IV gene. *Arch Biochem Biophys* 288:97–106
- Chomyn A, Attardi G (1987) Mitochondrial gene products. *Curr Top Bioenerg* 15:295–329
- Delson E (1992) Evolution of Old World monkeys. In: Jones JS, Martin RD, Pilbeam D, Bunney S (eds) *Cambridge encyclopedia of human evolution*. Cambridge University Press, Cambridge, pp 217–222
- Felsenstein J (1991) PHYLIP (phylogeny inference package). University of Washington, Seattle
- Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. *Science* 155:279–284
- Fleagle JG (1988) *Primate adaptation and evolution*. Academic Press, San Diego
- Fleagle JG (1990) New fossil platyrrhines from the Pinturas formation. *J Hum Evol* 19:61–85
- Gingerich PD (1984) Primate evolution: evidence from the fossil record, comparative morphology, and molecular biology. *Yearbook Phys Anthropol* 27:57–72
- Gingerich PD, Uhen MD (1994) Time of origin of primates. *J Hum Evol* 27:443–445
- Goodman M (1996) Epilogue: a personal account of the origins of a new paradigm. *Mol Phylogenet Evol* 5:269–285
- Goodman M (1985) Rates of molecular evolution: the hominoid slowdown. *Bioessays* 3:9–14
- Goodman M, Moore GW, Barnabas J, Matsuda G (1974) The phylogeny of human globin genes investigated by the maximum parsimony method. *J Mol Evol* 3:1–48
- Goodman M, Czelusniak J, Bailey WJ, Hayasaka K, Stanhope MJ, Slightom JL (1994) Molecular evidence on primate phylogeny from DNA sequences. *Am J Phys Anthropol* 94:3–24
- Gopalan G, Droste M, Kadenbach B (1989) Nucleotide sequence of cDNA encoding subunit IV of cytochrome *c* oxidase from fetal rat liver. *Nucleic Acids Res* 17:4376
- Grossman LI, Akamatsu M (1990) Nucleotide sequence of a mouse cDNA for subunit IV of cytochrome *c* oxidase. *Nucleic Acids Res* 18:6454
- Harada ML, Schneider H, Schneider MPC, Sampaio I, Czelusniak J, Goodman M (1995) DNA evidence on the phylogenetic systematics of New World monkeys: support for the sister-grouping of *Cebus* and *Saimiri* from two unlinked nuclear genes. *Mol Phylogenet Evol* 4:331–349
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro NH (ed) *Mammalian protein metabolism*. Academic Press, New York, pp 21–132
- Kadenbach B, Kuhn-Nentwig L, Buge U (1987) Evolution of a regulatory enzyme: cytochrome *c* oxidase (complex IV). *Curr Top Bioenerg* 15:114–162
- Kay RF (1990) The phyletic relationships of extant and fossil Platyrrhinae (Platyrrhini, Anthropeidea). *J Hum Evol* 19:175–208
- Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolutionary genetics analysis. The Pennsylvania State University, University Park, PA
- Li W-H, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326:93–96
- Lomax MI, Bachman NJ, Nasoff MS, Caruthers MH, Grossman LI (1984) Isolation and characterization of a cDNA clone for bovine cytochrome *c* oxidase subunit IV. *Proc Natl Acad Sci USA* 81: 6295–6299
- Lomax MI, Welch MD, Darras BT, Francke U, Grossman LI (1990) Novel use of a chimpanzee pseudogene for chromosomal mapping of human cytochrome *c* oxidase subunit-IV. *Gene* 86:209–216
- Lomax MI, Hewett-Emmett D, Yank TL, Grossman LI (1992) Rapid evolution of the human gene for cytochrome *c* oxidase subunit IV. *Proc Natl Acad Sci USA* 89:5266–5270
- MacFadden BJ (1990) Chronology Cenozoic primate localities in South America. *J Hum Evol* 19:7–21
- Maxam A, Gilbert W (1977) A new method for sequencing DNA. *Proc Natl Acad Sci USA* 74:560–564
- Novacek MJ (1992) Mammalian phylogeny: shaking the tree. *Nature* 356:121–125
- Rizzuto R, Nakase H, Darras B, Francke U, Fabrizi GM, Mengel T, Walsh F, Kadenbach B, DiMauro S, Schon EA (1989) A gene specifying subunit VIII of human cytochrome *c* oxidase is localized to chromosome 11 and is expressed in both muscle and nonmuscle tissues. *J Biol Chem* 264:10595–10600
- Romer AS (1966) *Vertebrate paleontology*. University of Chicago Press, Chicago
- Rosenberger AL, Setoguchi T, Shigehara N (1990) The fossil record of callitrichine primates. *J Hum Evol* 19:209–236
- Roux KH (1994) Using mismatched primer-template pairs in touchdown PCR. *Biotechniques* 16:812–814
- Schneider H, Schneider MP, Sampaio I, Harada ML, Stanhope M, Czelusniak J, Goodman M (1993) Molecular phylogeny of the New World monkeys (Platyrrhini, Primates). *Mol Phylogenet Evol* 2: 225–242
- Schneider H, Sampaio I, Harada ML, Barroso CML, Schneider MP, Czelusniak J, Goodman M (1996) Molecular phylogeny of the New World monkeys (Platyrrhini, Primates) based on two unlinked nuclear genes: IRBP intron 1 and ϵ -globin sequences. *Am J Phys Anthropol* 10:153–179
- Swofford DL (1990) PAUP: phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign, IL

- Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S (1995) Structures of metal sites of oxidized bovine heart cytochrome *c* oxidase at 2.8 Å. *Science* 269:1069–1074
- Van Kuilenburg ABP, Muijsers AO, Demol H, Dekker HL, Van Beeumen JJ (1988) Human heart cytochrome *c* oxidase subunit VIII. Purification and determination of the complete amino acid sequence. *FEBS Lett* 240:127–132
- Wu C-I, Li W-H (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci USA* 82:1741–1745
- Yamada M, Amuro N, Goto Y, Okazaki T (1990) Structural organization of the rat cytochrome *c* oxidase subunit-IV gene. *J Biol Chem* 265:7687–7692
- Zeviani M, Nakagawa M, Herbert J, Lomax MI, Grossman LI, Sherbany AA, Miranda AF, DiMauro S, Schon EA (1987) Isolation of a cDNA clone encoding subunit IV of human cytochrome *c* oxidase. *Gene* 55:205–217