## The Pattern of Mammalian Evolution and the Relative Rate of Molecular Evolution

## Simon Easteal

Human Genetics Group, John Curtin School of Medical Research, The Australian National University, Canberra ACT 2601, Australia

Manuscript received May 5, 1989 Accepted for publication September 15, 1989

#### ABSTRACT

The rates of nucleotide substitution at four genes in four orders of eutherian mammals are compared in relative rate tests using marsupial orthologs for reference. There is no evidence of systematic variation in evolutionary rate among the orders. The sequences are used to reconstruct the phylogeny of the orders using maximum likelihood, parsimony and compatibility methods. A branching order of rodent then ungulate then primate and lagomorph is overwhelmingly indicated. The nodes of the nucleotide based cladograms are widely separated in relation to the total lengths of the branches. The assumption of a star phylogeny that underlies Kimura's test for molecular evolutionary rate variation is shown to be invalid for eutherian mammals. Excess variance in nucleotide or amino acid differences between mammalian orders, above that predicted by neutral theory is explained better by variation in divergence time than by variation in evolutionary rate.

THE neutral theory prediction that, assuming a constant mutation rate, the mean rate of molecular evolution is stochastically uniform among lineages can be tested by comparing the interlineage variance in rate with that expected assuming a Poisson distribution (KIMURA 1968). Several authors (OHTA and KIMURA 1971; LANGLEY and FITCH 1974; KI-MURA 1983; GILLESPIE 1984, 1986a, b) have reported that the prediction is inconsistent with data for mammalian proteins.

This observation has presented a challenge to neutral theory. It has led, on the one hand, to the development of elaborate theories of molecular evolution episodically driven by natural selection (GILLESPIE 1984, 1986a, b); and on the other hand, to attempts to modify neutral theory to make it consistent with the observation by allowing for variable rates of mutation and variation in degrees of functional constraint among lineages (KIMURA 1987; TAKAHATA 1987).

Early approaches to testing for rate variance involved assumptions about either the divergence times of the compared species (OHTA and KIMURA 1971) or the branching order of the species (LANGLEY and FITCH 1974). More recently KIMURA (1983) devised a test for evolutionary rate variance based on the assumption that the species, proteins or genes being compared diverged from each other within a short period of time relative to the total length of time over which they have been evolving independently of each other, *i.e.*, that assumes a star phylogeny. The orders of eutherian mammals are often assumed to have diverged in this way, and it is for this reason (as well as for reasons of sequence availability) that the test has been applied to mammalian proteins.

The star-phylogeny assumption is critical to the appropriateness of the test. If the assumption is invalid then variance in nucleotide change might be explained by divergence-time variation rather than by evolutionary rate variation. There is in fact no sound empirical basis for assuming a star phylogeny for eutherian mammals. The pattern of early eutherian evolution is poorly documented in the fossil record (CLEMENS et al. 1979; NOVACEK 1982; KIELAN-JAWO-ROWSKA, BROWN and LILLEGRAVEN 1987) and the assumption is more a reflection of this lack of knowledge than of any pattern in the fossil data (Wyss, NOVACEK and MCKENNA 1987). The conclusion, derived from this or the earlier approaches, that the relative rates of mammalian protein evolution are inconsistent with neutral theory must remain tentative until the relative divergence times of mammalian orders can be more accurately determined.

Cladistic analysis of nucleotide sequences can be used to determine the branching order of taxa, but not their relative divergence times unless rate constancy is assumed. However, once a branching order is established, the hypothesis of rate constancy can be evaluated in another way-by the relative rate test (SARICH and WILSON 1967; WU and LI 1985; LI and TANIMURA 1987).

Using this approach to study published globin gene sequences in four mammalian orders (EASTEAL 1988) I demonstrated a branching sequence for these orders of rodents then ungulates then lagomorphs and primates. This branching order is consistent with some

The publication costs of this article were partly defrayed by the payment of page charges. This article must therefore be marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

derived from protein sequence data (CZELUSNIAK et al. 1982; GOODMAN et al. 1982) but not with others (PENNY and HENDY 1985). It is different from that proposed by KOOP and GOODMAN (1988) based on globin gene sequences using marsupial sequences as outgroups.

In my analysis I used maximum parsimony algorithms (HENDY and PENNY 1982; METROPOLIS et al. 1953) with paralogous genes as outgroups. The same branching order was indicated by three sets of genes ( $\alpha$ -globins,  $\beta$ -globins and  $\epsilon$ -globins), by a high proportion of bootstrap samples (FELSENSTEIN 1985) and by maximum likelihood (FELSENSTEIN 1981) analysis (my unpublished data). However as FELSENSTEIN (1978) and PENNY, HENDY and HENDERSON (1987) have pointed out, under some circumstances, particularly unequal rates of change, parsimony methods may result in convergence to an incorrect tree. LAKE's (1987) evolutionary parsimony method appears to overcome this problem; however, LI et al.'s (1987) computer simulations have shown that its performance is substantially worse than maximum parsimony unless rate differences are very great (approximately fourfold or more). Use of evolutionary parsimony would thus only seem appropriate when evolutionary rates are known to be highly variable.

As a check for rate constancy in my previous analysis (EASTEAL 1988) I compared the rate of divergence of orthologous (homologous via speciation) mouse and human genes with paralogs (homologs via gene duplication) from four different mammalian orders. No evidence of rate variation was observed; however, since the paralogs are substantially diverged, the stochastic error in the comparisons was high and moderate rate differences may not have been detected, although, if such differences exist, nonsignificant faster rates for mouse genes would be expected to be consistently observed, which was not the case.

Genomic sequences of marsupial  $\beta$ - and  $\epsilon$ -globin genes (KOOP and GOODMAN 1988) as well as cDNA sequences of marsupial  $\alpha$ -globin (WAINWRIGHT and HOPE 1985) and  $\alpha$ -lactalbumin (C. COLLET, unpublished data) are now available. These provide the most closely related sequences that can be used without question as references for comparison of the evolutionary rates of eutherian genes.

In this paper I compare these genes from four eutherian orders (rodents, ungulates, lagomorphs and primates) with their marsupial orthologs. I demonstrate that the rate of sequence evolution does not vary among the eutherian orders. I then use the same sequences to determine the phylogeny of the orders and to investigate whether the star-phylogeny assumption is valid.

## MATERIALS AND METHODS

Nucleotide sequences were obtained from the EMBL (1988 release 14) data base or directly from the published

literature. The sources of the sequences are in EASTEAL (1988) except for: opossum  $\epsilon^{m-}$  and  $\beta^{m-}$  (KOOP and GOODMAN 1988) and *Dasyurus*  $\alpha$ -globin (WAINWRIGHT and HOPE 1985), human and guinea pig  $\alpha$ -lactalbumin (HALL *et al.* 1982), goat  $\alpha$ -lactalbumin (KUMAGAI *et al.* 1987) and Tammar wallaby  $\alpha$ -lactalbumin (C. COLLET, unpublished data).

Multiway sequence alignment, using an iterative procedure (FENG and DOOLITTLE 1987) from the computer package ALIGN (D. SMITH, The Australian National University) was performed with a gap penalty of 2.5 and gaplength penalty of 0. Estimates of synonymous and nonsynonymous substitution rates in coding regions were made using the method of LI, WU and LUO (1985). In noncoding regions corrections were made for multiple substitutions as described by KIMURA (1983). Formal analysis of relative evolutionary rates was by the method of WU and LI (1985).

Phylogenetic analysis was performed using maximum parsimony, maximum likelihood and maximum compatibility methods. The algorithms used were all from the phylogenetic reconstruction computer package, PHYLIP (J. FELSEN-STEIN, University of Washington). The maximum parsimony algorithms, DNAMETRO, DNAPENNY and DNAPARS use Metropolis-annealing (METROPOLIS et al. 1953), branch and bound (HENDY and PENNY 1982) and the method of FITCH (1971), respectively, to find the most parsimonious trees. Maximum compatibility trees obtained using the DNACOMP algorithm did not differ in any substantial way from the maximum parsimony trees and they are not presented. Maximum likelihood trees (FELSENSTEIN 1981) were obtained using the DNAML algorithm.

Of the four genes for which marsupial sequences were available only the three globin genes had been sequenced in all four eutherian orders. No rabbit (lagomorph)  $\alpha$ -lactalbumin gene was available so that  $\alpha$ -lactalbumin was not included in the phylogenetic analysis.

## RESULTS

Relative rates of nucleotide substitution: The alignment of noncoding sequences from the  $\beta$ -like globin genes is shown in Figure 1. Alignment of the long second exons of these genes was problematic and they were not included in the analysis. For the  $\alpha$ globin and  $\alpha$ -lactalbumin genes only the coding regions were compared. The coding region of the Tammar wallaby  $\alpha$ -lactalbumin gene is three codons shorter than that of the goat and human genes and 39 codons shorter than that of the guinea pig gene at the 3' end. Only the regions of the eutherian genes that aligned within the boundaries of the marsupial gene were compared.

The substitution rates between the eutherian and marsupial genes for synonymous and nonsynonymous sites in coding regions and noncoding regions are shown in Tables 1, 2 and 3, respectively. In all comparisons the mean synonymous rates are similar to the noncoding rates suggesting that the synonymous sites are evolving largely without functional constraint.

For all three kinds of site, evolutionary rates were compared using the relative rate test. The results (Tables 4–6) show that there are only three significant differences between orders. These are the nonsynonymous sites for the combined globin genes and the globin and  $\alpha$ -lactalbumin genes combined between Mammalian molecular clock

		5' NON-CODING									
		10	20	30 4	10	50	60	70	80	90 10	0
OPOSSUM	в	GATAAAAGG-CAGAGC	TAGATTAGTTT	CAGCATCATA-	CTACTTTTG	ACACAGCTCT	GTGTTCAC	AAGTAAACT-	TTC	AAA	-
MU22000	۳ ۲	AATAAAAGG-CAGCCC	TAGCCCTACTO	CTGCATCAAC-	TACTTOTA	ACACTTCTG-	TGTTGAC	AAGCAAGCT-	TAC	AAGACCAACAT	C.
MOUCE	2		ACCATCACTTO	OT COTACAT	TCOTTOTO	ACATACTTC	TOTTOAC		0.00		č
MOUSE	β	CATATAAGG-TGAGGT	AGGAICAGIIG	CI-CLIACAI-	IGCITCIG	ACATAGITG-	IGIIGAC	ICACAAC		AGAAACAGACAT	5
MOUSE	γ2	AATAAAAGGCCACCAC	TTCTAGCAG	CAGTACGTACT	TCGCTTCTG	ACACTCCTG-	TGATCAC	CAGCAACC	TCC	CAGACTTGCCAT	С
MOUSE	BH1	AATAAAAGGACAGGTC	TTCAGCCTC	TTGAACATTC-	TGGCTT-TT	GCACACTTG-	AGATCAT	CTCCAACG	TTC	TAGACCTCACAC	:C
COW	Ŕ	CATAAAAGG-AAGAGC	TGGGCCAGCTG	CTGCTTACAC-	TTGCTTCTG	ACACAACCG-	TGTTCAC	TAGCAAC	TAC	ACAAACAGACAC	C
COAT	<b>~1</b>	AATAAAAGGCCACAGC	ATCCACCAG	CACCACACAC-	TTOCTTOTO	ATGOTTOTG-	TGATCAC	CTGTAAGC	TCC	ACGACTTGACAT	C
COAT	13				TTOCTTOTO	CCCCATTAT	COATCAC	CACTAACC	TCC		č
GUAT	η	AATAAAAGGCCATGCA	GIGAAGCAG	CGGCACAGAC-	TOCTO	GULUATIAT-	GGATCAC	CAGIAAGC	100		
HUMAN	β	CATAAAAGT-CAGGGC	AGAGCCAICIA	ITGCTTACAT-	HIGCHICIG	ACACAACIG-	IGHICAC	IAGCA	ACC	CAAALAGACAL	
HUMAN	3	AATAAAAGGCCAGACA	GAGAGGCAG	CAGCACATAT-	CTGCTTCCG	ACACAGCTG-	CAATCAC	TAGCAAGC	TCT	CAGGCCTGGCAT	С
HUMAN	Gγ	AATAAAAGGAAGCACC	CTTCAGCAG	TTCCACACAC-	TCGCTTCTG	GAACGTCTG-	AGGTTAT	CAATAAGC	TCC	TAGTCCAGACGC	:C
PARRIT	R1	CATAAAAGG-CAGAGC	AGGG-CAGCTG	CTGCTTACAC-	TTGCTTTTG	ACACAACTG-	TGTTTAC	TTGCAATC	000	CAAAACAGACAG	A
DADDIT	04		TTCAACCAC	CACCACAAAC-	CTCCTTCTC	ACACATTTCT	CATCOATCAC		000	CACACCTCACAC	2
RADDII	P4	AATAAAAGGCCAAGCC	TTICAAUCAG				ACACTAT				2
RABBII	β3	AATAAAAGGACGAGCC	TTAGAGCAG	STITCACATAC-	Indenteral	AGACATCIG-	AGACTAT	CAGCAAGCIC	AGCGAGCICC	AGACCAGACAT	C
		INTRON 1									
		110	120	130 14	10 1	150	160	170	180 1	90 20	0
000000	٥	CTAACTACCACCCCC	ATGGAAAAGGG	CC ATCA	AGA		CAGGGAATCA	CAACT-GCAA	TTAGTCGT	CGAATAGTA	A
000350M	p	CTAACTA TOOCAC			CARATTON		AACA_T	ACTOCCCT		TOTECOTONOTA	т
UPUSSUM	ε	GTAAGTA-TGGGAG		CAGIGGIA	GAAATIGA	ACCCAGGAGT	TOO A O TOO TOO		TAAAAAUT		Т
MOUSE	β	GITGGTATCCAGGTTA	CAAGGCAGCT-	CA	CAAGAAGAA	GIIGGGIGCT	IGGAGACAGA	GGICT	GUITICUA	G-CAGACAC	. 1
MOUSE	γ2	GTAAGAA-TTG	T	TTTTTAGTGTA	-CAAAGAG-	CTAGGG-A	AATC		AAGAATTC	IGAGGCTCCCT-	-
MOUSE	βH1	GTAAGGA-ATGGAGGG	AAATTAT-CCT	TATGCATGGCA	-GAAATTT-	CCAGGGTT	TCTA		TAGGGTTT	TGTGGCACACT-	-
COW	6	GTAGGTATCCCACTTA	CAAGGCAGGTT	TAAGGAGAG	TGAAATGCA	CCTGGGCGTG	TGAGGACAGA	GCCGTC-CCT	GAGATTCA	GAGAGCTGC	т
COAT	~1	GTAGAAACTCC		CCCCACCATCC	TGAATATCA	GCCTGGC AAA	TCGG-CCACA	AAAAT	TCTTCAAAA-	-ATCTGAGTTGC	T
GOAT	13	GTAGAAAGTGG		TACCACACCAC	TOTACAAAC		TOTT COACA		TCOTTACCTT	TOTTACATACTC	÷-
GUAT	η	GIAAGCAGIGG	ALALAGG	AGGAGAGGAG	IGIACAAAG		TOTTTCCAGA	AAAGAGGGAC			, I
HUMAN	β	GTTGGTATCAAGGTTA	CAAGACAGGII	IAAGGAGAC	CAATA-GAA	ACIGGGCAIG	IGGAGACAGA	GAAGACICII	GGGIIICI	GATAGGCAC	
HUMAN	3	GTAAGCATTGGTT	CTCAATGCATG	GGAATGAAGGG	TGAATATTA	CCCTAGCAAG	TTGA-TTGG-	GAAAG	TCCTCAAGAT	TTTTTGCATCTC	T
HUMAN	Gγ	GTAGGCT-CTGGTGAC	CAGGACA-AGG	GAGGGAAGGAA	GGACCCTGT	GCCTGGCAAA	AGTC-C	AGGTCGCT	TCTCAGGATT	TGTGGCACCTTC	Т
PARRIT	R1	GTTGGTATCCTTTTTA	CAGCACA-ACT	TAATGAGAC	AGATA-GAA	ACTGGTCTTG	TAGAAACAGA	GTAGTCGCCT	GCTTTTCT	GCCAGGTGC	T
DADDIT	01	CTAACTATTCCTT	CTCAATACTTC	CTACACAAACC	CAAATATCA	CCCTGGTAGA		GAAAT	TOCTCAA-AC	TTTTTGTATCTC	Ť
RADDII	P4	GTAAGTAIIGGTI	CICARTACITO				ACTO C	ACCCCCTT	TOTOACCACT	TOTOMACOTOTO	, ' T
KABBII	β3	GIAGGCC-CIGGGGIC	CAGGACA-AGG	CAGAGGAGGAA	GAAACIGA	GCCIGGCAGG	AUIC-C	Addcccri	ICICAGGACI	IdidAAdcicic	
					3' NO	N-CODING					
		210	220	230 24	3' NO! 10	N-CODING 250	260	270	280	290 3	00
MUSSORO	ß	210	220		3' NOM 40 AAG GCCTT	N-CODING 250 GCCCATCTGG	260 TGGTCTTCA-	270 TTGAGCTG	280 AGTCCACTAT	290 3 GTTCCATATAGT	00 CT
OPOSSUM	β	210 AAACTTGACT-AC		230 24	3' NOM 40 AAG GCCTT CAG ATATA	N-CODING 250 GCCCATCTGG	260 TGGTCTTCA-	270 TTGAGCTG	280 AGTCCACTAT	290 3 GTTCCATATAGT GAACACATGGAG	00 CT
OPOSSUM OPOSSUM	β ε	210 AAACTTGACT-AC GTGTATCTGCAAACTA	220 CTCCTTTCTAT ATTTTTCCCTT	230 24 TTCTTT-CCCA ICTCCTA	3' NOM 40 AAG GCCTT CAG ATATA	N-CODING 250 GCCCATCTGG GCCCCGTCTA	260 TGGTCTTCA- TG	270 TTGAGCTG GT	280 Agtccactat Gggggcctgt	290 3 GTTCCATATAGT GAACACATGGAG	00 CT
OPOSSUM OPOSSUM MOUSE	β ε β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT	230 24 TTCTTT-CCCA CTCCTA GTTTCCCTTTT	3' NOM 40 AAG GCCTT CAG ATATA TAG AC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCCTTTCC	260 TGGTCTTCA- TG TGCTCTTGCC	270 TTGAGCTG T TGT-GAACAA	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-T	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA STCCCCATAGAGA	
OPOSSUM OPOSSUM MOUSE MOUSE	β ε β γ2	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCA	230 24 TTCTTT-CCCA CTCCTA GTTTCCCTTTT ACTTAC-	3' NOP 40 AAG GCCTT CAG ATATA TAG AC CAG GC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCCTTTCC -CCTCTCTCT	260 TGGTCTTCA- TG TGCTCTTGCC AGCT	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-T TCCTGTGT	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCCGCTAT-GC	00 CT GC GC
OPOSSUM OPOSSUM MOUSE MOUSE MOUSE	β ε β γ2 βH1	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCA CTAATGCACTA	230 24 TTCTTT-CCCA CCTCCTA GTTTCCCTTTT ACTTAC-4 ATGTCT-	3' NOP AAG GCCTT CAG ATATA TAG AC CAG GC TAG AC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCCTTTCC -CCTCTCTC -TCCAT-TCC	260 ITGGTCTTCA- ITG ITGCTCTTGCC AGCT	270 TTGAGCTG GAT TGT-GAACAA -GTCCAGCAA ACACTGGCAA	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-T TCCTGTGT TCCCATGT	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC	
OPOSSUM OPOSSUM MOUSE MOUSE MOUSE COW	β ε β γ2 βH1 Β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GGCTTCCTCTGAC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCC CTAATGCACTA -CTTGTGCTGT	230 24 TTCTTT-CCCA CTCCTA GTTTCCCTTTT ACTTAC- TTTGTCT- TTTT-CTCCCCC	3' NOP AAG GCCTT CAG ATATA TAG AC CAG GC TAG AC TAG GC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCCTTTCC -CCTCTCTCT -TCCAT-TCC -TCCCTTTCC	260 ITGGTCTTCA- ITGCTCTTGCC AGCT AGCT	270 TTGAGCTG GAT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA	280 AGTCCACTATI GGGGGCCTGTI TGGTTAAT-TI TCCTGTGTI TCCCCATGTI AGGTTTTT-TI	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT	β ε β γ2 βH1 β ε1	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GGCTTCCTCTGAC GATTTTCCATCTG	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCC CTAATGCACT -CTTGTGCTGT CTATGTTTC	230 24 TTCTTT-CCCA CTCCTA GTTTCCCTTTT CTTAC- TTTGTCT- TTTT-CTCCCCC CATCTCA	3' NOM AAG GCCTTO CAG ATATAG TAG AC CAG GC TAG AC TAG AC TAG AT	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCTCTCTCT -TCCAT-TCC -TCCCTTTCC -TCCTTTCC	260 ITGGTCTTCA- ITGCTCTTGCC AGCT AGT ITGCTC	270 TTGAGCTG GACCA GTCCAGCAA -GTCCAGCAA ACACTGGCAA ACCCATTTTGT	280 AGTCCACTATI GGGGGCCTGTI TGGTTAAT-TI TCCTGTGTI TCCCATGTI AGGTTTTT-TI G	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT	β ε β γ2 βH1 β ε1 Γ	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCCTCGAC GACTTCTTATCTG	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCA CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTTAC- ITTT-CTCCCCC TTTT-CTCCCCCC ATGATCATCCCA	3' NOP 40 AAG GCCTTU CAG ATATAU TAG AC TAG GC TAG AT TAG TT	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCCTTTCC -CCTCTCTCT -TCCAT-TCC -TCCCTTTCC -TCTCT-TTA -GCC	260 TGGTCTTCA- TGCTCTTGCC AGCT AGT CAATTC TGCTC TGTC	270 TTGAGCTG GA-CT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTGT GCCTA	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-T TCCTGTGT TCCCATGT AGGTTTTT-T G CCATGCTGGT	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT	β ε β βH1 β ε1 η	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCCTCTGAC GACTTCTTATCTG GACTTCTTATCTG	220 CTCCTTTCTAT ATTITTCCCTT CCCTGT-CTAT GTCTTTCTCCA CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA	230 24 TTCTTT-CCCA ICTCCTAI ICTCCTTT ICGTCT- TTT-CTCCCCCC CATCTCA ITGATCATCCCA	3' NOM AAG GCCTTM CAG ATATAM TAG AC TAG AC TAG AC TAG GC TAG AT TAG CC	N-CODING 250 GCCCGTCTGG GCCCGTCTA -CCCTTTCC -CCTCTCTCT -TCCAT-TCC -TCCCTTTCC -TCCCTTTCC -TCCCTTTCT -TCCCTTTCT	260 TGGTCTTCA- TGCTCTTGCC AGCT AGT TGC TGC TGC	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTTGT GCCTA TTTCTATTAA	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-T TCCTGTGT TCCCATGT AGGTTTTT-T G CCATGCTGGT AGGTTCCT-T	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTTCCCAAGAGA GTCCGCTAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GCCTATCTGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN	β ε β βH1 β ε1 η β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTAACTG GATTTTCCATCTGAC GACTTCTTATCTG GACTTCTTATCTG	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCC CTAATGCACTA -CTTGTGCCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT	230 24 TTCTTT-CCCA TCTCCTA TGTTTCCCTTTT CCTTAC TTTCCCCCCC TTTT-CCCCCCCCCC TTTCCCCACCCT TTTCCCCACCCT	3' NOM AAG GCCTTM CAG ATATAM TAG AC TAG AC TAG AC TAG GC TAG AT TAG TT TAG GC	N-CODING 250 GCCCGTCTGG GCCCCGTCTA -CCCCTTTCC -CCTCTCTCT -TCCAT-TCC -TCCCTTTCC -TCCCTTTCC -TCCCT-TTA -GCC	260 TGGTCTTCA	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA ACACTGGCAA TTTCCAGGAA ACCATTTTGT GCCTA TTTCTATTAA	280 AGTCCACTATI GGGGGCCTGTI TGGTTAAT-TI TCCTGTGTI TCCCATGTI AGGTTTTT-TI G CCATGCTGGTI AGGTTCCT	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG TGTTCCCTAAGT GCCCTCACC-CC	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN	β ε β β β ε 1 η β ε	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC GATTTTTAACTG GGCTTCCTCGAC GACTTCTTATCTG GACTTCTTATCTG GACTTCTTATCTG AATTTTGTATCTG	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCC CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGT	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTTAC TTTT-CTCCCCA TTTT-CTCCCCA TTGATCATCCCA TTTTCCCACCCT	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG AC TAG TT TAG GC TAG GT	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCTCTCTCT -TCCAT-TCC -TCCCTTTCC -TCCTTTA -GCC -TCGCTTTCT -TCTCT-TCC	260 TGGTCTTCA- TGCTCTTGCC AGCT TGC TGC CAATTC TG TGCTGTCCAA AGTT	270 TTGAGCTG GA-CT TGT-GAACAA -GTCCAGCAA ACACTGGCAA ACCATTTGT ACCATTTTGT GCCTA TTTCTATTAA -TGCAGGTGT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT TCCCATGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG TGTTCCCTAAGT GACCCTGAC-AC	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT HUMAN HUMAN HUMAN	β ε β β H1 β ε 1 η β ε ζ Υ	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCCTCTGAC GACTTCTTATCTG GACTCTCTCGCC AATTTTGTATCTG GACTGTCAAACTG	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCA CTAATGCACTA CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGT TTCTTG-TCAA	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTTAC- TTT-CTCCCCC TTTT-CTCCCCCC TTTTCCCACCCT TTTTCCCACCCT TTTTCCCACCCT CATTTCA ATCTC-A	3' NOI AAG GCCTTI CAG ACTATAI TAG AC TAG GC TAG GC TAG TT TAG GC TAG GT CAG GC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCCTTTCC -TCCAT-TCC -TCCCTTTCC -TCTCT-TTA -TCCCTTTCC -TCCCTTCC -TCCCTTCC -TCCCTTCC -TCCCTTCC	260 TGGTCTTCA- TGCTCTTGCC AGCT CAGT CCAATTC TGCTGTCCAA AGTT CCA	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA ACACTGGCAA ACCATTTTGT GCCTA TTTCTATTAA -TGCAGGTGT GCCCAT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT TCCCATGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAT	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT	β εβ βH1 βε1 η βε β1	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCCTCTGAC GACTTCTTATCTG GACTCTCTCTGCC GACTGTCAAACTG GACT-TCTCCCCC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCA CTAATGCACTA -CTTGTGCTGT CTATGTTC TTCTTGTGACTA TATTGGTGT TTCTTG-TCAA CTGGGCTGT	230 24 TTCTTT-CCCA TCTCCTA TCTTAC- TTT-CTCCCCC TTT-CTCCCCCC TTT-CTCCCCCC TTTCCCCCCC TTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCCACCT TTTCCATTTCA	3' NOM AAG GCCTTM CAG AC CAG GC TAG AC TAG GC TAG GC TAG GC CAG GC CAG GC	N-CODING 250 GCCCGTCTA -CCCTTTCC -CCTCTCTCT -TCCAT-TCC -TCCCTTTCC -TCCCTTTCC -TCCCTTTCC -TCCCTTTCT -TCCCTTCT -TCCCTTCT -TCCT-TCC -TCTTTTCC	260 TGGTCTTCA- TGCTCTTGCC AGCT TGCT TGCTGTCCAA AGTT TGCTGTCCAA AGTT CCTCTGCCAAA	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTGT GCCTA TTCCATATAA -TGCAGGTGT GCCCAT	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T G CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTTCCCAAGAGA GTCCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAT	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT	β ε β β β β β β β β β β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC GATTTTTAACTG GGCTTCCTCTGAC GACTTCTTATCTG GACTTCTTATCTG GACTTCTCTCGCC GACTTGTCAAACTG GACTTCTCTCCCC GACTTTCTATTTG	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCCA CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGT TTCTTG-TCAA CTGGGCTGT CTGTTGTCC	230 24 TTCTTT-CCCA GTTTCCCTTTT GTTTCCCTTTT TTTGTCT- TTTT-CTCCCCCC TTTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCCACCTA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCACCCT TTTCCCCCCCCCC	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG GC TAG TT TAG GC TAG GC CAG GC CAG GA TAG AC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCCTTTCC -TCCAT-TCC -TCCCTTTCC -TCCCTTTCC -TCCCTTTCC -TCCCTTCC -TCCCTTCC -TCCTTCC -TCTCTTCC -TCTCTTCC	260 TGGTCTTCA- TGCTCTTGCC AGCT TGC TGC TGCTGTCCAA AGTTTC TGCTGTCCAA AGTT CTCTCTGCCAAA CCAGT	270 TTGAGCTG -GT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTTGT GCCTA TTTCTATTAA -TGCAGGTGT GCCCAT GCCCAT	280 AGTCCACTAT GGGGGGCCTGT TGCTGTGT TCCCATGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GCCCCTACTC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG TGTTCCCTAAG GACCCTGAC-AC GCTTTCAAGGAT 	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT DARBIT	β ε β β β β ε η β ε γ 1 β ε β 1 4 3	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC GATTTTTAACTG GGCTTCCTCGAC GACTTCTTATCTG GACTTCTTATCTG GACTTCTCATCG GACTTCTCATCG GATTTTCATTTG GATTTTCTATTTG GATATCTCCCCC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCC CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA ATATGGTGT TTCTTG-TCAA CTGGGCTGT CTGTTGTCC CTATTGTTCTC	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTTAC TTT-CTCCCCC TTTT-CTCCCCCA TGATCATCCCA TTTTCCCACCCT CTCCA TTTTCATTTCCT CATCATA CATCATA	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG AT TAG TT TAG GC TAG GC CAG GC TAG AC TAG AC TAG AC	N-CODING 250 GCCCATCTGG GCCCCTTTCC -CCTCTCTCT -TCCAT-TCC -TCCCTTTCC -TCCTTTCC -TCCCTTTCC -TCCCTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCTTCC -TCTCT-TCC	260 TGGTCTTCA- TGCTCTTGCC AGCT TGC TGC TGCTGTCCAA AGTT CTTGCCCAAA CCAGT CTCTCGCCAAA CAGT	270 TTGAGCTG TGAGCAG -GTCCAGCAA ACACTGGCAA ACCATTTGT GCCTA TTTCTATTAA -TGCAGGTGT GCCCAT GCCCAT GCCCAT	280 AGTCCACTAT GGGGGGCCTGT TGCTGTGT TCCTGTGT AGGTTTTT-T G CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GC-CCTACTC AGGACAGA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGGAGC -TCCCCAGT-GC GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAG ATCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT HUMAN HUMAN HUMAN RABBIT RABBIT	β ε β β β β ε η β ε γ β 1 β β β β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCCTCTGAC GACTTCTTATCTG GACTCTCTCTCCC GACTGTCAAACTG GACT-TCTCTCCC GATTTTCTATTTG GATACTCCCACCA	220 CTCCTTTCTAT ATTTTTCCCTT GTCTTTCTCCA CTAATGCACTA -CTTGTGCTGT CTATGTGTC- TTCTGTGACTA TATTGGTCTAT ATATGGTGTC- TTCTTG-TCAA CTGGGCTGT CTGTTGTCC- CTATTGTTCTC	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTCAC- TTT-CTCCCCCC TTT-CTCCCCCC TTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCATTTCA TTTCCATTTCT CATCATA GTCTCA	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG GC TAG GC TAG TT TAG GT CAG GC CAG GC TAG AC TAG AC TAG A	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -TCCAT-TCC -TCCCTTTCC -TCTCT-TTA -GCC -TCGCTTTCT -TCTCT-TCC -TCTCT-TCC -TCTTTTCC -TCTCT-TCC -TCTCT-TCC	260 TGGTCTTCA- TGCTCTTGCC AGCT AGT CCAATTC TGCTGTCCAA AGTT CTCTGCCAAA CAGT CTCTGCCAAA CAGT	270 TTGAGCTG TGT-GAACAA -GTCCAGCAA ACACTGGCAA ACCATTTGT GCCTA TTTCTATTAA -TGCAGGTGT GCCCAT CACTGT	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GCCCCTACTC AGGACAGA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC CCTATCTGAAG GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAT 	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT	β ε β γ2 β ε ε η β ε G γ β β β β β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTAACTG GACTTCTCTGAC GACTTCTCTGCC AATTTTGTATCTG GACTGTCAAACTG GACT-TCTCCCC GATTTTCTATTTG GATACTCCCACCA	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTTCTCCC CTAATGCACTA -CTTGTGCCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTCTAT CTGGGCTGT CTGTTGTCC CTATTGTTCTC	230 24 TTCTTT-CCCA TCTCCTA TGTTTCCCTTTT TTTCCCCCC TTTT-CTCCCCCC TTTCCCACCCA TTTCCCACCCA TTTCCCACCCA TTTCCCACCCA TTTCCCACCTA TTTCCTCCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCCACCTA TTTCCTCCA TTTCCCACCTA TTTCCCACCTA TTTCCCACCTA TTTCCCCACCTA TTTCCCACCTAC TTTCCCACCTAC TTTCCCACCTAC TTTCCCACCTAC TTTCCCACCTAC TTTCCCACCTAC TTTCCCACCTAC TTTCCCACCTAC TTTCCCACCTAC TTTCCCCACCTAC TTTCCCCACCTAC TTTCCCCACCTAC TTTCCCCACCTAC TTTCCCCACCTAC TTTCCCCACCTAC TTTCCCCACCTAC TTTCCCCCCCCACCTAC TTTCCCCCCCCCC	3' NOM AAG GCCTTM CAG ATATAM TAG AC TAG GC TAG GC TAG GC TAG GC TAG GC CAG GC CAG GA TAG A	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCTCTCTCT -TCCAT-TCC -TCCCTTTCC -TCCCTTTCC -TCCCTTTCC -TCCCTTCT -TCCTTCT -TCCTTCT -TCCTTTCC -TCTCTTCC -TCTCTTCC	260 TGGTCTTCA- TGCTCTTGCC AGCT AGT TGC TGCTGTCCAA AGTT TGCTGTCCAA AGTT CTCTGCCAAA CCAGT TT	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTGT GCCTA TTCTATTAA -TGCAGGTGT GCCCAT CACTGT	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T G	290 3 GTTCCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTTCCCAAGAGC GTCCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAT -TCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN RABBIT RABBIT RABBIT	β ε β β β ε 1 β ε G γ β β β β β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GACTTCTCTCTGAC GACTTCCTCTGAC GACTTCTTATCTG GACTGTCAAACTG GACTGTCAAACTG GACTTCTCTCCCC GATTTTCTATTTG GACTTCTCTCCCC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCC CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGT TTCTTG-TCAA CTGGGCTGT CTGTTGTCC CTATTGTTCTG	230 24 TTCTTT-CCCA GTTTCCCTTTT GCTTAC TTTGTCT TTTT-CTCCCCCC TTTT-CTCCCCCC TTTCCCACCCT TTTCCCACCCT CATCTCA TTTCCCACCCT CTCCA TTTTCATTTTCT CATCATA GTCTCTA	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG GC TAG GC TAG GC TAG GC CAG GC TAG AC TAG AC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCCTTTCC -TCCAT-TCC -TCCTTTCC -TCTCTTTC -TCCCTTTCC -TCCCTTTCC -TCTCTTCC -TCTCTTCC -TCTCTTCC -TCTCTTCC	260 TGGTCTTCA- TGCTCTTGCC AGCT TGC TGC TGCTGTCCAA AGTT CTCTGCCAAA CCAGT CTTGCCAAA	270 TTGAGCTG -GT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTGT GCCTA TTTCTATTAA -TGCAGGTGT GCCCAT CACTGT	280 AGTCCACTAT GGGGGGCCTGT TGCTGTGT TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GCCCCTACTC AGGACAGA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG TGTTCCCTAAGT GACCCTGAC-AC GCTTTCAAGGAT ATCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT	β ε β <b>2</b> β <b>1</b> β ε Γ β ε γ β β β β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCCTCGAC GACTTCTTATCTG GACTCTCTCTCCC GACTTCTCAAACTG GACT-TCTCTCCCC GATTTTCTATTTG GATACTCCCACCA 310	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCC CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGT TTCTTG-TCAA CTGGGCTGT CTGTTGTCC CTATTGTTCTC 320	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTTAC TTTT-CTCCCCA TTTT-CTCCCCA TTTTCCCACCCT CATCTCA TTTTCCCACCCT CTCTA TTTTCATTTTCT CTCTA STCTCA 330 34	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG GC TAG TT TAG GC TAG GC CAG GC CAG GA TAG AC TAG A 40	N-CODING 250 GCCCATCTGG GCCCCTTTCC -CCCTTTCC -CCTCTCTT TCCAT-TCC -TCCCTTTCC -TCCTTTCC -TCCCTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC	260 TGGTCTTCA- TGCTCTTGCC AGCT CGATTC TGC TGCTGTCCAA AGTT CTCTCGCCAAA CAGT CTTCGCCAAA CAGT CTT 360	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA ACCATTTTGT GCCTA TTTCTATTAA -TGCAGGTGT GCCCAT CACTGT 370	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-TI TCCTGTGT TCCCATGT CCATGCTGGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA  GCCCCTACTC AGGACAGA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC CATCCTCAGAGC GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAT ATCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN RABBIT RABBIT RABBIT	β ε β β β ε 1 β ε 6 γ 1 β ε 6 γ 1 β β 3 8 8 8 8 8	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCTTGCC GACTTCTTATCTG GACTGTCAAACTG GACTTCTCTCCCC GATTTTCTATTTG GATACTCCCACCA 310 ACCTT-CTGCACATGG	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCA CTAATGCACTA -CTTGTGCTGT CTATGTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGTAT- TTCTTG-TCAA CTGGGCTGT CTGTTGTCC CTATTGTTCTC 320 AATGAATAGGG	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTCCCA TTTGTCT- TTT-CTCCCCCC CATCTCA TTTTCCCACCCT CATTTCA TTTCCCACCCT CATCTCA ATCTC-A TTTCATTTCT CATCATA GTCTCA GTCTCA	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG GC TAG GC TAG GC TAG GT CAG GC CAG GC TAG AC TAG AC TAG AC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -TCCAT-TCC -TCCCTTTCC -TCTCT-TTA -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-CCC 350 CCCTGACCAA	260 TGGTCTTCA- TGCTCTTGCC AGCT CAATTCC TGCTGTCCAA CGTGTCCAA CGTT CTCTGCCAAA CAGTT CTTT 360 TAAACTGCAT	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA ACACTGGCAA ACCATTTGT GCCTA TTTCTATTAA -TGCAGGTGT CACTGT 370 TCCT	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-TI TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA  GCCCCTACTC AGGACAGA 380 TCCA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAT 	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT	β ε β γ β β ε η β ε ς γ 1 β β β β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC GATTTTCACTG GACTTCTCTGAC GACTTCTCTGCC GACTTCTCTGCC GACTGTCAAACTG GACTGTCAAACTG GACTACTCCCACCA 310 ACCTT-CTGCACATGG CATCTC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCCA CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGTC TTCTTG-TCA CTGTTGTCC CTATTGTCCC CTATTGTTCTC 320 AATGAATAGGG	230 24 TTCTTT-CCCA GTTTCCCTTTT CTTCC TTTCTCCCCC TTTT-CTCCCCCC TTTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCCCCCCACCCT TTTCCCCCCCCACCCT TTTCCCCCCCCCACCCCA	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG GC TAG GC TAG GT CAG GC TAG GC TAG AC TAG AC TAG AC TAG AC TAG AC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCCTTTCC -TCCAT-TCC -TCCCTTTCC -TCTCTTTC -TCTCTTCC -TCCCTTTCC -CCCTTTCC -CCCTCTCC -CCCC 350 CCCTGACCAA	260 TGGTCTTCA- TGCTCTTGCC AGCT TGC TGC TGCTGTCCAA AGTTTC TGCTGTCCAA AGTT CTTGCCATGCCAAA CCGGT CTTGCCCAAA CCGGT CTTGCCCAAA CCGGT CTTC	270 TTGAGCTG -GT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA TTTCCAGGAA TTTCTATTAA GCCCAT GCCCAT CACTGT 370 TCCT CACCCCTTGA	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT TCCCATGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GCCCCTACTC AGGACAGA 380 TCCA CCAA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTTCCCAAGAGA GTCCCCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAT -TCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT	β ε β β β β β β β β β β β β β β ε β β ε β β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε β ε ε β ε ε β ε ε β ε ε β ε ε β ε ε β ε ε β ε ε β ε ε β ε ε β ε ε β ε ε β ε ε ε β ε ε β ε ε ε β ε ε ε β ε ε ε β ε ε ε ε ε ε β ε	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCTCATCTG GACTTCTCATCTG GACTGTCAAACTG GACTGTCAAACTG GATTTTCTATTTG GATTTTCTATTTG GATACTCCCACCA 310 ACCTT-CTGCACATGG CCATCTC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCCA CTAATGCACTA -CTTGTGTGTCT- TTCTGTGACTA TATTGGTCTAT ATATGGTGT- TTCTTG-TCAA -CTGGGCTGT CTGTTGTCC CTATTGTTCTC 320 AATGAATAGGG TTACATC CCAAT-ATG	230 24 TTCTTT-CCCA GTTTCCCTTTT GCTTAC TTT-CTCCCCC TTTT-CTCCCCC TTTTCCCACCCT TTTCCCACCCT CATCTCA TTTCCCACCCT CATCTCA GTTTCCTTTCT CTCCA GTCTCA GT	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG GC TAG TT TAG GC TAG GC TAG GC TAG AC TAG AC TAG AC 40 GGGCATGGT GCA-ATGGT	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCTCTCTTC -TCCAT-TCC -TCCCTTTCC -TCCTTTCC -TCCTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-CCC 350 CCCTGACCAA TACGTACTGC	260 TGGTCTTCA- TGCTCTTGCC AGCT TGC TGC TGCTGTCCAA AGTTTC TGCTGTCCAAA CAGT CTCTGCCAAA CAGT TT 360 TTAAAAGCGCAT	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTGT GCCTA TTCTATTAA -TGCAGGTGT CACTGT 370 TCCT CACCCCTTGA TTAT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T CCCATGT AGGTTCCT-T TCCTGT GATGCAGA A	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG TGTTCCCTAAGT GACCCTGAC-AC GCTTTCAAGGAT -TCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT OPOSSUM OPOSSUM MOUSE	β ε β β β β ε η β ε γ 1 β ε γ 1 β ε β ε β ε β ε β ε β ε β ε β ε β ε β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GACTTTTTAACTG GACTTCTTACTG GACTTCTTATCTG GACTCTCTCTCCC GACTTCTCAAACTG GACT-TCTCCCC GATTTTCTATTTG GATACTCCCACCA 310 ACCTT-CTGCACATGG CCATCTC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCA CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGTC CTGTGTGTCC CTATTGTTCTC 320 AATGAATAGGG TTACATC GCAAAATG	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTCCTA TGTTTCCCTTTT ACGTCT- TTTT-CTCCCCA TGATCATCCCA TTTTCCCACCCT CATCTCA TTTCCATCTTCA TTTCCATCTCA GTCTGACCCTGT GTAAGGAAGGGA ATAGGACATTTGA CTCTCCCCCCC	3' NOI AAAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG AT TAG GC TAG GC TAG GC CAG GA TAG AC TAG AC 40 GGGCATGGT GCA-ATGGT AAATCTGTC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCTCTCTTTCC -TCCAT-TCC -TCCTTTCC -TCTCT-TTA -TCCTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC 350 CCCTGACCAA TACGTACTGC	260 TGGTCTTCA- TGCTCTTGCC AGCT CAATTC TGCTGTCCAA CGTT TGCTGTCCAA CGTT TGCTGTCCAA CGT TT 360 TTAAACTGCAT CTTGTGCCCC TTAAAAGCAT	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA ACCATTTTGT GCCTA TTTCTATTAA -TGCAGGTGT CACTGT 370 TCCT CACCCCTTGA TTCATGT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT TCCCATGT CCATGCTGGT AGGTTCCT-T CCTGT GATGCAGA GACAGA 380 TCCA CCAA GTTC ATCC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC GCCTATCTGAAG TGTTCCCTAAGT GACCCTGAC-AC GCTTTCAAGGAT 	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT OPOSSUM OPOSSUM MOUSE MOUSE	β ε β γ β β ε 1 η β ε G γ β β β β β β β β β β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCTCTGAC GACTTCTCTGCC GACTTCTCATCTG GACTGTCAAACTG GACTGTCAAACTG GACTACTCCCACCA 310 ACCTT-CTGCACATGG CCATCTC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTTCTCC/ CTAATGCACTA -CTTGTGTGACTA TATTGGTCTAT TATTGGTCTAT ATATGGTGT TTCTTG-TCAT CTGGGCTGT CTGTTGTCC CTATTGTTCTC 320 .AATGAATAGGG TTACATG GCAAAATG/ AATACTG-	230 24 TTCTTT-CCCA TCTCCTA TGTTTCCCTTTT TTTCCCCCC TTTT-CTCCCCCC TTTTCCCCCCCC TTTCCCCCCCC TTTCCCCCCCC	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG GC TAG GC TAG GC TAG GT CAG GC TAG GT CAG GC TAG AC TAG AC TAG AC TAG AC TAG AC TAG AC TAG AC GGGCATGGT AAATCTGTC -AGCACA	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -TCCAT-TCC -TCCCTTTCC -TCCCTTTCC -TCCTTTTC -TCCTTTC -TCCTTCC -TCCTTCC -TCCTTCC -TCCCTCC -TCCCTCC -TCCCTCC -TCCCTCC -TCCCC -TCCCCA -TCCCTGACCAA TCCTGACCAAA -TCCTGTCACCAAA	260 TGGTCTTCA- TGCTCTTGCC AGCT AGT CCAATTCC TGCTGTCCAA AGTT CTCTGCCCAAA CAGTT CTCTGCCCAAA CAGT CTCTGCCCAAA CAGT CTTGCTGCCCAA CAGT CTCTGCCCAAA CAGT CTCTGCCCAAA CAGT	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTTGT GCCCAT GCCCAT CACTGT 370 TCCT CGCACCCTTGA TTAT CCGAATCAAA	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-TI TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA  GCCCCTACTC AGGACAGA 380 TCCA CCAA GTTC ATCG	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAT -TCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT OPOSSUM OPOSSUM MOUSE MOUSE	β ε β β ε η β ε η β ε β β β β ε β β β ε β 2 11 β ε β ε β ε β 2 11 β ε β ε β 2 11 β ε β ε β 2 11 β ε β β 1 β ε β β 1 β ε β ε β β 1 β ε β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTAACTG GACTTCTCTGAC GACTTCTTATCTG GACTGTCAAACTG GACTGTCAAACTG GACTTCTCTCCC GATTTTCTATTTG GATTACTCCCACCA 310 ACCTT-CTGCACATGGT CCTCT-CTGCACATGTG CCTCT-CTGCACCATGT CCTCT-CTGCACCATGT	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCCA CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTCTAT -CTGGGCTGT CTGTTGTCC CTATTGTCCC- CTATTGTCCC- CTATTGTCCC- CTATTGTCCC- GCAAA-ATGG GCAAA-ATGG GCAAA-ATGG GGGACTGA	230 24 TTCTTT-CCCA TCTCCTA TCGTCT TTT-CTCCCCCC TTTT-CTCCCCCC TTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCA TTTCCTCCA TTTCCCCCCC TTTCCTCA TTTCCCCCCCCCC	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG GC TAG GC TAG GC TAG GC TAG GC TAG AC TAG AC TAG AC TAG AC TAG AC TAG AC GGGCATGGT AAATCTGTC -AGCACAA GAGCACAAA	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCCTTTCC -TCCAT-TCC -TCCCTTTCC -TCTCT-TTA -GCC -TCGCTTTCT -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-CCC 350 CCCTGACCAA TCTGACAAA -TCATGTTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT TGC TGC TGCTGTCCAA AGTTTC TGCTGTCCAA AGTT CCTCTGCCAAA CCTGCCCAAA CAACTGCAT TAAAACTGCAT TAAAAGCACCT TAAAAGCACCT TAAAATTGAT	270 TTGAGCTG -GT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTTGT GCCTA TTTCTATTAA -TGCAGGTGT CACTGT 370 TCCT CCACCCCTTGA TTAT CCGAATCAAA TCTTAATAAC	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GGCCCCTACTC AGGACAGA 380 TCCA CCCA GTTC ATCG ATCG	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC GCCTATCGAAGC -TCCCCAGT-GC GCCTATCTGAAGT GACCCTGAC-AC GCTTTCCAAGGAT ATCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT OPOSSUM OPOSSUM MOUSE MOUSE COW	β ε β β β β ε η β ε γ β β β ε β β ε β ε β ε β ε β ε β ε	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GACTTCTCTCTGAC GACTTCTTATCTG GACTGTCAAACTG GACTGTCAAACTG GATTTTCTATTTG GATTTTCTATTTG GATACTCCCACCA 310 ACCTT-CTGCACATGG CCATCTC- ATCTG-TCAGTTGTTG CCTCT-CTGCA-CATG AAAGA-TTGAATATGG	220 CTCCTTTCTAT ATTTTTCCCTT CCTGT-CTAT GTCTTTCTCCCA CTAATGCACTA -CTTGTGCTGT CTATGTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGT CTGTGGCCGT CTGTTGTCC CTATTGTTCTC 320 AATGAATAGGG TTACATG GCAAAATGA AATACTG- GGGGACTGAA	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTCCTA TGTTCCCTTTT ACCTCCA TTTT-CTCCCCC ATGATCATCCCA TTTTCCCACCCT CATCTCA ATTCAGAACGGA ATAGGAAGGGA ATAGGCATTGA ATAGGCATTGA ATAGGCATTGA ATAGGCATTGA	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG GC TAG GC TAG GT TAG GC TAG GC TAG GC TAG AC AG GA AGCACAGGT AAATCTGTC -AGCACAAA GCATCTGGC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTCC -CCTCTCTC -TCCAT-TCC -TCCCTTTCC -TCCTTTC -TCCTTTC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTGACAAA -TCATGTTAA TTTGTTTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT TGC TGC TGCTGTCCAA AGTT TGCTGTCCAA CAGT CTCTGCCAAA CAGT TT 360 TTAAACTGCAT TTAAAGCACCT TTAAAGCACCT AAAATTGAT	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTGT GCCTA TTCTATTAA -TGCAGGTGT CACTGT 370 TCAT CCGAATCAAA TCTTAATAAC TTAT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA ACCA GCCCTACTC AGGACAGA 380 TCCA CCAA GTTC ATCG ATCG ATCG ATTG TTTC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GCCTTCTAAGGAAG ATCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE COW GOAT GOAT HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT OPOSSUM OPOSSUM MOUSE COW GOAT	β ε β 2 β β β β β β β β β β β β β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GACTTTTTAACTG GACTTCTTACTG GACTTCTTATCTG GACTCTCTCTCCCC GACTTCTCAAACTG GACT-TCTCCCCC GATTTTCTATTTG GATACTCCCACCA 310 ACCTT-CTGCACATGG CCATCTC	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCA CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGTC CTGTGTGTCC CTATTGTTCTC 320 AATGAATAGGC TTACATC GCAAAATG/ GGAACTGA/ GGACTGA/ GGACTGA/	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTCCTA TGTTTCCCTTTT ACGTCT- TTTT-CTCCCCA TGATCATCCCA TTTTCCCACCCT CATCTCA TTTTCCATCTTCA TTTCCATCTCA GTCTGACCCTGT GTAAGGAAGGGA ATTAGGCATTGA ATTAGGCATTGA GTTTGGCCTTGT	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG AT TAG AT TAG GC TAG GC CAG GC CAG GA TAG AC TAG AC GAGCACAGAT GCA-ATGGT GAACCCAGAA	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCTCTCTTTC -TCCAT-TCC -TCCTTTCC -TCTCT-TTA -CCCTTTCT -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCTACCA 350 CCCTGACCAA TACGTACTGC TTCTGACAAA TTCTGCTTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT CAATC CCAT TGCTGTCCAA CGTT TGCTGTCCAA CAGT CTCTGCCAAA CAGT 360 TAAAACTGCAT CTTGTGCCCC TAAAATGCAC TAAAATGCAC TAAAATGCAC TAAAATGCAC TAAAATGCAC	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA ACCATTTTGT GCCTA TTTCTATTAA -TGCAGGTGT CACTGT 370 TCCT CGAATCAAA TCAT CCGAATCAAA TTAT TCTATTCAGT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT TCCCATGT AGGTTTTT-T G CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA  GCCCCTACTC AGGACAGA 380 TCCA CCAA GTTC ATCG ATCG ATCG ATTC GATC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GCCTTCCAAGGAT 	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT RABBIT OPOSSUM MOUSE MOUSE COW GOAT GOAT	β ε β 2 1 β ε 3 β ε 3 β ε β 2 1 β ε 3 β ε β 2 1 β ε β 2 1 β β ε η β ε β 2 1 β β ε η β ε β 2 1 β β ε β β ε β β β β ε β β β ε β β β ε β β ε β β ε β β ε β β ε β β ε β β ε β β ε β β ε β β ε ε γ β β β ε ε γ β β β ε β β β ε β β β ε β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTAACTG GACTTCTCTCTGAC GACTTCTTATCTG GACTGTCAAACTG GACTGTCAAACTG GACTTCTCACCCA 310 ACCTT-CTGCACATGG CCATCTC-TGACTCATG CCTCT-CTGCACATGG T-CTT-TGACTCCAGT AAGA-TTGAATATGG TCCTT-CTGCCCTTG CCAGTGTCCCAGA	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCC CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGT TTCTTG-TCA CTGTTGTCC CTGTTGTCC CTATTGTTCTC 320 .AATGAATAGGG TTACATG GCAAAATG/ AATACTG GGGACTGA/ GGACTGGGC 	230 24 TTCTTT-CCCA TCTCCTA TGTTTCCCTTTT TTTCCCCCC TTTT-CTCCCCCC TTTTCCCCCCCC TTTCCCCACCTT TTTCCCACCTT TTTCCCACCTT TTTCCTCCA TTTCCCACCTT TTTCCTCCA TTTCCTCCA TTTCCTCCA TTTCCTCCA TTTCCTCCA TTTCCTCCA TTTCCTCCA TTTCCTCCA TTTCCTCCA TTTCCTCCA TTTCCTCCA TTTCCCCCTCA TTAGGCATTGA TTTGCCCTTGT TTTCCTTCCCA TTTGCCCTTGT TTTCCCCCCC	3' NOI 40 AAG GCCTTI CAG ATATA TAG AC TAG AC TAG GC TAG GC TAG GT TAG GT CAG GC TAG AC TAG AC TAG AC TAG AC 40 GGGCATGGT AAATCTGTC -AGCACAAA GCATCTGGC GAACCCAGAA GCACCCAGAA GCACCCAGAA	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCCTTTCC -TCCAT-TCC -TCCCTTTCC -TCCTTTCC -TCTCT-TCA -TCTTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-CCC 350 CCCTGACCAAA -TCTGACCAAA -TCATGTTAA CTCTGCCTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT TGCTCTTGCC TGC TGCTGTCCAA CAGTTC TGCTGTCCAA CAGTT CTCTCGCCAAA CCTCGCCAAA CCGTTGCCCAA CCTTGCCCAAA CCTTGCCCAAA CCTTGCCCAAA CCTTGCCCAAA CCTTGCCCAAA CCTTGCCCAAA CCTTGCCCAAA CCTTGCCCAAA TTAAAACTGCAT TTAAAAGCACCT TTAAAAGCACCT TTAAAAGCACCT TTAAAAGCACCT TTAAAAGCACCT	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA TTTCCAGGAA TTCCAGGTGT GCCCAT CACTGT 370 TCCT CCGACCCTTGA TTAT CCGAATCAAA TCTTATACAGT GACATATAAC	280 AGTCCACTAT GGGGGCCTGT TGGTTAAT-TI TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA  GCCCCTACTC AGGACAGA 380 TCCA GTCC ATCG ATTG TTTC GATC TCCA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC -TCCCCAGT-GC GCCTATCTGAAG GACCCTGAC-AC GCTTTCAAGGAT 	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN HUMAN HUMAN OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN	β εβ β β β β β β β β β β β β β β β β ε β β β ε δ ε δ	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTAACTG GACTTCTCTGAC GACTTCTTATCTG GACTGTCAAACTG GACTGTCAAACTG GACT-TCTCCCC GATTTTCTATTTG GATACTCCCACCA 310 ACCTT-CTGCACATGGT CCTCT-CTGCACATGGT CCTCT-CTGCACATGGT AAGA-TTGAATATGG TCCTT-CTGCCCCTG	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTCTCCCA CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGTC CTGGTGTCC CTATTGTTCTC 320 AATGAATAGGC GCAAA-ATGA GGAACTGGA AAAAATTATG GGGACTGGA GGGACTGGA	230         24           ITTCTTT-CCCA.           IGTTCCCTTTT           IGTTCCCTTTC           IGTTCCCTTTC           IGTTCCCTTCA.           ITTT-CTCCCCCC.           ITTTCCCACCCT.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCATTTCA.           ITTCCATCTCA.           ITTCCATTTCA.           ITTCCATTTCA.           ITTCCATCTCA.           ITTCCATTTCA.           ITTCCATCTCA.           ITTCCATCTCA.           ITTCCCACCCT.           ITTCCTCCA.           ITTCCCACCCTGA.           ITAGACATTTGA.           ITAGGCATTGA.           ITTAGGCCTTGA.           ITTGGCCTTGA.           ITTGGCCTTGA.	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG AC TAG GC TAG GC TAG GC TAG GC TAG AC TAG AC TAG AC GGGCATGGT AAATCTGTC -AGCACAAA GAA-G-ACA GAA-G-ACA GCATCTGGA	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCCTTTCC -TCCAT-TCC -TCCTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-CCC 350 CCCTGACCAA TACGTACTGC TTCTGACAAA -TCATGTTAA TTCTGCCTAA GAGGGAGAGAGA	260 TGGTCTTCA- TGCTCTTGCC AGCT TGC TGC TGCTGTCCAA AGTTTC TGCTGTCCAA AGTT CCTCTGCCAAA CCAGT CAGT TT 360 TTAAAACTGCAT TAAAAACACT TAAAAAGCACCT TAAAAACTCGAT TAAAAACACAT	270 TTGAGCTG TGAGCAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTGT GCCTA TTTCTATTAA -TGCAGGTGT CACTGT 370 TCCACTGT 370 TCAT CCGAATCAAA TCTTAATAAC TTAT CACCATATAAC TTAT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GGTTCCTACTC AGGACAGA 380 TCCA GATC ATCG ATCG ATTG TTCC GATC TTCC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC GCCTATCTCAAGG TGTTCCCAGT-GC GCCTATCTGAAGT GACCCTGAC-AC GCTTTCAAGGAT ATCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE COW GOAT GOAT HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT RABBIT OPOSSUM	βεβ ΥβΗ βεη βεγ ββ β β β β β β β β β β β β β β β β	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GATTTTCCATCTGAC GACTCTCTTATCTG GACTGTCAAACTG GACTGTCAAACTG GATTTTCTATTTG GATACTCCCACAA 310 ACCTT-CTGCACATGG CCATCTC-CTGCA-CATG CCTCT-CTGCACCATGG CCTCT-CTGCACCATGG AAAGA-TTGAATATGG TCCTT-CTGCCCCTG CCAGTGTCCCAGA AACTA-CTAAACTGGG	220 CTCCTTTCTAT ATTTTCCCTT CCTGT-CTAT GTCTTTCTCCCA CTAATGCACTA -CTTGTGCTGT CTATGTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGT CTGTGGCCGT CTGTTGTCC CTATTGTTCTC GCAAAATGA AATACTG GGGACTGA GGGACTGG GGGATATTATG	230         24           CTTCTTT-CCCA.           GTTTCCCTTTT           GTTTCCCTTTT           AGTTCCCTTTT           AGTTCCCTTTC.           TTT-CTCCCCA.           ATGATCATCCCA.           ATTCCCACCCT.          CATCTCA.           ATTCCCACCCT.          CATCTCA.           ATTCCCACCCT.          CATCTCA.           ATTCCCACCCTGA.           330         34           ATAGGAAGGGA.           ATAGGAAGGGAA.           ATAGGCATTGA.           ATTAGGCATTGA.           ATTAGGCATTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG GC TAG GC TAG GT TAG GC TAG GC TAG GC TAG AC AGCACAGGG GAACCTGGC GAACCCAGA GCATCTGGC GAACCCAGA GCATCTGGC GAGAAAGCC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTCCT-TCC -CCTCTCTTCC -TCCCTTTCC -TCCTTTCC -TCCTTTCC -TCCTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTGACAAA -TCTGACCAAA -TCTGGCTAA TTCTGCTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT TGCTCTTGCC TGC TGCTGTCCAA CAGTTC TGCTGTCCAA CAGTT CCTCTGCCAAA CAGTT 360 TTAAACTGCAT TTAAAGCACCT TTAAAGCACCT TTAAAGCACT TTAAAATACAT ATAAAAAACAT TTAAAGTACAT	270 TTGAGCTG TGAGCAA -GTCCAGCAA ACACTGGCAA TTCCAGGAA ACCATTTGT GCCTA TTCTATTAA -TGCAGGTGT CACTGT 370 TCAT CCGAATCAAA TCTAATAAC TTAT TCTATTCAGT GACATATAAC TTAT TCTATCAGT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA  GCCCCTACTC AGGACAGA 380 TCCA GTTC ATCG ATTG TTTC GATC TTTC GATC TTTC AATC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC GCCTATCTCAAGAGC -TCCCCAGT-GC GCCTATCTGAAG GCTTTCAAGGAAG ATCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE COW GOAT GOAT HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT RABBIT OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN	βεβγββεηβεγ143 βεβ2Η βεγ1443 βεβγββεηβεγ	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTAACTG GACTTCTCTGAC GACTTCTTATCTG GACTCTCTCTGCC GACTGTCAAACTG GACT-TCTCCCCC GATTTTCTATTTG GACTACTCCACCA 310 ACCTT-CTGCACATGG CCATCTC-TGCACATGG CCTT-CTGCACATGG AAAGA-TTGAATATGG TCCTT-CTGCACCTGG CCAGTGTCCCAGA AACTA-CTAAACTGGG TCCTT-CTGCA-CATGG	220 CTCCTTTCTAT ATTTTTCCCTT GTCTTTCTCCCA CTAATGCACTA -CTTGTGTGACTA TATTGGTGTT TTCTGTGACTA -TTCTGGTCATA -CTGGGCTGT CTGTTGTCC- CTATTGTTCTC 320 AATGAATAGGG TTACATG GCAAAATGA AAAAATATGC-GG GGACTGA GGGACTGGG GGGACTGGG GGGACTGGG GGGACTGGG	230 24 TTCTTT-CCCA GTTTCCCTTTT ACTCCA TGTTTCCCTTTT ACGTCT- TTT-CTCCCCA TGATCATCCCA TTTTCCCACCCT CATCTCA TTTCCCACCCT CATCTCA ATTAGGAAGGGA ATAGGAAGGGA ATTAGGCATTGA A-AGTGTTTTGA GTTTGGCCTTGA GCTTGGCCTTGA ATAACCTTGA CCTGGCCTTGA	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG AC TAG AT TAG TT TAG GC TAG GC CAG GA TAG AC TAG AC TAG AC TAG AC GAGCACAGA GAGCACAGAG GAACCCAGAA GAAAGCCC GAACCCAGAA GAAAGCCC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCTCTCTTCC -TCCCTTTCC -TCCCTTTCC -TCCTTTCT -TCCTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCTAC 350 CCCTGACCAA TCCGCTAA TCTGCTAA TTCTGCTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT CAATTC CAATTC CAATTC TGCTGTCCAA AGTT TGCTGTCCAA CAGTT CTCTGCCAAA CAGTT 360 TAAAACTGCAT TAAAAGCACT TAAAAGCACT TAAAAGCACAT TAAAAGACAT TAAAAGACAT TAAAAGACAT TAAAAGACAT TAAAAGACAT TAAAAGACAT	270 TTGAGCTG TGAGCTA -GTCCAGCAA -GTCCAGCAA ACACTGGCAA ACACTGGCAA ACCATTTTGT GCCTA TTTCTATTAA -TGCAGGTGT CACTGT 370 TCCT CGAATCAAA TCAT CCGAATCAAA TCTAATAAC TTAT TCTATTCAGT GACATATAAC TTAT TTTCTTCAGT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT TCCCATGT AGGTTTTT-T GC CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GTTCCA GCCCCTACTC AGC-GACAGA 380 TCCA GTTC ATCG ATTC GATC TTTC GATC TCCA TTTC GATC TCCA	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC GCCTATCTGAAG GCCTATCTGAAG GCCTTCCAAGGAT 	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT RABBIT COPOSSUM MOUSE MOUSE COW GOAT GOAT GOAT GOAT HUMAN HUMAN HUMAN	βεβΥββεηβεGβββ βεβΥββεηβεGΥ	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC GATTTTCAGTGTC GATTTTTAACTG GACTTCTCTGAC GACTTCTTATCTG GACTGTCAAACTG GACT-CTCTCCCC GATTTTCTATTTG GATTATCTATTTG GATTACTCCCACCA 310 ACCTT-CTGCACATGG CCATCTC	220 CTCCTTTCTAT ATTTTTCCCTT CCTGT-CTAT GTCTTTCTCCC CTAATGCACTA -CTTGTGCTGT CTATGGTCTAT TATTGGTCTAT ATATGGTGT TTCTTG-TCAT CTGTTGTCC CTGTTGTCC CTGTTGTCC CTATTGTTCTC 320 .AATGAATAGGG TTACATG GGAACTGA .GGGACTGA .GGGACTGGG .AATACAA	230 24 TTCTTT-CCCA TCTCCTA TGTTTCCCTTTT TTTCTCCCCC TTTT-CTCCCCCC TTTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCCACCCT TTTCCTCCA TTTCCCACCCT TTTCCTCCA TTTCCCACCCT TTTCCTCCA TTTCCCCCCCCC 330 34 STCTGACCCTGA STCTGACCCTGA ATAGACATTTGA CTGTCCCTGA STTGGCCTTGA STTGGCCTTGA STTGGCCTTGA STTGGCCTTGA STTGGCCTTGA STTGGCCTTGA	3' NOI AAG GCCTTI CAG ATATA TAG AC TAG AC TAG AC TAG GC TAG GC TAG GC TAG GC TAG GC TAG AC TAG AC TAG AC TAG AC AGCACAAA GCATCTGGC GAACCAGAA GCATCTGGA GAAAGCCAAA GCATCTGGA	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCCTTTCC -TCCAT-TCC -TCCCTTTCC -TCCTTTCC -TCCTTTCC -TCCTTTCC -TCCTTTCC -TCCTTTCC -TCTCTTCC -CCCC 350 CCCTGACCAA TCCTGACCAA TCTGGTAA TCTGCTAA TTCTGCTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT TGCTCTTGCC TGC TGCTGTCCAA CAGTTC TGCTGTCCAA AGTT TGCTGTCCAA CAGT TT 360 TTAAAGCACAT TAAAAGCACCT TTAAAAGCACCT TTAAAAGCACAT TAAAAACAT TAAAAACAT TAAAAACAT TAAAAACAT	270 TTGAGCTG GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA TTTCCAGGAA TTTCTATTAA GCCCAT CACTGT 370 TCCT CCGAATCAAA TCTTATAAAAC TTAT CCGAATCAAAA TCTTATTCAGT GACATATAAC TTAT TTTCTTCAGT GAG	280 AGTCCACTAT GGGGGGCCTGT TGCTGTGT TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GGCCCCTACTC AGGACAGA 380 TCCA GCCA GTTC ATCG ATCG ATCG ATCG ATCG TTTC GATC TTCC ATCC TTCC ATCC TTCC ATCC TTCC ATCC TTCC ATCC TTCC ATCC TTCC ATCC TTCC ATCC TTCC ATCC TTCC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC GCCTATCTCAAGA GCCTATCTGAAG GCTTTCCAAGTAG GCTTTCAAGGAT ATCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT OPOSSUM MOUSE MOUSE MOUSE COW GOAT HUMAN HUMAN HUMAN RABBIT	βεβγβΗ βεηβεηβεGγ143 βεβγβΗ βεβγβββ βεβγββεηβεγ ββ1 βεβγ	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC GATTTTCT-ACTT GATTTTTTAACTG GACTTCTCTGAC GACTTCTTATCTG GACTGTCAAACTG GACTGTCAAACTG GATTTTCTATTTG GATTTCTCATTTG GATACTCCCACCA 310 ACCTT-CTGCACATGGG CCATCTC-TGCACATGGG T-CTT-TGACTCCATG AAGA-TTGAATATGG TCCTT-CTGCCCAGA AACTA-CTAAACTGGG TCCTT-CTGCCCAGA AACTA-CTAAACTGGG CCTT-CTGCCCAGA AACTA-CTAAACTGGG CCTT-CTGCCCAGA AACTA-CTAAACTGGG CCTT-CTGCCCAGA AACTA-CTAAACTGGG CCTT-CTGCCACATGG	220 CTCCTTTCTAT ATTTTCCCTT CCCTGT-CTAT GTCTTTCTCCCA CTAATGCACTA -CTTGTGCTGT CTATGTTTC TTCTGTGACTA TATTGGTCTAT ATATGGTGTC CTGGTGCCC CTATTGTTCTG CGCAAA-ATGA GGAACTA-GA GGGACTGA GGGACTGA GGGACTGGG AATACAA GGGACTGGG	230         24           TTCTTT-CCCA.           IGTTCCCTTTC           IGTTCCCTTTC           IGTTCCCTTCA.           ITT-CTCCCCCC.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCCACCCT.           ITTCCATTTCA.           ITTCCATTTCA.           ITTCCATTTCA.           ITTCCATTTCA.           ITTCCATTTCA.           ITTCCATTTCA.           ITTCCATTTCA.           ITTCCATCTCA.           ITTCCATCTCA.           ITTCCATCTCA.           ITTAGGCATTGA.           ITTAGGCCTTGA.           ITTGGCCTTGA.           ITTGGCCTTGA.           ITTGGCCTTGA.           ITTGGCCTTGA.           ITTGGCCTTGA.           ITTAGCCCTTGA.           ITTAGCCTTGA.           ITTAGCCTTGA.	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG GC TAG TT TAG GC TAG GC TAG GC TAG GC TAG AC TAG AC TAG AC TAG AC GAGCATGGT AAATCTGTC GAACCAAAA GCATCTGGAC GAACCCAGAA GAA-G-ACAA GCATCTGGAC GAGAAAGCC AAATCTA GCATCTGGAC	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTCCT-TCC -CCTCTCTTCC -TCCCTTTCC -TCCTTTCC -TCCTTTCC -TCCTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCTCCCCC 350 CCCTGACCAA TCTGACAAA -TCATGTTAA CTCTGCCTAA TTCTGCTAA- TTCTGCTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT TGCT TGC TGCTGTCCAA AGTTTC TGCTGTCCAA AGTT CCTCGCCAAA CAGT CTTGCCAAA CAGT 360 TTAAAACTGCAT TTAAAAGCACT TTAAAAGCACT TTAAAAGCACAT TTAAAAGCACAT TTAAAAGCACAT TTAAAAGCACAT	270 TTGAGCTG TGAGCAA -GTCCAGCAA ACACTGGCAA TTCCAGGAA ACCATTTGT GCCTA TTCTATTAA -TGCAGGTGT CACTGT 370 TCCACTGT 370 TCAT CCGAATCAAA TCTTAATCAGT GACATATAAC TTAT TTTCTTCAGT GAG TTAT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA ACCCTACTC AGGACAGA 380 TCCA CCAA GTTC ATCG ATTC GATC TTTC AATCA TTTC AATCA TTTC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC GCCTATCTCAAGG TGTTCCCTAAGT GACCCTGAC-AC GCTTTCAAGGAT ATCCCCAGC-GT GCTTCTAGGAAG	
OPOSSUM OPOSSUM MOUSE COW GOAT GOAT HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT OPOSSUM MOUSE MOUSE COW GOAT GOAT HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN	βεβΥββεηβεγ143 βεβ2Η βεγββεγ34 βεβΥββεηβεγ143 βεβΥββεηβεγ14	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTTAACTG GATTTTCCATCTG GACTCTCTCTCCCC GACTCTCTATCTG GACTGTCAAACTG GACT-TCTCCCCCACA 310 ACCTT-CTGCACATGG CCATCTC-TGCACATGG CCTCT-CTGCACCATGG CCTCT-CTGCACCATGG CCAGTGTCCCAGA AACTA-CTAAACTGGG TCCTT-CTGCACCATGG CCTT-CTGCACCATGG	220 CTCCTTTCTAT ATTTTCCCTT GTCTTTCTCCA CTAATGCACTA -CTTGTGCACTA -CTTGTGTGTC- TTCTGTGACTA TATTGGTCTAT ATATGGTGTA- -CTGGGCTGT CTGTTGTCC CTATTGTTCTC- GGCAAAATGA AATGAATAGGG GCAAAATGA GGGACTGA GGGACTGA GGGACTGA GGGACTGA GGGACTGA GGGACTGA GGGACTGA	230         24           CTTCTTT-CCCA.           GTTTCCCTTTT           CTCCTA.           GTTTCCCTTTT           ACTTAC           TTT-CTCCCCA.           TTTCCCATCTCA.           TTTCCCACCCT          CATCTCA.           TTTCCCACCCT          CATCTCA.           TTTCCCACCCT          CATCTCA.           TTTCCCACCCT          CTCA.           330           34           TCTGACCCTGA.           TTAGGGATTGA.           ATAGGACATTGA.           ATAGGCCTTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.           ATTAGGCCTTGA.           ATAGGCCTTGA.           ATAGCCCTTGA.           ATAGGCCTTGA.           ATAGCCCTTGA.	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG GC TAG GC TAG AT TAG TT TAG GC TAG GC TAG GC TAG GC TAG AC TAG AC AGCACACAGT GAGCACACA GCACCCAGA GAA-G-ACA GCATCTGGC GAGCACAGAT GCATCTGGAC GAGCACAGAT GCATCTGGAC GAGCACAGAT	N-CODING 250 GCCCATCTGG GCCCCTTCC -CCTCTCTTC -CCCTTCC -TCCCTTCC -TCCTTTC -TCCTTCC -TCCTTCC -TCCTTCC -TCCTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCTCCCCA -TCTGACAAA -TCTGCCTAA TTCTGCTAA- TTCTGCTAA- TTCTGCTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT TGCTCTTGCC AGCT TGC TGCTGTCCAA CAGTT CTCTGCCAAA CAGTT CTTGCCAAA CAGTT CTTGCCAAA CAGT TT 360 TTAAAACTGCAT TAAAAGCACT TAAAAGCACT TAAAAGCACT TAAAAGCACT TAAAAGCACT TAAAAGCACT TAAAAGCACT TAAAAGCACT TAAAAGCACT TAAAAGCACT TAAAAGCACAT TAAAAGCACAT TAAAAGCACAT	270 TTGAGCTG TGAGCAGCAA -GTCCAGCAA ACACTGGCAA ACACTGGCAA ACCATTTGT GCCTA TTCCATTAA -TGCAGGTGT CACTGT 370 TCAT CCGAATCAAA TCTTAATAAC TTAT TCTATTCAGT GAG TTAT TCTATTCAGT GAG TTAT TCTATTCAGT	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-TI TCCTGTGT AGGTTTTT-T CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GACAGA 380 TCCA GCCCTACTC AGTCC ATCG ATCG ATCG ATC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC GCCTATCTGAAG GCCTATCTGAAG GCTTTCAAGGAAG	
OPOSSUM OPOSSUM MOUSE COW GOAT GOAT HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT OPOSSUM MOUSE COW GOAT HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN HUMAN RABBIT RABBIT RABBIT RABBIT	βεβΥββεηβεGβ143 βεβ2H1 βεGβ143 βεβ2H1 βεGβ143	210 AAACTTGACT-AC GTGTATCTGCAAACTA AACTTTCAGTGTC TATATTCT-ACTT GATTTTTAACTG GACTTCTCTGAC GACTTCTTATCTG GACTCTCTCGCC ATTTTGTATCTG GACT-TCTCCCCC GATTTTCTATTTG GACTACTCCACATGG CCATCTC-CTGCACATGG CCTCT-CTGCACATGG CCTCT-CTGCACATGG CCTTCTCTGCA-CATG CCTTCTCTGCA-CATG GCTTTATTCTGCAAGCG AATTATGG TCCTT-CTGCACATGG GCTTTATCCGCACATG GCTTTATCCCCCAAAT	220 CTCCTTTCTAT ATTTTTCCCTT CCCTGT-CTAT GTCTTTTCTCC/ CTAATGCACT/ CTAATGCACT/ CTATGGTCTAT TATTGGTCTAT ATATGGTCTAT CTGGGCTGT CTGTTGTCC CTATTGTTCTC 320 .AATGAATAGGACTGA/ GGAAAATG/ AATACTG- GGGACTGG GGACTGG GGACTGG GGGACTGA/ GGGACTAGG AATAATGAAA	230         24           TTCTTTT-CCCA.           TGTTTCCTTTCC           TGTTTCCCTTTT           TGTTCCCTTACGTCTGTCTCA.           TGTTCCCACCT           TTTCCCACCCT           TTTCCCACCCT           TTTCCCACCCT           TTTCCCACCT           TTTCCCACCT           TTTCCCACCT           TTTCCCACCT           TTTCCCACCT           TTTCCCACCT           TTTCCCACCT           TTTCCCACCTGT           STCTGACCCTGT           STCTGACCCTGTA           ATAGGACATTTGA          CTGTTCCTGA           ATAGGACATTTGA           A-AGTGTTTTGACTTGT           STTTGGCCTTGT           STTTGGCCTTGA           GTTGGCCTTGA           GCTCAGCCTTGA           GCTCAGCCTTGA           ATAAT           ATAAT	3' NOI AAG GCCTTI CAG ATATAI TAG AC TAG AC TAG AC TAG AC TAG AT TAG TT TAG GC TAG GC CAG GC CAG GC CAG GA TAG AC TAG AC GAGCACAGA GAGCACAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA GAACCCAGA	N-CODING 250 GCCCATCTGG GCCCCGTCTA -CCCTTTCC -CCTCTCTTTCC -TCCCTTTCC -TCCTTTCC -TCCTTTCT -TCCTTTCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCT-TCC -TCTCTACCCA TCTGGCTAA TTCTGCTAA TTCTGCTAA TTCTGCTAA TTCTGCTAA	260 TGGTCTTCA- TGCTCTTGCC AGCT CAATC CCAATC TGCTGTCCAA CGTT TGCTGTCCAA CAGTT TGCTGTCCAA CAGTT 360 TAAAACTGCAT TAAAAGCACT TAAAAGCACT TAAAAGCACAT TAAAAGACAT TAAAAGACAT TAAAAGACAAT TAAAAGACAAT TAAAAGACAAT TAAAAGACAAT TAAAAGACAAT TAAAAGACAAT TAAAAGACAAT TAAAAGACAAT TAAAAGAAAAT TAAAAGAAAAT TAAAAGAAAAT	270 TTGAGCTG GT TGT-GAACAA -GTCCAGCAA ACACTGGCAA TTTCCAGGAA ACCATTTTGT GCCTA TTCTATTAA CACTGT 370 TCCACTGT 370 TCCACTGT 370 TCCACTGT 370 TCAT CCGAATCAAA TCTATTCAGT GACATTATAAC TTAT TCTATTCAGT CACTGT 	280 AGTCCACTAT GGGGGGCCTGT TGGTTAAT-T TCCTGTGT AGGTTTTT-T G CCATGCTGGT AGGTTCCT-T TCCTGT GATGCAGA GTTCCA GACCCTACTC AGC-GACAGA 380 TCCA GCCCCTACTC AGCG ATTC ATCG ATCG ATC	290 3 GTTCCATATAGT GAACACATGGAG GTTCCCAAGAGA GTCCGCTAT-GC GTCTATGAT-GC CATCCTCAGAGC TCCTCCAGAG GCCTATCTGAAG GCCTTCCAAGGAT 	

FIGURE 1.—Multiway alignment of 5' (nucleotides 1–100) 3' (nucleotides 242–381) noncoding sequences and intron 1 (nucleotides 101–241) for the mammalian functional  $\beta$ -like globin genes analyzed. A gap penalty of 2.5 was used for alignment. Gaps are indicated by dashes.

rodents and ungulates, and for the combined genes between rodents and primates. In all three cases the rodent genes have evolved faster. The differences are all significant at the 0.05 level. Since these differences are nonsynonymous, they may result from the differential effects of natural selection; however, three differences at this level of significance may be expected by chance from the total number of comparisons made particularly since they are not independent of each other. There are no significant differences in silentsite or noncoding region substitution rates. Overall there is no consistent pattern of variation among

### S. Easteal

## TABLE 1

## Rates of nucleotide substitution at synonymous sites between eutherian and marsupial genes

		Eutherian order				
Gene	Ν	Rodent	Ungulate	Lagomorph	Primate	
α-Globin	104	$102.7 \pm 20.4$	$97.1 \pm 17.6$	$106.5 \pm 19.6$	$114.9 \pm 24.1$	
$\beta$ -Globin	100	$90.0 \pm 16.8$	$106.9 \pm 19.4$	$101.2 \pm 17.4$	$103.1 \pm 17.9$	
€-Globin	101	$92.6 \pm 16.3$	$61.5 \pm 11.2$	$69.3 \pm 12.4$	$67.2 \pm 12.2$	
Globin mean	305	$96.2 \pm 10.1$	$85.2\pm8.5$	$89.2 \pm 9.0$	$91.2 \pm 9.4$	
$\alpha$ -Lactalbumin	80	$86.8 \pm 17.5$	$97.2 \pm 20.0$		$97.0 \pm 20.1$	
Mean	385	$94.3 \pm 8.7$	$87.5 \pm 7.8$		$92.1 \pm 8.0$	

Rodent genes are all from mouse except  $\alpha$ -lactalbumin from guinea pig. Ungulate genes are all from goat except  $\beta$ -globin from cow. Lagomorph genes are from rabbit and primate genes are from human.

#### **TABLE 2**

## Rates of nucleotide substitution at nonsynonymous sites between eutherian and marsupial genes

		Eutherian order					
Gene	Ν	Rodent	Ungulate	Lagomorph	Primate		
α-Globin	324	$17.5 \pm 2.6$	$13.9 \pm 2.2$	$17.9 \pm 2.6$	$13.8 \pm 2.2$		
$\beta$ -Globin	343	$23.3 \pm 3.0$	$20.7 \pm 2.9$	$20.3 \pm 2.7$	$20.8 \pm 2.8$		
€-Globin	344	$13.2 \pm 2.1$	$10.5 \pm 1.9$	$11.5 \pm 1.9$	$12.2 \pm 2.0$		
Globin mean	1012	$18.0 \pm 1.5$	$14.9 \pm 1.3$	$16.3 \pm 1.4$	$15.5 \pm 1.4$		
$\alpha$ -Lactalbumin	334	$42.6 \pm 4.7$	$38.5 \pm 4.3$		$36.4 \pm 4.1$		
Mean	1347	$23.3 \pm 1.5$	$20.1 \pm 1.4$		$20.1 \pm 1.4$		

**TABLE 3** 

Rates of nucleotide substitution in noncoding regions between eutherian and marsupial genes

			Eutheri	an order	
Gene	Ν	Rodent	Ungulate	Lagomorph	Primate
β-Globin	314	$98.3 \pm 11.2$	$93.4 \pm 10.3$	$101.9 \pm 11.8$	$91.4 \pm 9.7$
e-Globin	311	$82.4 \pm 8.8$	$77.3 \pm 8.1$	$74.0 \pm 7.5$	$80.3 \pm 8.3$
Mean	625	$89.9 \pm 6.9$	$84.9 \pm 6.4$	$85.8 \pm 6.5$	$85.6 \pm 6.3$

orders; all four orders have the highest and the lowest rate in different comparisons. There does not appear therefore to be any taxon-specific effect on rate such as cell-generation time or global mutation rate.

The magnitude of rate differences between taxa can be estimated approximately. Following LI and TANIMURA (1987), for a relative rate test the number of substitutions per site (K) between the reference sequence (3) and the point of divergence (0) of the two compared sequences (1 and 2) is obtained as:  $K_{03}$  $= (K_{13} + K_{23} - K_{12})/2$ . The numbers of substitutions per site in the lineages leading to the compared sequences are estimated as:  $K_{01} = K_{13} - K_{03}$  and  $K_{02} =$  $K_{23} - K_{03}$ . The ratios,  $K_{01}$ :  $K_{02}$  for the comparisons showing significant differences, range from 1.65 to 1.81. For other "total" comparisons (Tables 4-6) and for the different genic regions combined (data not shown) the ratios are closer to 1.0. Overall the degree of differences in rate between taxa appear to be substantially less than twofold. For this degree of difference the simulations of LI et al. (1987) indicate that

maximum parsimony is a more effective method of estimating phylogenies than is evolutionary parsimony.

**Phylogenetic relationships:** The single most parsimonious branching order of the functional genes of the  $\beta$ -globin complex in the four eutherian orders and in marsupials (obtained using DNAMETRO) is shown in Figure 2. The numbers at the branches in the tree indicate the percentage of bootstrap samples (of 100 samples) that gave the branch or cluster indicated. All branches are found in a high percentage of bootstrap samples except those leading to: (1) human  $\beta$  and rabbit  $\beta$ 1 (58%); (2) human G $\gamma$ , rabbit  $\beta$ 3, goat  $\eta$  and opossum  $\epsilon$  (33%); (3) goat  $\eta$  and opossum  $\epsilon$  (52%).

The topology of the tree is broadly consistent with that obtained previously by the same method (EAS-TEAL 1988) for the eutherian genes only. The only difference is that goat  $\eta$  ( $\epsilon$ 2) now appears to cluster with human G $\gamma$  and rabbit  $\beta$ 3, whereas previously it clustered with the other nonadult expressed genes. The two branches leading to goat (and opossum  $\eta$ ) are

## TABLE 4

	Gene						
	a-Globin	β-Globin	e-Globin	Globin total	α-Lactalbumin	Total	
N	104	105	101	310	80	390	
$K_{ru}$ $K_{rm} - K_{um}$ $K_{r0}:K_{l0}$	$66.9 \\ 5.6 \pm 21.1 \\ 1.18$	$56.6 - 16.9 \pm 18.5 0.54$	76.7 $31.1 \pm 16.5$ 2.36	67.1 11.0 ± 10.1 1.39	$ \begin{array}{r}     69.3 \\     -10.4 \pm 20.7 \\     0.85 \end{array} $	$66.8 \pm 8.9 \\ 1.23$	
$K_{\tau l} K_{\tau m} - K_{lm} K_{\tau 0}: K_{l0}$	74.0 -3.8 ± 22.6 0.90	$59.5 -11.2 \pm 17.7 0.68$	66.9 23.3 ± 15.9 2.07	66.4 7.0 ± 10.3 1.23			
$K_{\tau p} K_{\tau m} - K_{p m} K_{\tau 0}: K_{p 0}$	$79.8 - 12.2 \pm 26.5 0.73$	$\begin{array}{r} 48.7 \\ -13.1 \pm 16.7 \\ 0.58 \end{array}$	57.0 $25.4 \pm 15.1$ 2.61	$62.5 \\ 5.0 \pm 10.4 \\ 1.17$	$54.2 -10.2 \pm 19.1 0.68$	60.6 $2.2 \pm 8.5$ 1.07	
$K_{ul}$ $K_{um} - K_{lm}$ $K_{u0}:K_{l0}$	$35.1 - 9.4 \pm 15.4 0.58$	37.8 5.7 ± 14.6 1.36	$56.3 - 7.8 \pm 12.8 0.76$	$ \begin{array}{r} 41.8 \\ -4.0 \pm 7.7 \\ 0.82 \end{array} $			
$K_{up} K_{um} - K_{pm} K_{u0}:K_{p0}$	$38.5 - 17.8 \pm 20.1 0.37$	36.6 $3.8 \pm 14.7$ 1.23	$46.0 -5.7 \pm 12.0 0.78$	$40.0 - 6.0 \pm 8.0 0.74$	34.1 $0.2 \pm 16.7$ 1.01	$39.4 -4.6 \pm 6.6 0.94$	
$K_{lp} K_{lm} - K_{pm} K_{l0}:K_{p0}$	31.5 0.0 ± 23.6 0	$35.6 - 1.9 \pm 13.4 0.90$	55.4 2.1 ± 13.2 1.08	40.7 -2.0 ± 8.2 0.91			

# Comparisons of the numbers of synonymous substitutions per 100 nucleotides between orthologous eutherian genes by the relative rate test using marsupial genes for reference

 $K_{ij}$  is the substitution rate between orders *i* and *j*; *i*, *j* = *m*, *r*, *u*, *p* and *l* where *m* = marsupial, *r* = rodent, *u* = ungulate, *p* = primate, *l* = lagomorph.  $K_{i0}:K_{j0}$  is the ratio of substitution rates between orders *i* and *j* since their divergence (at point 0) estimated as described in the text.

## TABLE 5

## Comparisons of the numbers of nonsynonymous substitutions per 100 nucleotides between orthologous eutherian genes by the relative rate test using marsupial genes for reference

	Gene						
	α-Globin	β-Globin	e-Globin	Globin total	α-Lactalbumin	Total	
Ν	320	337	345	1002	337	1339	
$K_{ru} K_{rm} - K_{um} K_{r0}:K_{l0}$	9.4 3.6 ± 1.9 2.24	15.7 $2.6 \pm 2.7$ 1.40	7.4 2.7 ± 1.7 2.15	10.7 $3.1 \pm 1.2*$ 1.81	$19.6 \\ 4.1 \pm 3.9 \\ 1.52$	13.1 3.2 ± 1.3* 1.65	
$K_{rl} \\ K_{rm} - K_{lm} \\ K_{r0}:K_{l0}$	$   \begin{array}{r}     11.1 \\     -0.4 \pm 2.2 \\     0.93   \end{array} $	12.8 $2.5 \pm 2.5$ 1.48	10.4 $1.7 \pm 1.9$ 1.11	11.4 $1.7 \pm 1.3$ 1.35			
$K_{rp} \\ K_{rm} - K_{pm} \\ K_{r0}:K_{p0}$	8.7 3.7 ± 1.9 2.49	12.5 $3.0 \pm 2.3$ 1.63	8.8 1.0 ± 1.8 1.26	9.9 2.5 ± 1.3 1.49	$17.0 \\ 6.2 \pm 3.6 \\ 2.15$	11.6 3.2 ± 1.2* 1.76	
$K_{ul} \\ K_{um} - K_{lm} \\ K_{u0}:K_{l0}$	$10.3 -4.0 \pm 2.0 0.44$	$9.2 \\ 0.1 \pm 2.1 \\ 1.02$	$7.6 -1.0 \pm 1.7 0.77$	9.1 $-1.4 \pm 1.1$ 0.73			
$K_{up} \\ K_{um} - K_{pm} \\ K_{u0}:K_{p0} \\ K_{lp} \\ K_{lm} - K_{pm} \\ K_{l0}:K_{p0}$	$8.6 \\ 0.1 \pm 1.8 \\ 1.02 \\ 11.3 \\ 4.1 \pm 2.1 \\ 2.14$	$8.3 \\ 0.4 \pm 1.9 \\ 1.10 \\ 5.2 \\ 0.5 \pm 1.5 \\ 1.21$	$4.9 \\ -1.7 \pm 1.4 \\ 0.49 \\ 7.7 \\ -0.7 \pm 1.6 \\ 8.33$	7.1 -0.6 $\pm$ 1.0 0.84 8.0 0.8 $\pm$ 0.1 1.22	15.2 2.1 ± 3.3 1.32	9.0 0±1.1 1.0	

Abbreviations are as in Table 4. \* P < 0.05.

the least reliable of all those in the tree. The clustering of goat  $\eta$  with opossum  $\epsilon$  and of these genes with human G $\gamma$  and rabbit  $\beta$ 3 is therefore uncertain.

An important feature of the tree is that the eutherian branching order indicated by the adult-expressed genes, shown branching to the left, and the

TABLE 6

Comparisons of the numbers of nucleotide changes per 100 sites in noncoding regions between orthologous eutherian genes by the relative rate test using marsupial genes for reference

		Gene	<u> </u>
	β-Globin	€-Globin	Total
N	314	311	625
K <sub>ru</sub> K <sub>rm</sub> -K <sub>um</sub> K <sub>r0</sub> :K <sub>u0</sub>	52.4 $4.9 \pm 11.7$ 1.21	$69.1 \\ 5.1 \pm 9.7 \\ 1.16$	$60.0 \\ 5.0 \pm 7.4 \\ 1.18$
$egin{array}{l} K_{ au l} \ K_{ au m} - K_{ au m} \ K_{ au 0} \end{array}$	44.8 -3.6 ± 11.5 0.85	73.0 $8.5 \pm 9.6$ 1.26	62.2 $4.1 \pm 7.4$ 1.14
$K_{rp}$ $K_{rm} - K_{pm}$ $K_{r0}$ : $K_{p0}$	$44.8 \\ 6.9 \pm 10.8 \\ 1.36$	80.7 2.1 ± 10.1 1.05	59.8 $4.3 \pm 7.2$ 1.15
$K_{ul}$ $K_{um} - K_{im}$ $K_{u0} - K_{l0}$	$35.1 - 8.5 \pm 10.4 0.61$	36.1 $3.3 \pm 7.2$ 1.20	$35.9 - 0.9 \pm 6.0 0.95$
$egin{array}{l} K_{up} \ K_{um} - K_{pm} \ K_{u0} : K_{p0} \end{array}$	30.3 2.0 ± 9.2 1.14	44.7 -3.0 ± 8.2 0.87	$36.9 -0.7 \pm 6.0 0.96$
$K_{lp} \\ K_{lm} - K_{pm} \\ K_{l0}: K_{p0}$	31.4 $10.5 \pm 9.6$ 2.0	34.0 -6.3 ± 7.3 0.69	33.0 $0.2 \pm 5.8$ 1.01

Abbreviations are as in Table 4.

embryonic-expressed genes, shown branching to the top right, is the same. The order is rodent then ungulate then lagomorph and primate.

The tree was constructed by comparison of coding and noncoding sequences. The alignment of the noncoding sequences (Figure 1) was made using an iterative multiway method. While this is a more rational approach than the intuitive combining of two-way alignments, the alignments produced are not necessarily correct as they depend on an arbitrary choice of gap penalties. The noncoding sequences were included in the analysis to maximize the length of the compared sequences. However, because of alignment problems, their inclusion does introduce some uncertainty about the results. For this reason the coding and noncoding sequences were analysed separately.

The maximum likelihood trees, which have the same topologies as the corresponding maximum parsimony trees (using DNAPENNY) for the  $\beta$ - and  $\alpha$ globin genes and for the coding regions of three classes of genes combined (Figure 3) are the same as those in Figure 2. The proportions of bootstrap samples indicating this branching order for  $\beta$ -globin and total coding regions are approximately the same as the coding and noncoding regions combined (Figure 2). For the  $\alpha$ -globin coding regions the branching order is indicated by a smaller proportion of bootstrap samples (59% indicate a primary branching of rodents and 46% indicate a lagomorph-primate monophyletic group). A different branching order is indicated for the  $\epsilon$ -globin gene coding regions with the ungulates and primates forming a monophyletic group separate from the rodents and lagomorphs which form a separate group. Both of these groups however are indicated by a minority of bootstrap samples. Their validity is thus highly questionable and, rather than indicating an alternative branching order, analysis of the globin coding region alone would appear to be insufficient to resolve the branching order.

The maximum likelihood trees for the  $\beta$ - and  $\epsilon$ globin noncoding regions (Figure 4) both have topologies consistent with Figure 2. In both cases, a primary branching of rodents is indicated in a high proportion of bootstraps (100% and 95% for  $\epsilon$ -globin and  $\beta$ globin, respectively). Primates and lagomorphs cluster in 84% of bootstrap samples of  $\beta$ -globin noncoding regions but in only a low proportion of bootstrap samples of  $\beta$ -globin noncoding regions; the proportion is not shown in Figure 4 because an equally parsimonious topology exists that groups ungulates with primates, and which was indicated by a greater proportion of bootstrap samples (36%).

The rodent/ungulate/primate and lagomorph branching order is thus clearly indicated by one ( $\beta$ globin) of three coding regions as well as the combined coding regions and by one ( $\epsilon$ -globin) of two noncoding regions. The  $\alpha$ -globin coding region also indicates this topology but with lower probability, and the  $\beta$ -globin noncoding region indicates this topology when analyzed by maximum likelihood and indicates that it is one of two most parsimonious topologies. Only one region ( $\epsilon$ -globin coding region) positively identifies an alternative topology with both maximum likelihood and maximum parsimony methods, however the validity of this alternative topology has low probability as indicated by bootstrap sampling. The branching order is thus indicated by different genes and by both coding and noncoding regions. It does not appear to be an artifact arising from analysis of particular genes or from the incorrect alignment of noncoding sequences.

The topologies of the maximum parsimony and maximum likelihood trees based on the combined coding and noncoding regions are the same (Figure 5). A primary branching of rodents and a clustering of lagomorphs and primates are indicated in 100% and 86% of bootstrap samples respectively. The nodes of both trees are widely separated. The maximum parsimony and likelihood methods, respectively, indicate that in the ungulate, lagomorph and primate lineages, 43% and 37% of the average genetic change since their divergence from rodents occurred before the divergence of ungulates. Similarly the two methods respectively indicate that 34% and 22% of the average genetic change in the lagomorph and primate lineages, arising since their divergence from ungulates, occurred before they split (Figure 5).

A number of branching orders different from the



FIGURE 3.—Maximum likelihood cladograms for four eutherian orders ( $\mathbf{R}$  = rodent, U = ungulate, L = lagomorph, P = primate) rooted by marsupial (M) outgroup based on the *coding regions* of individual and combined globin genes. The percentage of bootstrap samples showing the branches in the trees are indicated. The scale indicates the expected number of substitutions per site.





FIGURE 4.—Maximum likelihood cladograms for four eutherian orders rooted by marsupial outgroup based on the *noncoding regions* of  $\epsilon$  and  $\beta$ -globin genes. The percentage of bootstrap samples showing the branches in the trees are indicated. Abbreviations are as in Figure 3. The scale indicates the expected number of substitutions per site.

one obtained here have been proposed. Three of these are compared in Table 7. Tree 1 is that obtained in the present study. Tree 2 is that proposed by KOOP and GOODMAN (1988) in which ungulates diverge first,

FIGURE 2.—Single most parsimonious unrooted cladogram of mammalian  $\beta$ -like globin genes (obtained using DNAMETRO). Coding sequences, and noncoding sequences shown in Figure 1 were analyzed. the percentage of (100) bootstrap samples showing the branches in the tree are indicated.

followed by primates then rodents and lagomorphs. Tree 2 also indicates that primates and rodents are more closely related to each other than either is to ungulates, and thus that ungulates are an appropriate reference species for a relative rate test of rodents and primates as suggested by WU and LI (1985). Trees 3 and 4 are two other trees indicating a relatively close relationship between rodents and primates. Tree 3 has a branching order of ungulate then lagomorph then rodent and primate. In tree 4 there are two separate monophyletic groups, one consisting of ungulates and lagomorphs and the other of primates and rodents.

The number of nucleotide changes occurring in each of the trees was determined using DNAPARS with the topologies defined. For each of the three globin genes, for the combined coding regions and the combined noncoding regions, and for all regions combined, the numbers of nucleotide changes occurring in trees 2, 3 and 4 are similar and substantially greater than in tree 1. Trees 2, 3 and 4 do not therefore appear to be probable alternatives to tree 1.

## DISCUSSION

An attempt has been made here to investigate the phylogenetic relations of four orders of eutherian mammals, using maximum parsimony, likelihood and compatibility methods. The appropriateness of these methods depends on there being approximate uniformity of evolutionary rate among the orders. A lack of rate variation has been demonstrated by relative rate comparisons of the eutherian genes with their marsupial orthologues. The phylogenetic analysis indicates that the orders are related in a dichotomously rather than a polychotomously branching pattern and that their branching points are widely separated; i.e., that they do not form a star phylogeny. The observation (KIMURA 1983; GILLESPIE 1984, 1986a, b) that there is greater variance in evolutionary rate among eutherian orders than is predicted by neutral theory depends on the assumption of a star phylogeny. Since this assumption is not valid the increased variance cannot be taken as evidence of evolutionary rate variation (resulting either from natural selection or other S. Easteal



Г	A	BI	Æ	7
---	---	----	---	---

Numbers of nucleotide changes for phylogenetic trees with different topologies

			Tr	Tree <sup>a</sup>	
Sequence	Length	1	2	3	4
α-Globin	429	239	244	247	247
$\beta$ -Globin	841	706	728	724	730
€-Globin	841	689	735	729	730
Coding	1317	697	715	717	717
Noncoding	794	937	992	983	990
Total	2111	1634	1707	1700	1707

<sup>a</sup> Tree topologies are as follows:

172

1. (Marsupial, (((Lagomorph, Primate), Ungulate), Rodent)).

2. (Marsupial, (((Lagomorph, Rodent), Primate), Ungulate)).

3. (Marsupial, (((Primate, Rodent), Lagomorph), Ungulate)).

4. (Marsupial, ((Primate, Rodent), (Lagomorph, Ungulate)).

factors), and may be better explained as reflecting variation in divergence time.

Earlier studies (OHTA and KIMURA 1971; LANGLEY and FITCH 1974) also reported greater variance in the rate of mammalian protein evolution than expected from neutral theory, although different approaches were used in arriving at this conclusion. OHTA and KIMURA's (1971) approach relied on assumptions about the geological divergence times of species derived from fossil record interpretation which may not be reliable. For instance they assumed a polychotomous divergence of eutherian orders (occurring 80 mya) which, as has been shown here, is an incorrect assumption. Their other assumed divergence times may be similarly incorrect.

LANGLEY and FITCH's (1974) analysis is based on an assumed eutherian phylogeny that has been shown here to be incorrect. Their approach to testing for evolutionary rate variation should be more sensitive than the relative rate test. However, although both methods involve assumptions about phylogeny, the only assumption made here in applying the relative rate test has been that eutherians comprise a monophyletic group when compared with marsupials. On the other hand, LANGLEY and FITCH's (1974) method requires a complete knowledge of the phylogeny of all compared species. It would be inappropriate to use the phylogeny derived from the DNA sequence data FIGURE 5.—Maximum parsimony (A) and maximum likelihood (B) cladograms for four eutherian orders rooted by marsupial outgroup based on the coding regions of  $\alpha$ -,  $\beta$ - and  $\epsilon$ -globin genes and the noncoding regions of  $\beta$ - and  $\epsilon$ -globin genes combined. The percentage of bootstrap samples showing the branches of the maximum parsimony tree are indicated. Abbreviations are as in Figure 3. The scale for the maximum likelihood tree indicates the expected number of substitutions per site.

as a basis for testing for rate variation for the same data using the LANGLEY and FITCH approach.

The branching order indicated by the cladistic analysis (rodent then ungulate then primate and lagomorph) is quite different from that indicated by KOOP and GOODMAN's (1988) analysis of many of the same genes studied here (ungulate then primate then rodent and lagomorph), and from that assumed by WU and LI (1985) in their investigation of evolutionary rate variation (ungulate then primate and rodent).

The difference with KOOP and GOODMAN's (1988) phylogeny may result from their use of "local-branchswapping" algorithm (GOODMAN et al. 1979) which is not guaranteed to find the most parsimonious tree. The branch-and-bound algorithm used here (DNA-PENNY) is guaranteed to do so. In neither of the other studies was any attempt made to assess the reliability of the trees obtained. In the present study the appropriateness of the use of maximum parsimony, compatibility and likelihood methods was established by a lack of evidence of evolutionary rate variation among taxa. This was assessed independently of the cladistic analysis. The tree obtained here is shown to be reliable by its being indicated from analysis of different genes and of both coding and noncoding regions, and by a high proportion of bootstrap samples. It is also a substantially more parsimonious tree than the alternatives discussed above.

The results reported here emphasize the importance of accurate phylogenetic information in investigations of evolutionary processes and demonstrate how this information can be reliably obtained from DNA sequence data. Previous comparative studies of molecular data from different eutherian orders (LANGLEY and FITCH 1974; KIMURA 1983; GILLESPIE 1984, 1986a, b; WU and LI 1985) have concluded that the rate of molecular evolution is variable among lineages. These conclusions depend on phylogenetic assumptions that appear from the present analysis to be incorrect. The phylogeny presented here is consistent with the results of the relative rate tests which showed that overall there is stochastic uniformity of molecular evolutionary rate among the eutherian orders.

the Tammar wallaby  $\alpha$ -lactalbumin gene before publication and to J. H. GILLESPIE and D. PENNY for their valuable comments. I also thank D. SMITH, W.-H. LI and J. FELSENSTEIN for providing their computer programs for sequence alignment, estimation of rates of synonymous and nonsynonymous substitutions and phylogenetic reconstruction respectively.

## LITERATURE CITED

- CLEMENS, W. A., J. A. LILLEGRAVEN, E. H. LINDSAY and G. G. SIMPSON, 1979 Where, when and what-a survey of known Mesozoic mammal distribution. pp. 7-58 in *Mesozoic Mammals*, edited by J. A. LILLEGRAVEN, Z. KIELAN-JAWOROWSKA and W. A. CLEMENS. University of California Press, Berkeley.
- CZELUSNIAK, J., M. GOODMAN, D. HEWETT-EMMETT, M. L. WEISS, P. J. VENTA and R. E. TASHIAN, 1982 Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature 298: 297–300.
- EASTEAL, S., 1988 Rate constancy of globin gene evolution in placental mammals. Proc. Natl. Acad. Sci. USA 85: 7622– 7626.
- FELSENSTEIN, J., 1978 Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27: 401-410.
- FELSENSTEIN, J., 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17: 368–376.
- FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.
- FENG, D.-F., and R. F. DOOLITTLE, 1987 Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25: 351-360.
- FITCH, W. M., 1971 Towards defining the course of evolution: minimum change for a specified tree topology. Syst. Zool. 20: 406-416.
- GILLESPIE, J. H., 1984 The molecular clock may be an episodic clock. Proc. Natl. Acad. Sci. USA 81: 8009-8013.
- GILLESPIE, J. H., 1986a Natural selection and the molecular clock. Mol. Biol. Evol. 3: 138–155.
- GILLESPIE, J. H., 1986b Variability of evolutionary rates of DNA. Genetics 113: 1077–1091.
- GOODMAN, M., J. CZELUSNIAK, G. W. MOORE, A. E. ROMERO-HERRERA and G. MATSUDA, 1979 Fitting the gene lineage into its species lineage. A parsimony strategy illustrated by cladograms constructed from globin sequences. Syst. Zool. 28: 132-163.
- GOODMAN, M., A. E. ROMERO-HERRERA, H. DENE, J. CZELUSNIAK and R. E. TASHIAN, 1982 Amino acid sequence evidence on the phylogeny of primates and other eutherians, pp. 115-191 in Macromolecular Sequences in Systematic and Evolutionary Biology, edited by M. GOODMAN. PLENUM, NEW YORK.
- HALL, I. L., R. K. CRAIG, M. R. EDBROOKE and P. N. CAMPBELL, 1982 Comparison of the nucleotide sequence of cloned human and guinea-pig pre-alpha-lactalbumin cDNA with that of chicken pre-lysozyne cDNA suggests evolution from a common ancestral gene. Nucleic Acids Res. 10: 3503–3515.
- HENDY, M. D., and D. PENNY, 1982 Branch and bound algorithms to determine minimal evolutionary trees. Math. Biosci. 59: 277-290.
- KIELAN-JAWOROWSKA, A., T. M. BROWN and J. A. LILLEGRAVEN, 1979 Eutheria, pp. 221-258 in *Mesozoic Mammals*, edited by

J. A. LILLEGRAVEN, Z. KIELAN-JAWOROWSKA and W. A. CLE-MENS. University of California Press, Berkeley.

- KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature 217: 624–626.
- KIMURA, M., 1983 The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.
- KIMURA, M., 1987 Molecular evolutionary clock and the neutral theory. J. Mol. Evol. 26: 24-33.
- KOOP, B. F., and M. GOODMAN, 1988 Evolutionary and developmental aspects of two hemoglobin  $\beta$ -chain genes ( $\epsilon^m$  and  $\beta^m$ ) of opossum. Proc. Natl. Acad. Sci. USA **85:** 3893–3897.
- KUMAGAI, I., Ε. ΤΑΜΑΚΙ, S. KAKINUMA and K. MIURA, 1987 Molecular cloning and sequencing of cDNA encoding goat pre α-lactalbumin. J. Biochem. 101: 511–517.
- LAKE, J. A., 1987 A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. Mol. Biol. Evol. 26: 59-73.
- LANGLEY, C. H., and W. M. FITCH, 1974 An examination of the constancy of the rate of molecular evolution. J. Mol. Evol. 3: 161–177.
- LI, W.-H., and M. TANIMURA, 1987 The molecular clock runs more slowly in man than in apes and monkeys. Nature **326**: 93-96.
- LI, W.-H., C.-I. WU and C. C. Luo, 1985 A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2: 150–174.
- LI, W.-H., K. H. WOLFE, J. SOURDIS and P. M. SHARP, 1987 Reconstruction of phylogenetic trees and estimation of divergence times under nonconstant rates of evolution. Cold Spring Harb. Symp. Quant. Biol. 52: 847–856.
- METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER and E. TELLER, 1953 Equation of state calculations by fast computing machines. J. Chem. Phys. 21: 1087–1092.
- NOVACEK, M. J., 1982 Information for molecular studies from anatomical and fossil evidence on higher eutherian phylogeny, pp. 3-41 in Macromolecular Sequences in Systematics and Evolutionary Biology, edited by M. GOODMAN. Plenum, New York.
- OHTA, T., and M. KIMURA, 1971 On the constancy of the evolutionary rate of cistrons. J. Mol. Evol. 1: 18-25.
- PENNY, D., and M. D. HENDY, 1985 Estimating the reliability of evolutionary trees. Mol. Biol. Evol. 3: 403-417.
- PENNY, D., M. D. HENDY and I. M. HENDERSON, 1987 Reliability of evolutionary trees. Cold Spring Harbor Symp. Quant. Biol. 52: 847-856.
- SARICH, V. M., and A. C. WILSON, 1967 Immunological time scale for hominid evolution. Science 158: 1200–1203.
- TAKAHATA, N., 1987 On the over-dispersed molecular clock. Genetics 116: 169–179.
- WAINWRIGHT, B., and R. HOPE, 1985 Cloning and chromosomal location of the  $\alpha$  and  $\beta$ -globin genes from a marsupial. Proc. Natl. Acad. Sci. USA 82: 8105–8108.
- WU, C.-I., and W.-H. LI, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 82: 1741-1745.
- WYSS, A. R., M. J. NOVACEK and M. C. MCKENNA, 1987 Amino acid sequence versus morphological data and the interordinal relationships of mammals. Mol. Biol. Evol. 4: 99–116.

Communicating editor: A. H. D. BROWN