A Molecular View of Primate Phylogeny and Important Systematic and Evolutionary Questions [1]

Ben F. Koop,[*,2] *Danilo A. Tagle,*[*] *Morris Goodman,*[*,†] *and Jerry L. Slightom*[†‡]

Departments of *Molecular Biology and Genetics and †Anatomy and Cell Biology, Wayne State University; and ‡Division of Molecular Biology, The Upjohn Company

Phylogenetic analysis of extensive nucleotide sequence data from primate β-globin gene clusters elucidates the systematics and evolution of the order Primates and reveals that rates of accumulation of mutations vary by as much as a factor of seven among different primate lineages. The picture of primate phylogeny from DNA sequences clarifies many ambiguities of the morphological picture. In the molecular picture, dwarf and brown lemurs group together into superfamily Lemuroidea, Lemuroidea and Lorisoidea into suborder Strepsirhini, and *Tarsius* and Anthropoidea into suborder Haplorhini. The molecular picture also privides both significant evidence for a human-chimpanzee clade that narrowly excludes gorilla and overwhelming evidence for the gorilla-chimpanzee-human clade within Hominoidea. Rates of DNA sequence evolution appear to have been fastest in the early primates ancestral to Anthropoidea and next fastest on the lorisoid branch. Rates were slowest over the past 25 Myr of hominoid descent, suggesting that mechanisms lowering the mutation rate evolved in correlation with lengthened life spans.

Introduction

The order Primates comprises six major monophyletic groups, Lemuroidea (Malagasy lemurs), Lorisoidea (Asian lorises and African galagos and pottos), Tarsioidea (Indonesian and Philippine tarsiers), Ceboidea (new-world monkeys), Cercopithecoidea (old-world monkeys), and Hominoidea (apes and humans). The first three groups are collectively referred to as *prosimian primates,* and the last three are termed *simian primates.* The latter are characterized by such derived primate features as larger brains and prolonged fetal stages of life. There is virtually unanimous agreement that the three simian groups form a larger monophyletic assemblage (Anthropoidea) which excludes the prosimians and subdivides into platyrrhines (ceboids) and catarrhines (cercopithecoids and hominoids). There is also agreement, although not as unanimous, that lemuroids and lorisoids are more closely related to each other than to Anthropoidea. Tarsiers, however, have been alternatively depicted as the closest relative (sister group) of lorisoids (Schwartz 1984, 1986; Schwartz and Tattersall 1985), of lemuroids and lorisoids (Simpson 1945; Napier and Napier 1967), of Anthropoidea (Pocock 1918; Baba et al. 1975, 1982; Luckett and Szalay 1978; Bonner et al. 1980; Delson and Rosenberger 1980; MacPhee and Cartmill 1986; Miyamoto and Goodman 1986; Szalay et al. 1987), of an assemblage grouping Anthropoidea and a lemuroid-

1. Key words: primate phylogeny, beta hemoglobin, molecular evolution, evolutionary rates.

2. Present address: Division of Biology, California Institute of Technology.

Address for correspondence and reprints: Dr. Ben F. Koop, Division of Biology, 147-75, California Institute of Technology, Pasadena, California, 91125.

Mol. Biol. Evol. 6(6):580–612. 1989.

0737-4038/89/0606-0002$02.00

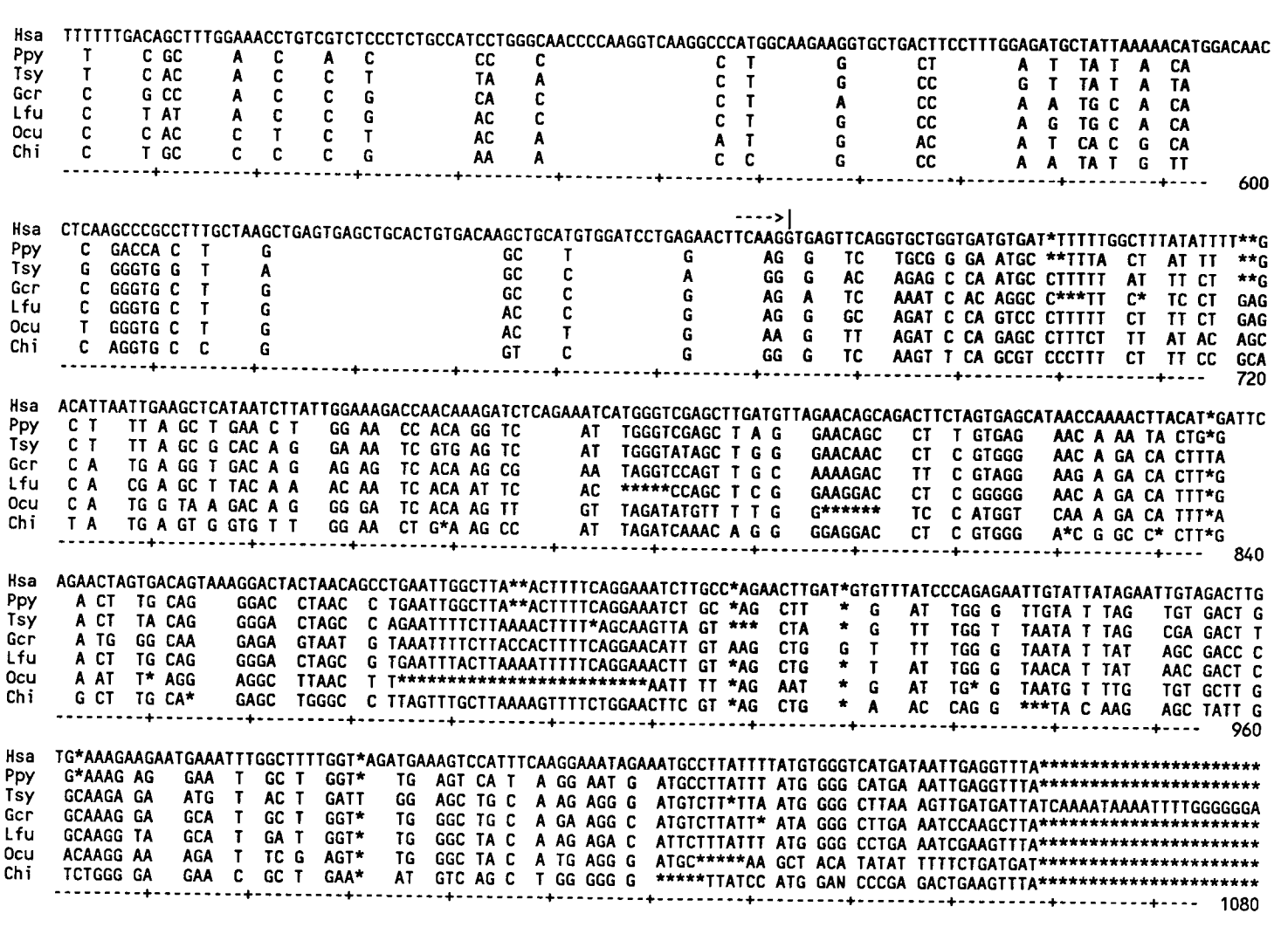
lorisoid group (Gingerich 1975; Schwartz, 1978), and of humans (Jones 1920). To obtain evidence on the phylogenetic position of tarsiers and thus establish which primate taxon is the sister group of Anthropoidea, we decided to gather a large body of comparative nucleotide sequence data on the major branches of primates and to extract the phylogenetic information contained in the DNA sequences. Toward this end we have sequenced the five β-type globin genes (ε, γ, ψη, δ, and β) of *Tarsius syrichta* (the Philippine tarsier) and have compared by cladistic methods these DNA sequences to orthologues (the closest matching sequences) from galago, lemurs, simians, and nonprimate mammals.

In *T. syrichta,* as in all other primate and placental mammals so far examined, the genes of the β-globin family are found clustered in a 40–60-kb chromosomal region (Collins and Weissman 1984; Hardies et al. 1984; Hardison 1984; Hill et al. 1984). Comparisons of >90 β-type globin gene sequences and of linkage maps showing the locations of these genes within the β clusters have provided evidence (1) that an initial duplication in a stem species of Mammalia gave rise in a common ancestor of Metatheria (marsupials) and Eutheria (placentals) to embryonically expressed proto ε and adult-expressed proto β (Koop and Goodman 1988), (2) that proto ε was the ancestor of ε-, γ-, and η-globin genes and that proto β was the ancestor of δ- and β-globin genes, and (3) that ε-, γ-, η-, δ-, and β-gene lineages existed prior to the separation of eutherian orders Artiodactyla, Rodentia, Lagomorpha, and Primates (Czelusniak et al. 1982; Goodman et al. 1984, 1987; Harris et al. 1984, 1986; Koop and Goodman 1988). Evolutionary reconstructions demonstrate that the η locus was silenced in stem primates (Jeffries et al. 1982; Harris et al. 1984; Koop et al. 1986*a*) and that the developmental expression of the γ locus in the lineage leading to the common ancestor of simian primates was delayed from embryonic to fetal life (Goodman et al. 1987; Tagle et al. 1988). Although lagomorphs and rodents lost the η locus and artiodactyls lost the γ locus, primates retained all five loci. The tarsier linkage map (Koop et al. 1989) has the same gene arrangement (5′-ε-γ-ψη-δ-β3′) as do the lorisoid (galago) (Tagle et al. 1988) and ceboid (owl monkey) (Harris et al. 1984) linkage maps. In lemurs, however, a deletion fused the 5′ sequence of ψη with the 3′ sequence of δ (Jeffries et al. 1982), and in simians the fetal γ locus duplicated in the stem-catarrhines to yield the cluster 5′-ε-γ^1-γ^2-ψη-δ-β-3′ found in cercopithecoids and hominoids (Slightom et al. 1980; Barrie et al. 1981). In addition, retropositioning of repetitive sequences such as Alu family members and insertions and deletions of smaller DNA pieces have occurred in introns and in flanking and intergenic DNA (i.e., in the noncoding regions) of the primate β-globin gene clusters during evolution (Sawada et al. 1985; Koop et al. 1986*b*). Otherwise, the various surviving primates share, as a primitive eutherian feature, the same linkage arrangement for the five β-type globin genes.

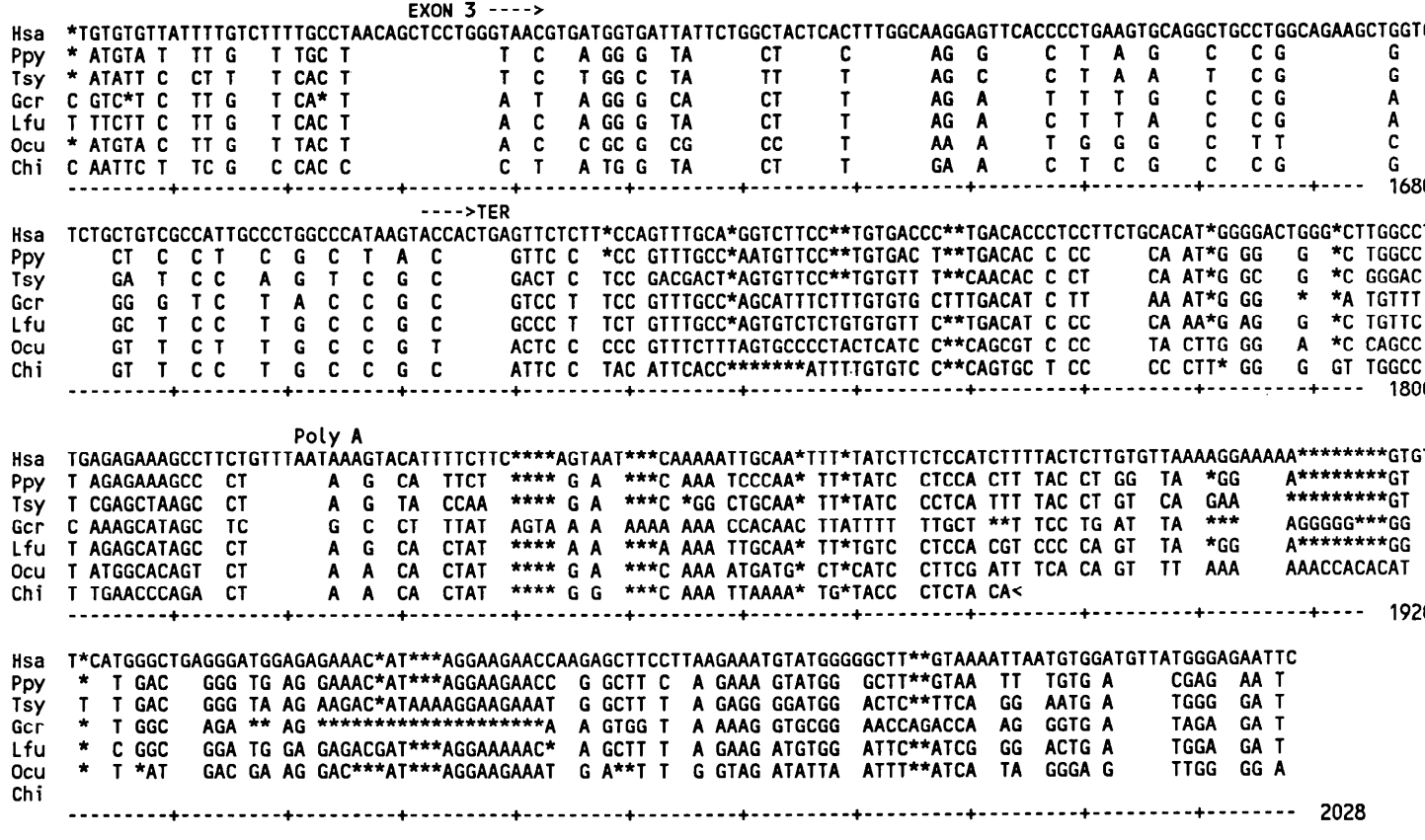
Only the conserved genic sequences from coding and 5′ promoter (regulatory) regions clearly reveal the common ancestry of ε-, γ-, η-, δ-, and β-globin genes, i.e., that these homologous genes are paralogues (descendants of duplicated genes). As the five paralogous loci had already originated and diverged from each other before the separation of eutherian orders and primate lineages but have retained their relative positions in the β-globin gene clusters of the descendant species, they serve as guide posts for aligning orthologous nucleotide sequences from the nongenic noncoding portions of the clusters as well as from the genic portions. Homologous sequences are orthologous if their ancestral separations coincide with the ancestral separations of the species in which they occur. Results on the β-globin gene clusters, both those reported elsewhere and those described below, indicate that orthologous alignments

ALIGNED EPSILON SEQUENCES

```
                                                                         CACCCC                       CCAAT
Hsa  CACATTATCACAAACTTAGTGTCCATCCATCACTGCTGACCCTCTCCGGACCTGACTCCACCCCTGA*GGA*CACAGGTCAGCCTTGACCAATGACTTTTAAGTACCATGGAGAACAGGG
Ppy  CACA      AC AACTTAG GTC ATCCATCACTG       C  TCCA AC TGA         C   A  A*  C GGTC GCC              TT  AT CCATG A  A  G
Gcr  CACA      TC AT****G GTG CTACTTCACTG       C  TGCT AC AGG         C   *  GA  G GCTT GCT              CT  CT CCACG A  A  A
Lfu  CACA      AC AT****G TTG ATACTTCACTG       C  TCCT AC TGG         C   *  GA  G GCTC TCC              CT  CT ACACA A  G  A
Ocu  GCCA      AC A*CTCAG GAC A*******CCA       A  TGGT ** TTA         *   *  A*  C *TCC GTC              **  GT *TGAA A  A  A
Chi  CAGG      AT CACTCAA ATA ATC**TCCTTG       A  AGCT AC TGA         C   *  GA  C GCCT ACC              CA  GG CAAGG G  G  A
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+----  120

                         AATAAAA                                                                                     INIEXO
Hsa  GGCCAGAACTTCGGCAGTAAAGAATAAAAGGCCAGACAGAGAGGCAGCAGCACATATCTGCTTCCGACACAGCTGCAATC****ACTAGCAAGCTCTC****AGGCCTGGCATCATGGTG
Ppy    CC   ACT  G      GAA A          CACAGAGAGG       C T TC GC   C ACA AGC  CA   ****A CA C  AC CTC****A G C  G  T
Tsy                                                                             >   ****A TA C  GT GCC****A A C  A  T
Gcr    CT   ACT  A      GGA G          CATAGAAAAT       T C CC AT   T GTA AGC  TG   ****A CA C  GC CCC****A A T  A  C
Lfu    CC   ACA  A      GGA A          CACAGAGAAG       C C CC GC   T ACA AGC  TG   ****G CA C  GC CCCGCCCA A T  A  C
Ocu    CC   CCA  A      GGG A          AGCCTTGAAG       C A GC GC   T ACA ATT  TG   GATCA CA C  GC CCC****A A G  A  C
Chi    GC   AGT  A      AAA A          CAGCATCCAG       C G CT GC   T ATG TTC  TG   ****A CT T  GC CCA****C A T  A  T
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+----  240

     N 1 ---->                                                                    ---->|
Hsa  CATTTTACTGCTGAGGAGAAGGCTGCCGTCACTAGCCTGTGGAGCAAGATGAATGTGGAAGAGGCTGGAGGTGAAGCCTTGGGCAGGTAAGCATTGGTTCTCAATGCATGGGAATGAAGG
Ppy    TT       G T  G  G   GCTGCCG CACTAGC T    A CA GA G  TG G  A  GG       T  A  CT            AGC TT GTTCTCAA G A  GGAATGA G
Tsy    TC       G T  A  A   TCTTCCG TACTAGC T    G TA GA G  TG G  T  AG       T  A  TT            AGC TC GTTCTCAA A A  AGAATGA G
Gcr    TT       G T  G  A   GCTATTA CATGAGC T    G AA GG G  TA A  A  GG       A  A  CT            AAC CT GTTCTCAG G A  GGAATAA G
Lfu    TT       G T  G  G   TCCACCA CCTGAGC T    G CA GG G  TG G  A  GG       C  A  CT            AAC TT TTTCTCGA G A  GGAATAA G
Ocu    CT       C T  G  A   TGCATTA CAGTAAG A    G CC GA G  CA C  T  GA       T  A  TT            AGT TT GTTCTCAA A T  GTAGAGA G
Chi    TT       G C  G  G   GCTGCTA CACTGGC T    G CA AG C  TG G  A  GG       C  G  TC            GAA GT G*****AC T A  GGGGAGG T
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+----  360

                                                                                                  EXON 2 ---->
Hsa  GTGAATATTACCCTAGCAAGTTGATTGGGAAAGTCCTCAAGATT*TTTTGCATCTCTAATTTTGTATCTGATATGGTGTCATTTCATAGACTCCTCGTTGTTTACCCCTGGACCCAGAGA
Ppy  GTG    GT CC   A C AGTT ATTGGG A G C   CA GATT* TTT AA CTC A     GTATC  A A GG GT A TTC T  G      C     T  C         C  G  A
Tsy  GTA    TG TG   G C AGTT ACCCAG T A C   TA AATTT TGA TA CTT A     TTGTC  C A GG GT A CTT C  G      C     T  T         C  G  A
Gcr  GGG    TA CT   G C AACT ACCAGG A G C   AA GATT* TGA CA CAC A     CCACC  T A GG GA G ATC T  G      T     C  C         C  G  A
Lfu  AGG    TA CC   G C AATT ACCAGG A G C   TA GATT* TGA CA CTC A     CCATC  C A GG GC A CTC C  G      C     C  C         A  G  A
Ocu  GCA    TG CC   G T GACT ACCAGG A T G   CA AGTT* TT* TA CTC G     CTATT  C G TG CC A CAT T  G      T     C  C         C  A  A
Chi  GTG    TG GC   G C AATC GCCAGA A A T   TC AAAA* CT* AG TGC G     CCATC  C A GT TC A CTC T  G      G     C  C         C  G  G
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+----  480
```

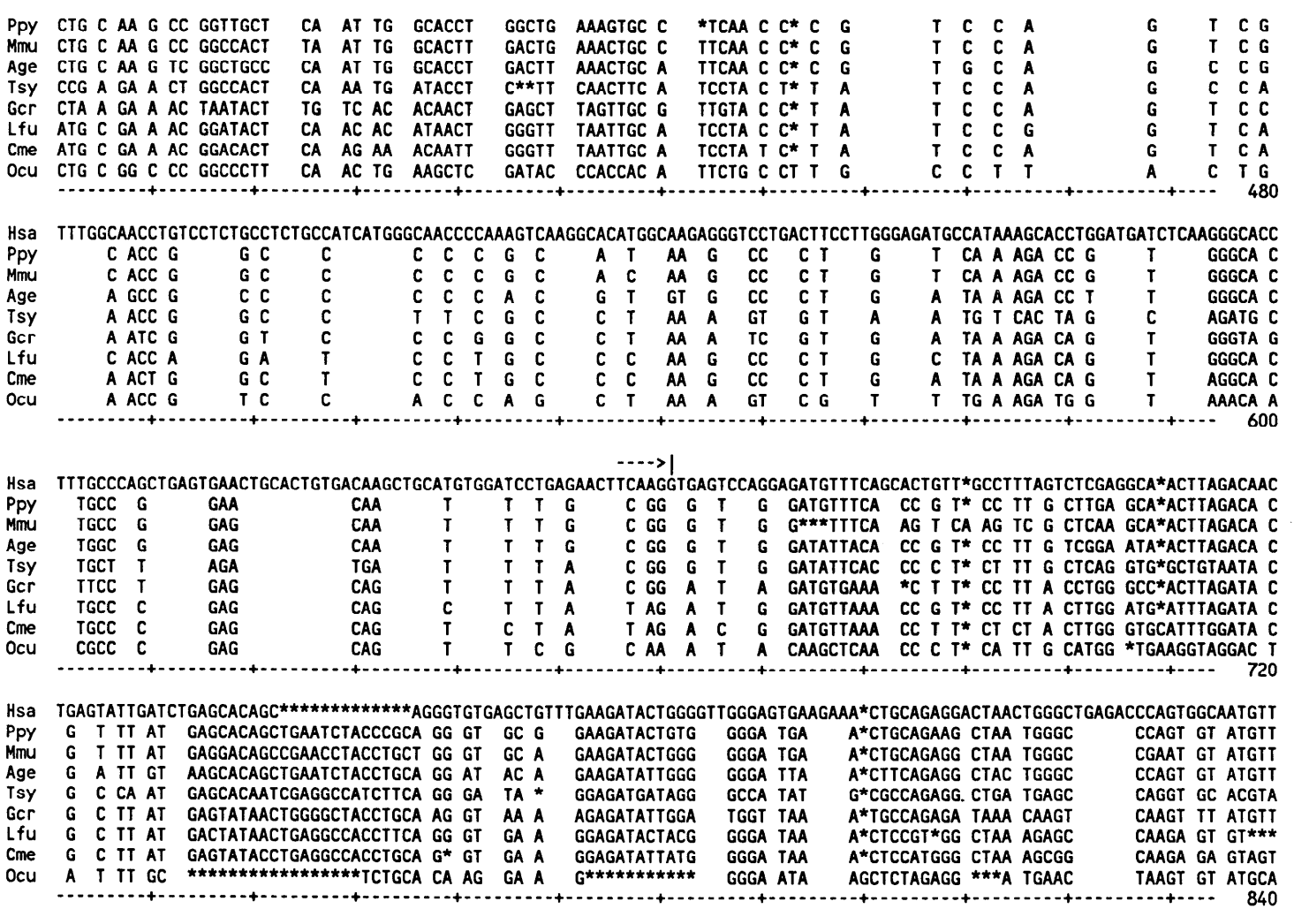

Hsa TTTTTTGACAGCTTTGGAAACCTGTCGTCTCCCTCTGCCATCCTGGGCAACCCCAAGGTCAAGGCCCATGGCAAGAAGGTGCTGACTTCCTTTGGAGATGCTATTAAAAACATGGACAAC
Ppy T C GC A C A C CC C C T G CT A T TA T A CA
Tsy T C AC A C C T TA A C T G CC G T TA T A TA
Gcr C G CC A C C G CA C C T A CC A A TG C A CA
Lfu C T AT A C C G AC C C T G CC A G TG C A CA
Ocu C C AC C T C T AC A A T G AC A T CA C G CA
Chi C T GC C C C G AA A C C G CC A A TA T G TT
600
Hsa CTCAAGCCCGCCTTTGCTAAGCTGAGTGAGCTGCACTGTGACAAGCTGCATGTGGATCCTGAGAACTTCAAGGTGAGTTCAGGTGCTGGTGATGTGAT*TTTTTGGCTTTATATTTT**G
Ppy C GACCA C T G GC T G AG G TC TGCG G GA ATGC **TTTA CT AT TT **G
Tsy G GGGTG G T A GC C A GG G AC AGAG C CA ATGC CTTTTT AT TT CT **G
Gcr C GGGTG C T G GC C G AG A TC AAAT C AC AGGC C***TT C* TC CT GAG
Lfu C GGGTG C T G AC C G AG G GC AGAT C CA GTCC CTTTTT CT TT CT GAG
Ocu T GGGTG C T G AC T G AA G TT AGAT C CA GAGC CTTTCT TT AT AC AGC
Chi C AGGTG C C G GT C G GG G TC AAGT T CA GCGT CCCTTT CT TT CC GCA
720
Hsa ACATTAATTGAAGCTCATAATCTTATTGGAAAGACCAACAAAGATCTCAGAAATCATGGGTCGAGCTTGATGTTAGAACAGCAGACTTCTAGTGAGCATAACCAAAACTTACAT*GATTC
Ppy C T TT A GC T GAA C T GG AA CC ACA GG TC AT TGGGTCGAGC T A G GAACAGC CT T GTGAG AAC A AA TA CTG*G
Tsy C T TT A GC G CAC A G GA AA TC GTG AG TC AT TGGGTATAGC T G G GAACAAC CT C GTGGG AAC A GA CA CTTTA
Gcr C A TG A GG T GAC A G AG AG TC ACA AG CG AA TAGGTCCAGT T G C AAAAGAC TT C GTAGG AAG A GA CA CTT*G
Lfu C A CG A GC T TAC A A AC AA TC ACA AT TC AC *****CCAGC T C G GAAGGAC CT C GGGGG AAC A GA CA TTT*G
Ocu C A TG G TA A GAC A G GG GA TC ACA AG TT GT TAGATATGTT T T G G****** TC C ATGGT CAA A GA CA TTT*A
Chi T A TG A GT G GTG T T GG AA CT G*A AG CC AT TAGATCAAAC A G G GGAGGAC CT C GTGGG A*C G GC C* CTT*G
840
Hsa AGAACTAGTGACAGTAAAGGACTACTAACAGCCTGAATTGGCTTA**ACTTTTCAGGAAATCTTGCC*AGAACTTGAT*GTGTTTATCCCAGAGAATTGTATTATAGAATTGTAGACTTG
Ppy A CT TG CAG GGAC CTAAC C TGAATTGGCTTA**ACTTTTCAGGAAATCT GC *AG CTT * G AT TGG G TTGTA T TAG TGT GACT G
Tsy A CT TA CAG GGGA CTAGC C AGAATTTTCTTAAAACTTTT*AGCAAGTTA GT *** CTA * G TT TGG T TAATA T TAG CGA GACT T
Gcr A TG GG CAA GAGA GTAAT G TAAATTTTCTTACCACTTTTCAGGAACATT GT AAG CTG G T TT TGG G TAATA T TAT AGC GACC C
Lfu A CT TG CAG GGGA CTAGC G TGAATTTACTTAAAATTTTTCAGGAAACTT GT *AG CTG * T AT TGG G TAACA T TAT AAC GACT C
Ocu A AT T* AGG AGGC TTAAC T T**********AATT TT *AG AAT * G AT TG* G TAATG T TTG TGT GCTT G
Chi G CT TG CA* GAGC TGGGC C TTAGTTTGCTTAAAAGTTTTCTGGAACTTC GT *AG CTG * A AC CAG G ***TA C AAG AGC TATT G
960
Hsa TG*AAAGAAGAATGAAATTTGGCTTTTGGT*AGATGAAAGTCCATTTCAAGGAAATAGAAATGCCTTATTTTATGTGGGTCATGATAATTGAGGTTTA**********
Ppy G*AAAG AG GAA T GC T GGT* TG AGT CA T A GG AAT G ATGCCTTATTT ATG GGG CATGA AATTGAGGTTTA**********
Tsy GCAAGA GA ATG T AC T GATT GG AGC TG C A AG AGG G ATGTCTT*TTA ATG GGG CTTAA AGTTGATGATTATCAAAATAAAATTTTGGGGGGA
Gcr GCAAAG GA GCA T GC T GGT* TG GGC TG C A GA AGG C ATGTCTTATT* ATA GGG CTTGA AATCCAAGCTTA**********
Lfu GCAAGG TA GCA T GA T GGT* TG GGC TA C A AG AGA C ATTCTTTATTT ATG GGG CCTGA AATCGAAGTTTA**********
Ocu ACAAGG AA AGA T TC G AGT* TG GGC TA C A TG AGG G ATGC*****AA GCT ACA TATAT TTTTCTGATGAT**********
Chi TCTGGG GA GAA C GC T GAA* AT GTC AG C T GG GGG G *****TTATCC ATG GAN CCCGA GACTGAAGTTTA**********
1080

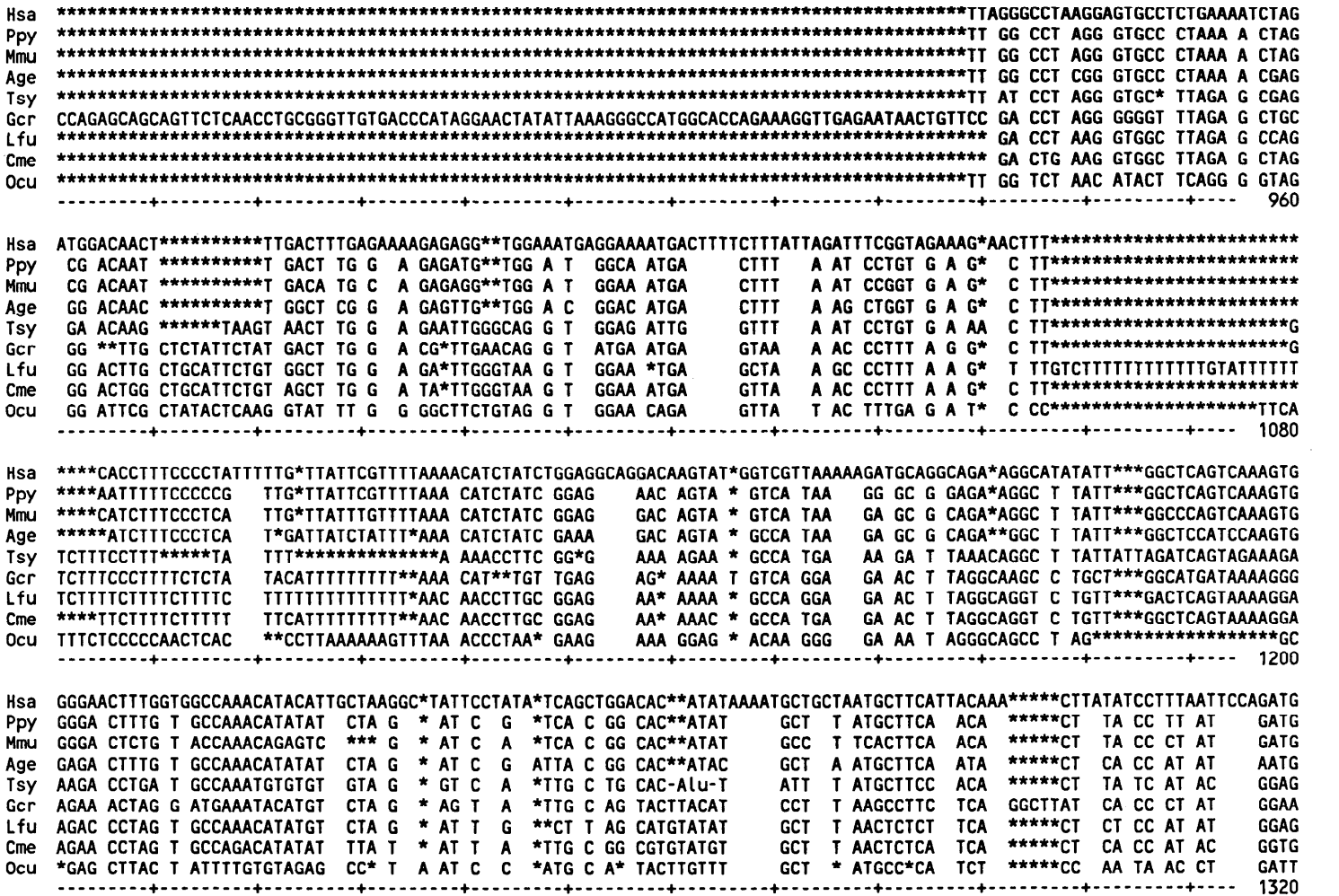

Hsa GA**AGAGATTTTTGCAAAAAAAATAAA*******AGATTTGCTCAAAGAAAAATAAGACACATTTTCTAAAATATGTTAAATTTCCCATCAGTATTGTGACCAAGTGAAGGCTTGTTTC
Ppy GA*AAGAG TT TTGCAAAA AAAAAAAAAAAAA G TTT CA G A TAGGA CAT T CTAAA TATGTTAAA TC CA T T G CCAAGTGAAGGCTTGTTTC
Tsy AATATAAA TA TTGGGGTA AAATTAAAAG**** T TTT CA G T TAAGA AAT C TTAAA TATGCTAAA TC TG T T C C******************
Gcr G*GAAAAG TT TGTTAAAA TTTTAAAGGT**** G TTT CA G A TACAG AAT C CTAAA TAT******A CT TG T C G TGAAATTGAGGCTTGTTAG
Lfu GAGAAAAG TT CGGTGAAG ATTTAAAGAC**** G TTT CG G A CACAA AAT C CTAAA TACATTAAA TT TG T T G CCAAATGGAGACTTTTTCC
Ocu GAACTAGG CT GATTAGGG CGATTTATGCG*** G CT* TA A A TACAA AAT T CTGAA AATATTAAA CT ** T T G CCAAATATAAACTTGTTCC
Chi GG***AAG TA TTGGGAGA TAATTATTAGCC** G CAT CA G A TTGAT ATA C CAAGG A******** AC CA A C G CTAGGTGGAGGCTTATTGT
1200
Hsa CGAATTT*GTTGGGGATTT**TAAACTCCCGCTGAGAACTCTTGCAGCACTCACATTCTACATTTACAAAAATTAGACAATTGCTTAAAGAAAAACAGGGAGAGAGGGAACCCAATAATA
Ppy TGAA T*G AGG AT ** A A CTGCTG GA T T CAGCA CAC TCT CATT TCAG A TTAGACAACTGCTTAAAGAAAAACAGGG GAG GGGAGCCCAA AATA
Tsy ***A G*G AAA AT ** A * TCACCA GA C T TAGCA GAA TCT TGTT ATGG T TTAAACAATTGTTTAAAGAGAAA**GAG GAA AGGAGGCAAT GGCA
Gcr TACA G*G GGA AT ** C A CCACTG AA T T ATGAA GAA CTG TATT ACAG A T*****AAT******AATAAGAA*AGGG GAA GGGAGAGAAA GACA
Lfu TGCA G*G GGA AT ** A A CCACTG GA C T CAGCA GAA CCT TGTT ACAG A TTAGACAATTG*GTAAATAAGAA*AGGA GAG AGAAGGCAAA AACA
Ocu TGCT G*G GGA AC ** A A C***** AG C T CAACA GGC TCT TGTT ACTG A CTAAACAATTA*TTAAATGAAAA**AGA AAA AGAAAGAAGA AAAT
Chi TGCA GAA GAG GT AC A G ATTCTA CA C A CAGCC GA* CCT TGAA A*TA A TTAGA**********AGG GGG AAAGGCAAC* AAAA
1320
Hsa CTGGTAAAATGGGGA*AGGGGGTGAGGGTGTAGGTAGGTAGAATGTTGAATGTAGGGCTCATAG******AATAAAATTGAA*****CCTAAGCTCATCTGAATTTTTTGGGTGGGCACA
Ppy C G AA GGA GAGG GGTGAGGATGTAGGTGGGT GA G A TGT G GC T A****** TA T G *****CCT ATC CA C GAATT TTT AGTAGGCACA
Tsy A G GA GAG GAAA GTTGAGGATTTTTGTAGGT GA A C TGT G GC T AA***** TG T G *****CCT AGC TA T GAATG ATT TACAGCAACA
Gcr A G GG GAA GATA AGTGAGGATATAGTTAAGC AG G A TAT G GT A AAA***** AA C A TTAGGCCT TGT TA T AAACA ATT GA********
Lfu A G TA GGG GAGA GGT********AGGTAGGC GG G A TGT G GC T AC***** TG T G *****CCT TGC TA C AAACT ATT GATAGGTAGA
Ocu A G GA GTG AGGA GAT*AAGATATAGGTAGTC GA G A CGG A TC T AA***** TG T G *****TCC TTC TG C TGGTT A*A GGTAGGCACA
Chi A A GA *AG GAGA GCAAGGGATATAGGCAGAC AA A T TGG G GC T GGATTTA TT T G ****GGAC AGC CA C GAGTT ATT TATAGGTACA
1440
Hsa AACCTTGGAACAGTTTGAGGTCAGGGTTGTCTAGGAATGTAGGTATAAAGCCGTTTTTGTTTGTTTGTTTGTTTTTTCAT*CAA*GTTGTTTTCGGAAACTTCTACTCAACATGCC****
Ppy AACCTTGGAACAGTTTGAGG CA GGTTGTCTA GAATG AGGT TAAA C A TTTTGTTTGTTTGTTTGTTTTTTCAT*CAA*GTTGTTTTCGGAAACTTCTACT ACA C ****
Tsy AAACTTGAAGCAGTTTGAGG CA AGTCGCCAA GATTG AGGT TAAC C A CTTT**********TTCT**G TCA C ****
Gcr ********** CA GGTTGTCA* GAA*G A**T TGAA C C TCTT**********TTTTTTCCCTTACAACATTCTTCTTGGAAACTTCTACT ACA C A*TT
Lfu ACCCTTGGAGCAGTTTGAGG CA AGTAGTCAG GAA*G G**C TAAA C A TTT**********ATTTTTCCTGCAA*ATTCATTTTGAAAACTTTTGCT ATA C ATTT
Ocu ATCCTTGGAGTAATTTGAGG CA AGTTTTCAG GAATG AGGT CATA T G TTC**********TTTCTT*TGTAA*ATTC*TTTTGGGAACACCTACT ATA C ****
Chi ACCCATGGAGAAGTTTAAGA GT GACTTGGGA TGGTT AGGT CTAA C A **********TTTC*TGTA**ACTCTTTTAGCAAACTT***** ACT G CTAC
1560

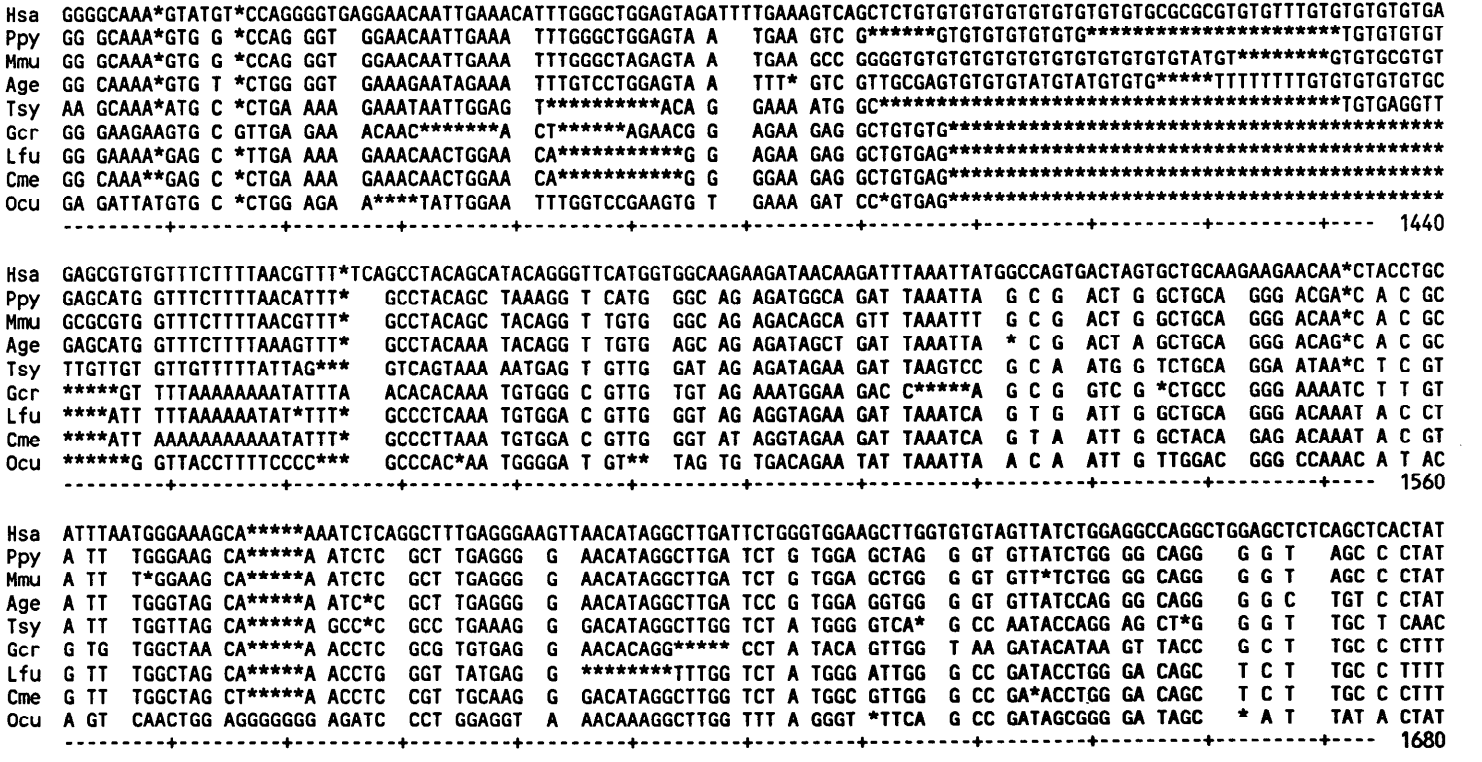

EXON 3 ---->
Hsa *TGTGTGTTATTTTGTCTTTTGCCTAACAGCTCCTGGGTAACGTGATGGTGATTATTCTGGCTACTCACTTTGGCAAGGAGTTCACCCCTGAAGTGCAGGCTGCCTGGCAGAAGCTGGTG
Ppy * ATGTA T TT G T TGC T T C A GG G TA CT C AG G C T A G C C G G
Tsy * ATATT C CT T T CAC T T C T GG C TA TT T AG C C T A A T C G G
Gcr C GTC*T C TT G T CA* T A T A GG G CA CT T AG A T T T G C C G A
Lfu T TTCTT C TT G T CAC T A C A GG G TA CT T AG A C T T A C C G A
Ocu * ATGTA C TT G T TAC T A C C GC G CG CC T AA A T G G G C T T C
Chi C AATTC T TC G C CAC C C T A TG G TA CT T GA A C T C G C C G G
1680
---->TER
Hsa TCTGCTGTCGCCATTGCCCTGGCCCATAAGTACCACTGAGTTCTCTT*CCAGTTTGCA*GGTCTTCC**TGTGACCC**TGACACCCTCCTTCTGCACAT*GGGGACTGGG*CTTGGCCT
Ppy CT C C T C G C T A C GTTC C *CC GTTTGCC*AATGTTCC**TGTGAC T**TGACAC C CC CA AT*G GG G *C TGGCC
Tsy GA T C C A G T C G C GACT C TCC GACGACT*AGTGTTCC**TGTGTT T**CAACAC C CT CA AT*G GC G *C GGGAC
Gcr GG G T C T A C C G C GTCC T TCC GTTTGCC*AGCATTTCTTTGTGTG CTTTGACAT C TT AA AT*G GG * *A TGTTT
Lfu GC T C C T G C C G C GCCC T TCT GTTTGCC*AGTGTCTCTGTGTGTT C**TGACAT C CC CA AA*G AG G *C TGTTC
Ocu GT T C T T G C C G T ACTC C CCC GTTTCTTTAGTGCCCCTACTCATC C**CAGCGT C CC TA CTTG GG A *C CAGCC
Chi GT T C C T G C C G C ATTC C TAC ATTCACC*******ATTTTGTGTC C**CAGTGC T CC CC CTT* GG G GT TGGCC
1800
Poly A
Hsa TGAGAGAAAGCCTTCTGTTTAATAAAGTACATTTTCTTC****AGTAAT***CAAAAATTGCAA*TTT*TATCTTCTCCATCTTTTACTCTTGTGTTAAAAGGAAAAA********GTGT
Ppy T AGAGAAAGCC CT A G CA TTCT **** G A ***C AAA TCCCAA* TT*TATC CTCCA CTT TAC CT GG TA *GG A********GT
Tsy T CGAGCTAAGC CT A G TA CCAA **** G A ***C *GG CTGCAA* TT*TATC CCTCA TTT TAC CT GT CA GAA *********GT
Gcr C AAAGCATAGC TC G C CT TTAT AGTA A A AAAA AAA CCACAAC TTATTTT TTGCT **T TCC TG AT TA *** AGGGGG***GG
Lfu T AGAGCATAGC CT A G CA CTAT **** A A ***A AAA TTGCAA* TT*TGTC CTCCA CGT CCC CA GT TA *GG A********GG
Ocu T ATGGCACAGT CT A A CA CTAT **** G A ***C AAA ATGATG* CT*CATC CTTCG ATT TCA CA GT TT AAA AAACCACACAT
Chi T TGAACCCAGA CT A A CA CTAT **** G G ***C AAA TTAAAA* TG*TACC CTCTA CA<
1920
Hsa T*CATGGGCTGAGGGATGGAGAGAAAC*AT***AGGAAGAACCAAGAGCTTCCTTAAGAAATGTATGGGGGCTT**GTAAAATTAATGTGGATGTTATGGGAGAATTC
Ppy * T GAC GGG TG AG GAAAC*AT***AGGAAGAACC G GCTT C A GAAA GTATGG GCTT**GTAA TT TGTG A CGAG AA T
Tsy T T GAC GGG TA AG AAGAC*ATAAAAGGAAGAAAT G GCTT T A GAGG GGATGG ACTC**TTCA GG AATG A TGGG GA T
Gcr * T GGC AGA ** AG **********A A GTGG T A AAAG GTGCGG AACCAGACCA AG GGTG A TAGA GA T
Lfu * C GGC GGA TG GA GAGACGAT***AGGAAAAAC* A GCTT T A GAAG ATGTGG ATTC**ATCG GG ACTG A TGGA GA T
Ocu * T *AT GAC GA AG GAC***AT***AGGAAGAAAT G A**T T G GTAG ATATTA ATTT**ATCA TA GGGA G TTGG GG A
Chi
2028

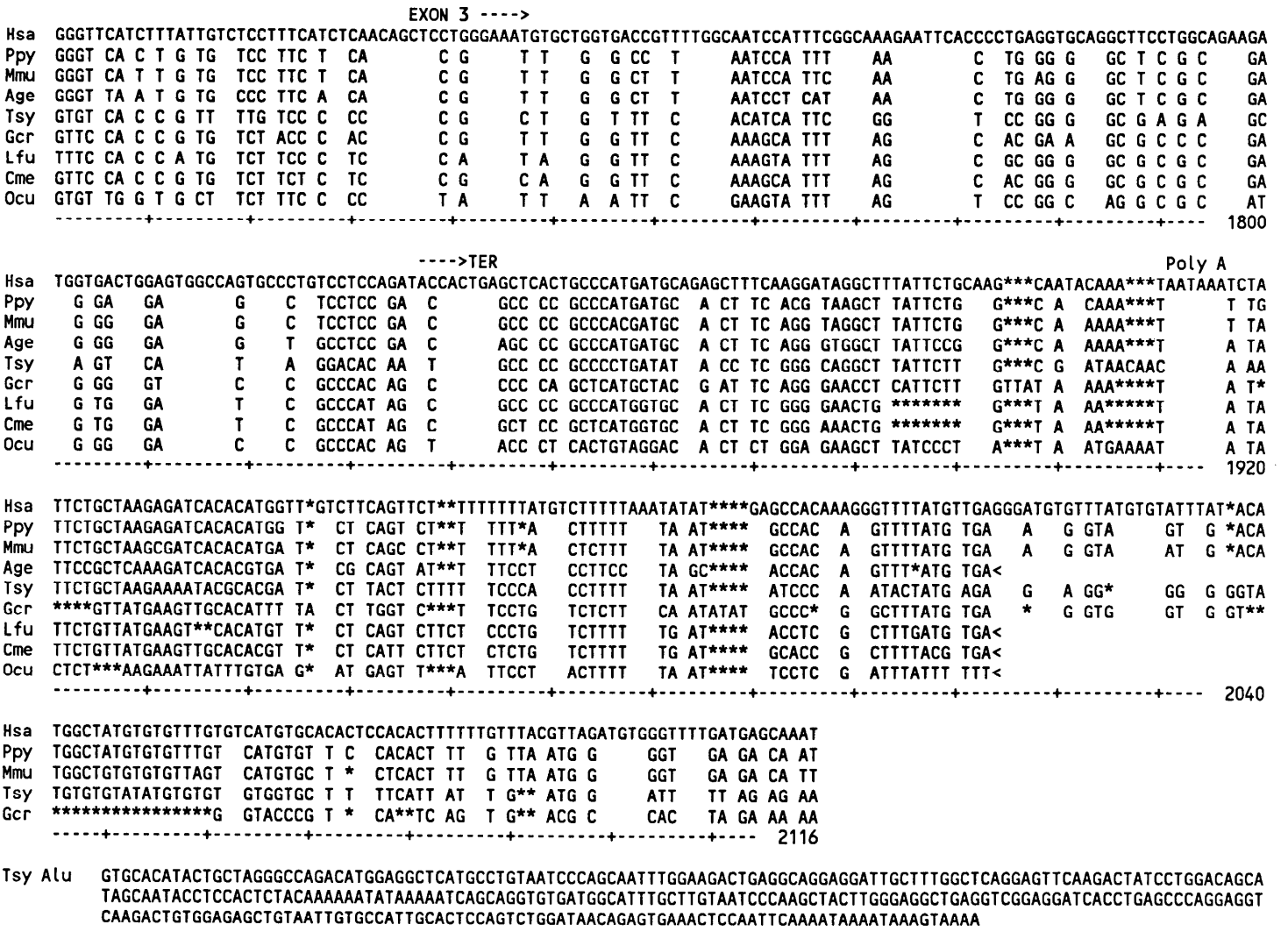
ALIGNED GAMMA SEQUENCES

```
                CACCC                                CCAAT                                CCAAT
Hsa  TAAACTCCACCCATGGGTTGGCCAGCCTTGCCTTGACCAATAGCCTTGACAAGGCAAACTTGACCAATAGTCTTAGAG*TATCCAGTGAGGCC*AGGGGCCGGCGGCTGGCTAGGGATGA
Ppy       TC     AT G T  G   GC    CCTTGACCAATAGCCTTGACAAGGAAAACT     A    T  T G  *TATCCG T  GG C*A    C GGCG CTGGCT GG ATGA
Mmu       TC     AT G T  G   GT    CCTTGACCAATAGCCTTGACAAGGCAACCT     A    T  T G  *TATCAG T  GG C*A    C GGCG CTGGCT GG ATGA
Age            AT G T  G   GC    CCTTGACCAATAGCTTTGACAAGGCAACCT     A    T  T G  *TATCGG T  GG CCG    C GGTG GTGGCT GG ATGA
Tsy       CC     CT T T  A   GC    **************************CCT     G    C  C G  CCCTCAG T  GG C*A    T AAAG TCA*** AT ACAA
Gcr       CA     CT T T  A   TC    CCTTGACCAATAGGCTTGACAAAGTGACCT     A    C  C G  *CACCAT T  GG A*A    C AGGG GCC*** AG ATAG
Lfu          >   A* G T  A   GC    CCTTGACCAA************GCAACCT     A    C  C G  *CACCAT T  GG A*A    C AGGG ATC*** GG ATGG
Cme          >   A* G T  A   GC    CCTTGACCAATAGCTTTGACCAAGTGACCT     A    C  C T  *CACCAT A  GG A*A    C AGGG GTC*** GG TTGG
Ocu       CT     CT C C  A   GC    CCTTGACCAATAGTCGTTACACAAAAACAC     A    C  C G  *AACACG C  AA A*A    C AGAT TCC*** GC AGGA
          --------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------+----  120

          AATAAA                                                                                    INIEXON 1 ---->
Hsa  AG*AATAAAAGGAAGCACCCTTCAGCAGTTCCACACACTCGCTTCTGGAACGTCTGAGGTTATCAATAAGCTC*********CTAGTCCAGACGCCATGGGTCATTTCACAGAGGAGGAC
Ppy   *          A CA  C CC G     CC CAC TTCG  T   GAA GT     GT ATCAAT      ********* TAGT CA   GC    GG C    C  A AG  G  C
Mmu   *          A CA  C CC G     CC CAC CTCG  T   GAA GG     AT ATCAAT      ********* TAGT CA   GC    GG C    C  A AG  G  C
Age   *          A CA  C CC T     CC CAT CTCG  *   AAA GT     AT ATCAAT      ********* TTGT CA   GC    AG A    C  A CT  G  C
Tsy   *          A CA  T CT G     GC CAT CAT*  T   AAA AT     AT A*GGAC      ********* CAAT CA   AG    GT C    T  A CT  G  G
Gcr   A          A CA  T CT G     GC CTT CTTG  T   CTT GT     AT ATCAAC      ********* TTTT TA   AC    GT C    T  A CT  A  G
Lfu   A          A CC  T CT G     GC CAT CTTG  T   ACA GT     AT ACCGAC      ********* CAAT CT   AC    GT C    T  A CG  A  G
Cme   A          A CC  T CT G     TT TAC TTTG  T   ACA AT     AT GTCAAC      ********* TGAT CT   AC    GT C    T  T TG  A  G
Ocu   *          C AG  T AG G     TC CAT CTTG  T   AGA AT     AC ATCAGC      AGCGAGCTC TAGA CA   AT    GT C    C  G CT  G  A
          --------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------+----  240

                                                                          ---->|
Hsa  AAGGCTACTATCACAAGCCTGTGGGGCAAGGTGAATGTGGAAGATGCTGGAGGAGAAACCCTGGGAAGGTAGGCTCTGGTGACCAGGACAAGGGA*GGGAAGGAAGGAC****CCTGTGC
Ppy       AC       A   CTG   GGCAAGT  A T  G     TGCT  A G    A T  G        G CT  G TGACCA GACA  GG * G AAG   GG C****T TGTGC
Mmu       AC       A   CTG   GGCAAGG  A T  G     TGCT  A G    A C  G        G CT  G TGACCA GACA  GA * G AAG   GG C****C TGTGC
Age       GC       T   CTG   GGCAAGG  A T  G     TGCT  G G    A C  G        G CT  G TGACCA GACG  GG * G AAG   GG A****C TATGC
Tsy       AT       A   CTG   GCTAAGG  A T  G     GACT  A G    G C  G        A TT  A GTACTG AATG  GG * G AAA   AT T****C *ATGC
Gcr       AT       A   CTG   GGTAAAG  A T  A     GGAT  A G    G T  G        T CT  G GGTGGT GGCA  AC A G AGG   GG ATGAAC TATGC
Lfu       GT       A   CTG   GGCAAGG  A T  G     GGCT  A G    G T  G        T CT  G GGGCCA GACA  GG * G ATA   GG A****C TGTAT
Cme       GT       A   CTG   GGCAAGG  A T  G     GGCT  A G    G C  G        C CT  G GGGCCA GACA  GG * G ATG   TG A****C TGTGT
Ocu       GC       A   ACA   AAGCTGG  G C  A     TGCC  A C    G C  C        G CC  G GGTCCA GACA  GC * A GAG   TG A****A TGAGC
          --------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------+----  360
```

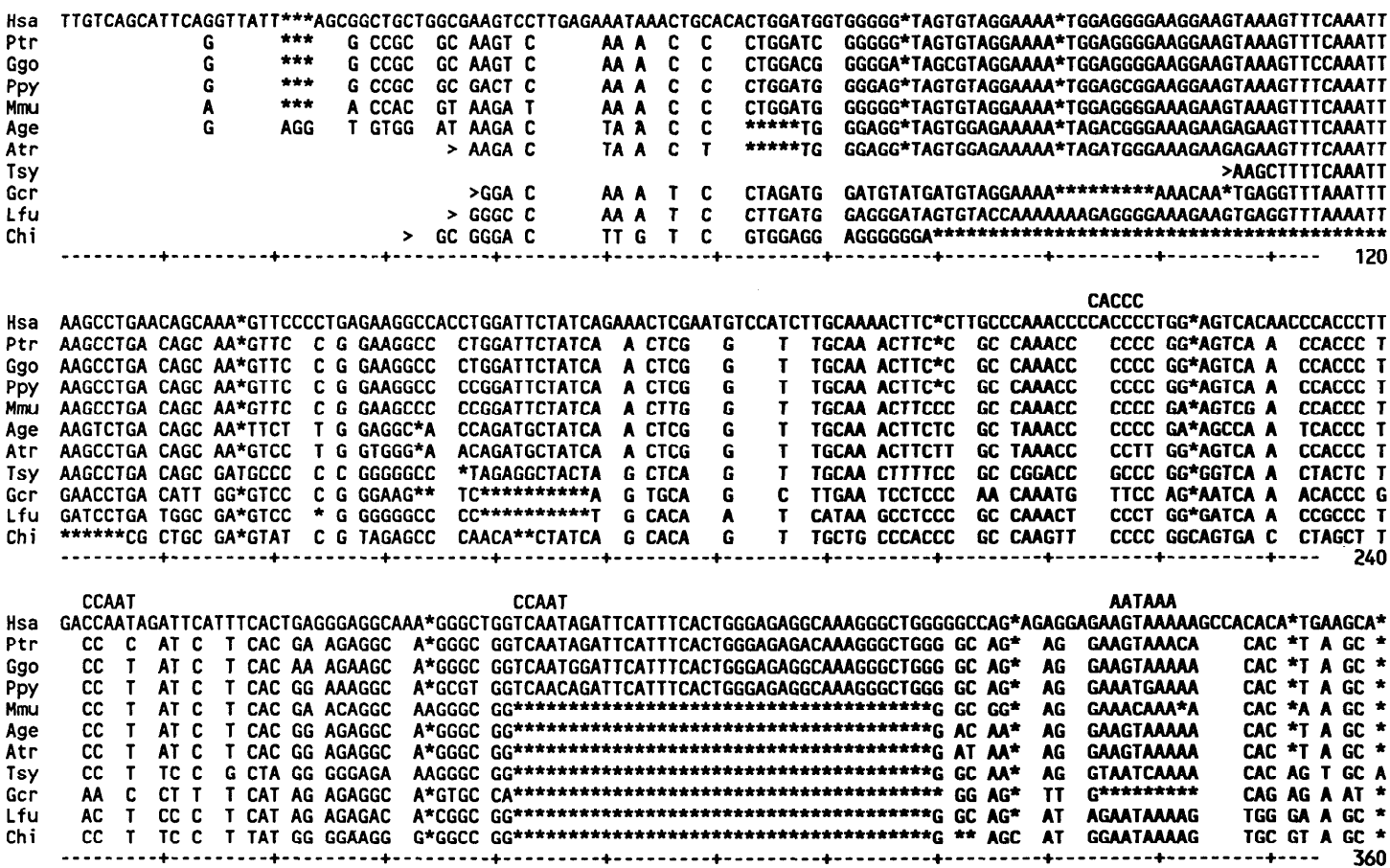

Ppy CTG C AA G CC GGTTGCT CA AT TG GCACCT GGCTG AAAGTGC C *TCAA C C* C G T C C A G T C G
Mmu CTG C AA G CC GGCCACT TA AT TG GCACTT GACTG AAACTGC C TTCAA C C* C G T C C A G T C G
Age CTG C AA G TC GGCTGCC CA AT TG GCACCT GACTT AAACTGC A TTCAA C C* C G T G C A G C C G
Tsy CCG A GA A CT GGCCACT CA AA TG ATACCT C**TT CAACTTC A TCCTA C T* T A T C C A G C C A
Gcr CTA A GA A AC TAATACT TG TC AC ACAACT GAGCT TAGTTGC G TTGTA C C* T A T C C A G T C C
Lfu ATG C GA A AC GGATACT CA AC AC ATAACT GGGTT TAATTGC A TCCTA C C* T A T C C G G T C A
Cme ATG C GA A AC GGACACT CA AG AA ACAATT GGGTT TAATTGC A TCCTA T C* T A T C C A G T C A
Ocu CTG C GG C CC GGCCCTT CA AC TG AAGCTC GATAC CCACCAC A TTCTG C CT T G C C T T A C T G
480
Hsa TTTGGCAACCTGTCCTCTGCCTCTGCCATCATGGGCAACCCCAAAGTCAAGGCACATGGCAAGAGGGTCCTGACTTCCTTGGGAGATGCCATAAAGCACCTGGATGATCTCAAGGGCACC
Ppy C ACC G G C C C C C G C A T AA G CC C T G T CA A AGA CC G T GGGCA C
Mmu C ACC G G C C C C C G C A C AA G CC C T G T CA A AGA CC G T GGGCA C
Age A GCC G C C C C C C A C G T GT G CC C T G A TA A AGA CC T T GGGCA C
Tsy A ACC G G C C T T C G C C T AA A GT G T A A TG T CAC TA G C AGATG C
Gcr A ATC G G T C C C G G C C T AA A TC G T G A TA A AGA CA G T GGGTA G
Lfu C ACC A G A T C C T G C C C AA G CC C T G C TA A AGA CA G T GGGCA C
Cme A ACT G G C T C C T G C C C AA G CC C T G A TA A AGA CA G T AGGCA C
Ocu A ACC G T C C A C C A G C T AA A GT C G T T TG A AGA TG G T AAACA A
600
---->|
Hsa TTTGCCCAGCTGAGTGAACTGCACTGTGACAAGCTGCATGTGGATCCTGAGAACTTCAAGGTGAGTCCAGGAGATGTTTCAGCACTGTT*GCCTTTAGTCTCGAGGCA*ACTTAGACAAC
Ppy TGCC G GAA CAA T T T G C GG G T G GATGTTTCA CC G T* CC TT G CTTGA GCA*ACTTAGACA C
Mmu TGCC G GAG CAA T T T G C GG G T G G***TTTCA AG T CA AG TC G CTCAA GCA*ACTTAGACA C
Age TGGC G GAG CAA T T T G C GG G T G GATATTACA CC G T* CC TT G TCGGA ATA*ACTTAGACA C
Tsy TGCT T AGA TGA T T T A C GG G T G GATATTCAC CC C T* CT TT G CTCAG GTG*GCTGTAATA C
Gcr TTCC T GAG CAG T T T A C GG A T A GATGTGAAA *C T T* CC TT A CCTGG GCC*ACTTAGATA C
Lfu TGCC C GAG CAG C T T A T AG A T G GATGTTAAA CC G T* CC TT A CTTGG ATG*ATTTAGATA C
Cme TGCC C GAG CAG T C T A T AG A C G GATGTTAAA CC T T* CT CT A CTTGG GTGCATTTGGATA C
Ocu CGCC C GAG CAG T T C G C AA A T A CAAGCTCAA CC C T* CA TT G CATGG *TGAAGGTAGGAC T
720
Hsa TGAGTATTGATCTGAGCACAGC**********AGGGTGTGAGCTGTTTGAAGATACTGGGGTTGGGAGTGAAGAAA*CTGCAGAGGACTAACTGGGCTGAGACCCAGTGGCAATGTT
Ppy G T TT AT GAGCACAGCTGAATCTACCCGCA GG GT GC G GAAGATACTGTG GGGA TGA A*CTGCAGAAG CTAA TGGGC CCAGT GT ATGTT
Mmu G T TT AT GAGGACAGCCGAACCTACCTGCT GG GT GC A GAAGATACTGGG GGGA TGA A*CTGCAGAGG CTAA TGGGC CGAAT GT ATGTT
Age G A TT GT AAGCACAGCTGAATCTACCTGCA GG AT AC A GAAGATATTGGG GGGA TTA A*CTTCAGAGG CTAC TGGGC CCAGT GT ATGTT
Tsy G C CA AT GAGCACAATCGAGGCCATCTTCA GG GA TA * GGAGATGATAGG GCCA TAT G*CGCCAGAGG.CTGA TGAGC CAGGT GC ACGTA
Gcr G C TT AT GAGTATAACTGGGGCTACCTGCA AG GT AA A AGAGATATTGGA TGGT TAA A*TGCCAGAGA TAAA CAAGT CAAGT TT ATGTT
Lfu G C TT AT GACTATAACTGAGGCCACCTTCA GG GT GA A GGAGATACTACG GGGA TAA A*CTCCGT*GG CTAA AGAGC CAAGA GT GT***
Cme G C TT AT GAGTATACCTGAGGCCACCTGCA G* GT GA A GGAGATATTATG GGGA TAA A*CTCCATGGG CTAA AGCGG CAAGA GA GTAGT
Ocu A T TT GC **********TCTGCA CA AG GA A G********** GGGA ATA AGCTCTAGAGG ***A TGAAC TAAGT GT ATGCA
840

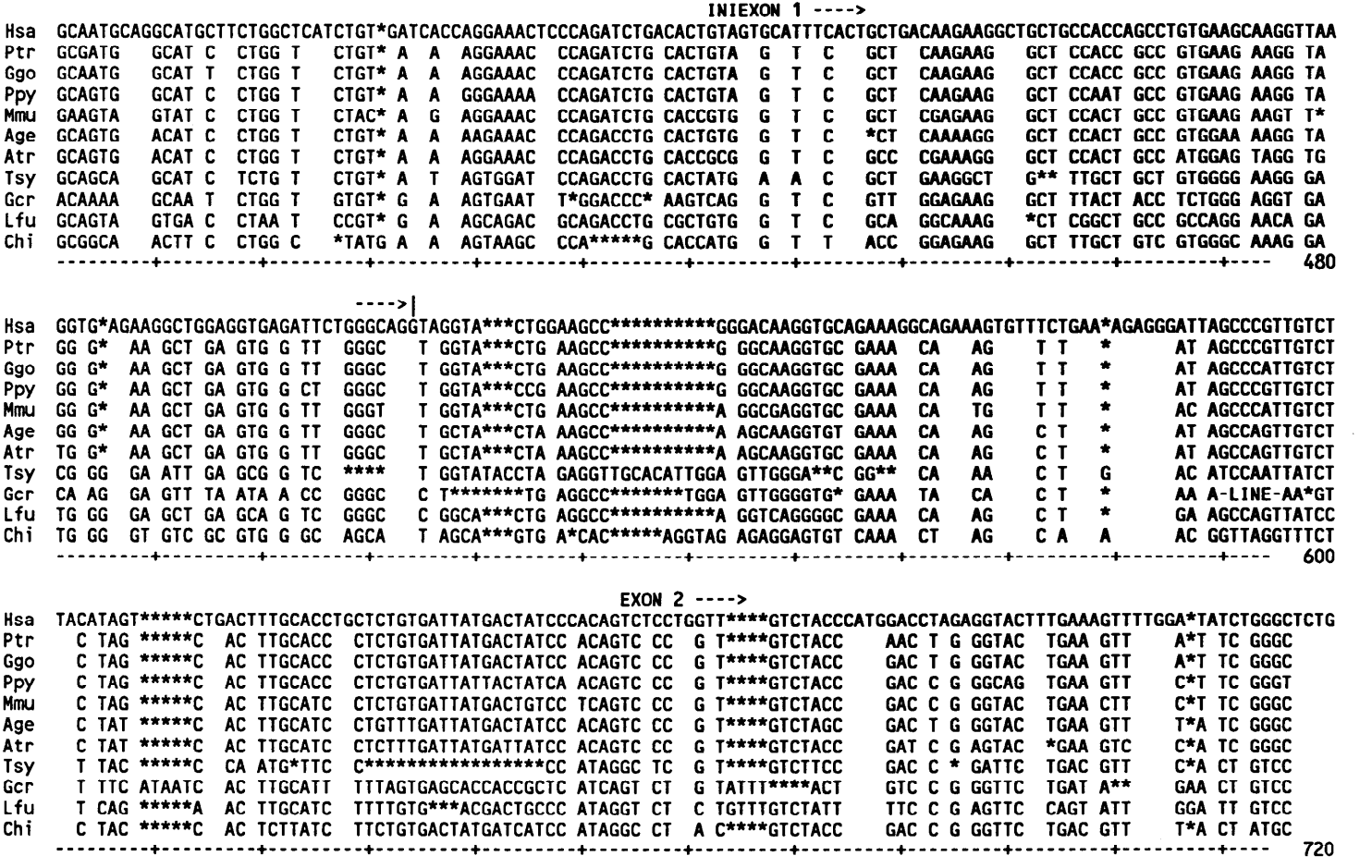

Hsa **********TTAGGGCCTAAGGAGTGCCTCTGAAAATCTAG
Ppy **********TT GG CCT AGG GTGCC CTAAA A CTAG
Mmu **********TT GG CCT AGG GTGCC CTAAA A CTAG
Age **********TT GG CCT CGG GTGCC CTAAA A CGAG
Tsy **********TT AT CCT AGG GTGC* TTAGA G CGAG
Gcr CCAGAGCAGCAGTTCTCAACCTGCGGGTTGTGACCCATAGGAACTATATTAAAGGGCCATGGCACCAGAAAGGTTGAGAATAACTGTTCC GA CCT AGG GGGGT TTAGA G CTGC
Lfu ********** GA CCT AAG GTGGC TTAGA G CCAG
Cme ********** GA CTG AAG GTGGC TTAGA G CTAG
Ocu **********TT GG TCT AAC ATACT TCAGG G GTAG
960
Hsa ATGGACAACT**********TTGACTTTGAGAAAAGAGAGG**TGGAAATGAGGAAAATGACTTTTCTTTATTAGATTTCGGTAGAAAG*AACTTT**********
Ppy CG ACAAT **********T GACT TG G A GAGATG**TGG A T GGCA ATGA CTTT A AT CCTGT G A G* C TT**********
Mmu CG ACAAT **********T GACA TG C A GAGAGG**TGG A T GGAA ATGA CTTT A AT CCGGT G A G* C TT**********
Age GG ACAAC **********T GGCT CG G A GAGTTG**TGG A C GGAC ATGA CTTT A AG CTGGT G A G* C TT**********
Tsy GA ACAAG ******TAAGT AACT TG G A GAATTGGGCAG G T GGAG ATTG GTTT A AT CCTGT G A AA C TT**********G
Gcr GG **TTG CTCTATTCTAT GACT TG G A CG*TTGAACAG G T ATGA ATGA GTAA A AC CCTTT A G G* C TT**********G
Lfu GG ACTTG CTGCATTCTGT GGCT TG G A GA*TTGGGTAA G T GGAA *TGA GCTA A GC CCTTT A A G* T TTGTCTTTTTTTTTTTTGTATTTTTT
Cme GG ACTGG CTGCATTCTGT AGCT TG G A TA*TTGGGTAA G T GGAA ATGA GTTA A AC CCTTT A A G* C TT**********
Ocu GG ATTCG CTATACTCAAG GTAT TT G G GGCTTCTGTAG G T GGAA CAGA GTTA T AC TTTGA G A T* C CC**********TTCA
1080
Hsa ****CACCTTTCCCCTATTTTTG*TTATTCGTTTTAAAACATCTATCTGGAGGCAGGACAAGTAT*GGTCGTTAAAAAGATGCAGGCAGA*AGGCATATATT***GGCTCAGTCAAAGTG
Ppy ****AATTTTTCCCCCG TTG*TTATTCGTTTTAAA CATCTATC GGAG AAC AGTA * GTCA TAA GG GC G GAGA*AGGC T TATT***GGCTCAGTCAAAGTG
Mmu ****CATCTTTCCCTCA TTG*TTATTTGTTTTAAA CATCTATC GGAG GAC AGTA * GTCA TAA GA GC G CAGA*AGGC T TATT***GGCCCAGTCAAAGTG
Age *****ATCTTTCCCTCA T*GATTATCTATTT*AAA CATCTATC GAAA GAC AGTA * GCCA TAA GA GC G CAGA**GGC T TATT***GGCTCCATCCAAGTG
Tsy TCTTTCCTTT*****TA TTT**********A AAACCTTC GG*G AAA AGAA * GCCA TGA AA GA T TAAACAGGC T TATTATTAGATCAGTAGAAAGA
Gcr TCTTTCCCTTTTCTCTA TACATTTTTTTTT**AAA CAT**TGT TGAG AG* AAAA T GTCA GGA GA AC T TAGGCAAGC C TGCT***GGCATGATAAAAGGG
Lfu TCTTTTCTTTTCTTTTC TTTTTTTTTTTTTT*AAC AACCTTGC GGAG AA* AAAA * GCCA GGA GA AC T TAGGCAGGT C TGTT***GACTCAGTAAAAGGA
Cme ****TTCTTTTCTTTTT TTCATTTTTTTTT**AAC AACCTTGC GGAG AA* AAAC * GCCA TGA GA AC T TAGGCAGGT C TGTT***GGCTCAGTAAAAGGA
Ocu TTTCTCCCCCAACTCAC **CCTTAAAAAAGTTTAA ACCCTAA* GAAG AAA GGAG * ACAA GGG GA AA T AGGGCAGCC T AG**********GC
1200
Hsa GGGAACTTTGGTGGCCAAACATACATTGCTAAGGC*TATTCCTATA*TCAGCTGGACAC**ATATAAAATGCTGCTAATGCTTCATTACAAA*****CTTATATCCTTTAATTCCAGATG
Ppy GGGA CTTTG T GCCAAACATATAT CTA G * AT C G *TCA C GG CAC**ATAT GCT T ATGCTTCA ACA *****CT TA CC TT AT GATG
Mmu GGGA CTCTG T ACCAAACAGAGTC *** G * AT C A *TCA C GG CAC**ATAT GCC T TCACTTCA ACA *****CT TA CC CT AT GATG
Age GAGA CTTTG T GCCAAACATATAT CTA G * AT C G ATTA C GG CAC**ATAC GCT A ATGCTTCA ATA *****CT CA CC AT AT AATG
Tsy AAGA CCTGA T GCCAAATGTGTGT GTA G * GT C A *TTG C TG CAC-Alu-T ATT T ATGCTTCC ACA *****CT TA TC AT AC GGAG
Gcr AGAA ACTAG G ATGAAATACATGT CTA G * AG T A *TTG C AG TACTTACAT CCT T AAGCCTTC TCA GGCTTAT CA CC CT AT GGAA
Lfu AGAC CCTAG T GCCAAACATATGT CTA G * AT T G **CT T AG CATGTATAT GCT T AACTCTCT TCA *****CT CT CC AT AT GGAG
Cme AGAA CCTAG T GCCAGACATATAT TTA T * AT T A *TTG C GG CGTGTATGT GCT T AACTCTCA TCA *****CT CA CC AT AC GGTG
Ocu *GAG CTTAC T ATTTTGTGTAGAG CC* T A AT C C *ATG C A* TACTTGTTT GCT * ATGCC*CA TCT *****CC AA TA AC CT GATT
1320


Hsa GGGGCAAA*GTATGT*CCAGGGGTGAGGAACAATTGAAACATTTGGGCTGGAGTAGATTTTGAAAGTCAGCTCTGTGTGTGTGTGTGTGTGTGTGCGCGCGTGTGTTTGTGTGTGTGTGA
Ppy GG GCAAA*GTG G *CCAG GGT GGAACAATTGAAA TTTGGGCTGGAGTA A TGAA GTC G******GTGTGTGTGTGTG**********TGTGTGTGT
Mmu GG GCAAA*GTG G *CCAG GGT GGAACAATTGAAA TTTGGGCTAGAGTA A TGAA GCC GGGGTGTGTGTGTGTGTGTGTGTGTGTATGT*********GTGTGCGTGT
Age GG CAAAA*GTG T *CTGG GGT GAAAGAATAGAAA TTTGTCCTGGAGTA A TTT* GTC GTTGCGAGTGTGTGTATGTATGTGTG*****TTTTTTTTGTGTGTGTGTGC
Tsy AA GCAAA*ATG C *CTGA AAA GAAATAATTGGAG T**********ACA G GAAA ATG GC**********TGTGAGGTT
Gcr GG GAAGAAGTG C GTTGA GAA ACAAC*******A CT******AGAACG G AGAA GAG GCTGTGTG**********
Lfu GG GAAAA*GAG C *TTGA AAA GAAACAACTGGAA CA**********G G AGAA GAG GCTGTGAG**********
Cme GG CAAA**GAG C *CTGA AAA GAAACAACTGGAA CA**********G G GGAA GAG GCTGTGAG**********
Ocu GA GATTATGTG C *CTGG AGA A****TATTGGAA TTTGGTCCGAAGTG T GAAA GAT CC*GTGAG**********
--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---- 1440
Hsa GAGCGTGTGTTTCTTTTAACGTTT*TCAGCCTACAGCATACAGGGTTCATGGTGGCAAGAAGATAACAAGATTTAAATTATGGCCAGTGACTAGTGCTGCAAGAAGAACAA*CTACCTGC
Ppy GAGCATG GTTTCTTTTAACATTT* GCCTACAGC TAAAGG T CATG GGC AG AGATGGCA GAT TAAATTA G C G ACT G GCTGCA GGG ACGA*C A C GC
Mmu GCGCGTG GTTTCTTTTAACGTTT* GCCTACAGC TACAGG T TGTG GGC AG AGACAGCA GTT TAAATTT G C G ACT G GCTGCA GGG ACAA*C A C GC
Age GAGCATG GTTTCTTTTAAAGTTT* GCCTACAAA TACAGG T TGTG AGC AG AGATAGCT GAT TAAATTA * C G ACT A GCTGCA GGG ACAG*C A C GC
Tsy TTGTTGT GTTGTTTTTATTAG*** GTCAGTAAA AATGAG T GTTG GAT AG AGATAGAA GAT TAAGTCC G C A ATG G TCTGCA GGA ATAA*C T C GT
Gcr *****GT TTTAAAAAAAATATTTA ACACACAAA TGTGGG C GTTG TGT AG AAATGGAA GAC C*****A G C G GTC G *CTGCC GGG AAAATC T T GT
Lfu ****ATT TTTAAAAAATAT*TTT* GCCCTCAAA TGTGGA C GTTG GGT AG AGGTAGAA GAT TAAATCA G T G ATT G GCTGCA GGG ACAAAT A C CT
Cme ****ATT AAAAAAAAAAATATTT* GCCCTTAAA TGTGGA C GTTG GGT AT AGGTAGAA GAT TAAATCA G T A ATT G GCTACA GAG ACAAAT A C GT
Ocu ******G GTTACCTTTTCCCC*** GCCCAC*AA TGGGGA T GT** TAG TG TGACAGAA TAT TAAATTA A C A ATT G TTGGAC GGG CCAAAC A T AC
--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---- 1560
Hsa ATTTAATGGGAAAGCA*****AAATCTCAGGCTTTGAGGGAAGTTAACATAGGCTTGATTCTGGGTGGAAGCTTGGTGTGTAGTTATCTGGAGGCCAGGCTGGAGCTCTCAGCTCACTAT
Ppy A TT TGGGAAG CA*****A ATCTC GCT TGAGGG G AACATAGGCTTGA TCT G TGGA GCTAG G GT GTTATCTGG GG CAGG G G T AGC C CTAT
Mmu A TT T*GGAAG CA*****A ATCTC GCT TGAGGG G AACATAGGCTTGA TCT G TGGA GCTGG G GT GTT*TCTGG GG CAGG G G T AGC C CTAT
Age A TT TGGGTAG CA*****A ATC*C GCT TGAGGG G AACATAGGCTTGA TCC G TGGA GGTGG G GT GTTATCCAG GG CAGG G G C TGT C CTAT
Tsy A TT TGGTTAG CA*****A GCC*C GCC TGAAAG G GACATAGGCTTGG TCT A TGGG GTCA* G CC AATACCAGG AG CT*G G G T TGC T CAAC
Gcr G TG TGGCTAA CA*****A ACCTC GCG TGTGAG G AACACAGG***** CCT A TACA GTTGG T AA GATACATAA GT TACC G C T TGC C CTTT
Lfu G TT TGGCTAG CA*****A ACCTG GGT TATGAG G ********TTTGG TCT A TGGG ATTGG G CC GATACCTGG GA CAGC T C T TGC C TTTT
Cme G TT TGGCTAG CT*****A ACCTC CGT TGCAAG G GACATAGGCTTGG TCT A TGGC GTTGG G CC GA*ACCTGG GA CAGC T C T TGC C CTTT
Ocu A GT CAACTGG AGGGGGGG AGATC CCT GGAGGT A AACAAAGGCTTGG TTT A GGGT *TTCA G CC GATAGCGGG GA TAGC * A T TAT A CTAT
--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---- 1680

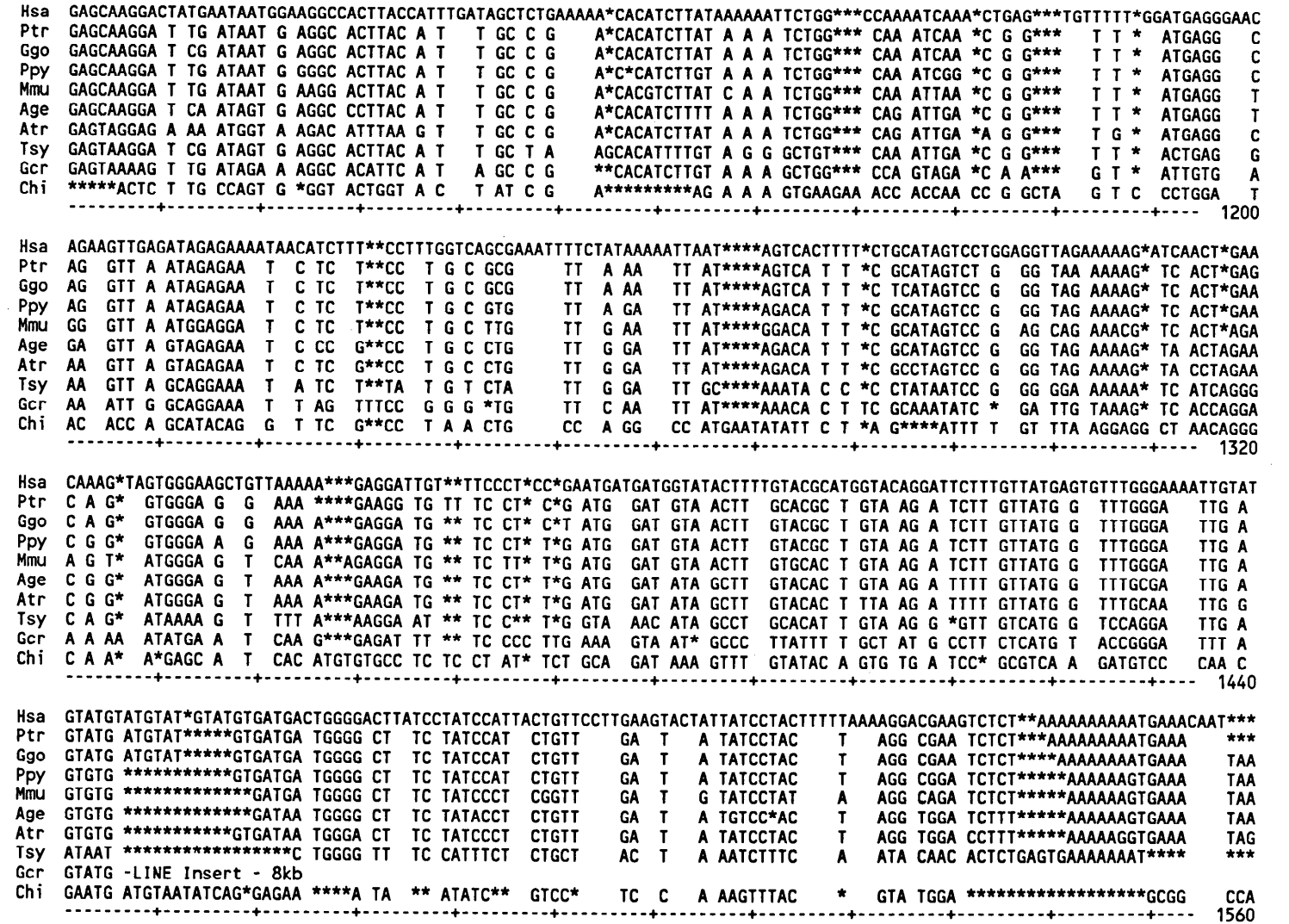

EXON 3 ---->
Hsa GGGTTCATCTTTATTGTCTCCTTTCATCTCAACAGCTCCTGGGAAATGTGCTGGTGACCGTTTTGGCAATCCATTTCGGCAAAGAATTCACCCCTGAGGTGCAGGCTTCCTGGCAGAAGA
Ppy GGGT CA C T G TG TCC TTC T CA C G T T G G CC T AATCCA TTT AA C TG GG G GC T C G C GA
Mmu GGGT CA T T G TG TCC TTC T CA C G T T G G CT T AATCCA TTC AA C TG AG G GC T C G C GA
Age GGGT TA A T G TG CCC TTC A CA C G T T G G CT T AATCCT CAT AA C TG GG G GC T C G C GA
Tsy GTGT CA C C G TT TTG TCC C CC C G C T G T TT C ACATCA TTC GG T CC GG G GC G A G A GC
Gcr GTTC CA C C G TG TCT ACC C AC C G T T G G TT C AAAGCA TTT AG C AC GA A GC G C C C GA
Lfu TTTC CA C C A TG TCT TCC C TC C A T A G G TT C AAAGTA TTT AG C GC GG G GC G C G C GA
Cme GTTC CA C C G TG TCT TCT C TC C G C A G G TT C AAAGCA TTT AG C AC GG G GC G C G C GA
Ocu GTGT TG G T G CT TCT TTC C CC T A T T A A TT C GAAGTA TTT AG T CC GG C AG G C G C AT
1800
----->TER
Poly A
Hsa TGGTGACTGGAGTGGCCAGTGCCCTGTCCTCCAGATACCACTGAGCTCACTGCCCATGATGCAGAGCTTTCAAGGATAGGCTTTATTCTGCAAG***CAATACAAA***TAATAAATCTA
Ppy G GA GA G C TCCTCC GA C GCC CC GCCCATGATGC A CT TC ACG TAAGCT TATTCTG G***C A CAAA***T T TG
Mmu G GG GA G C TCCTCC GA C GCC CC GCCCACGATGC A CT TC AGG TAGGCT TATTCTG G***C A AAAA***T T TA
Age G GG GA G T GCCTCC GA C AGC CC GCCCATGATGC A CT TC AGG GTGGCT TATTCCG G***C A AAAA***T A TA
Tsy A GT CA T A GGACAC AA T GCC CC GCCCCTGATAT A CC TC GGG CAGGCT TATTCTT G***C G ATAACAAC A AA
Gcr G GG GT C C GCCCAC AG C CCC CA GCTCATGCTAC G AT TC AGG GAACCT CATTCTT GTTAT A AAA****T A T*
Lfu G TG GA T C GCCCAT AG C GCC CC GCCCATGGTGC A CT TC GGG GAACTG ******* G***T A AA*****T A TA
Cme G TG GA T C GCCCAT AG C GCT CC GCTCATGGTGC A CT TC GGG AAACTG ******* G***T A AA*****T A TA
Ocu G GG GA C C GCCCAC AG T ACC CT CACTGTAGGAC A CT CT GGA GAAGCT TATCCCT A***T A ATGAAAAT A TA
1920
Hsa TTCTGCTAAGAGATCACACATGGTT*GTCTTCAGTTCT**TTTTTTTATGTCTTTTTAAATATAT****GAGCCACAAAGGGTTTTATGTTGAGGGATGTGTTTATGTGTATTTAT*ACA
Ppy TTCTGCTAAGAGATCACACATGG T* CT CAGT CT**T TTT*A CTTTTT TA AT**** GCCAC A GTTTTATG TGA A G GTA GT G *ACA
Mmu TTCTGCTAAGCGATCACACATGA T* CT CAGC CT**T TTT*A CTCTTT TA AT**** GCCAC A GTTTTATG TGA A G GTA AT G *ACA
Age TTCCGCTCAAAGATCACACGTGA T* CG CAGT AT**T TTCCT CCTTCC TA GC**** ACCAC A GTTT*ATG TGA<
Tsy TTCTGCTAAGAAAATACGCACGA T* CT TACT CTTTT TCCCA CCTTTT TA AT**** ATCCC A ATACTATG AGA G A GG* GG G GGTA
Gcr ****GTTATGAAGTTGCACATTT TA CT TGGT C***T TCCTG TCTCTT CA ATATAT GCCC* G GCTTTATG TGA * G GTG GT G GT**
Lfu TTCTGTTATGAAGT**CACATGT T* CT CAGT CTTCT CCCTG TCTTTT TG AT**** ACCTC G CTTTGATG TGA<
Cme TTCTGTTATGAAGTTGCACACGT T* CT CATT CTTCT CTCTG TCTTTT TG AT**** GCACC G CTTTTACG TGA<
Ocu CTCT***AAGAAATTATTTGTGA G* AT GAGT T***A TTCCT ACTTTT TA AT**** TCCTC G ATTTATTT TTT<
2040
Hsa TGGCTATGTGTGTTTGTGTCATGTGCACACTCCACACTTTTTTGTTTACGTTAGATGTGGGTTTTGATGAGCAAAT
Ppy TGGCTATGTGTGTTTGT CATGTGT T C CACACT TT G TTA ATG G GGT GA GA CA AT
Mmu TGGCTGTGTGTGTTAGT CATGTGC T * CTCACT TT G TTA ATG G GGT GA GA CA TT
Tsy TGTGTGTATATGTGTGT GTGGTGC T T TTCATT AT T G** ATG G ATT TT AG AG AA
Gcr **********G GTACCCG T * CA**TC AG T G** ACG C CAC TA GA AA AA
2116
Tsy Alu GTGCACATACTGCTAGGGCCAGACATGGAGGCTCATGCCTGTAATCCCAGCAATTTGGAAGACTGAGGCAGGAGGATTGCTTTGGCTCAGGAGTTCAAGACTATCCTGGACAGCA
TAGCAATACCTCCACTCTACAAAAAATATAAAAATCAGCAGGTGTGATGGCATTTGCTTGTAATCCCAAGCTACTTGGGAGGCTGAGGTCGGAGGATCACCTGAGCCCAGGAGGT
CAAGACTGTGGAGAGCTGTAATTGTGCCATTGCACTCCAGTCTGGATAACAGAGTGAAACTCCAATTCAAAATAAAATAAAGTAAAA

ALIGNED ETA SEQUENCES

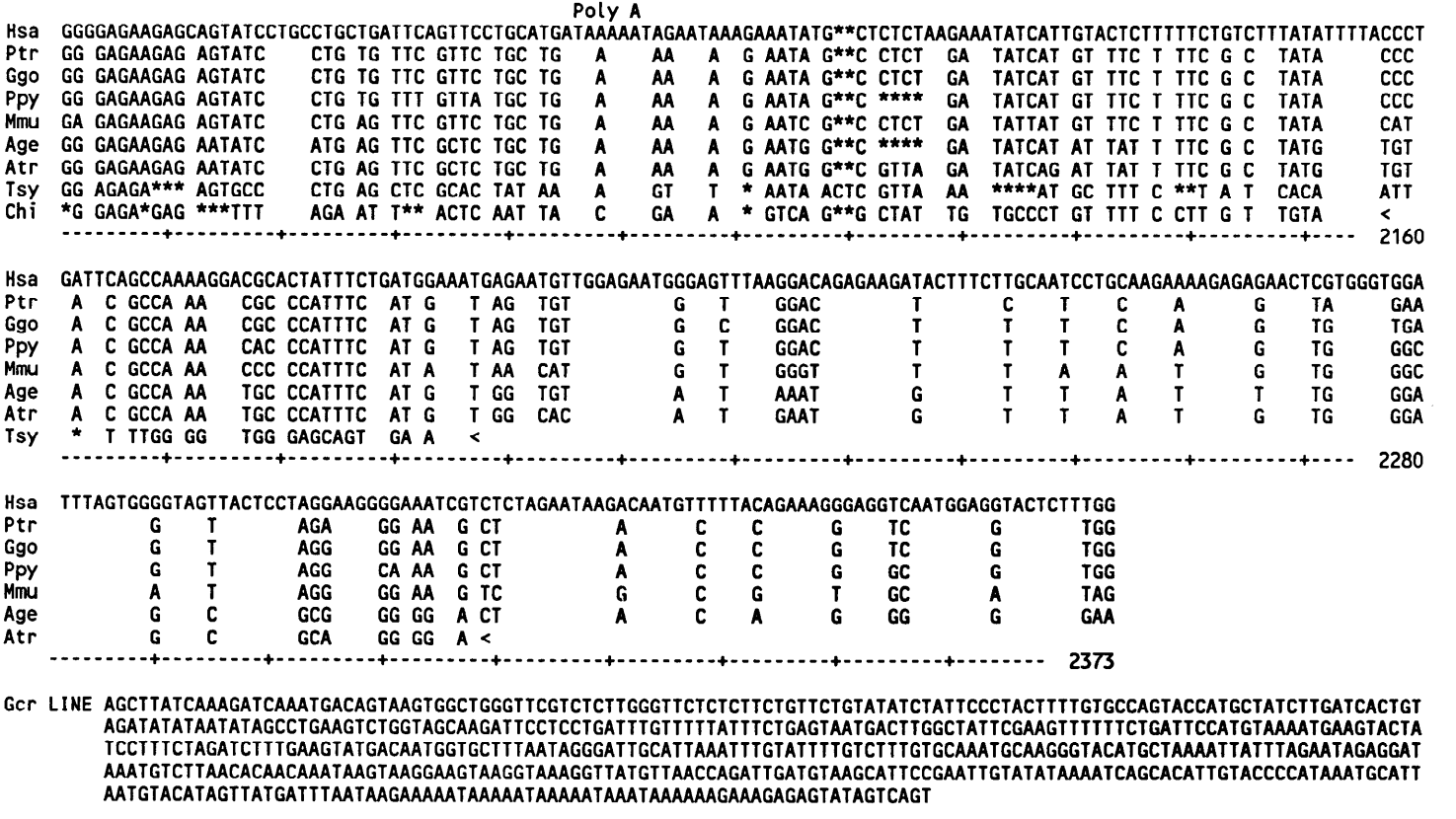



INIEXON 1 ---->
Hsa GCAATGCAGGCATGCTTCTGGCTCATCTGT*GATCACCAGGAAACTCCCAGATCTGACACTGTAGTGCATTTCACTGCTGACAAGAAGGCTGCTGCCACCAGCCTGTGAAGCAAGGTTAA
Ptr GCGATG GCAT C CTGG T CTGT* A A AGGAAAC CCAGATCTG CACTGTA G T C GCT CAAGAAG GCT CCACC GCC GTGAAG AAGG TA
Ggo GCAATG GCAT T CTGG T CTGT* A A AGGAAAC CCAGATCTG CACTGTA G T C GCT CAAGAAG GCT CCACC GCC GTGAAG AAGG TA
Ppy GCAGTG GCAT C CTGG T CTGT* A A GGGAAAA CCAGATCTG CACTGTA G T C GCT CAAGAAG GCT CCAAT GCC GTGAAG AAGG TA
Mmu GAAGTA GTAT C CTGG T CTAC* A G AGGAAAC CCAGATCTG CACCGTG G T C GCT CGAGAAG GCT CCACT GCC GTGAAG AAGT T*
Age GCAGTG ACAT C CTGG T CTGT* A A AAGAAAC CCAGACCTG CACTGTG G T C *CT CAAAAGG GCT CCACT GCC GTGGAA AAGG TA
Atr GCAGTG ACAT C CTGG T CTGT* A A AGGAAAC CCAGACCTG CACCGCG G T C GCC CGAAAGG GCT CCACT GCC ATGGAG TAGG TG
Tsy GCAGCA GCAT C TCTG T CTGT* A T AGTGGAT CCAGACCTG CACTATG A A C GCT GAAGGCT G** TTGCT GCT GTGGGG AAGG GA
Gcr ACAAAA GCAA T CTGG T GTGT* G A AGTGAAT T*GGACCC* AAGTCAG G T C GTT GGAGAAG GCT TTACT ACC TCTGGG AGGT GA
Lfu GCAGTA GTGA C CTAA T CCGT* G A AGCAGAC GCAGACCTG CGCTGTG G T C GCA GGCAAAG *CT CGGCT GCC GCCAGG AACA GA
Chi GCGGCA ACTT C CTGG C *TATG A A AGTAAGC CCA*****G CACCATG G T T ACC GGAGAAG GCT TTGCT GTC GTGGGC AAAG GA
480
---->|
Hsa GGTG*AGAAGGCTGGAGGTGAGATTCTGGGCAGGTAGGTA***CTGGAAGCC**********GGGACAAGGTGCAGAAAGGCAGAAAGTGTTTCTGAA*AGAGGGATTAGCCCGTTGTCT
Ptr GG G* AA GCT GA GTG G TT GGGC T GGTA***CTG AAGCC**********G GGCAAGGTGC GAAA CA AG T T * AT AGCCCGTTGTCT
Ggo GG G* AA GCT GA GTG G TT GGGC T GGTA***CTG AAGCC**********G GGCAAGGTGC GAAA CA AG T T * AT AGCCCATTGTCT
Ppy GG G* AA GCT GA GTG G CT GGGC T GGTA***CCG AAGCC**********G GGCAAGGTGC GAAA CA AG T T * AT AGCCCGTTGTCT
Mmu GG G* AA GCT GA GTG G TT GGGT T GGTA***CTG AAGCC**********A GGCGAGGTGC GAAA CA TG T T * AC AGCCCATTGTCT
Age GG G* AA GCT GA GTG G TT GGGC T GCTA***CTA AAGCC**********A AGCAAGGTGT GAAA CA AG C T * AT AGCCAGTTGTCT
Atr TG G* AA GCT GA GTG G TT GGGC T GCTA***CTA AAGCC**********A AGCAAGGTGC GAAA CA AG C T * AT AGCCAGTTGTCT
Tsy CG GG GA ATT GA GCG G TC **** T GGTATACCTA GAGGTTGCACATTGGA GTTGGGA**C GG** CA AA C T G AC ATCCAATTATCT
Gcr CA AG GA GTT TA ATA A CC GGGC C T********TG AGGCC********TGGA GTTGGGGTG* GAAA TA CA C T * AA A-LINE-AA*GT
Lfu TG GG GA GCT GA GCA G TC GGGC C GGCA***CTG AGGCC**********A GGTCAGGGGC GAAA CA AG C T * GA AGCCAGTTATCC
Chi TG GG GT GTC GC GTG G GC AGCA T AGCA***GTG A*CAC*****AGGTAG AGAGGAGTGT CAAA CT AG C A A AC GGTTAGGTTTCT
600
EXON 2 ---->
Hsa TACATAGT*****CTGACTTTGCACCTGCTCTGTGATTATGACTATCCCACAGTCTCCTGGTT****GTCTACCCATGGACCTAGAGGTACTTTGAAAGTTTTGGA*TATCTGGGCTCTG
Ptr C TAG *****C AC TTGCACC CTCTGTGATTATGACTATCC ACAGTC CC G T****GTCTACC AAC T G GGTAC TGAA GTT A*T TC GGGC
Ggo C TAG *****C AC TTGCACC CTCTGTGATTATGACTATCC ACAGTC CC G T****GTCTACC GAC T G GGTAC TGAA GTT A*T TC GGGC
Ppy C TAG *****C AC TTGCACC CTCTGTGATTATTACTATCA ACAGTC CC G T****GTCTACC GAC C G GGCAG TGAA GTT C*T TC GGGT
Mmu C TAG *****C AC TTGCATC CTCTGTGATTATGACTGTCC TCAGTC CC G T****GTCTACC GAC C G GGTAC TGAA CTT C*T TC GGGC
Age C TAT *****C AC TTGCATC CTGTTTGATTATGACTATCC ACAGTC CC G T****GTCTAGC GAC T G GGTAC TGAA GTT T*A TC GGGC
Atr C TAT *****C AC TTGCATC CTCTTTGATTATGATTATCC ACAGTC CC G T****GTCTACC GAT C G AGTAC *GAA GTC C*A TC GGGC
Tsy T TAC *****C CA ATG*TTC C**********CC ATAGGC TC G T****GTCTTCC GAC C * GATTC TGAC GTT C*A CT GTCC
Gcr T TTC ATAATC AC TTGCATT TTTAGTGAGCACCACCGCTC ATCAGT CT G TATTT****ACT GTC C G GGTTC TGAT A** GAA CT GTCC
Lfu T CAG *****A AC TTGCATC TTTTGTG***ACGACTGCCC ATAGGT CT C TGTTTGTCTATT TTC C G AGTTC CAGT ATT GGA TT GTCC
Chi C TAC *****C AC TCTTATC TTCTGTGACTATGATCATCC ATAGGC CT A C****GTCTACC GAC C G GGTTC TGAC GTT T*A CT ATGC
720

```
Hsa  ACTGTGCAATAATGGGCAACCCCAAAGTCAAGGCACATGGCAAGAAGGTGCTGATCTCCTTCGGAAAAGCTGTTATGCTCACGGATGACCTCA***AAGGCACCTTTGCTACACTGAGTG
Ptr  AC G  CAA A TG GC  CCCC AA     GG AC T GC A   GTG   TC CCT TGG   A CTG   TGCTCACGG TGACCTCA*** A G ACC    CTACA  G GT
Ggo  AC G  CAA A TG GC  CCCC AA     GG AC T GC A   GTG   TC CCT CGG   A CTG   TGCTCACGG TGACCTCA*** A G ACC    CTACG  G GT
Ppy  AC G  CAA A TG GC  CCCT AA     GG AC T GC A   GTG   TC *** CGG   A CTG   TGCTCATGG TGACCTCA*** A G ACC    CTACA  G GT
Mmu  AC G  CAA A TG GC  CCTC AA     GG AC T GA A   GTG   TC CCT CGA   A CTG   TGCTCATGG TGACCTCA*** A G ACC    CTATG  G GT
Age  AC G  TAA A TG *C  CCCC AA     GG AC T G* A   CTG   TC CCT CAG   A CTG   TGCTCTTGT TGACCTCA*** A G ACC    CTATG  G GT
Atr  AC G  TAA A TG *C  CCCC AA     GG AC T *C A   GTG   TC CCT CAG   A CTG   TGCTCATGG TGCTCATAGTA A G ACC    CTATA  G GT
Tsy  AC C  CAA C TG GT  CCCC GA     AG TC T GC A   GTG   CT TTT TGG   A TTG   TGCACATGG TGAACTCA*** A A GCT    CTAAG  G GT
Gcr  AC C  TGG A CA AC  TTCT AG     GT TC T GA A   CCA   CC CTT TGA   * ATG   CACATAGGG ACACATCT*** G G AAC    ATAAG  G AG
Lfu  GC C  TGC A TG CC  CCCC AG     GT TT T GC A   CAA   CC *CT TGG   A ATG   CGTGCACTG TGATCTCA*** A G AAC    CTGAG  G GT
Chi  AG C  CCA A TG GC  CCCC AG     GG CC C GC G   GTG   AC CCT TGG   T CCA   AGCACATGG TGATCTCA*** G G ACC    CAGAT  A GC
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+----  840

                                              ---->|
Hsa  ACCTGCACTGTAACAAGCTGCACGTGGACCCTGAGAACTTCCTGGTGAGTAGTAAGTACACTC**ACGCTTTCTTCTT****TACCCTTAGATATTTGCACTATGGGTACTTTTGAAAGC
Ptr   CCGGC C   A C   C GCATG GG CCC GA  A  TCCTGGT  G AC AAGT CACTC**ACGC   CT CT ****TACCCTT GA A TT CACTA   GTAC  T GA AGC
Ggo   CCTGC C   A C   C GCACG GG CCC GA  A  TCCTGGT  G AC AAGT CACTC**ACAC   CT CT ****TACCCTT GA A TT CACTA   GCAC  T GA AGC
Ppy   CCTGC C   A C   C GCACG GG CCC GA  A  TCCTGGT  G GC AAGT CACTC**ACGC   CT CT ****TACCCTT GA A TT CACTA   GTAC  T GA AGC
Mmu   CCTGT C   A C   C GCACG GG CCC GA  A  TCCTGGT  G AC AAGT TACTC**ATGC   CT CT ****TACCCTT GC A TT CACCA   GTAC  T GA AGC
Age   CCTGC C   G C   C ACACA GG CCC GA  A  TCCTGGT  G TC AAGC CACTC**ATGC   CT CT ****TCCCCTT AA A TC CACGA   CTAC  T GA AGC
Atr   CCTGC C   G C   C ACACA GG CCC GA  A  TCCTTGA  A TC AAGC CACTC**ATGC   CT CC ****TCCCCTT AA A TT CACGA   CTAC  T GA AGC
Tsy   GCTTC C   A T   C GCACG TG TTC GA  A  TCAGGGT  G TC GAGC CAATGTAATGA   CT TC TTTTTCCCCTT GA C TT TGCCA   TTAC  T GG AAC
Gcr   GTTAC G   G C   C GAGTG GA TTT GA  G  TTCCATT  G TC GAAC CACTT**ATGC   AC CT *********** GG G CT TGTCA   ATAC  T GA ACC
Lfu   GCTGC C   G C   T GCATG GG TCC GA  A  GCAGG&lt;
Chi   GCTGC C   G C   T GCACG GG TCC CC  A  TCAGGGT  G TC GGGT TGCCT**GTGC   GT CT ****TCATC*C GG T CA TGCTG   ****  A *A TAG
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+----  960

Hsa  AGA**GGTGGCT*TTCTCTTGTGTTAT**GAGTCAGCTATGGGATATGATATTTCAGCAGTGGGAT****TTTGAGAGTTATGTT***GCTGTAAATAACATA*ACTAA**AATTTGGTA
Ptr    A**GGT GCT*TTC    GTGTTA **GA  C GC A G  ATA GA    CAGCAGTGG AT****T TGAG GT AT   *** CTGTA  TAACATA*A TA **A   TGGTA
Ggo    A**GGT GCT*TTC    GTGTTA **GA  C GC G G  ATA AA    CAGCAGTGG AT****T TGAG GT AT   *** CTGTA  TAACATA*A TA **A   TGGTA
Ppy    A**GGT GCT*TTC    GTGTTA **GA  C GC A G  ATA GA    CAGCAGTGG AT****T TGAG GT AT   *** TTGTA  TAACATA*A TA **A   TGGTG
Mmu    A**GGT GGT*TTC    GTGCTA **GA  C GC A G  ATA GA    CAGCAGTGG ATCGTTT TGAG GT AT   *** CTCTA  TAACATA*A TG **T   TGGTA
Age    A**GGC GCT*TTC    GTGCTA **GA  C GC G G  **A GG    CAGTGATGG ATCGAAT TGAG GT AT   *** CTCTA  TAACATC*C TA **A   TGGTG
Atr    A**GGT GCT*TTC    GTGCTA **GA  C GC G G  ATA GA    CAGTGTTGG ACTGATT TTAG AT AG   *** CTCTA  TGATGTA*A TA **A   CAGTA
Tsy    G**GGT TTC*ATC    GTATTA AGGG  T GC G A  ATA GA    TGGCAATGG ATAGATT TGAG GT AT   *** GTCCA  TAGCATG*C TA **A   TGGTA
Gcr    A**GGT TCC*ATC    ATTTCA AGGA  T GC G G  ATA GA    TGATGATGA GTTGATT GGAG CT AG   *** CTCTA  TGGCACA*G CA **A   TGGTA
Chi    AACACT ACTCAAG    AAGCTG TGAA  * TT T T  AGG AA    CAGAGAAAG CTTGATC TGGT GG *T   CCA AGGTC  *AGGTTAGA CA GTA   *****
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+----  1080
```


Hsa GAGCAAGGACTATGAATAATGGAAGGCCACTTACCATTTGATAGCTCTGAAAAA*CACATCTTATAAAAAATTCTGG***CCAAAATCAAA*CTGAG***TGTTTTT*GGATGAGGGAAC
Ptr GAGCAAGGA T TG ATAAT G AGGC ACTTAC A T T GC C G A*CACATCTTAT A A A TCTGG*** CAA ATCAA *C G G*** T T * ATGAGG C
Ggo GAGCAAGGA T CG ATAAT G AGGC ACTTAC A T T GC C G A*CACATCTTAT A A A TCTGG*** CAA ATCAA *C G G*** T T * ATGAGG C
Ppy GAGCAAGGA T TG ATAAT G GGGC ACTTAC A T T GC C G A*C*CATCTTGT A A A TCTGG*** CAA ATCGG *C G G*** T T * ATGAGG C
Mmu GAGCAAGGA T TG ATAAT G AAGG ACTTAC A T T GC C G A*CACGTCTTAT C A A TCTGG*** CAA ATTAA *C G G*** T T * ATGAGG T
Age GAGCAAGGA T CA ATAGT G AGGC CCTTAC A T T GC C G A*CACATCTTTT A A A TCTGG*** CAG ATTGA *C G G*** T T * ATGAGG T
Atr GAGTAGGAG A AA ATGGT A AGAC ATTTAA G T T GC C G A*CACATCTTAT A A A TCTGG*** CAG ATTGA *A G G*** T G * ATGAGG C
Tsy GAGTAAGGA T CG ATAGT G AGGC ACTTAC A T T GC T A AGCACATTTTGT A G G GCTGT*** CAA ATTGA *C G G*** T T * ACTGAG G
Gcr GAGTAAAAG T TG ATAGA A AGGC ACATTC A T A GC C G **CACATCTTGT A A A GCTGG*** CCA GTAGA *C A A*** G T * ATTGTG A
Chi *****ACTC T TG CCAGT G *GGT ACTGGT A C T AT C G A*********AG A A A GTGAAGAA ACC ACCAA CC G GCTA G T C CCTGGA T
1200
Hsa AGAAGTTGAGATAGAGAAAATAACATCTTT**CCTTTGGTCAGCGAAATTTTCTATAAAAATTAAT****AGTCACTTTT*CTGCATAGTCCTGGAGGTTAGAAAAAG*ATCAACT*GAA
Ptr AG GTT A ATAGAGAA T C TC T**CC T G C GCG TT A AA TT AT****AGTCA T T *C GCATAGTCT G GG TAA AAAAG* TC ACT*GAG
Ggo AG GTT A ATAGAGAA T C TC T**CC T G C GCG TT A AA TT AT****AGTCA T T *C TCATAGTCC G GG TAG AAAAG* TC ACT*GAA
Ppy AG GTT A ATAGAGAA T C TC T**CC T G C GTG TT A GA TT AT****AGACA T T *C GCATAGTCC G GG TAG AAAAG* TC ACT*GAA
Mmu GG GTT A ATGGAGGA T C TC T**CC T G C TTG TT G AA TT AT****GGACA T T *C GCATAGTCC G AG CAG AAACG* TC ACT*AGA
Age GA GTT A GTAGAGAA T C CC G**CC T G C CTG TT G GA TT AT****AGACA T T *C GCATAGTCC G GG TAG AAAAG* TA ACTAGAA
Atr AA GTT A GTAGAGAA T C TC G**CC T G C CTG TT G GA TT AT****AGACA T T *C GCCTAGTCC G GG TAG AAAAG* TA CCTAGAA
Tsy AA GTT A GCAGGAAA T A TC T**TA T G T CTA TT G GA TT GC****AAATA C C *C CTATAATCC G GG GGA AAAAA* TC ATCAGGG
Gcr AA ATT G GCAGGAAA T T AG TTTCC G G G *TG TT C AA TT AT****AAACA C T TC GCAAATATC * GA TTG TAAAG* TC ACCAGGA
Chi AC ACC A GCATACAG G T TC G**CC T A A CTG CC A GG CC ATGAATATATT C T *A G****ATTT T GT TTA AGGAGG CT AACAGGG
1320
Hsa CAAAG*TAGTGGGAAGCTGTTAAAAA***GAGGATTGT**TTCCCT*CC*GAATGATGATGGTATACTTTTGTACGCATGGTACAGGATTCTTTGTTATGAGTGTTTGGGAAAATTGTAT
Ptr C A G* GTGGGA G G AAA ****GAAGG TG TT TC CT* C*G ATG GAT GTA ACTT GCACGC T GTA AG A TCTT GTTATG G TTTGGGA TTG A
Ggo C A G* GTGGGA G G AAA A***GAGGA TG ** TC CT* C*T ATG GAT GTA ACTT GTACGC T GTA AG A TCTT GTTATG G TTTGGGA TTG A
Ppy C G G* GTGGGA A G AAA A***GAGGA TG ** TC CT* T*G ATG GAT GTA ACTT GTACGC T GTA AG A TCTT GTTATG G TTTGGGA TTG A
Mmu A G T* ATGGGA G T CAA A**AGAGGA TG ** TC TT* T*G ATG GAT GTA ACTT GTGCAC T GTA AG A TCTT GTTATG G TTTGGGA TTG A
Age C G G* ATGGGA G T AAA A***GAAGA TG ** TC CT* T*G ATG GAT ATA GCTT GTACAC T GTA AG A TTTT GTTATG G TTTGCGA TTG A
Atr C G G* ATGGGA G T AAA A***GAAGA TG ** TC CT* T*G ATG GAT ATA GCTT GTACAC T TTA AG A TTTT GTTATG G TTTGCAA TTG G
Tsy C A G* ATAAAA G T TTT A***AAGGA AT ** TC C** T*G GTA AAC ATA GCCT GCACAT T GTA AG G *GTT GTCATG G TCCAGGA TTG A
Gcr A A AA ATATGA A T CAA G***GAGAT TT ** TC CCC TTG AAA GTA AT* GCCC TTATTT T GCT AT G CCTT CTCATG T ACCGGGA TTT A
Chi C A A* A*GAGC A T CAC ATGTGTGCC TC TC CT AT* TCT GCA GAT AAA GTTT GTATAC A GTG TG A TCC* GCGTCA A GATGTCC CAA C
1440
Hsa GTATGTATGTAT*GTATGTGATGACTGGGGACTTATCCTATCCATTACTGTTCCTTGAAGTACTATTATCCTACTTTTTAAAAGGACGAAGTCTCT**AAAAAAAAAATGAAACAAT***
Ptr GTATG ATGTAT*****GTGATGA TGGGG CT TC TATCCAT CTGTT GA T A TATCCTAC T AGG CGAA TCTCT***AAAAAAAAATGAAA ***
Ggo GTATG ATGTAT*****GTGATGA TGGGG CT TC TATCCAT CTGTT GA T A TATCCTAC T AGG CGAA TCTCT****AAAAAAAATGAAA TAA
Ppy GTGTG **********GTGATGA TGGGG CT TC TATCCAT CTGTT GA T A TATCCTAC T AGG CGGA TCTCT*****AAAAAAGTGAAA TAA
Mmu GTGTG **********GATGA TGGGG CT TC TATCCCT CGGTT GA T G TATCCTAT A AGG CAGA TCTCT*****AAAAAAGTGAAA TAA
Age GTGTG **********GATAA TGGGG CT TC TATACCT CTGTT GA T A TGTCC*AC T AGG TGGA TCTTT*****AAAAAAGTGAAA TAA
Atr GTGTG **********GTGATAA TGGGA CT TC TATCCCT CTGTT GA T A TATCCTAC T AGG TGGA CCTTT*****AAAAAGGTGAAA TAG
Tsy ATAAT **********C TGGGG TT TC CATTTCT CTGCT AC T A AATCTTTC A ATA CAAC ACTCTGAGTGAAAAAAAT**** ***
Gcr GTATG -LINE Insert - 8kb
Chi GAATG ATGTAATATCAG*GAGAA ****A TA ** ATATC** GTCC* TC C A AAGTTTAC * GTA TGGA **********GCGG CCA
1560

```
Hsa  *CACAATATGTTGGGG*TAGTGAGTTGGCATAGCAAGTAAGAGAAGGATAGGACACAATGGGAGGTGCAGGGCTGCCAGTCATATTGAAGCTGATATCTAGCCCATAATGGTG*******
Ptr  *CACAATA GCTGGGG*TAGTGA  TGG  T  C  G AAGAGA G   AGG  A  A   GA  GGCA GG  GCC     TAT   AG TA  AT TAGC CA     G  *******
Ggo  TCACAATA GCTGGGG*TAGTGA  TGG  T  C  G AAGAGA G   AGG  A  A   GA  TGCA GG  GCC     TAT   AG TG  AT TAGC CA     G  *******
Ppy  TCACAATA GCTGGGG*TAGTGA  TGG  T  C  G AAGAGA G   AGG  A  A   GA  TGCA GG  GCC     TAT   AG TC  GT TAGC CA     G  *******
Mmu  TCACACTG GCTGGGG*TAGTGA  TGG  T  C  G AAAGGA G   AGG  A  A   GA  TGCA CG  GCC     TAT   AG TG  GT TATC CA     G  AACTTAC
Age  TCACAATG GCTGTGG*TAGTGA  TGG  T  C  G AAGAGA G   AGG  A  A   AA  TGCA GA  GGC     TAT   AG TG  GT TAGC CA     G  TACTTAC
Atr  TCACAATG GCTGTGG*TAATGA  TGG  T  C  G AAGAGA G   AGG  A  A   AA  TGCA GG  GGC     TAT   AG TG  GT TAGC CA     G  TACTTAC
Tsy  *CTCAATG T*******CAGTGA  TAG  *  C  G AAACAG A   TGA  C  A   GA  TGTG GA  TGG     CAG   GG TG  GT AAGC CA     A  TACTTAA
Chi  TGAGCA*A GCTCAGGCTTGAAC  CGA  G  A  A GGGAGA G   GTA  A  G   G*  TTCA G*  GGG     **C   AT CA  GC AGGT TC     G  CACTT**
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---- 1680

Hsa  **********************AGAGTTGCTCAAACTCTGGTCAAAAAGGATGTAAGTGTTATATCTATTTAC*TGCAAGTCCAGCTTGAGGCCTTCTATTCACTATGTACCATTTTCTTT
Ptr  ********************** GA T GC   AACT  GG C  A A GATGTA G GT A A C   T AC* GC   TC  G T    G CTTC AT CAC A  T  CA T T
Ggo  ********************** GA T GC   AACT  GG C  A A GATGTA G GT A A C   T AC* GC   TC  G T    G CTTC AT CAC A  T  CA T T
Ppy  ********************** GA T GC   AACT  GG C  A A GATGTA G GT G A C   T AC* GC   TC  G T    G CTTC AT CAC A  T  CA T T
Mmu  CTTTGTGAGAATAAGACT**G GA T GC   AACT  GG C  A A GAAGTA G GT G A C   G AT* GC   TC  A T    G CTTC AT CAC A  T  CA T T
Age  CTTTGTGAGAAGAAGACTCAG GA T GC   AACT  TG C  A A AAAATA A GC G A T   T AT* GC   TC  G T    G CTAT AT CAC A  T  CA T C
Atr  CTTTGTGAGAATAAGACTCTG GA T GC   AACT  TG C  A A AAAATA G GT G A T   T AT* GC   TC  G T    G CTGT AT CAC A  T  CA T C
Tsy  TCTTGTAGGAGCAAGACCCTG GG T TT   TTTC  GG C  G A AAAATA G AT G G C   G AT* AC   GT  G T    G ACTC GC TTC G  G  CA T T
Chi  **TTGTGGGAACATGGCCATG AA C GC   CCCT  TA G  G G ACCGGC * GT G A C   T GCT GG   GG  G C    A CTTC CT CAT A  T  TG C T
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---- 1800

                          EXON 3 ---->
Hsa  TTTATCTTCACTCCCTCCCCAGCTCTTAGGCAACGTGATATTGATTGTTTTGGCAACCCACTTCAGCGAGGATTTTAC*CCTACAGATACAGGCTTCTTGGCAGTAACTAACTAAATGCTG
Ptr  T T  CTTCACTCCCTC C   C  T AG C  CG    A     TG T  G   A C       CG G  G     *C TA  C  A      T TT    GTAA TA CA ATGCTG
Ggo  T *  CTTCACTCCCTC C   C  T AG C  CG    A     TG T  G   A C       CG G  G     *C TA  G  A      T CT    GTAA TA CA ATGCTG
Ppy  T *  CTTCACTCCCTC C   C  T AG C  CG    A     TG T  G   A C       CG G  G     *C TA  G  A      T TT    GGAA TA CA ATGCTG
Mmu  T *  CTTCACTCCCTC C   C  T AA C  CG    A     TA T  G   A C       *A A  G     *C TA  G  A      T TT    GGAA CC CA ATGCTG
Age  T *  CTTCACTCCCTC C   C  T AG C  CG    A     GG T  G   A C       CG G  G     *C TA  G  A      T CA    AAGA CA CA A*****
Atr  T *  CTTCACTCCCTC C   T  T AG C  TG    A     GG T  G   G C       CG G  G     *C TA  G  A      T CG    AAGA CA TA A*****
Tsy  C *  CTCTCCTCCTCC A   C  C GG C  CG    C     TG C  A   A T       CA G  A     TT CA  G  G      T CT    GAAA TG CA CTGCTG
Chi  T *  T*********** C   C  C AG A  CA    A     TG C  G   A C       CA G  A     *C CG  G  G      G TT    GAAG TG CC ACGCTG
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---- 1920

                           ---->TER
Hsa  TGGTTAATGCTGTAGCCCACAAG*ACCACTGAGTTCCCTGTCC*ACTATGTTTGTACCTA*TGGTCCACTATGTTTGTACCTATGTCCCAAAATCTCATCTCCTTTAGATGGGGGAGGTT
Ptr  TGGTTAA G TG A    A   G* C  C G  TTCCCTGTCC*ACTATGTTTGTACCTA*TGGTCCACTATGTTT**A CTATGT CCGAAATC   TC    TT G TG GGGAGGTT
Ggo  TGGTTAA G TG A    A   G* C  C G  TCCCCTGTCC*ACTATGTTTGTACCTACTGGTCCACTATGTTTGTA CTATGT CCCAAATC   TC    TT G TG GGGAGGTT
Ppy  TGGTTAA G TG A    A   GG C  C G  TTCCCC**************************GTCCACTATGTTTGTA CTATGT CCAAAATC   TC    TT G TG GGGAGGTT
Mmu  TGGCTAA G TG A    T   GG C  T G  TTCCCT**************************GTCCACTATGTTTGTA CTATGT CCAAAAGC   TT    TT G CG GGGACACT
Age  ******* G TG G    A   TT T  C G  CAACTTGTCC*ACTGTGTTCGTACCTGCTGGTCCACTATGTTTGTA CCATGT CTAAAAGC   TC    TT G TG AGGATGTT
Atr  ******* G CA G    A   TT T  C G  CTCCCGGTCC*ACTGTGTGTGTACCTACTGGTCCGCCATGTTTGTA CCATAT CTAAAAGC   TC    TT G TG AGGATGTT
Tsy  TGGCTAA T TC G    A   GT C  C G  TTGTCTGTCT*ATTATGTTGGTGCCTCCTGG**********AGGGG TTGCGC TCAAGCGT   CC    TA A TG AGGGATGT
Chi  TGGCTAA G TC G    A   GT C  C *  TTGCCTGGCCTACCATGCTGGTGCCTA*********TCTGAAGGC* CAGTGT CCAGAAGT   TC    GA G CA A*******
     ---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---- 2040
```


Poly A
Hsa GGGGAGAAGAGCAGTATCCTGCCTGCTGATTCAGTTCCTGCATGATAAAAATAGAATAAAGAAATATG**CTCTCTAAGAAATATCATTGTACTCTTTTTCTGTCTTTATATTTTACCCT
Ptr GG GAGAAGAG AGTATC CTG TG TTC GTTC TGC TG A AA A G AATA G**C CTCT GA TATCAT GT TTC T TTC G C TATA CCC
Ggo GG GAGAAGAG AGTATC CTG TG TTC GTTC TGC TG A AA A G AATA G**C CTCT GA TATCAT GT TTC T TTC G C TATA CCC
Ppy GG GAGAAGAG AGTATC CTG TG TTT GTTA TGC TG A AA A G AATA G**C **** GA TATCAT GT TTC T TTC G C TATA CCC
Mmu GA GAGAAGAG AGTATC CTG AG TTC GTTC TGC TG A AA A G AATC G**C CTCT GA TATTAT GT TTC T TTC G C TATA CAT
Age GG GAGAAGAG AATATC ATG AG TTC GCTC TGC TG A AA A G AATG G**C **** GA TATCAT AT TAT T TTC G C TATG TGT
Atr GG GAGAAGAG AATATC CTG AG TTC GCTC TGC TG A AA A G AATG G**C GTTA GA TATCAG AT TAT T TTC G C TATG TGT
Tsy GG AGAGA*** AGTGCC CTG AG CTC GCAC TAT AA A GT T * AATA ACTC GTTA AA ****AT GC TTT C **T A T CACA ATT
Chi *G GAGA*GAG ***TTT AGA AT T** ACTC AAT TA C GA A * GTCA G**G CTAT TG TGCCCT GT TTT C CTT G T TGTA <
2160
Hsa GATTCAGCCAAAAGGACGCACTATTTCTGATGGAAATGAGAATGTTGGAGAATGGGAGTTTAAGGACAGAGAAGATACTTTCTTGCAATCCTGCAAGAAAAGAGAGAACTCGTGGGTGGA
Ptr A C GCCA AA CGC CCATTTC AT G T AG TGT G T GGAC T C T C A G TA GAA
Ggo A C GCCA AA CGC CCATTTC AT G T AG TGT G C GGAC T T T C A G TG TGA
Ppy A C GCCA AA CAC CCATTTC AT G T AG TGT G T GGAC T T T C A G TG GGC
Mmu A C GCCA AA CCC CCATTTC AT A T AA CAT G T GGGT T T A A T G TG GGC
Age A C GCCA AA TGC CCATTTC AT G T GG TGT A T AAAT G T T A T T G TG GGA
Atr A C GCCA AA TGC CCATTTC AT G T GG CAC A T GAAT G T T A T G TG GGA
Tsy * T TTGG GG TGG GAGCAGT GA A <
2280
Hsa TTTAGTGGGGTAGTTACTCCTAGGAAGGGGAAATCGTCTCTAGAATAAGACAATGTTTTTACAGAAAGGGAGGTCAATGGAGGTACTCTTTGG
Ptr G T AGA GG AA G CT A C C G TC G TGG
Ggo G T AGG GG AA G CT A C C G TC G TGG
Ppy G T AGG CA AA G CT A C C G GC G TGG
Mmu A T AGG GG AA G TC G C G T GC A TAG
Age G C GCG GG GG A CT A C A G GG G GAA
Atr G C GCA GG GG A <
2373
Gcr LINE AGCTTATCAAAGATCAAATGACAGTAAGTGGCTGGGTTCGTCTCTTGGGTTCTCTCTTCTGTTCTGTATATCTATTCCCTACTTTTGTGCCAGTACCATGCTATCTTGATCACTGT
AGATATATAATATAGCCTGAAGTCTGGTAGCAAGATTCCTCCTGATTTGTTTTTATTTCTGAGTAATGACTTGGCTATTCGAAGTTTTTTCTGATTCCATGTAAAATGAAGTACTA
TCCTTTCTAGATCTTTGAAGTATGACAATGGTGCTTTAATAGGGATTGCATTAAATTTGTATTTTGTCTTTGTGCAAATGCAAGGGTACATGCTAAAATTATTTAGAATAGAGGAT
AAATGTCTTAACACAACAAATAAGTAAGGAAGTAAGGTAAAGGTTATGTTAACCAGATTGATGTAAGCATTCCGAATTGTATATAAAATCAGCACATTGTACCCCATAAATGCATT
AATGTACATAGTTATGATTTAATAAGAAAAATAAAAATAAAAATAAATAAAAAAGAAAGAGAGTATAGTCAGT

```
Age Alu  TTATTTATTTATTTATTTATTTATTTTTTATTGCATTTTAGGTTTGGGGGTACATGTGAAGAACGTGGAAGATTGTTGCGTAGGTACACACATGGCAGTGTGATGTGCTGCCTTCC
         ACCCCATCACCTATATCTGGCATTTCTCCCCATGCTATCCCTCCCCAACTCCCCACCCCCCACTATCCCTCCCCTTTTTACCCCCGACAGACCCCAGTGTGTGATACTCCCCTCCC
         TGTGTCCATGTGTTCCCATTGTTCAACACCCGCCTATGAGTGAGAACATGCGGTGTTTGATTTTCTGTTCTTGTGTCAGTTTGCTGAGAATGATGGTTTCCATGTTCATCCATGTC
         CCTACAAAGGACACGAACTCATCGTTTTTGATGGCTACATAATATTCCATGGTGTATATGTGCCACATTTTCCCTGTCCAGTCTATCATCGATGGGCATTTGGAGTCTTT
```

FIG. 1.—The nucleotide sequences of tarsier ε-, γ-, η-, and galago η-globin genes aligned with orthologous sequences from other primates and from nonprimates. Primate sequences include human (*Homo sapiens*—Hsa), chimpanzee (*Pan troglodytes*—Ptr), gorilla (*Gorilla gorilla*—Ggo), orangutan (*Pongo pygmaeus*—Ppy), rhesus monkey (*Macaca mulatta*—Mmu), spider monkey (*Ateles geoffroyi*—Age), owl monkey (*Aotus trivirgatus*—Atr), tarsier (*Tarsius syrichta*—Tsy), galago (*Galago crassicaudatus*—Gcr), brown lemur [*Lemur macaco* (*fulvus*)—Lfu], dwarf lemur (*Cheirogaleus medius*—Cme), and nonprimate sequences are from rabbit (*Oryctolagus cuniculus*—Ocu) and goat (*Capra hircus*—Chi). For the sake of clarity, only the human sequence is complete; nucleotides are shown for other species at positions where one of the sequences differs from that of human. Gaps (*) have been inserted into the alignment to increase sequence similarity. The location of exons and known promoter sequences (CACCC, CCAAT, AATAAA, and poly A recognition sequence) are noted above the human sequence. The > and < symbols mark the beginning and end of sequences, respectively.

spanning thousands of nucleotide positions can readily be achieved between different primate species, even between those belonging to different suborders. Such extensive alignments of orthologous sequences contain the genetic information needed to resolve the genealogical relationships of the species represented by the sequences.

Material and Methods

Globin gene–containing *Eco*RI endonuclease restriction fragments from previously isolated lambda clones [Tsy Ch35-15.8, Tsy Ch40-15.5 (Koop et al. 1989) and Gcr Ch35-16.1 (Tagle et al. 1988)] were subcloned into pUC-19 plasmids (Maniatis et al. 1982) and propagated in *Escherichia coli* K-12 strain JM 109. A restriction map of the β-globin cluster of *Tarsius syrichta* has been presented by Koop et al. (1989), and a map of *Galago crassicaudatus* ε-, γ-, and η-globin genes has been presented by Tagle et al. (1988). Plasmid Tsy Ch35-5.8p7.1_E contained the ε gene, Tsy Ch35-15.8p1.8_E and p1.0_E contained the γ gene, and Tsy Ch40-15.5p4.4_E contained the η gene of tarsier. Plasmids Gcr Ch35-16.1p1.8_E and p4.3_E contained exons 1 and 2 of the η-globin gene of galago (exon 3 of the galago η gene was separated by a LINE insert >8 kb in length, and we have yet to locate its exact location). Gene regions were sequenced by chemical cleavage methods (Maxam and Gilbert 1980). Approximately 85% of the sequence was obtained on both strands, and all were sequenced at least twice. A more detailed account of the sequencing procedures has been described elsewhere (Slightom et al. 1985, 1987, 1988).

The nucleotide sequences of tarsier ε-, γ-, η-, and galago η-globin genes were aligned with orthologous sequences from other primates and nonprimates. Primate sequences include human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), rhesus monkey (*Macaca mulatta*), spider monkey (*Ateles geoffroyi*), owl monkey (*Aotus trivirgatus*), tarsier (*T. syrichta*), galago (*Galago crassicaudatus*), brown lemur [*Lemur macaco*(*fulvus*)], dwarf lemur (*Cheirogaleus medius*), and nonprimate sequences are from rabbit (*Oryctolagus cuniculus*) and goat (*Capra hircus*). References for sequences other than those presented in the present paper can be found in the work of Miyamoto et al. (1987, 1988), Fitch et al. (1988), Tagle et al. (1988), and Koop et al. (1989). Sequence alignments follow those presented in the above papers. Gaps were inserted into the alignment to increase sequence similarity.

Using two computational methods, we examined the molecular phylogeny of each of the five sets of aligned orthologous sequences (ε, γ, η, δ, and β). In the first method, a neighbor-joining tree (Saitou and Nei 1987) was computed from a distance matrix (number of sequence differences found in pairwise comparisons of the aligned sequences). In computing all pairwise sequence divergence, we treated all gaps as a single difference, since a single event is hypothesized. A gap of one or more positions was counted as a single event or change irrespective of the length of the contiguous unpaired region. As noncoding sequences account for 80%–100% of the total sequence and are generally not subject to varying degrees of selection pressure, they can be treated in a straightforward manner. For this reason we eliminated coding regions from divergence analyses. One advantage of using this distance method is that the same orthologous sequence in each of the compared taxa is not required. Only a complete distance matrix is required. In the second method, a maximum-parsimony tree (Fitch 1971; Goodman et al. 1979) was computed from analysis of the character states provided by the alignment positions. This tree accounted for the descent of the sequences by the minimum number of nucleotide substitutions, insertions, and dele-

tions, i.e., by the minimum number of fixed mutational events. This maximum-parsimony procedure does not require clock-like behavior in rates of change but only that, among taxa, matches in nucleotide sequence result more often from shared common ancestry than from convergencies, reversals, or parallelisms (Fitch 1971; Goodman et al. 1979, 1984, 1987; Czelusniak 1982).

Results and Discussion

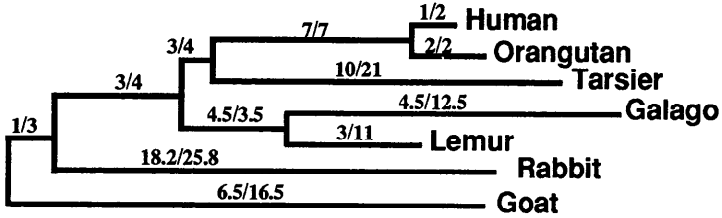
Tarsier ε, γ, and η and galago η nucleotide sequences and their alignments with both other primate and nonprimate mammalian sequences are presented in figure 1. Alignments of tarsier δ- and β-globin gene sequences with corresponding sequences from other primate and nonprimate species have been presented in a separate study that details the gene conversions (nonreciprocal recombination) occurring between the coding, intron 1, and 5′ regulatory regions of tarsier, lemur, ancestral anthropoid, and ancestral primate δ- and β-globin genes (Koop et al. 1989). Only δ and β sequences outside of these converted regions were used in the present study.

The tarsier ε and γ genes both have coding regions, splice sites, initiation codons, and termination codons that appear functional. The γ gene 5′ regulatory region, however, is unique among globin genes in that it lacks a functional CCAAT element. Even though both TATA and CACCC regulatory elements of the tarsier γ gene appear functional, the loss of both CCAAT elements would suggest a lower level of γ expression. This, however, remains to be confirmed in prenatal tarsier tissues. An additional unique feature of the tarsier γ gene is a type I Alu sequence in intron 2. As in all other primate η genes, the tarsier η appears nonfunctional. Two frameshift deletions and a single base insertion result in several premature stop codons within the η globin-coding regions. A mutation in the TATA promoter region of the η gene has also occurred. The galago η, like other primate η genes, has an altered initiation codon in addition to frameshift deletions resulting in premature stop codons. The galago η gene also contains two LINE sequences, the first of which is 533 bp long and inserted into intron 1 and the second of which is ≥8 kb long and inserted ~530 bp 3′ of the end of exon 2. We have yet to locate the 3′ end of the large LINE and exon 3 of the galago η gene.

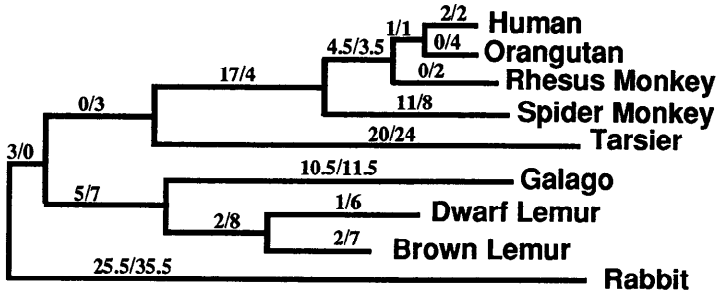
Pattern of Primate Phylogeny

The maximum-parsimony trees computed for ε-, γ-, η-, δ-, and β-globin genes show complete concordance with respect to the branching arrangement of the six major groups of primates (fig. 2). The neighbor-joining-method trees based on sequence divergences (when both corrected and uncorrected divergence values from total sequence and strictly noncoding sequences are used) were concordant for four of the five genes, with ε, γ, δ, and β trees agreeing with parsimony results and with the η tree grouping tarsier with both lemur and galago rather than with Anthropoidea (the sister group of tarsier in the parsimony trees).

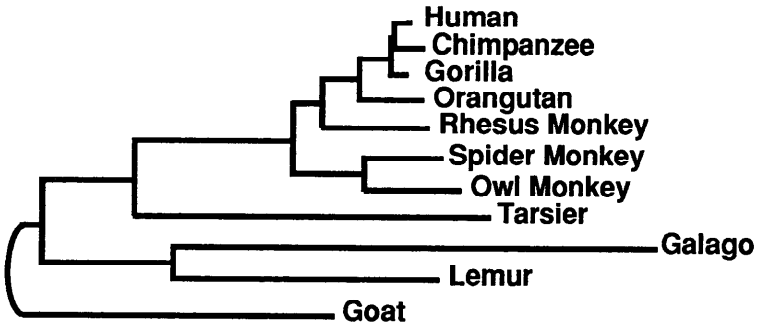
As in previous studies, hominoids group most closely with cercopithecoids to form the catarrhine branch, which in turn joins the ceboid branch to form the anthropoid assemblage. This phylogenetic arrangement is found with each of the three genes (γ, η, and δ) for which hominoid, cercopithecoid, and ceboid representatives are available (fig. 2). Within Anthropoidea, alternative groupings of hominoids, cercopithecoids, and ceboids require at least 69 more mutations than the number required in the most parsimonious γ, η, and δ trees. To break up Anthropoidea with tarsier or a lemuroid-lorisoid branch requires 230 more mutations than the number required in the most parsimonious score. Also, the dwarf lemur and brown lemur, which are



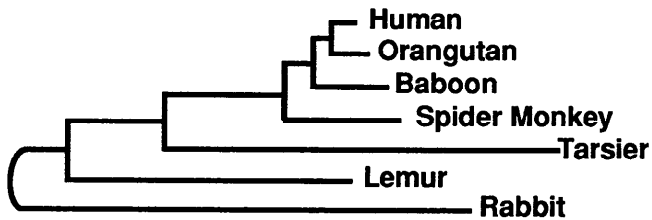
ε (1488)



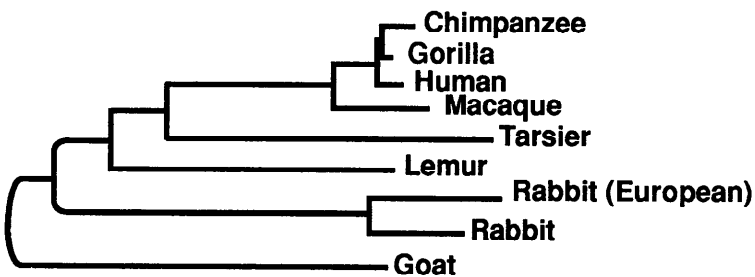
γ (1664)



η (1779)



δ (911)



β (1314)

represented by γ sequences, group together very closely. To break up this Cheirogalidae-Lemuridae group adds a minimum of 38 mutations to the score of the most parsimonious γ-gene tree (fig. 2). Thus, in agreement with many previous studies (Simpson 1945; Napier and Napier 1967; Bonner et al. 1980; Harris et al. 1986; Tagle et al. 1988), Cheirogalidae and Lemuridae are depicted as sister groups within Lemuroidea. In the ε-, γ-, and η-gene trees, the lemuroid branch then groups most closely with the galago or lorisoid branch. Ninety-one additional mutations over the most parsimonious score are required to prevent lorisoids from grouping with lemuroids.

The parsimony trees constructed for ε-, γ-, η-, δ-, and β-gene sequences also provide information on some of the more controversial phylogenetic relationships among primates. With regard to the disputes on the position of the tarsier branch, the most parsimonious tree for each of the five β-globin genes places tarsier closer to anthropoids than to any other primate group. This arrangement is consistent with the classification scheme which divides Primates into Haplorhini and Strepsirhini (table 1). Alternative placements of *Tarsius* require at least 28 more mutations than the number of mutations for all of the trees shown in figure 2. These 28 additional mutations are required to join the tarsier branch to the stem of a lemuroid-lorisoid group. To join tarsier to the stem of an anthropoid-lemuroid-lorisoid group requires 35 more mutations, and to join tarsier to galago requires 97 more mutations than the number required for the haplorhine grouping. The concordance of the maximum-parsimony trees for all five β-globin genes gives considerable support to grouping Anthropoidea and Tarsiodea in Haplorhini.

Molecular findings also help answer another long-standing question of primate phylogeny. This question concerns the branching pattern within Hominoidea, in particular whether the two African ape lineages (chimpanzee and gorilla) constitute a monophyletic group by themselves, as many primate morphologists believe (Tuttle

---

FIG. 2.—Species relationships as indicated by separate parsimony analysis of ε-, γ-, η-, δ-, and β-globin gene sequences. δ and β sequences are from flanking and intron 2 regions that have remained unaffected by gene conversions (Koop et al. 1989). In the η tree, catarrhine species are represented by the γ^1 locus, which appears to be the primary donor in the gene conversions between the duplicated γ^1 and γ^2 genes (Slightom et al. 1985, 1987, 1988). Branching patterns are based on complete sequences. The number in parentheses indicates the total length of the most parsimonious branching arrangement. Branch lengths are proportional to the number of substitutions in noncoding regions where all sequences are represented. The fractions on the ε and γ branch lengths refer to the number of amino acid changing substitutions over the number of silent substitutions occurring in the coding portions of the gene (the η-globin gene is silent in all primates). The trees presented for ε, γ, η, δ, and β are congruent with respect to branching arrangements of hominoids (human, chimpanzee, gorilla, and orangutan), old-world monkeys (rhesus and macaques), new-world monkeys (spider and owl monkeys), tarsiers, lemuroids (brown and dwarf lemurs), and lorisoids (galago). Within hominids there is some discrepancy with regard to the branching arrangement of human, chimpanzee, and gorilla. In the β-globin sequences examined, one sequence position supported grouping chimpanzee with gorilla, whereas in the η-globin sequences three sequence positions support grouping human with chimpanzee. This matter is explored further in fig. 3. Of 1,883 unambiguous substitutions identified in the haplorhine branch of most-parsimonious trees, 20.9% were G → A, 4.7% G → T, 4.6% G → C, 14.6% A → G, 3.7% A → T, 4.1% A → C, 3.3% T → G, 3.0% T → A, 13.3% T → C, 3.7% C → G, 5.4% C → A, 18.6% C → T [relative frequencies calculated as per Gojobori et al. (1982) and Tajima and Nei (1982)]. Approximately 67.3% of the substitutions were transitions (C ↔ T or A ↔ G); therefore, transitions occur twice as frequently as transversions (C or T ↔ A or G). Using the observed substitution frequencies to calculate the base composition of noncoding DNA under equilibrium conditions (Wright 1969), we find that the expected equilibrium nucleotide frequencies are very close to those observed within noncoding sequences (expected frequencies are as follows: G, 19.52%; A, 27.73%; T, 30.66%; and C, 22.10%; observed frequencies are G, 18.84%; A, 29.62%; T, 32.10%, and C, 19.44%).

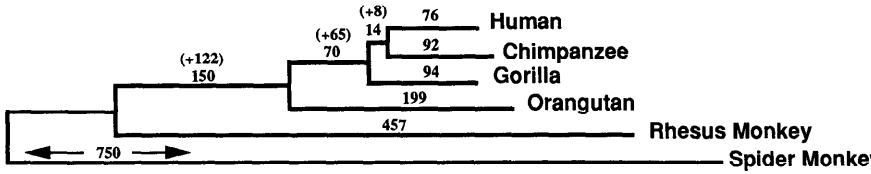
Table 1
Classification of Living Primates

Order Primates
Suborder Strepsirhini
Superfamily Lemuroidea—brown and dwarf lemurs
Superfamily Lorisoidea—galago, lorises
Suborder Haplorhini
Semisuborder Tarsiiformes
Superfamily Tarsioidea—tarsiers
Semisuborder Anthropoidea
Infraorder Platyrrhini
Superfamily Ceboidea—new-world monkeys, spider monkey
Infraorder Catarrhini
Superfamily Cercopithecoidea—old-world monkeys, rhesus and crab-eating macaques, baboons
Superfamily Hominoidea—humans and apes

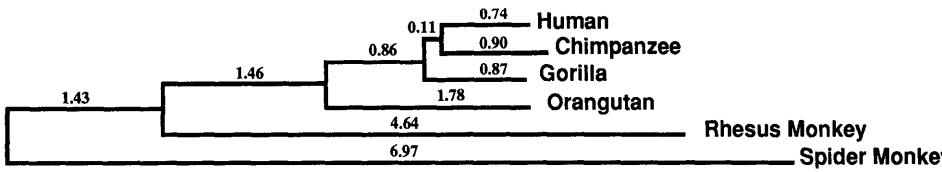
1967, 1969; Andrews 1986; Marks 1983), or whether one of the two lineages is closer to humans. Parsimony analysis of the η sequences indicates that one insertion, one deletion, and one transitional substitution group chimpanzee closest to the human lineage (Koop et al. 1986). In the β sequences examined, however, one transitional substitution groups chimpanzee with gorilla (Savatier et al. 1987; Koop et al. 1989). To further examine the branching pattern within Hominoidea, we have combined the data from five recent studies (Koop et al. 1986*a;* Miyamoto et al. 1987, 1988; Fitch et al. 1988; Maeda et al. 1988). These data consist of orthologous noncoding nucleotide sequences from human, chimpanzee, gorilla, orangutan, rhesus monkey, and spider monkey. The sequences for each species represent a contiguous 10.8-kb genomic region that spans the ψη-globin locus and extends 3′ toward the δ locus. In both the neighbor-joining tree and the maximum-parsimony tree computed for these noncoding DNA orthologues, the human and chimpanzee lineages are genealogically closest to each other (fig. 3). To first group chimpanzee with gorilla required an additional eight mutations, and to first group human with gorilla required nine mutations more than the number required in the maximum-parsimony score. A minimum of 65 additional mutations were required to place orangutan within the human-chimpanzee-gorilla branch. When the trees were ordered according to increasing mutational length, the first tree that did not group orangutan with the African ape–human branch added 122 mutations to the maximum-parsimony score. We found no sequence character uniquely supporting either a great-ape (orangutan-gorilla-chimpanzee) branch or a human-orangutan branch. Clearly these results provide overwhelming evidence for a monophyletic human-chimpanzee-gorilla branch, and, together with other studies (Sibley and Ahlquist 1984, 1987; Groves 1986; Miyamoto et al. 1987, 1988; Fitch et al. 1988; Holmquist et al. 1988; Maeda et al. 1988), provide significant evidence that within this branch the human and chimpanzee lineages share the most recent common ancestor.

Rates of Noncoding DNA Sequence Evolution in Primates

Our results not only elucidate the pattern of primate phylogeny but also provide important information on rates of DNA sequence evolution of humans and other primates. Theoretical findings on the spread of mutant genes through populations indicate that mutation pressure drives molecular evolution (Kimura 1983, 1987; Britten 1986). In previous studies the estimated rates of evolution among mammals have



Parsimony - number of substitutions over ~ 10,000 bp



NJ - corr (% Divergence)

C	1.64				
G	1.72	1.88			
Or	3.48	3.61	3.55		
Rh	7.85	8.01	7.80	7.86	
Sp	11.62	11.80	11.59	11.71	13.09
	H	C	G	Or	Rh

% Divergence - corr

FIG. 3.—Phylogenetic trees determined by parsimony (*top*) and neighbor-joining methods (*bottom*) for the 10.8 kb of noncoding DNA encompassing the ψη-globin gene of human, chimpanzee, gorilla, orangutan, rhesus monkey, and spider monkey. Branch lengths represent the number of changes incorporated along each lineage; in the neighbor-joining tree, this number is presented as the number of changes/100 base positions. The divergence matrix used to determine the neighbor-joining tree is presented below the two trees. Raw divergence values were calculated as the number of substitutions plus the number of gaps divided by the total number of shared nucleotide positions plus the total number of gaps (see text). Raw divergence values were then corrected for superimposed changes by using the Jukes and Cantor (1969) method; $K = -\frac{3}{4} \times \ln[1 - \frac{4}{3}(d)]$, where d = raw divergence and K = corrected divergence. The numbers in parentheses indicate the number of additional mutations required to break up the indicated grouping (see text for details).

ranged anywhere from 1×10^{-9} to 10×10^{-9} substitutions/site/year in unconstrained nucleotide sequences (Wu and Li 1985; Li and Tanimura 1987). This range of estimates has led to considerable controversy as to whether the maximum rate of evolution varies in different species lineages—or whether it is generally constant (Wilson et al. 1977, 1987), with mutations accumulating as a linear function of time (Zuckerkandl and Pauling 1962; Sarich and Wilson 1967*a*, 1967*b*). The results of the present study clearly demonstrate that marked nonuniformities in the accumulation of mutations in noncoding DNA have occurred in different primate lineages.

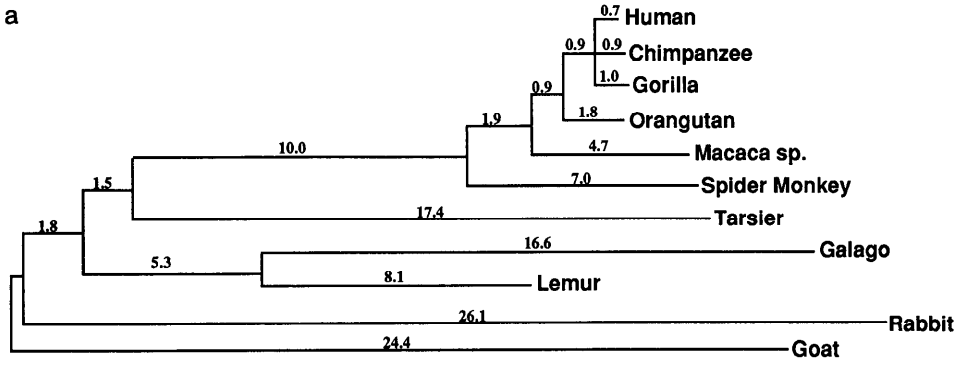
In the most parsimonious ε, γ, η, δ, and β trees shown in figure 2, the length of a branch represents the number of nucleotide substitutions found at noncoding positions. When one follows the branches from the ancestral primate (strepsirhine-haplorhine) node to the extant species, each gene tree shows smaller numbers of nucleotide substitutions at noncoding positions in lemur and hominoid lineages than in other primate lineages. In turn, the largest numbers of noncoding substitutions are consistently found in tarsier and galago lineages. Moreover, in those trees represented by all

Table 2
Divergences of Noncoding β-Cluster Sequence

	Human	Chimpanzee	Gorilla	Orangutan	Macaca species[a]	Spider Monkey	Tarsier	Galago	Lemur	Rabbit	Goat
Human		0.10	0.13	0.14	0.21	0.33	0.74	1.16	0.83	1.23	1.37
		15537	10902	17474	17259	11903	7902	4902	5607	4354	3492
Chimpanzee	1.61		0.13	0.18	0.22	0.36	0.96	1.51	1.20	1.99	1.78
			10920	11455	16822	10654	5176	2674	2932	1912	2165
Gorilla	1.72	1.80		0.18	0.26	0.36	1.07	1.52	1.23	2.08	1.82
				11474	12226	10682	4116	2674	2876	1856	2088
Orangutan	3.44	3.70	3.72		0.27	0.33	0.87	1.16	0.92	1.36	1.43
					11808	11908	5824	4100	4569	3611	3143
Macaca species[a]	7.18	7.32	7.77	7.80		0.37	0.93	1.54	1.22	1.99	1.87
						11052	5634	2673	3113	2031	2167
Spider Monkey	11.74	12.05	11.93	11.80	13.80		1.08	1.58	1.24	1.91	1.93
							4328	2577	3022	2159	1867
Tarsier	31.07	33.22	33.00	31.57	33.99	34.65		1.26	.93	1.41	1.58
								3750	5237	4000	3192
Galago	36.91	39.62	39.74	36.75	40.60	40.87	38.98		0.96	1.76	1.90
									34.63	2689	2505
Lemur	28.44	30.56	30.94	28.59	32.45	32.73	32.20	24.75		1.23	1.60
										4261	2316
Rabbit	41.43	45.57	47.30	41.84	47.16	46.57	47.05	48.34	41.17		2.40
											1556
Goat	41.44	42.63	42.97	40.96	45.66	43.01	47.13	50.74	38.70	50.52	

NOTE.—Divergences are presented below the null diagonal, and SE (Kimura 1983) and number of nucleotide sequence positions compared are given above the diagonal. Divergence and SE are expressed as substitutions/100 nucleotides, corrected for superimposed substitutions by using the formula indicated in the legend to fig. 3. Similar values are obtained using an alternative formula (Kimura 1983). Even though each pairwise divergence value was calculated for a minimumm of 1,500 base positions and SEs are given, divergence estimates may vary more than expected simply because different regions of the β-globin cluster were compared.

[a] Crab-eating macaque (*Macaca cynomalgus*) and rhesus monkey sequences were combined.



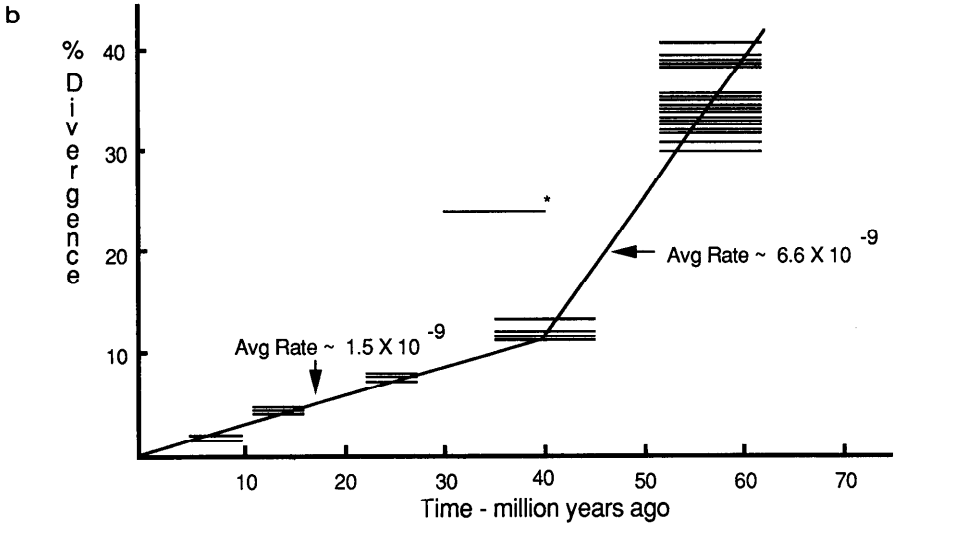


FIG. 4.—a, Summary of ε-, γ-, η-, δ-, and β-globin gene DNA sequence phylogeny. Only species represented by two or more globin gene sequences were used in this larger comparison. Branching arrangements reflect only the congruent brancing patterns found in fig. 2. Branch lengths (changes/100 base positions) were obtained using an additive method of apportioning pairwise genetic divergencies (from table 2) to the individual branch links. b, Sequence divergence (table 2) plotted against estimated time of separation (in millions of years ago). Human-chimpanzee-gorilla divergence was 4–9 Mya; orangutan diverged from human, chimpanzee, and gorilla 12–16 Mya; hominoids diverged from old-world monkeys 22–28 Mya; catarrhines diverged from platyrrhines 35–45 Mya; anthropoids diverged from tarsiers and strepsirhines 50–60 Mya (Gingerich 1984; Fleagle 1986). The asterisked (*) line denotes that lemurs may have diverged from galagos 30–40 Mya (see text). No reliable fossil evidence sufficiently delimits, for purposes of this plot, the time frame for the divergence of primates from lagomorphs or artiodatyls.

three anthropoid branches, the ceboid and cercopithecoid branches consistently diverge the most—and hominoids diverge the least—from the ancestral anthropoid (platyrrhine-catarrhine) node. That the hominoid rate of noncoding sequence evolution is slower than the cercopithecoid rate is also evident from the branch lengths of the maximum-parsimony tree for the orthologous DNA sequences of the extended ψη region (fig. 3).

The differences in branch lengths observed among primate lineages in the maximum-parsimony trees were also observed both in the individual neighbor-joining trees (e.g., see fig. 3) and in a summary tree for the combined results from all orthologous noncoding nucleotide sequence positions. These combined results in some cases represent more than 17,000 sequence positions (table 2). The branch lengths for this summary tree were derived (Fitch and Margoliash 1967, Saitou and Nei 1987) from the pairwise distances shown in table 2. The most striking difference in branch lengths is found between lemur and galago. After their ancestral separation, the galago accumulated nearly twice as many mutations as did the lemur lineage. It is also evident from the branch lengths of the summary tree that the rate of accumulation of mutations was slower in hominoid evolution than in either spider monkey or macaque lineages (fig. 4).

The interspecies sequence divergences estimated from noncoding nucleotide sequences (table 2) agree closely with the interspecies DNA divergences estimated by cross-hydridization of total genomic single-copy DNAs (table 3). Despite criticism of the analytical procedures used in DNA hydridization methods (Lewin 1988*a*, 1988*b*; Sarich et al. 1989), the hybridization results of Sibley and Ahlquist (1984, 1987) on catarrhines and of Bonner et al. (1980) on lemuroid, lorisoid, tarsier, and simian branches concur with the nucleotide sequence results shown in table 2 and in figures 3 and 4. The remarkable agreement of these values, obtained by very different methods, validates both of the methods and indicates that some of the criticisms of Sarich et al. are not of great moment.

In addition, Bonner et al. (1980), using DNA hybridization methods, found a much slower rate of DNA evolution in the lemuroid branch than in either the lorisoid or tarsier branches. In that DNA hybridization study, the lemuroid branch was represented by four diverse genera of Malagasy lemurs and the lorisoid branch was represented by three different genera (one from Africa and two from Asia). The excellent agreement between divergence values and different branch lengths obtained from total single-copy DNA and our sequence data (table 3) adds to evidence that rates of DNA sequence evolution vary markedly among species lineages and to the validity of using both methods.

Using paleontological estimates of primate divergence times, we converted the branch lengths of the summary tree for noncoding sequences (fig. 4a) into numbers of fixed changes per site per year, i.e., rates of fixation of mutations. Approximate divergence times employed are 55 Myr ago (Mya) for the strepsirhine-haplorhine split, 40 Mya for the platyrrhine-catarrhine split, and 25 Mya for the hominoid-cercopithecoid split (Gingerich 1984). On the basis of these times, the line of descent from the primate node to the human species shows a striking slowdown in evolutionary rates, from a high of 7.7×10^{-9} fixed changes/site/year for the first 15 Myr (55–40 Mya) to only 1.3×10^{-9} for the next 15 Myr (40–25 Mya) to a low of 1.0×10^{-9} for the last 25 Myr. The average evolutionary rate for the whole hominoid branch is also very slow (1.1×10^{-9}), whereas the cercopithecoid (macaque) and platyrrhine (spider monkey) rates (1.9×10^{-9} and 1.8×10^{-9}, respectively) are nearly twice as fast as

Table 3
Comparison of Sequence Divergence, as Estimated from Nucleotide Sequences from the β-Globin Cluster and as Estimated from DNA Hybridization Methods

	β-Globin Gene Cluster Noncoding Sequences (% Divergence)[a]	Genomic DNA-DNA Hybridization (T_{50} H)	
		From Sibley and Ahlquist (1987)	From Bonner et al. (1980)
Human–chimpanzee	1.6	1.6	...
Human–gorilla	1.7	2.3	...
Human–orangutan	3.4	3.6	...
Human–rhesus monkey	6.9	7.3	6.9
Human–spider monkey	10.9	...	11.2
Human–tarsier	25.4	...	25.4
Human–galago	29.2	...	28.0
Human–lemur	23.7	...	22.3
Tarsier–galago	30.4	...	30.2
Tarsier–lemur	26.2	...	25.8
Galago–lemur	21.1	...	22.1

[a] Values are not corrected for superimposed substitutions. All divergence values are multiplied by 100.

the hominoid rate but still much slower than the earlier stem-simian rate. The sister branch of Anthropoidea, in its descent from the primate node to *Tarsius syrichta,* shows an evolutionary rate of 3.4×10^{-9} fixed changes/site/year, which is approximately double the average simian rate but approximately half the stem-simian rate. Whether the early tarsier rate was fast and then became slower, as is to be seen in the comparison of the anthropoid stem to simians, cannot be determined, because, except for the sibling Philippine and Indonesian species of the extant genus *Tarsius,* only extinct fossil species provide evidence of the earlier branchings within Tarsioidea.

With regard to rates of evolution in strepsirhine primates, fossil evidence (Gingerich 1984; Beard et al. 1988; Martin 1988) indicates that lemuroids and lorisoids may have split 30–40 Mya, though some estimates go as far back as 50 Mya. Using these time estimates, we calculate a rate of 2.0×10^{-9} to 2.7×10^{-9} fixed changes/site/year in lemurs, 4.2×10^{-9} to 5.5×10^{-9} fixed changes/site/year in galago, and 2.1×10^{-9} to 3.5×10^{-9} fixed changes/site/year in stem-strepsirhines. These rate differences suggest that elevated rates of evolution occur not only in the stem-simian lineage but also in the lorisoid lineage.

The elevated stem-simian rate was further indicated by plotting pairwise divergence values from table 2 against estimated species divergence times (fig. 4b). In this plot, the average evolutionary rate of the simian lineages is $\sim 1.5 \times 10^{-9}$ fixed changes/site/year, whereas the line describing the divergence of simians from earlier primates shows a rate of $\sim 6.6 \times 10^{-9}$ fixed changes/site/year. These results agree with the rates determined from the branch lengths in figure 4a.

Mechanisms proposed to account for lineage differences in rates of fixation of mutations involve selection, population structure, generation time (more specifically the number of germ-line replications), and DNA repair mechanisms (Wu and Li 1985; Britten 1986; Li and Tanimura 1987). As the lineage differences in rates described above pertain to noncoding nucleotide sequences, neither selection nor population

structure would appear to be the primary cause of the rate variations observed. Only a small percent of noncoding sequences consist of promoter and enhancer elements with known regulatory functions. The majority of noncoding sequences, such as those from the η-globin pseudogene locus, are relatively free of selective constraints and thus represent selectively neutral DNA. Current theory suggests that the rate of drift of substitutions in selectively neutral DNA is independent of population size but that the rate of occurrence of neutral mutations determines that rate of fixation by drift (Kimura 1983, 1987). To the extent that new mutations result from errors of DNA replication, increasing generation times by decreasing the per-year number of germline replications should tend to decrease the rate of occurrence of new mutations. Similarly, an increased efficiency of DNA repair mechanisms would decrease the mutation rate (Mellon et al. 1987; Downes 1988).

Lineage differences in mutation rates, due either to differences in generation times or to differences in efficiency of DNA repair or to both factors, would appear to be the primary cause of lineage differences in rates described above for noncoding sequences. The twofold-higher rate of evolution in the lorisoid lineage vis-à-vis the lemuroid lineage, during the period of descent from the strepsirhine ancestor to the present, cannot be easily explained by variations in either generation time or population size. Lorisoids and lemurs have very similar life histories and generation times (Nowak and Paradiso 1983). Although the fixation of neutral mutations by random drift is thought to be independent of population size (Kimura 1983), this view continues to be debated. The alternative view is that fixation by random drift occurs more rapidly in small populations, such as founder populations, than in large populations (Wright 1969). By this latter view we might expect lemuroids, which are restricted to the island of Madagascar, to have incorporated more mutations than have the lorisoids of continental Africa and Asia. However, just the opposite is observed. Thus, we attribute the faster fixation rate of lorisoids to a higher mutation rate possibly due to less efficient DNA repair mechanisms. The lorisoid lineages of continental Africa and Asia perhaps faced more severe challenges from competing species than did the isolated lemuroids of Madagascar. Those lorisoids with a high mutation rate may have been the lorisoids in which phenotypically adaptive mutations first occurred—and therefore may have been the lorisoids that left surviving descendants.

Such an interpretation may also apply to the highly elevated substitution rate in the β-globin gene cluster of stem-simians. In this lineage the γ-globin gene was delayed in its developmental expression from embryonic life to fetal life, and the β-globin gene was delayed in its full expression until after birth (Tagle et al. 1988). Selected *cis*-acting regulatory mutations may have helped shape the altered pattern of developmental expression of γ- and β-globin genes. The primary cause of the elevated substitution rate, however, was probably that the mutation rate itself was high and that a high rate proved adaptive in these early continental primates. Both increased generation times and improved DNA repair mechanisms in later simian lineages could then explain the reduced rate of evolution in the descent of hominoids. As recently pointed out by James Neel (Marx 1988), it would make evolutionary sense for humans, with their long interval from birth to reproduction, to have evolved smarter DNA repair enzymes than have animals, which have short generation times.

Acknowledgments

Our thanks to M. Weiss, O. J. Miller, D. A. Miller, D. Fitch, W. M. Fitch, J. Czelusniak, W. H. Li, R. Britten, G. W. Lasker, and E. Mayr for their comments and

suggestions. We would also like to thank P. Wright and E. Simons of the Duke University Primate Center for their contribution of *Tarsius* tissues. Support for this study was provided by NSF grant BSR-8607202, NIH grant HL33940, and the Center for Molecular Biology, Wayne State University School of Medicine.

LITERATURE CITED

ANDREWS, P. 1986. Molecular evidence for catarrhine evolution. Pp. 107–129 *in* B. A. WOOD, L. MARTIN, and P. ANDREWS, eds. Major topics in primate and human evolution. Cambridge University Press, Cambridge.

BABA, M., M. GOODMAN, H. DENE, and G. W. MOORE. 1975. Origins of the Ceboidea viewed from an immunological perspective. J. Hum. Evol. **4**:89–102.

BABA, M., M. L. WEISS, M. GOODMAN, and J. CZELUSNIAK. 1982. The case of tarsier hemoglobin. Syst. Zool. **31**:89–102.

BARRIE, P. A., A. J. JEFFREYS, and A. F. SCOTT. 1981. Evolution of the β-globin gene cluster in man and primates. J. Mol. Biol. **149**:319–336.

BEARD, K. C., M. DAGOSTO, D. L. GEBO, and M. GODINOT. 1988. Interrelationships among primate higher taxa. Nature **331**:712–714.

BONNER, T. I., R. HEINEMANN, and G. J. TODARO. 1980. Evolution of DNA sequences has been retarded in Malagasy primates. Nature **286**:420–423.

BRITTEN, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science **231**:1393–1398.

COLLINS, F., and S. WEISSMAN. 1984. The molecular genetics of human hemoglobins. Prog. Nucleic Acids Res. Mol. Biol. **31**:315–462.

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.

DELSON, E., and A. L. ROSENBERGER. 1980. Phyletic perspectives in platyrrhine origins and anthropoid relationships. Pp. 445–458 *in* R. L. CIOCHON and A. B. CHIARELLI, eds. Evolutionary biology of new world monkeys and continental drift. Plenum, New York.

DOWNES, C. S. 1988. Views of unity and diversity. Nature **332**:208–209.

FITCH, D. H. A , C. MANINONE, J. L. SLIGHTOM, and M. GOODMAN. 1988. The spiker monkey ψη-globin gene region and surrounding sequence: recent or ancient insertions of LINES and SIGNS? Genomics **3**:237–255.

FITCH, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **19**:406–416.

FITCH, W. M., and E. MARGOLIASH. 1967. Construction of phylogenetic trees. Science **155**: 279–284.

FLEAGLE, J. G. 1986. The fossil record of early catarrhine evolution. Pp. 130–149 *in* B. A. WOOD, L. MARTIN, and P. ANDREWS, eds. Major topics in primate and human evolution. Cambridge University Press, Cambridge.

GINGERICH, P. D. 1975. Systematic position *Plesiadapsis.* Nature **253**:111–113.

———. 1984. Primate evolution: evidence from the fossil record, comparative morphology, and molecular biology. Yearbook Phys. Anthropol. **27**:57–72.

GOJOBORI, T., W. H. LI, and D. GRAUR. 1982. Patterns of nucleotide substitutions in pseudogenes and functional genes. J. Mol. Evol. **18**:360–369.

GOODMAN, M., J. CZELUSNIAK, B. F. KOOP, D. A. TAGLE, and J. L. SLIGHTOM. 1987. Globins: a case study in molecular phylogeny. Cold Spring Harbor Symp. Quant. Biol. **52**:875–900.

GOODMAN, M., J. CZELUSNIAK, G. W. MOORE, A. J. ROMERO-HERRARA, and G. MATSUDA. 1979. Fitting the gene lineage into its species lineages: a parsimony strategy illustrated by cladograms constructed from globin sequences. Syst. Zool. **28**:132–163.

GOODMAN, M., B. F. KOOP, J. CZELUSNIAK, M. L. WEISS, and J. L. SLIGHTOM. 1984. The η-globin gene: its long evolutionary history in the β-globin gene family of mammals. J. Mol. Biol. **180**:803–823.

GROVES, P. 1986. Systematics of the great apes. Pp. 187–217 *in* D. R. SWINDLER and J. ERMIN, eds. Comparative primate biology. Vol. **1**: Systematics, evolution and anatomy. Alan R. Liss, New York.

HARDIES, S. C., M. H. EDGELL, and C. A. HUTCHINSON III. 1984. Evolution of the mammalian β-globin gene cluster. J. Biol. Chem. **259**:3748–3756.

HARDISON, R. C. 1984. Comparison of the β-like globin gene families of rabbits and humans indicates that the gene cluster 5′-ε-γ-δ-β-3′ predates the mammalian radiation. Mol. Biol. Evol. **1**:390–410.

HARRIS, S., P. BARRIE, M. L. WEISS, and A. J. JEFFREYS. 1984. The primate ψβ1 gene: an ancient β-globin pseudogene. J. Mol. Biol. **180**:785–801.

HARRIS, S., J. R. THACKERAY, A. J. JEFFREYS, and M. L. WEISS. 1986. Nucleotide sequence analysis of the lemur β-globin gene family: evidence for major rate fluctuations in globin polypeptide evolution. Mol. Biol. Evol. **3**:465–484.

HILL, A., S. C. HARDIES, S. J. PHILLIPS, M. G. DAVIS, C. A. HUTCHINSON III, and M. H. EDGELL. 1984. Two mouse early embryonic β-globin gene sequences. J. Biol. Chem. **259**: 3739–3747.

HOLMQUIST, R., M. M. MIYAMOTO, and M. GOODMAN. 1988. Analysis of higher-primate phylogeny from transversion differences in nuclear and mitochondrial DNA by Lake's methods of evolutionary parsimony and operator metrics. Mol. Biol. Evol. **5**:217–236.

JEFFREYS, A., P. BARRIE, S. HARRIS, D. FAWCETT, Z. NUGENT, and C. BOYD. 1982. Isolation and sequence analysis of a hybrid δ-globin pseudogene from the brown lemur. J. Mol. Biol. **156**:487–503.

JONES F. W. 1920. Discussion on the zoological position and affinities of *Tarsius*. Proc. Zool. Soc. [Lond.] **1920**:491–949.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–123 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Sinauer, Sunderland, Mass.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

———. 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 β-hemoglobin genes (ε^M and β^M) of opossum. Proc. Natl. Acad. Sci. USA **85**:3893–3897.

KOOP, B. F., M. GOODMAN, P. XU, K. CHAN, and J. L. SLIGHTOM. 1986*a*. Primate η-globin DNA sequences and man's place among the great apes. Nature **319**:234–238.

KOOP, B. F., M. M. MIYAMOTO, J. E. EMBURY, M. GOODMAN, J. CZELUSNIAK, and J. L. SLIGHTOM. 1986*b*. Nucleotide sequence and evolution of the orangutan ε globin gene region and surrounding Alu repeats. J. Mol. Evol. **24**:94–102.

KOOP, B. F., D. SIEMIENIAK, J. L. SLIGHTOM, M. GOODMAN, J. DUNBAR, P. C. WRIGHT, and E. L. SIMONS. 1989. *Tarsius* δ- and β-globin genes: conversions, evolution and systematic implications. J. Biol. Chem. **264**:68–79.

LE GROS CLARK, W. E. 1962. The antecedents of man. 2d ed. Edinburgh University Press, Edinburgh.

LEWIN, R. 1988*a*. Conflict over DNA clock continues. Science **241**:1756–1759.

———. 1988*b*. Conflict over DNA clock results. Science **241**:1598–1600.

LI, W. H., and M. TANIMURA. 1987. The molecular clock runs more slowly in man than in apes and monkeys. Nature **326**:93–96.

LUCKETT, W. P., and F. S. SZALAY. 1978. Clades versus grades in primate phylogeny. Pp. 227–235 *in* D. J. CHIVERS and K. A. JOYSEY, eds. Recent advances in primatology. Vol. **3**. Academic Press, New York.

MACPHEE, R. E. E., and M. CARTMILL. 1986. Basicranial structures and primate systematics.

Pp. 219–275 *in* D. R. SWINDLER and J. ERVIN, eds. Comparative primate biology. Vol. **1:** Systematics, evolution and anatomy. Alan R. Liss, New York.

MAEDA, N., C.-I. WU, J. BLISKA, and J. RENEKE. 1988. Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences. Mol. Biol. Evol. **5**:1–20.

MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MARKS, J. 1983. Hominoid cytogenetics and evolution. Yearbook Phys. Anthropol. **26**:131–159.

MARTIN, R. D. 1988. Several steps forward for Eocene primates. Nature **331**:660–661.

MARX, J. 1988. Lower radiation levels found. Science **241**:1286.

MAXAM, A., and W. GILBERT. 1980. Sequencing end-labelled DNA with base-specific chemical cleavage. Methods Enzymol. **68**:499–560.

MELLON, I., G. SPIVAK, and P. C. HANAWALT. 1987. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. Cell **51**:241–249.

MIYAMOTO, M. M., and M. GOODMAN. 1986. Biomolecular systematics of eutherian mammals: phylogenetic patterns and classification. Syst. Zool. **35**:230–240.

MIYAMOTO, M. M., B. F. KOOP, J. L. SLIGHTOM, M. GOODMAN, and M. R. TENNANT. 1988. Molecular systematics of higher primates: genealogical relationships and classification. Proc. Natl. Acad. Sci. USA **85**:7627–7631.

MIYAMOTO, M. M., J. L. SLIGHTOM, and M. GOODMAN. 1987. Phylogenetic relations of humans and African apes from DNA sequences in the ψη-globin region. Science **238**:369–373.

NAPIER, J. R., and P. H. NAPIER. 1967. A handbook of living primates. 3d ed. Academic Press, New York.

NOWAK, R. M., and J. L. PARADISO. 1983. Walker's mammals of the world. 4th ed. The Johns Hopkins University Press, Baltimore.

POCOCK, R. I. 1918. On the external characters of lemurs and of *Tarsius*. Proc. Zool. Soc. [Lond.] **1918**:19–53.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SARICH, V. M., C. W. SCHMID, and J. MARKS. 1989. DNA hybridization as a guide to phylogenies: a critical analysis. Cladistics **5**:3–32.

SARICH, V. M., and A. C. WILSON. 1967*a*. Immunological time scale for hominoid evolution. Science **158**:1200–1203.

———. 1967*b*. Rates of albumin evolution in primates. Proc. Natl. Acad. Sci. USA **58**:142–148.

SAVATIER, P., G. TRABUCHET, Y. CHEBLOUNE, C. FAURE, G. VERDIER, and V. M. NIGON. 1987. Nucleotide sequence of the delta-beta globin intergenic segment in the macaque: structure and evolutionary rates in higher primates. J. Mol. Evol. **24**:297–308.

SAWADA, I., C. WILLARD, C.-K. J. SHEN, B. CHAPMAN, A. C. WILSON, and C. W. SCHMID. 1985. Evolution of Alu family repeats since the divergence of human and chimpanzee. J. Mol. Evol. **22**:316–322.

SCHWARTZ, J. H. 1978. If *Tasius* is not a prosimian, is it a haplorhine? Pp. 195–202 *in* D. J. CHIVERS and K. A. JOYSEY, eds. Recent advances in primatology. Vol. **3**. Academic Press, London.

———. 1984. What is a tarsier? Pp. 38–49 *in* N. ELDREDGE and S. M. STANLEY, eds. Living fossils. Springer, New York.

———. 1986. Primate systematics and a classification of the order. Pp. 1–41 *in* D. R. SWINDLER and J. ERVIN, eds. Comparative primate biology. Vol. **1:** Systematics, evolution and anatomy. Alan R. Liss, New York.

SCHWARTZ, J. H., and I. TATTERSALL. 1985. Evolutionary relationships of living lemurs and lorises (Mammalia, Primates) and their potential affinities with European eocene Adapidae. Anthropol. Pap. Am. Museum Nat. Hist. **60**:1–100.

SIBLEY, C. G., and J. E. AHLQUIST. 1984. The phylogeny of hominoid primates, as indicated by DNA-DNA hybridization. J. Mol. Evol. **20**:2–15.

———. 1987. DNA evidence of hominoid phylogeny: results from an expanded data set. J. Mol. Evol. **26**:99–121.

SIMPSON, G. G. 1945. The principles of classification and a classification of mammals. Bull. Am. Museum Nat. Hist. **85**:1–350.

SLIGHTOM, J. L., A. E. BLECHL, and O. SMITHIES. 1980. Human $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**:627–638.

SLIGHTOM, J. L., L.-Y. E. CHANG, B. F. KOOP, and M. GOODMAN. 1985. Chimpanzee fetal $^{G}\gamma$ and $^{A}\gamma$ globin gene nucleotide sequences provide further evidence of gene conversions in hominine evolution. Mol. Biol. Evol. **2**:370–389.

SLIGHTOM, J. L., B. F. KOOP, P. XU, and M. GOODMAN. 1988. Rhesus β-globin genes: concerted gene evolution in the descent of higher primates. J. Biol. Chem. **263**:12427–12438.

SLIGHTOM, J. L., T. W. THEISEN, B. F. KOOP, and M. GOODMAN. 1987. Orangutan fetal globin genes. J. Biol. Chem. **262**:7472–7483.

SZALAY, F. S., A. L. ROSENBERGER, and M. DAGOSTO. 1987. Diagnosis and differentiation of the order Primates. Yearbook Phys. Anthropol. **30**:75–105.

TAGLE, D. A., B. F. KOOP, M. GOODMAN, J. L. SLIGHTOM, D. L. HESS, and R. T. JONES. 1988. Embryonic ε and γ globin genes of a prosimian primate (*Galago crassicaudatus*): nucleotide and amino acid sequences, developmental regulation, and phylogenetic footprints. J. Mol. Biol. **203**:439–455.

TAJIMA, F., and M. NEI. 1982. Biases in the estimates of DNA divergence obtained by the restriction enzyme technique. J. Mol. Evol. **18**:115–120.

TUTTLE, R. H. 1967. Knuckle-walking and the evolution of hominoid hands. J. Phys. Anthropol. **26**:171–206.

———. 1969. Quantitative and functional studies on the hands of the Anthropoidea. J. Morphol. **128**:309–364.

WILSON, A. C., S. S. CARLSON, and T. J. WHITE. 1977. Biochemical evolution. Annu. Rev. Biochem. **46**:573–639.

WILSON, A. C., H. OCHMAN, and E. M. PRAGER. 1987. Molecular time scale for evolution. Trends Genet. **3**:241–247.

WRIGHT, S. 1969. Evolution and genetics of populations. Vol. **2**. University of Chicago Press, Chicago.

WU, C.-I., and W. H. LI. 1985. Evidence for higher rates of nucleotide substitutions in rodents than in man. Proc. Natl. Acad. Sci. USA **82**:1741–1745.

ZUCKERKANDL, E., and L. PAULING. 1962. Molecular disease, evolution, and genetic heterogeneity. Pp. 189–225 *in* M. KASKA and B. PULLMAN, eds. Horizons in biochemistry. Academic Press, New York.

WALTER M. FITCH, reviewing editor

Received February 10, 1989; revision received July 17, 1989

Accepted July 18, 1989