J Mol Evol (1991) 32:128–144

Journal of
Molecular Evolution

© Springer-Verlag New York Inc. 1991

Evolution of the Cytochrome *b* Gene of Mammals

David M. Irwin, Thomas D. Kocher,* and Allan C. Wilson

Division of Biochemistry and Molecular Biology, University of California, Berkeley, CA 94720, USA

Summary. With the polymerase chain reaction (PCR) and versatile primers that amplify the whole cytochrome *b* gene (~1140 bp), we obtained 17 complete gene sequences representing three orders of hoofed mammals (ungulates) and dolphins (cetaceans). The fossil record of some ungulate lineages allowed estimation of the evolutionary rates for various components of the cytochrome *b* DNA and amino acid sequences. The relative rates of substitution at first, second, and third positions within codons are in the ratio 10 to 1 to at least 33. For deep divergences (>5 million years) it appears that both replacements and silent transversions in this mitochondrial gene can be used for phylogenetic inference. Phylogenetic findings include the association of (1) cetaceans, artiodactyls, and perissodactyls to the exclusion of elephants and humans, (2) pronghorn and fallow deer to the exclusion of bovids (i.e., cow, sheep, and goat), (3) sheep and goat to the exclusion of other pecorans (i.e., cow, giraffe, deer, and pronghorn), and (4) advanced ruminants to the exclusion of the chevrotain and other artiodactyls. Comparisons of these cytochrome *b* sequences support current structure–function models for this membrane-spanning protein. That part of the outer surface which includes the Q_o redox center is more constrained than the remainder of the molecule, namely, the transmembrane segments and the surface that protrudes into the mitochondrial matrix. Many of the amino acid replacements within the transmembrane segments are exchanges between hydrophobic residues (especially leucine, isoleucine, and valine). Replacement changes at first and second positions of codons approximate a negative binomial distribution, similar to other protein-coding sequences. At four-fold degenerate positions of codons, the nucleotide substitutions approximate a Poisson distribution, implying that the underlying mutational spectrum is random with respect to position.

Key words: Polymerase chain reaction — Direct sequencing — Mitochondrial DNA — Hoofed mammals — Cetaceans — Molecular phylogeny — Evolutionary rates — Base composition — Structure–function relationships

Introduction

Mitochondrial DNA (mtDNA) is a valuable molecule for understanding the evolutionary relationships among individuals, populations, and species. Until recently such understanding was achieved mainly by restriction analysis because that technique is simpler than conventional cloning and sequencing (Wilson et al. 1985, 1989; Harrison 1989). With the advent of the polymerase chain reaction (PCR) (White et al. 1989), it became possible to obtain mtDNA sequences directly from many taxa; these sequences can be aligned across greater time spans than can restriction maps (Kocher et al. 1989; Kocher and White 1989; Thomas et al. 1989; Simon et al. 1990). However, short sequences amplified in this manner could not resolve ancient evolutionary relationships (Kocher et al. 1989). Here we describe primers that amplify the whole cytochrome *b* gene from a wide variety of hoofed mammals (ungulates) as well as cetaceans (e.g., dolphins) and evaluate the utility of these sequences for resolving the phylogeny of ungulates.

The ungulates have a good fossil record and many morphological distinctions (Savage and Russell

* *Present address:* Department of Zoology, University of New Hampshire, Durham, NH 03824, USA
Offprint requests to: D.M. Irwin

1983; Shoshani 1986; Carroll 1988; Janis 1988). Molecular comparisons of ungulates are available for a few proteins, including albumin (Carlson et al. 1978; Sarich 1985, and personal communications), lens crystallins (de Jong 1985), globins (Koop and Goodman 1988; Czelusniak et al. 1990; Easteal 1990), fibrinopeptides (Mross and Doolittle 1967), and ribonucleases (Beintema et al. 1988), and also for mitochondrial rDNA (Miyamoto and Boyle 1989). Nevertheless, relationships among many groups are uncertain. There has been a diversity of opinion about how elephants are related to other ungulates (Simpson 1945; Novacek 1989). Also troublesome are the relationships of true ruminants to camels and pigs (Janis 1988). Resolution of the phylogenetic relationships among these latter species may clarify the origin and evolution of ruminant stomach lysozymes (Irwin and Wilson 1989, 1990; Irwin et al. 1989; Jollès et al. 1989, 1990).

The cytochrome *b* gene was chosen as a phylogenetic probe because it may be easier to align a protein-coding sequence that has evolved over the period spanning the origin of mammalian orders than to align either mitochondrial rDNA or noncoding sequences from distant relatives. This particular gene, together with a mitochondrial rDNA, has been found valuable for addressing even deeper phylogenetic questions, such as the origin of tetrapods (Meyer and Wilson 1990). In addition, our cytochrome *b* approach complements analogous studies on the rDNAs of ungulate mitochondria (Miyamoto and Boyle 1989).

Furthermore, the protein-coding gene we have chosen has an advantage over those studied before by cloning and sequencing from a taxonomically graded series of mammalian mitochondria. We refer to the ND4 and ND5 genes, parts of which have been compared extensively in both closely and distantly related primates (Brown et al. 1982; Hayasaka et al. 1988). Relatively little is known of the structure–function relationships in the ND4 and ND5 proteins. Cytochrome *b*, by contrast, is one of the best known of the 9–10 proteins that make up complex III of the mitochondrial oxidative phosphorylation system (Hatefi 1985), and it is the only one of them encoded by the mitochondrial genome. Complex III transfers electrons from dihydroubiquinone to cytochrome *c*, and this reaction is coupled to translocation of protons across the mitochondrial inner membrane (Hatefi 1985). Cytochrome *b* is believed to contain both redox centers, Q_o and Q_i, involved in electron transfer (Hatefi 1985; Howell and Gilbert 1988). Mutational and evolutionary studies have facilitated the development of a model structure of cytochrome *b* as well as the definition of the sites of electron transfer and inhibitor action (Howell and Gilbert 1988; Howell 1989; di Rago et al. 1990). Our knowledge of structure–function relationships in this protein enhances the utility of its gene for evolutionary investigations.

The 17 complete sequences of the cytochrome *b* gene reported here, together with the published human (Anderson et al. 1981), mouse (Bibb et al. 1981), and rat (Gadaleta et al. 1989) sequences, provide a total of 20 mammalian sequences.[1] This allows comparisons of genes with a wide range of divergence times from less than 5 to more than 65 million years. These comparisons permit an assessment of the tempo and mode of the evolution of the cytochrome *b* gene as well as the phylogenetic utility of replacement, silent, transition, and transversion changes within this gene.

Materials and Methods

DNA Sources. Total genomic DNA from domestic cow (*Bos taurus*), domestic sheep (*Ovis aries*), pronghorn antelope (*Antilocapra americana californica*), giraffe (*Giraffa camelopardalis*), fallow deer (*Dama dama*), black-tailed deer (*Odocoileus hemionus*), larger Malay chevrotain (*Tragulus napu*), dromedary camel (*Camelus dromedarius*), domestic pig (*Sus scrofa*), collared peccary (*Tayassu tajacu*), Grévy's zebra (*Equus grevyi*), and a long-beaked dolphin (1a, *Stenella longirostris*) were from a previous study (Irwin et al. 1989). Tissues for preparation of DNA as described by Irwin et al. (1989) from black rhinoceros (*Diceros bicornis*), African elephant (*Loxodonta africana*), and another long-beaked dolphin (1b, *S. longirostris*) were provided by O.A. Ryder, San Diego Zoo. Narrow-snouted dolphin DNA (2, *Stenella attenuata*) was from A.E. Dizon (Southwest Fisheries Center, La Jolla, CA) and domestic goat (*Capra hircus*) DNA from D.E. Dobson (Boston University School of Medicine). The DNA sequences for cytochrome *b* were from published sources for human (Anderson et al. 1981), mouse (*Mus domesticus*, Bibb et al. 1981), and rat (*Rattus norvegicus*, Gadaleta et al. 1989).

Cytochrome b Sequences. Mitochondrial sequences containing the cytochrome *b* gene were isolated via PCR. Primers for amplification and sequencing are shown in Table 1 and were designed to correspond to conserved areas of the human, mouse, and cow (Anderson et al. 1982) sequences, as suggested by Kocher et al. (1989). Typically the entire cytochrome *b* gene was amplified with the flanking tRNA primers L14724 and H15915 (Table 1 and Fig. 1). Conditions for amplification and DNA sequencing were as described by Kocher et al. (1989). Double-stranded DNA was purified by electrophoresis in agarose of low gelling temperature (Maniatis et al. 1982) and used for single-stranded PCR or M13 cloning. Single-stranded DNA for sequencing was produced via asymmetric PCR (Gyllensten and Erlich 1988) of the entire gene or part of it using the double-stranded product as the template. Double-stranded PCR products from the cow, camel, and pig were cloned and sequenced in M13 (Messing 1983).

Data Analysis. Parsimony trees were constructed from character-state data using PAUP (Swofford 1989). Confidence values

[1] A cow sequence has also been published before on the basis of classical cloning and sequencing methods (Anderson et al. 1982); the one we determined proved to be identical to that previously reported, although others have found small variations among the mtDNAs of domestic cattle (Watanabe et al. 1985, 1989)

Table 1. Twelve primers for amplification and sequencing of mammalian cytochrome *b* genes

Name of primer	Sequence
L14724[a]	5′-CGAAGCTTGATATGAAAAACCATCGTTG-3′
L14724B	5′-CGAGATCTGAAAAACCATCGTTG-3′
L14841[b]	5′-AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA-3′
H14927	5′-GTGACAGAGGAGAATGCTGT-3′
L14979	5′-GACGTCAACTACGGCTGAAT-3′
H15149[b]	5′-AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA-3′
L15162[a]	5′-GCAAGCTTCTACCATGAGGACAAATATC-3′
L15408	5′-ATAGACAAAATCCCATTCCA-3′
L15513	5′-CTAGGAGACCCTGACAACTA-3′
L15775	5′-GTAAAACGACGGCCATACATGAATTGGAGGACAACCAGTC-3′
H15915[a]	5′-AACTGCAGTCATCTCCGGTTTACAAGAC-3′
H15915R	5′-GGAATTCATCTCTCCGGTTTACAAGAC-3′

Names identify the DNA strand (H or L) and the position of the 3′ end of the oligonucleotide according to the numbering system for the human sequence (Anderson et al. 1981)
[a] Primers obtained from S. Pääbo
[b] Primers described by Kocher et al. (1989)

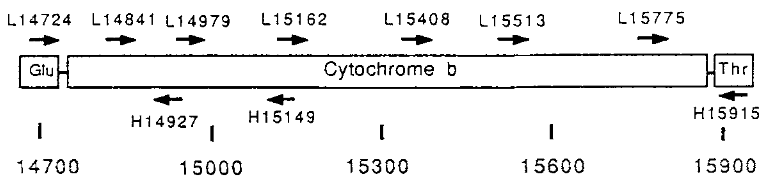

Fig. 1. Strategy for amplification and sequencing of the cytochrome *b* gene. Bars depict the cytochrome *b* gene and flanking tRNA genes and arrows denote primers (see Table 1). Numbering is according to the human mtDNA sequence (Anderson et al. 1981).

for internal lineages were assessed with the bootstrap option (Felsenstein 1985). Trees were constructed using various components of the sequences and various weighting options, as described in the Results and Discussion. In addition, trees were constructed using the number of transversion differences as a measure of distance in the neighbor-joining algorithm (Saitou and Nei 1987).

Base compositions of cytochrome *b* genes were analyzed by the method of Prager and Wilson (1988). To determine the spatial pattern of substitutions within the cytochrome *b* sequence, the number of inferred nucleotide substitutions at each codon position was compared to the number expected according to Poisson and negative binomial distributions (Holmquist et al. 1983). In order to analyze just replacement changes at first positions of codons, only those codons that did not encode leucine in a master sequence (either cow or human) were used. Similarly, at third positions of codons, only fourfold degenerate codons in the master sequence (cow or human) were analyzed when attention was being confined to silent changes.

Results and Discussion

Cytochrome b Sequences

The entire mitochondrial cytochrome *b* gene was amplified via the flanking tRNA primers (see Fig. 1). In all cases, a single DNA fragment of approximately 1100–1200 bp resulted, with no apparent differences in size being evident among species. The PCR products from three of the species (cow, camel, and pig) were cloned and sequenced using M13 vectors. Because cloned material amplified via PCR typically contains mutations (Saiki et al. 1988), for the species from which the cytochrome *b* gene was cloned sequences were determined from three independent clones, as recommended by Irwin and Wilson (1990), and the consensus sequence reported (Fig. 2). The number of nucleotide differences between the clones within the same species was found to be similar to the number expected from known error rates for *Taq* polymerase (Saiki et al. 1988; Tindall and Kunkel 1988; Irwin and Wilson 1989; Keohavong and Thilly 1989). The remaining 14 sequences were determined by direct sequencing of the single-stranded PCR product (Gyllensten and Erlich 1988), in which case errors induced by *Taq* polymerase are not seen (Pääbo and Wilson 1988).

Mitochondrial DNA fragments have been found in the nuclear genome of humans (Fukuda et al. 1985; Nomiyama et al. 1985; Kamimura et al. 1989). If a complete cytochrome *b* gene accompanied by the two flanking tRNA genes had been transferred to the nuclear genome, a product of indistinguishable size might be amplified via PCR. If this second product formed a significant portion (>10%) of the amplified material, the sequence determined should reveal some positions at which there were two different bases. If the nuclear sequence were the only one amplified, this sequence should show nucleotide differences from the other species analogous to those of nuclear pseudogenes, i.e., an increased number of replacement substitutions and a lower transition bias at silent sites (Li et al. 1985). Because we found neither double sequences nor any sequence appearing to behave as a pseudogene, we conclude that all

```
                                                                                         20
             M   T   N   I   R   K   S   H   P   L   M   K   I   V   N   N   A   F   I   D   L   P   A   P   S   N   I   S   S   W   W   N
Cow         ATG ACT AAC ATT CGA AAG TCC CAC CCA CTA ATA AAA ATT GTA AAC AAT GCA TTC ATC GAC CTT CCA GCC CCA TCA AAC ATT TCA TCA TGA TGA AAT
Sheep       ATG ATC AAC ATC CGA AAA ACC CAC CCA CTA ATA AAA ATT GTA AAC AAC GCA TTC ATT GAT CTC CCA GCT CCA TCA AAT ATT TCA TCA TGA TGA AAC
Goat        ATG ACC AAC ATC CGA AAG ACC CAC CCA TTA ATA AAA ATT GTA AAC AAC GCA TTT ATT GAC CTC CCA ACC CCA TCA AAC ATC TCA TCA TGA TGA AAC
Pronghorn   ATG ATC AAC ATC CGA AAA TCC CAC CCA TTA ATA AAA ATT GTA AAC AAC GCA TTC ATT GAC CTC CCA GCC CCA TCA AAC ATC TCA TCT TGA TGA AAC
Giraffe     ATG ATC AAC ATC CGA AAG TCC CAC CCA CTA ATA AAA ATT GTA AAT AAC GCA CTA ATC GAT CTA CCA GCC CCA TCA AAT ATC TCA TCA TGA TGA AAC
Fallow      ATG ATC AAC ATC CGA AAA TCT CAC CCA TTG ATA AAA ATC GTA AAC AAC GCA TTC ATT GAT CTC CCA GCC CCA TCA AAT ATT TCA TCC TGA TGA AAT
Black-tail  ATG ACC AAC ATC CGA AAA ACC CAC CCA CTC ATA AAA ATT GTA AAC AAC GCA TTC ATT GAT CTT CCT GCC CCA TCA AAC ATC TCG TCA TGA TGA AAC
Chevrotain  ATG ATC AAT ATC CGA AAA TCA CAC CCA CTA ATA AAA ATT GTC AAC AAT GCA TTT ATT GAC CTC CCA GCC CCA TCA AAC ATC TCT TCA TGG TGA AAC
Camel       ATG ACA AAC ATC CGA AAA TCA CAC CCA CTT CTA AAA ATT ATA AAC GAC GCA TTC ATT GAC CTT CCA GCC CCC TCC AAT ATT TCA TCA TGA TGA AAC
Dolphin 1a  ATG ACC AAC ATC CGA AAA ACA CAC CCT CTA ATA AAA ATC CTC AAT GAC GCA TTC ATT GAT CTC CCC ACC CCA TCT AAT ATC TCC TCT TGA TGA AAT
Dolphin 1b  ATG ACC AAC ATC CGA AAA ACA CAC CCA CTA ATA AAA ATC CTC AAT GAC GCA TTC ATT GAT CTA CCC ACC CCA TCT AAT ATC TCC TCT TGA TGA AAT
Dolphin 2   ATG ACC AAC ATC CGA AAA ACA CAC CCA CTA ATA AAA ATC CTC AAT GAT GCA TTC ATT GAT CTA CCC ACT CCA TCT AAC ATC TCC TCT TGA TGA AAT
Peccary     ATG ACC AAT ATC CGA AAA TCC CAC CCA CTA ATA AAA ATT ATT AAC AAC ACA TTC ATC GAC TTA CCA ACC CCA TCA AAT ATT TCA TCA TGA TGA AAC
Pig         ATG ACC AAC ATC CGA AAA TCA CAC CCA CTA ATA AAA ATT ATC AAC AAC GCA TTC ATT GAC CTC CCA GCC CCC TCA AAC ATC TCA TCA TGA TGA AAC
Zebra       ATG ACA AAC ATC CGA AAA TCC CAC CCG CTA ATT AAA ATC ATC AAT CAT TCT TTC ATC GAC CTA CCA GCC CCC TCA AAC ATC TCA TCA TGA TGA AAC
Rhino       ATG ACT AAC ATC CGT AAA TCC CAC CCA CTA ATC AAA ATT ATC AAT CAC TCA TTC ATC GAC CTA CCC ACC CCA TCA AAC ATT TCA GCC TGA TGA AAT
Elephant    ATG ACC GAC ATT CGA AAA TCT CAC CCC TTA CTT AAA ATC ATC AAT AAA TCC TTC ATT GAT CTA CCT ACC CCA TCC AAC ATA TCA ACA TGA TGA AAT
Rat         ATG ACA AAC ATC CGA AAA TCT CAC CCC CTA TTC AAA ATC ATC AAC CAC TCC TTT ATC GAC CTA CCG GCC CCA TCT AAC ATC TCA TCA TGA TGA AAC
Mouse       ATG ACA AAC ATA CGA AAA ACA CAC CCA TTA TTT AAA ATT ATT AAC CAC TCA TTC ATT GAC CTA CCT GCC CCA TCC AAC ATT TCA TCA TGA TGA AAC
Human       ATG ACC CCA ATA CGC AAA ATT AAC CCC CTA ATA AAA TTA ATT AAC CAC TCA TTC ATC GAC CTC CCC ACC CCA TCC AAC ATC TCC GCA TGA TGA AAC
                                                                             14800
```

```
                                         40                                                                              60
             F   G   S   L   L   G   I   C   L   I   L   Q   I   L   T   G   L   F   L   A   M   H   Y   T   S   D   T   T   T   A   F   S
Cow         TTC GGT TCC CTC CTG GGA ATC TGC CTA ATC CTA CAA ATC CTC ACA GGC CTA TTC CTA GCA ATA CAC TAC ACA TCC GAC ACA ACA ACA GCA TTC TCC
Sheep       TTT GGC TCT CTC CTA GGC ATT TGC TTA ATT TTA CAG ATT CTA ACA GGC CTA TTC CTA GCA ATA CAC TAT ACA CCT GAC ACA ACA ACA GCA TTC TCC
Goat        TTT GGA TCC CTC CTA GGA ATT TGC CTA ATC TTA CAA ATC CTG ACA GGC CTA TTC CTA GCA ATA CAC TAT ACA TCC GAC ACA ATA ACA GCA TTT TCC
Pronghorn   TTC GGC TCT CTA CTA GGC ATC TGC CTA ATC CTA CAA ATC CTA ACA GGC CTA TTT CTA GCA ATA CAC TAC ACA GCA GAC ACA ACA ACA GCA TTC TCT
Giraffe     TTC GGC TCC CTA CTA GGC ATC TGC CTC ATT TTA CAA ATT CTA ACA GGC CTA TTT CTA GCA ATA CAC TAC ACA CCT GAC ACA ACA ACA GCG TTC TCC
Fallow      TTT GGC TCC CTA CTA GGA ATT TGC TTA ATC CTA CAA ATC CTC ACA GGC CTA TTC CTA GCA ATA CAC TAC ACA TCA GAT ACA ATA ACA GCA TTC TCC
Black-tail  TTC GGC TCT CTG CTA GGA ATT TGC TTA ATT CTA CAA ATT CTT ACC GGA CTA TTT CTA GCA ATA CAC TAT ACA TCC GAC ACA ATA ACA GCA TTT TCT
Chevrotain  TTC GGC TCC CTG TTA GGC ATC TGC CTC ATC CTT CAA ATC CTA ACA GGC CTA TTC CTG GCA ATG CAC TAC ACA TCA GAC ACA TCC ACA GCC TTC TCC
Camel       TTC GGC TCT CTA CTA GGA GTT TGC TTA ATT ATA CAA ATC CTG ACA GGA CTA TTC CTG GCA ATA CAT TAT ACA TCC GAC ACA ACA ACA GCC TTC TCT
Dolphin 1a  TTT GGT TCC TTA CTA GGC CTT TGC CTA ATT ATA CAA ATC CTA ACA GGC CTA TTT CTA GCA ATA CAT TAC ACG CCA GAC ACA ACA ACT GCT TTC TCA
Dolphin 1b  TTT GGT TCC TTA CTA GGC CTT TGC CTA ATT ATA CAA ATC CTA ACA GGT TTA TTT CTA GCA ATA CAT TAC ACG CCA GAC ACC TCA ACT GCT TTC TCA
Dolphin 2   TTT GGT TCC TTA CTA GGC CTC TGC CTA ATT ATA CAA ATC CTA ACA GGA TTA TTT CTA GCA ATA CAT TAC ACG CCA GAC ACC TCA ACT GCT TTC TCA
Peccary     TTC GGA TCC CTC CTG GGC ATC TGC CTG CTA CTA CAG ATC CTA ACA GGC CTA TTC CTA GCC ATA CAC TAC ACA CCA GAC ACA ACT ACC GCC TTT TCA
Pig         TTC GGT TCC CTC TTA GGC ATC TGC CTA ATC TTG CAA ATC CTA ACA GGC CTG TTC TTA GCA ATA CAT TAC ACA TCA GAC ACA ACA ACA GCT TTC TCA
Zebra       TTT GGC TCC CTC CTA GGA ATT TGC CTA ATC CTC CAA ATC CTA ACA GGC CTA TTC CTA GCT ATA CAC TAC ACA TCA GAC ACA ACA ACT GCC TTC TCA
Rhino       TTT GGC TCT CTA CTA GGA ATC TGC CTA ATC CTA CAA ATC CTA ACC GGA CTA TTT CTT GCT ATA CAT TAT ACA CCA GAC ACA ACA ACT GCC TTC TCA
Elephant    TTC GGC TCA CTA CTA GGA GCA TGC CTA ATT ACA CAA ATC CTA ACA GGA TTA TTC CTA GCC ATA CAT TAT ACA CCC GAC ACA ATA ACT GCA TTT TCA
Rat         TTC GGT TCT CTA CTA GGA GTA TGC CTC ATA GTA CAA ATC CTC ACA GGC TTA TTC CTA GCA ATA CAC TAC ACG TCT GAT ACC ATA ACA GCA TTC TCA
Mouse       TTT GGG TCC CTT CTA GGA GTC TGC CTA ATA GTC CAA ATC ATT ACA GGT CTT TTC TTA GCC ATA CAC TAC ACA TCA GAT ACA ATA ACA GCC TTT TCA
Human       TTC GGC TCA CTC CTT GGC GCC TGC CTG ATC CTC CAA ATC ACC ACA GGA CTA TTC CTA GCC ATG CAC TAC TCA CCA GAC GCC TCA ACC GCC TTT TCA
                                                                                     14900
```

```
                                                                         80
             S   V   T   H   I   C   R   D   V   N   Y   G   W   I   I   R   Y   M   H   A   N   G   A   S   M   F   F   I   C   L   Y   M
Cow         TCT GTT ACC CAT ATC TGC CGA GAC GTG AAC TAC GGC TGA ATC ATC CGA TAC ATA CAC GCA AAC GGA GCT TCA ATG TTT TTT ATC TGC TTA TAT ATG
Sheep       TCT GTA ACC CAC ATT TGC CGA GAC GTA AAC TAT GGC TGA ATT ATC CGA TAT ATA CAC GCA AAC GGG GCA TCA ATA TTT TTT ATC TGC CTA TTT ATG
Goat        TCT GTA ACT CAC ATT TGT CGA GAT GTA AAT TAT GGC TGA ATC ATC CGA TAC ATA CAC GCA AAC GGA GCA TCA ATA TTC TTT ATC TGC CTA TTC ATA
Pronghorn   TCT GTC ACC CAT ATT TGC CGA GAC GTA AAC TAC GGC TGA ATC ATC CGA TAC ATG CAC GCC AAC GGA GCA TCA ATA TTC TTC ATC TGC CTA TTT ATG
Giraffe     TCT GTC ACC CAT ATT TGC CGA GAT GTT AAC TAC GGT TGA ATC ATC CGA TAT ATA CAC GCA AAT GGG GCA TCC ATA TTC TTC ATC TGC TTA TTC ATG
Fallow      TCT GTC ACC CAT ATC TGC CGA GAC GTC AAC TAC GGC TGA ATC ATC CGA TAC ATG CAC GCA AAC GGA GCA TCA ATA TTC TTC ATC TGC CTA TTT ATG
Black-tail  TCT GTC ACC CAT ATC TGC CGA GAC GTC AAC TAT GGC TGA ATC ATT CGA TAT ATA CAT GCC AAT GGA GCA TCC ATA TTC TTC ATC TGC CTA TTT ATA
Chevrotain  TCA GTA ACC CAC ATC TGT CGA GAC GTG AAT TAT GGA TGA ATT ATC CGT TAT ATA CAC GCA AAT GGA GCC TCC ATA TTC TTT ATT TGC CTA TAC ATG
Camel       TCA GTC GCA CAC ATC TGC CGA GAT GTT AAC TAC GGC TGA ATC ATT CGA TAT TTA CAT GCT AAC GGA GCT TCC ATA TTC TTC ATT TGC TTA TAT ATT
Dolphin 1a  TCA GTC GCA CAT ATC TGC CGA GAC GTC AAC TAT GGC TGA TTT ATC CGC TAT CTA CAT GCA AAC GGA GCC TCC ATA TTC TTC ATC TGC CTA TAC GCC
Dolphin 1b  TCA GTC GCA CAT ATC TGC CGA GAC GTC AAC TAT GGC TGA TTT ATC CGC TAT CTA CAT GCA AAC GGA GCC TCC ATA TTC TTC ATC TGC CTA TAC GCC
Dolphin 2   TCA GTC GCA CAC ATC TGC CGA GAC GTC AAC TAT GGC TGA TTC ATC CGC TAT TTA CAT GCA AAC GGA GCC TCC ATA TTC TTC ATC TGT CTA TAC GCC
Peccary     TCA GTA ACC CAC ATC TGT CGA GAC GTA AAC TAC GGA TGA ATT ATC CGC TAC CTA CAT GCA AAC GGA GCA TCC ATA TTC TTT ATC TGC CTC TTT ATT
Pig         TCA GTT ACA CAC ATT TGT CGA GAC GTA AAT TAC GGA TGA GTT ATT CGC TAT CTA CAT GCA AAC GGA GCA TCC ATA TTC TTT ATT TGC CTA TTC ATC
Zebra       TCC GTC ACC CAC ATC TGC CGA GAC GTT AAC TAC GGA TGA ATC ATT CGC TAT CTC CAT GCC AAC GGA GCA TCC ATA TTT TTC ATC TGC CTC TTC ATC
Rhino       TCC GTT GCC CAC ATC TGT CGA GAG GTA AAC TAC GGC TGA ATT ATC CGC TAC CTA CAT GCA AAC GGA GCA TCC ATA TTT TTT ATC TGC CTA TTC ATC
Elephant    TCT ATA TCC CAT ATT TGC CGA GAT GTA AAC TAC GGC TGA ATT ATT CGA CAA CTA CAC TCA AAC GGA GCA TCC ATT TTC TTC CTC TGC CTA TAC ACA
Rat         TCA GTC ACC CAC ATC TGC CGA GAC GTA AAC TAC GGC TGA CTA ATC CGA TAC CTA CAC GCC AAC GGC GCC TCA ATA TTT TTC ATC TGC CTA TTC CTC
Mouse       TCA GTA ACA CAC ATT TGT CGA GAC GTA AAT TAC GGG TGA CTA ATC CGA TAT ATA CAC GCA AAC GGA GCC TCA ATA TTT TTT ATT TGC TTA TTC CTT
Human       TCA ATC GCC CAC ATC ACT CGA GAC GTA AAT TAT GGC TGA ATC ATC CGC TAC CTT CAC GCC AAT GGC GCC TCA ATA TTC TTT ATC TGC CTC TTC CTA
                                                                                         15000
```

```
                         100                                                                             120
             H   V   G   R   G   L   Y   Y   G   S   Y   T   F   L   E   T   W   N   I   G   V   I   L   L   L   T   V   M   A   T   A   F
Cow         CAC GTA GGA CGA GGC TTA TAT TAC GGG TCT TAC ACT TTT CTA GAA ACA TGA AAT ATT GGA GTA ATC CTT CTG CTC ACA GTA ATA GCC ACA GCA TTT
Sheep       CAT GTA GGA CGA GGC CTA TAC TAT GGA TCA TAT ACC TTC CTA GAA ACA TGA AAC ATC GGA GTA ATC CTC CTA TTT GCG ACA ATA GCC ACA GCA TTC
Goat        CAT ATC GGA CGA GGT CTA TAT TAT GGA TCA TAT ACC TTT CTA GAA ACA TGA AAC ATT GGA GTA ATC CTC CTG CTC GCG ACA ATG GCC ACA GCA TTC
Pronghorn   CAT GTA GGA CGA GGC CTA TAC TAC GGA TCG TAT ATA TTC CTA GAA ACA TGA AAC ATT GGA GTA ATC CTC CTA TTC ACA GTT ATA GCT ACA GCA TTC
Giraffe     CAT GTA GGA CGG GGC CTA TAC TAC GGA TCG TAT ACC TTC CTA GAA ACA TGA AAC ATT GGA GTA ATC CTC CTA TTC ACA GTA ATA GCC ACA GCT TTT
Fallow      CAT GTA GGA CGA GGC CTA TAC TAT GGA TCG TAT ATA TTC CTA GAA ACA TGA AAC ATT GGA GTA ATT CTC CTA TTC ACA GTT ATA GCT ACA GCA TTC
Black-tail  CAT GTG GGA CGA GGA CTG TAC TAC GGA TCA TAC ACT TTC CTA GAA ACA TGA AAT ATT GGA GTA ATT CTC CTA TTC ACA GTT ATA GCT ACA GCA TTT
Chevrotain  CAT GTA GGA CGA GGC TTA TAC TAC GGA TCC TAC ACC TTC CTA GAA ACG TGA AAC ATT GGA GTA ATC CTC CTA CTT ACA GTG ATA GCC ACA GCA TTC
Camel       CAC GTG GGT CGC GGG CTT TAT TAC GGC TCG TAT ACC TTT TCA GAA ACC TGA AAC GTT GGA ATG GTT TTA TTG TTC ACA GTA ATA GCA ACA GCT TTC
Dolphin 1a  CAC ATG GGA CGA GGC CTG TAC TAT GGC TCC TAT ATA TTC CAA GAA ACA TGA AAC ATT GGT GTG CTC CTA CTA TTA ACA GTC ATA GCC ACT GCA TTC
Dolphin 1b  CAC ATT GGA CGT GGC CTG TAC TAT GGC TCC TAT ATA TTC CAA GAA ACA TGA AAC ATT GGC GTG CTC CTA CTA TTA ACA GTC ATA GCC ACT GCA TTC
Dolphin 2   CAC ATT GGA CGT GGC CTA TAC TAT GGC TCT TAC ATA TTC CAA GAA ACA TGA AAC ATT GGT GTA CTC CTA CTA TTA ACA GTC ATA GCT ACT GCA TTC
Peccary     CAC GTA GGC CGA GGA CTA TAC TAT GGG TCC TAC CTA TTT CTA GAA ACA TGA AAT ATC GGA GTC ATT CTC CTA CTA ACA GTA ATA GCA ACA GCC TTT
Pig         CAC GTA GGC CGA GGT CTA TAC TAC GGA TCC TAT ATA TTC CTA GAA ACA TGA AAC ATT GGA GTA GTC CTA CTA TTT ACC GTT ATA GCA ACA GCC TTC
Zebra       CAC GTA GGA CGT GGC CTC TAC TAT GGC TCT TAC ACA TTC CTA GAA ACA TGA AAC ATT GGA ATT ATC CTA CTT CTC ACA GTA ATA GCC ACA GCA TTC
Rhino       CAC ATA GGA CGC GGC CTC TAT TAC GGA TCC TAC ACC TTC CTA AAA ACC TGA AAC ATC GGA GTT ATT CTA CTA CTC ACA GTA ATA GCC ACA GCA TTC
Elephant    CAC ATT GGA CGA AAC ATC TAC TAT GGG TCC TAC CTA TAC TCG GAA ACT TGA AAT ACC GGC ATT ATA TTA CTA CTA ATC ACC ATA GCC ACC GCC TTC
Rat         CAT GTG GGA CGA GGA CTA TAC TAT GGA TCC TAC ACT TTC CTA GAA ACC TGA AAC ATT GGG ATC ATC CTA CTA TTT GCA GTC ATA GCA ACT GCA TTC
Mouse       CAT GTC GGA CGA GGC TTA TAT TAT GGA TCA TAT ACA TTT ATA GAA ACC TGA AAC ATT GGA GTA CTT CTA CTG TTC GCA GTC ATA GCC ACA GCA TTT
Human       CAC ATC GGG CGA GGC CTA TAT TAC GGA TCA TTT CTC TAC TCA GAA ACC TGA AAC ATC GGC ATT ATC CTC CTG CTT GCA ACT ATA GCA ACA GCC TTC
                                                                                             15100
```

Fig. 2. Sequences of mammalian cytochrome *b* genes. The DNA sequences of 17 ungulate cytochrome *b* genes are shown along with those from human (Anderson et al. 1981), mouse (Bibb et al. 1981), and rat (Gadaleta et al. 1989). Numbering is according to the human sequence. The predicted amino acid sequence of the cow protein is shown above in the single-letter amino acid code. Dashes between codons 325 and 326 indicate gaps introduced to maximize sequence identity. [Codon 83 in the rat is shown as CAC, correcting an error in the published sequence (Saccone, personal communication).]

140   160

	M G Y V L P W G Q M S F W G A T V I T N L L S A I P Y I G T N L
Cow	ATA GGA TAC GTC CTA CCA TGA GGA CAA ATA TCA TTC TGA GGA GCA ACA GTC ATC ACC AAC CTC TTA TCA GCA ATC CCA TAC ATC GGC ACA AAT TTA
Sheep	ATA GGC TAT GTC TTA CCA TGA GGA CAA ATA TCA TTC TGA GGA GCA ACA GTT ATT ACC AAC CTC CTT TCA GCA ATT CCA TAT ATT GGC ACA AAC CTA
Goat	ATA GGC TAT GTC CTA CCA TGA GGA CAA ATA TCA TTT TGA GGG GCA ACA GTC ATC ACC AAT CTC CTC TCA GCA ATC CCA TAT ATT GGC ACA AAC CTA
Pronghorn	ATA GGA TAC GTC CTA CCA TGA GGA CAA ATA TCA TTC TGA GGG GCA ACA GTC ATT ACT AAC CTA CTC TCA GCT ATC CCA TAC ATT GGT ACT AAC CTA
Giraffe	ATA GAG TAC GTC CTA CCA TGA GGA CAA ATA TCA TTT TGA GGA GCA ACA GTC ATT ACT AAT CTC CTA TCA GCA ATC CCA TAT ATC GGC ACA AAT CTA
Fallow	ATA GGA TAC GTC CTA CCA TGA GGA CAA ATA TCA TTC TGA GGA GCA ACA GTC ATT ACT AAC CTC CTC TCA GCA ATC CCA TAC ATT GGT ACA AAC CTA
Black-tail	GTA GGG TAC GTC CTA CCA TGA GGA CAA ATA TCA TTC TGA GGG GCA ACA GTC ATC ACA AAT CTC CTC TCA GCA ATC CCA TAT ATC GGC ACA AAT CTA
Chevrotain	ATA GGA TAT GTA CTA CCC TGA GGG CAA ATA TCT TTT TGA GGA GCT ACA GTC ATC ACT AAC CTT CTT TCA GCA ATC CCC TAT ATC GGC ACC GAA CTA
Camel	ATG GGC TAC GTC CTT CCA TGA GGA CAG ATA TCA TTC TGA GGA GCA ACA GTA ATT ACT AAT TTG CTC TCA GCA ATC CCC TAT ATC GGC ACA ACA CTA
Dolphin 1a	GTA GGC TAC GTC CTA CCC TGA GGA CAA ATA TCA TTC TGA GGC GCA ACG GTC ATT ACC AAC CTC CTA TCA GCA ATC CCT TAT ATT GGC ACT ACC CTA
Dolphin 1b	GTA GGC TAC GTC CTA CCC TGA GGA CAA ATA TCA TTC TGA GGC GCA ACG GTC ATT ACC AAC CTC CTA TCA GCA ATC CCT TAT ATT GGC ACT ACC CTA
Dolphin 2	GTA GGC TAC GTC CTG CCC TGA GGA CAA ATA TCA TTC TGA GGC GCA ACG GTA ATT ACA AAT CTC CTA TCA GCA ATC CCT TAT ATT GGC ACT ACC CTA
Peccary	ATA GGA TAC GTC CTA CCA TGA GGA CAA ATA TCT TTT TGA GCA GCT ACA GTC ATC ACC AAC CTC CTA TCA GCC ATC CCA TAC ATT GGT ACC GAT CTG
Pig	ATA GGC TAC GTC CTG CCC TGA GGA CAA ATA TCA TTC TGA GGA GCT ACG GTC ATC ACA AAT CTA CTA TCA GCT ATC CCT TAT ATC GGA ACA GAC CTC
Zebra	ATA GGC TAC GTC CTA CCA TGA GGC CAA ATA TCC TTC TGA GGA GCA ACA GTT ATT ACA AAT CTC CTG TCA GCA ATT CCC TAC ATC GGC ACT ACC CTC
Rhino	ATA GGC TAT GTC CTA CCA TGA GGC CAA ATA TCC TTT TGA GGA GCA ACA GTC ATT ACA AAC CTA CTC TCA GCT ATC CCG TAT ATC GGC ACA ACT CTC
Elephant	ATA GGA TAT GTC CTT CCG TGA GGA CAA ATA TCA TTC TGA GGG GCA ACC GTA ATC ACT AAC CTT CTC TCA GCA ATC CCT TAT ATC GGC ACA AAC CTA
Rat	ATG GGC TAT GTA CTC CCA TGA GGA CAA ATA TCA TTC TGA GGA GCT ACA GTA ATT ACA AAC CTA TTA TCA GCT ATC CCT TAC ATT GGG ACT ACC CTA
Mouse	ATA GGC TAC GTC CTT CCA TGA GGA CAA ATA TCA TTC TGA GGT GCC ACA GTT ATT ACA AAC CTC CTA TCA GCC ATC CCA TAT ATT GGA ACA ACC CTA
Human	ATA GGC TAT GTC CTC CCG TGA GGC CAA ATA TCA TTC TGA GGG GCC ACA GTA ATT ACA AAC TTA CTA TCC GCC ATC CCA TAC ATT GGG ACA GAC CTA

15200

180

	V E W I W G G F S V D K A T L T R F F A F H F I L P F I I M A I
Cow	GTC GAA TGA ATC TGA GGC GGA TTC TCA GTA GAC AAA GCA ACC CTT ACC CGA TTC TTC GCT TTC CAT TTT ATC CTT CCA TTT ATC ATC ATA GCA ATT
Sheep	GTC GAA TGA ATC TGG GGA GGA TTC TCA GTA GAC AAA GCT ACC CTC ACC CGA TTT TTC GCC TTT CAC TTT ATT TTC CCA TTC ATC ATC GCA GCC CTC
Goat	GTC GAA TGA ATC TGA GGG GGA TTC TCA GTA GAC AAA GCC ACT CTC ACC CGA TTC TTC GCC TTC CAC TTT ATC CTC CCA TTC ATC ATC ACA GCC CTC
Pronghorn	GTA GAA TGA ATC TGA GGG GGA TTC TCA GTA GAC AAA GCA ACC CTC ACC CGA TTC TTC GCA TTC CAC TTT ATC CTC CCA TTC ATC ATT GCA GCA CTA
Giraffe	GTC GAA TGA ATC TGA GGG GGC TTC TCA GTA GAT AAA GCA ACC CTT ACC CGA TTC TTC GCC TTC CAC TTC ATT CTC CCA TTC ATC ATT ATA GCA CTC
Fallow	GTT GAA TGA ATC TGA GGG GGC TTT TCA GTA GAC AAA GCA ACC TTA ACT CGA TTC TTC GCT TTC CAC TTT ATT CTA CCA TTT ATC ATT GCA GCA CTT
Black-tail	GTC GAA TGA ATC TGA GGA GGC TTT TCC GTA GAC AAA GCA ACC CTA ACT CGA TTC TTC GCT TTC CAC TTT ATC CTC CCA TTT ATC ATC GCA GCA CTT
Chevrotain	GTT GAA TGA ATC TGA GGC GGG TTC TCA GTA GAC AAA GCA ACC CTT ACA CGA TTT TTT GCC TTC CAC TTC ATC CTC CCA TTT GTC ATT ACA GCC CTA
Camel	GTA GAA TGG ATT TGA GGT GGT TTC TCC GTA GAT AAA GCC ACC CTC ACA CGA TTC TTC GCC TTC CAT TTC ATC CTA CCA TTC ATT ATC ACG GCT CTA
Dolphin 1a	GTT GAA TGA ATC TGA GGT GGA TTT TCC GTA GAC AAA GCA ACA TTA ACA CGC TTT TTC GCT TTC CAT TTT ATC CTC CCA TTC ATC ATC ACA GCA TTA
Dolphin 1b	GTT GAA TGA ATC TGA GGA GGA TTT TCC GTA GAC AAG GCA ACA TTA ACA CGC TTT TTC GCT TTC CAT TTT ATC CTC CCA TTC ATC ATC ACA GCA TTA
Dolphin 2	GTT GAA TGA ATC TGA GGA GGA TTT TCC GTA GAC AAG GCA ACA CTA ACA CGA TTC TTC GCT TTC CAC TTT ATC CTC CCG TTC ATC ATT ACA GCA TTA
Peccary	GTA GAA TGA ATT TGA GGG GGA TTC TCA GTA GAT AAA GCT ACC CTC ACA CGA TTC TTC GCC TTC CAC TTC ATC CTC CCA TTT ATC ATC ACC GCC CTA
Pig	GTA GAA TGA ATC TGA GGG GGC TTT TCC GTC GAC AAA GCA ACC CTC ACA CGA TTC TTC GCC TTC CAC TTT ATC CTG CCA TTC ATC ATT ACC GCC CTC
Zebra	GTC GAG TGA ATC TGA GGC GGA TTC TCG GTA GAC AAA GCC ACC CTC ACC CGA TTT TTT GCC TTC CAC TTC ATC CTA CCA TTT ATC ATC ACA GCC CTG
Rhino	GTA GAA TGA ATC TGA GGA GGG TTC TCC GTC GAC AAA GCC ACA CTT ACA CGA TTC TTC GCT TTC CAC TTT ATT CTC CCC TTT ATC ATC TCA GCC CTA
Elephant	GTA GAA TGA ATC TGA GGA GGC TTT TCA GTA GAC AAA GCA ACC TTA AAT CGA TTT TTC GCC CTC CAT TTC ATT CTT CCA TTT ACT ATA ATT GCA CTA
Rat	GTC GAA TGA ATC TGA GGA GGC TTC TCA GTA GAC AAA GCA ACC CTA ACA CGC TTC TTC GCA TTC CAC TTC ATC CTC CCA TTC ATT ATC GCC GCC CTT
Mouse	GTC GAA TGA ATT TGA GGG GGC TTC TCA GTA GAC AAA GCC ACC TTG ACC CGA TTC TTC GCT TTC CAC TTC ATC TTA CCA TTT ATT ATC GCG GCC CTA
Human	GTT CAA TGA ATC TGA GGA GGC TAC TCA GTA GAC AGT CCC ACC CTC ACA CGA TTC TTT ACC TTT CAC TTC ATC TTG CCC TTC ATT ATT GCA GCC CTA

15300

200   220

	A M V H L L F L H E T G S N N P T G I S S D V D K I P F H P Y Y
Cow	GCC ATA GTC CAC CTA CTA TTC CTC CAC GAA ACA GGC TCC AAC AAC CCA ACA GGA ATT TCC TCA GAC GTA GAC AAA ATC CCA TTC CAC CCC TAC TAT
Sheep	GCC ATA GTT CAC CTA CTC TTC CTC CAC GAA ACA GGA TCC AAC AAC CCC ACA GGA ATT CCA TCG GAC ACA GAT AAA ATT CCC TTC CAC CCT TAT TAC
Goat	GCC ATA GTC CAC CTA CTC TTC CTC CAC GAA ACA GGA TCC AAC AAC CCC ACA GGA ATT CCA TCA GAC ACA GAT AAA ATC CCA TTT CAC CCT TAC TAC
Pronghorn	GCC ATA GTA CAC TTA CTA TTC CTC CAC GAA ACA GGA TCC AAC AAC CCC ACA GGA ATC CCA TCA GAC GCA GAC AAA ATC CCA TTC CAC CCA TAC TAC
Giraffe	ACT ATA GTC CAC CTA CTC TTT CTT CAC GAA ACA GGA TCA AAC AAC CCA ATA GGA ATT CCA TCA GAC ATA GAC AAA ATC CCG TTC CAC CCC TAC TAC
Fallow	GCT ATA GTA CAT TTA CTA TTC CTC CAC GAG ACA GGA TCC AAT AAC CCA ACA GGA ATC CCA TCA GAT GCA GAT AAA ATC CCA TTT CAC CCC TAC TAC
Black-tail	GCT ATA GTC CAT TTA CTC TTC CTC CAC GAA ACA GGA TCT AAT AAT CCA ACG GGA ATC CCA TCA GAC GCA GAC AAA ATC CCA TTC CAC CCC TAC TAC
Chevrotain	GCC CTA GTC CAT CTT TTA TTT CTC CAC GAG ACA GGA TCA AAT AAC CCC ACA GGA ATC CCC TCA GAC GCA GAC AAG ATC CCC TTC CAC CCA TAC TAC
Camel	GTG GCC GTA CAC CTA TTA TTT CTA CAC GAA ACA GGT TCT AAT AAC CCA ACA GGA ATT TCC TCA GAC ATA GAT AAA ATC CCA TTC CAT CCC TAT TAC
Dolphin 1a	GCA GCC GTC CAC CTA CTA TTC CTA CAC GAA ACA GGA TCC AAT AAC CCC ACA GGA ATC CCA TCC AAC ATA GAC ATA ATC CCA TTC CAC CCT TAT TAT
Dolphin 1b	GCA GCC GTC CAC CTA CTA TTC CTA CAC GAA ACA GGA TCC AAT AAC CCC ACA GGA ATC CCA TCC AAC ATA GAC ATA ATC CCA TTC CAC CCT TAT TAT
Dolphin 2	GCA GCC GTT CAC TTG CTA TTC CTA CAC GAA ACA GGA TCC AAC AAT CCC ACA GGA ATC CCA TCC AAC ATA GAC ATA ATC CCA TTC CAT CCT TAT TAT
Peccary	GTA ATC GTA CAT TTA CTC TTT CTT CAC GAA ACA GGA TCA AAC AAC CCC ACA GGA ATC CCA TCC AAC ATA GAC AAA ATT CCA TTC CAC CCA TAC TAT
Pig	GCA GCC GTA CAT CTC ATA TTC CTG CAC GAA ACC GGA TCC AAC AAC CCT ACC GGA ATC TCA TCA GAC ATA GAC AAA ATT CCA TTT CAC CCA TAC TAC
Zebra	GTA ATC GTC CAT TTA CTA TTC CTC CAC GAA ACA GGA TCC AAC AAC CCC TCA GGA ATT CCA TCT GAC ATA GAC AAA ATC CCA TTT CAC CCA TAC TAC
Rhino	GCA ATC ACC CAC CTA TTA TTC CTA CAC GAA ACA GGA TCC AAC AAC CCA TCA GGA ATC CCA TCC AAT ATA GAC AAA ATT CCA TTC CAC CCA TAC TAC
Elephant	GCA GGA GTA CAC CTA ACC TTT CTT CAC GAA ACA GGC TCA AAC AAT CCA CTG GGC CTC ACT TCA GAC TCA GAC AAA ATC CCC TTT CAC CCA TAC TAT
Rat	GCA ATT GTA CAT CTT CTT TTC CTC CAC GAA ACA GGA TCA AAT AAC CCC ACA GGA TTA AAC TCC GAC GCA GAC AAA ATC CCA TTC CAT CCA TAT TAT
Mouse	GCA ATC GTT CAC CTC CTC TTC CTC CAC GAA ACA GGA TCA AAC AAC CCA ACA GGA TTA AAC TCA GAT GCA GAT AAA ATT CCA TTT CAC CCC TAC TAT
Human	GCA ACA CTC CAC CTC CTA TTC TTG CAC GAA ACG GGA TCA AAC AAC CCC CTA GGA ATC ACC TCC CAT TCC GAT AAA ATC ACC TTC CAC CCT TAC TAC

15400

240

	T I K D I L G A L L L I L A L M L L V L F A P D L L G D P D N Y
Cow	ACC ATT AAG GAC ATC TTA GGG GCC CTC TTA CTA ATT CTA GCT CTA ATA CTA CTA GTA CTA TTC GCA CCC GAC CTC CTC GGA GAC CCA GAT AAC TAC
Sheep	ACC ATT AAA GAC ATC CTA GGT GCT ATC CTA CTA ATC CTC ATC CTC ATG CTA CTA GTA CTA TTC ACA CCT GAC TTA CTC GGA GAC CCA GAT AAC TAT
Goat	ACC ATT AAA GAT ATC TTA GGC GCC ATG CTA CTA ATT CTT GTC CTA ATA CTA CTA GTA CTA TTC ACA CCC GAC CTT CTC GGA GAC CCA GAT AAC TAC
Pronghorn	ACC ATC AAA GAC ATT CTA GGA GCA CTA CTA ATA ATC TTA GCC CTA ATA ATA CTA GTA CTA TTC TCA CCA GAC CTG TTA GGA GAC CCC GAC AAC TAC
Giraffe	ACT ATC AAA GAT ATC TTA GGG GCA CTT CTA TTA ATT CTA GTC CTA ATA CTT CTA GTC CTA TTT ACA CCC GAC CTG CTC GGA GAC CCA GAT AAC TAC
Fallow	ACC ATC AAA GAT ATC TTA GGG GCA CTA CTT ATA ATC TTA GTC CTA ATA ATA CTA GTA CTA TTC TCA CCA GAC GTG TTA GGA GAC CCC GAC AAC TAC
Black-tail	ACT ATT AAA GAT ATC CTA GGA GCC CTA CTT CTA ACT CTC TTC CTA ATA CTA TTA GTA CTA TTC GCA CCA GAC CTA CTC GGA GAC CCA GAT AAT TAT
Chevrotain	ACC ATC AAA GAT GTC CTA GGG GCT CTA GTC CTA ATA CTA GTC CTT CTA TTA CTA GTC CTA TTT TCA CCG GAC TTG TTG GGA GAC CCC GAC AAT TAC
Camel	ACG ATT AAA GAT ATC CTA GGA GCA CTA CTA CTG ATG CTA GCC CTA CTT ATC CTC GTA TTA TTC TCA CCA GAC TTA TTA GGA GAC CCT GAC AAC TAC
Dolphin 1a	ACA ATC AAA GAC ATC CTA GGT GGC TTA CTC TTA ATC TTA ACC CTA CTA GCA CTA ACC CTA TTC ACC CCT GAC TTA CTA GGA GAC CCT GAT AAC TAT
Dolphin 1b	ACA ATC AAA GAC ATC CTA GGT GGC TTA CTC TTA ATC TTA ACC CTA CTA GCA CTA ACC CTA TTC ACC CCT GAC TTA CTA GGA GAC CCT GAT AAC TAT
Dolphin 2	ACA ATC AAA GAT ATC CTT GGC GGA TTA CTC TTA ATC TTA ACC CTA CTA GCA CTA ACC CTA TTC ACC CCT GAC CTA CTA GGA GAC CCT GAC AAC TAT
Peccary	ACA ATT AAA GAC ATT CTA GGA GCC ACA CTC ATA ATC CTA ATC CTA CTA CTA CTA GTA CTA TTC TCA CCA GAC CTC CTA GGA GAC CCG GAC AAC TAC
Pig	ACT ATT AAA GAC ATT CTA GGA GCC TTA TTT ATA ATA CTA ATC CTA CTA ATC CTT GTA CTA TTC TCA CCA GAC CTA CTA GGA GAC CCA GAC AAC TAC
Zebra	ACA ATT AAA GAT ATC CTA GGA CTC CTC CTC CTA ATC CTA CTC CTA CTG ACC CTA GTA TTA TTC TCT CCT GAC CTC CTA GGA GAC CCG GAC AAC TAT
Rhino	ACA ATC AAA GAT ATC CTA GGA ATC CTA CTC CTA ATC CTA ACA CTA CTC ACC CTA GTC CTA TTC TCA CCC CAC CAT CTA GGA GAC CCT GAC AAC TAC
Elephant	ACC ATT AAG GAC TTC CTA GGA TTA CTT ATC CTA ATT TTA CTT CTT CTA CTC CTA GCC CTA CTA TCT CCT GAC CAT CTA GGA GAC CCT GAC AAC TAC
Rat	ACA ATT AAA GAC CTC CTA GGT GTA TTT ATA TTA CTA TTA TTC CTA ATA ACC CTA GTA CTA TTC TTC CCA GAC CTA CTA GGA GAC CCA GAC AAT TAT
Mouse	ACA ATC AAA GAT ATC CTA GGT ATC CTA ATC ATA TTC TTA ATT CTC ATA ACC CTA GTA TTA TTT TTC CCA GAC ATA CTA GGA GAC CCA GAC AAC TAC
Human	ACA ATC AAA GAC GCC CTC GGC TTA CTT CTC TTC CTT CTC TCC TTA ATG ACA TTA ACA CTA TTC TCA CCA GAC CTC CTA GGC GAC CCA GAC AAT TAT

15500

Fig. 2. Continued

```
                260                                                                                         280
            T   P   A   N   P   L   N   T   P   P   H   I   K   P   E   W   Y   F   L   F   A   Y   A   I   L   R   S   I   P   N   K   L
Cow         ACC CCA GCC AAT CCA CTC AAC ACA CCC CCT CAC ATC AAA CCC GAG TGA TAC TTC TTA TTT GCA TAC GCA ATC TTA CGA TCA ATC CCC AAC AAA CTA
Sheep       ACC CCA GCA AAC CCA CTT AAC ACA CCC CCT CAC ATT AAA CCT GAA TGA TAT TTC CTA TTT GCA TAC GCA ATC TTA CGA TCA ATC CCT AAT AAA CTA
Goat        ACC CCA GCA AAT CCA CTC AAC ACT CCC CCT CAC ATT AAG CCA GAA TGA TAT TTC CTA TTT GCA TAC GCA ATC CTA CGA TCA ATC CCC AAC AAA CTA
Pronghorn   ACA CCA GCT AAC CCA CTC AAC ACT CCC CCA CAC ATT AAG CCA GAA TGA TAT TTC CTA TTC GCA TAC GCA ATC CTA CGA TCA ATC CCT AAC AAA CTA
Giraffe     ACA CCA GCC AAC CCA CTC AAT ACT CCT CCC CAT ATT AAA CCA GAA TGA TAC TTT CTA TTT GCA TAC GCA ATC CTA CGA TCA ATC CCC AAC AAA CTA
Fallow      ACA CCA GCT AAC CCA CTC AAC ACT CCT CCA CTT ATT AAG CCA GAA TGA TAT TTC CTA TTC GCA TAC GCA ATC CTA CGA TCA ATC CCT AAC AAA CTA
Black-tail  ACT CCA GCA AAC CCA CTC AAT ACC CCT CCT CAC ATC AAA CCT GAA TGC TAC TTC TTG TTT GCA TAC GCA ATT CTA CGA TCA ATT CCA AAT AAA CTA
Chevrotain  ACT CCG GCA AAC CCC CTC AAC ACA CCA CCT CAT ATT AAG CCA GAG TGG TAT TTC CTA TTC GCA TAC GCA ATC CTA CGA TCA ATC CCC AAT AAA TTA
Camel       ACT CCC GCC AAC CCC CTC AAT ACA CCA CCA CAT ATC AAG CCG GAA TGA TAT TTC CTG TTC GCA TAC GCT ATT CTA CGA TCA ATC CCC AAC AAA TTA
Dolphin 1a  ACC CCA GCA AAT CCA CTA AGC ACC CCT CCA CAC ATC AAA CCA GAA TGA TAT TTC CTA TTC GCA TAC GCA ATC TTA CGA TCA ATC CCT AAC AAA CTT
Dolphin 1b  ACC CCA GCA AAT CCA CTA AGC ACC CCT GCA CAC ATC AAA CCA GAA TGA TAT TTC CTA TTC GCA TAC GCA ATC TTA CGA TCA ATC CCT AAC AAA CTT
Dolphin 2   ACT CCA GCA AAT CCA CTA AGC ACC CCT GCA CAC ATC AAA CCA GAA TGA TAC TTC CTA TTC GCA TAC GCA ATC TTA CGA TCA ATC CCT AAT AAA CTT
Peccary     ACC CCA GCA AAC CCA CTA AAC ACA CCT TCA CAC ATT AAA CCA GAA TGA TAC TTC CTA TTT GCC TAC GCA ATC CTA CGC TCA ATC CCA AAT AAA CTA
Pig         ACC CCA GCA AAC CCA CTA AAC ACC CCA CCC CAT ATT AAA CCA GAA TGA TAT TTC TTA TTC GCC TAC GCT ATT CTA CGT TCA ATT CCT AAT AAA CTA
Zebra       ACC CCA GCC AAC CCT CTC AGC ACT CCC CCT CAT ATT AAG CCA GAA TGA TAT TTC CTA TTT GCC TAC GCC ATC CTA CGC TCT ATT CCC AAC AAA CTA
Rhino       ACC CCC GCC ACC CCT CTC AAT ACC CCT CCA CAT ATC AAA CCA GAG TGA TAT TTC TTA TTT GCC TAT GCA ATC CTA CGA TCC GTC CCT AAC AAA CTA
Elephant    ACC CTG GCC AAC CCC TTA AAT AAC CCC CCT CAT ATC AAG CCA GAG TGA TAT TTT CTC TTT GCT TAC GCC ATC CTA CGA TCT GTA CCA AAC AAA CTA
Rat         ACA CCC GCT AAC CCC CTC AAC ACC CCA CCC CAC ATC AAA CCA GAA TGA TAC TTT CTC TTT GCC TAC GCT ATT CTA CGC TCC ATT CCC AAC AAA CTA
Mouse       ATA CCA GCT AAT CCA CTA AAC ACC CCA CCC CAT ATT AAA CCC GAA TGA TAT TTC CTA TTT GCA TAC GCC ATT CTA CGC TCA ATC CCC AAT AAA CTA
Human       ACC CTA GCC AAC CCC TTA AAC ACC CCT CCC CAC ATC AAG CCC GAA TGA TAT TTC CTA TTC GCC TAC ACA ATT CTC CGA TCC GTC CCT AAC AAA CTA
                                                                                                                          15600
```

```
                                                        300                                                                       320
            G   G   V   L   A   L   A   F   S   I   L   I   L   A   L   I   P   L   L   H   T   S   K   Q   R   S   M   M   F   R   P   L
Cow         GGA GGA GTA CTA GCC CTA GCC TTC TCT ATC CTA ATT CTT GCT CTA ATC CCC CTA CTA CAC ACC TCC AAA CAA CGA AGC ATA ATA TTC CGA CCA CTC
Sheep       GGA GGA GTC CTC GCC CTA GTC CTC TCG ATC CTA GTC CTA GTA ATT ATA CCC CTC CTC CAT ACA TCA AAG CAA CGG AGC ATA ATA TTC CGA CCA ATC
Goat        GGA GGA GTC CTA GCC CTA GTC CTC TCA ATC CTA ATC TTA GTA CTT GTA CCC TTC CTC CAC ACA TCT AAA CAA CGA AGC ATA ATA TTC CGC CCA ATC
Pronghorn   GGA GGA GTC CTA GCA CTA GTC CTC TCT ATT CTC ATC CTT ATC TTT ATA CCC CTG CTC CAC ACA TCC AAA CAG CGA AGC ATA ATA TTC CGA CCA TTC
Giraffe     GGA GGA GTC CTA GCC CTA GTT CTC TCC ATC CTC ATC CTC ATC TTC ATA CCC CTA CTC CAC ACA TCC AAA CAA CGA AGT ATG ATA TTT CGA CCA TTC
Fallow      GGA GGA GTC CTA GCA CTA GTC CTC TCT ATT CTC ATC CTT ATC TTT ATA CCC CTG CTC CAC ACA TCC AAA CAA CGA AGC ATG ATA TTT CGA CCA TTC
Black-tail  GGA GGA GTG TTA GCC CTA GTC CTC TCC ATC CTA ATC CTC GTC CTT ATA CCC CTA CTT CAC ACA TCC AAA CAA CGA AGT ATA ATA TTC CGA CCA TTC
Chevrotain  GGA GGG GTA CTA GCA CTC ATT GCA TCT ATC CTA ATT TTA CAG CTT ATA CCC CTG CTC CAC ACA TCC AAG CAA CGA AGC ATG ATA TTT CGA CCA ATC
Camel       GGA GGC GTA CTG GCC TTA GTC CTC TCC ATC TTA ATC CTA GCA TTT ATC CCC GCA CTA CAC ACA TCA AAA CAA CGA AGC ATA ACA TTC CGT CCC ATT
Dolphin 1a  GGA GGA GTA CTA GCA TTA TTA CTC TCC ATC CTT ATT CTA ATC TTT ATC CCA ATA CTT CAG ACA TCC AAA CAA CGA AGT ATA ATA TTC CGA CCC TTC
Dolphin 1b  GGA GGA GTA CTA GCA TTA TTA CTC TCC ATC CTT GTT CTA ATC TTT ATC CCA ATA CTT CAG ACA TCC AAA CAA CGA AGC ATA ATA TTC CGA CCC TTC
Dolphin 2   GCA GGA GTA CTA GCA CTA TTA CTT TCC ATC CTT GTC CTA ATC TTT ATC CCA ATG CTT CAG ACA TCC AAA CAA CGA AGC ATA ATA TTC CGA CCC TTC
Peccary     GGA GGA GTG CTA GCC CTA GCC CTC TCA ATC CTA ATC CTA GCC CTT GTG CCA GCA CTA CAC ACC TCC AAA CAA CGA AGC ATA ATG TTC CGA CCC CTC
Pig         GGT GGA GTG TTG GCC CTA GTA GCC TCC ATC CTA ATC CTA ATT TTA ATG CCC ATA CTG CAC ACA TCC AAA CAA CGA GGC ATA ATA TTT CGA CCA CTA
Zebra       GGC GGC GTA TTA GCC CTT ATC CTT TCC ATC TTA ATC TTA GCA CTC ATC CCT ACT CTA CAT ACA TCA AAA CAA CGA AGC ATA ATA TTC CGA CCT CTC
Rhino       GGC GGA GTA CTA GCC CTA GCA CTT TCC ATC CTA ATC CTA GCT CTT ATC CCC ATT CTC CAC ACA TCC AAA CAA CGA AGC ATA ATA TTC CGA CCC CTA
Elephant    GGA GGC GTC CTA GCC CTA CTC CTA TCA ATT CTA ATC CTA GGA TTA ATA CCA CTT CTC CAT ACA TCC AAG CAC CGA AGC ATA ATA CTC CGA CCT CTT
Rat         GGA GGG GTC GTA GCC CTA ATC TTA TCA ATC CTA ATC TTA GCC TTC CTA CCA TTC CTG CAT ACC TCA AAA CAA CGC AGC TTA ACA TTC CGC CCA ATC
Mouse       GGA GGT GTC CTA GCC TTA ATC TTA TCT ATC CTA ATT TTA GCC CTA ATA CCT TTC CTT CAT ACC TCA AAG CAA CGA AGC CTA ATA TTC CGC CCA ATC
Human       GGA GGC GTC CTT GCC CTA TTA CTA TCC ATC CTC ATC CTA GCA ATA ATC CCC ATC CTC CAT ATA TCC AAA CAA CAA AGC ATA ATA TTT CGC CCA CTA
                                                                                                                              15700
```

```
                                                                                    340
            S   Q   C   L   F   -   W   A   L   V   A   D   L   L   T   L   T   W   I   G   G   Q   P   V   E   H   P   Y   I   T   I   G
Cow         AGC CAA TGC CTA TTC --- TGA GCC CTA GTA GCA GAC CTA CTG ACA CTC ACA TGA ATT GGA GGA CAA CCA GTC GAA CAC CCA TAT ATC ACC ATC GGA
Sheep       AGT CAA TGT ATA TTC --- TGA ATC CTA GTA GCC GAC CTA TTA ACA CTC ACA TGA ATT GGA GGC CAG CCA GTT GAA CAC CCC TAC ATC ATT ATT GGA
Goat        AGC CAA TGC ATA TTC --- TGA ATC CTG GTA GCA GAT CTA TTA ACA CTC ACA TGA ATT GGA GGA CAG CCA GTC GAA CAT CCC TAC ATT ATT ATT GGA
Pronghorn   AGC CAA TGC TTA TTC --- TGA ATC CTA GTA GCA GAC CTA CTC ACC CTA ACA TGA ATC GGA GGC CAA CCA GTC GAA CAC CCA TTT ATT ATC ATT GGA
Giraffe     AGT CAA TGC CTA TTT --- TGA ATT CTA GTA GCG GAC CTG CTA ACA CTC ACA TGA ATT GGA GGA CAG CCA GTT GAA CAC CCA TTC ATC ATC ATC GGA
Fallow      AGC CAA TGC TTA TTT --- TGA ATC CTT GTA GCA GAC CTA CTA ACA CTA ACA TGA ATC GGA GGG CAA CCA GTT GAA CAC CCA TTT ATT ATC ATT GGA
Black-tail  AGC CAA TGC CTG TTT --- TGA ATT TTA GTA GCG GAC CTA CTA ACA CTC ACA TGA ATT GGA GGC CAA CCA GTT GAA CAC CCT TTC ATC ATC ATT GGA
Chevrotain  AGC CAA TGC TTA TTC --- TGA CTT CTA GCA GCC GAC CTA CTC ACA CTA ACA TGA ATT GGA GGC CAA CCT GTA GAA CAC CCC TAT GTC GTA ATT GGC
Camel       AGT CAA TGC CTG TTT --- TGA GTA CTA GTA GCT GAT CTA CTA ACC CTG ACA TGA ATT GGG GGA CAA CCA GTA GAA CCC CCA TTC ATC ATA ATT GGC
Dolphin 1a  AGC CAA CTC CTA TTT --- TGA GTA TTA ATC GCT GAC CTC CTA ACC TTA ACA TGA ATT GGA GGC CAA CCC GTA GAA CAC CCA TAT ATC ATT GTA GGC
Dolphin 1b  AGC CAA CTC CTA TTT --- TGA ACC TTA ATC GCT GAC CTC CTA ACC TTA ACA TGA ATT GGA GGC CAA CCC GTA GAA CAC CCA TAT ATC ATT GTA GGC
Dolphin 2   AGC CAA CTC CTA TTT --- TGA ACC CTA ATC GCT GAC CTC TTA ACC TTA ACA TGA ATT GGA GGC CAA CCC GTA GAA CAC CCA TAC ATT ATT GTA GGC
Peccary     AGC CAA CTC CTA TTT --- TGA ATA CTA GTG GCA GAC TTC CTC ACA CTG ACA TGA ATC GGA AGC CAG CCC GTC GAA CAT CCA TTC ATC ATC ATT GGA
Pig         AGT CAA TGC CTA TTC --- TGA ATA CTA GTA GCA GAC CTC ATT ACA CTA ACA TGA ATT GGA GGA CAA CCC GTA GAA CAC CCG TTC ATC ATC ATC GGC
Zebra       AGT CAA TGC GTG TTT --- TGA CTC TTA GTA GCA GAC TTA CTA ACA CTA ACA TGA ATC GGC GGC CAA CCA GTG GAA CAC CCA TAC ATA ATT ATC GGC
Rhino       AGC CAA TGT ATG TTC --- TGA CTA CTA GTA GCC GAC TTA CTC ACA CTT ACA TGA ATC GGA GGA CAA CCA GTC GAG CAC CCA TTC ATC ATT ATT GGC
Elephant    AGC CTA TGT GCC TAT TGC TGA ACT CTA ACA ATA GAT TTA CTA ACA CTT ACA TGA ATT GGC AGC CAA CCA GTA GAA TAT CCT TAC ATC ATT ATT GGT
Rat         ACC CAA ATC CTT TAC --- TGA ATC CTA GTA GCC AAC CTC CTA GTC TTA ACA TGA ATC GGA GGC CAA CCA GTA GAA CAC CCA TTT ATC ATT ATT GGT
Mouse       ACA CAA ATT TTG TAC --- TGA ATC CTA GTA GCC AAC CTA CTT ATC TTA ACC TGA ATT GGG GGC CAA CCA GTA GAA CAC CCA TTT ATT ATC ATT GGC
Human       AGC CAA TCA CTT TAT --- TGA CTC CTA GCC GCA GAC CTC CTC ATT CTA ACC TGA ATC GGA GGA CAA CCA GTA AGC TAC CCT TTT ACC ATC ATT GGA
```

```
                                        360                                                                 379
            Q   L   A   S   V   L   Y   F   L   L   I   L   V   L   M   P   T   A   G   T   I   E   N   K   L   L   K   W   *
Cow         CAA CTA GCA TCT GTC CTA TAC TTT CTC CTC ATC CTA GTG CTA ATA CCA ACG GCC GGC ACA ATC GAA AAC AAA TTA CTA AAA TGA AGA
Sheep       CAA CTA GCA TCT ATT ATA TAT TTC CTT ATC ATT CTA GTC ATA ATA CCA GTA GCT AGC ATC ATC GAA AAC AAC CTC CTA AAA TGA AGA
Goat        CAA CTA GCA TCT ATT ATA TAT TTC CTC ATC ATT CTA GTA ATA ATA CCA GTA GCT AGC ACC ATT GAA AAC AAC CTT CTA AAA TGA AGA
Pronghorn   CAA CTG GCA TCC ATC ATA TAT TTC CTC ATC ATC CTA GTA CTA ATA CCA GTT ACT AGC ACA ATC GAA AAC AAC CTA CTA AAA TGA AGA
Giraffe     CAA CTA GCA TCT ATT ATG TAC TTT CTC ATC ATC CTA GTG TTA ATA CCA GTC ACT AGT GCA ATC CAA AAC AAC CTT CTA AAA TGA AGA
Fallow      CAA CTA GCA TCT ATC TTA TAT TTT CTC ATT ATT CTA GTA CTA ATA CCA GCC ACC AGC ACA ATC CAG AAT AAC CTT CTA AAA TGA AGA
Black-tail  CAA CTA GCA TCA ATC CTA TAC TTT CTC ATT ATC CTA GTA CTT ATG CCA GTT ACT AGC ACA ATC GAA AAC AAC CTC TTA AAA TGA AGA
Chevrotain  CAG CTT GCA TCA ATT TTA TAT TTC TCT ATC ATT CTA GTT CTC ATG CCA GTA GCA GGA GTA ATT GAA AAC AAA ATA CTA AAA TGA AGA
Camel       CAA GTG GCA TCT ATC CTG TAC TTC TCC CTA ATC CTA ATT CTC ATG CCC GTA GCA GGA ATC ATT GAA AAC CGT ATC CTA AAA TGA AGA
Dolphin 1a  CAA TTA GCA TCG ATT CTA TAC TTC CTT CTA ATC CTA GTG CTA ATA CCA ACA GCC GGC CTC ATT GAA AAT AAA CTC CTA AAA TGA AGA
Dolphin 1b  CAA TTA GCA TCG ATT CTA TAC TTC CTT CTA ATC CTA GTG CTA ATA CCA ACA GCC GGC CTC ATT GAA AAT AAA CTC CTA AAA TGA AGA
Dolphin 2   CAA TTA GCA TCC ATC CTA TAC TTC CTC CTA ATC CTA GTG CTA ATA CCA ACA GCT GGC CTT ATT GAA AAT AAA CTT CTA AAA TGA AGA
Peccary     CAA CTG GCA TCC ATC CTA TAT TTT CTC ATT ATC CTT GTA CTC ATA CCA GTA GCT AAC ATT ATC GAA AAC AAC CTC CTA AAA TGA AGA
Pig         CAA CTA GCC TCC ATC TTA TAC TTC CTA ATC ATT CTA GTA TTG ATA CCA ATC ACT AGC ATC ATC GAA AAC AAC CTA TTA AAA TGA AGA
Zebra       CAA CTG GCT TCA ATC CTC TAC TTC TCC CTA ATT CTC ATC TTC ATA CCA CTC GCA AGC ACC ATT GAA AAC AAC CTT TTA AAA TGA AGA
Rhino       CAA CTA GCA TCA ATC CTA TAT TTC TCC CTA ATC CTA GTA CTT ATA CCC CTT GCA GGC ATT ATC GAA AAC AAC CTT CTG AAA TGA AGA
Elephant    CAA ATA GCC TCA ATT CTA TAT TTC TCC ATT ATC CTA GCC TTC CTG CCA ATT GCA GGA GTA ATC GAA AAC TAC CTC ATT AAG TAA
Rat         CAA CTA GCC TCC ATC AGT TAT TTT TCA ATT ATC CTC ATT CTC ATA CCA ATC TCT GGA ATT GTT GAA GAC AAA ATG TTA AAA TGA AAT TAA
Mouse       CAA CTA GCC TCC ATC TCA TAC TTC TCA ATC ATC TTA ATT CTT ATA CCA ATC TCA GGA ATT ATC GAA GAC AAA ATA CTA AAA TTA TAT CCA T
Human       CAA GTA GCA TCC GTA CTA TAC TTC ACA ACA ATC CTA ATC CTA ATA CCA ACT ATC TCC CTA ATT GAA AAC AAA ATA CTC AAA TGG GCC T
      15800
```

Fig. 2. Continued

of the sequences shown in Fig. 2 are from the mitochondrial genomes of these animals.

The complete DNA sequences of 17 ungulate mitochondrial cytochrome *b* genes, together with those from human, mouse, and rat, appear in Fig. 2. The three dolphin sequences are all very similar to a partial sequence for another dolphin, *Cephalorhynchus commersonii* (Southern et al. 1988), with the *Stenella* sequences being more similar to each other than to the *Cephalorhynchus* sequence. The sequences obtained from independent *S. longirostris* samples show, in contrast to the cow situation, that considerable intraspecific differences (17 nucleotide differences, 1.5%) can exist, though less than between different *Stenella* species (5.7% observed difference, on average).

The predicted translation products of all 17 of the ungulate and dolphin cytochrome *b* genes are 379 amino acids in length, beginning with a conserved initiating methionine codon and terminating with an AGA stop codon (TAA in the elephant). In the last respect these sequences differ from the translational termination signals produced by the polyadenylation of the processed mRNA found in the human and mouse sequences. The elephant sequence differs from the other mammalian sequences in possessing a 3-bp (1 amino acid) insertion between positions 15721 and 15722 (amino acid residues 325 and 326), an insertion not found in any other characterized cytochrome *b* sequence (Howell 1989). In addition the elephant has a 3-bp (1 amino acid) deletion relative to the other ungulate sequences at the C-terminal end, a region of length variation in other mammalian cytochrome *b* sequences (see Fig. 2).

Sequence Differences in Relation to Time

Transitions and Transversions

Sequence comparison of mitochondrial rRNA genes has shown that transversions accumulate linearly with time in the rRNA gene of artiodactyls (Miyamoto and Boyle 1989). To determine if similar relationships exist in the cytochrome *b* gene, transitions and transversions at third codon positions, representing chiefly silent sites, were plotted versus divergence times based on the fossil record (Fig. 3C and Table 2). Transversions accumulate nearly linearly with time at third positions, in a fashion similar to that between rRNA genes. In contrast to the divergence rate for transversions of 0.2% per million years (Myr) observed in rDNA genes, the divergence rate at third positions of codons is approximately 2.5 times greater (0.5% per Myr), a consequence of the fact that silent transversions can occur at many of the third positions of codons.

Estimates of transition to transversion ratios for silent positions in mammalian mtDNA range from 10:1 to 20:1 (or more) (Aquadro and Greenberg 1983; Brown 1985). Use of silent transversions at the third position of codons and a transition to transversion ratio of 10:1 for divergences up to 25 Myr (Table 2) yields an estimate of a silent divergence rate of approximately 10% per Myr, which is similar to the value reported for hominoids (Brown et al. 1982).

Replacement and Silent Substitutions

Similar analyses of first and second positions of codons (Fig. 3A and B) show that both transversions and total nucleotide differences accumulate nearly linearly with time. Using the most recent well-established divergence (sheep vs goat, see Table 2) we estimate the rate of replacement substitutions at first and second positions of codons to be about 0.4% per Myr. The majority of replacement substitutions occur at the first position of codons, as observed in other mitochondrial genes (Brown et al. 1982), at a rate approximately fourfold higher than at the second position of codons. The rate for replacement substitutions at the second position of codons estimated from Fig. 3B is similar to previous estimates for the cytochrome *b* gene of vertebrates (Saccone et al. 1987). Even though most third positions of codons are silent, a few amino acid replacements do occur, at a rate similar to those of second positions of codons. The increased replacement rate at first positions appears to be due to the fact that many more replacements at this position result in conservative amino acid changes than for replacements at second or third positions of codons.

Leucine codons pose a potential problem in an evolutionary analysis because they are encoded by six triplets, some of which differ by a first position silent change (TTR $\leftrightarrow$ CTR). Because the majority of cytochrome *b* leucine codons in mammals have purines in their third positions and the first position change is a transition, these changes will be very rapid (as many as 10% of leucine codons could change per Myr) and will obscure relationships. A typical cytochrome *b* gene has approximately 55 leucine codons, so that there could be 5.5 silent changes at first positions of codons per Myr. Because there is a large (~25-fold) difference in substitution rates at silent and replacement sites, the value of estimating an overall rate of change of the cytochrome *b* gene, or any other mitochondrial sequence, is questionable (see also Thomas and Beckenbach 1989). At silent sites, transitions saturate rapidly, within the first few million years, and subsequently the only observed changes are silent transversions and replacement substitutions. That is, after the first few million years the observed increase

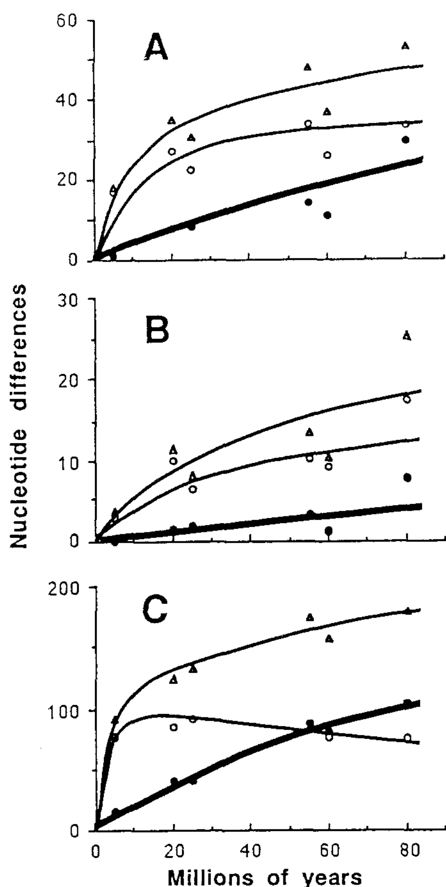

Fig. 3. Dependence of sequence difference on time of divergence for the three positions in codons. The number of nucleotide differences at first (A), second (B), and third (C) positions of codons due to transitions (○) and transversions (●) as well as the total differences (△) are plotted against time of divergence for the cytochrome *b* comparisons in Table 2. Each point shown is the average of all possible pairwise comparisons at the given divergence time.

in difference between two sequences is due to a tiny portion of the changing sites.

Base Composition

Animal mtDNA is noted for an extreme bias in base composition at silent sites, which manifests itself as strand bias, i.e., as a difference in composition between the two strands of mtDNA (Brown 1985). It is also known that different classes and phyla have different biases. Compositional bias could have two origins, selection or mutation pressure (Sueoka 1988). A consequence, independent of the origin of the bias, is that transition–transversion ratios will vary at different positions within codons due to differences in selection pressure. This consequence follows from the fact that third positions are more likely to be silent than first or second and from the likelihood that different nucleotide compositions will lead to different mutation spectra. If silent sites are to be used as phylogenetic markers, a requirement

Table 2. Approximate divergence dates, with ranges, from fossil record (Savage and Russell 1983)

Comparison	Divergence time (Myr) Probable	Range
Goat versus sheep	5	4–10
Cow versus goat/sheep	20	15–25
Cow versus pronghorn	20	15–25
Bovidae versus giraffe/deer	25	20–30
Ruminantia versus Camelidae	55	35–60
Ruminantia versus Suidae	60	50–65
Artiodactyls/perissodactyls versus rodents/primates	80	60–100

is that there should be little or no change in bias between the compared species. If a similar change in bias occurs in two different species, convergence at silent sites may result; alternatively, a similar bias may be the result of a unique event and may truly reflect phylogenetic information.

The base compositions at the three positions within cytochrome *b* codons are shown in Table 3. They differ greatly from one another, third positions having very little G, second positions being rich in T, and first positions having a rather unbiased composition. As expected, second positions of codons show the least interspecific variability (mean standard deviation = 0.72), with third positions showing the greatest (mean standard deviation = 2.74). First position substitutions tend to result in more conservative amino acid replacements than do nucleotide substitutions at second positions, and therefore more changes at the former sites are expected. Because third position changes tend to be silent, they should be the most variable.

The index of compositional difference described by Prager and Wilson (1988) is for mammalian cytochrome *b* smaller at second positions of codons than at first or third positions. Moreover, the mean values of this index, obtained for all possible pairs of the 20 sequences, were as small as those among mammalian lactalbumins (Prager and Wilson 1988). Indeed, at first and second positions of codons there does not appear to be any species with a differing base composition, as expected based on their similar amino acid compositions.

Because many differences at third codon positions reflect silent nucleotide substitutions, differences in composition should be most observable at this position. Such differences were found in the human, pronghorn, and zebra sequences (Table 3), which have unique base compositions (differing not only from all the remaining species but also from one another). The human and zebra sequences are unique in preferring C over A; the opposite bias (A over C) is observed in all other species considered. The pronghorn has an unusually low proportion of

Table 3. Base composition at first, second, and third positions of codons

Species	First				Second				Third			
	G	A	T	C	G	A	T	C	G	A	T	C
1 Cow	22.9	28.7	22.4	26.0	13.7	20.5	40.5	25.3	3.7	44.5	12.6	39.2
2 Sheep	22.1	31.3	21.1	25.5	13.7	20.3	42.1	23.9	3.2	42.9	18.7	35.3
3 Goat	21.6	31.6	20.8	26.1	13.7	20.3	41.8	24.2	3.2	44.2	15.8	36.8
4 Pronghorn	22.1	30.3	21.6	26.1	13.7	20.0	42.4	23.9	2.9	47.4	10.3	39.5
5 Giraffe	21.1	30.5	21.1	27.4	13.4	20.3	43.2	23.2	4.7	43.2	16.6	35.5
6 Fallow deer	21.8	30.0	22.9	25.3	13.7	19.7	42.9	23.7	3.7	46.3	17.1	32.9
7 Black-tail deer	22.1	29.2	21.8	26.8	13.7	20.0	42.1	24.2	3.7	43.4	19.7	33.2
8 Chevrotain	23.9	27.1	22.9	26.1	13.7	20.8	41.1	24.5	7.6	40.3	17.1	35.0
9 Camel	23.7	27.9	24.2	24.2	13.9	19.5	40.8	25.8	7.1	40.8	19.2	32.9
10 Dolphin 1a	21.6	28.7	23.2	26.6	13.9	20.0	41.1	25.0	2.4	43.4	17.4	36.8
11 Dolphin 1b	21.8	28.4	23.7	26.1	13.9	20.0	40.8	25.3	2.4	43.2	17.1	37.4
12 Dolphin 2	21.8	28.4	23.4	26.3	13.7	20.0	40.8	25.5	2.6	43.7	18.2	35.5
13 Pig	22.1	30.3	23.2	24.5	13.7	20.0	42.4	23.9	3.7	44.2	14.2	37.9
14 Peccary	21.1	30.0	20.3	28.7	13.2	20.3	42.4	24.2	3.9	44.5	12.9	38.7
15 Zebra	20.0	29.5	22.9	27.6	13.9	19.7	41.6	24.7	3.2	37.4	15.3	44.2
16 Rhinoceros	20.3	29.7	21.1	28.9	13.7	19.7	41.1	25.5	2.1	41.1	15.8	41.1
17 Elephant	19.5	29.5	23.4	27.6	13.7	22.1	39.2	25.0	3.4	40.8	22.1	33.7
18 Rat	22.0	28.1	24.7	25.2	13.1	20.7	42.0	24.1	2.6	42.0	15.0	40.5
19 Mouse	21.3	30.2	25.5	23.1	12.6	20.5	42.8	24.2	2.6	44.6	19.2	33.6
20 Human	19.5	29.5	23.4	27.6	12.9	20.0	40.0	27.1	3.7	36.3	12.1	47.9
Mean	21.6	29.4	22.7	26.3	13.6	20.2	41.6	24.7	3.6	42.7	16.3	37.4
Standard deviation	1.17	1.15	1.38	1.47	0.32	0.57	1.04	0.96	1.43	2.68	2.91	3.92
Bias[a]	0.076				0.221				0.401			

[a] Bias in base composition is calculated as

$$C = (2/3) \sum_{i=1}^{4} |c_i - 0.25|$$

where C is the compositional bias and c_i is the frequency of the ith base

Ts. Artiodactyls tend to have fewer pyrimidines in the light strand than do the nonartiodactyls, and therefore silent sites may tend to associate them to the exclusion of other species. A notable exception to this generalization is encountered with alanine codons, where GCA predominates in true ruminants, camel, and dolphins but GCC does so in the remaining species, i.e., pig, peccary, perissodactyls, rodents, elephant, and human.

Phylogenetic Analysis

Parsimony Trees

Within the 20 mammalian cytochrome *b* sequences there are 477 positions that have substitutions of any kind at first or second positions or a transversion difference at third positions; 361 of these are phylogenetically informative. Thus, about one-third of the 1140 bp of the cytochrome *b* gene may provide information on the phylogeny of these mammals. Figure 4 shows a tree built by the parsimony method, which represents the variability in the cytochrome *b* gene with the fewest events (1346). These events can be divided into the three categories, namely, 447 at first positions, 150 at second positions, and 749 at third positions.

Because the rate of nucleotide substitution at second positions is about one-fourth the rate of first position replacements or third position transversions (see above), it might be argued that they should be weighted more than first or third position changes. Accordingly, second positions were given a weight of 4, and first and third positions, which have similar substitution rates, were given weights of 1. The most parsimonious tree built using these weights is identical to the tree in Fig. 4.

To determine if other factors were influencing sequence similarities, the two types of silent sites were removed. Third positions were first removed, as differences in base composition occur among species (Table 3). The resulting parsimony tree was similar to that in Fig. 4, suggesting that the bias at third positions does not significantly influence the phylogeny in Fig. 4. In an attempt to determine the effect of silent first position changes, these were removed, but a similar tree again resulted.

Distance Trees

In another approach to determining relationships, the numbers of transversion differences were used to construct trees using the neighbor-joining method (Saitou and Nei 1987), shown in Fig. 5. Some differences in branching order are seen between the parsimony and distance trees, though if

only first and second position transversions are used as a distance measure (Fig. 5B), fewer differences are observed. This result implies that the silent third position transversions may be confusing distance phylogenies due to convergence in base composition. Most of the differences are with respect to a series of divergences separating perissodactyls, artiodactyls, and cetaceans, as well as early artiodactyl radiations. All these common ancestral lineages are very short and thus of doubtful statistical significance. Both the distance (Fig. 5) and parsimony (Fig. 4) methods indicate that these early radiations and divergences may have occurred within a short period, and the cytochrome *b* data cannot resolve them convincingly.

Approximate Molecular Time Scale

Because several of the divergences within artiodactyls have well-established divergence dates from the fossil record (Table 2) and some of the distance measures are approximately linear with time (see above, Fig. 3A and B), we can estimate additional dates of divergences roughly. Within the advanced ruminants, which are classified as pecorans (cow, sheep, goat, pronghorn, giraffe, and deer), it appears that all the species radiated from one another 20–25 Myr ago, except the following pairs: sheep–goat (see Table 2) and pronghorn–fallow deer (~10 Myr). Earlier sets of divergences appear to have occurred 45–50 Myr ago and to include the divergence of the chevrotain from pecorans, dolphins from camel, peccary from pig, and rhinoceros from zebra. These results indicate that the species chosen for the present survey represent three major radiations within mammals: (1) initial radiation of placental mammals roughly 80 Myr ago, (2) divergence of artiodactyls, perissodactyls, and cetaceans and their initial radiations 45–60 Myr ago, and (3) the radiation within Pecora 20–25 Myr ago. Thus the majority of internodes within the tree represent short time periods within these three radiations (see also Irwin and Wilson 1991).

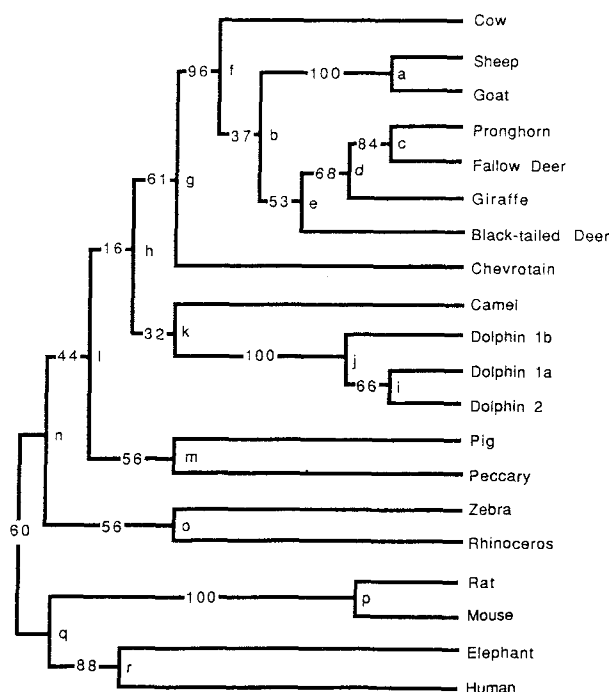

Fig. 4. Phylogenetic relationships of 20 mammalian cytochrome *b* genes. This parsimony tree is based on conservative changes, i.e., all changes at first and second positions as well as transversions at third positions. The number of mutational steps is 1346; the number of variable sites is 477, with 155, 81, and 241, respectively, at first, second, and third positions. The number of phylogenetically informative sites is 361 (115, 42, and 204, respectively, at first, second, and third positions). The consensus tree resulting from 100 bootstrap trials is identical with the parsimony tree. The value on each internal branch is the number of trials in which that branch was found. Dolphins 1a and 1b refer to the two *S. longirostris* samples, and dolphin 2 refers to *S. attenuata*. The rat, mouse, elephant, and human sequences were used to root the remainder of the tree, and the branching order of these species does not imply a rooting of the complete mammalian tree. Letters refer to each of the nodes, with the lengths of each branch following: cow—f, 34; sheep—a, 21; goat—a, 13; pronghorn—c, 22; fallow deer—c, 13; giraffe—d, 27; black-tailed deer—e, 29; chevrotain—g, 59; camel—k, 84; dolphin 1b—j, 2; dolphin 1a—i, o; dolphin 2—i, 15; pig—m, 61; peccary—m, 53; zebra—o, 54; rhinoceros—o, 52; rat—p, 55; mouse—p, 53; elephant—r, 115; human—r, 123; ab, 21; cd, 15; de, 11; eb, 12; bf, 17; fg, 31; gh, 18; ij, 8; jk, 78; kh, 26; hl, 17; ml, 20; ln, 25; on, 21; nq, 36; pq, 52; rq, 53.

Bootstrap Analysis

The consensus tree constructed after 100 bootstrap replications (Fig. 4) is identical to the shortest parsimony tree, and Table 4 summarizes the relationships that are supported strongly (>95%). Three clades were identified in 100% of the bootstrap samples: those including the sheep and goat, the rat and mouse, and the three dolphin sequences. The associations of rat with mouse and of the three dolphins with each other are supported by long common lineages (pq and jk in Fig. 4). These results, together with the sheep–goat case (ab), suggest that stem lineages that are at least 15 Myr long (see asterisks in Table 4) can be identified with near certainty. One other clade was supported in a high (96%) proportion of the bootstrap samples, that including all pecorans. There is uncertainty in the date of divergence of the chevrotain based on the fossil record (Savage and Russell 1983) but, based on distance data (see above) and Fig. 3, we suggest that the divergence was approximately 45 Myr ago. This would imply that the common lineage for all pecorans (fg in Fig. 4) is approximately 20 Myr long. By contrast, as argued above, the majority of the internal lineages in Fig. 4 are probably shorter than 15 Myr.

Because the characters that support one clade

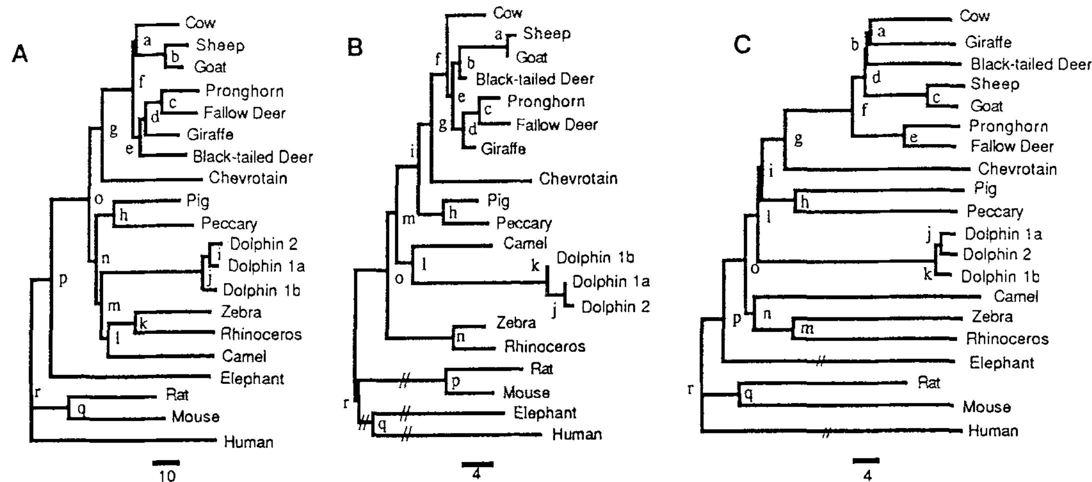

Fig. 5. Three neighbor-joining trees for 20 mammalian cytochrome *b* genes. These trees were built as described by Saitou and Nei (1987) using the number of transversion differences observed at (**A**) all positions, (**B**) first and second positions of codons, and (**C**) third positions of codons. As in Fig. 4, the rat, mouse, elephant, and human sequences were used to root the remaining ungulate sequences, and as such these trees do not imply a root for mammals. The small bars with each tree indicate distance in number of transversions. Branch lengths (number of transversions) for each of the trees are as follows: (A) cow—a, 21.3; sheep—b, 9.5; goat—b, 6.5; pronghorn—c, 13.4; fallow deer—c, 12.6; giraffe—d, 14.7; black-tailed deer—e, 22.0; chevrotain—g, 47.1; pig—h, 34.9; peccary—h, 39.1; dolphin 2—i, 5.1; dolphin 1a—i, 2.9; dolphin 1b—j, 4.8; zebra—k, 35.4; rhinoceros—k, 36.6; camel—1, 50.4; elephant—p, 79.2; mouse—q, 46.1; rat—q, 38.9; human—r, 89.2; ab, 16.7; cd, 10.3; de, 3.0; af, 0.3; ef, 3.0; fg, 15.7; go, 5.3; hn, 8.5; ij, 4.3; jm, 50.3; kl, 14.6; lm, 2.7; mn, 2.8; no, 2.5; op, 13.6; pr, 5.6; qr, 19.8. (B) cow—f, 3.1; sheep—a, 1.0; goat—a, 0.0; black-tailed deer—b, 0.9; pronghorn—c, 1.9; fallow deer—c, 3.1; giraffe—d, 1.3; chevrotain—g, 9.6; pig—h, 3.7; peccary—h, 4.3; camel—l, 6.8; dolphin 1b—k, 0.1; dolphin 1a—j, 0.0; dolphin 2—j, 1.0; zebra—n, 2.4; rhinoceros—n, 3.6; mouse—p, 4.9; rat—p, 8.1; elephant—q, 18.2; human—q, 24.8; ab, 5.0; be, 0.4; cd, 1.7; de, 0.9; ef, 0.7; fg, 1.4; gi, 1.0; hi, 1.6; im, 1.9; jk, 1.9; kl, 12.3; lm, 1.2; mo, 0.6; no, 5.7; or, 3.2; pr, 12.8; qr, 8.0. (C) cow—a, 17.2; giraffe—a, 15.8; black-tailed deer—b, 19.5; sheep—c, 8.7; goat—c, 6.3; pronghorn—e, 11.1; fallow deer—e, 9.9; chevrotain—g, 37.0; pig—h, 34.2; peccary—h, 31.8; dolphin 1a—j, 2.9; dolphin 2—j, 4.1; dolphin 1b—k, 4.8; camel—n, 44.0; zebra—m, 33.4; rhinoceros—m, 32.6; elephant—p, 55.8; mouse—q, 40.6; rat—q, 29.4; human—r, 57.6; ab, 1.5; bd, 0.6; cd, 11.9; df, 3.1; ef, 9.7; fg, 12.4; gi, 4.5; hi, 6.8; il, 0.1; jk, 2.2; kl, 37.4; lo, 3.4; mn, 7.0; no, 1.8; op, 8.0; pr, 4.7; qr, 7.4.

Table 4. Phylogenetic inferences based on sequences of the cytochrome *b* gene

Phylogenetic inference	Internal branch[a]	Bootstrap value
Monophyletic group		
Sheep and goat	ab*	100
Dolphins	jk*	100
Rat and mouse	pq*	100
Pecorans	fg	96
Significant internal branch[b]		
Pig and peccary versus elephant and human	mlnqr	99
Chevrotain and pecorans versus elephant and human	ghlnqr	99
Pronghorn and fallow deer versus cow, sheep, and goat	cdeb	98
Zebra and rhinoceros versus elephant and human	onqr	98
Camel and dolphins versus elephant and human	khlnqr	98
Dolphins and artiodactyls versus elephant and human	hlnqr	98
Artiodactyls, perissodactyls, and dolphins versus elephant and human	nqr	95
Artiodactyls and perissodactyls versus elephant and human	nqr	95

[a] Identified with respect to the branching points labeled in Fig. 4. Each internal branch unites the groups of species mentioned, to the exclusion of others. Asterisks (*) mark common ancestral lineages that are at least 15 Myr long (cf. Table 2 and Fig. 4)

[b] Significant internal branches are determined with the equation

$$B = 100\left[1 - \prod_{i=1}^{n} (1 - b_i)\right],$$

where n is the number of contiguous internal branches to be combined, b_i is the bootstrap value observed for a particular internal branch (Fig. 4), and Π signifies the product. This equation should be regarded as approximate only (Irwin and Wilson 1990)

should be independent of characters that support a higher or lower level in the phylogeny, it should be possible to obtain estimates of support for combined internal lineages. As shown in Table 4, eight additional examples of significant phylogenetic resolutions were achieved by combining bootstrap values. For example, significant associations of placental mammals were obtained between two or three orders to the exclusion of two other orders of placental mammals. Specifically, the artiodactyls, perissodactyls, and cetaceans separate cleanly as a group from the primates (human) and proboscidians (elephant). An implication is that the elephants are only distantly related to the other ungulates tested.

Comparison to Morphological and Other Molecular Data

The phylogenetic results highlighted in Table 4 are generally consistent with other morphological and molecular data (Mross and Doolittle 1967; de Jong 1985; Sarich 1985, and personal communications; Shoshani 1986; Beintema et al. 1988; Miyamoto and Boyle 1989; Novacek 1989; Czelusniak et al. 1990). The most significant difference is in the placement of the elephant. Recent morphological analysis suggested that elephants are related more closely to perissodactyls than are artiodactyls (Janis 1988; Novacek 1989). In contrast, our cytochrome *b* results implying that the elephant is only distantly related to the other ungulates (artiodactyls, perissodactyls) and cetaceans fit well with other molecular data (de Jong 1985; Shoshani 1986; Czelusniak et al. 1990) and with the classical morphological view (Simpson 1945). For a more comprehensive treatment of this point, see Irwin and Wilson (1991).

Comparisons of proteins encoded by nuclear genes (Sarich 1985, and personal communications; Shoshani 1986) also support our observation that cetaceans are closely related to artiodactyls, as do some morphological data (Novacek 1989), though other morphological traits suggest that cetaceans are distant from artiodactyls (Shoshani 1986). Although, as a general rule, the cytochrome *b* gene sequences do not resolve the radiations that occurred 45–65 Myr ago and 20–25 Myr ago, the pronghorn result is notable. Morphological data as traditionally interpreted (Janis 1988) and fibrinopeptide sequences (Mross and Doolittle 1967) suggested that the pronghorn is a bovid. By contrast, ribonuclease sequences group it weakly with the giraffe (Beintema et al. 1988), and a new view of the morphological evidence groups it with deer (Janis 1988). Our results, too, ally the pronghorn to certain deer.

A finding of methodological interest concerns the three dolphin species subjected to mtDNA sequencing. The tree in Fig. 4 implies that *S. longirostris* is a paraphyletic species, even though the two *S. longirostris* samples (dolphins 1a and 1b) are more similar to each other (1.5% different) than to *S. attenuata* (dolphin 2) (5.7% observed difference, on average). Phylogenetic analysis of the three *Stenella* sequences together with the partial *Cephalorhynchus* dolphin sequence (Southern et al. 1988), including third position transitions, reveals that the *S. longirostris* sequences are phylogenetically more related to each other (results not shown) than to a sequence from another species. Thus, exclusion of transitions at third positions can lead to misleading inferences about very closely related sequences.

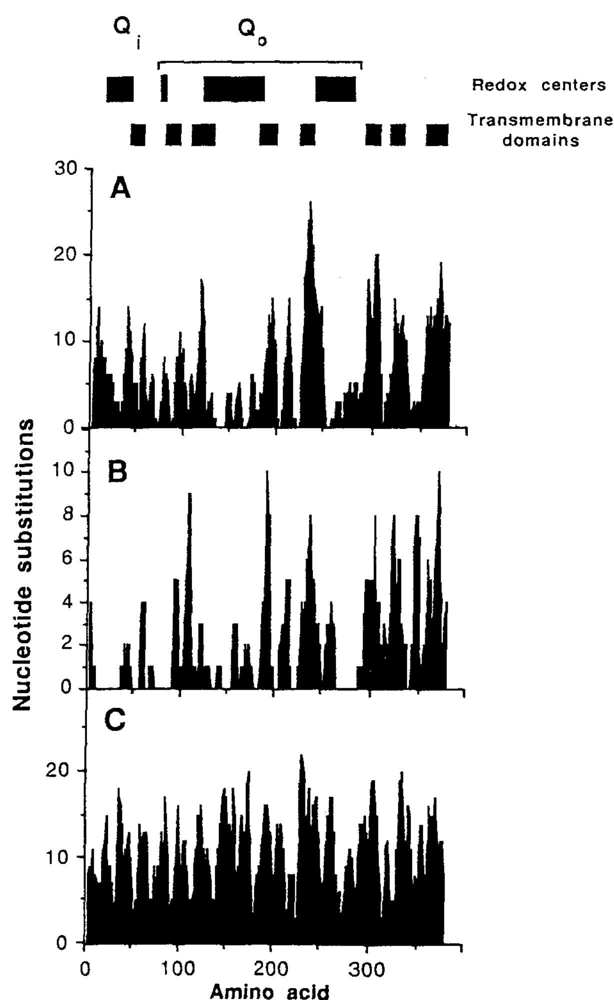

Fig. 6. Variation across the cytochrome *b* gene. Variability in the cytochrome *b* gene is illustrated by showing the sum of phylogenetically inferred nucleotide substitutions, from Fig. 4, in a moving window of five amino acid positions through the 20 cytochrome *b* sequences. (**A**) First positions, (**B**) second positions, and (**C**) third positions of codons. Only transversions are depicted at third positions. The locations of redox centers and transmembrane domains, based on Howell (1989), are indicated at the top of the figure.

Possible Limitations on Phylogenetic Resolving Power

Despite the abovementioned evidence of phylogenetic utility, the cytochrome *b* gene sequences did

Table 5. Observed and expected frequencies of nucleotide substitutions at each codon position throughout the cytochrome *b* molecule

	First positions					
	All			Replacement		
Assigned events	Observed	Poisson	Negative binomial	Observed	Poisson	Negative binomial
0	226	118	220	224	152	217
1	46	138	66	37	113	43
2	29	81	34	19	42	21
3	24	32	21	14		12
4	28		13	15		8.0
5	15		8.5	6	12.8	5.4
6	7	12.1		3		
7	6		18.8	2		13.5
8+	0			0		
k[a]			0.4			0.25
χ^2		354.8	31.5		155.6	13.1*
df[b]		3	5		2	5

At first and second positions of codons all nucleotide interconversions were counted, whereas at third positions only transversions were counted. First positions in the replacement category exclude those encoding leucine in the cow sequence. Third positions in the silent category are those with silent transversions relative to the cow sequence, including all leucine codons. The assigned events are the inferred number of nucleotide changes based on the tree in Fig. 4. Brackets in the columns listing expectations based on the Poisson and negative binomial models mark those classes that were pooled for the chi-square (χ^2) test so as to avoid small expected values. Significance levels in the chi-square test are denoted as follows: *, $P > 0.025$; **, $P > 0.05$; P is the probability that the observed distribution fits the expected distribution under one of the two models examined

[a] From the negative binomial distribution, where $k = m^2/(s^2 - m)$, with m being the mean number of inferred events per position and s^2 the variance. See Larson and Wilson (1989) for further discussion

[b] Degrees of freedom for the chi-square test

not affirm the monophyly of such well-recognized groups as the Bovidae (which includes the cow, sheep, and goat) and the Cervidae (which includes the two deer). (Nor did these sequences contradict the monophyly of these two families.) Although this failure could reflect the likely brevity of the periods of common ancestry for bovids and cervids,[2] it might instead imply that this mitochondrial gene has limited resolving power. If so, the basis for this apparently limited resolving power requires further investigation. We suspect that it may reside in the biased base composition displayed by the most conservative positions of this fast-evolving gene. Despite the low interspecific variability in base composition (see above), the overall bias at second positions (Table 3) is rather high. This bias is 0.221 on a scale where 0 means no bias and 1.0 refers to a completely biased sequence (such as one having only a single kind of base). In general, all genes for mitochondrial proteins have a compositional bias at the second positions of codons (Gadaleta et al. 1989), with only the three cytochrome oxidase genes (COI, II, and III) exhibiting less bias than the cytochrome *b* gene (calculations not shown). At biased positions, the pattern of evolution is less random than at unbiased positions, which means that parallelisms are more likely and that the assumptions of the parsimony method about the process of molecular evolution are in danger of being violated (Sidow and Wilson 1990).

[2] The family Bovidae is traditionally divided into subfamilies, one containing the cow and another containing sheep and goats. The period of common bovid ancestry shared by these two subfamilies could be less than 5 Myr (Savage and Russell 1983). Molecular phylogenies based on stomach lysozyme cDNA sequences reinforce this possibility (Irwin and Wilson 1990). Likewise, the two deer studied here belong to distinct subfamilies (Cervinae and Odocoileinae) and could share only a brief period of common ancestry before their common ancestral lineage intersects with noncervid lineages (Carroll 1988); indeed, according to Janis (1988), the cervids may not be monophyletic

Variability Along the Cytochrome b Gene

Figure 6 shows the strikingly nonrandom distribution of positions at which nucleotides change in the cytochrome *b* gene—a phenomenon observed for other parts of the mtDNA genome as well (Thomas and Beckenbach 1989). The inferred number of nucleotide substitutions (based on the tree in Fig. 4) is given for first, second, and third (transversions only) positions in a moving window that is five codons long. This approach emphasizes small segments of the sequence rather than single amino acid residues. When a window of 10 codons was used, it tended to mask variable amino acids within conserved regions and conserved residues in variable regions. By contrast, a window of five residues tends to define both conserved and variable sequences, without too much sensitivity to single positions.

Table 5. Continued

Second positions			Third positions: All			Third positions: Silent	
Observed	Poisson	Negative binomial	Observed	Poisson	Negative binomial	Observed	Poisson
300	257	300	140	53	108	7	8
45	101	46	46	105	92	24	26
15	20	18	49	103	65	39	43
12		8.2	48	68	43	47	47
6			49	33	28	46	38
3	2.9	8.8	33	13	17	31	25
0			13		11	13	13
			2	5.8	16.1	2	10.0
			1			1	
		0.25			1.5		
	152.5	2.29**		266.6	78.8	8.7**	
	2	3		5	6	6	

At first positions, prominent peaks of variation are associated with each of the eight proposed transmembrane domains (Howell 1989; see Fig. 6). This is due to both silent first position transitions in leucines together with frequent transversions that exchange isoleucine, valine, or methionine for leucine. There are regions of few first position changes near amino acids 140 and 260 associated with the Q_o redox center. At second positions there are fewer nucleotide changes, and several areas exhibit no variation. Two of the latter regions include the Q_o redox center; another highly conserved region is near position 30, which includes the Q_i redox center. Third codon positions show high variability throughout the molecule.

Variance in Substitutions among Positions

From the tree in Fig. 4, we can count the number of changes assigned to each position in the sequence over evolutionary history. The number of sites experiencing a particular number of changes is given in Table 5 for each codon position and for amino acid substitutions. If the substitutions had occurred at random, a Poisson distribution would be observed. However, as has been frequently noted, evolutionary substitutions rarely approximate a Poisson distribution (Uzzell and Corbin 1971; Holmquist et al. 1983). The probability of substitution varies among sites, so that a negative binomial distribution is often a better description of the data. The fit of second positions to a negative binomial is excellent (Table 5; $P > 0.50$). The fit for the first and third positions is quite poor and indicates a bimodal distribution of substitutional probabilities, i.e., there is a class of sites that is experiencing many more substitutions than expected.

At the first positions of codons, there can be silent changes in leucine codons; these would be expected to occur faster than replacement changes and may distort the distribution. To test this idea, we removed all amino acid positions that encode leucine in one sequence (e.g., cow in Table 5) and thereby removed most of these hypervariable sites. The resulting distribution of nucleotide substitutions then almost fit a negative binomial (Table 5; $P > 0.025$). [Similar results were found after removing amino acid positions with leucine in the human sequence ($\chi^2 = 12.5$, $P > 0.025$).]

Similarly, at third positions of codons, because only transversions were included, two classes of positions are expected: those where transversions are silent and those where they result in amino acid substitutions. To explore the possibility that replacement transversions are altering the random assortment of mutations, we analyzed only those sites at which silent transversions could occur (i.e., all fourfold degenerate sites and all leucine codons). In the cow there are 210 such sites, and these appear to have a Poisson distribution (Table 5; $P > 0.15$),

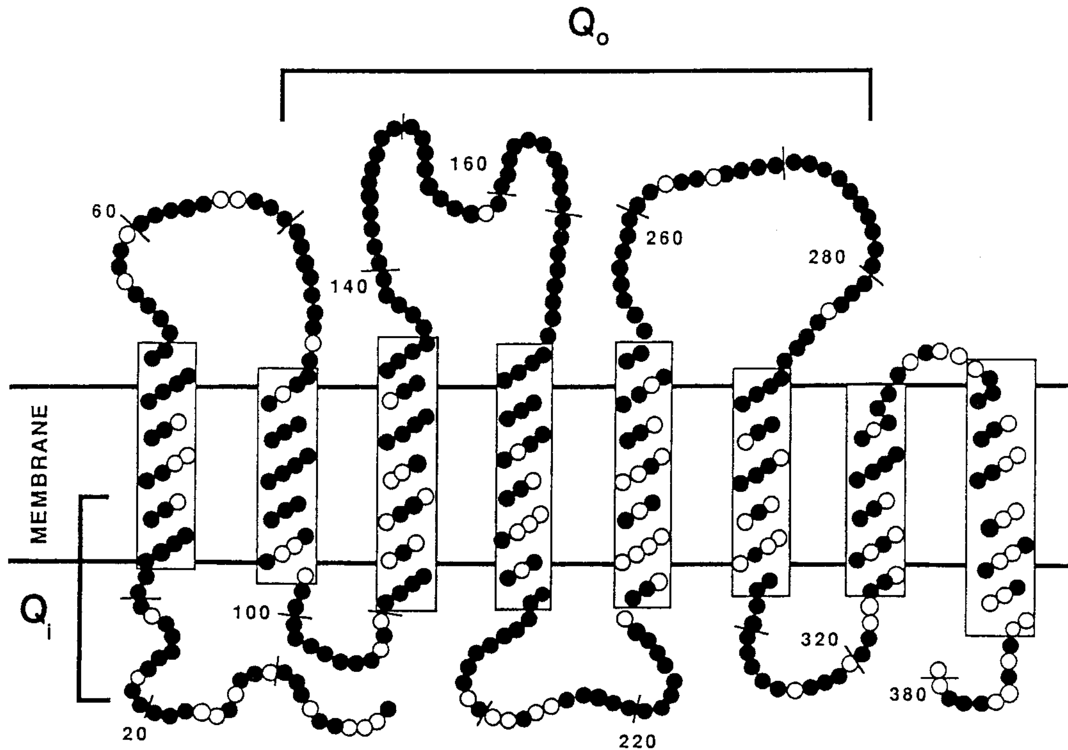

Fig. 7. Amino acid variability in mammalian cytochromes *b*. In this structural model of cytochrome *b* (adapted from Howell 1989), conservative residues are solid circles, whereas hypervariable residues are empty circles. The latter are defined as residues that exist in three or more amino acid states (see Fig. 2) or that are phylogenetically inferred to have been replaced at least twice on the lineages leading to the 20 cytochromes *b* in Fig. 4. Q_i and Q_o are the redox centers (cf. also Fig. 6). The outer and inner surfaces of the membrane correspond, respectively, to the top and bottom sections of the figure. [The amino acid inserted in the elephant sequence between residues 325 and 326 (Fig. 2) is not included in this figure.]

Table 6. Distribution of slowly and rapidly evolving amino acid positions with respect to the structural model for cytochrome *b*

Domain	Number of sites		
	Slow	Fast	Sum
Transmembrane	116	56	172
Outer surface	99	12	111
Inner surface	71	27	98
Total	286	95	381

The data tabulated came from Fig. 7

implying that there is a random distribution of transversions at these sites; using the human sequence, similar results were found ($\chi^2 = 4.48$, $P >$ 0.50).

Evolution of Cytochrome b

Previous comparisons of cytochrome *b* used amino acid sequences that diverged relatively early in the history of life (Howell 1989). Here we compare sequences of less than 5 to about 80 Myr divergence and can observe fine-tuning of the cytochrome *b* protein. Figure 6 has already shown how sequence variability is distributed along the length of the gene. Figure 7 now shows where in the structural model for cytochrome *b* (Howell 1989) the invariant and variable amino acid residues occur. Based on the tree in Fig. 4, it can be inferred that 225 (59.1%) residues are invariant and 61 (16.0%) change only once, whereas the remaining 95 (24.9%) change at least twice. Most of the variable positions are located within the transmembrane segments, or at the amino and carboxy ends of the protein (see Table 6 and Fig. 7). The outer surface of the protein appears to be evolving more slowly than either the transmembrane region ($\chi^2 = 17.5$, $P < 0.0001$) or the inner surface ($\chi^2 = 9.61$, $P < 0.005$). Because there is no difference in the rate of change of the transmembrane or inner surface ($\chi^2 = 0.76$, $P <$ 0.3), it is the outer surface that is evolving slowly, rather than the transmembrane region evolving fast.

Most of the variable changes in the transmembrane portion are changes between hydrophobic (leucine, isoleucine, and valine) residues. Much of the inner surface is composed of the amino and carboxy terminals of the protein, and may have little functional importance (see Fig. 7). Regions that had been defined as Q_o and Q_i by mutational studies (Howell and Gilbert 1988) or by distant evolutionary comparisons (Howell 1989) are very conservative within mammals (Fig. 7). The Q_i redox center involves a short portion of the first transmembrane

segment (see Fig. 7) and therefore contributes only a small fraction of the transmembrane region. On the other hand, most of the outer surface is implicated in the Q_o redox center (Howell 1989; see Fig. 7), and this appears to be a major contributor to the reduced evolutionary rate for the outer surface (Table 6).

Acknowledgments. We thank P. Arctander, S.V. Edwards, J.R. Kornegay, A. Meyer, S. Pääbo, E.M. Prager, C. Saccone, V.M. Sarich, A. Sidow, W.K. Thomas, and L. Vigilant for helpful advice, suggestions, and discussions. This work received support from National Institutes of Health and National Science Foundation grants to A.C.W. and postdoctoral fellowships from the Medical Research Council of Canada to D.M.I. and the National Institutes of Health to T.D.K. The nucleotide sequences reported in this paper have been submitted to the GenBank® with accession numbers X56281–X56296.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG (1982) Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J Mol Biol 156:683–717

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103:287–312

Beintema JJ, Schüller C, Irie M, Carsana A (1988) Molecular evolution of the ribonuclease superfamily. Prog Biophys Mol Biol 51:165–192

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180

Brown WM (1985) The mitochondrial genome of animals. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 95–130

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol 18:225–239

Carroll RL (1988) Vertebrate paleontology and evolution. WH Freeman, New York

Carlson SS, Wilson AC, Maxson RD (1978) Do albumin clocks run on time? Science 200:1183–1185

Czelusniak J, Goodman M, Moncrief ND, Kehoe SM (1990) Maximum parsimony approach to construction of evolutionary trees from aligned homologous sequences. Methods Enzymol 183:601–615

de Jong WW (1985) Superordinal affinities of rodentia studied by sequence analysis of eye lens proteins. In: Luckett WP, Hartenberger J-L (eds) Evolutionary relationships among rodents. A multidisciplinary analysis. Plenum, New York, pp 211–226

di Rago J-P, Netter P, Slonimski PP (1990) Pseudo-wild type revertants from inactive apocytochrome *b* mutants as a tool for the analysis of the structure/function relationships of the mitochondrial ubiquinol–cytochrome *c* reductase of *Saccharomyces cerevisiae.* J Biol Chem 265:3332–3339

Easteal S (1990) The pattern of mammalian evolution and the relative rate of molecular evolution. Genetics 124:165–173

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fukuda M, Wakasugi S, Tsuzuki T, Nomiyama H, Shimada K, Miyata T (1985) Mitochondrial DNA-like sequences in the human nuclear genome. Characterization and implications in the evolution of mitochondrial DNA. J Mol Biol 186:257–266

Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sbisà E, Saccone C (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J Mol Evol 28:497–516

Gyllensten UB, Erlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the *HLA-DQA* locus. Proc Natl Acad Sci USA 85:7652–7656

Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol & Evol 4:6–11

Hatefi Y (1985) The mitochondrial electron transport and oxidative phosphorylation system. Annu Rev Biochem 54:1015–1069

Hayasaka K, Gojobori T, Horai S (1988) Molecular phylogeny and evolution of primate mitochondrial DNA. Mol Biol Evol 5:626–644

Holmquist R, Goodman M, Conroy T, Czelusniak J (1983) The spatial distribution of fixed mutations within genes coding for proteins. J Mol Evol 19:437–448

Howell N (1989) Evolutionary conservation of protein regions in the proton-motive cytochrome *b* and their possible roles in redox catalysis. J Mol Evol 29:157–169

Howell N, Gilbert K (1988) Mutational analysis of the mouse mitochondrial cytochrome *b* gene. J Mol Biol 203:607–618

Irwin DM, Wilson AC (1989) Multiple cDNA sequences and the evolution of bovine stomach lysozyme. J Biol Chem 264: 11387–11393

Irwin DM, Wilson AC (1990) Concerted evolution of ruminant stomach lysozymes. Characterization of lysozyme cDNA clones from sheep and deer. J Biol Chem 265:4944–4952

Irwin DM, Wilson AC (1991) Limitations of molecular methods for establishing the phylogeny of mammals, with special reference to the position of elephants. In: Szalay FS, Novacek MJ, McKenna MC (eds) American Museum of Natural History symposium on mammalian phylogeny. Princeton University Press, Princeton NJ (in press)

Irwin DM, Sidow A, White RT, Wilson AC (1989) Multiple genes for ruminant lysozymes. In: Smith-Gill SJ, Sercarz EE (eds) The immune response to structurally defined proteins: the lysozyme model. Adenine Press, Schenectady NY, pp 73–85

Janis CM (1988) New ideas in ungulate phylogeny and evolution. Trends Ecol & Evol 3:291–297

Jollès J, Jollès P, Bowman BH, Prager EM, Stewart C-B, Wilson AC (1989) Episodic evolution in the stomach lysozymes of ruminants. J Mol Evol 28:528–535

Jollès J, Prager EM, Alnemri ES, Jollès P, Ibrahimi IM, Wilson AC (1990) Amino acid sequences of stomach and nonstomach lysozymes of ruminants. J Mol Evol 30:370–382

Kamimura N, Ishii S, Liandong M, Shay JW (1989) Three separate mitochondrial DNA sequences are contiguous in human genomic DNA. J Mol Biol 210:703–707

Keohavong P, Thilly WG (1989) Fidelity of DNA polymerases in DNA amplification. Proc Natl Acad Sci USA 86:9253–9257

Kocher TD, White TJ (1989) Evolutionary analysis via PCR. In: Erlich HA (ed) PCR technology. Principles and applications for DNA amplification. Stockton Press, New York, pp 137–147

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in mammals: amplification and sequencing with conserved primers. Proc Natl Acad Sci USA 86:6196–6200

Koop BF, Goodman M (1988) Evolutionary and developmental aspects of two hemoglobin β-chain genes (ϵ^M and β^M) of opossum. Proc Natl Acad Sci USA 85:3893–3897

Larson A, Wilson AC (1989) Patterns of ribosomal RNA evolution in salamanders. Mol Biol Evol 6:131–154

Li W-H, Luo C-C, Wu C-I (1985) Evolution of DNA sequences. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 1–94

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor NY

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101:20–78

Meyer A, Wilson AC (1990) Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J Mol Evol 31:359–364

Miyamoto MM, Boyle SM (1989) The potential importance of mitochondrial DNA sequence data to eutherian mammal phylogeny. In: Fernholm B, Bremer K, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 437–450

Mross GA, Doolittle RF (1967) Amino acid sequence studies on artiodactyl fibrinopeptides. II. Vicuna, elk, muntjak, pronghorn antelope, and water buffalo. Arch Biochem Biophys 122:674–684

Nomiyama H, Fukuda M, Wakasugi S, Tsuzuki T, Shimada K (1985) Molecular structures of mitochondrial-DNA-like sequences in human nuclear DNA. Nucleic Acids Res 13:1649–1658

Novacek MJ (1989) Higher mammal phylogeny: the morphological–molecular synthesis. In: Fernholm B, Bremer K, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 421–435

Pääbo S, Wilson AC (1988) Polymerase chain reaction reveals cloning artefacts. Nature 334:387-388

Prager EM, Wilson AC (1988) Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. J Mol Evol 27:326–335

Saccone C, Attimonelli M, Lanave C, Gallerani R, Pesole G (1987) The evolution of mitochondrially coded cytochrome genes: a quantitative estimate. In: Papa S, Chance B, Ernster L (eds) Cytochrome systems. Plenum, New York, pp 103–109

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sarich VM (1985) Rodent macromolecular systematics. In: Luckett WP, Hartenberger J-L (eds) Evolutionary relationships among rodents. A multidisciplinary analysis. Plenum, New York, pp 423–452

Savage DE, Russell DE (1983) Mammalian paleofaunas of the world. Addison-Wesley, Reading MA

Shoshani J (1986) Mammalian phylogeny: comparison of morphological and molecular results. Mol Biol Evol 3:222–242

Sidow A, Wilson AC (1990) Compositional statistics: an improvement of evolutionary parsimony and its application to deep branches in the tree of life. J Mol Evol 31:51–68

Simon C, Pääbo S, Kocher TD, Wilson AC (1990) Evolution of mitochondrial ribosomal RNA in insects as shown by the polymerase chain reaction. In: Clegg M, Clark S (eds) Molecular evolution. UCLA symposia on molecular and cellular biology, new series, vol 122. Alan R Liss, New York, pp 235–244

Simpson GG (1945) The principles of classification and a classification of mammals. Bull Am Mus Nat Hist 85:1–350

Southern SO, Southern PJ, Dizon AE (1988) Molecular characterization of a cloned dolphin mitochondrial genome. J Mol Evol 28:32–42

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA 85:2653–2657

Swofford DL (1989) PAUP: phylogenetic analysis using parsimony, version 3.0b. Illinois Natural History Survey, Champaign IL

Thomas WK, Beckenbach AT (1989) Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitution. J Mol Evol 29:233–245

Thomas WK, Maa J, Wilson AC (1989) Shifting constraints on tRNA genes during mitochondrial DNA evolution in animals. New Biologist 1:93–100

Tindall KR, Kunkel TA (1988) Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry 27: 6008–6013

Uzzell T, Corbin KW (1971) Fitting discrete probability distributions to evolutionary events. Science 172:1089–1096

Watanabe T, Hayashi Y, Semba R, Ogasawara N (1985) Bovine mitochondrial DNA polymorphism in restriction endonuclease cleavage patterns and the location of the polymorphic sites. Biochem Genet 23:947–957

Watanabe T, Masangkay JS, Wakana S, Saitou N, Tomita T (1989) Mitochondrial DNA polymorphism in native Philippine cattle based on restriction endonuclease cleavage patterns. Biochem Genet 27:431–438

White TJ, Arnheim N, Erlich HA (1989) The polymerase chain reaction. Trends Genet 5:185–189

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn Soc 26:375–400

Wilson AC, Zimmer EA, Prager EM, Kocher TD (1989) Restriction mapping in the molecular systematics of mammals: a retrospective salute. In: Fernholm B, Bremer K, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 407–419

Received May 9, 1990/Revised and accepted August 25, 1990