J Mol Evol (1982) 18:225–239

Journal of
Molecular Evolution
© Springer-Verlag 1982

Mitochondrial DNA Sequences of Primates: Tempo and Mode of Evolution

Wesley M. Brown[1], Ellen M. Prager, Alice Wang[2], and Allan C. Wilson

Department of Biochemistry, University of California, Berkeley, California 94720, USA

Summary. We cloned and sequenced a segment of mitochondrial DNA from human, chimpanzee, gorilla, orangutan, and gibbon. This segment is 896 bp in length, contains the genes for three transfer RNAs and parts of two proteins, and is homologous in all 5 primates. The 5 sequences differ from one another by base substitutions at 283 positions and by a deletion of one base pair. The sequence differences range from 9 to 19% among species, in agreement with estimates from cleavage map comparisons, thus confirming that the rate of mtDNA evolution in primates is 5 to 10 times higher than in nuclear DNA. The most striking new finding to emerge from these comparisons is that transitions greatly outnumber transversions. Ninety-two percent of the differences among the most closely related species (human, chimpanzee, and gorilla) are transitions. For pairs of species with longer divergence times, the observed percentage of transitions falls until, in the case of comparisons between primates and non-primates, it reaches a value of 45. The time dependence is probably due to obliteration of the record of transitions by multiple substitutions at the same nucleotide site. This finding illustrates the importance of choosing closely related species for analysis of the evolutionary process. The remarkable bias toward transitions in mtDNA evolution necessitates the revision of equations that correct for multiple substitutions at the same site. With revised equations, we calculated the incidence of silent and replacement substitutions in the two protein-coding genes. The silent substitution rate is 4 to 6 times higher than the replacement rate, indicating strong functional constraints at replacement sites. Moreover, the silent rate for these two genes is about 10% per million years, a value 10 times higher than the silent rate for the nuclear genes studied so far. In addition, the mean substitution rate in the three mitochondrial tRNA genes is at least 100 times higher than in nuclear tRNA genes. Finally, genealogical analysis of the sequence differences supports the view that the human lineage branched off only slightly before the gorilla and chimpanzee lineages diverged and strengthens the hypothesis that humans are more related to gorillas and chimpanzees than is the orangutan.

Key words: DNA sequencing – Transitions – Multiple-hit corrections – Silent substitutions – Replacement substitutions – Transfer RNA – Mutation pressure – Functional constraints – Hominoid phylogeny – Start codons

Introduction

The mitochondrial genome of animals is an excellent one with which to examine the tempo and mode of molecular evolution. Detailed comparisons of this genome among appropriately chosen species will also contribute important information about the structure-function relationships of mitochondrial genes and gene products. Based on the results of cleavage mapping and annealing studies, it was suggested that animal mtDNA is accumulating nucleotide substitutions 5 to 10 times faster than single-copy nuclear DNA (Brown et al. 1979). To explain this high rate of substitution in a genome so tightly packed with functional genes (Anderson et al.

[1] *Present address:* Division of Biological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

[2] *Present address:* Cetus Corporation, 600 Bancroft Way, Berkeley, California 94710, USA

Offprint requests to: E.M. Prager

Abbreviations: mtDNA = mitochondrial DNA; bp = base pair; URF = unidentified reading frame

0022–2844/82/0018/0225/$ 03.00

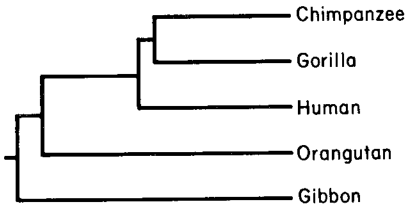


Fig. 1. Evolutionary tree for apes and humans based on mitochondrial DNA. The branching order shown is derived from a parsimony analysis of cleavage maps of the whole mitochondrial genome (Ferris et al. 1981a) and is supported by the nucleotide sequences reported in the present paper

1981; Bibb et al. 1981), it was proposed that animal mtDNA has an unusually high mutation rate (Brown et al. 1979). In order to assess the rate of base substitution in a more direct manner, and also to test the hypothesis of a higher mutation rate as the cause, we have cloned and sequenced a homologous piece of mtDNA from 5 species of hominoid primates whose evolutionary relationship is depicted in Fig. 1. Because the most distantly related hominoid species probably diverged from one another no more than 10 million years ago (Sarich and Wilson 1967; Wilson et al. 1977), this group of primates can provide information on mtDNA evolution that is relatively unobscured by multiple substitutions at the same nucleotide position.

The mtDNAs of human (Anderson et al. 1981), cow (Anderson et al. 1982), and house mouse (Bibb et al. 1981) have been completely sequenced. Comparisons of the mtDNA sequences from these species add to our understanding of mtDNA. However, the evolutionary information gained from them is limited, for the following reasons: (1) The divergence time between any two of these three species is about 80 million years, hence DNA divergence at only one time point is being sampled; (2) the degree of nucleotide difference between the mtDNAs of these species will not be appreciably different from that between species which diverged only 20 million years ago, because the readily-substituted positions in the mtDNA have become "saturated" by this time, as is evident from Fig. 3 of Brown et al. (1979). To understand the dynamics of the evolution of any macromolecule, it is necessary to choose a series of species for comparison whose divergence times are (1) different and (2) lie within a time range that is short enough to give a favorable signal-to-noise ratio. The primate species compared in this paper satisfy these conditions for mtDNA.

Comparison of the primate sequences confirms our previous estimate of the substitution rate. Moreover, by allowing us to consider the nature and precise location as well as the exact number of substitutions, the sequences provide us with new insight into the mechanism of mtDNA mutation and evolution. Finally, the differences observed among the sequences add greatly to the number of known genetic differences between apes and humans, which allows a more reliable test of hypotheses about the genealogical relationships of humans and apes.

Materials and Methods

Materials and MtDNA Preparation. Human embryonic liver was a gift from M. Golbus, University of California, San Fransisco. Gibbon (*Hylobates lar*) mtDNA was from the source described in Ferris et al. (1981a); mtDNAs from common chimpanzee (*Pan troglodytes*), pygmy chimpanzee (*P. paniscus*), and gorilla (*Gorilla gorilla*) came from the individuals designated number 1 in Figs. 3 and 5 of Ferris et al. (1981b); orangutan (*Pongo pygmaeus*) mtDNA came from the individual designated 2a in Fig. 1 of Ferris et al. (1981b). MtDNAs were prepared from liver (4 species), liver plus kidney (gorilla), or a cultured cell line (common chimpanzee) as described (Brown et al. 1979), including those from pygmy chimpanzee, gorilla, and gibbon, which were gifts from S.D. Ferris.

Restriction endonucleases were from New England BioLabs and Bethesda Research Labs. T_4 DNA ligase was obtained from P-L Biochemicals and as a gift grom W.S. Davidson and D. Julin. Bacterial alkaline phosphatase was obtained from Bethesda Research Labs and P-L Biochemicals and T_4 polynucleotide kinase from P-L Biochemicals and New England BioLabs. Calf thymus DNA and yeast tRNA were purchased from Sigma. Gamma-labelled ^{32}P-ATP was a gift from R. Myers. D. Rio, and R. Tjian or was purchased from New England Nuclear. Dimethyl sulfate, hydrazine, and piperidine were from Aldrich, Eastman, and Sigma, respectively.

DNA Cloning and Sequencing. 20 to 500 ng of mtDNA digested to completion with Hind III were added to 20 to 500 ng of Hind III digested, alkaline phosphatase treated plasmid pBR322 DNA and the mixture incubated for 4–12 h at 10°C with T_4 DNA ligase. The ligation products were used to transform *E. coli* K12 (strains χ1776 or HB101). All recombinant DNA procedures were carried out using P2 containment and EK1 or EK2 host-vector systems, in accordance with NIH guidelines.

Transformant colonies showing a tetracycline-sensitive, ampicillin-resistant phenotype were screened for the presence of mtDNA sequences using the colony hybridization procedure of Hanahan and Meselson (1980). The hybridization probe was human mtDNA labelled with P^{32} by nick translation (Cordell et al. 1979). Putative recombinants were confirmed by size analysis and Southern blot hybridization of Hind III digests of the plasmids and by further restriction endonuclease analysis of the mitochondrial inserts, using the cleavage maps published by Ferris et al. (1981a,b) as references.

After digestion of recombinant plasmids with Hind III, the mtDNA fragments were separated by preparative gel electrophoresis in 1% agarose. DNA fragments were retrieved from the gels by band excision and electroelution, followed by two extractions each of phenol, chloroform – 4% isoamyl alcohol, and ether (see Smith (1980) for details). Essentially the same procedures were used for retrieval of radioactively labelled DNA fragments from 5% polyacrylamide gels except that in the electroelution step yeast tRNA (20 μg/ml) was added to the buffer within the dialysis tubing as carrier. Early in this investigation some Hind III fragments were isolated by R.M. Watson with a preparative gel elution apparatus.

DNA sequencing was performed according to methods described by Maxam and Gilbert (1980).

Calculations. Transfer RNAs were located and cloverleaf structures determined by inspection and analogy with the human sequence (Anderson et al. 1981) and checked with the TRNA program of Staden (1980). Regression lines through the origin (Figs. 4A and 7; text) were calculated by the method of Steel and Torrie (1960). For all interordinal comparisons in this work, the 5 primates were individually compared to cow and to mouse and in each case the resulting values averaged; the three possible interordinal comparisons were then given equal weight. The methods used for correction for multiple substitutions at the same site are presented in the Appendix.

Results and Discussion

Cloning of a Homologous Region. Since the Hind III sites at 24 and 30 map units are highly conserved in the mitochondrial genome of both primates and rodents (Brown and Vinograd 1974; Brown et al. 1979; Ferris et al. 1981a,b), we cloned the DNA segment bounded by these two sites from several primate species (Fig. 2). Sequencing indicated that this region, including the terminal Hind III recognition sites, contained 896 bp and corresponded to nucleotides 11680–12575 in the human mtDNA sequence published by Anderson et al. (1981). The homologous Hind III region in mouse and cow contains 894 bp (Bibb et al. 1981) and 899 bp (Anderson et al. 1982), respectively.

The region was cloned as a single Hind III fragment in the case of human, gorilla, orangutan, and gibbon (cf. Fig. 2). This was not possible for either the common chimpanzee or the pygmy chimpanzee, both of which have an intermediate Hind III site at 26.3 map units (Fig. 2). In addition, common chimpanzee lacks a Hind III site at 24 map units. Because the estimated intrageneric differences in mtDNA between the two chimpanzee species are smaller than the intergeneric differences (Ferris et al. 1981a,b), we combined the complementary sequence data from the two respective chimpanzee clones and treated them as a single entity, hereafter designated chimpanzee.

Gene Content of the Sequenced Region

As previously determined by both transcription mapping (Attardi et al. 1980) and direct DNA sequence analysis (Anderson et al. 1981), this region contains information about five mitochondrial genes (Fig. 2). In addition to the complete genes for histidyl-, seryl-, and leucyl-transfer RNA (tRNA), portions of two genes [designated unidentified reading frame (URF) 4 and URF 5 by Anderson et al. (1981)] that code for two putative mitochondrial proteins (designated Proteins 4 and 5 in this work) are present. These portions correspond to 458 bp at the 3′ end of URF 4 and 239 bp and the 5′ end of URF 5.

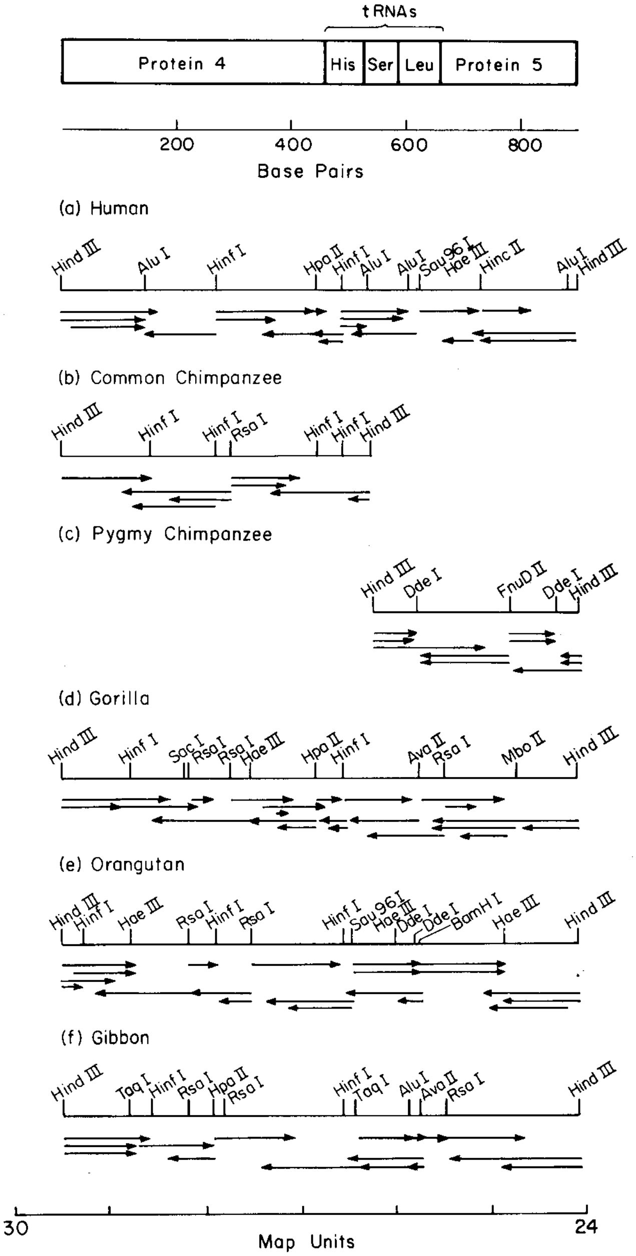


Fig. 2. Cleavage maps, sequencing strategy, and gene content of cloned segments of hominoid mitochondrial DNA. For each species all sites for the restriction endonucleases actually employed during sequencing are shown. The *arrows* in the diagram indicate the direction and length of the fragments sequenced. On the average 83% of the sequence was determined at least twice, either in a single direction or along opposite strands. For each individual species the following percentages were sequenced 2 or more times: human, 81%; common chimpanzee, 65%; pygmy chimpanzee, 97%; gorilla, 95%; orangutan, 86%; gibbon, 73%. Five different genes are represented in this piece of DNA, whose total length is 896 bp: the carboxy-terminal third of Protein 4, the complete genes for 3 different transfer RNAs, and the amino-terminal eighth of Protein 5. Map units are indicated as in Ferris et al. (1981a,b)

Sequence Comparisons

The nucleotide sequences for this mtDNA segment appear in Fig. 3. Differences among the primate sequences were observed at 284 positions. Their distribution among the 5 genes is shown in Fig. 3. Nucleotide substitutions account for the differences at 283 of these positions. In addition to substitutions, there was a deletion of a single base pair from orangutan mtDNA in a portion of the seryl-tRNA gene. Based on secondary structure considerations, the most reasonable position for the deletion is 560.

The pairwise sequence differences range from 0.2% to 19% (as Table 1 shows), in agreement with estimates made from cleavage map comparisons of the whole mtDNA genome (Ferris et al. 1981a). The two human sequences are most similar, differing (see Fig. 3) by only 2 substitutions (0.2%), a difference that is about as small as would be expected from Brown's (1980) restriction analysis of 21 human mtDNAs. Among species, the human, chimpanzee, and gorilla sequences are most similar to one another, the orangutan sequence is less similar, and the gibbon sequence is least similar. These data are consistent with the evolutionary tree shown in Fig. 1.

Comparison of the primate sequences with those of the corresponding regions in cow (Anderson et al. 1982) and mouse (Bibb et al. 1981) mtDNA provides a similar picture with respect to the relative proportions of deletions and substitutions, although both kinds of differences occur more frequently in the comparisons with these more distantly related species. The average percent sequence differences are: primate-mouse, 33; primate-cow, 30; mouse-cow, 31. Compared to the primate sequences (excluding that of orangutan), the mouse sequence exhibits 2 deletions and the cow sequence 4 additions and one deletion. All addition and deletion events in this region are confined to single base pairs and occur exclusively in or between tRNA genes.

High Substitution Rate and the Saturation Effect

A plot of percent sequence difference against divergence time (curve A, Fig. 4) illustrates how quickly the highly variable positions in mtDNA become saturated by substitutions. Sequences that have diverged for only 10 million years differ by nearly as much as do those which have diverged 80 million years ago. This graph shows how misleading it can be to infer absolute rates of evolution in mtDNA from the comparison of such distantly related sequences as those of human, cow, and mouse (see curve A, Fig. 4). The initial slope of the curve corresponds to a 2% change in sequence per million years for a pair of species. This estimate is in agreement with the value of 2% per million years inferred from cleavage mapping of the whole mitochondrial genome (Brown et al. 1979). Such a rate is 5 to 10 times higher than that for the single-copy fraction of nuclear DNA (Brown et al. 1979) and for the globin gene regions (Zimmer 1980; Barrie et al. 1981; Martin et al. 1981) of higher primates.

Predominance of Transitions

The predominance of transitions over transversions is probably the most striking result to emerge from our primate sequence comparisons (see Table 2). This evolutionary bias toward transitions is evident in both the tRNA and protein-coding genes. Moreover, it is about equally apparent at silent and replacement sites in the protein-coding genes (Table 2). Since the selection pressures operating on these three classes of nucleotide positions are extremely different, we suggest that the high proportion of transitions is due chiefly to a bias in the mutation process rather than to selection at the level of gene products.

Notable also is the time dependence of the observed proportion of transitions (curve B, Fig. 4). For the three

Table 1. Sequence differences among mammalian mitochondrial DNAs at 896 positions[a]

Species compared	Human A	Human B	Chimpanzee	Gorilla	Orangutan	Gibbon	Cow	Mouse
Human A	---	2	79	92	144	162	272	297
Human B	0.22	---	79	92	144	162	273	298
Chimpanzee	8.8	8.8	---	95	154	169	276	296
Gorilla	10.3	10.3	10.6	---	150	169	277	297
Orangutan	16.1	16.1	17.2	16.7	---	169	248	307
Gibbon	18.1	18.1	18.9	18.9	18.9	---	255	301
Cow	30.3	30.4	30.8	30.9	27.6	28.4	---	277
Mouse	33.2	33.3	33.1	33.2	34.3	33.6	30.9	---

[a] The number of sequence differences between any two mtDNAs appears in the upper right-hand section of the matrix, and the percent sequence difference is given in the lower left-hand section. Human A refers to the sequence of Anderson et al. (1981) and human B to the sequence determined in this work. The cow and mouse sequences were determined by Anderson et al. (1982) and Bibb et al. (1981), respectively. The orangutan actually has 895 nucleotides in this segment, the cow 899, and the mouse 894 (see text); the deletions and additions involved have been included along with base substitutions in calculation of sequence differences

```
                              20                  40                  60                  80                 100
                                                   G
Human       AAGCTTCACCGGCGCAGTCATTCTCATAATCGCCCACGGACTTACATCCTCATTACTATTCTGCCTAGCAAACTCAAACTACGAACGCACTCACAGTCGC
Chimpanzee                  A T  C                 A               T                      T  T        C
Gorilla                       TG    T     T        A        A      T                               A  C     C
Orangutan                   AC  CC     G  T     T  A  C        CC    G                             A  C     C
Gibbon            T  A  T   AC G C                 A  A  C  T  CC G           T                    A        C

                             120                 140                 160                 180                 200
Human       ATCATAATCCTCTCTCAAGGACTTCAAACTCTACTCCCACTAATAGCTTTTTGATGACTTCTAGCAAGCCTCGCTAACCTCGCCTTACCCCCCACTATTA
Chimpanzee          T     C                                C           C                        C       T  C
Gorilla             T              C     C                 CC             G           C                    C
Orangutan                       C                 C        CC C                     A       T   C    A     C  C
Gibbon                 A    G   G  C   G CT          G     C  C        CGC                      C

                             220                 240                 260                 280                 300
Human       ACCTACTGGGAGAACTCTCTGTGCTAGTAACCACGTTCTCCTGATCAAATATCACTCTCCTACTTACAGGACTCAACATACTAGTCACAGCCCTATACTC
Chimpanzee   T  C  A  G        C            T A                C            C      T           A          G
Gorilla            A     G     C  A           A              C C   C  TT           TCT         A T        G
Orangutan       T  A           C  A   A  G   TA     T     T  C     CA       A                  A     A
Gibbon          C  A  T       TC  A   A GG  T C        GG    C CT   A TAC   C  C  G      G     A    G

                             320                 340                 360                 380                 400
Human       CCTCTACATATTTACCACAACACAATGGGGCTCACTCACCCACCACATTAACAACATAAAACCCTCATTCACACGAGAAAACACCCTCATGTTCATACAC
Chimpanzee           G                 A                       T        G                    T  T     A  TT
Gorilla        T  T                    A   C       A        C C                  T             T      A     G
Orangutan   T     T     C            C A  TA   C   A        C              T  T        C     T         C
Gibbon         T          T  T       C A   A    T  A           A               C               TAT A  AC T  G

                             420                 440                 460                 480                 500
Human       CTATCCCCCATTCTCCTCCTATCCCTCAACCCCGACATCATTACCGGGTTTTCCTCTTGTAAATATAGTTTAACCAAAACATCAGATTGTGAATCTGACA
Chimpanzee             C     T           T  T  T     C  T  A  CA    C
Gorilla                C                       T  T  C        CA    C                                          T
Orangutan              C      T              AG      CG T     CG   AC                         T             A T
Gibbon        C T    C C         A          TA          T  C   A TC C      C         T        T             A

                             520                 540                 560                 580                 600
                                                                                        A
Human       ACAGAGGCTTACGACCCCTTATTTACCGAGAAAGCTCACAAGAACTGCTAACTCATGCCCCCATGTCTGACAACATGGCTTTCTCAACTTTTAAAGGATA
Chimpanzee           C                          T T            T    AT       C
Gorilla              C  A                        GT    G            A     G  CT
Orangutan    T  G C CC  A                                         TCA T-      G                   G
Gibbon       T       CGAA   T   GC             C                   CTAT       A

                             620                 640                 660                 680                 700
Human       ACAGCTATCCATTGGTCTTAGGCCCCAAAAATTTTGGTGCAACTCCAAATAAAAGTAATAACCATGCACACTACTATAACCACCCTAACCCTGACTTCCC
Chimpanzee       C    G                                                       T T     C         T       A   C  T
Gorilla                           A                                       T   T  G    C         T  G    A      T
Orangutan             C           AT                                  C G     TTT  C  C     TG    C   T A
Gibbon                            A                                     G A   T    C  C   G   TT     G  A  C

                             720                 740                 760                 780                 800
Human       TAATTCCCCCCATCCTTACCACCCTCGTTAACCCTAACAAAAAAAACTCATACCCCCATTATGTAAAATCCATTGTCGCATCCACCTTTATTATCAGTCT
Chimpanzee        T         C         A                         T           G         A    G           C  T  C
Gorilla               T            T  A C  T             G               C        T  C                 C     C
Orangutan       C        TACCG T      A       C            C          C         A GGCCA        G      C      C
Gibbon                   TACAG       TA       C  T     G    T      G  C  C      ATG CCA T  C  T        A     C

                             820                 840                 860                 880         896
Human       CTTCCCCACAACAATATTCATGTGCCTAGACCAAGAAGTTATTATCTCGAACTGACACTGAGCCACAACCCAAACAACCCAGCTCTCCCTAAGCTT
Chimpanzee  T                    A                C         A     G        A
Gorilla                       TC A                C         A G            A             TT  A
Orangutan   TA    A           T  C       GA      ACC  CG  A A      TG    A A  C     G   CTA  A     A
Gibbon      A  T       T             AC          ACC     T  A            A TG          GCTAG A
```

Fig. 3. Sequences of 896 bp fragments of primate mitochondrial DNAs. Common and pygmy chimpanzee are considered as one entity (see text). The published (Anderson et al. 1981) human mtDNA sequence is shown in the uppermost line in small capital letters. Differences from the published human sequence are indicated with large capital letters. Since our human mtDNA differs in sequence from the published one only at positions 40 and 569, the human is listed only once, with the changes observed here indicated as just described. Sequence differences exist at 284 positions among the primates studied. At 177 positions the changes are unique – i.e., there is a different base in only one species. The distribution of these unique changes is human, 17; chimpanzee, 19; gorilla, 22; orangutan, 55; gibbon, 64. At the remaining 107 sites the changes are non-unique – i.e., in at least two lineages there is a base which is different from that existing in the remaining lineages. The only deletion was at position 560 in the orangutan, and there were no additions. The 5 genes present encompass the following nucleotides: Protein 4, 1–458; histidine tRNA, 459–527; serine tRNA, 528–586; leucine tRNA, 587–657; Protein 5, 658–896

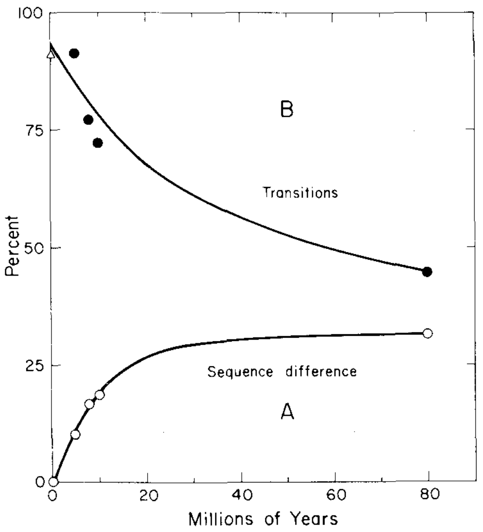
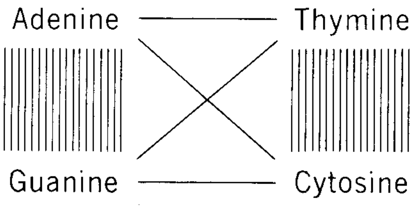


Fig. 4. Dependence of sequence difference and proportion of transitions on divergence time for a cloned segment of mitochondrial DNA. In curve A the uncorrected percentage difference in nucleotide sequence is plotted against time of divergence, by analogy with Fig. 3 of Brown et al. (1979). For interspecific comparisons divergence times were based on Sarich and Wilson (1967), Brown et al. (1979), and Pilbeam (1979). The intrahuman divergence is placed at 40,000 years, i.e., the time when modern humans appeared (Brues 1977). In curve B the observed percentage of the sequence differences that are transitions is plotted against divergence time. The solid circles are for interspecific comparisons of the segment of mtDNA we sequenced. The empty triangle is based on intraspecific comparisons of two types: (1) The human segment (896 bp) we sequenced and (2) the rat sequence of 169 bp determined by Castora et al. (1980), De Vos et al. (1980), and Goddard et al. (1981). In both curves A and B each point is the average for all comparisons made for a particular divergence time. The method of averaging the interordinal comparisons is stated in Materials and Methods. Most of the values on which the figure is based appear in Tables 1 and 2

most recently diverged pairs of species, an average of 92% of all point-mutational differences are transitions. For pairs with greater divergence times, the proportion of transitions declines until, as shown by curve B of Fig. 4, one reaches cases in which transversions appear to outnumber transitions.

We explain this time-dependence by supposing that

(1) the evolutionary process is heavily biased toward transitions, as shown in the following diagram:



(2) multiple substitutions occur at the same site when the time is long, obscuring the record of transitions. This last point is illustrated by the following hypothetical case of evolution by 10 point mutations at the same site along two diverging lineages:

Table 2. Percent transitions in mammalian mitochondrial DNA[a]

Portion of genome compared	Divergence time, millions of years			
	5	8	10	80
Protein 4				
Silent sites	93.9	78.6	73.4	49.1
Replacement sites	67.9	80.2	68.5	39.3
Protein 5				
Silent sites	96.2	75.8	73.8	46.8
Replacement sites	93.3	70.4	70.0	32.8
Transfer RNAs	94.2	75.8	75.8	67.0
Total	91.5	76.8	72.1	44.5

[a] For the first three columns, each value in the table is the average of the 3 or 4 pairwise comparisons of primates which diverged from one another 5, 8, or 10 million years ago (cf. Sarich and Wilson 1967). The final column shows comparisons of different orders (computed as described in Materials and Methods), which diverged 80 million years ago (cf. Brown et al. 1979). The last line is for the entire 896 bp fragment sequenced in the primates, save for those few positions where a deletion or addition occurs. Silent sites in protein-coding genes are those at which base substitutions can occur without changing the amino acid encoded, while replacement sites are those at which base substitutions can change the amino acid encoded

$$\begin{tikzcd}
 & C \arrow[r] & T \arrow[r] & C \arrow[r] & T \arrow[r] & C \\
G \arrow[ru] \arrow[rd] & & & & & \\
 & A \arrow[r] & G \arrow[r] & A \arrow[r] & G \arrow[r] & A
\end{tikzcd}$$

Nine of the 10 substitutions are transitions, but the overall process would be scored in an evolutionary analysis as one C ↔ A transversion. By contrast, there is no simple way in which one transition can erase the record of transversions. Hence, in the zone of saturation, i.e., from 10 million years on (curve A, Fig. 4), one expects and observes (curve B, Fig. 4) the ratio of scored transitions to scored transversions to be rather low, even when the actual number of transitions exceeds the actual number of transversions by a big factor.

Further evidence that transitions predominate in mtDNA evolution comes from comparison of individuals within a species. In rats (Castora et al. 1980; De Vos et al. 1980; Goddard et al. 1981), a total of 6 substitutions were identified, 5 of which were transitions. In humans (Walberg and Clayton 1981; this work) all 7 of the substitutions detected are transitions; further work by B. Greenberg (pers. commun.) also shows a preponderance of transitions in non-coding regions.

By extrapolating curve B, as shown in Fig. 4, to zero time of divergence, we eliminate the problem of multiple substitutions at a site and thereby get an approximate estimate of the actual proportion of evolutionary substitutions that are transitions. Our extrapolation is a conservative one and the resulting estimate is 90 ± 5%.

In both bacterial and nuclear DNA evolution, a slight bias toward transitions has long been known (Derancourt et al. 1967; Jukes 1980; Fitch 1980; Nichols et al. 1980). Although more work with very closely related DNA sequences is needed to measure the extent of the substitutional bias in these types of DNA, it is apparent already that the percentage of transitions in globin genes is about 70 for short times of divergence (Martin et al. 1981) and below 50 for long divergence times (Derancourt et al. 1967; Martin, Vincent, and Wilson, unpub. work).

A practical consequence of the observation that transitions and transversions occur with profoundly unequal probabilities is that this must be taken into account in any attempt to correct an observed sequence difference for multiple substitutions at the same site. Several widely used methods of correction are based on the assumption that all base substitutions are equally likely and thus that the ratio of transitions to transversions is 1 to 2. Therefore, we have revised one of these methods, as described in the Appendix, so that corrections can be made for any transition/transversion ratio and also to take into account the differences between the mitochondrial and "universal" genetic codes.

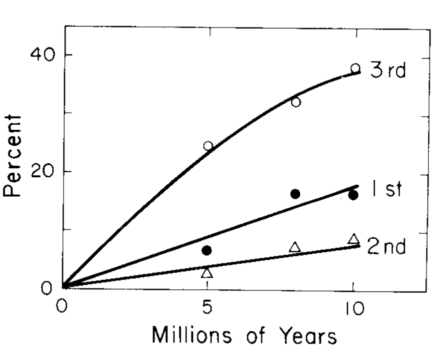


Fig. 5. Sequence difference as a function of divergence time for the three positions in primate mitochondrial codons. This figure is based on the nucleotide sequences for the two coding regions (URF 4 and 5) in an 896-bp segment of mtDNA from 5 primates. Altogether 229 codons were considered. For each position, the number of observed differences was divided by 229 and converted to percent. Each point on the graph is the average for all comparisons made for a given divergence time. The divergence times are taken from Sarich and Wilson (1967)

Protein-Coding Sequences

The sequences of the two protein-coding regions were examined in the three possible reading frames for each primate species. In each region (i.e., for bases 1–458 and 658–896, see Fig. 3), only one reading frame was open; the other two frames each contained several stop codons. Visual inspection of the aligned sequences (Fig. 3) suggests that most of the differences occur with a spacing of 3 bp. Quantitative analysis confirms this and, further, shows (see Fig. 5) that sequence differences occur most often at third positions of codons and least often at second positions of codons in the open reading frame. The open frames for URF 4 and 5 are translated in Fig. 6.

Consistent with the result illustrated in Fig. 5, most of the base substitutions in the two coding sequences of primate mtDNAs are silent, i.e., do not cause amino acid substitutions (Fig. 6 and Table 3). By contrast, when distantly related mtDNAs are compared, silent substitutions are ostensibly less common than replace-

Table 3. The observed number of silent (*S*) and replacement (*R*) nucleotide substitutions in portions of the URF 4 and URF 5 genes of mammalian mtDNAs[a]

Species compared	Number of substitutions observed					
	URF 4			URF 5		
	S	R	S/R	S	R	S/R
Human/chimpanzee	40	7	5.71	17	6	2.83
Human/gorilla	42	10	4.20	18	10	1.80
Human/orangutan	48	20	2.40	28.5	26.5	1.08
Human/gibbon	54.5	33.5	1.63	29.5	23.5	1.26
Chimpanzee/gorilla	46	9	5.11	17	10	1.70
Chimpanzee/orangutan	54	22	2.45	27.5	25.5	1.08
Chimpanzee/gibbon	57.5	30.5	1.89	33.5	21.5	1.56
Gorilla/orangutan	48	24	2.00	26.5	29.5	0.90
Gorilla/gibbon	57.5	32.5	1.77	29	26	1.12
Orangutan/gibbon	61	33	1.85	30.5	22.5	1.35
Primate/cow	77.7	56.9	1.37	33.1	57.3	0.58
Primate/mouse	75.6	75.0	1.01	31.2	75.4	0.41
Cow/mouse	71	62	1.15	28	74	0.38

[a] URF 4 was analyzed at 152 codons and URF 5 at 77 codons. Within primates fractional values are sometimes tabulated for the number of changes because some changes are half-silent and half-replacement (cf. Appendix, *Corrected Divergence at Silent and Replacement Sites*). Interordinal averages were computed as described (Materials and Methods)

PROTEIN 4

			10										40										70							
Human	Ser	Phe	Thr	Gly	Ala	Val	Ile	Leu	Met	Ile	Ala	His	Gl*y*	Leu	Thr	Ser	Ser	Leu	Leu	Phe	Cys	Leu	Ala	Asn	Ser	Asn	Tyr	Glu	Arg	Thr
Chimpanzee						Ile	Il*e*						Gl*y*						*L*eu							As*n*	Ty*r*			Th*r*
Gorilla						Va*l*	Val	Le*u*		Il*e*			Gl*y*			Se*r*			*L*eu										Ar*g*	Th*r*
Orangutan						Thr	Thr		Me*t*	Il*e*		Hi*s*	Gl*y*	Le*u*			Se*r*	*L*e*u*	Le*u*										Ar*g*	Th*r*
Gibbon		Phe	Th*r*	Gl*y*		Thr	Va*l*						Gl*y*	Le*u*	Th*r*	Se*r*	Se*r*	*L*e*u*				Le*u*							Ar*g*	

			100										130										160							
Human	His	Ser	Arg	Ile	Met	Ile	Leu	Ser	Gln	Gly	Leu	Gln	Thr	Leu	Leu	Pro	Leu	Met	Ala	Phe	Trp	Trp	Leu	Leu	Ala	Ser	Leu	Ala	Asn	Leu
Chimpanzee						Il*e*		Se*r*	Gln				Thr						Al*a*	Phe			Le*u*	Leu				Ala		
Gorilla		Se*r*				Il*e*			Gln		Le*u*		Th*r*						Al*a*	Leu				Le*u*				Al*a*		
Orangutan		Se*r*							Gln	Gl*y*			Thr			Pr*o*			Al*a*	Leu				Leu				Thr		Le*u*
Gibbon		Se*r*					Le*u*		Arg	Gl*y*	Le*u*		Ala	*L*eu			Le*u*		Al*a*	Phe			Le*u*	Ala				Ala		

			190										220										250							
Human	Ala	Leu	Pro	Pro	Thr	Ile	Asn	Leu	Leu	Gly	Glu	Leu	Ser	Val	Leu	Val	Thr	Thr	Phe	Ser	Trp	Ser	Asn	Ile	Thr	Leu	Leu	Leu	Thr	Gly
Chimpanzee		*L*eu		Pr*o*	Th*r*		As*n*	Le*u*	Le*u*	Gl*y*			Se*r*			Val	Thr	Ser				Ser		Thr		Leu	Leu	Le*u*		
Gorilla					Th*r*				Le*u*		Gl*u*		Se*r*	Va*l*		Val	Thr	Th*r*				Ser	As*n*	Thr	Th*r*	Le*u*	*L*eu			
Orangutan		*L*eu	Pr*o*		Th*r*	Il*e*		Le*u*	Leu				Se*r*	Va*l*		Met	Ala	Met		Se*r*		Se*r*	As*n*	Ile	Th*r*	Ile	Leu	Le*u*		
Gibbon		*L*eu						Le*u*	Le*u*	Gl*y*			Phe	Va*l*		Me*t*	Ala	Se*r*			Tr*p*	Ala	As*n*	Th*r*		Il*e*	Thr	Le*u*	Th*r*	Gl*y*

			280										310										340							
Human	Leu	Asn	Met	Leu	Val	Thr	Ala	Leu	Tyr	Ser	Leu	Tyr	Met	Phe	Thr	Thr	Thr	Gln	Trp	Gly	Ser	Leu	Thr	His	His	Ile	Asn	Asn	Met	Lys
Chimpanzee	Phe		Met		Ile		Ala	Le*u*					Me*t*		Thr	Thr			Tr*p*		Ser	Leu					As*n*			Ly*s*
Gorilla	Ser		Met		Il*e*		Ala	Le*u*			Le*u*	Ty*r*			Thr	Thr			Tr*p*		Pro	Leu	Th*r*			Il*e*	Thr			
Orangutan	Leu		Met		Ile		Thr			Se*r*		Ty*r*		Ph*e*	Thr	Thr			Arg	Gl*y*	Thr	Pro	Th*r*			Il*e*	Asn			
Gibbon	Leu		Val		Ile	Th*r*	Ala				Le*u*				Ile	Met			Arg		Thr	Le*u*	Th*r*				Lys			

			370										400										430									
Human	Pro	Ser	Phe	Thr	Arg	Glu	Asn	Thr	Leu	Met	Phe	Met	His	Leu	Ser	Pro	Ile	Leu	Leu	Leu	Ser	Leu	Asn	Pro	Asp	Ile	Ile	Thr	Gly	Phe	Ser	Ser
Chimpanzee			Phe				As*n*	Th*r*		Me*t*	Ph*e*	Leu			Ser		Il*e*		Le*u*		Ser		As*n*	Pr*o*	As*p*		Ile	Th*r*	Gl*y*	Ph*e*	Thr	Se*r*
Gorilla			Ph*e*					Ile		Me*t*	Phe	Me*t*			Ser		Il*e*				Ser				As*p*	Il*e*	Il*e*	Thr		Ph*e*	Thr	Se*r*
Orangutan	Pr*o*	Se*r*	Phe		Ar*g*		As*n*	Thr			Leu	Met			Ser		Il*e*			*L*eu	Ser				Ser		Il*e*	Ala		Ph*e*	Ala	Tyr
Gibbon			Leu					Met	*L*eu	Me*t*	Le*u*	Me*t*		Le*u*	Phe		Leu				Thr			Pr*o*	Asn			Th*r*	Gl*y*		Th*r*	Pro

PROTEIN 5

	660										690										720									
Human	Met	Thr	Met	His	Thr	Thr	Met	Thr	Thr	Leu	Thr	Leu	Thr	Ser	Leu	Ile	Pro	Pro	Ile	Leu	Thr	Thr	Leu	Val	Asn	Pro	Asn	Lys	Lys	Asn
Chimpanzee	Met	Thr		Tyr	Thr	Th*r*		Thr	Thr	*L*eu	Thr	Le*u*		Pro	*L*eu		Leu			Le*u*	Thr		Leu	Ile						Asn
Gorilla	Met	Th*r*		Ty*r*	Ala	Th*r*		Thr	Thr	*L*eu	Ala	Le*u*		Ser	*L*eu		Pro	Pr*o*		Leu	Thr		Phe	Il*e*	As*n*					Ser
Orangutan	Thr	Ala		Phe	Th*r*	Th*r*		Th*r*	Ala	Le*u*	Thr	*L*e*u*		Ser		Il*e*	Pro		Il*e*	Thr	Ala		Leu	Ile		Pr*o*				Asn
Gibbon	Met	Al*a*		Ty*r*	Th*r*	Th*r*		Ala	Ile		Th*r*	Le*u*	Th*r*	Ser			Pro		Il*e*	Thr	Al*a*		Le*u*	Ile		Pr*o*	As*n*		Ly*s*	Asn

	750										780										810									
Human	Ser	Tyr	Pro	His	Tyr	Val	Lys	Ser	Ile	Val	Ala	Ser	Thr	Phe	Ile	Ile	Ser	Leu	Phe	Pro	Thr	Thr	Met	Phe	Met	Cys	Leu	Asp	Gln	Glu
Chimpanzee	Ser	Ty*r*				Va*l*		Ser	Ile	Ile	Al*a*		Thr		Il*e*	Il*e*	Se*r*	Le*u*	Phe			Thr			Me*t*		Leu	Asp		
Gorilla	Ser				Ty*r*			Se*r*	Il*e*	Val			Thr		Il*e*		Se*r*		Phe			Thr		Ph*e*	Leu		Leu	Asp		
Orangutan	Pro			His				Thr	Ala	Ile			Ala		Thr		Se*r*	Le*u*	Ile	Pr*o*		Thr		Ph*e*	Ile		Leu	Gly		
Gibbon	Leu		Pr*o*	His	Ty*r*			Met	Thr	Il*e*	Al*a*	Se*r*	Thr		Met		Se*r*	Le*u*	Ph*e*			Met			Met		Thr	Asp		

	840										870								
Human	Val	Ile	Ile	Ser	Asn	Trp	His	Trp	Ala	Thr	Thr	Gln	Thr	Thr	Gln	Leu	Ser	Leu	Ser
Chimpanzee	Ala		Ile	Se*r*	Asn	Tr*p*	His		Al*a*	Thr				Thr	Gln				
Gorilla	Ala		Ile	Se*r*	Ser		His		Al*a*	Thr				Ile	Gl*n*				
Orangutan	Thr	Il*e*	Val	Thr	Asn		Cys		Thr	Th*r*		Gl*n*		Leu	Gl*n*		Se*r*		
Gibbon	Thr		Il*e*	Se*r*	Asn		His		Th*r*	Ala			Th*r*	Leu	Glu				

Fig. 6. Predicted amino acid sequence variation and silent substitutions in Proteins 4 and 5 of primates. In the upper portion of the figure the 456 nucleotides from residues 2–457 coding for the 152 carboxy-terminal amino acids of Protein 4 are translated. For the 41 codons which are invariant in all 5 species, only the human sequence is shown. For the 75 codons where only silent substitutions occur relative to the published (Anderson et al. 1981) human sequence, the sequence is shown for those species in which such a silent change occurs; *italics* at the third, first, or both letters of the 3-letter amino acid abbreviations are used to indicate that silent substitutions occur at the corresponding positions of the codon. At the 36 codons where amino acid replacements occur in one or more species, the sequence of all 5 species is shown, with any additional silent changes among 2 or more species with the same amino acid at that position indicated with *italics* as just described. The numbers above the human sequence correspond to those in Fig. 3. In the lower portion of the figure the 237 nucleotides from residues 658–894 coding for the 79 amino-terminal residues of Protein 5 are translated. Invariant codons, silent substitutions, amino acid replacements, and nucleotide residue numbers are indicated as for Protein 4. 16 codons are invariant in all 5 species, 29 contain only silent substitutions, and 34 code for amino acid changes in one or more species. Note that this portion of the orangutan sequence does not begin with a methionine (cf. text and Fig. 8)

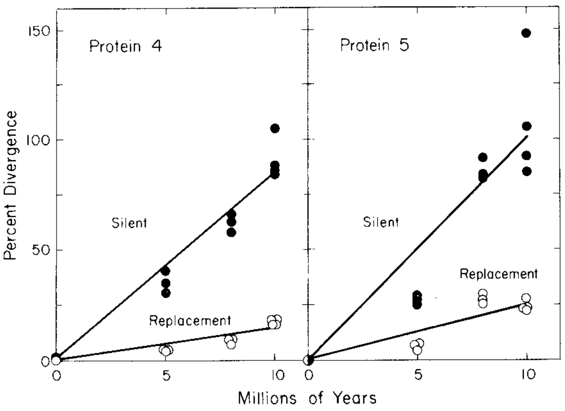


Fig. 7. Corrected divergence at silent and replacement sites in two protein-coding genes of mitochondrial DNA. The corrected values for the percent divergence were calculated as described in the Appendix and plotted versus the divergence times given in Fig. 5. *Protein 4:* The regression lines for silent and replacement substitutions are, respectively, $y = 8.5x$ and $y = 1.4x$, where x = divergence time in millions of years and y = percent divergence. From the slopes of these lines the ratio of silent to replacement change is calculated as 6.1. *Protein 5:* The regression lines for silent and replacement substitutions are, respectively, $y = 10.1x$ and $y = 2.5x$; the calculated ratio of silent to replacement change is 4.0. (URF 4 was analyzed at 456 sites, corresponding to residues 2–457 of the fragment and encoding 152 amino acids. URF 5 was analyzed at 231 sites, corresponding to residues 664–894 and encoding 77 amino acids; the 6 additional nucleotides encoding 2 more amino acids at the amino-terminus of Protein 5 were omitted because of the uncertainty about the position of the initiation codon in the orangutan)

ment substitutions (Table 3), because of multiple substitutions at silent sites.

Fig. 7 compares the estimated rates of silent and replacement evolution in the two protein-coding genes, URF 4 and 5. These estimates were made by correcting for multiple substitutions at the same site with revised equations (see Appendix). For URF 4, the silent rate exceeds the replacement rate by a factor of 6, and for URF 5 the corresponding value is 4. The fact that these values are far greater than 1 is indicative of strong functional constraints on the protein-coding sequences in mtDNA. It appears from the results summarized in Table 4 that the constraints operating on URF 4 and 5 are nearly as strong as those on nuclear genes coding for globin chains and preproinsulin.

It is important to emphasize how high the absolute rate of evolution is for silent substitutions in these two protein-coding sequences. The silent rates shown in Fig. 7 are both about 10% per million years, which is 10 times higher than the silent rate estimated for globin and other nuclear genes (Perler et al. 1980; Efstratiadis et al. 1980; Martin et al. 1981), as shown in Table 4.

Next, we consider the kinetics of amino acid substitution. The dependence of the observed number of amino acid substitutions on divergence time is roughly like that shown for nucleotide substitutions in curve A of Fig. 4. That a saturation effect exists is evident from comparing the number of amino acid sequence differences for 10 million years of divergence with that for 80 million years of divergence. The former number is over half (58%) of the latter for URF 4 and nearly half as large (43%) in the case of URF 5. We conclude that the kinetics of amino acid substitution are biphasic, as they are at the nucleotide level. This kinetic behavior is probably the result of the accumulation of multiple amino acid substitutions at some of the amino acid sites.

It is notable also that the saturation effect for amino acids is not as pronounced as at the nucleotide level. For nucleotide substitutions the values at 10 million years relative to those at 80 million years are 65% and 55% for URF 4 and 5, respectively. This is presumably because the rate of amino acid substitution is lower than the rate of nucleotide substitution.

Table 4. Rates of evolution in nuclear and mitochondrial genes

Type of site	Mean rate of evolution[a]	
	Nuclear	Mitochondrial
Silent	1.0	9.4
Replacement	0.13	2.0
Transfer RNA	<0.01	1.7

[a] The rate of evolution is given as the corrected percent divergence per million years. The silent and replacement rates for nuclear genes (i.e., preproinsulin and globin genes) come from Perler et al. (1980), while those for mitochondrial genes (URF 4 and 5) come from Fig. 7. The transfer RNA rates come from the text and Fig. 10

Start and Stop Signals: An Anomalous Initiation Codon?

As has been observed before (Anderson et al. 1981, 1982; Bibb et al. 1981), the termination codon for URF 4 (TAA) is incomplete in the coding sequence, which contains only the initial T residue; this codon is presumably completed by polyadenylation of the messenger RNA transcript after it is processed. These same investigators observed that the start codon for URF 5 is ATA in human and cow, and ATC in the mouse, instead of the more normal ATG. This ATA start codon is also present in 4 of the 5 primate species that we have investigated. In one species, however, we observed a different potential start codon, one not previously seen by other investigators.

In the orangutan sequence, as documented in Fig. 8, ACA is the first codon in URF 5, if this coding region starts at position 658 [as Anderson et al. (1981) have inferred]. An ACA codon has not been observed in an

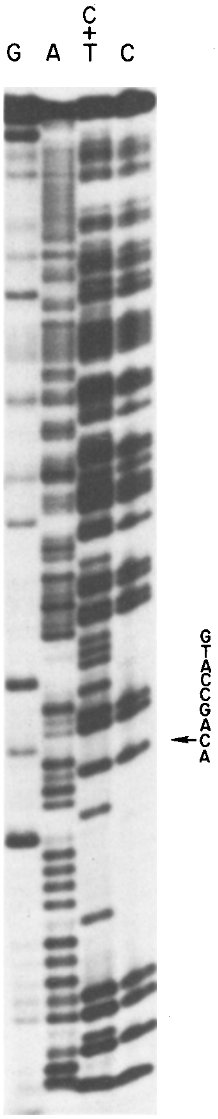


Fig. 8. Documentation of the nucleotide sequence in the initiation region of the orangutan gene for Protein 5. This autoradiogram of a sequencing gel shows DNA fragments produced by the Maxam and Gilbert (1980) method for a portion of the cloned segment of orangutan mtDNA (from residues 640 to 772). The fragments were separated electrophoretically in gels containing 8% polyacrylamide at 1.6 kV for 2.5 h. The orangutan, unlike all other hominoids examined, has a C rather than a T at position 659 (arrow)

initiation position for any other mitochondrial protein gene (Anderson et al. 1981; Bibb et al. 1981). We note the presence of a canonical ATG start codon, in phase, two codons downstream from the ACA in the orangutan sequence, Figs. 3, 6, and 8, and at the same position in the sequences from the other primate species, Figs. 3 and 6. Acceptable start codons are also present at this second position in both cow (ATA: Anderson et al. 1982) and mouse (ATT: Bibb et al. 1981) mtDNAs.

Comparison of tRNA Genes

The locations of sequence differences and probable secondary structures of the three tRNAs encoded within the 896 bp fragment appear in Fig. 9. Considering all three tRNAs together, it can be seen that no major structural feature has been immune to substitution. Interspecific differences are seen in all loops, in all stems, and in all regions that interconnect stems. The lack of a dihydrouridine loop in human mitochondrial seryl-tRNA, originally described by de Bruijn et al. (1980), is also shown by the other hominoid species. Four of the 5 nucleotides which take the place of this loop are seen to be highly variable, as are the 4 nucleotides across the center of the cloverleaf from them, Fig. 9. The 1 bp deletion observed in the orangutan also has almost certainly occurred among these latter 4 bases. The sequence of the pseudouridine loop appears, Fig. 9, to be particularly variable in the histidyl-tRNA gene, which shows differences at 5 of its 7 nucleotide positions. Finally, though the leucyl-tRNA gene is the least variable of the tRNA genes among these hominoid species, it is the only gene to show substitutions in the anticodon loop and an imperfect base pair (in orangutan) at the very top of the amino acid stem.

One of the most remarkable results of mtDNA sequence studies is the discovery of the tremendous variability shown by its tRNA genes, compared to the tRNA genes in nuclear and bacterial DNA. As illustrated in Fig. 10, identical nuclear tRNA sequences occur in organisms whose lineages diverged over 500 million years ago (Sprinzl et al. 1980). In contrast, the mitochondrial tRNAs of species that diverged only 5 million years ago differ significantly in nucleotide sequence. This suggests that mitochondrial tRNA genes evolve at least 100 times faster than their nuclear counterparts. For the three mitochondrial tRNA genes combined, the mean corrected rate of substitution among the primates, 1.7% per 10^6 years, is comparable to the rate at the replacement

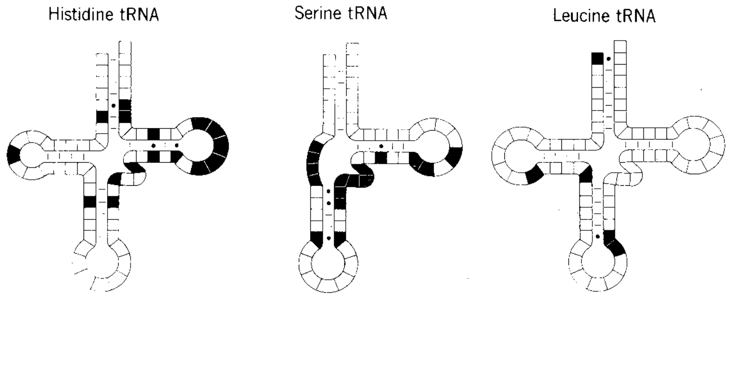


Fig. 9. Sequence variation in mitochondrial transfer RNA genes of primates. Dark areas indicate positions at which base substitutions have taken place. A solid line indicates an AT or GC base pair, while a dot indicates that in some species another base pair occurs

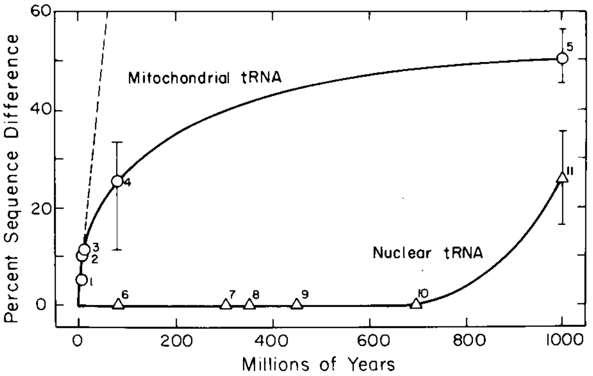


Fig. 10. Accumulation of point mutations in transfer RNA sequences specified by mitochondrial and nuclear DNA. Uncorrected percentage differences in nucleotide sequence among tRNAs reading the same codon(s) are plotted versus divergence time. The tRNA sequences were all converted to the parent DNA sequences, and differences arising during and after transcription to yield bases other than A, G, U, and C in the final tRNA products were not considered. Average values are shown, and the thin vertical lines indicate the range of values observed at a given time point. The dashed line shows the initial rate of change in mt tRNA. Where tRNAs being compared differed greatly in length (cf. point 5), only the homologous regions were compared. The data points are based on comparisons of the tRNAs listed below, with the term primate used to refer collectively to the five hominoids whose mtDNA was sequenced in this work. *Mt tRNAs:* (1) His, ser, and leu (CUX) (human, chimpanzee, and gorilla); (2) His, ser, and leu (orangutan vs. human, chimpanzee, and gorilla); (3) His, ser, and leu (gibbon vs. human, chimp, gorilla, and orangutan); (4) His, ser, and leu (primate, cow, and mouse); Thr (human and cow); Phe, val, and leu (UU$^{A}_{G}$) (human and mouse); (5) His (primate and yeast); Ser (primate and cow vs. yeast and *Aspergillus*); Leu (CUX) (primate vs. *Neurospora* and *Aspergillus*); Thr (human and cow vs. *Neurospora* and *Aspergillus*). *tRNAs Specified by Nuclear DNA:* (6) Asn (human and rat); Lys and met (mouse and rabbit); Met initiator (human, sheep, mouse, and rabbit); Phe (human, cow, and rabbit); Val (human, mouse, and rabbit); (7) Trp (cow and chicken); (8) Met initiator (mammalian (cf. point 6) and *Xenopus*); (9) Met initiator (mammalian (cf. point 6) and *Xenopus* vs. salmon); (10) Lys (mouse and rabbit vs. *Drosophila*); (11) Gly (human and yeast); Lys and met (mouse and rabbit vs. yeast); Met initiator (vertebrate (cf. point 9) vs. yeast and *Neurospora*); Phe (mammalian (cf. point 6) and yeast); Trp (vertebrate (cf. point 7) and yeast); Val (mammalian (cf. point 6) and yeast). All the cytoplasmic tRNA sequences were from Sprinzl et al. (1980). The mt tRNA sequences came from the following sources: vertebrates, this work, Anderson et al. (1981, 1982), and Bibb et al. (1981); yeast, Martin et al. (1980) and Sprinzl et al. (1980); *Neurospora*, Heckman et al. (1980); *Aspergillus*, Köchel et al. (1981). The divergence times were taken from Romer (1966), Sarich and Wilson (1967), Holmquist et al. (1976), Clemmey (1976), Brown et al. (1979), and Pilbeam (1979)

sites in the protein-coding regions (cf. Table 4). Some appreciation of the high rate at which mitochondrial tRNAs can evolve was already obtained from the comparison of distantly related mtDNA sequences (Anderson et al. 1981; Bibb et al. 1981; Saccone et al. 1981). A full appreciation, however, had to await the comparison of such closely related sequences as those of the hominoid primates.

The relative rates of substitution of the individual tRNA genes (seryl > histidyl > leucyl) vary inversely with the frequency of occurrence of their respective codons in the mtDNA. The generality of this correlation is doubtful, however, because of the report that in rodents the infrequently used leucyl tRNA^{UUR} evolves at a relatively low rate (Saccone et al. 1981). More kinds of tRNA genes need to be sequenced from closely related species in order to test whether codon usage should be regarded as a significant functional constraint on tRNA evolution.

The tRNA genes of mitochondria contrast also with those of bacteria. In bacteria the rate of tRNA evolution is much lower than the replacement rate in protein-coding genes. This is evident from comparing *Salmonella* and *Escherichia*. Four types of tRNA have been compared and in every case the two bacteria are identical (Singer and Smith 1972; Dayhoff 1973, 1976). Yet the protein-coding genes so far compared differ by about 10% in sequence between these two bacteria (Cocks and Wilson 1972; Nichols et al. 1980). Thus for both nuclear and bacterial genomes the rate of evolutionary substitution in tRNA genes is extremely low relative to the rate of protein evolution.

Mechanisms of MtDNA Evolution

Two observations about the high rate of evolutionary change in mtDNA call for explanation. First, the mitochondrial genes coding for proteins evolve 5 to 10 times faster than typical genes in the nucleus; there is direct evidence for this from our sequence studies and indirect evidence from restriction analysis of the whole genome (Brown et al. 1979; Ferris et al. 1981a,b), 70% of which appears to code for proteins. Second, the tRNA genes evolve at least 100 times faster in the mitochondrial genome than in the nuclear genome.

Any attempt to account for the rate of evolution (*E*) can benefit from recalling the equation

$$E = MF \qquad (1)$$

where *M* is the mutation rate per population and *F* is the fraction of mutations fixed. The high rate at which the whole mitochondrial genome evolves can be attributed, in principle, to either or both of the following, *viz* an elevated mutation rate and an enhanced probability of fixation.

The possibility of an elevated mutation rate for mtDNA was suggested by Brown et al. (1979) and Brown (1981) and remains attractive because our sequencing work has produced three results consistent with this suggestion:

(1) The high incidence of transitions, which could be a manifestation of enhanced mutation pressure (cf. Freese and Yoshida 1965; Topal and Fresco 1976; Sinha and Haimes 1981).

(2) The high rate of silent substitution, which may indicate a high rate of neutral mutation (cf. Kimura 1981).

(3) The high ratio of silent to replacement substitutions, which may indicate that the fraction of mutations fixed at replacement sites is nearly as low as for nuclear genes.

The factors responsible for a higher mutation rate in mtDNA than in nuclear DNA could include (1) greater exposure to oxidative damage, (2) a more error-prone system of replication, (3) less efficient editing or repair functions, and (4) a higher rate of turnover.

We should not, however, disregard the possibility that the fixation probability is also elevated for mtDNA mutations. The probability of fixation is in theory affected by factors such as population size and mode of inheritance. Since mtDNA differs conspicuously from nuclear DNA in both of these respects, it is important to devise ways of assessing the effect of these parameters on the rate of mtDNA evolution.

The 100-fold elevation of the rate of tRNA gene evolution in mtDNA requires an additional explanation. Ten of this 100-fold elevation can presumably be accounted for by the mechanisms discussed above. We attribute the additional factor of 10 to relaxed functional constraints on tRNA genes (Cann et al. 1982). Nuclear tRNA genes have three major functions: they are specifically transcribed by RNA polymerase III and associated transcriptional factors, the transcripts are modified specifically by numerous enzymes, and the mature tRNAs engage in protein synthesis and numerous regulatory processes. Some of these functions are almost certainly lacking in the case of mitochondrial tRNA genes. The small size of mitochondrial tRNA genes and their apparently low degree of base modification (Cedergren et al. 1981) are consistent with this possibility. Furthermore, as pointed out by Jukes (1981) and Cann et al. (1982), the protein-synthesizing machinery is likely to be less constrained in a small genome specifying only 13 polypeptides than in one specifying thousands of polypeptides. Thus the factor which allows the mitochondrial tRNA genes to evolve quickly may also be the one which allows the genetic code of mtDNA to drift (Cann et al. 1982).

Evolutionary Tree

In one respect the present study has confirmed the phylogenetic conclusion drawn in an earlier study (Ferris et al. 1981a) based on cleavage map comparisons of the whole mitochondrial genome. The parsimony method of tree construction, when applied to the 896 bp sequences, favors the branching order (A) which associates chimpanzee and gorilla most closely (Fig. 11), but some alternative branching orders (namely B, C, and D in Fig. 11) cannot be ruled out be the present data.

In another respect the mtDNA sequence data are more decisive. They appear to rule out tree E, which was suggested by some anatomical work (Kluge 1982). Using the sequence data, 28 more substitutional events are required for tree E than for the most parsimonious tree (A in Fig. 11). On the basis of cleavage maps for whole mtDNA, 4 more events were required for tree E than for tree A (Ferris et al. 1981a).

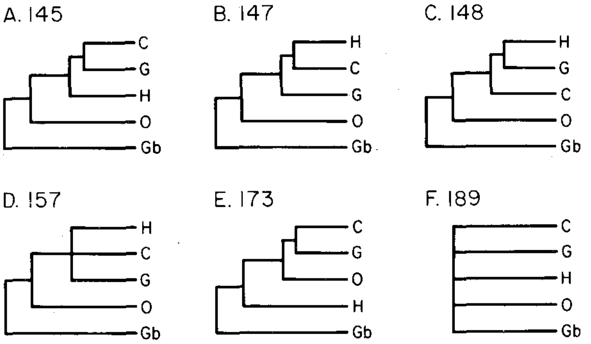


Fig. 11. Possible evolutionary trees for humans and apes. Six possible phylogenetic relationships (A-F) are shown among the five higher primates considered here. The figure indicates for each tree the minimum number of events required to produce that tree by a parsimony analysis of the sequence data for the 90 phylogenetically informative sites within the 896 bp fragment of mtDNA. The *abbreviations* used are H, human; C, chimpanzee; G, gorilla; O, orangutan; Gb, gibbon. Trees analogous to A-E were constructed also by subjecting the intraprimate data in Table 1 to phylogenetic analysis by the matrix methods of Fitch and Margoliash (1967) and Farris (1972); both methods favored trees A and B and indicated tree E was far less likely than A-D

The preponderance of transitions helps us to understand why it has been hard to establish the branching order for hominoid lineages by mtDNA comparisons. A high incidence of transitions inevitably produces parallel and back mutations at the same site among lineages. The problem of parallelisms and reversals was recognized in an earlier restriction mapping study (Ferris et al. 1981a), but its molecular basis has only now become apparent.

Acknowledgements. We thank those who provided purified samples, tissues, and reagents used in this investigation; J. Drouin, S. Anderson, and D.A. Clayton for information on the complete human, cow, and mouse mtDNA sequences, respectively, prior to publication; J. Kent for assistance with computer analyses; J. Adams, B.N. Ames, S.M. Beverley, S.S. Brown, R.L. Cann, S.M. Carr, B.S. Chapman, W.S. Davidson, L. Ernster, S.D. Ferris, J. Gallant, M. George, B. Greenberg, K.M. Helm-Bychowski, C. Higgins, R.G. Higuchi, J. Hixson, R. Holmquist, T.H. Jukes, A.G. Kluge, S.W. Krauss, S.L. Martin, R. Myers, R.M. Watson, R.T. White, D.A. Wilson, B. Wu, and E.A. Zimmer for helpful advice and discussion; and S. Schaerer for assistance in preparing the manuscript. This work was supported by grants from the National Science Foundation.

Appendix

Correction for Multiple Substitutions. It was necessary to revise the equation traditionally used (Jukes and Cantor 1969) to correct for multiple substitutions at the same nucleotide site, namely

$$\Lambda = -\frac{3}{4} \ln (1 - \frac{4}{3}\lambda) \quad (1),$$

where λ is the observed fraction of the sites at which a nucleotide sequence differs from another sequence of the same length, and Λ is the inferred number of substitutions per site. Equation 1 rests on the assumption that all base interchanges are equally likely, which implies that transversions are twice as frequent as transitions. For mtDNA, Prager and Wilson revised this equation to reflect the finding that transitions outnumber transversions by a factor of about 9 to 1.

The revised method of calculation employs two kinds of λ, one being the fraction of sites that differ by transitions (λ_i) and the other being the fraction that differ by transversions (λ_v). Thus we use two equations

$$\Lambda_i = -\frac{i+v}{i} [\frac{1}{2} \ln (1 - 2\lambda_i)] \quad (2)$$

$$\Lambda_v = -\frac{i+v}{v} [\frac{2}{3} \ln (1 - \frac{3}{2}\lambda_v)] \quad (3)$$

which give two estimates, Λ_i and Λ_v, of the number of substitutions per nucleotide site. [Equation 2 is based on equation 26 of Holmquist (1972).] It is assumed that transitions account for a constant fraction, *i*, of the inferred substitutions and that transversions account for the remaining fraction, *v*. For the 896-bp segment of mtDNA we studied, *i* = 0.9 and *v* = 0.1, so that equations 2 and 3 reduce to

Table 5. Silent and replacement sites in the genetic code for mammalian mitochondrial DNA

Amino acid (and codons)[a]	Position in codon	Number of potential changes			
		Silent sites		Replacement sites	
		Category *i*, transitions	Category *v*, transversions	Category *i*, transitions	Category *v*, transversions
Asp, Asn, Cys, Glu, Gln,	First	0	0	1	2
His, Ile, Lys, Met, Phe,	Second	0	0	1	2
Ser(AGY), Trp, Tyr	Third	1	0	0	2
Ala, Arg, Gly, Leu(CTY),	First	0	0	1	2
Pro, Ser(TCX), Thr, Val	Second	0	0	1	2
	Third	1	2	0	0
Leu(CTR)	First	1	0	0	2
	Second	0	0	1	2
	Third	1	2	0	0
Leu(TTR)	First	1	0	0	2
	Second	0	0	1	2
	Third	1	0	0	2

[a] Codons are written in those cases where different codons for the same amino acid lead to a different classification of potential changes. Y is a pyrimidine, R is a purine, and X is any base

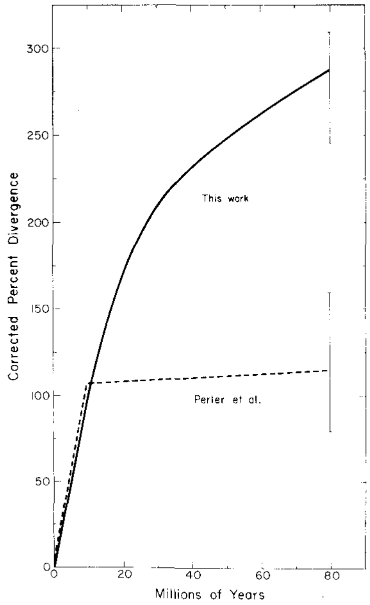


Fig. 12. Comparison of results of two methods for calculating corrected divergence at silent sites. The nucleotide sequences compared are 231 residues of the mitochondrial gene coding for Protein 5 of human, chimpanzee, gorilla, orangutan, gibbon, and cow. The heavy line indicates results computed by our method, and the dashed line indicates results computed by the Perler et al. (1980) method. The thin vertical lines at 80 million years indicate the range of values obtained for the 5 individual primates compared with the cow. The initial slopes are based on the primate comparisons (0–10 million years of divergence, Table 3 and Fig. 7). In this time range, the Perler et al. (1980) method gives results similar to those of our method. For large divergence times the two methods give contrasting results. A similar picture emerged when the two methods were applied to silent sites in the gene coding for Protein 4 and to replacement sites in the two genes. To explain why with our method the relationship between corrected divergence and time becomes nonlinear at times approaching 80 million years, we draw attention to an assumption of both methods, namely that all of the sites in a given category (see Table 5) are equally able to accept a nucleotide substitution. To the extent that this assumption is unlikely to be valid, both methods will underestimate the incidence of multiple substitutions for long divergence times (Holmquist and Pearl 1980)

$$\Lambda_i = -\frac{5}{9}\ln(1-2\lambda_i) \quad (4)$$

$$\Lambda_v = -\frac{20}{3}\ln(1-\frac{3}{2}\lambda_v) \quad (5).$$

As the best estimate of Λ we recommend the weighted average of Λ_i and Λ_v, given by equation 6.

$$\Lambda = \frac{1}{2}[\frac{1}{2}(\Lambda_i + \Lambda_v) + (\Lambda_i\delta_i + \Lambda_v\delta_v)(\delta_i + \delta_v)^{-1}] \quad (6)$$

where δ_i is the number of nucleotide sites observed to differ by transitions and δ_v is the corresponding number for transversions.

This approach was used to compute corrected divergence for tRNA genes. 189 of the 199 sites encoding the three tRNAs were analyzed collectively; the 9 sites comprising the anticodons were considered invariant, and the one site at which a deletion occurs in the orangutan was also omitted.

Corrected Divergence at Silent and Replacement Sites. Multiple-hit corrections were also made according to the above principle for sequences in which there are two definable classes of sites differing as to the probability of evolutionary substitution, e.g., silent and replacement sites in genes coding for proteins. Prager and Wilson modified the approach introduced by Perler et al. (1980) to take account of the fact that mtDNA evolves mainly by way of transitions and that the genetic code for mammalian mtDNA differs from that for nuclear DNA (Barrell et al. 1980; Anderson et al. 1981).

Table 5 lists the number of silent and replacement substitutions that each position in every codon can undergo. To compute the corrected divergence at silent sites, for example, one first counts the number of sites in silent category *i* in each of the two sequences being compared and takes the average. The next step is to sum the observed number of silent substitutions by which the two sequences differ at these sites and to divide this sum by the average number of sites in silent category *i*, thereby obtaining a λ value, λ_{is}. Similarly, one obtains a λ value termed λ_{vs} for sites in silent category *v*. (In some cases a substitution is classified as half-silent and half-replacement, as when one compares the third position of the serine codon TCA with that of the phenylalanine codon TTC. The substitution of C for A in the serine codon would be silent, but the substitution of A for C in the phenylalanine codon would produce a leucine codon.) Corrected divergence at silent sites is then calculated with equations 7 and 8,

$$\Lambda_{is} = -\frac{5}{9}\ln(1-2\lambda_{is}) \quad (7)$$

$$\Lambda_{vs} = -\frac{20}{3}\ln(1-\frac{3}{2}\lambda_{vs}) \quad (8)$$

which are analogous to equations 4 and 5. Weighted averages are calculated as described by Perler et al. (1980): the average number of sites (*n*) and the number of observed differences (δ) in each category are used separately as weighting factors and the arithmetic mean of the two weighted averages is taken, as specified in equation 9.

$$\Lambda_s = \frac{1}{2}[(\Lambda_{is}n_{is} + \Lambda_{vs}n_{vs})(n_{is} + n_{vs})^{-1} + (\Lambda_{is}\delta_{is} + \Lambda_{vs}\delta_{vs})(\delta_{is} + \delta_{vs})^{-1}] \quad (9)$$

By an analogous method one calculates Λ_r, the corrected divergence at replacement sites.

Comparison of Correction Methods. Figure 12 shows how the results of our method of calculation compare with those obtained by the Perler et al. (1980) method. When the observed extent of sequence difference is small, the two methods give similar values for the corrected divergence. But when the observed sequence difference is large, the Perler et al. (1980) method gives a far lower value than our method does. Further, whereas the Perler et al. (1980) approach leads to curiously complex kinetics for divergence, our method provides a more reasonable picture of divergence with time (see Fig. 12).

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) The complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J Mol Biol (in press)

Attardi G, Cantatore P, Ching E, Crews S, Gelfand R, Merkel C, Montoya J, Ojala D (1980) The remarkable features of gene organization and expression of human mitochondrial DNA. In: Kroon AM, Saccone C (eds) The organization and expression of the mitochondrial genome. Elsevier/North Holland Biomedical Press, Amsterdam, pp 103–119

Barrell BG, Anderson S, Bankier AT, de Bruijn MHL, Chen E, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1980) Different patterns of codon recognition by mammalian mitochondrial tRNAs. Proc Natl Acad Sci USA 77: 3164–3166

Barrie PA, Jeffreys AJ, Scott AF (1981) Evolution of the β-globin gene cluster in man and the primates. J Mol Biol 149:319–336

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 (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc Natl Acad Sci USA 77:3605–3609

Brown WM (1981) Mechanisms of evolution of animal mitochondrial DNA. Annals NY Acad Sci 361:119–134

Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76: 1967–1971

Brown WM, Vinograd J (1974) Restriction endonuclease cleavage maps of animal mitochondrial DNAs. Proc Natl Acad Sci USA 71:4617–4621

Brues AM (1977) People and races. Macmillan, New York, pp 1–336

Cann RL, Brown WM, Wilson AC (1982) Evolution of human mitochondrial DNA: Molecular, genetic and anthropological implications. Proc Sixth Internat Congress Human Genetics, Vol I, in press

Castora FJ, Arnheim N, Simpson MV (1980) Mitochondrial DNA polymorphism: Evidence that variants detected by restriction enzymes differ in nucleotide sequence rather than in methylation. Proc Natl Acad Sci USA 77:6415–6419

Cedergren RJ, Sankoff D, LaRue B, Grosjean H (1981) The evolving tRNA molecule. CRC Crit Rev Biochem 11:35–103

Clemmey H (1976) World's oldest animal traces. Nature 261: 576–578

Cocks GT, Wilson AC (1972) Enzyme evolution in the Enterobacteriaceae. J Bacteriol 110:793–802

Cordell B, Bell G, Tischer E, DeNoto FM, Ullrich A, Pictet A, Rutter WJ, Goodman HM (1979) Isolation and characterization of a rat insulin gene. Cell 18:533–543

Dayhoff MO (1973) Atlas of protein sequence and structure, Vol 5, Supp I. Nat Biomed Res Found, Georgetown Univ Med Center, Wash DC, p S–101

Dayhoff MO (1976) Atlas of protein sequence and structure, Vol 5, Supp 2. Nat Biomed Res Found, Georgetown Univ Med Center, Wash DC, pp 283–284

de Bruijn MHL, Schreier PH, Eperon IC, Barrell BG, Chen EY, Armstrong PW, Wong JFH, Roe BA (1980) A mammalian mitochondrial serine transfer RNA lacking the "dihydrouridine" loop and stem. Nucleic Acids Res 8:5213–5222

Derancourt J, Lebor AS, Zuckerkandl E (1967) Séquence des acides aminés, séquence des nucléotides et évolution. Bull Soc Chim Biol 49:577–607

De Vos WM, Bakker H, Saccone C, Kroon AM (1980) Further analysis of the type differences of rat liver mitochondrial DNA. Biochim Biophys Acta 607:1–9

Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, DeRiel JK, Forget BG, Weissman SM, Slightom JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot NJ (1980) The structure and evolution of the human β-globin gene family. Cell 21: 653–668

Farris JS (1972) Estimating phylogenetic trees from distance matrices. Am Natur 106: 645–668

Ferris SD, Wilson AC, Brown WM (1981a) Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc Natl Acad Sci USA 78:2432–2436

Ferris SD, Brown WM, Davidson WS, Wilson AC (1981b) Extensive polymorphism in the mitochondrial DNA of apes. Proc Natl Acad Sci USA 78: 6319–6323

Fitch WM (1980) Estimating the total number of nucleotide substitutions since the common ancestor of a pair of homologous genes: Comparison of several methods and three beta hemoglobin messenger RNAs. J Mol Evol 16:153–209

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Freese E, Yoshida A (1965) The role of mutations in evolution. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 341–355

Goddard JM, Masters JN, Jones SS, Ashworth WD, Wolstenholme DR (1981) Nucleotide sequence variants of *Rattus norvegicus* mitochondrial DNA. Chromosoma 82:595–609

Hanahan D, Meselson M (1980) Plasmid screening at high colony density. Gene 10:63–67

Heckman JE, Sarnoff J, Alzner-DeWeerd B, Yin S, RajBhandary UL (1980) Novel features in the genetic code and codon reading patterns in *Neurospora crassa* mitochondria based on sequences of six mitochondrial tRNAs. Proc Natl Acad Sci USA 77:3159–3163

Holmquist R (1972) Theoretical foundations for a quantitative approach to paleogenetics. Part I: DNA. J Mol Evol 1:115–133

Holmquist R, Jukes TH, Moise H, Goodman M, Moore GW (1976) The evolution of the globin family genes: Concordance of stochastic and augmented maximum parsimony genetic distances for α hemoglobin, β hemoglobin and myoglobin phylogenies. J Mol Biol 105:39–74

Holmquist R, Pearl D (1980) Theoretical foundations for quantitative paleogenetics. Part III: The molecular divergence of nucleic acids and proteins for the case of genetic events of unequal probability. J Mol Evol 16:211–267

Jukes TH (1980) Silent nucleotide substitutions and the molecular evolutionary clock. Science 210:973–978

Jukes TH (1981) Amino acid codes in mitochondria as possible clues to primitive codes. J Mol Evol 18:15–17

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism, Vol III, Academic Press, New York, pp 21–132

Kimura M (1981) Possibility of extensive neutral evolution under stabilizing selection with special reference to nonrandom usage of synonymous codons. Proc Natl Acad Sci USA 78:5773–5777

Kluge AG (1982) Reclassification of the great apes. In: Ciochon RL, Corruccini RS (eds) New interpretations of ape and human ancestry. Plenum Press, New York, in press

Köchel HG, Lazarus CM, Basak N, Küntzel H (1981) Mitochondrial tRNA gene clusters in *Aspergillus nidulans*: Organization and nucleotide sequence. Cell 23:625–633

Martin NC, Miller D, Hartley J, Moynihan P, Donelson JE (1980) The $\mathrm{tRNA}^{\mathrm{Ser}}_{\mathrm{AGY}}$ and $\mathrm{tRNA}^{\mathrm{Arg}}_{\mathrm{CGY}}$ genes form a gene cluster in yeast mitochondrial DNA. Cell 19:339–343

Martin SL, Zimmer EA, Davidson WS, Wilson AC, Kan YW (1981) The untranslated regions of β-globin mRNA evolve at a functional rate in higher primates. Cell 25:737–741

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Meth Enzymology 65: 499–560

Nichols BP, Miozzari GF, Van Cleemput M, Bennett GN, Yanofsky C (1980) Nucleotide sequences of the trp G regions of *Escherichia coli, Shigella dysenteriae, Salmonella typhimurium* and *Serratia marcescens*. J Mol Biol 142:503–517

Perler F, Efstratiadis A, Lomedico P, Gilbert W, Kolodner R, Dodgson J (1980) The evolution of genes: The chicken preproinsulin gene. Cell 20: 555–566

Pilbeam D (1979) Recent finds and interpretations of Miocene hominoids. Ann Rev Anthrop 8:333–352

Romer AS (1966) Vertebrate paleontology. Univ of Chicago, Chicago, pp 1–468

Saccone C, Cantatore P, Gadaleta G, Gallerani R, Lanave C, Pepe G, Kroon AM (1981) The nucleotide sequence of the large ribosomal RNA gene and the adjacent tRNA genes from rat mitochondria. Nucleic Acids Res 9:4139–4148

Sarich VM, Wilson AC (1967) Immunological time scale for hominid evolution. Science 158:1200–1203

Singer CE, Smith GR (1972) Histidine regulation in *Salmonella typhimurium*. XIII. Nucleotide sequence of histidine transfer ribonucleic acid. J Biol Chem 247:2989–3000

Sinha NK, Haimes MD (1981) Molecular mechanisms of substitution mutagenesis, J Biol Chem 256:10671–10683

Smith HO (1980) Recovery of DNA from gels. Meth Enzymology 65:371–380

Sprinzl M, Grueter F, Spelzhaus A, Gauss DH (1980) Compilation of tRNA sequences. Nucleic Acids Res 8:r1–r22

Staden R (1980) A computer program to search for tRNA genes. Nucleic Acids Res 8:817–825

Steel RGD, Torrie JH (1960) Principles and procedures of statistics–with special reference to the biological sciences. McGraw-Hill, New York, pp 1–481

Topal MD, Fresco JR (1976) Complementary base pairing and the origin of substitution mutations. Nature 263:285–289

Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411–5421

Wilson AC, Carlson SS, White TJ (1977) Biochemical evolution. Annu Rev Biochem 46:573–639

Zimmer EA (1980) Evolution of primate globin genes. PhD Thesis, Univ of California, Berkeley, pp 1–366

Note Added in Proof

G.G. Brown and M.V. Simpson inform us that they too have discovered a high incidence of transitions and silent substitutions in mtDNA evolution. An account of their work, which is based on sequencing of a cloned segment of mtDNA from two species of rat, will appear in Proc Natl Acad Sci USA 79 (1982).

Received January 20/Revised February 5, 1982