Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis

David R.Foran, James E.Hixson[+] and Wesley M.Brown*

Department of Biology, University of Michigan, Ann Arbor, MI 48109-1048, USA

ABSTRACT

The mitochondrial DNA (mtDNA) control regions for common chimpanzee, pygmy chimpanzee and gorilla were sequenced and the lengths and termini of their D-loop DNA's characterized. In these and all other species for which there are data, 5' termini map to sequences that contain the trinucleotide YAY. 3' termini are 25-51 nucleotides downstream from a sequence that is moderately conserved among vertebrates. Substitutions were >1.5 times more frequent in the control region than in regions encoding structural genes. Additions and deletions were also frequent, especially in gorilla. Sequences of promoters and of two of four transcription factor binding sites were highly conserved. Comparisons of sequence similarity and transition/ transversion ratios suggest that human and chimpanzees may be more closely related to each other than either is to gorilla, if substitution rates are approximately equal among these species.

INTRODUCTION

The control region of vertebrate mtDNA, located between the genes for tRNAPro and tRNAPhe, contains sites that regulate mtDNA replication and transcription (reviewed in 1-5). A three-stranded DNA structure, the D-loop, is formed within this region by displacement synthesis of a short DNA strand complementary to the L-strand of mtDNA. This short strand, called D-loop DNA, ranges in size from ca. 500 to 1700 nucleotides (nt) among ver-tebrate species, and two or more discrete size classes are usually occur within individuals (6-12). Direct evidence for a precursor-product relationship between D-loop DNA and newly replicated H-strand DNA is lacking; however, the 5' termini of D-loop DNA's and the 5' termini of replicating H-strands are identical (13). The same sequences are probably responsible for initiating both D-loop DNA synthesis and mtDNA replication.

Control region sequences are known for several vertebrate

mtDNA's, and relatively conserved sequence elements within and
flanking the D-loop have been identified (8,12,14-20). Some of
these elements appear to function in D-loop DNA synthesis, but
the functions of others are unknown. Six other elements, with
less conserved primary sequences, appear to function in the
transcription of human mtDNA (1,5,21). Two are promoters for
the mtDNA H- and L-strands. There is disagreement about the
exact locations of these (22,23). The remaining four are in
vitro binding sites for a mitochondrial transcription factor
(21); however, two are physically removed from either promoter
and are of questionable function.

A number of elements upstream of the 5' termini of D-loop
DNA's have been proposed to be important in initiating mtDNA
replication (4,19,24-26), but functional correlations exist for
only two. The L-strand promoter appears to prime D-loop DNA
synthesis (24,26), and an endoribonuclease cleavage site may be
important in switching synthesis from RNA to DNA (25).

A sequence element 25 to 65 nt upstream (5') of the 3'
termini of D-loop DNA's may be a signal for the termination of
their synthesis (12,27). The variability of the position of
this termination associated sequence relative to the 3' termini
suggests that additional signals are needed, but no candidate
sequences have been identified. The sequence of this element is
highly variable among mammalian species. Given this, it is
possible that other regulatory sequences could go unrecognized
except in comparisons between very closely related taxa. The
mitochondrial promoters appear to be examples of such sequences
(10,22,26,28,29). In order to identify such labile elements, we
determined the sequences of mtDNA control regions and character-
ized D-loop DNA's from three species of hominoid primates: pygmy
chimpanzee, common chimpanzee, and lowland gorilla. These and
the sequences published for human (14,19,27,30,31), provide
comparisons among four closely related hominoid species.

MATERIALS AND METHODS
mtDNA Preparation, Cloning and Sequence Analysis
The preparation of mtDNA's from pygmy chimpanzee (Pan
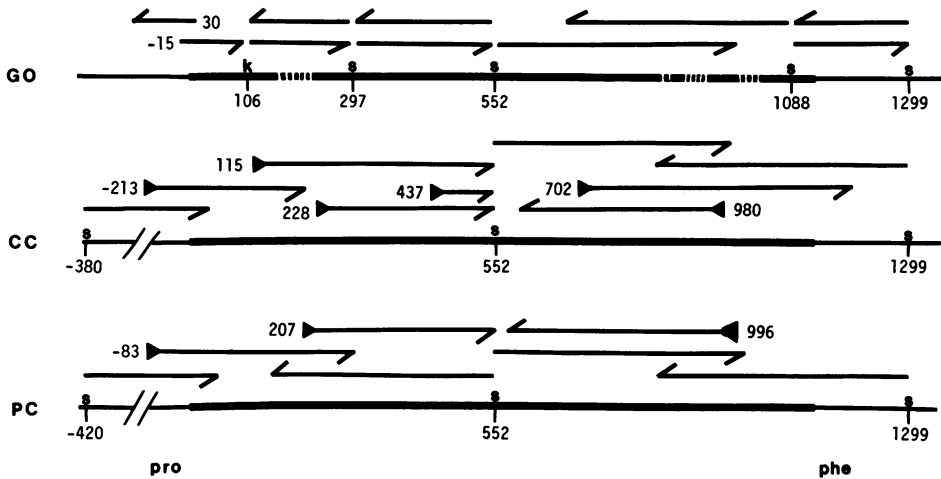paniscus), common chimpanzee (Pan troglodytes), lowland gorilla

Figure 1. Sequencing strategies for the control regions of
gorilla (GO), common chimpanzee (CC) and pygmy chimpanzee (PC)
mtDNA's. The control regions (heavy lines) are flanked by the
tRNA genes (light lines) for proline (pro) and phenylalanine
(phe). Lengths and directions of sequencing runs are indicated
by arrows. s and k denote, respectively, SauIIIA and KpnI
restriction sites. Arrow shafts with triangular ends denote
deletion clones; associated numbers indicate the position of the
sequencing startpoints for each of these. Small gaps shown in
the gorilla control region approximate sites of insertions or
deletions. Positions are numbered in agreement with Fig. 2.

(Gorilla gorilla) and human (Homo sapiens), and the cloning of

KpnI and EcoRI/AvaI fragments that contain the control regions

and rRNA genes have been described (32-34). The cloned mtDNA

fragments were cleaved with SauIIIA and subcloned into the BamHI

site of M13 vectors for sequence determination (34). Sequencing

of large fragments was accomplished by construction of over-

lapping deletion clones (35). Sequencing strategies used are

summarized in Fig. 1. Sequencing was by the chain termination

method (36,37), except that the gorilla sequence from a position

within the tRNAPhe gene through position 30 of Fig. 1 was

first obtained using the chemical method (38). Based on this, a

primer homologous to positions 16009-16023 of the human sequence

(14) was synthesized and used in obtaining the gorilla sequence

from positions 1 through 105, Fig. 1, by the chain termination

method (37).

```
            10        20        30        40        50        60        70        80        90       100
HO TTCTTTCATG GGGAAGCAGA TTTGGGTACC ACCCAAGTAT TGACTCACCC ATCAACAACC GCTATGTATT TCGTACATTA CTGCCAGCCA CCATGAATAT
GO           GA  A      A          T          --         AG A       -T TT ATC   G        G           A               A
CC           A          A          T          C          G  TT      T -
PC           A          AA  G                 G          TT         CT -T

           110       120       130       140       150       160       170       180       190       200
HO TGTACGGTAC CATAAA-TAC TTGACCACCT GTAGTACATA AAAACCCA-A TCCACATCAA AACCCCCTCC CCATGCTTAC AAGCAAGTAC AGCAATCAAC
GO            ACG CCA T         C    [          ---------- acaacccccc ccttcccccc cccccgcac tg-------- ----------
CC C     A   -T C      CCA  T    A          -        T CT C            TT A T              C C        A
PC AC TA     T    T C T    A          A  AC         T C            T              C         A T        C A        AT  G

           210       220       230       240       250       260       270       280       290       300
HO CCTCAACTAT CACACATCAA CTGCAACTCC AAAGCCACCC CTCACCCACT AGGATACCAA CAAACCTACC CACCCTTAAC AGTACATAGT ACATAAAGCC
GO ---------- ---------- ---------- ---------- ---------- ---]       AC G A       T   C T  CT ACA  A     AC       GAT
CC T C       G   GA     A   ACA  T     C A         C       C CC       G      T  TC     G       A              C A
PC  TC       G   GA     -   ACA CC C     A   T   C C       C CC                GA  AGT                 C         C ATT

           310       320       330       340       350       360       370       380       390       400
HO ATTTACCGTA CATAGCACAT TACAGTCAAA TCCCTTCTCG TCCCCATGGA TGACCCCCCT CAGATAGGGG TCCCTTGACC ACCATCCTCC GTGAAATCAA
GO      T  C             CC   T     -ACC         C         C          G  AA      A
CC    AC                           C   C       C   C     CT          AA         GT
PC    A                            A C         C   C      C          AA         G

           410       420       430       440       450       460       470       480       490       500
HO TATCCCGCAC AAGAGTGCTA CTCTCCTCGC TCCGGGCCCA TAACACTTGG GGGTAGCTAA AGTGAACTGT ATCCGACATC TGGTTCCTAC TTCAGGGTCA
GO                     -
CC                     --                      TC                                          G             C         C
PC                     -                                            C                                    C         C

           510       520       530       540       550       560       570       580       590       600
HO TAAAGCCTAA ATAGC--CCA CACGTTCCCC TTAAATAAGA CAT-CACGAT GGATCACAGG TCTATCACCC TATTAACCAC TCACGGGAGC TCTCCATGCA
GO           CA       GC TT-
CC G    TTC  AGA  TC                                      G    T                           G             CT
PC G    TTC  GGA  TC                            G    T                                                   G

           610       620       630       640       650       660       670       680       690       700
HO TTTGGTATT- TTCGTCTGGG GGGTATGCAC GCGATAGCAT TGCGAGACGC TGGAGCCGGA GCACCCTATG TCGCAGTATC TGTCTTTGAT TCCTGCCTCA
GO          T          -         G                  ---       --           A           AC                       C
CC                               G                         A           CC                                       C
PC                               G                         A           CC                                       C

                                                                          CSB-1
                                                                     ┌─────────────────┐
           710       720       730       740       750       760       770       780       790       800
HO TCCTATTATT TATCGCACCT ACGTTCAATA TTACAGGCGA ACAT-ACTTA CTAAAGTGTG TTAATTAATT AATGCTTGTA GGACATAATA ATAA-CAATT
GO G   CC                            C          G GC       AGT G TC TG            C          T         -    C C  GG
CC    TG                            GAC T  G    G     C                G                     C          C C G  G-C
PC    TACG                          TTAC T  G   G T              C                           C          G      G-C

                                                          CSB-2
                                                     ┌────────────────┐
           810       820       830       840       850       860       870       880       890       900
HO GA-ATGTCTG CACAGCCACT TTCCACACAG ACATCATAAC AAAAAATTTC CACCCAAACCC CCCCTCCCCC GCTTCTGGCC ACAGCACTTA AACACATCTC
GO G  C  GAA      A     C C T      CA  AA  C    -          --        ----         --         --------   -----    AG    A   G C
CC A A    CTCA   T A  TG          CA                 C         -          T       ---- C             C    A
PC A A    AC-CA  GT A  TG          CA                 G                            A  C      T        C    TT

      CSB-3
   ┌─────────────┐
           910       920       930       940       950       960       970       980       990      1000
HO TGCCAAACCC CAAAAACAAA GAACCCTAAC ACCAGCCTAA CCAGATTTCA AATTTTATCT TTTGGCGGTA TGCACTTTTA ACAGTCACCC CCCAACTAAC
GO CA       T ---------- ---------- ---------- ---------      C                                              T
CC             AG  G       G         C         C           C        A                                        T T
PC             AG  T                 CC        C                              T                              T

          1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
HO ACATTATTTT CCCCTCCCAC TCCCATACTA CTAATCTCAT CAATACAACC CCC-GCCCAT CCTACCCAGC ACACACACAC CGCTGCTAAC CCCATACCCC
GO T G[----- --cagcccac ---------- ----------  ] G        -    GC    ----    -          T     G T  T
CC TGCCC CCC   -  A-         T      GC C  G    CGT        TA T  C    T A- -    T T                          T
PC TG CC CCC   -  A-         T C    GC C  A    C T        T      C   C T  -- -  T T                  T      T

          1110      1120      1130
HO GAACCAACCA AACCCCAAAG ACACCCCC-C ACA
GO    TT                    -   A
CC                         TA
PC A  G                 T T    A
```

D-loop DNA End-labelling, Isolation and Analysis

D-loop DNA preparation and analysis was modified from (6). The mtDNA samples used for sequence determinations were also used for preliminary D-loop DNA analyses. Additional mtDNA samples were obtained from human, gorilla and common chimpanzee placentas and from gorilla and pygmy chimpanzee livers. D-loop DNA's from 6 humans and 2 individuals from each of the other species were analyzed. 30-100 ng of each mtDNA sample was heated at 80 C for 5 min. to dissociate D-loop DNA, then treated with calf-intestinal alkaline phosphatase. 5' ends were labeled using [γ-32P]ATP and polynucleotide kinase (39), and 3' ends using [α-32P]ddATP and terminal transferase (40).

End-labeled D-loop DNA's were electrophoretically separated in denaturing (7M urea) 4% polyacrylamide gels. Bands corresponding to D-loop DNA's were excised from the gel and extracted as described in (38). After centrifugation for 15 min at 13,000 g, the supernatants were removed. The pellets were checked for radioactivity and re-extracted, if necessary, until >80% of the label was recovered. Following two cycles of ethanol precipitation, a portion of each sample was analyzed electrophoretically to verify that the isolated DNA's were identical in mobility to the D-loop DNA's in the original samples.

Equal amounts of the D-loop DNA samples, based on radioactivity, were digested with a large excess of HaeIII, ethanol precipitated, resuspended, and analyzed by electrophoresis in denaturing 8% polyacrylamide gels. To ensure the generality of the results, these experiments were repeated using mtDNA's isolated from different tissues (liver, placenta) and individuals.

Figure 2. Mitochondrial DNA control region sequences for four hominoids. In the orientation shown, the tRNA gene for proline precedes position 1 and that for phenylalinine follows position 1133. The published human (HO) sequence (14) is shown at the top of each set of four lines. Its numbering has been offset from that in (14); thus, human position 1 corresponds to 16024 of (14). The gorilla (GO), common chimpanzee (CC) and pygmy chimpanzee (PC) sequences are aligned below it, but only differences from the human sequence are shown. A minus (-) indicates that no nucleotide is present. Lower case letters in the gorilla sequence denote nucleotides whose alignment cannot be determined; their placement at positions 151-182 and 1013-1020 is wholly arbitrary. Brackets labeled CSB-1 through CSB-3 denote the conserved sequence blocks defined in (19).

HaeIII digests of ΦX174 viral DNA (36) and HinfI or HaeII
digested, heat-denatured pBR322 DNA (39) were used as single
stranded DNA size standards. The ΦX174 digests reproducibly
contained two extra bands, of 504 and 544 nt, which presumably
resulted from incomplete digestion of the HaeIII sites between
the 310 nt fragment and the adjacent 194 and 234 nt fragments.

RESULTS
Control Region Sequences
    The aligned control region sequences of common chimpanzee,
pygmy chimpanzee and gorilla mtDNA's are presented in Fig. 2,
along with the human sequence from (14). The control regions of
common and pygmy chimpanzee and human are similar in sequence
organization and size (1122, 1113 and 1121 nt, respectively),
although numerous base changes and small (1-4 nt) deletions and
additions have occurred (Fig. 2; Table 1). The gorilla control
region is smaller (982 nt), exhibiting sixteen small (1-4 nt)
and at least four large (13-73 nt) unique deletions, but only
three unique (1 nt) additions (Fig. 2). In gorilla, sequences
of 32 and 8 nt are present in two of the large deleted regions
(respectively, positions 139 to 244 and 1005 to 1041, Fig. 2),
but it was impossible to determine the alignment of these with
the other species. Deletions in gorilla have also occurred in
two conserved sequence blocks, CSB-2 and CSB-3, Fig. 2.
D-loop DNA Sizes and Relative Abundance
    Sizes of D-loop DNA's were estimated by denaturing poly-
acrylamide gel electrophoresis. Each mtDNA sample contained two
to four size classes of D-loop DNA's, the relative abundance of
which varied between species, individuals and tissues (compare
panels A and B of Fig. 3). Three size classes, of 574, 610 and
652 nt, were present in all human samples (lanes 1, Fig. 3A and
3B), and some contained a fourth, of 630 nt. These four size
classes were also found in some common chimpanzee samples (lane
2, Fig. 3A), while in others only the three largest size classes
were present (lane 2, Fig. 3B). Pygmy chimpanzee samples
contained either the two smallest or three largest size classes
(lanes 3, Fig. 3A and 3B). Gorilla samples yielded three size
classes, of 555, 565 and 590 nt, which were smaller than those
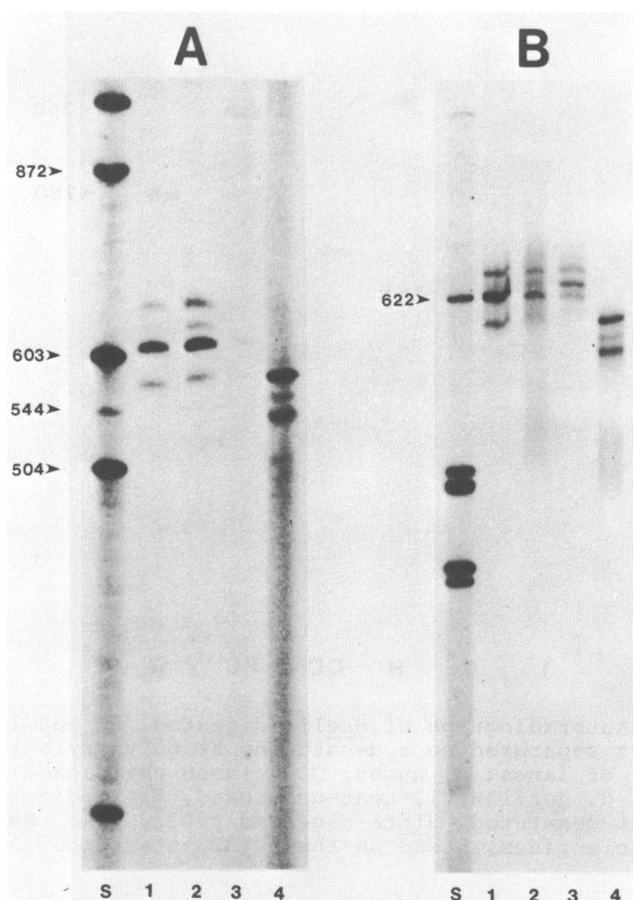
Figure 3. Autoradiograph of full-length, end-labeled D-loop DNA's from two different sets of individuals and separated on denaturing 4% polyacrylamide gels. DNA content of lanes for A and B: 1, human; 2, common chimpanzee; 3, pygmy chimpanzee; 4, gorilla; S, size standards (in A, single-stranded, HaeIII-digested ◆X-174 DNA; in B, heat-denatured, HaeII-digested pBR322 DNA). Numbers are sizes in nucleotides. D-loop DNA's from the gel displayed in B were eluted and used to map 5' and 3' termini. A samples: human, #WMB151; chimpanzees and gorilla as given in (34). B samples: human, #WMB153; common chimpanzee, PTR/KB#5007"A"; pygmy chimpanzee, SD#21327; gorilla, #YN85-105.

found among the other species (lanes 4, Fig. 3A and 3B).

Positions of the 3' Termini

The 3' termini of the D-loop DNA's were determined relative to the HaeIII site that begins at position 435 of Fig. 2 in all
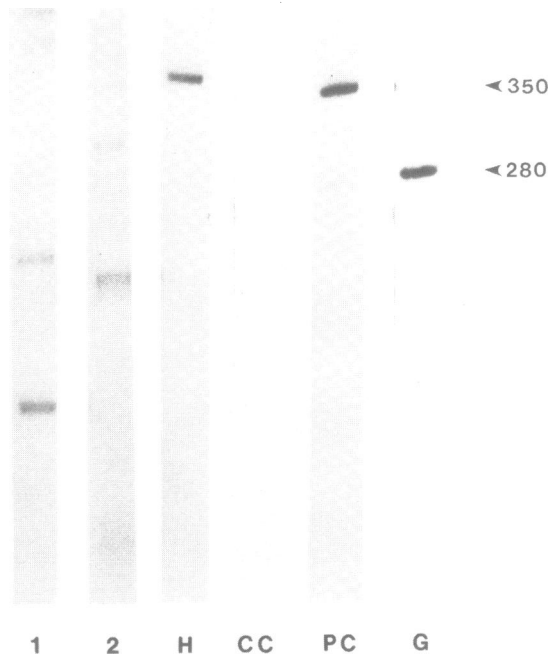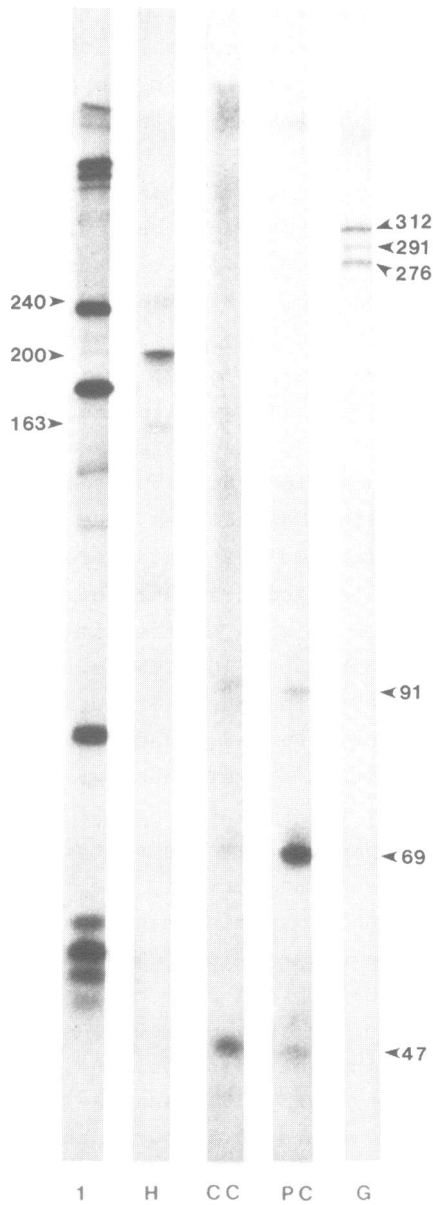
Figure 4. Autoradiograph of HaeIII-digested, 3' end-labeled
D-loop DNA's separated on a denaturing 8% polyacrylamide gel.
DNA content of lanes: H, human; CC, common chimpanzee; PC, pygmy
chimpanzee; G, gorilla; 1, heat-denatured, HaeII-digested pBR322
DNA; 2, heat-denatured, HinfI-digested pBR322 DNA. Numbers are
sizes in nucleotides, based on the pBR322 standards (39,p.497).

four species. 3' end-labeled D-loop DNA's were HaeIII digested
and the sizes of the labeled fragments were determined. Those
from common chimpanzee, pygmy chimpanzee and human migrated
identically. In each digest, a fragment of ca. 350 nt was
observed, Fig. 4. A second, less intense band corresponding to
a fragment of ca. 410 nt was often present in chimpanzee and
human digests and is probably a product of incomplete digestion,
as reported previously (6). These data indicate that D-loop
DNA's in common and pygmy chimpanzees and in human have one,
identically-located 3' terminus that maps to position 84 ($\pm$4 nt)
in Fig. 2. The labeled 3' end fragments in gorilla also migrate
as one band, but their size, ca. 280 nt, is 70 nt smaller than
in the other species. This size decrement agrees well with that

predicted from the sequence comparisons, in which the gorilla
control region lacks 73 nt between positions 139-244 relative to
the other species. The 3' termini of the gorilla D-loop DNA's
thus map to the same position as those of the other species.

Positions of the 5' Termini

The positions of the 5' termini of D-loop DNA's were also
determined relative to nearby HaeIII sites. Within the D-loop
DNA's, the location of the HaeIII site proximal to the 5' ends
differs among the hominoids. In human, the site is at positions
496-499 (Fig. 2). Although this HaeIII site is absent in the
human mtDNA sequence shown in Fig. 2, it is present in many
human samples (6,9,30,31). In gorilla, the 5'-proximal HaeIII
site is at positions 435-438, and in common and pygmy
chimpanzees it is at positions 652-655 (see Fig. 2).

5' end-labeled D-loop DNA samples were digested with HaeIII
and the sizes of the labeled fragments were determined. The
human digests yielded labeled fragments of 163, 200 and 240 nt,
Fig. 5. These differed in size by the same amounts as the cor-
responding full-length human D-loop DNA's, Fig. 3. From the Hae
III site at position 496, the human 5' termini map to positions
644, 701 and 741 ($\pm$3 nt), Fig. 2. HaeIII-digestion of common
and pygmy chimpanzee mtDNA's yielded major labeled fragments of
47, 69 and 91 nt, Fig. 5. From the HaeIII site at position 652,
two of these termini map to the same approximate positions (700
and 744) as two of the human termini, and the third maps to
position 722. Although 5' termini mapping to position 722 were
not observed among the human D-loop DNA samples used in this
analysis, the fourth D-loop DNA size class (630 nt) that was
seen in some human samples (see preceding section) would presum-
ably have its 5' terminus at this position. Similarly, if the
chimpanzee DNA's shown in Fig. 3 (lanes A-2 and A-3) had been
used, they would have generated a fourth band, of 12 nt, whose
5' terminus would be at position 664. A few faint bands differ-
ing in size by 1-3 nt from the major bands appear in human and
chimpanzee samples in Fig. 5. These could reflect size varia-
tion in the mtDNA control region within an individual, slight
variation in the exact position of the 5' terminus, or variation
due to degradation during the manipulation of the samples.

Figure 5. Autoradiograph of HaeIII-digested, 5' end-labeled D-loop DNA's separated on a denaturing 8% polyacrylamide gel. Lane contents as in Fig. 4. The sizes of the human D-loop DNA fragments are indicated on the left, those of the gorilla and the chimpanzees' on the right. The faint bands in the standard (lane 1) are undenatured HaeIII-digested pBR322 DNA fragments.

TABLE 1. Pairwise comparisons of control region sequences[a]

Species	Number of nt compared[b]	Percentage sequence difference	Numbers of		S/V
			Transitions (S)	Transversions (V)	
HO/GO	903	13.5	82	40	2.1
HO/CC	969	11.8	95	19	5.0
	(1107)	(13.4)	(116)	(32)	(3.6)
HO/PC	975	12.4	102	19	5.4
	(1113)	(14.0)	(124)	(32)	(3.9)
GO/CC	897	17.2	108	46	2.3
GO/PC	898	17.3	106	49	2.2
CC/PC	973	8.4	70	12	5.8
	(1111)	(9.6)	(87)	(19)	(4.6)

a. Species abbreviated as in Fig. 2. All numbers are from direct comparisons of the sequences shown in Fig. 2, and are uncorrected for multiple substitution events.
b. Positions 140-244 and 1005-1041 and all other additions and deletions have been excluded from the comparisons. The values for species pairs between which positions 140-244 and 1005-1041 (Fig. 2) can be included are shown in parentheses.

In gorilla D-loop DNA's, the 5' HaeIII fragments are of 276, 291 and 312 nt, Fig. 5. The 5' termini thus map to positions 719, 734 and 756 (+3 nt) when referenced to the HaeIII site beginning at position 435 in Fig. 2. The terminus estimated to be at position 720 may correspond to the chimpanzee terminus at 722, but the other termini are apparently unique to gorilla. Noting the numerous sequence differences between gorilla and the other species in this region (particularly in the intervals 700-710 and 735-760), this difference is not surprising.

DISCUSSION
Sequence Comparisons of the Hominoid Control Regions
    The control region contains sequences vital for transcription and replication. However, substitutions, additions and deletions accumulate more rapidly in this than in other mtDNA regions (2,3,41). This is due both to some apparently functionless DNA segments and to others whose primary sequence is unimportant (3). The divergence among hominoids, apparent in Fig.

2, is summarized in Table 1. Two sets of numbers are presented, based on inclusion or exclusion of the sequence intervals 140-244 and 1005-1041 (Fig.2). Most of the gorilla sequence in these intervals is deleted; the remainder cannot be meaningfully aligned. Accordingly, these intervals were not included in any comparison involving gorilla. When these two intervals are ex-cluded from all comparisons, the sequences of common and pygmy chimpanzee are most similar, human and chimpanzees' are next most similar, and gorilla is least similar (Table 1). If rela-tionships between species and sequence are proportional, then human and chimpanzees are closer to one another than either is to gorilla. This interpretation is supported by transition/transversion ratios (S/V, Table 1) which, among closely related species, are inversely proportional to the product of divergence time and substitution rate (2,3,42). This is the strongest support that mtDNA comparisons have provided in differentiating the order of evolutionary relatedness among human, chimpanzee and gorilla (see 34), but it is valid only if mtDNA substitution rates are approximately equal among these species.

In these comparisons, the S/V ratios in the control region are much lower than in regions containing protein, tRNA and rRNA genes (34,42). Given the higher substitution rate observed in the control region (30,41), an increased proportion of multiply-substituted positions and, thus, a greater proportion of transversions is expected (see 30 and 42 for details).

Extensive Length Difference in the Gorilla Control Region

A deletion in or adjacent to the control region of gorilla was inferred from cleavage map comparisons of hominoid mtDNA's (43). As the present study shows, there are actually a large number of deletions in gorilla (Fig. 2). The largest of these (i.e., those on the intervals 140-244 and 1005-1041 of Fig. 2) probably represent the fusion of many small deletions, rather than single events. This also appears to be the case for a portion of the control region sequence from African green monkey (44). Deletion sites in gorilla are often adjacent to homopolymer sequences, suggesting that they may have arisen by slipped-mispairing during mtDNA replication (34,45,46).

Promoter Sequences

The human mtDNA L- and H-strand promoters lie, respectively,
either at positions 991-1005 and 1091-1112 of Fig. 2 (23), or at
956-972 and 1101-1123 (22,47). Among the hominoids, 991-1005
and 956-972 differ by 8 and 2 substitutions, respectively, at 6
of 15 and 2 of 17 positions. Based on sequence conservation,
the data suggest that the L-strand promoters are on the interval
956-972. No such conservation-based discrimination is possible
for the H-strand promoters: the sequences 1091-1112 and 1101-
1123 differ by 7 and 5 substitutions at 6 of 22 and 5 of 23
positions, respectively. However, positions 1106-1121 are
identical in these species, and include the nucleotides
(1113-1115) identified by in vitro mutagenesis as being the most
important for H-strand transcription (47). The comparisons thus
support the promoter locations identified in (22).

Transcription Factor Binding Sequences

Four sites bind, in vitro, a partially purified mitochond-
rial transcription factor (mtTF) in the human control region
(21). Sites mtTF-L (positions 973-1000 in Fig. 2) and mtTF-H
(1081-1108) are immediately upstream of the respective promot-
ers; sites mtTF-X (831-858) and mtTF-Y (786-816) are between
conserved sequence blocks 1 and 2 (21). A consensus sequence
has been suggested (21) in which 7 positions are absolutely con-
served. However, 6 of the 7 are not conserved in the mtDNA's of
gorilla and the chimpanzees, and two are polymorphic in a sample
of 7 humans (30). Based on the alignments in Fig. 2, average
sequence similarities (mean$\pm$SD) for all pairwise comparisons of
orthologous mt-TF binding sites are: L=94$\pm$2%, H=85$\pm$7%, X=79$\pm$15%,
and Y=61$\pm$14%. Sites L and H show no deletions, and the SD's are
small. Sites X and Y show deletions and larger SD's. On
average, X is 94% similar among human and chimpanzees, but only
65% among these and gorilla. X has 1 nt deleted in common
chimpanzee, and 7 nt in gorilla. Y is located in one of the
most variable parts of the control region. Pairwise, this site
differs, on average, at 12 nt, including two deletions of 1 nt
each. These comparisons indicate that sites L and H probably
provide similar functions in chimpanzees, gorilla and human, and

that X and Y, if functional, are under more relaxed constraints
than L and H.

Sequences Associated with D-loop DNA Synthesis

There is presently no in vitro system for directly investi-
gating mammalian mtDNA replication and D-loop DNA formation, and
insights into these processes must come from comparative data.
The control region sequences from a number of distantly related
mammalian species are known, and their comparison has resulted
in identification of several conserved elements that could be
important for D-loop DNA initiation and termination (10,12,13,
15,19,27). Comparisons of control region sequences from pygmy
chimpanzee, common chimpanzee and gorilla add information about
some of the sequences necessary for these processes.

Little is known about how the 3' and 5' termini of D-loop
DNA's are formed. Published evidence indicates that D-loop DNA
synthesis is primed by an RNA whose transcription is initiated
at the L-strand promoter (24,26), and that the 5' termini of
D-loop DNA's are formed by subsequent processing events (25).
Although it is currently assumed that the 3' termini are formed
by synthesis termination, the possibility that these ends are
also processed has not been excluded.

Sizes for the HaeIII-fragments of the D-loop DNA's were
estimated from calibration curves obtained with precisely-known
size standards. Variability was generally $\leq$ 1% of mean fragment
length, based on repeated determinations and on comparisons of
the estimates for the human D-loop DNA fragments, whose precise
lengths are known from sequencing (19,25,27).

Sequences Near D-loop DNA 3' Termini. A moderately conser-
ved, termination associated sequence (TAS) has been identified
upstream from the 3' end of each D-loop DNA (12,27). Species
with multiple 3' termini have a matching number of TAS (12,27).
One TAS-like element was identified in mtDNA's of chimpanzees
and gorilla (positions 135-151, Fig. 6). As summarized in Fig.
7 for 37 vertebrate TAS, the consensus sequence appears to be:

TACATtAAAaYYYAAT

[Y = C or T; bases at positions shown in lower case are invar-
iant, but occur in <50% of the TAS]. A large amount of sequence
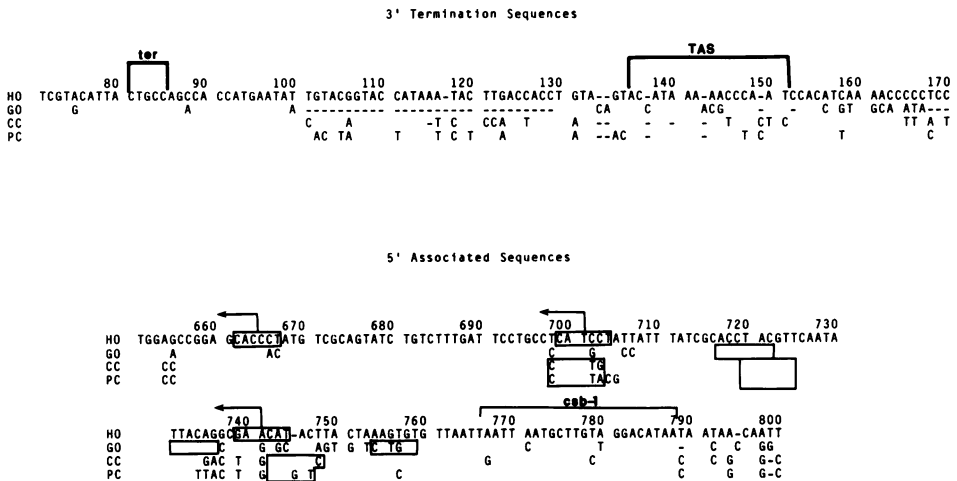
Figure 6. L-strand sequences in regions complementary to the 3' and 5' ends of the hominoid D-loop DNA's. Abbreviations and notation as in Fig. 2. Upper figure: The bracket labeled ter denotes the sequence to which 3' termini map; that labeled TAS is the termination-associated sequence. The portion of gorilla mtDNA from positions 101-140 in Fig. 2 has been renumbered as 130-167 and a 29 nt gap inserted between positions 100 and 130, above, in order to align the TAS. Lower figure: Boxed sequences indicate the sites to which 5' termini map. The arrows above the human sequence indicate the direction of D-loop DNA synthesis. csb-1, conserved sequence block 1 (19).

and length diversity is present among different individuals in the same species and among TAS that co-occur in the same mtDNA (see Fig. 7). The gorilla TAS is perhaps the most aberrant.

Positions of the 3' termini appear identically located among hominoids. Sequences associated with both the 3' and 5' termini are displayed in Fig. 6. In the two chimpanzee species the 3' termini are ca. 51 nt downstream from the TAS. This is the same as in human (27), in which the 3' terminus maps to positions 81-83 (CTG, Fig. 6). Gorilla exhibits both length and sequence divergence in this interval. Positions occupied by the TAS in human and the chimpanzees are mostly deleted in gorilla, which has, instead, a TAS sequence 25 nt further downstream [positions 108-124 of Fig. 2 (= 135-151 of Fig. 6)]. Thus, the 3' termini map to the same position, although in gorilla this position is

```
HUMAN(1)    TAC-AT-AAAAACCC--A-AT
  "  (2)    TAC-AT-AAAAACCC--A-AC
  "  (3)    TAC-AT-AAAAATCC--A-AC
  "  (4)    TAC-AT-AGAAACCC--A-AT
  "  (5)    TAC-AT-AAAAACCC--A-AT
  "  (6)    TAC-AT-AAAAACCC--A-AT
  "  (7)    TAC-AT-AAAAACCC--A-AT
CCHIMP      TAC-AT-AAAA-TCC--AC-T
PCHIMP      CAC-AT-AAAAACCT--ACAT
GORILLA     TACCAT-AAAA-CGCCCA-AT
MOUSE       CAC-ATTAAA--CT---A-AT
  "         AAC-AT-AAA--TC---A-AT
  "         TAC-ATTAAA--TC---A-AT
  "         TAC-ATTAAA--CT-----AT
RAT(RE)     TAC-ATTAA---TTT----AT
  "         TAT-AT-ATAA-TT---A-AT
  "         AACTA--AAAA-TTT--AAAC
  "         TAC-ATTAAGA-T----A-AT
RAT(NB)     TAC-ATTAAA--TT-----AT
  "         TAT-AT-ATA--TCT--A-AT
  "         AACTAT-AAA--TTC----AT
  "         TAC-ATTAGAA-T----A-AT
RAT(NC)     TAC-ATTAAA--TT-----AT
  "         TAT-AT-ATA--TTT--A-AT
  "         AACTAT-AAA--TTC----AT
  "         TAC-ATTAAAA-T----A-AT
RAT(NA)     TAC-ATTAAA--TT-----AT
  "         TAT-AT-ATA--TCT--A-AT
  "         AACTAT-AAA--TTC----AC
  "         TAC-ATTAGA--T----A-AT
PIG         TAC-AT-A----TT-----AT
  "         TAC-AT-A----TC-----AT
COW         TAC-AT-AA---C------AT
  "         AAC-AT-AA---T------AT
  "         TAC-ATTAAA--TT-----AT
XENOPUS     TACAAT-AAAA-C------AT
  "         TACAATTAAA--C------AT
```

CONSENSUS:

$$T_{78}A_{100}C_{89}A_{100}T_{97}A_{100}A_{75}A_{84}A_{43}Y_{100}Y_{81}Y_{51}A_{62}A_{97}T_{89}$$

Figure 7. Alignment of 36 termination associated sequences (TAS) from 11 species. Subscripts in the consensus sequence denote the percent frequency of occurrence of each nucleotide. HUMAN (#1-7) are TAS from 7 different individuals. CCHIMP and PCHIMP are common and pygmy chimpanzee, respectively; MOUSE is Mus domesticus, RAT(RE) is Rattus rattus, and RAT(NA,NB,NC) three different R. norvegicus. Some species have multiple TAS per mtDNA (mouse, rats = 4; pig = 2; cow = 3). Sequence sources: gorilla, common chimpanzee and pygmy chimpanzee, this study; human, (14,19,27,30,31; numbered as in Table 1 of ref. 30); mouse, (27); cow and pig, (12); rats, (16); Xenopus, (17).

only 25 nt from the TAS. Despite this change, the variation in distance from the TAS to the 3' termini (25-51 nt) is within the range reported for other mammalian species (12,27).

As an alternative hypothesis, termination might be directed by a sequence element at or directly adjacent to the termination site, like that found in the interval 76-87 of Fig. 2. Except for the TAS, however, there are no notable sequence similarities

in the regions to which the 3' termini of other mammalian D-loop
DNA's map. Moreover, in species with multiple 3' termini (e.g.,
mouse, rat, cow, Xenopus), the sequences within which the
individual termini map are different (12,17,27).

It has been suggested that computer-derived secondary struc-
tures flanking the D-loop might regulate its formation (8,16,
48). Based on analyses of sequences near the 3' termini in cow,
pig, mouse, human, and Xenopus, one group (8) hypothesized that
secondary structures containing the conserved sequence TACAT (at
positions 135-139 of the human sequence, Fig. 2) may play a role
in termination. Similar structures can be hypothesized for the
other hominoid sequences in this region, but with TACAT changed
to CACAT in gorilla and pygmy chimpanzee. However, among verte-
brate species these hypothetical structures and the 3' termini
are uncorrelated in either number or positional relationship
(the structures may be either 5' or 3' to the termini, depending
on the species chosen). Also, there is no basic similarity in
the form of the hypothetical structures between species.

Because the number of TAS and the number of 3' termini are
identical, and because the TAS are always upstream from the 3'
termini (albeit at a variable distance), the TAS hypothesis is
presently better supported than any other. Further sequence
comparisons from additional individuals and species will aid in
the refinement of the questions that must be addressed regarding
TAS, other (as yet unidentified) sequences, and the relevance of
secondary structures. However, it appears unlikely that such
comparisons will lead to a definitive understanding of D-loop
DNA termination. For this, functional assays must be developed.

Sequences Near D-loop DNA 5' Termini. In several vertebrate
species, three conserved sequence blocks (CSB's) are present up-
stream from the 5' termini (19,20). These CSB's are also in the
chimpanzees and gorilla, Fig. 2. There is evidence from human
and mouse mtDNA that a switch from RNA to DNA synthesis can
occur at CSB-1 (24,26). The average similarity among these
hominoids is 91±6% for CSB-1, and the 5' termini of the largest
D-loop DNA's are located just 3' to it (Fig. 6). However, the
data for the chimpanzees and gorilla (Figs. 5 and 6), as well as
those for human (6,9,24), rodents (18,19,26), and Xenopus (17)

indicate that the 5' termini of most D-loop DNA's originate
10-100 nt downstream from CSB-1 (Fig. 6). Thus, even if CSB-1
is the signal for the switch from RNA to DNA synthesis, most of
the 5' termini must be formed by subsequent processing events,
and additional sequence elements must be involved.

Among the hominoids, substitutions have occurred in all
three CSB's. Also, parts of CSB-2 and 35% of CSB-3 have been
deleted from the gorilla sequence. The reconstituted CSB-2
sequence in gorilla differs at only 2 of 17 positions from CSB-2
in the other hominoids, an amount of variation that is tolerated
between different individuals and species (19,30,49). No func-
tions are known for either CSB-2 or CSB-3, and part of CSB-2 and
all of CSB-3 have been deleted from cow mtDNA (15; also see Fig.
3 in ref. 3 and Fig. 11 in ref. 10). Although no conclusions
concerning CSB-2 may be drawn, the marked divergence of CSB-3
between gorilla and the other hominoids and the absence of CSB-3
from cow argue against its involvement in D-loop DNA initiation.

When the sequences to which the 5' termini of the hominoid
D-loop DNA's map are examined (Fig. 6), all are seen to contain
an adenine flanked on each side by a pyrimidine (YAY). This
correlation holds for the 5' termini of gorilla D-loop DNA's
that map to positions different from those of the other species
and also for the major 5' termini of other mitochondrial D-loop
DNA's [e.g., mouse (26), rat (18), and Xenopus (8,17)]. The
trinucleotide YAY also occurs at an analogous position in the
yeast mtDNA origin of replication, where it marks the point of
transition from RNA to DNA synthesis (50). In contrast, YAY is
not usually present at the sites of RNA to DNA transition in
bacterial origins of replication (51).

Without functional evidence there can be no rigorous conclu-
sion that YAY is required for 5' terminus formation. However,
for current data, the correlation between the positions of all
major 5' termini and YAY is perfect [Xenopus (8,17); rat (18);
mouse (19,26); cow (10,11; W. W. Hauswirth, personal communica-
tion); pig (W. W. Hauswirth, personal communication); human
(19); three ape species (this study)]. One possibility is that
YAY, in concert with other mtDNA sequences or structures, forms
part of a site that is recognized by a processing enzyme. In

this regard, we note that there are a number of YAY sequences in this region to which no 5' termini map, indicating that if YAY plays a role in 5' terminus formation, then additional sequence elements must also be required.

*To whom correspondence should be addressed

+Present address: Department of Genetics, Southwest Foundation for Biomedical Research, P.O. Box 28147, San Antonio, TX 78284

REFERENCES
 1. Attardi, G. (1985) In Reanney, D. and Chambon, P. (eds), Genome Evolution in Prokaryotes and Eukaryotes, Academic Press, New York, pp. 93-145.
 2. Brown, W.M. (1983) In Nei, M. and Koehn, R.K. (eds), Evolution of Genes and Proteins, Sinauer Associates, Sunderland, Massachusetts, pp. 62-88.
 3. Brown, W.M. (1985) In MacIntyre, R.J. (ed), Molecular Evolutionary Genetics, Plenum Press, New York, pp. 95-130.
 4. Clayton, D.A. (1982) Cell 28, 693-705.
 5. Clayton, D.A. (1984) Annu. Rev. Biochem. 53, 573-594.
 6. Brown, W.M., Shine, J. and Goodman, H.M. (1978) Proc. Natl. Acad. Sci. USA 75, 735-739.
 7. Densmore, L.D., Wright, J.W. and Brown, W.M. (1985) Genetics 110, 689-707.
 8. Dunon-Bluteau, D.C. and Brun, G.M. (1987) Biochem. Int. 14, 643-657
 9. Gillum, A.M. and Clayton, D.A. (1978) Proc. Natl. Acad. Sci. USA 75, 677-681.
10. King, T.C. and Low, R.L. (1987) J. Biol. Chem. 262, 6204-6213.
11. King, T.C. and Low, R.L. (1987) J. Biol. Chem. 262, 6214-6220.
12. MacKay, S.L.D., Olivo, P.D., Laipis, P.J. and Hauswirth, W.W. (1986) Mol. Cell. Biol. 6, 1261-1267.
13. Tapper, D.P. and Clayton, D.A. (1981) J. Biol. Chem. 256, 5109-5115.
14. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature 290, 457-465.

15. Anderson, S., de Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982) J. Mol. Biol. <u>156</u>, 683-717.
16. Brown, G.G., Gadaleta, G., Pepe, G., Saccone, C. and Sbisa, E. (1986) J. Mol. Biol. <u>192</u>, 503-511.
17. Cairns, S.S. and Bogenhagen, D.F. (1987) J. Biol. Chem. <u>261</u>, 8481-8487.
18. Sekiya, T., Kobayashi, M., Seki, T. and Koike, K. (1980) Gene <u>11</u>, 53-62.
19. Walberg, M.W. and Clayton, D.A. (1981) Nucleic Acids Res. <u>9</u>, 5411-5420.
20. Wong, J.F.H., Ma, D-P, Wilson, R.K. and Roe, B.A. (1983) Nucleic Acids Res. <u>11</u>, 4977-4995.
21. Fischer, R.P., Topper, J.N. and Clayton, D.A. (1987) Cell <u>50</u>, 247-258.
22. Chang, D.D. and Clayton, D.A. (1984) Cell <u>36</u>, 635-643.
23. Montoya, J., Christianson, T., Levens, D., Rabinowitz, M. and Attardi, G. (1982) Proc. Natl. Acad. Sci. USA <u>79</u>, 7195-7199.
24. Chang, D.D. and Clayton, D.A. (1985) Proc. Natl. Acad. Sci. USA <u>82</u>, 351-355.
25. Chang, D.D. and Clayton, D.A. (1987) EMBO J. <u>6</u>, 409-417.
26. Chang, D.D., Hauswirth, W.W. and Clayton, D.A. (1985) EMBO J. <u>4</u>, 1559-1567.
27. Doda, J.N., Wright, C.T. and Clayton, D. A. (1981) Proc. Natl. Acad. Sci. USA <u>78</u>, 6116-6120.
28. Bogenhagen, D.F., Yoza, B.K. and Cairns, S.S. (1986) J. Biol. Chem. <u>261</u>, 8488-8494.
29. Chang, D.D., Wong. T.W, Hixson, J.E. and Clayton, D.A. (1985) In Quagliariello, E., Slater, E.C., Palmieri, F., Saccone, C. and Kroon, A.M. (eds), Achievements and Perspectives of Mitochondrial Research, Elsevier, Amsterdam, pp. 135-144.
30. Aquadro, C.F. and Greenberg, B.D. (1983) Genetics <u>103</u>, 287-312.
31. Greenberg, B.D., Newbold, J.E. and Sugino, A. (1983) Gene <u>21</u>, 33-49.
32. Brown, W.M., George, M. and Wilson, A.C. (1979) Proc. Natl. Acad. Sci. USA <u>76</u>, 1967-1971.
33. Brown, W.M. and Vinograd, J. (1974) Proc. Natl. Acad. Sci. USA <u>71</u>, 4617-4621.
34. Hixson, J.E. and Brown, W.M. (1986) Mol. Biol. Evol. <u>3</u>, 1-18.
35. Hong, C.F. (1982) J. Mol. Biol. <u>158</u>, 539-549.
36. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M. and Smith, M. (1977) Nature <u>265</u>, 687-695.
37. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA <u>74</u>, 5463-5467.
38. Maxam, A. and Gilbert, W. (1980) Methods in Enzymology <u>65</u>, 499-560.
39. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
40. Yousaf, S.I., Carroll, A.R. and Clarke, B.E. (1984) Gene <u>27</u>, 309-313.

41. Upholt, W.B. and Dawid, I.B. (1977) Cell 11, 571-583.
42. Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C. (1982) J. Mol. Evol. 18, 225-239.
43. Ferris, S.D., Wilson, A.C. and Brown, W. M. (1981) Proc. Natl. Acad. Sci. USA 78, 2432-2436.
44. Karawya, E.M. and Martin, R.G. (1987) Biochim. Biophys. Acta 909, 30-34.
45. Albertini, A.M., Hofer, N., Calos, M.P. and Miller, J.H. (1982) Cell 29, 319-328.
46. Streisinger, G., Okada, Y., Emrich J., Newton, J., Tsugita, A., Terzaghi, E. and Inouye, M. (1966) Cold Spring Harbor Symp. Quant. Biol. 31, 77-84.
47. Hixson, J.E. and Clayton, D.A. (1985) Proc. Natl. Acad. Sci. USA 82, 2660-2664.
48. Saccone, C., Attimonelli, M. and Sbisa, E. (1985) In Quagliariello, E., Slater, E.C., Palmieri, F., Saccone, C. and Kroon, A.M. (eds), Achievements and Perspectives of Mitochondrial Research, Elsevier, Amsterdam, pp. 37-47.
49. Hauswirth, W.W., Van De Walle, M.J., Laipis, P.J. and Olivo, P.D. (1984) Cell 37, 1001-1007.
50. Baldacci, G., Cherif-Zahar, B. and Bernardi, G. (1984) EMBO J. 3, 2115-2120.
51. Zyskind, J.W. and Smith, D.W. (1986) Cell 46, 489-490.