LENGTH MUTATIONS IN HUMAN MITOCHONDRIAL DNA

R. L. CANN AND A. C. WILSON

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

Manuscript received March 1, 1983
Revised copy accepted April 19, 1983

ABSTRACT

By high-resolution, restriction mapping of mitochondrial DNAs purified from 112 human individuals, we have identified 14 length variants caused by small additions and deletions (from about 6 to 14 base pairs in length). Three of the 14 length differences are due to mutations at two locations within the D loop, whereas the remaining 11 occur at seven sites that are probably within other noncoding sequences and at junctions between coding sequences. In five of the nine regions of length polymorphism, there is a sequence of five cytosines in a row, this sequence being comparatively rare in coding DNA. Phylogenetic analysis indicates that, in most of the polymorphic regions, a given length mutation has arisen several times independently in different human lineages. The average rate at which length mutations have been arising and surviving in the human species is estimated to be many times higher for noncoding mtDNA than for noncoding nuclear DNA. The mystery of why vertebrate mtDNA is more prone than nuclear DNA to evolve by point mutation is now compounded by the discovery of a similar bias toward rapid evolution by length mutation.

EARLY studies of genetic variation among mitochondrial DNAs (mtDNAs) of different individuals and species of animals seemed to show that nearly all of the mitochondrial mutations compatible with survival were base substitutions (BROWN 1981). Only in the D loop was there evidence for length mutations, *i.e.*, additions and deletions of bases (UPHOLT and DAVID 1977; CREWS *et al.* 1978; BROWN and WRIGHT 1979; HAYASHI *et al.* 1979; FAURON and WOLSTENHOLME 1980; WALBERG and CLAYTON 1981; FERRIS, WILSON and BROWN 1981; BROWN and SIMPSON 1981; Greenberg et al., 1982). This picture, which contrasted with that for fungal mtDNA, began to change when base sequences from different animal species became available. By then, it was evident that distantly related species also differed by small length mutations at many sites outside the D loop (BIBB *et al.* 1981; ANDERSON *et al.* 1982; BROWN *et al.* 1982; CLARY *et al.* 1982).

We now report the results of a search for length differences in a series of closely related mtDNAs from different human individuals. Initial restriction studies of human variation in mtDNA revealed about 55 different polymorphisms, and all of these were attributed to base substitutions (POTTER *et al.* 1975; BROWN 1980; GILES *et al.* 1980; CASE and WALLACE 1981; DENARO *et al.* 1981). A more extensive survey brought the total to about 165 (CANN, BROWN and WILSON 1982), all of which were provisionally attributed to base substitution. Closer

scrutiny of the fragment patterns, using large numbers of size standards made available by the complete base sequence for mtDNA of one human individual (termed the Cambridge sequence), has allowed us to detect a new class of length polymorphism at several sites outside the D loop. This paper describes the length variants, their location in the mitochondrial genome and their incidence in human populations. Finally, we discuss their possible significance for our understanding of the driving force for mtDNA divergence.

MATERIALS AND METHODS

Human mtDNA: mtDNA was isolated to high purity from placenta, liver and tissue culture cells by methods described previously (BROWN, GEORGE and WILSON 1979). All tissue was collected from donors in accordance with the National Institutes of Health guidelines established for protection of human subjects. The 112 samples included in this study had the following broad racial distribution: 12 Australian aborigines, 19 individuals of African descent, 35 individuals of Asian descent and 46 individuals from Europe, North Africa or the Middle East.

Restriction mapping and gel electrophoresis: All samples in this study were mapped for restriction site polymorphisms with 12 restriction enzymes: *Alu*I, *Ava*II, *Dde*I, *Fnu*DII, *Hae*III, *Hha*I, *Hin*fI, *Hpa*I, *Hpa*II, *Mbo*I, *Rsa*I and *Taq*I. Reactions were carried out under conditions specified by the supplier (New England Biolabs or Bethesda Research Laboratories). Each individual was mapped at approximately 352 sites. Of the individuals examined in this analysis, 21 had been initially compared for restriction fragment differences (BROWN 1980). These samples were included in a more extensive analysis for evidence of molecular polymorphisms at an additional 121 sites. Details of all restriction site polymorphisms detected will appear in a separate publication (R. L. CANN, W. M. BROWN and A. C. WILSON).

Fragments were labeled at the ends with ^{32}P as described by DROUIN and SYMONS (1979) and modified by BROWN (1980). After this step they were run into either 3.5% acrylamide or 1.2% agarose vertical gels for separation. Autoradiography was performed on vacuum-dried gels using Kodak X-Omat RP film and an intensifying screen at −70°.

Denaturing gel conditions were used to verify length mutations in fragments. Urea was added to a 3.5% acrylamide gel solution so that the final concentration was 7 M. Samples were treated as usual for labeling but were dissolved in a final loading buffer containing 90% formamide, heated for 3 min at 90° in a temperature bloc, and quick chilled on dry ice before loading.

Mapping strategy: Our approach takes advantage of the fact that one human mtDNA has been completely sequenced (ANDERSON *et al.* 1981). Any human drawn at random can be expected to differ from this known sequence by three or four base substitutions per kilobase (kb) on the average (BROWN 1980; CANN, BROWN and WILSON 1982; AQUADRO and GREENBERG 1983). Maps for restriction sites in humans of unknown sequence can be derived by comparing their fragment patterns with that predicted from the published DNA sequence. Using this technique, we can detect and rapidly map sites to within a few base pairs (bp) for a large number of individuals without recourse to the double digestion method (DANA, SACK and NATHANS 1973) which had traditionally provided the best approach to rigorous restriction mapping. The high degree of sequence homology among human mtDNAs makes the mapping of single digests with respect to the Cambridge sequence both accurate (to 2 bp) and convenient. (By contrast, it is virtually impossible with the double-digestion method to map the large number of fragments produced by restriction enzymes that recognize a specific sequence of 4 bp.)

RESULTS

Locations of cleavage sites: Figure 1 shows the locations of the 443 cleavage sites mapped in this study. They are widely distributed among all 13 protein-coding genes, both rRNA genes and 21 of the 22 tRNA genes, as well as in the D loop and other noncoding regions. Two hundred and seventy-eight of these

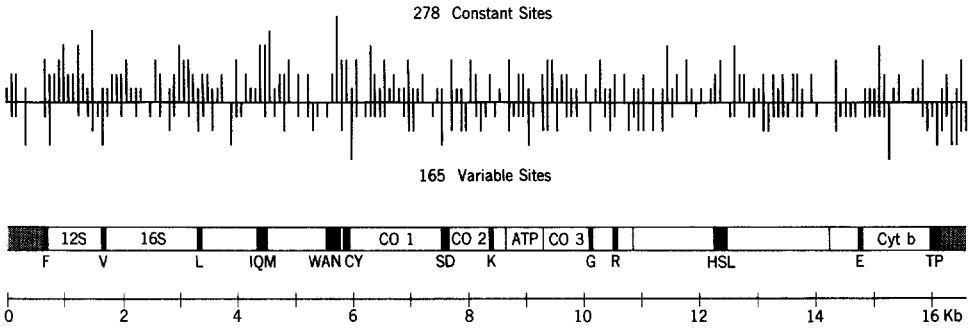


FIGURE 1.—Locations of cleavage sites and functional regions in human mtDNA. The 16,569-bp circular genome sequenced by ANDERSON *et al.* (1981) is drawn in linear form. The major bar shows the regions of known function: 22 tRNA genes, each represented by a single letter and black shading, two rRNA genes (*12s* and *16s*) and 13 genes coding for proteins, eight of which are unidentified and five of which are known, namely, three cytochrome oxidase subunits (Co 1, 2 and 3), one ATPase subunit (ATP) and cytochrome b (Cyt b). Diagonal lines represent the large noncoding region, extending from 16,024 to 576 bp. The upper panel shows the locations of cleavage sites found in mtDNAs from 112 humans with the aid of 12 restriction enzymes. Vertical lines *below* the horizontal line show the variable sites, *i.e.*, those present in some but not all of the 112 mtDNAs. The vertical lines *above* the horizontal line show those sites present in all of the human mtDNAs examined. Height of the vertical lines is proportional to the number of sites found within an 80-bp segment.

sites are invariant, *i.e.*, present in all 112 of the human mtDNAs sampled. The remaining 165 sites are variable, *i.e.*, present in only some of these human mtDNAs. Further details as to the locations, nature and geographic distribution of these sites will appear elsewhere (R. L. CANN, W. M. BROWN and A. C. WILSON).

Length polymorphisms: Our survey of 112 humans revealed length polymorphisms in nine regions of the mtDNA molecule. The locations of these regions, designated I-IX, are given in Table 1 and Figure 2. The short deletions and additions occurring in these regions range in size from about 6 to 14 bp (see Table 1).

To illustrate the type of evidence for these polymorphisms, we focus attention on length mutations in region V. Eight of the 112 individuals gave restriction fragment patterns that differed from other common patterns in the size of a single restriction fragment bearing this region. In seven of these cases, the fragment was shorter, and in one case it was longer than in normal mtDNAs. This shift in size could not be due to a point mutation causing a restriction site to be gained or lost, because all other fragments were accounted for and appeared to line up. This is evident from the comparison of lanes 1 and 3 in Figure 3, which shows the fragments produced by testing three samples of human mtDNA with *Hae*III, a restriction enzyme that cuts the Cambridge sequence at 50 sites. The fragment sizes shown on the left for sample 1 agree with expectation for the Cambridge sequence except for the changes brought about by 6-point mutations (see Table 2). The fragment sizes shown on the right for sample 3 differ in that the 141-bp fragment is replaced by one that is about

TABLE 1

Length mutations in mitochondrial DNA of 112 humans

Polymorphic region	Location[a] (bp)	Length change (bp)	No. of human variants	No. of inferred mutations[b]	Geographic area[c]			
					1	2	3	4
I	37–585	+7	12	9	+	+	+	+
I	37–585	−7	16	12	−	+	+	+
II	3958–4428	+10	2	2	−	+	+	−
III	5261–5552	+7	3	2	−	+	−	+
III	5261–5552	−7	5	2	+	+	+	−
IV	5877–5978	+14	2	1	−	−	+	−
V	8250–8303	+7	1	1	−	+	−	−
V	8250–8303	−7	7	3	+	+	+	−
VI	10352–10598	+10	2	2	−	+	−	−
VII	14608–14802	+7	2	2	+	+	−	−
VII	14608–14802	−7	1	1	+	−	−	−
VIII	15883–15994	+6	1	1	−	−	−	+
VIII	15883–15994	−6	8	6	+	+	+	+
IX	15925–16303	−10	4	1	+	−	+	−

[a] bp locations are numbered as in ANDERSON *et al.* (1981).

[b] The number of inferred mutations is estimated by phylogenetic analysis as exemplified in Figure 5 for the deletion of seven base pairs at site I and the addition of seven base pairs at the same site.

[c] The four aboriginal geographic areas are as follows: (1) Sub-Saharan Africa; (2) China plus Vietnam, the Philippines and Tonga; (3) Australia; (4) Europe plus North Africa and Western Asia.

134 bp long. Sample 2 resembles sample 3 in this respect (see legend, Figure 3). Since this mobility difference was reproducible and did not disappear with longer digestion, it could be attributed to a deletion of 7 bp or to an alteration in secondary structure caused by substitution or modification of a base.

When the experiment was repeated under denaturing conditions, the same mobility difference was seen, which implies that secondary structure is not the cause of the mobility difference. A deletion of about 7 bp from the region giving rise to the 141-bp fragment, *i.e.*, the region between 8250 and 8391 bp, is, therefore, implicated.

This suggestion was confirmed by examining the same region with eight additional restriction enzymes, as illustrated in Figure 4. Two of these enzymes, *Alu*I and *Dde*I, produced a fragment embracing this region and having a size less than 552 bp. In each case, the individuals with the postulated 7-bp deletion exhibited a fragment that was smaller by about 7 bp than that produced by individuals with a conventional fragment size. The remaining six enzymes produced fragments that were too big for reliable detection of a size change of 7 bp (see Figure 4). The variable region must be one common to the three fragments that show the shift, namely, the region from 8250 to 8303. In this region, there is a noncoding segment, from 8270 to 8290 bp, which separates the cytochrome oxidase 2 gene from the lysine tRNA gene; this is considered to be the most likely location for the deletion (see DISCUSSION).

With this procedure, we documented the occurrence of 14 length polymorphisms in regions I–IX of the human mitochondrial genome (Figure 2 and Table

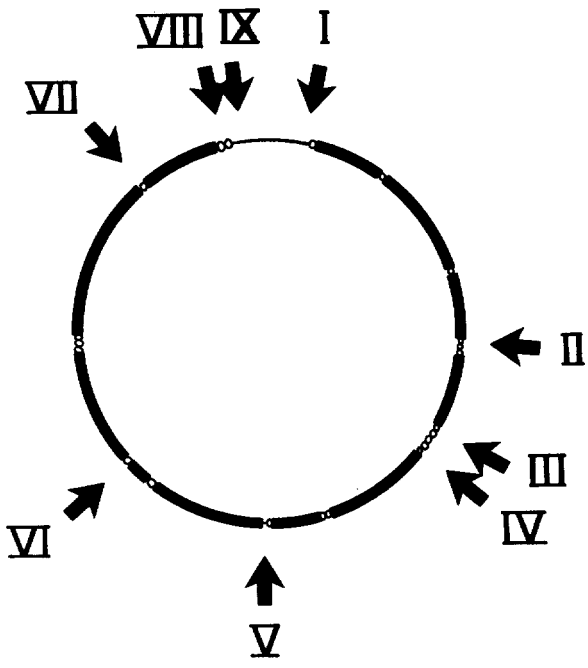


FIGURE 2.—Nine regions of length polymorphism in human mtDNA. The solid bars refer to genes specifying proteins or ribosomal RNA, the empty circles to transfer RNAs and the thin line to the large noncoding region from 16,024 to 576 bp. Arrows show the approximate locations of the nine polymorphic regions, labeled I-IX. For further information about these locations, see Table 1 and Figure 1.

1). In five of the regions (I, III, V, VII and VIII) both additions and deletions were demonstrated. In all cases, it appears probable that the length changes have taken place in or near noncoding DNA, but, as will be evident from Table 1, exact localization has not been achieved and the possibility that such changes have occurred in tRNA genes, the carboxy termini of unidentified reading frames (specifically Urfs 1, 2 and 3), or even at the amino terminus in the case of urf 6 is not excluded (see Table 3).

DISCUSSION

Direct repeats: Some of these length mutations may result from the existence of short, directly repeated sequences in or near noncoding regions. In region V, for example, there is a tandemly repeated, noncoding sequence of 9 bp in the Cambridge mtDNA and (by inference from the maps) in most human mtDNAs, as follows: ACCCCCTCT ACCCCCTCT. An addition or deletion of one of these repeats could be expected by slipped mispairing during replication or repair (JONES and KAFATOS 1982). The observed addition and deletions of about 7 bp in this region agree, within the error of measurement (3 bp), with this expectation.

In region IV, there is a noncoding sequence containing a run of five cytosines and a nontandem 3-bp repeat (CAC). Slipped mispairing within the C series or

TABLE 2

HaeIII sites in four human mtDNAs[a]

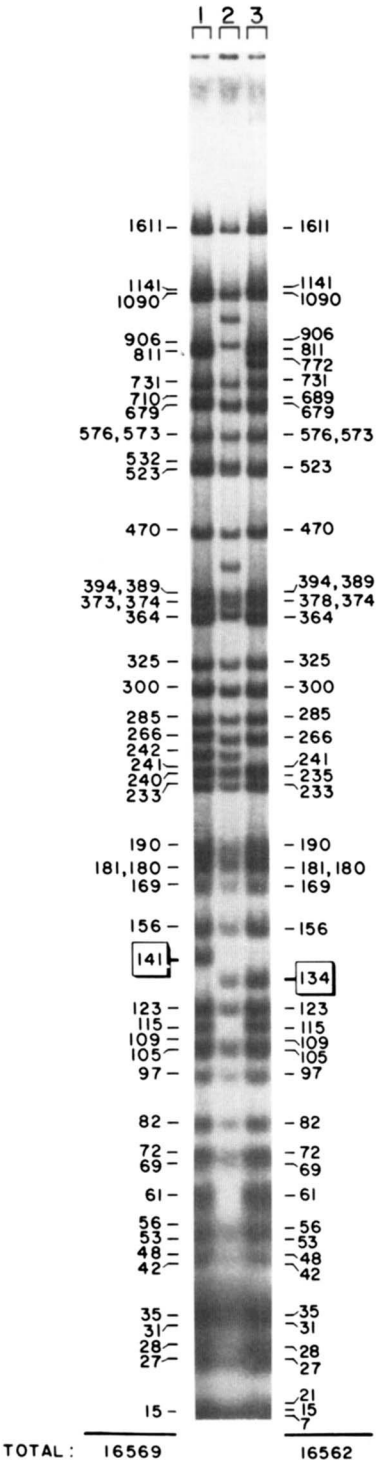
Common sites				Variant sites				
					Presence of sites			
Positions (bp)				Position	1	2	3	C
322	4481	7886	9438	1484	−	+	+	−
1463	4563	8250	10364	3090	+	+	+	−
2173	4848	8391	10689	3842	−	−	+	−
2567	5226	8572	12301	6425	+	+	+	−
3146	5261	8838	12979	6957	+	+	−	+
3315	5837	8994	13051	9553	+	−	+	+
3412	6027	9025	13957	10725	−	−	−	+
3427	6260	9266	15047	13702	−	−	−	+
3607	6383	9294	15152	15172	−	−	−	+
3849	7197	9342	15883	16517	+	−	+	−
3958	7497	9411	16456					
4428								

[a] The four mtDNAs compared include the three (1-3) whose *HaeIII* fragments are displayed in Figure 3 plus (C) the Cambridge mtDNA whose complete base sequence is known (ANDERSON et al. 1981).

between the CAC sequences could generate additions or deletions of 1 to 6 bp (as illustrated by JONES and KAFATOS 1982), but it would be hard to account for the observed 14-bp addition by one slipped mispairing event.

Information about the occurrence of tandemly and nontandemly repeated sequences at other regions of length polymorphism appears in Table 3. A notable feature is the run of five cytosines which occurs not only in region IV and V but also in regions I, II and IX in or near noncoding regions. Additional evidence suggesting that CCCCC is correlated with length variation in mtDNA of primates comes from comparisons of the tRNA serine^{AGY} gene in apes and humans (BROWN *et al.* (1982), where the only length mutation seen in a 896-bp segment is localized to this sequence. This CCCCC sequence is also present in the repeat reported in the 5′ flanking region of the human insulin gene, a region highly polymorphic in length (BELL, SELBY and RUTTER 1982), as well as in regions of yeast mtDNA which undergo frequent length changes (BALDACCI and

FIGURE 3.—Restriction fragments of mtDNA from three humans. This shows an autoradiograph of a 3.5% polyacrylamide gel containing end-labeled fragments of three human mtDNAs digested with the restriction enzyme *HaeIII*. The gel size was 40 × 15 × 0.08 cm, and the running time was 14 hr at 4.5 MA in a buffer containing Tris (0.05 M), boric acid (0.05 M) and EDTA (0.001 M) at pH 8.3 and room temperature (BROWN 1980). Lane 1 contains fragments that conform to the most common pattern seen for this enzyme, and the sizes of the fragments produced are indicated on the left in base pairs. Lane 3 contains fragments of a variant mtDNA, in which there appears to have been a 7-bp deletion causing a particular fragment (134 bp), marked with a horizontal bar on the right, to migrate further than does the corresponding fragment (141 bp) in lane 1. The sizes of the fragments displayed in lane 3 sum to 16,562 bp, whereas those in lane 1 sum to the standard size of the human mitochondrial genome (16,569 bp). An individual whose fragments are displayed in lane 2 also has this variant fragment. The other differences in fragment patterns among lanes are accounted for by point mutations, resulting in either site gains or site losses (see Table 2).


1 2 3
1611
1141
1090
906
811
731
710
679
576,573
532
523
470
394,389
373,374
364
325
300
285
266
242
241
240
233
190
181,180
169
156
141
123
115
109
105
97
82
72
69
61
56
53
48
42
35
31
28
27
15
TOTAL: 16569
1611
1141
1090
906
811
772
731
689
679
576,573
523
470
394,389
378,374
364
325
300
285
266
241
235
233
190
181,180
169
156
134
123
115
109
105
97
82
72
69
61
56
53
48
42
35
31
28
27
21
15
7
16562

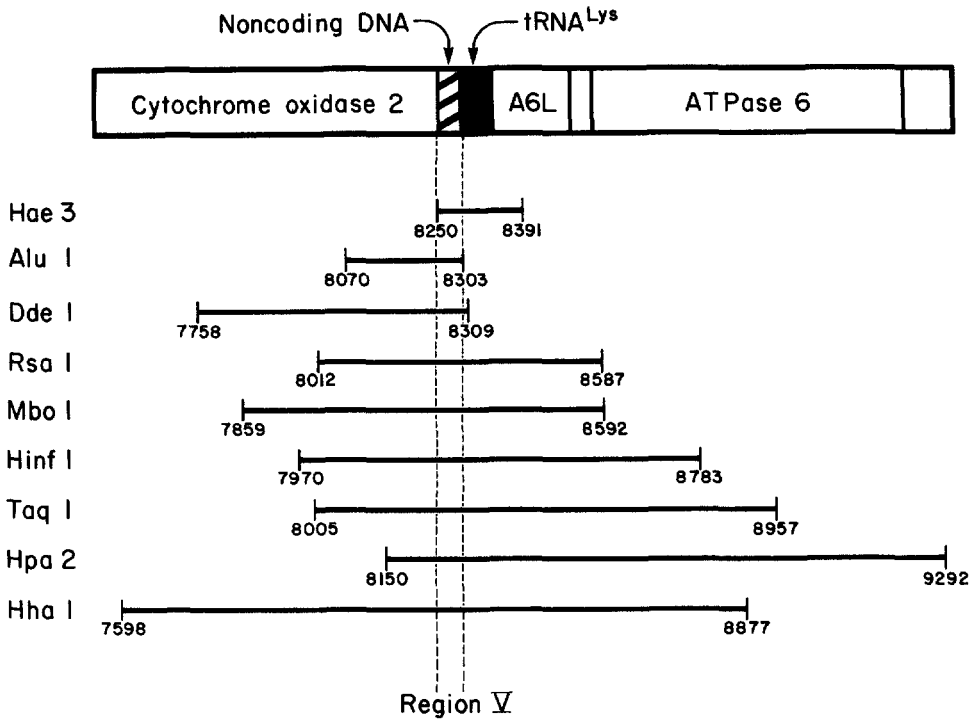


FIGURE 4.—Mapping of length mutations in region V of human mtDNA. The DNA shown, spanning from 7598 to 9292 bp, contains part of the cytochrome oxidase 2 gene, a noncoding region (cross-hatched), the lysine tRNA gene (solid black), an unidentified reading frame (A6L), the partially overlapping ATPase 6 gene, and a portion of the cytochrome oxidase 3 gene (ANDERSON *et al.* 1981). The nine restriction fragments shown include region V, in which deletions and additions of 7 bp have been found. These mutations cause readily detectable changes in the electrophoretic mobility of the two smallest fragments, produced by the enzymes *Hae*III and *Alu*I, barely detectable changes in the *Dde*I fragment and undetectable changes in the larger fragments.

BERNARDI 1982). However, it is also clear that some of the regions of length polymorphism do not contain obvious candidates for slipped mispairing in noncoding regions. Further assessment of the role of repeated sequences in generating mtDNA length mutations will require the sequencing of these regions from variant individuals.

Parallel mutations: Phylogenetic analysis allows an estimate of the extent to which parallel and back mutations have occurred at the nine polymorphic regions. A tree relating the 110 types of mtDNA encountered in our survey appears in Figure 5. It is based on a parsimony analysis that takes account of both length and point mutational variation among these mtDNAs. In building this tree we assumed that mtDNA's inheritance is strictly maternal and clonal. In region I, for example, there have probably been at least 12 independent deletions and nine independent additions. The highest incidences of parallel mutations are in regions I, V and VIII. There is only one region (IX) at which a single mutation accounts for the variation seen. Altogether in the nine regions examined, a total of 45 length mutations were inferred (Table 1), 31 (69%) being

TABLE 3

Direct repeats in regions of length polymorphism

Region	Repeated sequences[a]	Location
I	Many (>50)	Noncoding
II	tccccctc	Urf 1 (3′)
III	ttatactaataatcttata	Urf 2 (3′)
IV	CACCCCCAC	Noncoding
	ccatttttacc	Tyr tRNA (5′)
V	TAGCACCCCCTCTACCCCCTCTAG	Noncoding
	tatttacccctaT	Co 2 (3′)
VI	gaaccgaattgg	Urf 3 (3′), Arg tRNA (5′)
	aaatgccc	Arg tRNA (3′), Urf 4L (5′)
VII	aaacaaa	Urf 6 (5′)
	catacatcat	Urf 6 (5′)
	acACCAatgacccaa	Glu tRNA (5′), Cyt b (5′)
VIII	AAA	Noncoding
	aaatggg	Cyt b (3′)
IX	Many (>30)	Noncoding

[a] In the light strand of the human mtDNA whose sequence has been reported (ANDERSON *et al.* 1981). Capital letters are for noncoding sequences; lower case letters refer to sequences specifying tRNAs or polypeptides. Tandem repeats are underlined; nontandem repeats are overlined; only sequences of more than two contiguous bases are considered. The list includes only those in, or within 20 bp of, a noncoding region or an intergenic junction.

FIGURE 5.—Multiple occurrences of deletions and additions during divergence of 110 human mitochondrial lineages. This tree was derived by the parsimony method (FARRIS, 1970; FERRIS, WILSON and BROWN 1981) from 89 informative characters, of which we attribute 10 to length mutations and 79 to base substitutions. Details about these base substitutions will appear in a separate publication (R. L. CANN, W. M. BROWN and A. C. WILSON). The following symbols designate individuals of different racial heritages: ▲, Australian; △, Asian; ●, African and ○, European, North African or Middle Eastern, as defined in the legend of Table 1. The heavy lines show 12 lineages that have independently experienced a deletion of about 7 bp in region I of the mitochondrial genome. The dotted lines show nine lineages that have independently experienced an addition of about 7 bp in region I.

parallelisms or reversals. Twenty-one of the total are additions, and 24 are deletions. We also note that these mutations, like point mutations (CANN, BROWN and WILSON 1982), show no strong correlation with race or geography, consistent with multiple origins for the length mutations. In a recent analysis of published

sequence data, one of the three sites of length polymorphism examined in the D loop exhibited parallelism, and this study dealt with only seven types of mtDNA (AQUADRO and GREENBERG 1983).

These phylogenetically based estimates of the incidence of length changes give limited support to the idea that length polymorphisms are due only to short repeated sequences. The correlation between number of inferred mutations (Table 1) and the incidence of short repeated sequences in noncoding DNA is approximate at best. The implication is that other factors may be required to account fully for the incidence of length variation.

Incidence of length mutations in nuclear and chloroplast DNA: The incidence of length change in mtDNA may be compared with that in chloroplast and nuclear DNA. Table 4 presents a summary of results for noncoding regions, from which it is evident that the ratio of length mutation to point mutation is at least as high (0.67 to 1.0) for mtDNA as for other types of DNA (mean for nuclear DNA, 0.28; mean for chloroplast DNA, 0.57). When this conclusion is considered, it is worthwhile to note that the estimate of 0.67 for mtDNA is based on restriction mapping, whereas all of the other estimates in Table 4 come from DNA sequencing. The mapping approach is likely, we suggest, to underestimate the incidence of length changes because it is probably poor at detecting the smallest of them. From the mtDNA sequence comparisons made by ANDERSON *et al.* (1982), it is clear that additions and deletions of one or two bases are frequent. So, the proposal that noncoding regions of mtDNA experience a high ratio of length to point mutational change seems sound.

Rates of length change in evolution: Since point mutations are known to accumulate five to ten times more rapidly in mtDNA than in nuclear or chloroplast DNA (BROWN, GEORGE and WILSON 1979; BROWN *et al.* 1982), the absolute rate at which length mutations accumulate in noncoding regions of mtDNA probably exceeds that in nuclear or chloroplast DNA by a large factor also. The implication is that whatever forces are responsible for the high rate of point mutational evolution in mtDNA are also elevating the rate of length change in this same molecule.

The high rate of point mutational evolution in animal mtDNA is due, in part, to an elevation of the rate at sites throughout this genome. Superimposed on this is a further elevation of the rate for tRNA and rRNA genes, owing probably to relaxed constraints on components of the protein-synthetic apparatus (BROWN *et al.* 1982; CANN, BROWN and WILSON 1982). Two classes of explanation exist for the general elevation of the evolutionary rate in mtDNA. One possibility, stressed by BROWN, GEORGE and WILSON (1979), BROWN *et al.* (1982), BROWN and SIMPSON (1982) and MIYATA *et al.* (1982), is that the mutation rate is higher for mtDNA than for nuclear DNA. Another possibility, being developed by R. L. CANN, W. M. BROWN and A. C. WILSON (unpublished results) is that the probability of fixation of mildly deleterious mutations is higher for mtDNA. The latter hypothesis is attractive because it accounts for accelerated evolution by both length and point mutations, regardless of the mechanisms generating the mutations. The challenge that now faces investigators is how to test these competing hypotheses.

TABLE 4

Incidence of length and point mutations in noncoding DNA

Gene region and reference[a]	Species	Noncoding bases compared	Length mutations per kb (a)	Point mutations[b] per kb (b)	Ratio (a/b)
Mitochondrial					
Maps (this work)	Human	1214	12	18	0.67[c]
D loop (1)	Human	1118	3.5	3.5	1.00
Chloroplast					
Interribosomal (2)	Plants	2294	16	28	0.57
Nuclear					
Histones (3)	Sea urchins	295	64	159	0.40
Actins (4)	Yeast	720	1.4	2.8	0.50
Chorions A, B (5)	Silk moth	1232	18	56	0.32
Tubulins (6)	Rat	293	3	17	0.18
Immunoglobulins (7)	Mouse	912	2	4	0.50
Metallothioneins (8)	Human	214	9	65	0.14
ζ-globins (9)	Human	1433	3	94	0.03
β-globins (10)	Goat	1656	14	144	0.10
δ-globins (11)	Human	1267	0.8	1.6	0.50
δ-globins (12)	Primates	535	9	63	0.14

[a] References: (1) GREENBERG, NEWBOLD and SUGINO 1983; (2) TAKAIWA and SUGIURA 1982; (3) BUSSLINGER, RUSCONI and BIRNSTIEL 1982; (4) NELLEN *et al.* 1981; (5) JONES and KAFATOS 1982; (6) LEMISCKHA and SHARP 1982; (7) OLLO and ROUGEON 1982; (8) KARIN and RICHARDS 1982; (9) PROUDFOOT, GIL and MANIATIS 1982; (10) SCHON *et al.* 1981; (11) KIMURA *et al.* 1982; (12) MARTIN, VINCENT and WILSON 1983.

[b] To avoid saturation effects, attention is confined in these calculations to sequences differing at fewer than 16% of the bases compared.

[c] This calculation assumes that all the length mutations are in noncoding DNA. Since it is not certain that these mutations are confined to noncoding regions in human mtDNA, an alternative calculation was made for the regions in which the length mutations have been localized (see Table 1). For these regions, which sum to 2229 bp, the ratio of length mutations to point mutations is 0.50.

We thank E. BLACKBURN, W. BROWN, S. CARR, B. CHAPMAN, A. J. CLARK, M. GEORGE, M. HAMMER, K. HELM-BYCHOWSKI, R. HIGUCHI, E. PRAGER, W. RUTTER, V. SARICH, R. T. WHITE and E. ZIMMER for helpful discussion and valuable comments. We also thank S. ANDERSON, D. CLAYTON and B. GREENBERG for sharing information prior to publication and R. SPRITZ for revised information about human delta globin sequences.

LITERATURE CITED

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. H. L. DEBRUIJN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, F. SANGER, P. H. SCHREIER, A. J. H. SMITH, R. STADON and I. G. YOUNG, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

ANDERSON, S., M. H. L. DEBRUIJN, A. R. COULSON, F. SANGER and I. G. YOUNG, 1982 The complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J. Mol. Biol. **156:** 683–717.

AQUADRO, C. F. and B. D. GREENBERG, 1983 Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics **103:** 287–312.

BALDACCI, G. and G. BERNARDI, 1982 Replication origins are associated with transcription initiation sequences in the mitochondrial genome of yeast. EMBO J. **1:** 987–994.

BELL, G. I., M. J. SELBY and W. J. RUTTER, 1982 The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature **295:** 31–35.

BIBB, M. J., R. A. VAN ETTEN, C. WRIGHT, M. W. WALBERG and D. A. CLAYTON, 1981 Sequence and gene organization of mouse mitochondrial DNA. Cell **26:** 167–180.

BROWN, G. G. and M. V. SIMPSON, 1981 Intra- and interspecific variation of the mitochondrial genome in *Rattus norvegicus* and *Rattus rattus*. Genetics **97:** 125–143.

BROWN, G. G. and M. V. SIMPSON, 1982 Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit 2 genes. Proc. Natl. Acad. Sci. USA **79:** 3246–3250.

BROWN, W. M. and J. W. WRIGHT, 1979 Mitochondrial DNA analyses and the origin and relative age of parthenogenetic lizards (genus Cnemidophorus). Science **203:** 1247–1249.

BROWN, W. M., 1980 Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc. Natl. Acad. Sci. USA **77:** 3605–3609.

BROWN, W. M., 1981 Mechanisms of evolution in animal mitochondrial DNA. Ann. N.Y. Acad. Sci. **361:** 119–132.

BROWN, W. M., M. GEORGE, JR. and A. C. WILSON, 1979 Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:** 1967–1971.

BROWN, W. M., E. M. PRAGER, A. WANG and A. C. WILSON, 1982 Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225–239.

BUSSLINGER, M., S. RUSCONI and M. L. BIRNSTIEL, 1982 An unusual evolutionary behavior of a sea urchin histone gene cluster. EMBO J. **1:** 27–33.

CANN, R. L., W. M. BROWN and A. C. WILSON, 1982 Evolution of human mitochondrial DNA: a preliminary report. p. 157. In: *Human Genetics, part A: The Unfolding Genome*, Edited by B. BONNE-TAMIR, P. COHEN and R. N. GOODMAN. Alan R. Liss, New York.

CASE, J. T. and D. C. WALLACE, 1981 Maternal inheritance of mitochondrial DNA polymorphisms in cultured human fibroblasts. Somatic Cell Genet. **7:** 103–108.

CLARY, D. O., J. M. GODDARD, S. C. MARTIN, C. M-R. FAURON and D. R. WOLSTENHOLME, 1982 *Drosophila* mitochondrial DNA: a novel gene order. Nucleic Acids Res. **10:** 6619–6637.

CREWS, S., D. OJALA, J. POSAKONY, J. NISHIGUCHI and G. ATARDI, 1978 Nucleotide sequence of a region of human mitochondrial DNA containing the precisely identified origin of replication. Nature **277:** 192–198.

DANNA, K. J., G. H. SACK, JR. and D. NATHANS, 1973 Studies of simian virus 40 DNA part 7: a cleavage map of the SV40 genome. J. Mol. Biol. **78:** 363–376.

DENARO, M., H. BLANC, M. J. JOHNSON, K. H. CHEN, E. WILMSEN, L. L. CAVALLI-SFORZA and D. C. WALLACE, 1981 Ethnic variation in Hpa 1 cleavage patterns of human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **78:** 5768–5772.

DROUIN, J. and R. H. SYMONS, 1979 Cloning of human mitochondrial DNA. pp. 471. In: *ICN-UCLA Symposium Proceedings on Extrachromosomal DNA*, Vol. 15, Edited by D. CUMMINGS, P. BORST, I. DAVID, S. WEISSMAN and C. F. FOX. Academic Press, New York.

FARRIS, J. S., 1970 Methods for computing Wagner trees. Syst. Zool. **19:** 83–92.

FAURON, C. M.-R. and D. R. WOLSTENHOLME, 1980 Intraspecific diversity of nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules of *Drosophila mauritiana*, *Drosophila melanogaster* and *Drosophila simulans*. Nucleic Acids Res. **8:** 5391–5410.

FERRIS, S. D., A. C. WILSON and W. M. BROWN, 1981 Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc. Natl. Acad. Sci. USA **78:** 2432–2436.

GILES, R. E., H. BLANC, H. M. CANN and D. C. WALLACE, 1980 Maternal inheritance of human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **77:** 6715–6719.

GREENBERG, B. D., J. E. NEWBOLD and A. SUGINO, 1983 Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene **21:** 33–49.

HAYASHI, J. I., H. YONEKAWA, O. GOTOH, Y. TAGASHIRA, K. MORIWAKI and T. H. YOSIDA, 1979 Evolutionary aspects of variant types of rat mitochondrial DNAs. Biochim. Biophys. Acta **564:** 202–211.

JONES, C. W. and F. C. KAFATOS, 1982 Accepted mutations in a gene family: evolutionary diversification of duplicated DNA. J. Mol. Evol. **19:** 87–103.

KARIN, M. and R. I. RICHARDS, 1982 Human metallothionein genes: primary structure of the metallothionein-2 gene and a related processed gene. Nature **299:** 797–802.

KIMURA, A., E. MATSUNAGA, Y. OHTA, T. FUJIYOSHI, T. MATSUO, T. NAKAMURA, T. IMAMURA, T. YANASE and Y. TAKAGI, 1982 Structure of cloned δ-globin genes from a normal subject and a patient with δ-thalassemia: sequence polymorphisms found in the δ-globin gene region of Japanese individuals. Nucleic Acids Res. **10:** 5725–5732.

LEMISCKHA, I. and P. A. SHARP, 1982 The sequences of an expressed rat α-tubulin gene and a pseudogene with an inserted repetitive element. Nature **300:** 330–335.

MARTIN, S. L., K. A. VINCENT and A. C. WILSON, 1983 Rise and fall of the delta globin gene. J. Mol. Biol. **164:** 513–528.

MIYATA, T., H. HAYASHIDA, R. KIKUNO, M. HASEGAWA, M. KOBAYASHI and K. KOIKE, 1982 Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J. Mol. Evol. **19:** 28–35.

NELLEN, W., C. DONATH, M. MOOS and D. GALLWITZ, 1981 The nucleotide sequences of the actin genes from *Saccharomyces carlsbergensis* and *S. cerevisiae* are identical except for their introns. J. Mol. Appl. Genet. **1:** 239–244.

OLLO, R. and F. ROUGEON, 1982 Mouse immunoglobulin allotypes: post duplication divergence of γ2a and γ2b chain genes. Nature **296:** 761–763.

POTTER, S., J. E. NEWBOLD, C. A. HUTCHINSON and M. H. EDGELL, 1975 Specific cleavage analysis of mammalian mitochondrial DNA. Proc. Natl. Acad. Sci. USA **72:** 4496–4500.

PROUDFOOT, N. J., A. GIL and T. MANIATIS, 1982 The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell **31:** 553–563.

SCHON, E. A.,M. L. CLEARY, J. R. HAYNES and J. B. LINGREL, 1981 Structure and evolution of goat $\gamma,\beta,^{c}$ and β^{a}-globin genes: three developmentally regulated genes contain inserted elements. Cell **27:** 359–369.

TAKAIWA, F. and M. SUGIURA, 1982 Nucleotide sequence of the 16s-23s spacer region in an rRNA gene cluster from tobacco chloroplast DNA. Nucleic Acids Res. **10:** 2665–2676.

UPHOLT, W. B. and I. B. DAVID, 1977 Mapping of mitochondrial DNA of individual sheep and goats: rapid evolution in the D loop region. Cell **11:** 571–583.

WALBERG, M. W. and D. A. CLAYTON, 1981 Sequence and properties of the human KB cell line and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res. **9:** 5411–5421.

Corresponding editor: M. T. CLEGG