Mitochondrial DNA Migration Events in Yeast and Humans: Integration by a Common End-joining Mechanism and Alternative Perspectives on Nucleotide Substitution Patterns

Jeffrey L. Blanchard[1] *and Gregory W. Schmidt*

Department of Botany, University of Georgia, Athens

In contrast to extensive infiltration of plant nuclear genomes by mitochondrial and chloroplast DNA fragments, a computer assessment method could only detect seven mitochondrial DNA integration events in *Saccharomyces cerevisiae* chromosomes and five examples of DNA migration into mammalian nuclear genes. No evidence could be detected for mitochondrial DNA insertion into chromosome III of *Caenorhabditis elegans* or in nuclear DNA sequences of *Drosophila* sp. or *Plasmodium falciparum.* Thus, the quantity of organellar DNA in the nucleus appears to vary amongst organisms and is lower in *Saccharomyces cerevisiae* than suggested by experimental plasmid systems. As in plants, migratory mitochondrial DNA fragments in yeast and mammals are found in intergenic regions and introns. Although many of these insertions are located near retroelements, mitochondrial DNA incorporation appears to be independent of retroelement insertion. Comparison of the mitochondrial DNA fragments with mitochondrial transcription maps suggest that two fragments may have transposed through DNA-based and one through RNA-based mechanisms. Analyses of the integration sites indicate that organellar DNA sequences are incorporated by an end-joining mechanism common to yeast, mammals, and plants. The transferred sequences also provide a novel perspective on rates and patterns of nucleotide substitution. Analysis of the D-loop region including a nuclear copy of mitochondrial DNA supports a progressive reduction in D-loop length within both monkey and great apes mitochondrial lineages. Relative distance tests polarized with nuclear copies of the mitochondrial 12S/16S rRNA region suggest that a constant number of transversions has accumulated within the great ape clade, but the number of transitions in orangutan is elevated with respect to members of the human/chimp/gorilla clade. In addition to DNA migration events, 29 nuclear/mitochondrial genes were identified in GenBank that appear to result from inadvertent ligation of nuclear and mitochondrial mRNA transcripts during the cloning process.

Introduction

The presence in the nucleus of DNA sequences also found in organellar genomes was first detected in filter hybridization studies (Richards 1967; du Buy and Riley 1967) and was confirmed by DNA sequence comparison between nuclear and organellar genomes (van den Boogaart, Samallo, and Agsterbibe 1982; Farrelly and Butow 1983; Gellissen et al. 1983). Stable transposition of organellar DNA over time is supported by nuclear-localized mitochondrial DNA (numtDNA) sequences that show a range in divergence from the original mitochondrial sequences (Fukuda et al. 1985). Variable lengths of organellar-derived sequences have been reported ranging from 30 bp (Fukuchi et al. 1991) to 300–600 kb of a tandemly repeated 7.9-kb segment (Lopez et al. 1994). The insertions have been localized to introns (Pichersky and Tanksley 1988), flanking regions of nuclear genes (Bernatzky, Mau, and Clarke 1989), and to telomeric regions (Louis and Haber 1991).

Although numtDNA has led to aberrant mitochondrial restriction fragment length polymorphism (RFLP) maps due to the additional nuclear fragments (Quinn and White 1987) and to unexpected polymerase chain reaction products (Smith, Thomas, and Patton 1992), it can be useful in inferring other genetic and evolutionary processes. For example, through sequence comparisons between mitochondrial DNA (mtDNA), edited mitochondrial RNA transcripts, and numtDNA, transfer to the nucleus was suggested to involve RNA intermediates (Nugent and Palmer 1991; Grohmann, Brennicke, and Schuster 1992; Blanchard and Schmidt 1995). Similarly, support for DNA-based transfer comes from the assimilation of large tracts of chloroplast DNA containing intergenic regions (Ayliffe and Timmis 1992) and of nontranscribed areas of mtDNA (Lopez et al. 1994) into the nuclear genome. Analyses of plant nuclear integration sites before and after DNA incorporation have been used to suggest a model of illegitimate recombination similar to the uptake of "foreign DNA" through transformation (Sun and Callis 1993; Blanchard and Schmidt 1995). However, in fungi (Farrelly and Butow 1983), animals (Zullo et al. 1991), and protists (Ossario, Sibley, and Boothroyd 1991), transfer to the nucleus has been suggested to be facilitated by transposable or viral elements because of their proximity to many numtDNA sequences.

[1]Present address: Department of Biology, University of Oregon.

Key words: DNA migration, mitochondria, genomic evolution, gene transfer, substitution rate, *Saccharomyces cerevisiae, Homo sapiens,* illegitimate recombination.

Address for correspondence and reprints: Jeffrey L. Blanchard, Department of Biology, University of Oregon, Eugene, Oregon 97403. E-mail: jeffb@oregon.uoregon.edu

Mol. Biol. Evol. 13(3):537–548. 1996

We developed a computer assessment method that identified over 25 fragments of organellar DNA in plant nuclear genes (Blanchard and Schmidt 1995). In this report, we have extended our analyses to include other groups of organisms for which a sufficient database has been generated and define additional features of DNA migration events. We identify examples of mtDNA transfer into introns of human nuclear genes, flanking regions of rat genes, and intervening regions of yeast and human genes. To determine whether transfer was RNA or DNA based, numtDNA fragments were compared to yeast (De Zamaroczy and Bernardi 1986) and mammalian transcript maps (Clayton 1992). A 400-bp sequence of 12S rRNA from a number of Old and New World monkeys (van der Kuyl et al. 1995) and the sequencing of complete mitochondrial genomes from chimpanzee, gorilla, and orangutan (Horai et al. 1995) enabled the use of numtDNA sequences in a phylogenetic context. Finally, nuclear/mitochondrial chimeric sequences were detected, which appear to be artifacts of the cloning process.

Materials and Methods

Computer Resources

DNA sequence databases were accessed through the BioSciences Computational Resource at the University of Georgia. The GenBank (version 82.0) and EMBL (version 37.0) databases were queried with the algorithms in the BLASTN search program (National Center for Biotechnology, National Institutes of Health, Bethesda, MD; Altschul et al. 1990), FASTA (University of Wisconsin, Madison, Genetics computer Group [GCG] version 7.0; Deveraux, Haeberli, and Smithies 1984) or FastDB (IntelliGenetics, Mountain View, CA). Percent identity between the nuclear and organellar sequences was calculated from pairwise alignments using FASTA. Sequence alignments were conducted using PILEUP (GCG) and GENALIGN (IntelliGenetics, Mountain View, CA) and when necessary were modified by hand. Phylogenetic analyses were performed using PHYLIP version 3.5 (Felsenstein 1994) and PAUP version 3.1.1 (Swofford 1993). MacClade (Madison and Madison 1992) was used as an intermediary between computer packages and to calculate transitions and transversions among taxa.

Search Strategy

Complete mitochondrial genomic sequences from *Homo sapiens* (J01415), *Rattus norvegicus* (X14848), *Mus musculus* (J01420), *Saccharomyces cerevisiae* (M62622), *Caenorhabditis elegans* (X54252), *Drosophila melanogaster* (J01404, J01409, M37275), and *Plasmodium falciparum* (M76611) were retrieved from GenBank. These files were used sequentially to query sequence libraries in GenBank using BLASTN or GCGFASTA. The mitochondrial genome of *S. cerevisiae* was first divided into 10-kb blocks overlapping by 100 bp. Separate searches were performed with regions that have been extensively sequenced for systematic purposes. GCGFASTA searches were run with the gap penalty set at the default value (12) and then at the maximum value (24). All other parameters were left at the default values.

Sequences were collected that fulfilled the following selection criteria: (1) lengths greater than 40 nucleotides and (2) at least 80% identity to mtDNA when the length is between 40 and 100 nucleotides and at least 70% identity when the length is greater than 100 nucleotides. These cutoff values generally gave "Expect" probabilities, using BLASTN, smaller than 10^{-10} between the nuclear and mitochondrial sequences. The "Expect" value represents the probability of finding a similar sequence with this degree of identity or higher in the database due to random chance. Expressed sequence tags resulting from cDNA sequencing projects and adenine- and thymidine-rich regions were not included in the analyses because of uncertainties regarding their genomic origin.

Results

Nuclear Sequences Similar to mtDNA

A summary of yeast and mammalian nuclear sequences similar to mtDNA identified through searches of the databases is presented in tables 1 and 2. Sequences similar to mtDNA, varying in length from 42 to 3,110 nucleotides, were identified in five different yeast chromosomes and 29 mammalian nuclear genes. No instances of mtDNA migration were observed in chromosome III of *C. elegans* or in genomic clones of *Drosophila* sp., or *P. falciparum.* Because the characteristics of mtDNA sequences in cDNA libraries suggest that they may be derived from the cloning process, their analysis will be presented separately from the genomic numtDNA sequences.

Five mitochondrial sequences were identified in human genomic clones and six previously unidentified insertions were found in yeast chromosomal DNA (tables 1 and 2). In yeast, the detected numtDNA sequences comprise 1,098 bp of the 3,583,808 bp (0.031%) from these seven chromosomes. The two mitochondrial fragments found in the telomeric region of chromosome IX are also located on telomeres of chromosomes IV and X in strain A364a and chromosomes IX and X in strain YP1 but are not found in other *S. cerevisiae* strains (Louis and Haber 1991). Three of the numtDNA sequences identified in yeast are composed of multiple fragments derived from disparate regions of the mitochondrial genomes, which appear to have ligated together prior to or during insertion. We also detected the numtDNA sequence reported by Farrelly and Butow (1983) downstream of a yeast Prp39 gene (accession

Table 1
Regions of *S. cerevisiae* Chromosomes Similar to MtDNA

Chromosome						
Number	Size (bp)	Accession no.	Position*	% Id	Size (bp)	Origin of Segment†
I	223,111	L28920	13814–13953 (#1)	97	140	cytb intron
I	223,111	L28920	13954–13995 (#2)	93	42	coxI
II	807,188	Z35929	3660–3824	75	165	21S rRNA
III	315,338	X59720	272148–272206	90	59	coxIII
V	569,202	None detected	—	—	—	—
VIII	592,638	None detected	—	—	—	—
IX	439,885	Z38061	41272–41380	80	109	coxII
IX	439,885	Z46921	6947–7015 (#1)	88	69	cytb + intron
	439,885	Z46921	7015–7245 (#2)‡	100	109	coxI intron§
IX	439,885	Z38060	18538–18629	91	92	IG + coxII
XI	666,446	Z28218	236–483 (#1)	66–72	248	IG
	666,446	Z28218	484–548 (#2)	97	65	21S rRNA

* Position according to the author's numbering in the sequencing file.
† IG, intergenic region; cox, cytochrome oxidase; cytb, cytochrome b.
‡ Similarity noted by Louis and Haber (1991).
§ Similar to *Saccharomyces douglassi* (M97514).

#L29224). In humans, a 71-bp region similar to a segment of the mitochondrial 16S rRNA gene is located in intron 18 of two human β-myosin heavy chain genes (table 2). The finding of mtDNA sequences in two different clones isolated by two separate research groups further supports that these sequences are derived from intracellular DNA migration events.

Incorporation of mtDNA by End-joining

The location of numtDNA fragments with respect to nuclear genes, pseudogenes, and transposable elements is illustrated in figure 1. To estimate the nuclear DNA sequence at the insertion site prior to the integration of numtDNA we relied on two different tactics. First, the same gene in another taxon was identified that lacked the numtDNA sequences. The myosin heavy chain gene in rats lacks numtDNA and the numtDNA appears to be responsible for a length polymorphism between the human and rat intron (accession #L12104—alignment not shown). Therefore the rat intron was used to approximate the site before the integration event (fig. 2*A*), although some of the differences could be due to changes in the rat or human intron after their divergence or changes in mtDNA or numtDNA sequence after the integration event. Second, in yeast two of the numtDNA sequences split one member of a multicopy gene family (or possibly pseudogene family). In these two cases, the copy of the gene most similar to the region surrounding the numtDNA insertion was used to represent the site prior to numtDNA integration (fig. 2*B,C*). From these alignments shown in figure 2 some base complementarity (0–3 bp) is observed between the mitochondrial and nuclear sequences and all insertions are flanked by short regions of either direct or inverted repeats. The direct repeats are only apparent in the flanking regions of the numtDNA after the integration occurred. In addition, the insertion into *S. cerevisiae* YKRS1 appears to have accompanied a small deletion. For most of the numtDNA sequence in tables 1 and 2 comparable sequences lacking the numtDNA could not be identified.

Because the contiguous numtDNA fragments in *S. cerevisiae* chromosome I are separated by 20 nucleotides from a TY element, it appears that the mtDNA integration event occurred separately from TY element insertion (fig. 3). In addition, most numtDNA fragments observed in this study are not adjacent to any previously described transposable elements. The exception is the numtDNA in the telomere of *S. cerevisiae* chromosome IX, but in this case we cannot determine if the retroelements and numtDNA sequences were derived from single or multiple integration events.

Evidence for RNA- and DNA-based Transfer

To determine whether transfer was RNA or DNA based we relied on yeast (De Zamaroczy and Bernardi 1986) and mammalian transcript maps (Clayton 1992). If the ends of the numtDNA sequenced matched the ends of human or yeast mitochondrial transcripts we infer that the transfer process involved an RNA intermediate. If the numtDNA sequence contained a region of nontranscribed DNA we infer the transfer process to be solely DNA based. The fragment at the end of the microsatellite sequence is similar to one end of a mitochondrial transcript, indicating a possible RNA-based transfer. On the other hand, the mitochondrial D-loop region found in the T-cell receptor (TCR) locus is not transcribed in humans and the yeast chromosome IX insertion (#Z38060) contains five nucleotides of non-

Table 2
Mammalian Nuclear Genomic and cDNA Sequences Similar to mtDNA

DNA Source	Sequence Description	Accession No.	Position*	Size (bp)	% Id	Origin of Segment
Homo sapiens						
genomic DNA	β-Myosin heavy chain	X52889	In: 11586–11655	70	87	16S rRNA
	β-Myosin heavy chain	M57965	In: 12903–12971	68	87	16S rRNA
	Complement component C2	L09706	In: 5775–5831	57	80	coxI
	T-cell receptor β locus	L36092	G: 148071–151331	3,110	80	D-loop; tRNA-Phe; 12S rRNA; RNA-Val; 16S rRNA
	Microsatellite DNA	Z23381	U: 310–371E	62	100	tRNA-Ala(END)
Homo sapiens cDNA ...	Hox1.8 homeobox gene	S41211	5′: 2–114	113	100	nad5
	Lymphocyte-specific protein 1	M33552	5′: 1–46	46	100	16S rRNA
	Cortex mRNA containing Alu	X51525	5′: 1–208	208	100	16S rRNA (BEG)
	Monocyte chemotactic protein 3†	X72308	5′: 4–281	278	99	12S rRNA
	Lung surfactant protein D†	X65108	5′: 4–141	138	99	cytb
	Glucagon-like peptide-1†	U01157	3′: 3071–3179E	109	100	16S rRNA
	κ-casein†	M73628	5′: 1–55	55	100	16S rRNA
	Replication factor C†	L23320	5′: 10–78	69	100	tRNA-Gly (COMP)
	hLON ATP-dependent protease	U02389	5′: 22–70	49	96	nad6
	Break point cluster protein†	M24603	5′: 1–44	44	100	nad2
	Break point cluster protein†	X02596	5′: 1–44	44	100	nad2
	Transcription initiation factor IIB†	X95268	5′: 26–319	294	99.7	nad2 (END)
	S-adenosylmethionine synthetase†	D11332	5′: 1–258	258	100	coxI
	Semaphorin	L26081	3′: 2531–2594E	64	100	16S rRNA
	β-Adducin	X58199	5′: 1–108	108	100	L-strand (COMP)
	von Willebrand factor II†	M17588	5′: 2–59	58	97	nad5
Mus musculus cDNA ...	Perforin†	X60165	3′: 2103–2332E	230	99	coxII
	Cux homeodomain protein†	X75013	5′: 9–167	159	96	12S rRNA
	MPT δ-tyrosine phosphatase†	D13904	5′: 1–198	198	99	nad3
	Stromal cell derived protein-1†	D16847	5′: 65–434	370	99	coxI
Rattus norviegicus						
cDNA	Frizzled gene	L02529	5′: 245–353	109	98	16S rRNA
	Guanine nucleotide releasing protein (GNRP)	L10336	5′: 143–716	572	89	16S rRNA
	Protein-tyrosine phosphatase†	L11587	3′: 6101–6229 (1)	129	100	cytb
			6228–6258 (2)	31	100	cytb
	Protein-tyrosine phosphatase†	L12329	3′: 5015–5143 (1)	129	100	cytb
			5142–5172 (2)	31	100	cytb
	Tropomyosin†	L24776	3′: 810–1101E	292	98	nad2
	Syntaxin-binding protein	L26264	3′: 2327–2369E	43	100	16S rRNA
	Protein S†	U06230	5′: 1–35	36	100	coxI
	D-β-Hydroxybutyrate dehydrogenase	M89902	5′: 2–84	83	100	atp8
	3-β-Hydroxysteroid dehydrogenase†	S63167	5′: 3–1022	1,020	99	D-loop; tRNA-Phe; 12S rRNA (BEG)
Cavis sp. cDNA	Hox1.7 homeobox gene	X13537	5′: 1–220	220	>72‡	nad4 (END)

* Position according to the author's numbering in the sequence file; IN, introl; U, relation to nuclear gene unknown; G, intergenic; 5′, 5′ of coding region; 3′, 3′ of coding region; E, end of sequence entry.

† Other cDNAs are available from this same species without the mtDNA region.

‡ The mitochondrial nad4 sequence has not been reported for *Cavia* sp.

transcribed DNA upstream of the mitochondrial gene's transcription start site. Therefore, these two fragments appear to be derived from a DNA-based transfer process. Because the rest of the genomic fragments listed in tables 1 and 2 are internal to transcribed regions we could not infer an RNA or DNA intermediate.

Perspectives on Nucleotide Substitution Measurements Provided by Using TCR numtDNA as an Outgroup

To estimate the period in which the mitochondrial fragment in the TCR was assimilated into the nuclear genome maximum parsimony analysis was performed

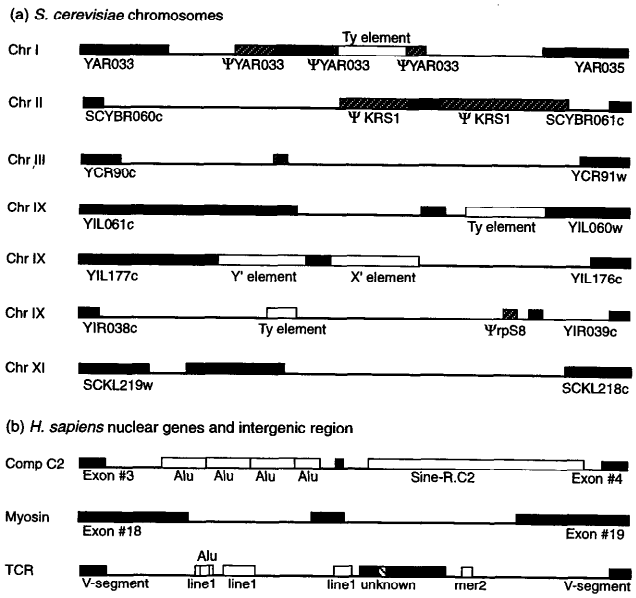


FIG. 1.—Location of mtDNA insertions relative (*a*) to genes in *S. cerevisiae* chromosomes and (*b*) to human nuclear genes and an intergenic region. Black, gray, clear, and striped boxes represent, respectively, putative nuclear genes, numtDNA, transposable elements, nuclear pseudogenes. Position of the transposable elements is taken from the GenBank entries listed in tables 1 and 2. The three chromosome IX regions are Z38061, Z46921, and Z38060, respectively.

on primate and mammalian mtDNA spanning the 12S/16S rRNA region (fig. 4*A*). The analysis strongly supports a transfer point prior to the divergence of the great ape clade. A tree was also constructed with a shorter 12S rRNA region from which New and Old World monkey sequences are available (fig. 4*B*). The topology of this tree is not well supported by the bootstrap analysis nor is it in concordance with current primate mitochondrial trees. The Old World monkeys branched off prior to the split up New World monkey and great ape clades and the branching order of the great ape clade is as shown in figure 4*A* (Ruvolo 1993; Adkins and Honeycutt 1994). However, this analysis does provide further support for a transfer prior to the divergence of the great ape clade and possibly prior to the ape and monkey clades.

The number of inferred changes along the TCR numtDNA branch should represent nuclear changes following the integration of the mitochondrial fragment into the nuclear genome. The PAUP analysis infers 205 changes (7.9%) in figure 4*A* and 25 changes (6.0%) in figure 4*B*. As a test of the accuracy of PAUP in reconstructing nuclear changes in the TCR numtDNA the number of changes was determined at sites invariant among the taxa listed in figure 4*A* and *B*. Out of 1,242 invariant sites in the larger data set there are 71 nucleotide substitutions (5.7%) and in the smaller 12S rRNA

```
(a)
mtDNA    ACCGAGCCTGGTGATAGCT...AATCCCCTTGTAAATTTA
H-myo    TACAACCCCTTGGATAGCT...AATCCCCCAATGAGACAT
R-myo    AATAGCCCCT---------...--------AATGAGGCAT
(b)
mtDNA    CTTCAGCAGATAGGAACCA...TATATTACCCTGTTCAAA
ΨKRS1    GATTGGCAGATAGAAACCA...TATATCACATTGGAAAAA
SCKRS1   GATTG-AAGA---------...--------ATTGGAAAAA
(c)
mtDNA    AAATAAATTAATTAATTAT...ATTTGAGCTAATTTTTCA
ΨYAR033  AGATGAACGAATAAATAAT...ATCTGAGCGTATTTATTT
YAR033   AGATGAACCA---------...--------GTATTTATTT
```

FIG. 2.—Comparison of the integration site sequences before and after insertion of numtDNA to appropriate mitochondrial sequences. (*a*) The rat myosin intron (R-myo L12104) represents the site of insertion for the human myosin heavy chain gene (H-myo). (*b, c*) The *S. cerevisiae* nuclear genes YAR033 (L28920) and SCKRS1 (X56259) are inferred to represent the integration site before the transfer event. Gaps are shown by a dash and periods represent intervening sequence. The dashed lines represent an inverted repeat. Solid lines highlight direct repeats. The boxed region represents a region that may have been deleted during the integration event.

data set there are 13 substitutions across 218 invariant sites (6.0%). The distribution of these changes is mapped on the 12S rRNA alignment (fig. 4*C*).

The numtDNA fragment in the TCR locus harbors a 127-bp region that is not similar to any reported mitochondrial Old World monkey or great ape D-loop sequence (fig. 5*A*). The nucleotide composition of this gap is strikingly characteristic of mtDNA. The heavy strand AT composition is 60% (compared with 60% in the preceding D-loop region) and G content is 8% (vs. 12%). To further investigate insertion/deletion processes in the D-loop region, parsimony principles were applied to reconstruct length changes (fig. 5*B*). Along all branches there is a net reduction in D-loop size, indicating that the D-loop is being progressively shortened in these primate mitochondrial lineages.

Because no primate outgroup is available for evaluating mitochondrial nucleotide substitution patterns in the great ape clade in the 12S/16S rRNA region, the TCR numtDNA sequence was employed for this purpose. From either the TCR numtDNA a greater number of substitutions have occurred in orangutans than in the

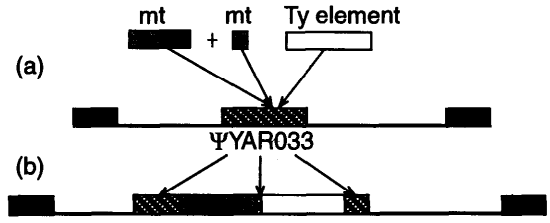
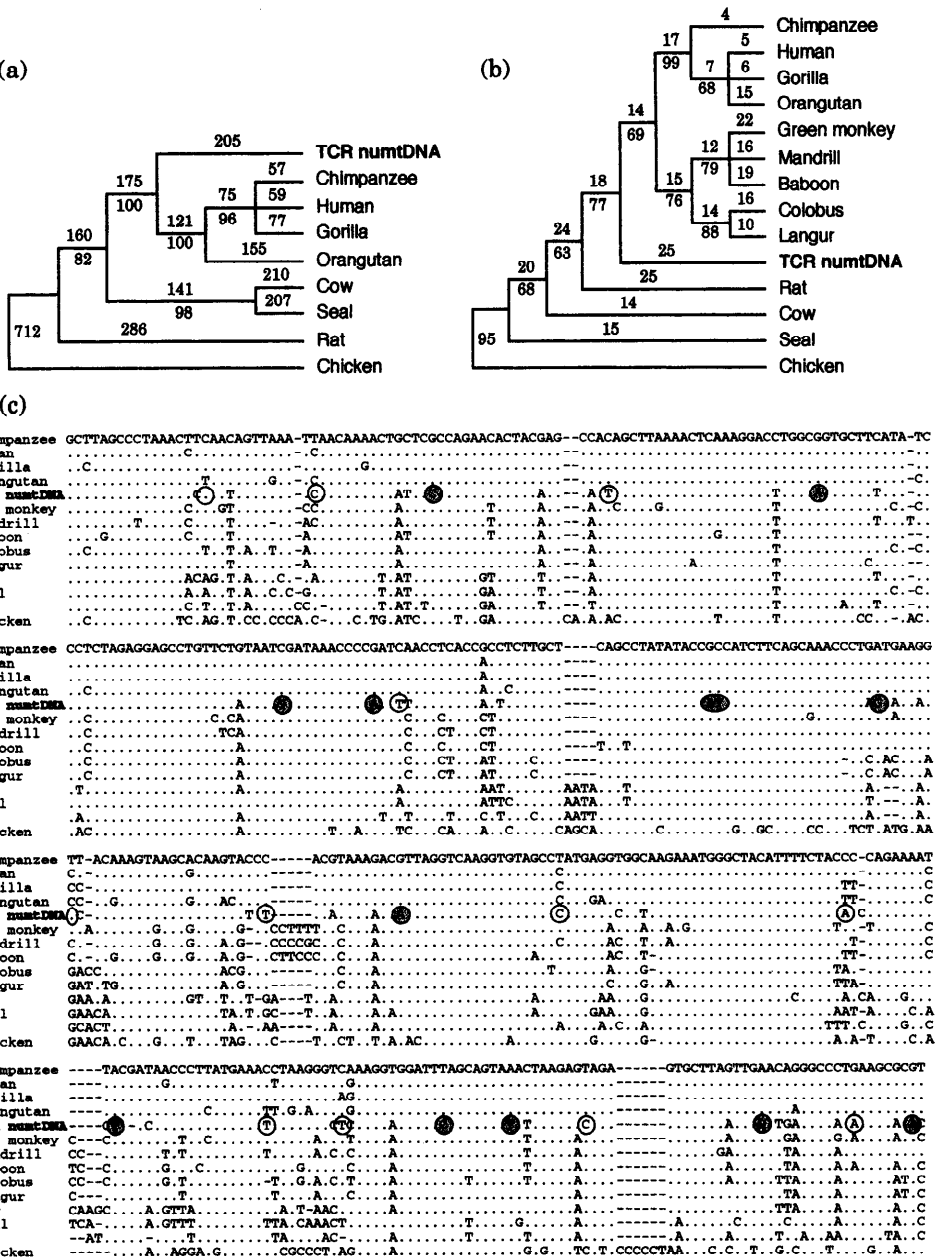


FIG. 3.—Mitochondrial DNA insertion is independent of retroelement insertion. (*a*) Two mitochondrial fragments have inserted into the *S. cerevisiae* YYAR033 (L28920) gene 20 bp away from the site of a TY element insertion. (*b*) These insertions split the YYAR033 into three pieces. Nucleotides surrounding the mtDNA insertion site are shown in figure 2.


(a)
205
TCR numtDNA
57
Chimpanzee
175
100
75
59
Human
121
96
77
Gorilla
160
100
155
Orangutan
82
210
Cow
141
207
Seal
98
712
286
Rat
Chicken
(b)
4
Chimpanzee
17
5
Human
99
7
6
Gorilla
68
15
Orangutan
14
22
Green monkey
69
12
16
Mandrill
18
79
19
Baboon
15
77
16
Colobus
76
14
10
Langur
88
24
25
TCR numtDNA
63
20
25
Rat
68
14
Cow
15
95
Seal
Chicken
(c)
Chimpanzee GCTTAGCCCTAAACTTCAACAGTTAAA-TTAACAAAACTGCTCGCCAGAACACTACGAG--CCACAGCTTAAAACTCAAAGGACCTGGCGGTGCTTCATA-TC
Human .....C.....-.C.....
Gorilla ..C.....-.....G.....
Orangutan .....T.....G..-.C.....
TCR numtDNA .....C.T.....-C.....AT.....A...A.T.....T.....T...-.
Gr. monkey .....C..GT.....-CC.....A.....T.....A...A.C...G.....T.....C.-C.
Mandrill .....T.....C....T....-..-AC.....A.....T.....A...A.....T.....T..T..
Baboon ....G.....C....T.....-A.....AT.....T.....A...A.....G.....T.....--.
Colobus ..C.....T..T.A..T...-A.....A.....A...A.....T.....C.-C.
Langur .....T.....-A.....A.....A...A.....A.....T.....C.....--.
Cow .....ACAG.T.A...C.-.A.....T.AT.....GT....T...A.....T.....T...-.
Seal .....A.A..T.A..C.C-G.....T.AT.....GA....T...A.....T.....C.-C.
Rat .....C.T..T.A.....CC.-.....T.AT.T.....GA....T...T.....T.....A..T...-.
Chicken ..C.....TC.AG.T.CC.CCCA.C-...C.TG.ATC....T..GA.....CA.A.AC.....T.....T.....CC..-.AC.

Chimpanzee CCTCTAGAGGAGCCTGTTCTGTAATCGATAAACCCCGATCAACCTCACCGCCTCTTGCT----CAGCCTATATACCGCCATCTTCAGCAAACCCTGATGAAGG
Human .....A.....----.....
Gorilla .....A.....----.....
Orangutan ..C.....A..C.....----.....
TCR numtDNA .....A.....A.....A.T.....----.....A.....A..A.
Gr. monkey ..C.....C.CA.....C...C...CT.....----.....G.....A....
Mandrill ..C.....TCA.....C...CT...CT.....----.....
Baboon ..C.....A.....C...C...CT.....----T.T.....
Colobus ..C.....A.....C...CT...AT....C.....----.....-..C.AC...A
Langur ..C.....A.....C...CT...AT....C.....----.....-..C.AC...A
Cow .T.....A.....A.....AAT.....AATA...T.....A.---.A.
Seal .....A.....ATTC.....AATA...T.....T.---.A.
Rat .A.....A.....T..T....T...C.T...C...AATT.....A.---.A.
Chicken .AC.....A.....T..A....TC...CA...A..C....CAGCA.....C.....G..GC....CC...TCT.ATG.AA

Chimpanzee TT-ACAAAGTAAGCACAAGTACCC-----ACGTAAAGACGTTAGGTCAAGGTGTAGCCTATGAGGTGGCAAGAAATGGGCTACATTTTCTACCC-CAGAAAAT
Human C.-.....G.....C.....-.....C
Gorilla CC-.....C.....TT-.....C
Orangutan CC-..G.....G...AC.....C...GA.....TT-.....C
TCR numtDNA C-.....T.....A...A.....C.....C..T.....A.C.....
Gr. monkey .A.....G...G.....G-..CCTTTT..C...A.....A...A..A.G.....T..-T.....C
Mandrill C.-.....G...G...A.G--.CCCCGC..C...A.....C.....AC..T..A.....T-.....C
Baboon C.-..G....G...G...A.G-..CTTCCC..C...A.....A.....AC..T-.....TT-.....C
Colobus GACC.....ACG.....C...A.....T.....A...G-.....TA.-.....
Langur GAT.TG.....A.G.....C...A.....C...G..A.....TTA-.....
Cow GAA.A.....GT..T..T-GA---T..A...A.....A.....AA...G.....C....A.CA...G...
Seal GAACA.....TA.T.GC---T..A...A.A.....A.....GAA...G.....AAT-A....C.A
Rat GCACT.....A.-.AA-----.A...A.....A...A.C..A.....TTT.C....G..C
Chicken GAACA.C...G...T...TAG...C----T..CT..T.A.AC.....A.....G.....G-.....A.A-T....C.A

Chimpanzee ----TACGATAACCCTTATGAAACCTAAGGGTCAAAGGTGGATTTAGCAGTAAACTAAGAGTAGA------GTGCTTAGTTGAACAGGGCCCTGAAGCGCGT
Human ----.....G.....T.....G.....------.....
Gorilla ----.....AG.....------.....
Orangutan ----.....C.....TT.G.A...G.....------.....A.....
TCR numtDNA ----.....C.....T.....TT.....A.....T.....C.....------.....A.....TGA...A.A...A.....
Gr. monkey C---C.....T..C.....A..T...A.....T...A.....------.....A.....GA.....G.A...A..C
Mandrill CC--.....T.T.....T.....A.C.C....A.....T.....A.....------.....GA.....TA.....A.....
Baboon TC--C.....G...C.....G.....C....A.....T.....A.....------.....A.....TA.....A.A....A..C
Colobus CC--C.....G.T.....-T..G.A.C.T....A.....T.....T.....A.....------.....A.....TTA.....A.....AT.C
Langur C---.....T.....T....A...C....A.....A.....------.....TA.....A.....AT.C
Cow CAAGC....A.GTTA.....A.T-AAC.....A.....A.....------.....TTA.....A.....A..C
Seal TCA-.....A.GTTT.....TTA.CAAACT.....T.....G.....A.....------.A.....C.....C.....A.....A..C
Rat --AT.....-....T.....TA....AC-.....A.....T.....T.....A.....------.A.....A.....T..A..AA.....TA..C
Chicken -------...A..AGGA.G.....CGCCCT.AG....A.....G.G..TC.T.CCCCCTAA...C.C..T..G.C...T....G..A...

FIG. 4.—(*a*) Maximum parsimony consensus tree of the 12S rRNA/tRNA-Val/16S rRNA mtDNA region (2,649 sites) including available primate mtDNA. (*b*) Shortest parsimony tree for a 12S rRNA sequence from which DNA sequence is available for New and Old World monkey (411 sites). Trees were constructed by performing heuristic (tree bisection and reconnection) searches using the phylogenetic analysis program PAUP (Swofford 1993). Inferred number of nucleotide changes are shown along each branch. Bootstrap values (generated in PAUP, 100 replications) are shown below the branches. (*c*) Alignment of the 12S rRNA region. Circles enclose nucleotide changes along the TCR numtDNA branch identified in the parsimony analysis. Shaded circles indicate sites where nucleotide changes have occurred in TCR numtDNA at sites invariant among these animals. Periods indicate an identical nucleotide and dashes indicate alignment gaps. The species names, common names, and GenBank accession number for the listed genera are: chimpanzee, *Pan troglodytes,* D38113; human, *Homo sapiens,* J01415; gorilla, *Gorilla gorilla,* D38114; orangutan, *Pongo pygmaeus,* D38115; green monkey, *Cercopithecus neglectus,* L35182; mandrill, *Mandillus spinx,* L35196; baboon, *Papio ursinus,* L35206; colobus, *Colobus guerza,* L35195; langur, *Presybytis crisatus,* L35200; cow, *Bos taurus,* J01394; seal, *Phoca vitulina,* X72004; rat, *Rattus novegicus,* X14848; chicken, *Gallus gallus,* X52392.

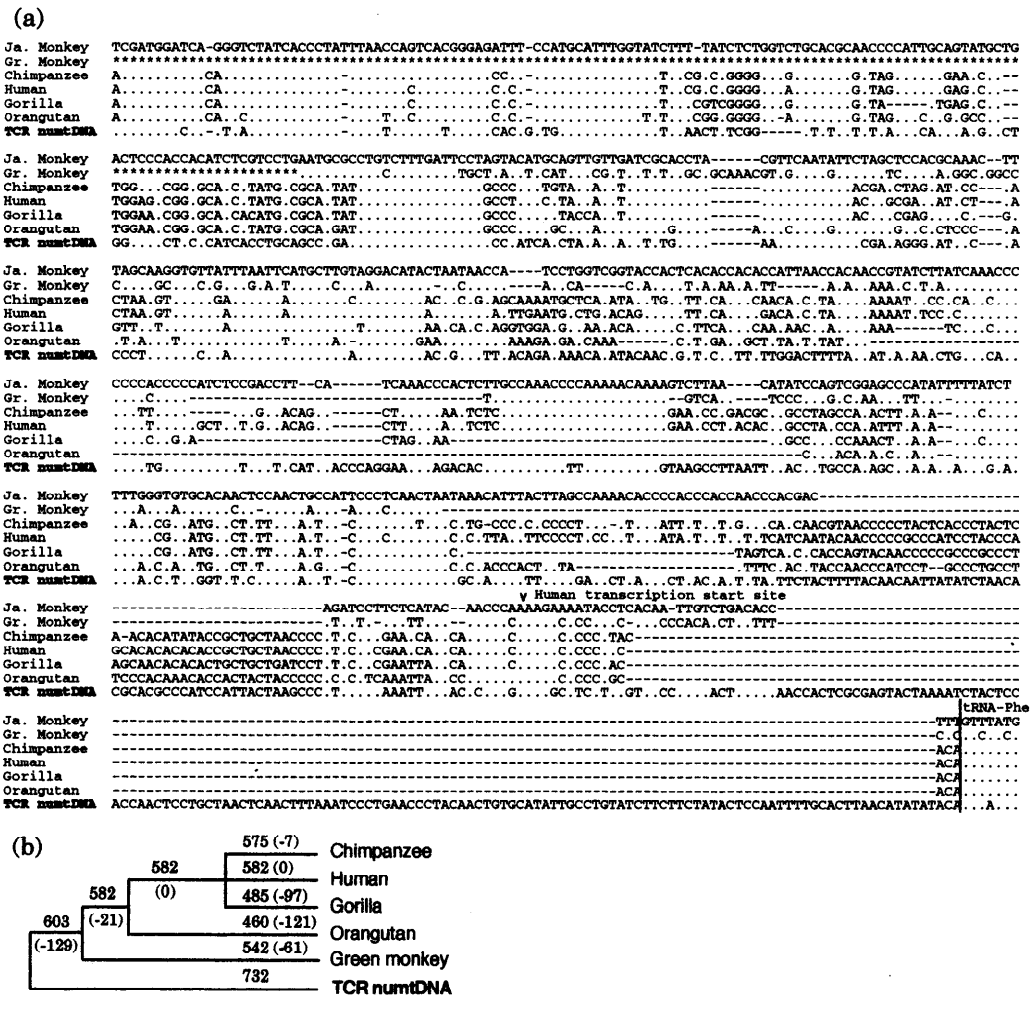


FIG. 5.—(*a*) Alignment TCR numtDNA with available primate D-loop sequences. This alignment does not encompass the complete D-loop region. The portion of the green monkey D-loop sequence that has not been reported is represented by asterisks. The human transcription start site was identifed by Chang and Clayton (1984). The species names and GenBank accession number for the given sequences are; Japanese monkey, *Macaca fuscata,* M80209; green monkey, *Cercopithecus aethiops,* M25659; all others the same as in figure 4. (*b*) Reconstruction of the D-loop region shown above during primate evolution by parsimony. The nucleotide length of this region is shown along the branches with parentheses indicating the length difference.

human/chimp/gorilla clade (table 3). This difference is due to an increase in transitions with the number of transversions being remarkably similar among the four taxa. This pattern is also supported using the numtDNA sequence reported by Nomiyama et al. (1985). Phylogenetic information from the smaller 12S rRNA region (fig. 4*B,C*) is not sufficient to confidently determine the transfer node relative to New and Old World monkeys, although this limited data set (table 3) shows a constant number of transitions among Old World monkeys (green monkey, mandrill, and baboon) and the great apes. Other numtDNA sequences are too short to place in a phylogenetic context or, in the case of *S. cerevisiae,* suitable sequences from closely related fungi have not been reported.

Mitochondrial Sequences in cDNA Libraries

Twenty-six mitochondrial-derived regions are found at the ends of cDNA sequences and are nearly identical to mtDNA. For many of these cDNAs an additional cDNA sequence or genomic sequence lacking the mitochondrial fragment has been filed in GenBank (table 2). These chimeric sequences appear to have arisen during the cloning process through ligation of the mitochondrial and nuclear transcripts through complementary base-pair interactions (data not shown). As expected none of the cDNA clones contain nontranscribed regions of mtDNA. Mitochondrial fragments in cDNA clones that have ends similar to the ends of mitochondrial transcripts are listed in table 2.

Table 3
Nucleotide Difference between TCR numtDNA and mtDNA Sequences

Sequence compared	TCR locus numtDNA* 2,649 sites				TCR locus numtDNA† 411 sites			
	All	CT	AG	V	All	CT	AG	V
Chimpanzee	395	166	104	125	57	26	21	10
Human	399	168	109	122	60	28	22	10
Gorilla	400	164	112	124	59	27	21	11
Orangutan	456	202	132	122	69	32	27	10
Green monkey ..	—	—	—	—	67	36	22	9
Mandrill	—	—	—	—	62	30	22	10
Baboon	—	—	—	—	62	30	23	9
Colobus	—	—	—	—	68	35	19	14
Langur	—	—	—	—	68	34	18	16
Cow	610	174	142	294	63	17	21	25
Seal	612	179	139	294	82	25	23	34
Rat	681	164	144	373	63	23	18	22
Chicken	904	248	186	470	126	35	25	66

Abbreviations: CT = cytosine ↔ thymine changes; AG = adenine ↔ guanine changes; V = transversions; — indicates this sequence has not been reported.

* From tRNA-Phe to the end of the insertion.

† Region shown in figure 4.

Mitochondrial sequences not at the 5′ or 3′ ends of cDNAs also show characteristics of artifactual chimeras. Both the frizzled gene and the two protein-tyrosine phosphatase genes contain a sequence identical to commercially available ECORI linkers used in constructing cDNA libraries (fig. 6). Alternative cDNA sequences lacking the mitochondrial fragments have been reported for both the stromal cell-derived protein and protein-tyrosine phosphatase.

In contrast to the previously mentioned cDNAs, the mtDNA fragment found in the cDNA for the rat guanine

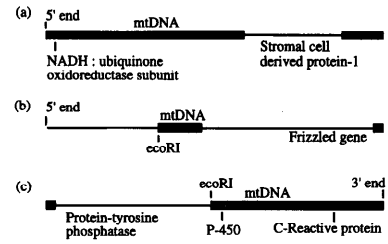


FIG. 6.—Location of the mitochondrial-derived region within four cDNA clones. (*a*) Mouse stromal cell-derived protein-1 (D16847). (*b*) Rat frizzled gene (L02529). (*c*) Rat protein-tyrosine phosphatase genes (L11587) and (L12329). Black, gray, and striped boxes represent, respectively, the nuclear gene, mtDNA, and a fragment from a different nuclear gene.

nucleotide releasing protein (GNRP) has diverged substantially from rat mtDNA (fig. 7*B*). Parsimony analysis indicates that the migration event occurred soon after the divergence of rats and mice (results not shown). Transcripts of 1.5 kb and 1.7 kb (fig. 7*A*) corresponding to the two GNRP cDNAs are observed in several different rat tissues (Burton et al. 1993). However, Burton et al. (1993) did not determine whether the transcripts are derived from alternative splicing or from duplicated genes.

Discussion

Frequency of Organellar DNA Transfer

During our study of organellar DNA migration in plants (Blanchard and Schmidt 1995), we recognized the possibility that cDNA sequences could be artifactual chimeras between nuclear and organellar DNA and thereby were excluded from the analysis. Chimeric cDNAs have been previously reported to be derived from the cloning process (Claverie and Makalowski

FIG. 7.—(*a*) Comparison of the two rat GNRP cDNAs sequenced by Burton et al. (1993). The striped, gray, and white boxes represent, respectively, a region of nuclear DNA, numtDNA, and the nuclear GNRP gene. (*b*) Alignment of the numtDNA fragment in the rat GNRP with rat mtDNA. The dashes indicate gaps in the alignment. The triangle marks the first nucleotide of a shorter cDNA clone (Burton et al. 1993). The methionine (met) represents the putative translation start site.

1994). The quantity of chimeric constructs can vary in accordance with the cDNA library and appears to be dependent on the cloning strategy (A. R. Kerlvage, personal commnication). In addition, expressed sequence tag (EST) projects have deposited over 250 human and over 80 *C. elegans* mitochondrial sequences in the database (data not shown). Nearly all of these sequences were identical to mtDNA, indicating mitochondrial transcripts are present in cDNA libraries. A subset of the human cDNA sequences shown in table 2 was also recently independently identified by Wegner and Gassman (1995), who also commented on their probable artifactual origin. In the present study, only the cDNA of the rat GNRP gene with a 5′ mtDNA fragment is likely to have arisen from DNA migration because of the numerous nucleotide differences from the rat mitochondrial sequence (fig. 7). In two previous reports of chimeric mitochondrial/nuclear transcripts, both in human cell lines (Koch, Hofschneider, and Koshy 1989; Shay et al. 1991), the mitochondrial fragment is at the 5′ end of the cDNA clone.

Although artifactual chimeras of mitochondrial and nuclear DNA are clearly present in GenBank, the computer-based search strategy described in this study has been successful at identifying DNA transfer events in mammals and yeast as well as those previously reported in plants (Blanchard and Schmidt 1995). Because 4%–7% of plant genomic sequence files contain organellar-derived DNA (Blanchard and Schimdt 1995), the paucity of detectable mtDNA insertions in this report is surprising. However, a direct comparison cannot be made between the composition of humans and plant genomes because the amount of human genomic DNA in the database was not quantified. Although, one might expect there to be more human than plant genomic sequences in GenBank.

The analysis of the seven *S. cerevisiae* chromosomes detected only 1,098 bp of mitochondrial-derived DNA in 3,583,808 bp of nuclear DNA (0.031%). However, the actual amount could be slightly higher because sequences similar to AT repeats in the mitochondrial IG/spacer regions were not described as numtDNA sequences. Thorsness and Fox (1990) measured, via transfection of *S. cerevisiae* mitochondria with a plasmid containing a selectable *uridA* marker, that the plasmid escaped from the mitochondrion and integrated into the nuclear genome at a rate of one transfer event per 10^5 generations. Because mitochondrial integration events would in most cases be expected to be neutral or deleterious, the observed discrepancy may reflect differences in the fixation of a neutral or deleterious mtDNA fragment versus selection for an advantageous plasmid integration event. At the present time there are insufficient sequence data to make quantitative comparisons between taxa of the amount of DNA of mitochondrial origin in the nucleus. Because no evidence could be detected for mtDNA insertions into chromosome III of *C. elegans* or in nuclear DNA sequences of *Drosophila* sp. or *P. falciparum,* there appears to be more detectable organellar-derived sequences in the plant sequence database than in other groups. A difference in the quantity of organellar DNA in the nucleus between taxa has been previously suggested based on Southern blot data (Gellissen and Michaelis 1987).

Mechanisms of DNA Migration

From this broad survey of DNA migration events, we have identified a few cases from which we can make inferences about the transfer process. DNA-based transfer is supported by the movement of one nontranscribed region in yeast and one in humans. However, a mitochondrial-derived segment in the human microsatellite sequence offers support for RNA-based transfer. Previously, RNA-based transfer has been reported only in plants (Nugent and Palmer 1991; Grohmann, Brennicke, and Schuster 1992; Blanchard and Schmidt 1995). Lopez et al. (1994) suggest that mtDNA may form extrachromosomal episomes that can be tandemly amplified subsequent to their integration into the nuclear genome. Although their data are solid support for the occasional involvement of episomes in the transfer process, the numtDNA sequences in the yeast and human databases do not show signs of circularization prior to their integration.

The mtDNA integration process appears to be independent of retroelement insertion (figs. 1, 2, and 3) in contrast to previous suggestions based on the proximity of retroelements to numtDNA sequences (Farrelly and Butow 1983; Ossario, Sibley, and Boothroyd 1991; Zullo et al. 1991). The common observation of organellar fragments near retroelements is to be expected based on the near ubiquitous presence of retroelements in noncoding regions. The characteristics surrounding the insertion junctions (fig. 2) are similar to those described for end-joining processes that require limited or no sequence homology between the foreign DNA and chromosomal DNA but often involve free single-stranded ends (Roth and Wilson 1988). In figure 2*A* and *B* the 4–5-bp direct repeats are only apparent in the flanking regions of the numtDNA after the integration occurred, suggesting a step involving single-stranded ends. Mitochondrial DNA then could have been ligated to the chromosomal ends using limited base complementarity (0–3 bp). Roth and Wilson (1988) suggest that free chromosomal ends are generated from errors in DNA metabolism and at these sites segments of foreign DNA can integrate regardless of their terminal sequences. During this process, end-joining of multiple DNA fragments can occur, as may have happened with the (1)

three regions containing multiple mitochondrial fragments in yeast (fig. 1); (2) the multiple fragments reported by Farrelly and Butow (1983); and (3) in the three noncontiguous mitochondrial segments found in human DNA (Kamimura et al. 1989). This process has also been observed in plants (Sun and Callis 1993; Blanchard and Schmidt 1995) and suggests that organellar DNA may integrate into the nuclei of plants, yeast, and mammals by similar end-joining mechanisms.

Patterns of Nucleotide Substitution Inferred from the Human TCR Locus

Nuclear changes in the TCR numtDNA sequence were inferred from parsimony analysis using PAUP (Swofford 1993) and by calculating the percent change at nucleotides invariant among the taxa in figure 4*A* and *B*. These two estimates are identical in the shorter 12S rRNA region (6.0%), but a small difference was observed over the entire 12S/16S region (7.9% vs. 5.7%). Over the larger region no monkey sequences were available to include in the analysis. This may have resulted in a less accurate reconstruction of the TCR numtDNA branch point using PAUP, thereby increasing the length of the TCR branch and decreasing the mitochondrial branch to the great ape node.

The numtDNA fragment in the TCR locus harbors a 127-bp region that is not similar to any reported mitochondrial or nuclear sequence (fig. 5*A*). Enlarged mitochondrial D-loops have often been associated with the tandem duplication of mtDNA (Hayasaka, Ishida, and Horai 1991), but this 127-bp TCR numtDNA segment is not similar to any other region of mtDNA. A mitochondrial origin rather than a subsequent nuclear insertion is inferred for this sequence based on a nucleotide composition characteristic of mtDNA. The parsimony analysis also supports the possibility that this sequence may have been in an ancestor of great ape and monkey clades (fig. 4*B*). Apparently a reduction in D-loop size is a general trend in primate mtDNA evolution, at least within both Old World monkeys and great apes (fig. 5*B*). This trend is in agreement with studies using fungi and on human mitochondrial diseases that report a replicative advantage to a smaller mitochondrial haplotype (Wallace 1994). However, it is also possible that deletions are generated more frequently than insertions in the D-loop region and the observed trend is reflective of a more neutral fixation process.

An interesting perspective on the nucleotide substitution patterns is also provided by polarizing relative distances within the great ape clade using numtDNA (table 3). The data support an analysis by Hasegawa et al. (1990) that suggests variation in transition rates among primates but a constant rate of transversions. However, by comparing a 4.8-kb region of mtDNA from chimpanzee, human, gorilla, orangutan, and siamang, Horai et al. (1992) suggested an increase has occurred in both transition and transversion rates in the orangutan lineage relative to the human/chimp/gorilla clade. Comparison of complete mitochondrial genomes from great apes shows that transition and transversion rates vary between rRNA, tRNA, and protein-coding genes as well as among sites within genes (Horai et al. 1995). Therefore, other areas of the mitochondrial genome need to be more thoroughly investigated to determine whether the relative number of transition substitutions viewed by the TCR numtDNA is a local phenomenon restricted to the 12S/16S rRNA region.

Conclusions

The results of this study complete a survey of eukaryotic groups for migratory organellar DNA sequences that began with analysis of the plant genomic DNA database. This computer-based strategy offers advantages over a clone-based strategy in that: searches can be rapidly done across taxonomic boundaries; short moderately divergent regions are detected that would be difficult to discern from random noise signals on Southern blots; the sequence files often contain the complete insert; and insertions are more likely to be found in relatively better studied gene families, thus facilitating a comparative analysis. The data compiled represent a threefold increase in the number of transfer events that have been identified and characterized at the DNA sequence level. From this large data set a number of interesting mechanistic and evolutionary processes are inferred. All transferred sequences were found in noncoding regions and because most numtDNA are gene fragments they are unlikely to be intermediates in a gene transfer process. The quantity of organellar DNA in the nucleus appears to vary considerably amongst organisms and is much lower in *S. cerevisiae* than suggested by experimental plasmid systems. Tandemly arranged fragments from disparate regions of the mitochondrial genome indicate that fragments join together from an intracellular pool of RNA and/or DNA before they integrate into the nuclear genome. Comparisons of integrated sequences to genes lacking the insertions, as well as the occurrence of coligated fragments, support a model of random integration by end-joining common to plants, animals, and fungi. Finally, in addition to illuminating intracellular DNA transfer processes, the nuclear-localized mitochondrial sequences provide a unique time-line for reconstructing mitochondrial genomic evolution as evidenced by support for a concomitant reduction of D-loop size within primates.

Acknowledgments

We thank the following scientists for their helpful discussions and suggestions: Mary Hagen, Shawn White, Tom Bureau, Liz Bachman, Matt Hare, Aleksander Popadic, Kurt Wollenberg, Chris Babcock, Russell Malmberg, Masami Hasegawa, Robert Price, and John Avise. We also gratefully acknowledge the thoughtful criticism provided by two anonymous reviewers.

LITERATURE CITED

ADKINS, R. M., and R. L. HONEYCUTT. 1994. Evolution of the primate cytochrome oxidase subunit II gene. J. Mol. Evol. **38**:215–231.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

AYLIFFE, M. A., and J. N. TIMMIS. 1992. Plastid DNA sequence homologies in the tobacco nuclear genome. Mol. Gen. Genet. **236**:105–112.

BERNATZKY, R., S.-L. MAU, and A. E. CLARKE. 1989. A nuclear sequence associated with self-incompatibility in *Nicotiana alata* has homology with mitochondrial DNA. Theor. Appl. Genet. **77**:320–324.

BLANCHARD, J. L., and G. W. SCHMIDT. 1995. Pervasive migration of organellar DNA to the nucleus in plants. J. Mol. Evol. **41**:397–406.

BURTON, J., D. ROBERTS, M. MONTALDI, P. NOVICK, and P. DECAMILLI. 1993. A mammalian guanine-nucleotide-releasing protein enhances function of yeast secretory protein Sec4. Nature **361**:464–467.

CHANG, D. D., and D. A. CLAYTON. 1984. Precise identification of individual promoters for transcription of each stand of human mitochondrial DNA. Cell **19**:27–35.

CLAVERIE, J.-M., and W. MAKALOWSKI. 1994. Alu alert. Nature **371**:752.

CLAYTON, D. A. 1992. Transcription and replication of animal mitochondrial DNAs. Int. Rev. Cytol. **141**:217–232.

DEVEREAUX, J., P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.

DE ZAMAROCZY, M., and G. BERNARDI. 1986. The primary structure of the mitochondrial genome of *Saccharomyces cerevisae*—a review. Gene **47**:155–177.

DU BUY, H. G., and F. L. RILEY. 1967. Hybridization between the nuclear and kinetoplast DNA's of *Leishmania enriettii* and between nuclear and mitochondrial DNA's of mouse liver. Proc. Natl. Acad. Sci. USA **57**:790–797.

FARRELLY, F., and R. A. BUTOW. 1983. Rearranged mitochondrial genes in the yeast nuclear genome. Nature **301**:296–301.

FELSENSTEIN, J. 1994. PHYLIP manual version 3.5. University of Washington, Seattle, WA.

FUKUCHI, M., T. SHIKANAI, V. G. KOSSYKH, and Y. YAMADA. 1991. Analysis of nuclear sequences homologous to the B4 plasmid-like DNA of rice mitochondria: evidence for sequence transfer from mitochondria to nuclei. Curr. Genet. **20**:487–494.

FUKUDA, M., S. WAKASUGI, T. TSUZUKI, H. NOMIYAMA, and K. SHIMADA. 1985. Mitochondrial DNA-like sequences in the human nuclear genome. J. Mol. Biol. **186**:257–266.

GELLISSEN, G., J. Y. BRADFIELD, B. N. WHITE, and G. R. WYATT. 1983. Mitochondrial DNA sequences in the nuclear genome of a locust. Nature **301**:631–634.

GELLISSEN, G., and G. MICHAELIS. 1987. Gene transfer: mitochondria to nucleus. Ann. N.Y. Acad. Sci. **503**:391–401.

GROHMANN, L., A. BRENNICKE, and W. SCHUSTER. 1992. The mitochondrial gene encoding ribosomal protein S12 has been translocated to the nuclear genome in *Oenothera*. Nucleic Acids Res. **20**:5641–5646.

HASEGAWA, M., H. KISHINO, K. HAYASAKA, and S. HORAI. 1990. Mitochondrial DNA evolution in primates: transition rate has been extremely low in lemur. J. Mol. Evol. **31**:113–121.

HAYASAKA, K., T. ISHIDA, and S. HORAI. 1991. Heteroplasmy and polymorphism in the major noncoding region of mitochondrial DNA in Japanese monkeys: association with tandemly repeated sequences. Mol. Biol. Evol. **8**:399–415.

HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE, and N. TAKAHATA. 1995. Recent African origin of modern humans revealed by compete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA **92**:532–536.

HORAI, S., Y. SATTA, K. HAYASAKA, R. KONDO, T. INOUE, T. ISHIDA, S. HAYASHI, and N. TAKAHATA. 1992. Man's place in hominoidea revealed by mitochondrial DNA genealogy. J. Mol. Evol. **35**:32–43.

KAMIMURA, N., S. ISHII, M. LIANDONG, and J. W. SHAY. 1989. Three separate mitochondrial DNA sequences are contiguous in human genomic DNA. J. Mol. Biol. **210**:703–707.

KOCH, I., P. H. HOFSCHNEIDER, and R. KOSHY. 1989. Expression of a hepatitis B virus transcript containing fused mitochondrial-like domains in human hepatoma cells. Virology **170**:591–594.

LOPEZ, J. V., N. YUHKI, R. MASUDA, W. MODI, and S. J. O'BRIEN. 1994. *NUMT*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol. **39**:174–190.

LOUIS, E. J., and J. E. HABER. 1991. Evolutionarily recent transfer of a group I mitochondrial intron to telomere regions in *Saccharomyces cerevisiae*. Curr. Genet. **20**:411–415.

MADISON, W. P., and D. R. MADISON. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer Associates, Sunderland, MA.

NOMIYAMA, H., M. FUKUDA, S. WAKASUGI, T. TSUZUKI, and K. SHIMADA. 1985. Molecular structures of mitochondrial DNA-like sequences in human nuclear DNA. Nucl. Acids Res. **13**:1649–1658.

NUGENT, J. M., and J. D. PALMER. 1991. RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell **66**:473–481.

OSSARIO, P. N., D. L. SIBLEY, and J. C. BOOTHROYD. 1991. Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. J. Mol. Biol. **22**:525–536.

PICHERSKY, E., and S. D. TANKSLEY. 1988. Chloroplast DNA sequences integrated into an intron of a tomato nuclear gene. Mol. Gen. Genet. **215**:65–68.

QUINN, T. W., and B. N. WHITE. 1987. Analysis of DNA sequence variation. Pp. 163–198 *in* F. COOKE and P. A. BUCKLEY, eds. Avian genetics. Academic Press, London.

RICHARDS, O. C. 1967. Hybridization of *Euglena gracilis* chloroplast and nuclear DNA. Proc. Natl. Acad. Sci. USA **57**: 153–163.

ROTH, D., and J. H. WILSON. 1988. Illegitimate recombination in mammalian cells. Pp. 621–653 *in* R. KUCHERLAPATI and G. SMITH, eds. Genetic recombination. American Society for Microbiology, Washington, DC.

RUVOLO, M. 1993. Molecular evolutionary processes and conflicting gene trees: the hominoid case. Am. J. Phys. Anthropol. **94**:89–114.

SHAY, J. W., T. BABA, Q. ZHAN, N. KAMIMURA, and J. A. CUTHBERT. 1991. HeLaTG cells have mitochondrial DNA inserted into the *c-myc* oncogene. Oncogene **6**:1869–1874.

SMITH, M. F., W. K. THOMAS, and J. L. PATTON. 1992. Mitochondrial DNA-like sequences in the nuclear genome of an akodontine rodent. Mol. Biol. Evol. **9**:204–215.

SUN, C.-W., and J. CALLIS. 1993. Recent stable insertion of mitochondrial DNA into an *Arabidopsis* polyubiquitin gene by nonhomologous recombination. Plant Cell **5**:97–107.

SWOFFORD, D. L. 1993. Phylogenetic analysis using parsimony (PAUP). Version 3.1.1. Illinois Natural History Survey, Champaign, IL.

THORSNESS, P. E., and T. D. FOX. 1990. Escape of DNA from mitochondrial to the nucleus in *Saccharomyces cerevisiae*. Nature **346**:376–379.

VAN DEN BOOGAART, P., J. SAMALLO, and E. AGSTERBIBE. 1982. Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondrial genomes of *Neurospora crassa*. Nature **298**:187–189.

VAN DER KUYL, C. A., C. L. KUIKEN, J. T. DEKKER, and J. GOUDSMIT. 1995. Phylogeny of African monkeys based upon mitochondrial 12S rRNA sequences. J. Mol. Evol. **40**: 173–180.

WALLACE, D. C. 1994. Mitochondrial DNA sequence variation in human evolution and disease. Proc. Natl. Acad. Sci. USA **91**:8739–8746.

WEGNER, R. H., and M. GASSMAN. 1995. Mitochondria contaminate databases. Trends Genet. **11**:167–168.

ZULLO, S., C. S. LEANG, J. L. SLIGHTON, H. I. HADLER, and J. M. EISENSTADT. 1991. Mitochondrial D-loop sequences are integrated in the rat nuclear genome. J. Mol. Biol. **221**: 1223–1235.

BARBARA A. SCHAAL, reviewing editor

Accepted December 5, 1995