J Mol Evol (1991) 32:396–404

Journal of
Molecular Evolution

Polymorphisms in Tandemly Repeated Sequences of *Saccharomyces cerevisiae* Mitochondrial DNA

P.J. Skelly and G.D. Clark-Walker

Molecular and Population Genetics Group, Research School of Biological Sciences,
Australian National University, GPO Box 475, Canberra, ACT 2601, Australia

Summary. A spontaneously arising mitochondrial DNA (mtDNA) variant of *Saccharomyces cerevisiae* has been formed by two extra copies of a 14-bp sequence (TTAATTAAATTATC) being added to a tandem repeat of this unit. Similar polymorphisms in tandemly repeated sequences have been found in a comparison between mtDNAs from our strain and others. In 5850 bp of intergenic mtDNA sequence, polymorphisms in tandemly repeated sequences of three or more base pairs occur approximately every 400–500 bp whereas differences in 1–2 bp occur approximately every 60 bp. Some polymorphisms are associated with optional G+C-rich sequences (GC clusters). Two such optional GC clusters and one A+T repeat polymorphism have been discovered in the tRNA synthesis locus. In addition, the variable presence of large open reading frames are documented and mechanisms for generating intergenic sequence diversity in *S. cerevisiae* mtDNA are discussed.

Key words: Yeast—Mitochondrial DNA—Polymorphism—Repeated sequences

Introduction

Some years after experiments commenced on the mitochondrial genetics of baker's yeast a logical proposal was made to adopt a single strain for all studies. Fortunately for the advancement of knowledge on the evolution and plasticity of mitochondrial DNA (mtDNA) this suggestion was not accepted. As a consequence of using different strains we now know that the mitochondrial genome of *Saccharomyces cerevisiae* can harbor a variety of optional sequences such as introns, open reading frames, intergenic A+T-rich regions, and GC clusters (de-Zamaroczy and Bernardi 1985, 1986a,b; Seraphin et al. 1987). Attention has now shifted to discovering how such structures are acquired or lost. In this context we were curious to discover the nature of a restriction fragment length polymorphism (RFLP) that arose spontaneously in the mitochondrial genome of a yeast strain studied in this laboratory. One stock, cultured for 2 yr, had gained approximately 50 bp of sequence in a 2.15-kb CfoI fragment relative to the original stock maintained in liquid nitrogen. This polymorphism has been used as a marker in a study of conversion in *S. cerevisiae* mtDNA (Skelly and Clark-Walker 1990). By analogy with other variations in mtDNA, a gain of approximately 50 bp could be due to the acquisition of a GC cluster. These reportedly mobile elements range in size from 20 to 50 bp and are responsible for polymorphisms in the *var1* gene, in small rRNA and large rRNA genes, and in numerous places in intergenic regions (Dujon 1980; Sor and Fukuhara 1982; Hudspeth et al. 1984; Weiller et al. 1989). However, to our surprise, sequence determination has revealed that the spontaneous change arose from the acquisition of two repeats of 14 bp added to a multiple tandem repeat of TTAATTAAATTATC. This discovery led us to seek other polymorphisms and in particular those involving tandemly repeated elements by comparison of our strain to the published consensus sequence.

Offprint requests to: P. J. Skelly

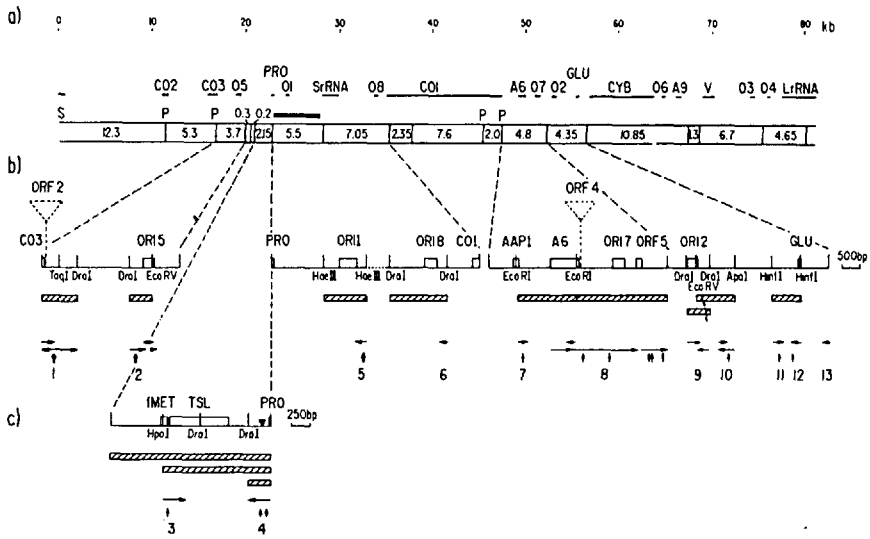

Fig. 1. Map of wild-type *S. cerevisiae* mtDNA (a) and detailed map of selected fragments (b, c). The boxes in the upper map show the position and size of fragments generated by CfoI–PvuII digestion, P identifies PvuII sites and S the unique SalI site in the large rRNA gene. The filled box above the map shows the position of the 5.0-kb deletion. Lines above the map show positions of genes identified by: CO 1–3, subunits 1–3 of cytochrome oxidase; S and L rRNA, small and large subunits of rRNA; A6 and 9, ATPase subunits 6 and 9; CYB, apocytochrome b; V, variant 1 protein. Also shown are positions of *ori/rep* sequences, O 1–8, and tRNAs glu and pro. Maps b and c additionally show the location of the tRNA synthesis locus, TSL, the ATPase subunit 8 gene, AAP1, and the fMet tRNA gene. The positions of the optional ORFs 2 and 4 are indicated. Cloned fragments are shown as hatched boxes below the detailed maps and arrows show the extent and direction of determined sequences (nos. 1–13). ↑ indicates the presence of 3-bp or larger A+T sequence polymorphisms. Map c depicts the polymorphic 2.15-kb CfoI fragment. The position of the 14-mer polymorphism is indicated by ▼.

Materials and Methods

Yeast Strains. Strain Dip 2 has been derived from mating D13.1a (α, rho$^+$, his3-532, trp1) with T3/3 (a, rho$^+$, ade1, arg4-16) (Evans et al. 1985). Strain Δ5.0 (a, rho$^+$−Δ5, ade1, his3-532) lacks a 5-kb mtDNA segment extending between the genes for proline tRNA and the small rRNA gene (Skelly and Clark-Walker 1990). This deletion mutant was obtained following a series of in vivo steps involving the crossing of spontaneously arisen petite mutants and the subsequent isolation of respiratory competent zygotes (Evans and Clark-Walker 1985). These zygotes produce petites at a high frequency and contain a direct duplication in the mitochondrial genome. Revertants to a normal level of petite production have lost one duplication and possess a smaller mitochondrial genome than wild type. Strain Δ5.0/a$^+$ is a spontaneous variant of Δ5.0 that arose following subculture of the strain over the course of 2 yr. Strain Δ5.0/a$^+$ contains a mtDNA RFLP due to an extra sequence in a 2.15-kb CfoI fragment. This larger fragment is referred to as the 2.2-kb CfoI fragment.

Mitochondrial DNA Isolation and Electrophoresis. Spheroplasts, produced by zymolyase, were broken in a French press and mtDNA was isolated by dye-bouyant density centrifugation using bisbenzamide and CsCl. Restriction endonuclease digestion and electrophoretic separation of fragments on 0.8% agarose gels were carried out as previously described (Evans et al. 1985) except that potassium glutamate buffer was used for all digestions (McClelland et al. 1988).

Cloning and Sequencing. Figure 1 shows a CfoI/PvuII restriction digestion map of wild-type strain Dip 2 mtDNA. The location of the 5.0-kb deletion in strain Δ5.0 is indicated. Fragments examined in this work and the sequenced regions are indicated (Fig. 1b). Sequences were obtained from 13 locations in the mitochondrial genome (numbered 1–13 in Fig. 1). Mitochondrial DNA fragments were isolated from agarose gels and cloned into the vectors pTZ18 or pTZ19 using standard techniques (Maniatis et al. 1982). The precise cloning protocols for many of the fragments used in this work have been described in detail elsewhere (Clark-Walker 1989; Skelly and Clark-Walker 1990). The 2.15-kb CfoI fragment in Δ5.0 and the 2.2-kb fragment in the aberrant strain were isolated, blunt ended, and cloned into the SmaI site of pTZ18. The 0.7-kb CfoI/HpaI (Fig. 1c) region common to both fragments was subsequently removed from the clones. Mapping studies on the remaining segments of mtDNA revealed that the polymorphism is present at the right-hand end of the fragment as drawn in Fig. 1c. Therefore the terminal DraI/CfoI fragment from both strains was cloned and sequenced. All sequencing was carried out by the dideoxy chain termination procedure (Sanger et al. 1977) with [^{35}S]dATP and Sequenase (United States Biochemicals, OH).

Attempts to Induce Mitochondrial Genome Polymorphisms. In an attempt to generate further polymorphisms in the mitochondrial genome, strain Dip 2 was subjected to continuous growth for 100 generations at the elevated temperatures of either 34°C or 37°C on nonfermentable GlyYP medium. GlyYP contains 2% glycerol, 0.5% yeast extract, and 1% Bacto peptone. Cells were plated on GlyYP solidified with 1.5% agar and incubated at 30°C for 5 days. Colonies were selected at random and total cellular DNA was isolated as described (Skelly and Clark-Walker 1990). Examination of mtDNA was performed on CfoI/PvuII digests

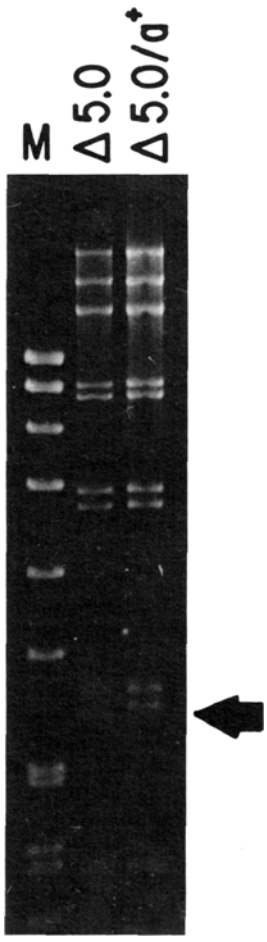

Fig. 2. CfoI digestion patterns of mtDNAs from strains Δ5.0 and Δ5.0/a+. Lane M is an EcoRI digest of phage Spp1 DNA showing size markers of (from top) 7.8, 7.0, 5.9, 4.7, 3.4, 2.7, 1.9, 1.8, 1.5, 1.3, and 1.1 kb. The arrow indicates the fragment with changed mobility.

of total DNA that had been electrophoretically separated on 0.8% agarose, transferred to nylon, and hybridized with ^{32}P-labeled mtDNA purified by CsCl-gradient centrifugation.

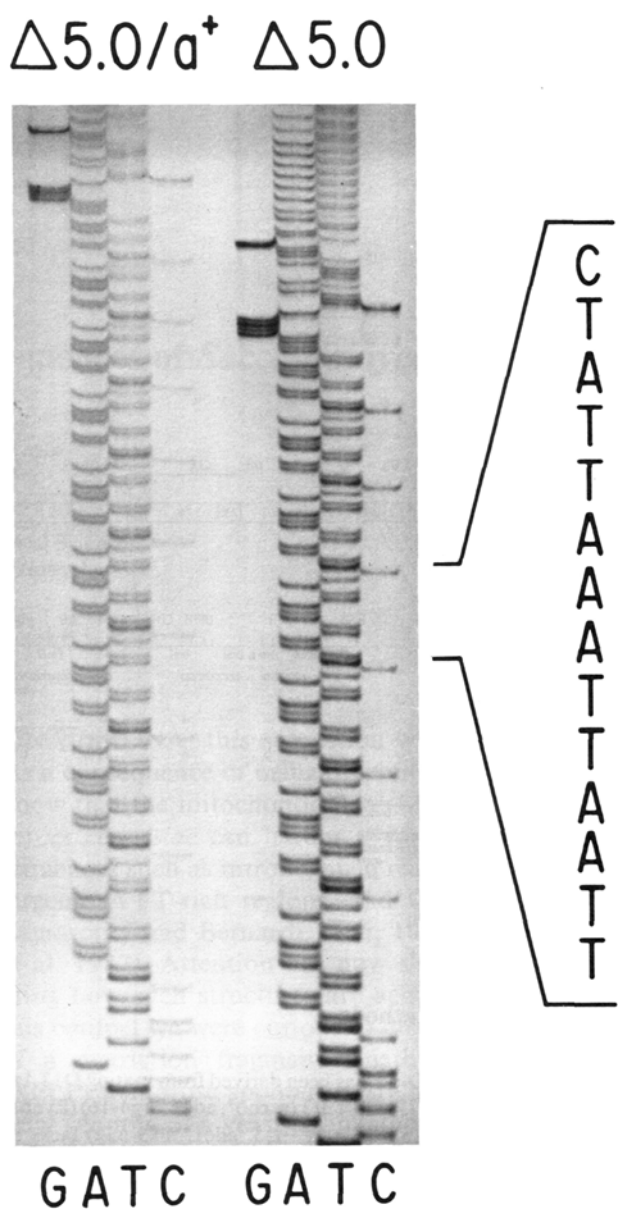

Fig. 3. Part of a sequencing gel showing the increased number of tandem repeats of the 14-bp sequence TTAATTAAATTATC in variant strain Δ5.0/a+ relative to its parent Δ5.0

Results

Characterization of the RFLP in Strain Δ5.0/a+

The CfoI restriction digestion profiles of mtDNA from strains Δ5.0 and Δ5.0/a+ are shown in Fig. 2. Strain Δ5.0/a+, which had been cultured in the laboratory for 2 yr, has gained sequence so that the wild-type 2.15-kb CfoI fragment now has a size of approximately 2.2 kb (Fig. 2). A more precise location of the alteration was obtained by comparing restriction digestion patterns from the cloned 2.2- and 2.15-kb fragments. This revealed that the polymorphism occurred in a terminal 0.3-kb DraI/CfoI fragment (not illustrated). Sequencing of this region from both yeast stocks showed that the sequence TTAATTAAATTATC is present in eight copies in strain Δ5.0/a+ and in six copies in the original strain (Fig. 3). The final nucleotide of the last 14-mer ends 112 bp upstream of the proline tRNA gene. The 14-mer does not exist in the equivalent region of another yeast strain (Miller et al. 1983). No other difference in sequence was found in the cloned CfoI fragments.

Attempts to Induce Further Mitochondrial Genome Polymorphisms

Because some actively transcribing cells exhibit enhanced recombination (Thomas and Rothstein 1989) and because recombination could account for the observed RFLP in the mtDNA of strain Δ5.0/a+, and for other polymorphisms in mtDNA (see later), attempts were made to induce further changes by growing cells continuously on nonfermentable medium where mitochondria are derepressed. In ad-

dition, growth at supraoptimal temperatures may destabilize the mtDNA double helix in A+T-rich regions (increasing the chance of DNA strand slippage) or the DNA polymerase could be more prone to err at the higher temperatures. In GlyYP medium Dip 2 grows at 0.46 generations/h at 30°C whereas at 34°C Dip 2 grows at 0.25 generations/h and at 0.18 generations/h at 37°C. The mtDNA of 44 individual colonies of the sample grown at 34°C and 38 colonies of the sample grown at 37°C were examined after 100 generations at these temperatures. None exhibited an altered mtDNA restriction digestion pattern relative to wild type. This observation carries the qualification that our screening technique would only detect changes of 20–30 bp in the smaller mtDNA fragments (<5 kb) and only larger changes of >50 bp in fragments above 5 kb.

Tandemly Repeated Sequence Polymorphisms in Intergenic Regions of the Mitochondrial Genome of Strain Dip 2

In the course of characterizing novel junctions in deleted and rearranged mitochondrial genomes derived from Dip 2 we have accumulated a large amount of sequence information (Clark-Walker 1989; Skelly and Clark-Walker 1990). This has enabled us to make direct visual comparisons between mtDNA sequences of this strain and those previously reported. As indicated in Figs. 1 and 4, polymorphisms involving tandemly repeated sequences are common. The examined regions are widely dispersed throughout the mitochondrial genome, yet almost all fragments contain at least one sequence polymorphism (Fig. 1). Among sequence polymorphisms 3 bp or larger, strains commonly differ by the number of times a particular tandemly repeated A+T sequence motif is reiterated. For instance the A+T motif ATTTAAT of sequence number 1 (Fig. 4ai) is present in two copies in our strain and in one copy in a second strain. In many cases the number of tandem repeats is increased by one unit when two strains are compared. However increases of 2, 3, 4, and up to 8 tandemly repeated motifs are also evident. In the 5850 bp of intergenic sequence compared, 14 such polymorphisms [comprising the RFLP of $\Delta 5.0/a^+$ (absent from reference strain M, Miller et al. 1983) and sequence numbers 1–13 inclusive (Fig. 4ai)] are observed, giving a frequency of one approximately every 420 bp. This figure represents a comparison between one strain (Dip 2) and a series of other yeast strains. If we compare the A+T-rich polymorphisms in Dip 2 and a single other strain (strain B) we arrive at a somewhat lower estimate of the polymorphic nature of the mitochondrial genome. Of the 1590-bp comparable intergenic sequence, three A+T repeat polymorphisms are detected, representing one approximately every 530 bp. Other polymorphisms 3 bp or larger comprise A+T-rich insertions/deletions of unique sequences (Fig. 4aii). These sequences almost always contain elements shared by neighboring sequences (underlined in Fig. 4aii). The locations of all interstrain differences of 3 bp or greater are illustrated in Fig. 1 by arrows.

The 2-bp differences in sequence (Fig. 4b) are heterogeneous in nature. In only two cases (sequences 26 and 28, Fig. 4b) are tandem repeat polymorphisms seen in this group. At three locations numerous small (1–2-bp) differences are clustered (Fig. 4c). As in the case of the 2-bp differences, it is difficult to assess whether these clustered sequence changes arose through single or multiple events, and the latter have not been included in the calculations of the degree of polymorphism (see below).

Among the 89 1-bp interstrain differences seen in the 5850 bp examined (not illustrated), 38 are base substitutions (42% transitions); insertions/deletions comprise the largest category (51/89), and most of these (46/51) consist solely of A and T bases (Fig. 4d). Additionally, most insertions/deletions (36/51) are of the same form as the larger A+T tandemly repeated sequence polymorphisms listed in Fig. 4a. Although both expansions and contractions of tandem repeats are seen in the mitochondrial genome of strain Dip 2 relative to the other strains examined, in most cases the genome of Dip 2 has fewer repeats than the comparison strain.

Tandem Repeats with Associated Optional GC Clusters

Three A+T tandemly repeated polymorphic sequences lie alongside GC clusters that are themselves optional. These repeated sequences are those that give rise to the RFLP in strain $\Delta 5.0/a^+$ and those numbered 12 and 13 (Fig. 4a). The sequences of the GC clusters together with the A+T tandem repeats are given in Fig. 5a–c.

Polymorphisms Involving Large Open Reading Frames (ORFs)

One or both of ORFs 2 and 4 are absent in some yeast strains (Seraphin et al. 1987). Our inability to detect these ORFs in the regions sequenced suggests that strain Dip 2 lacks both of them. ORF 4 and its flanking sequences, comprising 1557 bp, are replaced in strain Dip 2 by the 8-mer GGCTATAG whereas ORF 2 and its flanking sequences, comprising 1682 bp, are replaced in strain Dip 2 by the 53-bp sequence ATTTACATATGGAGTATTAA ACTATAATAAATACAATATACCCCATCCCCC CC.

(a) Differences of 3bp or longer: i. Tandem repeats (n=13)

No.	Fig.1 ref.[a]	Position[b]	Polymorphism[c]	Strain (ref)[d].
1	1	17040	aattt[ATTTAAT]$_2$attta aattt[ATTTAAT]$_1$attta	D2, (1)
2	1	17071	aatat[TAATA]$_4$atgta aatat[TAATA]$_2$atata	D2, (1)
3	2	18894	tataa[ATA]$_3$aaaaa tataa[ATA]$_2$aaaaa	A, (2)
4	3	21010	attat[ATAAT]$_1$aatta attat[ATAAT]$_2$aatta	M, (3)
5	4	22198	ttatt[TTT]$_1$attat ttatt[TTT]$_2$attat	M, (3)
6	5	24220	tcttt[AAT(T)]$_4$aatat tcttt[AAT(T)]$_7$aatat	A,J, (4,5)
7	5	24251	taata[TTTTA]$_1$taaaa taata[TTTTA]$_2$taaaa	A,J, (4,5)
8	7	44973	tattt[AAT]$_3$aaaat tattt[AAT]$_4$aaaat	J, (6)
9	8	48859	ttttt[ATT]$_4$aattt ttttt[ATT]$_5$aattt	D2, (7)
10	8	50048	cgtaa[GTATATA]$_2$gtata cgtaa[GTATATA]$_4$gtata	B, (8)
11	8	50412	gtttt[AAAAAT]$_9$aaaaa gtttt[AAAAAT]$_{12}$aaaaa	B, (8)
12	10	51947	gacta[TT(T)TTA]$_5$t-aaa gaaga[TT(T)TTA]$_9$ttaaa	B, (2)
13	12	53701	agaa-[TTTTA]$_2$attta agaag[TT(T)TTA]$_4$attta agaag[TT(T)TTA]$_{12}$attta	D1, (9) Cons, (10)

ii. Unique sequence differences (n=5)

14	2	18877	taata[A(ATA)$_3$]tttat taata[-------]tttat	A, (2)
15	4	22127	taaat[----]taaaa taaat[AAAA]taaaa	M, (3)
16	8	48138	tttat[TAAT(ATT)$_3$]atctt tttat[----(ATT)$_1$]atcat	D2, (7)
17	8	50146	aatat[GGG]gagct aatat[AAA]gagct	B, (8)
18	11	53301	aatat[---]aaata aatat[ATT]aaatt	B, (2)

(b) 2bp differences (n=10)

19	1	17268	gaaac[TT]ataaa gaaac[--]ataaa	D2, (1)
20	5	24140	aaata[AA]aataa aaata[GG]aataa	A, (4)
21	8	48179	aataa[C-]ttatt aataa[TT]ttatt	D2, (7)
22	8	48316	ttata[C-]ttaat ttata[TT]ttaat	D2, (7)
23	8	50333	ataaa[AA]aaaag ataaa[T-]aaaag	B, (8)
24	10	51927	taata[AC]tcctt taata[--]tcctt	Cons, (10)
25	10	51982	tatta[T-]cttct ta-ta[AA]cttct	B, (2)

Fig. 4. Caption continued on next page.

26	10	52085	tcaac[AT]$_1$tacce tcaa-[AT]$_2$tacce	B, (2)
27	11	53277	atatt[TG]ccttc atatt[AA]ccttc	B, (2)
28	11	53361	tattt[AT]$_4$aatcc tattt[AT]$_5$aatcc	B, (2)

(c) Clustered 1-2bp differences (n=3)

29	8	48115	tcttt[TATCTAATAAAA-]tattt tcttt[CATA-AATATATC]tattt	D2, (7)
30	9	51547	aactc[CGAAAGAGTAAGGAGATA]ttaat aactc[-GA--GAGTT--GAGA--]ttaat	Cons, (10)
31	11	53322	ttaat[AAGTA-]ttata ttaat[T-G-AG]ttata	B, (2)

(d) Single base differences (n=89)

Insertions/deletions: 51 (comprising 46 A-T; 5 G-C)
Base substitutions: 38
Transitions: 16
Transversions: 22 (comprising 19 A-T; 3 A-C/T-G; 0 G-C)

Fig. 4. Mitochondrial DNA sequence differences between yeast strain Dip 2 and other *S. cerevisiae* strains. (**a**) Differences of 3 bp or longer. (i) Tandem repeats ($n = 13$). (ii) Unique sequence differences ($n = 5$). (**b**) Two-base pair differences ($n = 10$). (**c**) Clustered 1–2-bp differences ($n = 3$). (**d**) Single base differences ($n = 89$). [a] Refers to the position of the polymorphism relative to the sequenced regions (1–13) of Fig. 1. [b] Indicates the precise position of the polymorphism on the consensus mitochondrial genome sequence of deZamaroczy and Bernardi (1986b). The number refers to the base, common to both strains, juxtaposing the polymorphism. [c] The polymorphic sequence is in uppercase letters within square brackets. The number of times it is repeated is indicated by the subscript. Single bases contained within curved brackets indicate that the base is not present in all repeats. Sequences flanking the polymorphism are in lowercase letters. Gaps, –, have been introduced in some sequences to facilitate alignment. All sequences are of the nontranscribed strand. [d] The Dip 2 sequences is uppermost in all cases. Reference strains are denoted with the single letter abbreviation system of deZamaroczy and Bernardi (1986b). Cons indicates sequence derived from the consensus mitochondrial genome sequence of deZamaroczy and Bernardi (1986b). References: (1) Thalenfeld and Tzagoloff (1980), (2) Goursot et al. (1982), (3) Miller et al. (1983), (4) Gaillard et al. (1980), (5) Maxwell et al. (1986), (6) Novitski et al. (1984), (7) Macino and Tzagoloff (1980), (8) Colin et al. (1985), (9) Bonitz et al. (1982), (10) deZamaroczy and Bernardi (1986b).

	Position[b]	Polymorphism[c]	Ref.[d]
a)	22116	nnnnn[AAAAATAAAAAGGGGTTCGGTCCCCCCCCTTCCGTATACTTACGGGAGGGGGGTCCCTCACTCCTTC(TTAATTAAATTATC)$_6$]-ttaa	
		tacce[----------]cttaa	M, (3)
b)	51945	gacta[----------(TT(T)TTA)$_5$]t-aaa	
		nnnnn[TATTAGTAGGAGTGCGGGACCCCGTGGGAACCGAACCCCTTTTTTATTTTAAAGAAGAAGA(TT(T)TTA)$_9$]ttaaa	B, (2)
c)	53642	accce[TATTTAAGAAGGAGTGCGAAACCCCGTGGGACCGAACCCCTTTTTTATTTGAAAAGAA(TTTTA)$_2$]tttaa	
		accce[----------GAAAAGAA(TT(T)TTA)$_4$]tttaa	D1, (9)
		accce[TATTTAAGAAGGAGTGCGAAACCCCGTGGGACCGAACCCCTTTTTTATTTAAAGAAGAAG(TT(T)TTA)$_{12}$]tttaa	Cons, (10)

Fig. 5. A+T tandemly repeated polymorphisms with associated optional GC clusters. [b,c,d] As Fig. 1; n indicates undetermined sequence.

Discussion

Polymorphisms Involving Tandem Repeats

Polymorphisms involving tandemly repeated A+T-rich sequences in the mitochondrial genome of *S. cerevisiae* are common and must arise by DNA replication errors or intramolecular/intermolecular recombination. One important feature of these inter-strain differences in tandem repeats is that there appears to be no breakage within any of the motifs. The misalignment of one repeated motif either with its complementary motif on the opposite strand (slipped mispairing, Fig. 6a) or else with a companion motif on a second strand (brought about by an invasive primary strand, Fig. 6b) could generate the differences seen at the "a" locus of strain Dip 2. It is noteworthy that two copies of the 14-mer, which comprises the "a" sequence, can associate to form a hairpin structure (Fig. 6aii). The significance of

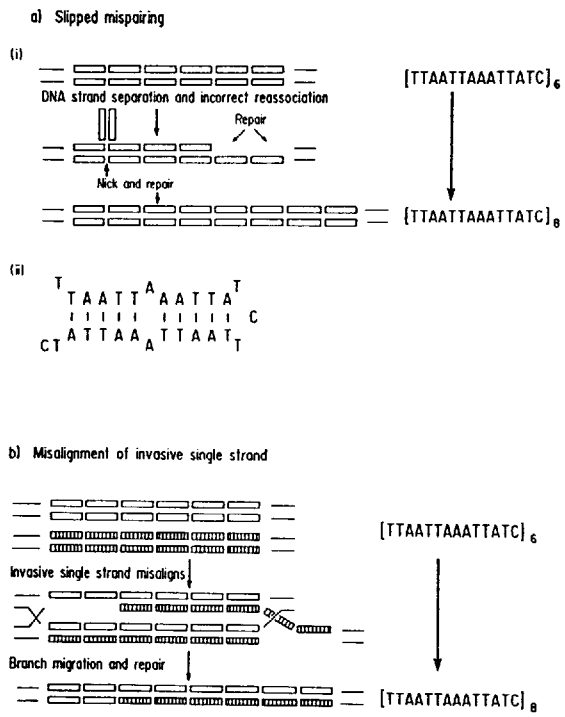

Fig. 6. Mechanisms of A+T tandemly repeated sequence expansion. Rectangles represent repeating sequence motifs. **a**(i) Slipped mispairing. (ii) Hairpin formed by two copies of the 14-mer at the "a" locus of strain Dip 2. **b** Misalignment of an invasive single strand.

this to the formation of polymorphisms in other tandem repeats is unclear as the ability to form hairpins is not present universally.

Because the unique A+T polymorphisms documented (Fig. 4aii) usually share sequence elements with flanking regions, processes similar to those illustrated in Fig. 6 could also account for many of these differences. For instance some imprecision in the alignment of repeats, resulting in what has been termed dislocation mutagenesis (Kunkel and Soni 1988), could generate optional sequences with a nonregular composition. Thus sequence 14, which has an additional A (bold type in Fig. 4b) that disrupts its ATA repeat motif, and sequence 18, which has an additional T (bold type, Fig. 4b) that disrupts its AT repeat, may have arisen by this means.

Although a proportion of the small (1–2-bp) sequence differences between strains may reflect sequencing errors, many small changes undoubtedly arise through the sorts of slipped mispairing interactions and recombinations listed above for the larger A+T tandem repeat polymorphisms. Such simple single base substitutions/additions/deletions may form larger repeating motifs by chance. If the repeats are not separated by too great a distance they too can then undergo slippage and recombinational events as outlined to generate larger sequence inserts. Such processes represent potential mechanisms of formation of the large, A+T-rich intergenic mtDNA regions of some yeasts (deZamaroczy and Bernardi 1987). Indeed these processes have been proposed as major mechanisms of DNA sequence evolution (Levinson and Gutman 1987).

Other small (1–2-bp) differences may reflect the absence of certain error correcting abilities by the mtDNA polymerase. The clustering of such differences in a number of locations suggests that regional variation in sequence composition, and therefore secondary structure, makes the polymerase more prone to error in certain parts of the genome.

The variant 1 gene displays a number of tandemly repeated A+T sequence polymorphisms of the sort listed in Fig. 4 involving the triplet sequence AAT. At one location the triplet has been found in either 5 or 8 copies and in a second location in 8, 10, or 12 copies (Butow et al. 1985). In the small rRNA gene of some yeast strains a 16-bp optional A+T insert has been reported (Huttenhofer et al. 1988). Other genes do not contain such polymorphisms. However as illustrated in Fig. 4 the tandemly repeated A+T polymorphism of sequence 5 is contained within the region defined as a tRNA synthesis locus (TSL). Previously a GC cluster polymorphism extending from position 21103 to 21151 (see deZamaroczy and Bernardi 1986b) has been identified at this locus (Miller et al. 1983). In addition the Dip 2 TSL is polymorphic for a second, 46-bp GC cluster (from position 20921 to 20967, deZamaroczy and Bernardi 1986b). The finding of these three polymorphisms (two GC clusters and an A+T tandem repeat) suggests that these regions are not essential in the operation of the locus.

Some size variation in animal mtDNAs has likewise been shown to be due to a varying copy number of a repeated element (La Roche et al. 1990 and references therein). Even in the few noncoding regions that exist in animal mtDNA, length variation is common (Hauswirth et al. 1984). However animal mtDNA repeated sequence units, in contrast to those of *S. cerevisiae* documented here, are substantially larger and range from 64 bp to 10.4 kb in lizards (Densmore et al. 1985; Moritz and Brown 1987; C. Moritz, personal communication).

Tandem Repeats with Associated Optional GC Clusters

Three of the tandemly repeated A+T polymorphic sequences identified in this study lie alongside optional GC clusters. Such optional GC clusters have been suspected to be mobile elements, although only one (the 46-bp GC cluster located in the coding region of the *var1* gene) has been shown to be directly involved in transfer (Strausberg and Butow 1981; Hudspeth et al. 1984). As many as 200 GC clusters exist in the mitochondrial genome of *S. cerevisiae* and these have been categorized according

to their primary and secondary structure (de-Zamaroczy and Bernardi 1986a; Weiller et al. 1989). Based on sequence characteristics, each of the optional GC clusters identified in this study belongs to the M3 category (Weiller et al. 1989). The occurrence of some M3 GC clusters in tandem arrays has led to the suggestion that they insert 3′ to the terminal AGGAG sequence of a preexisting cluster (Weiller et al. 1989). Although it seems clear that the mechanism of movement of M3 GC clusters differs from that of other GC clusters, the possibility should not be overlooked that, rather than transposing, tandemly arranged GC clusters can expand and contract in the same manner as tandem AT-repeated sequences as discussed above.

The tandem repeat units of sequence 12 and 13 are identical. In addition the GC clusters associated with these repeats are very similar in sequence (Fig. 5). This finding of similar optional GC clusters with associated A+T repeats at a number of locations in different yeast strains suggests that these sequences may have recombined/transposed as a unit.

Base Substitutions

In contrast to the pattern of base substitutions in animal mitochondrial genes, which is greatly biased towards transitions (DeSalle et al. 1987; Thomas and Beckenbach 1989), the pattern observed here for yeast intergenic sequences shows some bias toward transversions (58%, 22/38). In other studies involving comparison of mitochondrial protein-coding genes from related yeast species transversions exceed transitions by 2:1 (Clark-Walker, unpublished). A comparison of the mitochondrial-encoded large rRNA gene sequence of other fungi suggests about equal numbers of transitions and transversions (Bruns et al. 1989). Hence fungal mitochondrial genomes appear to differ from animal mtDNAs by lacking the extreme bias in favor of transitions. Whether this variation is due to a preference for A and T by the mtDNA polymerase or reflects differences in the mismatch repair mechanisms has yet to be decided.

Despite these comments, animal and fungal mtDNAs are similar in that the majority of changes in noncoding regions are due to small length changes rather than base substitutions (Moritz and Brown 1987; Buroker et al. 1990).

Polymorphism Involving the ORFs

The large ORFs 2 and 4 are replaced in strain Dip 2 by short sequences; ORF 4 by an 8-bp and ORF 2 by a 53-bp A+T-rich sequence. In the latter case, as there is little evidence of internal repeating motifs within the sequence it is unlikely that it arose by the types of expansion mechanisms considered for other A+T-rich optional sequences. Because ORFs have some characteristics of group 1 introns, it has been suggested that these elements may have resulted from transposition (Seraphin et al. 1987). It is possible that the insertion of the ORFs may have eliminated some sequence at their sites of entry. These sequences could still remain in Dip 2. Alternatively, an ancestral strain may have lacked both the ORFs and also the sequences now replacing them in Dip 2. Such a strain may, in one instance, gain an ORF by transposon insertion and, in another instance, gain a replacement sequence through recombination or possibly through an insertion process.

Conclusion

Interstrain sequence comparisons highlight the extreme polymorphism of mtDNA intergenic regions particularly with regard to A+T-rich tandem repeats. It is noteworthy that the observed variation at the "a" locus of strain Dip 2 is due to a gain of 28 bp. Although we have referred to other polymorphisms in tandem repeats between Dip 2 and other strains as a gain or loss it may well be that such variations are due to a gain in sequence only. From an evolutionary viewpoint it would be extremely interesting to learn whether tandem repeated structures can lose units. If a gain of tandem units predominates over loss then this may help to account for the presence of the large tracts of A+T-rich intergenic sequences in baker's yeast mtDNA. It can be imagined that base substitutions together with additions and deletions of single nucleotides will eventually destroy the repeated structure and hence mask the origin of much intergenic sequence. Furthermore we believe that the mitochondrial genome of *S. cerevisiae* can tolerate the high degree of intergenic polymorphism identified in this study because much of this sequence has no biological function. This view is supported by the observation that an *S. cerevisiae* mutant strain has been produced that lacks a total of 8.7 kb of intergenic mtDNA sequence from two regions but has growth characteristics indistinguishable from wild type (Skelly and Clark-Walker 1990).

Acknowledgments. We thank Erika Wimmer for skilled technical assistance and Dr. Craig Moritz for sharing unpublished data.

References

Bonitz SG, Homison G, Thalenfeld BE, Tzagoloff A, Nobrega FG (1982) Assembly of the mitochondrial membrane system. Processing of the apocytochrome b precursor RNAs in *Saccharomyces cerevisiae D273-10B*. J Biol Chem 257:6268–6274

Bruns TD, Fogel R, White TJ, Palmer JD (1989) Accelerated evolution of a false-truffle from a mushroom ancestor. Nature 339:140–142

Buroker NE, Brown JR, Gilbert TA, O'Hara PJ, Beckenbach AT, Thomas WK, Smith MJ (1990) Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics 124:157–163

Butow RA, Perlman PS, Grossman LI (1985) The unusual *var1* gene of yeast mitochondrial DNA. Science 228:1496–1501

Clark-Walker GD (1989) In vivo rearrangement of mitochondrial DNA in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 86:8847–8851

Colin Y, Baldacci G, Bernardi G (1985) A new putative gene in the mitochondrial genome of *Saccharomyces cerevisiae*. Gene 36:1–13

Densmore LD, Wright JW, Brown WM (1985) Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*). Genetics 110:689–707

DeSalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J Mol Evol 26:157–164

deZamaroczy M, Bernardi G (1985) Sequence organization of the mitochondrial genome of yeast—a review. Gene 37:1–17

deZamaroczy M, Bernardi G (1986a) The GC clusters of the mitochondrial genome of yeast and their evolutionary origin. Gene 41:1–22

deZamaroczy M, Bernardi G (1986b) The primary structure of the mitochondrial genome of *Saccharomyces cerevisiae*—a review. Gene 47:155–177

deZamaroczy M, Bernardi G (1987) The AT spacers and the *var1* genes from the mitochondrial genomes of *Saccharomyces cerevisiae* and *Torulopsis glabrata*: evolutionary origin and mechanism of formation. Gene 54:1–22

Dujon B (1980) Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the ω and *rib*-1 loci. Cell 20:185–197

Evans RJ, Clark-Walker GD (1985) Elevated levels of petite formation in strains of *Saccharomyces cerevisiae* restored to respiratory competence. II. Organization of mitochondrial genomes in strains having high and moderate frequencies of petite mutant formation. Genetics 111:403–432

Evans RJ, Oakley KM, Clark-Walker GD (1985) Elevated levels of petite formation in strains of *Saccharomyces cerevisiae* restored to respiratory competence. I. Association of both high and moderate frequencies of petite formation with the presence of aberrant mitochondrial DNA. Genetics 111:389–412

Gaillard C, Strauss F, Bernardi G (1980) Excision sequences in the mitochondrial genome of yeast. Nature 283:218–220

Goursot R, Mangin M, Bernardi G (1982) Surrogate origins of replication in the mitochondrial genomes of *ori*^{0} petite mutants of yeast. EMBO J 1:705–711

Hauswirth WW, Van de Walle MJ, Laipis PJ, Olivo PD (1984) Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. Cell 37:1001–1007

Hudspeth MES, Vincent RD, Perlman PS, Shumard DS, Treisman LO, Grossman LI (1984) Expandable *var1* gene of yeast mitochondrial DNA: in frame insertions can explain the strain-specific protein size polymorphisms. Proc Natl Acad Sci USA 81:3148–3152

Huttenhofer A, Sakai H, Weiss-Brummer B (1988) Site-specific AT-cluster insertions in the mitochondrial 15S rRNA gene of the yeast *S. cerevisiae*. Nucleic Acids Res 16:8665–8674

Kunkel TA, Soni A (1988) Mutagenesis by transient misalignment. J Biol Chem 29:14784–14789

La Roche J, Snyder M, Cook DI, Fuller K, Zouros E (1990) Molecular characterization of a repeat element causing large-scale variation in the mitochondrial DNA of the sea scallop *Placopecten megallanicus*. Mol Biol Evol 7(1):45–64

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4(3):203–221

Macino G, Tzagoloff A (1980) Assembly of the mitochondrial membrane system: sequence analysis of a yeast mitochondrial ATPase gene containing *oli-2* and *oli-4* loci. Cell 20:507–517

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning—a laboratory manual. Cold Spring Harbor Laboratories, Cold Spring Harbor, NY

Maxwell RJ, Devenish RJ, Nagley P (1986) The nucleotide sequence of the mitochondrial genome of an abundant petite mutant of *Saccharomyces cerevisiae* carrying the *ori1* replication origin. Biochem Int 13:101–107

McClelland M, Hanish J, Nelson M, Patel Y (1988) KGB: a single buffer for all restriction endonucleases. Nucleic Acids Res 16:364

Miller DL, Underbrink-Lyon K, Najarian DR, Krupp J, Martin NC (1983) Transcription of yeast mitochondrial tRNA genes and processing of tRNA gene transcripts. In: Schweyen RJ, Wolf K, Kaudewitz F (eds) Mitochondria 1983. Nucleo-mitochondrial interactions. De Gruyter, Berlin, pp 151–164

Moritz C, Brown WM (1987) Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc Natl Acad Sci USA 84:7183–7187

Novitski CE, Macreadie IG, Maxwell RJ, Lukins HB, Linnane AW, Nagley P (1984) Biogenesis of mitochondria: genetic and molecular analysis of the *oli2* region of mitochondrial DNA in *Saccharomyces cerevisiae*. Curr Genet 8:135–146

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-termination inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Seraphin B, Simon M, Faye G (1987) The mitochondrial reading frame RF3 is a functional gene in *Saccharomyces uvarum*. J Biol Chem 262:10146–10153

Skelly PJ, Clark-Walker GD (1990) Conversion at large intergenic regions of mitochondrial DNA in *Saccharomyces cerevisiae*. Mol Cell Biol 10:1530–1537

Sor F, Fukuhara H (1982) Nature of an inserted sequence in the mitochondrial gene coding for the 15S ribosomal RNA of yeast. Nucleic Acids Res 10:1625–1633

Strausberg RL, Butow RA (1981) Gene conversion at the *var1* locus of yeast mitochondrial DNA. Proc Natl Acad Sci USA 78:494–498

Thalenfeld BE, Tzagoloff A (1980) Assembly of the mitochondrial membrane system. Sequence of the *Oxi 2* gene of yeast mitochondrial DNA. J Biol Chem 255:6173–6180

Thomas BJ, Rothstein R (1989) Elevated recombination rates in transcriptionally active DNA. Cell 56:619–630

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

Weiller G, Schueller CME, Schweyen RJ (1989) Putative target sites for mobile G+C rich clusters in yeast mitochondrial DNA: single elements and tandem arrays. Mol Gen Genet 218:272–283

Received May 9, 1990/Revised and accepted November 10, 1990