Gene, 37 (1985) 1–17
Elsevier

GENE 1354

Sequence organization of the mitochondrial genome of yeast – a review

(*Saccharomyces cerevisiae*; genome unit size; base composition; coding and noncoding sequences)

Miklos de Zamaroczy and Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2, Place Jussieu, 75005 Paris (France) Tel. (1) 329-58-24; (1) 336-25-25, ext. 41.01

(Received February 25th, 1985)
(Accepted April 17th, 1985)

SUMMARY

We have compiled the available primary structural data for the mitochondrial genome of *Saccharomyces cerevisiae* and have estimated the size of the remaining gaps, which represent 12–13% of the genome. The lengths of sequenced regions and of gaps lead to a new assessment of genome sizes; these range (in round figures) from 85 000 bp for the long genomes, to 78 000 bp for the short genomes, to 74 000 bp for the supershort genome of *Saccharomyces carlsbergensis*. These values are 8–11% higher than those previously estimated from restriction fragments.

Interstrain differences concern not only facultative intervening sequences (introns) and mini-inserts, but also insertions/deletions in intergenic sequences. The primary structure appears to be extremely conserved in genes and *ori* sequences, and highly conserved in intergenic sequences. Since coding sequences represent at most 33–35% of the genome, at least two thirds of the genome are formed by noncoding and yet highly conserved sequences.

The G + C level of genes or exons is 25%, and that of intronic open reading frames (ORFs) 22%; increasingly lower values are shown by intronic closed reading frames (CRFs), 20%, *ori* sequences, 19%, intergenic ORFs, 17.5% and intergenic sequences, 15%.

INTRODUCTION

Data on the primary structure of the mitochondrial genome units of *S. cerevisiae* have been accumulating since the full sequence of a petite genome was first published six years ago (Gaillard and Bernardi, 1979; some very short nucleotide stretches had been sequenced previously, Van Kreijl and Bos, 1977; Cosson and Tzagoloff, 1979), but the goal of establishing the complete primary structure has not yet been reached. This may seem surprising if one considers that investigations on animal mitochondrial genomes, which were started at about the same time, have already produced a number of full sequences (human, Anderson et al., 1981; murine, Bibb et al., 1981; bovine, Anderson et al., 1982; *Xenopus laevis*, Wong et al., 1983; *Drosophila*

Abbreviations: bp, base pair(s); CRF, closed reading frame; nt, nucleotide(s); ORF, open reading frame; *ori*, origin of DNA replication.

yakuba, Clary and Wolstenholme, 1984). There are, however, several reasons for this delay: (i) the mitochondrial genome units of yeast are larger, by a factor of five, than those of animals; (ii) most investigations have been focused on genes, introns, and *ori* sequences, and not on the intergenic sequences which form the majority of the genome; (iii) cloning large restriction fragments, which contain long intergenic sequences, has not been achieved so far; (iv) petite genomes have been used in most primary structure studies, with the consequence of leaving unsequenced gaps on the wild-type genome map; (v) restriction sites are widely spaced in intergenic sequences. As a result, 12–13% of the genome, mostly formed by intergenic sequences, is still undetermined. Since, however, the vast majority of the genome is now known in its primary structure, we thought that a critical analysis of the published data might be useful at this point. This is not an easy task mainly because the data, which are widely scattered in the literature, concern different strains. This compilation provides a much needed overall view of the mitochondrial genome of yeast essentially based on primary structural data. Assessing the total lengths of sequenced regions and of the remaining gaps has led us to re-estimate the size of long, short and supershort genome units. Finally, this work has provided estimates of the relative amounts of coding and noncoding sequences in the mitochondrial genome of yeast, of the G + C contents of different genome regions, and of the copy number of genome units.

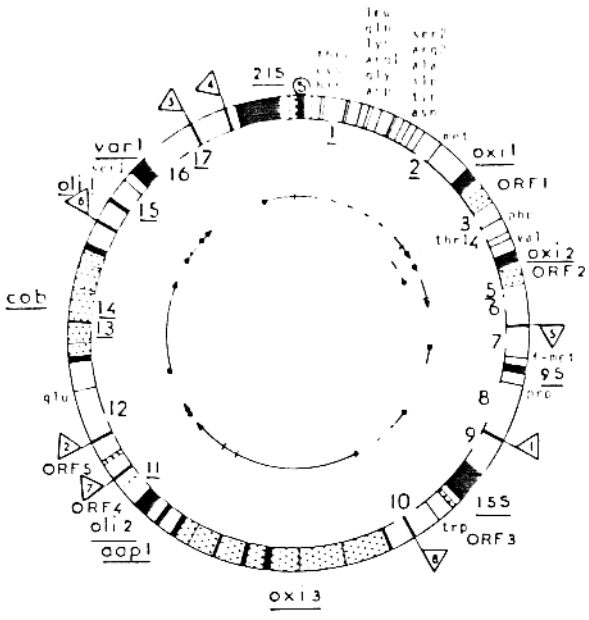

Fig. 1. Physical map of a long mitochondrial genome unit of wild-type *S. cerevisiae*, showing the remaining sequence gaps; these are numbered 1 through 17; underlined figures refer to intra-regional gaps, other figures to inter-regional gaps. In the sequenced regions, black areas correspond to mitochondrial genes or their exons (the last exon block of *oxi3* comprises exons A 6-8 and introns aI 7, 8); dotted areas to intervening sequences and intergenic open reading frames (ORF1-ORF5; Colin et al., 1985); radial lines indicate tRNA genes (the *thr1* gene is the only one to have an anticlockwise orientation). Among mitochondrial genes, *oxi1*, *2* and *3* encode sub-units II, III and I, respectively, of cytochrome *c* oxidase; *cob*, cytochrome *b*; *aap1*, *oli2* and *oli1*, sub-units 8, 6 and 9 of ATPase; *var1*, a protein associated with the small mitochondrial ribosome sub-unit; *9S* corresponds to the central part of tRNA synthesis locus; *15S* and *21S* are the genes for the small and large ribosomal RNAs, respectively. Triangles indicate the location of *ori* sequences 1-8; they point in the direction cluster C to cluster A. Mitochondrial transcripts are represented on an inner circle. They have a clockwise orientation except for two short transcripts in the ORF1-*thr1* region. Dots correspond to initiation sites, arrows to termination sites, cross-lines to processing sites. For some transcripts, initiation and/or termination sites are not known so far. Ⓢ corresponds to the *Sal*I site used as the reference point (origin) of the map.

RESULTS

(a) The sequenced regions

Our starting point was a critical compilation of the mitochondrial DNA regions that have been sequenced so far in a number of *S. cerevisiae* strains. These data are presented in Fig. 1 and in Table I. The rationale for putting together data for different strains will be provided in DISCUSSION. Basically, the data of Table I show that sequenced regions comprise 73 830 bp in long genomes, which contain a full complement of intervening sequences (introns), and 69 120 bp in short genomes. Only some short regions of the genome have been sequenced in more than one strain (see Table V).

(b) The sequence gaps

As shown in Fig. 1 and Table I, a number of gaps exist in the available sequences. These belong to two classes. Intra-regional gaps (underlined figures of Tables I and II and Fig. 1) are gaps within sequenced

TABLE I

Sequenced regions

Gap[a]	Region	Size[b]	Strain[c]	Ref.[d]
	thr2-cys-his	1380	D1	(1)
1	*leu-gln*	1320	D1	(2)
	lys-arg1-gly-asp-ser2-arg2*	3080	D1	(3)
		(630)	A	(4)
	ala-ile-tyr*	1840	S	(5)
		(360)	D1	(3)
2	*asn-met-oxi1*-ORF1	4440	D1	(6,7)
3	*phe*	360	D1	(8)
4	*thr1*	310	D1	(9)
5	*val-oxi2*-ORF2*	3990	D2	(10)
		(1790)	D2	(11)
6	*ori5*	1750	A	(12-14)
		(960)	K	(15)
7	*fmet-9S-pro*	2610	M	(16)
8	*ori1*	880	A	(17)
		(900)	J	(18)
		(1360)	B	(19)
9	*15S*	480	B	(13)
		2130	M	(20)
		(460)	A	(4,21)
	ORF3-*trp*	2200	A	(4,21)
10	*oxi3*	800	K	(22)
		10170	D2	(23)
		2900[e]	K	(24)
	aap1	1750	J	(25)
		(450)	D2	(23,26)
11	*oli2**-ORF4	2390	D2	(26)
		(950)	J	(25)
	*ori7**-ORF5-*ori2*	340	B	(13)
		170	D2	(26)
		3000	B	(27,28,14)
		(300)	D2	(26)
12	*glu*	1650	B	(13)
		1450	D1	(29)
		(1720)	A	(4)
13,14	*cob-ori6**	6130	D1	(30)
		1810[e]	F	(31)
		(250)	B	(27)
15	*oli1*-ser1-var1**	4230	D1	(32)
		(400)	K	(33)
		(1190)	E	(34)
		(500)	A	(14)
16,17	*ori3-ori4*	4580	A	(14,27)
		(710)	K	(35)
	21S	5690	I	(36)
		(1550)	I	(37)
	Total	69120 bp (short genomes)		
		73830 bp (long genomes)		

[a] Intra-regional (underlined) and inter-regional gaps are numbered as in Fig. 1. The former are gaps within published sequences; the latter separate independently sequenced regions. For the definition of gaps and more details, see section b of RESULTS and Table II. Inter-regional gaps also comprise (i) the mini-gaps before and after the *21S* gene (as detected when comparing strains A and I for the 5′ ends, or strains I and D1 for the 3′ ends); and (ii) the mini-gaps between the *tyr* and *asn* genes (as seen by comparing strains S and D1). These mini-gaps might also correspond to local sequence rearrangements.

[b] Sizes are given in bp; they comprise intra-regional mini-gaps, namely very short stretches mostly located between identical restriction sites present in GC clusters, which were neglected or lost in sequencing work. Numbers in parentheses concern regions, or subregions (indicated by asterisks) sequenced in other strains or by other authors (this information is not provided for tRNA sequences); these figures were not included in the totals.

[c] Strains are: A, D243-2B-R1; B, C982-19d; C, NCYC-74S; D1, D273-10B/A21; D2, D273-10B/A48; E, A10; F, 777-3A; I, IL8-8C; J, J69-1B; K, KL14-4A; M, MH41-7B; S, SM202. Note (i) that in the single capital letter abbreviation system used in our laboratory for wild-type strains, D indicates strain DM (Bernardi et al., 1968), which is unrelated to D1 and D2; and (ii) that strains A and J have a common origin (Faugeron-Fonty et al., 1984).

[d] References are: (1) Berlani et al. (1980a); (2) Berlani et al. (1980b); (3) Bonitz and Tzagoloff (1980); (4) Bordonné (1982); (5) Martin, R.P. (pers. comm.); (6) Coruzzi and Tzagoloff (1979); (7) Coruzzi et al. (1981); (8) Miller et al. (1979); (9) Li and Tzagoloff (1979); (10) Thalenfeld and Tzagoloff (1980); (11) Michel (1984); (12) de Zamaroczy et al. (1981); (13) Goursot et al. (1982); (14) de Zamaroczy and Bernardi (unpublished data); (15) Blanc (1984); (16) Miller et al. (1983); (17) Gaillard et al. (1980); (18) Nagley, P. (pers. comm.); (19) Faugeron-Fonty et al. (1984); (20) Sor and Fukuhara (1980); (21) Martin et al. (1983); (22) Osinga et al. (1984a); (23) Bonitz et al. (1980); (24) Hensgens et al. (1983); (25) Novitski et al. (1984); (26) Macino and Tzagoloff (1980); (27) de Zamaroczy et al. (1984); (28) Colin et al. (1985); (29) Bonitz et al. (1982); (30) Nobrega and Tzagoloff (1980); (31) Lazowska et al. (1980); (32) Tzagoloff et al. (1980); (33) Grivell et al. (1979); (34) Hudspeth et al. (1982); (35) Blanc and Dujon (1980); (36) Sor and Fukuhara (1983); (37) Dujon (1980); (38) de Zamaroczy (1984); (39) Jacq et al. (1982).

[e] These sequences account for the difference between long and short genomes (see Table VI).

regions. They (i) are short (< 500 bp), except for the gap corresponding to intron bI3 (which is only present in the *cob* gene of long genomes); (ii) are estimated in size on the basis of the corresponding restriction maps; and (iii) amount to a total of 3440 bp for long genomes and of 1490 bp for short genomes; since they are based on restriction maps, intra-regional gaps may be underestimated by up to

TABLE II

Gaps in the primary structure

Gap[a]	Size[b]	Strains[c]	Location[d]	References[e]
1	190	D1	530 *leu*	(2)
2	480	D1	210 *met*	(7)
3	420*	D1	130 *phe*	(7,8)
4	500*	D1	70 *thr1*	(9)
5	150	D2	*oxi2* 1990	(10)
6	450*	D2,A	1050 *ori5*	(10,27)
7	550*	A,M	1000 *fmet*	(16,27)
8	2650*	M/A,B,C,K	360 *ori1*	(4,38)
9	1000*	A/M	>410 *15S*	(4,38)
	(600)*	B,C,K/M	>410 *15S*	(13,38)
10	1500*	A/D2,K	*trp* 810	(4,21,22)
11	420	D2	in ORF4	(26)
12	400*	B	*ori2* 550	(13,14)
13	250	F	in bI2[f]	(31)
14	1700	F	bI3[f]	(39)
15	130	D1	*oli1* 150	(32)
16	400*	D1/A	2150 *ori3*	(14,32)
17	120*	A	260 *ori3*	(14)

1 490 bp intra-regional gaps (short genomes)
3 440 bp intra-regional gaps (long genomes)
7 870 bp inter-regional gaps (short or long genomes)
9 360 bp total gaps (short genomes)
11 310 bp total gaps (long genomes)

[a] Intra-regional (underlined) and inter-regional gaps are numbered as in Fig. 1 and Table I. The size in parentheses was not included in the totals. For more details, see section **b** of RESULTS and Table I.

[b] Gap sizes in bp are quoted values or our estimates (asterisks) from the references. All intra-regional gaps, and inter-regional gaps 6, 7 and 12 were estimated from restriction maps; the other inter-regional gaps from the mapped positions of flanking landmarks.

[c] See Table I for strain nomenclature. Only the 5′ or the 3′ ends of the inter-regional gaps were localized in some strains; those containing the 5′ ends are shown on the left, those containing the 3′ ends on the right of the slash.

[d] Numbers before or after genes (or *ori* sequences) indicate the distance (in bp) between their beginning or end, respectively, and the proximal end of a gap. For this purpose, *ori* sequences are considered to extend from cluster A to cluster C only, not including sequences *r* (de Zamaroczy et al., 1984).

[e] References as in Table I.

[f] These gaps only exist in long genomes and account for the difference in intra-regional gaps of short and long genomes.

12% (see section **c**). Inter-regional gaps are those which exist between independently sequenced regions. Their size is generally larger than that of intra-regional gaps, and is not as precisely estimated as the latter, because in most cases they bridge regions which have been sequenced in different strains; as a consequence, estimates are often based on the positions of relatively distant restriction sites (details are given in the footnotes to Table I). Inter-regional gaps amount to 7870 bp in both long and short genomes; inter-regional gaps may be either over- or under-estimated by as much as 20%; because of possible compensation this error is likely to affect to a smaller extent the overall estimates of inter-regional gaps; in any case, errors on both intra- and inter-regional gaps may lead to a maximum uncertainty on genome size of $\pm 2\%$. Table II presents the size and the precise location of both intra- and inter-regional gaps which represent 12% and 13% of short and long genomes, respectively. A third class of gaps is formed by mini-gaps; these are either intra-regional or inter-regional (see footnotes a and b of Table I). The latter are negligible in size; the former amount to less than 500 bp and have been

included in the corresponding sequenced regions (Table I).

(c) The unit sizes of short and long genomes

The sum of sequenced regions and gaps leads to an estimate of 85 140 bp for the size of long genomes (Table III). This value is 7% higher than our original estimate obtained by adding up the restriction fragments from the mitochondrial genome of strain A (Bernardi, 1975; Prunell et al., 1977).The underestimate appears to be larger, 12%, in the case of strains K and M as studied in other laboratories (Sanders et al., 1977; Morimoto and Rabinowitz, 1979a). By subtracting introns aI5α and β of the *oxi3* gene and introns bI1-3 of the *cob* gene from the size of long genomes one can calculate a value of 78 480 bp for the size of short genomes. This estimate is again higher, by 11%, than that previously derived from restriction fragments, 70 400 bp. Finally, by subtracting introns aI1, aI4 of the *oxi3* gene and intron rI of the *21S* gene, which are absent in the

TABLE III

Interstrain differences: genome unit sizes and facultative intervening sequences (introns)

Group[a]	Strain	Size (bp) as determined from			
		Restriction fragments[b]			Sequences[b,c]
Long[d]	A	79 500	(1–3)		85 140
	K	75 750	(4,5)		
	M	76 100	(6)		
"Insertions"	II (a,b)	2 600	(4,5)	Introns aI5 α,β (7)	2901
	III (a–c)	3 000		bI1-3 (8)	3757
		5 600			6658
Short	D1,D2	70 400	(9)		78 480
	JS1-3D	70 750	(4,5)		
"Insertions"	IV	1 500	(4,5)	Introns aI1 (10)	2,452
	I	900		aI4 (10)	1010
	VI	1 050		rI (11)	1143
		3 450			4605
Supershort	C	69 000	(1,2,12)		73 880
		68 000	(4,5)		

[a] The supershort, short and long groups were originally defined as groups I, II and III, respectively (Sanders et al., 1977). It should be noted that, since the *21S* intron rI may be present or absent in long genomes, these belong in two sub-groups (see footnote c).

[b] References are: (1) Bernardi (1975); (2) Prunell et al. (1977); (3) de Zamaroczy et al. (1984); (4) Sanders et al. (1977); (5) Borst et al. (1977); (6) Morimoto and Rabinowitz (1979a); (7) Hensgens et al. (1983); (8) Lazowska et al. (1980); (9) Morimoto et al. (1979); (10) Bonitz et al. (1980); (11) Dujon (1980); (12) Faugeron-Fonty et al. (1984).

[c] Present work. The value given for long genome units is the sum of sequenced regions (Table I) and total gaps (Table II). The facultative introns corresponding to the "insertions" of Sanders et al. (1977) are indicated. The *21S* intron rI is present in short genomes JS1-3D, D1, D2, in the long genomes of strains I, K, F, M and A, but absent in the long genomes of strains 55R5-3C and B and in the super-short genome of strain C.

[d] Strains K and M have very similar restriction maps, which are also shared by strains F (Van Ommen et al., 1980) and I (Dujon et al., 1976; Morimoto and Rabinowitz, 1979b; see also Grivell, 1984, for detailed information on restriction maps). Strain B showed a genome unit size of only 76 500 bp (Bernardi, 1975; Prunell et al., 1977), a value lower than that of strain A, as determined under identical conditions. The size difference between the genomes of strains A and B are exactly accounted for by the absence of both the *21S* intron and the *ori4* region (see Table IV) in strain B.

TABLE IV

Interstrain differences: intragenic and intergenic sequences

Localization	Size (bp)	Presence (in strains)	Absence (in strains)	Ref.[a]
Intragenic				
1. *15S* (mini-inserts)	40	M	FF1210-6C	(1)
	nd	M	A,K	(1,2)
2. *var1* (mini-inserts)	6 + 18	D1	E	(3,4)
	nd + 18	D1	A	(3,5)
3. *21S* (mini-insert)[b]	66	I	55R5-3C	(6)
Intergenic				
4. *ori*1-*15S*	400	A,J	B,C,K	(7)
5. *oli2*-ORF4	1800	A,J,D2,K	JM6, JC6, X4005-11A	(8,9)
	> 1300	K	SC167	(10)
6. *ori*4	1900	19 strains	B	(11)

[a] References are: (1) Sor and Fukuhara (1982); (2) Martin et al. (1983); (3) Tzagoloff et al. (1980); (4) Hudspeth et al. (1982); (5) de Zamaroczy and Bernardi (unpublished data); (6) Dujon (1980); (7) de Zamaroczy (1984); (8) Nagley et al. (1981); (9) Cobon et al., (1982); (10) Naggert (1983); (11) Faugeron-Fonty et al. (1984).

[b] The *21S* mini-insert usually accompanies intron rI.

nd, not determined.

S. carlsbergensis genome, from the size of the short genomes, one obtains a size of 73 880 bp for this supershort genome; this value is again higher by about 7–9% than that previously estimated from restriction fragments (Bernardi, 1975; Prunell et al., 1977; Sanders et al., 1977).

(d) Interstrain local differences in genome sequences

Apart from the differences due to the presence or absence of facultative introns (Table III), local intragenic and intergenic differences also exist among different strains (Table IV). (i) The former comprise facultative mini-inserts in the *15S*, *21S* and *var1* genes. These are formed by two different G + C clusters in the case of the *15S* and *21S* genes (M. de Zamaroczy and G. Bernardi, paper in preparation); in the case of the *var1* gene, mini-inserts comprise two short $(AAT)_2$ and $(AAT)_6$ sequences and, in some cases, a duplicated, inverted copy of the G + C cluster present in the gene (Zassenhaus et al., 1983). (ii) The latter concern two intergenic sequences (in the *ori*1-*15S* and *oli2*-ORF4 regions, respectively), and one *ori* sequence (*ori*4).

(e) Interstrain comparison of primary structures

Table V indicates that only about 10% of the genome have been sequenced in more than one strain. These sequences are about equally divided between genes (and *ori* sequences), and intergenic sequences. Percentagewise, however, the former represent 25%, the latter only 7–8% of the sequences. Neglecting facultative introns and the mini-inserts in the *var1*, *15S*, and *21S* genes, and the rearrangements in the *ori*1 region, differences in genes or *ori* sequences comprise extremely few single-base changes and single-base additions. In contrast, intergenic sequences differ not only by single-base changes and additions, but also by insertions/deletions and local rearrangements. Fig. 2 provides two examples of these situations.

(f) Size and base composition of genes and intergenic ORFs

Details on the sizes of exons and introns of *oxi3*, *cob* and *21S* genes are given in Table VI and Fig. 3. Table VII presents a compilation of the sizes and base composition of genes (or exons) of introns (with their closed and open reading frames), and of the open reading frames present in intergenic sequences.

(g) Coding and noncoding sequences

The extent of the coding and noncoding region in

TABLE V

Comparison of regions sequenced in two strains[a]

Region	Strains	Genes or *ori* size	Intergenic sequences		
			Size	% divergence	G + C cluster[b]
1. *asp*	D1; A	72	473	3.6	*
2. *ala*	D1; S	73	257	1.6	*
3. *ori5*	A; K	280	683	1.6	+
4. *ori1*	A,J; B	—[c]	—[c]		nd
5. *15S*	A; M	411[d]	—		
6. *oli2*	D2; J	780	502[e]	3.0	+ −
7. before *ori7*	D2; B	—	257	3.1	*
8. *glu-cob*	D1; A	70 + 323	1157[f]	1.1	− + +
9. *ori6*	D1; B	250	—		
10. *oli1*	D1; K	231	169	7.7[g]	
11. *var1*	D1; E	1191	—		+
	D1; A	(450)	52		+
12. *ori3*	A; K	280	430	3.5[h]	+
		3961	3980	2.4[i]	7 +, 2 −, 3*
% of short genomes		5.1	5.1		
% of genes and *ori*		25.3			
% of intergenic sequences			7.5		

[a] See Table I, values in parentheses, for the sizes of the regions concerned and for the references. Regions were split, whenever possible, into genes, or *ori* sequences, and intergenic regions. All sizes are in bp.

[b] Homologous G + C clusters (only differing by point mutations) are indicated by a plus sign; non-homologous G + C clusters by a minus sign. When more signs follow each other they refer to the clusters present in the region, as they appear in the clockwise direction. Asterisks indicate the presence of an incomplete or uncertain G + C cluster sequence in one of the two strains; such clusters have not been taken into account in the size of the intergenic sequence. G + C clusters of *ori* sequences (de Zamaroczy et al., 1984) were not taken into consideration. nd, not determined.

[c] This region, sequenced over 800 and 1600 bp in strains A, or J, and B, respectively, differs because of rearrangements in both the *ori* sequence and its flanking stretches (Faugeron-Fonty et al., 1984).

[d] A segment of 45 and 100 bp, respectively, downstream of the gene exhibits no homology between the two strains.

[e] This region comprises a "rearranged" 80-bp stretch comprising one of the two G + C clusters (see Fig. 3). This "rearrangement" might also be overestimated because of sequencing mistakes.

[f] This region comprises an inserted 100-bp stretch (including part of a G + C cluster) in one of the two strains; this stretch is not comprised in the size value given for intergenic sequence. One of the two homologous G + C clusters is followed by a 70 or 90-bp long AT stretch which is completely different in the two strains.

[g] This high divergence is due to the fact that a number of tri- and tetra-A are longer by one A in strain K, possibly because of sequence mistakes.

[h] If an apparently rearranged 40-bp region, which is located just after *ori3* and contains 11 single-base changes, is not taken into account, divergence is only 1%.

[i] Average value.

the mitochondrial genome of yeast is evaluated in Table VIII on the basis of all available data.

(h) Transcription initiation

In vitro capping experiments and S1 nuclease mapping have been used to identify and localize the sites corresponding to transcription initiations (see Table IX for references). These experiments have also shown the existence of other sites (particularly in the tRNA cluster region), but these have not been localized yet. The twelve transcription initiation sites localized so far on the mitochondrial genome correspond to the sequence $\frac{A}{T}$TATAAGTA, except for

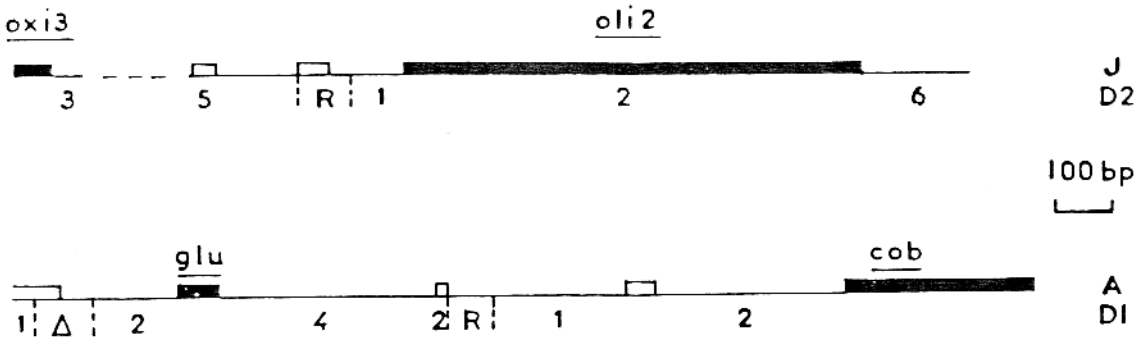

Fig. 2. Differences in two intergenic sequences, as determined in two pairs of different strains. Black bars correspond to genes or their exons; open bars to G + C clusters. The broken line between the *oxi3* and *oli2* genes indicates a 1400-bp stretch which has not been sequenced in strain D2. Symbols below the line indicate the differences found in strains D2 and D1, relative to strain J and A, respectively. Numerals indicate the number of single-base changes and/or additions in different sequence stretches; *Δ*, a deletion/insertion; R, rearranged regions.

that of the *glu* gene (ATATAGGTA) and that of ORF5 (AAATAAGTT). Most of these sites are isolated, except for two tandem pairs preceding *fmet* and *oli1*. Their locations are 10 to 600 bp upstream or immediately before the corresponding gene(s); in the case of the *21S* gene, the last position of the nonanucleotide is the first position of the gene. One site is present inside the *var1* gene. Finally, some sites correspond to the initiation of RNA primers used in the DNA replication (*ori*1, 2, 3 and 5). Fig. 1 and Table IX show the approximate and precise locations, respectively, of these sites. Most genes are cotranscribed into primary polycistronic transcripts (see Fig. 1).

(i) Transcript processing and termination

The sequence AATAAPyATTCTT (or ATATATAAACTT in the case of ORF5) corresponds to transcript cleavage sites for polycistronic transcripts or to terminations (see Table IX for references), the latter also possibly corresponding to cleavages (Osinga et al., 1984b). Finally, three sites do not correspond to any demonstrated cleavage. The locations of these sites are given in Fig. 1 and Table IX.

DISCUSSION

(a) The sequenced regions

As shown in Tables I–III, a total of 87–88% of the primary structure of the yeast mitochondrial genome have been determined in one or another strain. This percentage drops, however, by a factor of 2, to 47%, if one considers the sequences determined in the two closely related strains, D1 and D2, which have been mostly investigated by Tzagoloff and co-workers. We think, however, that there are good reasons for putting together into an overall compilation sequences determined in different strains. The rationale for such a compilation (as presented in Table I) should be discussed separately for genic (or *ori*) sequences and for intergenic sequences.

As far as genes and *ori* sequences are concerned, 25% of them have been determined in at least two strains (see Table V). Gene sequences are extremely conserved, exceptionally few point mutations and single-base deletions or additions being found in different strains. Only in some special cases (*15S*, *21S* and *var1* genes), small insertions/deletions have been found (Table IV). Taking into account the practical certainty that the sequences which have been determined are representative of all genic sequences as present in *S. cerevisiae* strains, it seems justified to consider that genic sequences are the same in different strains, except for the minor variations mentioned above and, obviously, for the major variations associated with the presence of facultative introns in some genes (Table III). In the case of *ori* sequences, conservation is almost perfect for *ori*3, 5 and 6, whereas the whole *ori1* region exhibits large differences in different strains and *ori*4 is missing in strain B (see Table IV).

If we now consider the same problem for inter-

TABLE VI

Structure of split genes *oxi3*, *cob*, and *21S*[a]

Exon	Intron	Exon size	Intron size				Gene size
			Total	ORF	CRF	ORF	
oxi3							
A1	[aI1]	168	2452	2337	100	15	
A2	aI2	36	2512	2358	61	93	
A3	aI3	39	1514	1005	485	24	
A4	[aI4]	477	1010	951	20	39	
A5α	(aI5α)	252	1365	921	363	81	
A5β	(aI5β)	135	1536	36	338		
				996	148	18	
A5γ	aI6	24	887	24	779	84	
A6	aI7	303	21	21			
A7	aI8	30	45	45			
A8		75					
Short genomes		1539	8441				9980
Long genomes		1539	11342				12881
cob							
B1	(bI1)	415	765		650	115	
B2	(bI2)	14	1292	846	446[b]		
B3	(bI3)	84[b]	1700[b]	1158[c]	542[c]		
B4	bI4	246[b]	1414	1158	256		
B5	bI5	54	730		730		
B6		351					
Short genomes		1164	2144				3308
Long genomes		1164	5901				7065
21S							
R1		2622					
[Mini-insert][d]		66					
[rI][d]			1143		295		
				708	140		
R2		585					
Supershort genomes		3207	—				3207
Short and some long genomes		3273	1143				4416

[a] Sizes are given in bp. Introns in parentheses are only present in long genomes; introns in square brackets are only present in long and short genomes (see Table III and Fig. 3). References for introns are given in Table I.

[b] The size of CRF in bI2, and of B3, B4 and bI3 are only estimated because of the presence of gaps 13 and 14, respectively.

[c] The sizes of ORF and CRF of bI3 can be estimated on the basis of the map of Rödel et al. (1983).

[d] Intron rI and the accompanying mini-insert account for the difference between short and super-short genomes (see Table III).

genic sequences, the justification for the overall compilation of Table I comes from two independent lines of evidence. The first emerges from the detailed comparison of the *Hae*III or *Hpa*II fragments derived from the mitochondrial genomes of four different yeast strains (A, B, C, D) and of crosses of two of them (A and B; see Table I for strain nomenclature). These numerous fragments (71–84 fragments in the case of *Hae*III; 107–113 fragments in the case of *Hpa*II) correspond to restriction sites

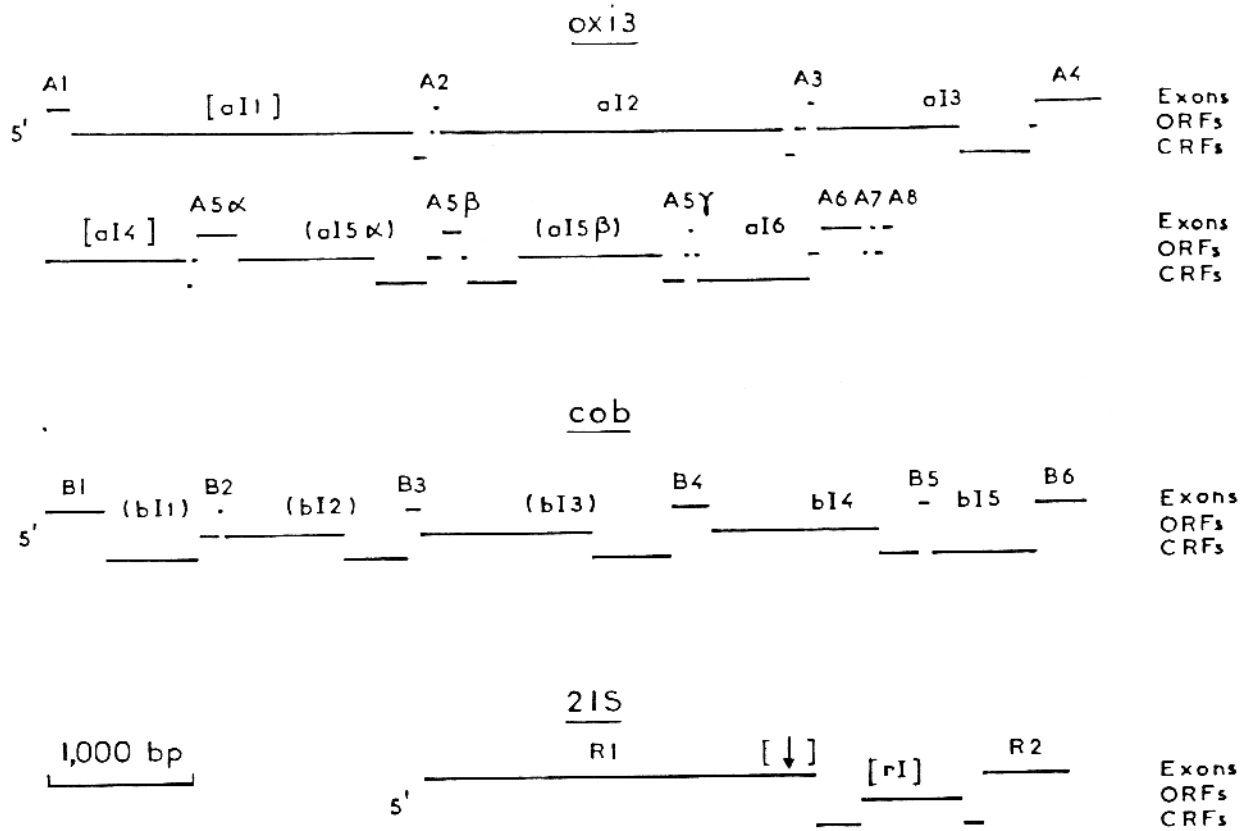

Fig. 3. A scheme of the mosaic genes *oxi3*, *cob* and *21S* (see also Table VI). Exons, ORFs and CRFs are shown on three different lines. Introns in parentheses are only present in long genomes; introns in square brackets are only present in long and short genomes. The arrow on the *21S* gene indicates the position of the mini-insert. References for introns are given in Table I.

located in G + C clusters of intergenic sequences. The study of *Hae*III and *Hpa*II fragments (Bernardi et al., 1975; 1976; Bernardi, 1975; 1976; Prunell et al., 1977; Fonty et al., 1978), not only provided the first biochemical estimation of mitochondrial genome size and the first evidence of interstrain genome size differences, but also indicated a high level of homology in intergenic sequences. Indeed, (i) the genomes of all strains released approximately the same number of fragments and comprised therefore about the same number of G + C clusters containing *Hae*III and *Hpa*II sites; (ii) the size distributions of *Hae*III and *Hpa*II fragments were similar indicating that the corresponding G + C clusters were similarly distributed in the genomes. These results extended previous ones (Bernardi et al., 1970; Bernardi and Timasheff, 1970; Ehrlich et al., 1972; Prunell and Bernardi, 1974) which indicated an essential identity of all physical and chemical properties of the DNAs under consideration (G + C level, buoyant density, melting transition, pyrimidine isostichs, A + T spacers, G + C clusters). The small interstrain differences in the sizes of *Hae*III and *Hpa*II fragments were accounted for by small additions/deletions. Other approaches along the line just mentioned were the demonstration (i) that extensive fragment size identity exists among larger fragments as obtained by degradation with other restriction enzymes (see for example Bernardi, 1975; Prunell et al., 1977; Sanders et al., 1977; Morimoto et al., 1979); (ii) that great similarities exist in the restriction maps of genomes from different strains, (Sanders et al., 1977; Morimoto et al., 1979; Morimoto and Rabinowitz, 1979a,b; Faugeron-Fonty et al., 1984). Moreover, a number of differences can be pinned down to point mutations and/or small deletions/insertions destroying restriction sites (Faugeron-Fonty et al., 1984).

Another line of evidence in favor of interstrain homology among intergenic sequences comes from the direct sequence comparisons of Table V. In this case, the comparison only concerned 7–8% of all

TABLE VII

Size and G + C contents of genes and intergenic ORFs[a]

Map position		Sizes				G + C
		Gene or exons	Introns		Intergenic ORF	
			ORF	CRF		
24 tRNA[b]		1778				32.8*
oxi1	(11100)	756				27
ORF1	(12100)				1161[c]	19
oxi2	(16100)	813				30
ORF2	(17100)				1278[c]	15.7
9S[b]	(23600)	431				13.2*
15S	(30250)	1649[d]				22
ORF3	(32350)				243	21.0*
oxi3	(36600)	1533				31.1*
			6996	1445		24.1*; 24.5*
			9048	2294[e]		22.2*; 20.8*[f]
aap1	(50450)	147				18.4*
oli2	(51300)	780				22.6*
ORF4	(nd)				500[c]	nd
ORF5	(55000)				141	15.7
cob	(61100)	1158				27.8
			1158	986		19.5*; 16.1*
			3276	2625[e]		19.0*; 16.0*[f]
oli1	(71400)	231				33.3
var1	(73600)	1,215				10.2*
		(1,191)[g]				10.4*
21S	(81400)	3273[d]				21.4*
			708	435[h]		20.0*; 20.9*
Totals		13764	8862	2866	3323	(short genomes)
		13764	13032	5354	3323	(long genomes)

[a] Values for G + C levels (%) are from the references or from our estimates (asterisks). Sizes are in bp and comprise termination codon(s); figures in parentheses were not included in the totals. Map positions (bp) are given relative to the *Sal*I site (which is in fact clustered with a *Hin*fI, a *Hin*cII and a *Taq*I site).

[b] A 25th tRNA gene (*trp2*) was localized by hybridization between tRNA *pro* and *ori1* of strain A (Martin et al., 1980); its sequence is not known. The G + C value is the average value of all tRNA genes, which range from 18 to 44%. The size of tRNA genes ranges from 70 to 87 bp. The *9S* "gene" corresponds to the central part of the tRNA synthesis locus of Miller et al. (1983).

[c] An extension of ORF1 towards the *oxi1* gene is possible (Michel, 1984; see Fig. 1). The size of ORF2 does not comprise its two G + C clusters (Michel, 1984); the size of ORF4 is only estimated (Macino and Tzagoloff, 1980) because of the presence of sequence gap 11 (see Table II).

[d] These values comprise facultative mini-inserts (see Table V). The size of the *15S* gene was determined by taking into account S1 degradation experiments on the 5′ end and the comparison with the *16S* RNA from *E. coli* for the 3′ end (Sor and Fukuhara, 1980; Osinga and Tabak, 1982; Bordonné, 1982).

[e] Intron size. The long genome of KL14-4A differs from the short one of D273-10B in having additional introns aI5α, aI5β. The long genome of 777-3A differs from the short one of D273-10B in having additional introns bI1,-2,-3 (see Table III). The size of intron bI3, which is not yet sequenced, is estimated as 1700 bp (Jacq et al., 1982); the sizes of its ORF and CRF can be estimated as 1158 and 542 bp on the basis of the map of Rödel et al. (1983).

[f] The additional ORFs and CRFs present in long genomes have G + C levels of 16.0% and 14.7% in the case of *oxi3*, of 18.5% and 15.9% undetermined G + C levels in the case of *cob*.

[g] Values given refer to the two different alleles, as sequenced in D273-10B and A10, respectively.

[h] This facultative intron is present in short genomes and some long genomes (see Table III; footnote c).

TABLE VIII

Size and G + C contents of coding and noncoding sequences[a]

	Genomes		Genomes	
	Short	Long	Short	Long
	Size (%)		G + C content (%)	
1. Total genome	100	100	17.8[b]	
2. Genes or exons	17.5	16.2	24.6	
3. Introns	14.9	21.6	23.0	20.8
a. ORF[c]	11.3	15.3	23.6	21.7
b. CRF	3.7	6.3	20.9	18.7
4. Intergenic regions (1-2-3)	67.6	62.2	15.0[d]	
a. Intergenic ORF	4.2	3.9	17.5	
b. *ori* sequences	2.9	2.6	18.7	
5. Coding regions				
a. Actual (2)	17.5	16.2		
b. Putative (3a)	11.3	15.3		
c. Potential (4a)	4.2	3.9		
d. Maximal (2 + 3a + 4a)	33.0	35.4		
6. Noncoding regions (1–5d)	67.0	64.6		

[a] All values are given as %. Some values are estimated by addition (+) or substraction (−), as indicated in parentheses. See Tables VI and VII for sizes in bp.

[b] This value, identical for both short and long genomes, was calculated by assuming that all remaining gaps (except those in ORF4 and in introns bI2 and bI3) have the G + C content of intergenic sequences.

[c] Proteins shown to be involved in RNA processing (also called "RNA maturases") are encoded by ORFs of introns aI1 (Groudinsky et al., 1983), aI2, aI4 (Grivell, 1983), aI5α, aI5β (Hensgens et al., 1983) of the *oxi3* gene and of introns bI2, bI3, bI4 of the *cob* gene (see Grivell, 1982, 1983, for reviews). (It should be noted (i) that strains totally deprived of *cob* introns can be constructed without altering the expression of the *cob* gene; Jacq et al., 1982; Labouesse and Slonimski, 1983; and (ii) that the presence of the bI4 intron is both necessary and sufficient for the expression of the *oxi3* gene; its absence can, however, be compensated by a nuclear mutation, *NAM2*; Dujardin et al., 1983). A protein involved in intron-transposition is encoded in the ORF of intron rI (Butow, 1985; B. Dujon, personal comm.).

[d] As based on the sequenced 2/3 of intergenic regions.

intergenic sequences. Even if this value is admittedly small, the sequences which have been compared were not selected in any way and are very likely to be representative of intergenic sequences. In this case, the degree of interstrain divergence is comprised between 1.1 and 3.6% (a higher value of 7.7% is accounted for in footnote g of Table V); these values can only be overestimates because of sequence errors which have unfortunately been frequent in intergenic sequences (see the examples quoted by de Zamaroczy et al., 1984). A realistic estimate of interstrain divergence would probably be 1–2%, which obviously corresponds to a remarkable level of sequence conservation. It is of interest to note that G + C clusters in homologous sequences from different strains are, in their majority, identical in primary structure, or, at least, in location. In the latter case, the rare large changes are in all likelihood the result of unequal crossing-overs (Prunell et al., 1977). The data of Table V also provides evidence (see footnotes c-f) for some insertions/deletions and rearrangements in intergenic sequences (see also Fig. 2).

(b) The size and copy number of genome units

The preceding section justifies the overall compilation of sequenced regions from different strains, as presented in Table I, as well as the estimation of the sizes of inter-regional gaps of Table II; intra-

TABLE IX

Initiation and processing sites for transcription

Initiation sites	Location[a]	Ref.[b]
ATATAAGTA	*21S*	(1,2)
TTATAAGTA	*phe* (− 25)	(2)
TTATAAGTA	*thr1* (+ 135)	(3)
TTATAAGTA	*fmet* (− 28)	(4)
ATATAAGTA	*fmet* (− 15)	(4)
ATATAAGTA	*15S*	(1,2)
ATATAAGTA	*oxi3* (− 540)	(3)
AAATAAGTT	ORF5(− 338)	(5)
ATATAGGTA	*glu* (− 391)	(6)
ATATAAGTA	*oli1* (− 621)	(2,7)
ATATAAGTA	*oli1* (− 544)	(2,7)
TTATAAGTA	*var1* (481)	(8)
ATATAAGTA	*ori5*	(2,3,9)
ATATAAGTA	*ori1*	(9)
ATATAAGTA	*ori2*	(2,3,9)
ATATAAGTA	*ori3*	(2,3,9)
Processing sites		
AATAATATTCTT	rI-ORF	(10)
AATAATATTCTT	*oxi1* (+ 63)	(10)
AATAATATTCTT	ORF1(+ 215)	(10)
AATAATATTCTT	ORF2(790)	(10)
AATAATATTCTT	ORF2(+ 120)	(11)
AATAATATTCTT	a15β-ORF(+ 144)	(10)
AATAATATTCTT	*oxi3* (+ 75)	(10)
AATAATATTCTT	ORF4(nd)	(10)
ATATATATTCTT	ORF5(+ 403)	(5)
AATAATATTCTT	*cob* (+ 98)	(10)
AATAACATTCTT	*oli1* (+ 89)	(10)
AATAATATTCTT	*var1* (+ 524)	(10)

[a] Values in parentheses give the distance in bp from the proximal end of the indicated landmark or from the 5′ end for sites located inside the landmark. Upstream and downstream positions are indicated by (−) and (+) signs, respectively.

[b] References are: (1) Osinga and Tabak (1982); (2) Christianson and Rabinowitz (1983a); (3) Osinga et al. (1984a); (4) Miller et al. (1983); (5) Colin et al. (1985); (6) Christianson and Rabinowitz (1983b); (7) Edwards et al. (1983); (8) Zassenhaus et al. (1983); (9) Baldacci and Bernardi (1982); (10) Osinga et al. (1984b); (11) Michel (1984).

regional gaps obviously raise no problem. These data can be combined to estimate the genome unit sizes presented in Table III. These estimates are larger than those obtained from restriction fragments by 7–11% for both short and long genomes and by 7–9% for the supershort genome of *S. carlsbergensis*.

It may be of interest now to calculate the copy number of mitochondrial genome units for strain A, a case in which we have a reliable estimate of the amount of mitochondrial DNA relative to total DNA (13.3%; Bernardi et al., 1972). If we use the nuclear genome size estimate of Lauer et al. (1977; see also Broach et al., 1983), $9 \pm 2 \times 10^9$ Dal, this copy number is comprised between 21 and 33 units per haploid cell, a number about half that usually quoted of 50. The relative amount of mitochondrial DNA may vary, in different strains from 14 to 24%, possibly, however, as a result of different growth conditions (Hall et al., 1976). If this latter upper estimate is considered, the copy number might reach values as high as 38–60 units per haploid cell under certain circumstances.

(c) Coding and noncoding sequences

The data of Table VIII indicate that an actual coding role has only been demonstrated so far for about 17% of the genome. This value can be increased to about 30% if one takes into account the intronic ORFs and to about 34% by considering, in addition, the intergenic ORFs.

As far as the latter are concerned, the available evidence for transcription can be summarized as follows (Fig. 1). Three intergenic ORFs are represented in the tails of transcripts originating from other genes. ORF1 is represented in the transcript of *oxi1* and in another 1500-nt transcript copied from the opposite strand (Coruzzi et al., 1981). ORF4 is represented in the *oxi3-aap1-oli2* transcript (Macino and Tzagoloff, 1980; Cobon et al., 1982; Osinga et al., 1984b) and ORF2 in the *oxi2* transcript (Thalenfeld and Tzagoloff, 1980). It should be noted (i) that ORF4 is absent in some strains (Nagley et al., 1981; Cobon et al., 1982); (ii) that ORF2 is made up of three short ORFs which are separated by two G + C clusters and are out of phase with each other (Michel, 1984); and that the corresponding transcript has been found to be shorter by 2000 nt on the 3′-end side in a particular yeast strain, LL20 (Thalenfeld et al., 1983). No information exists on the transcription of ORF3 (an ORF called RF2 by Martin et al., 1983). In contrast with the other intergenic ORFs, ORF5 corresponds to a specific transcript (Colin et al., 1985). Another class of potential coding sequences is represented by the short ORFs

present in *ori1*, *2*, *3* and *5* (de Zamaroczy et al., 1984).

In any case, even by maximally stretching the potential coding sequences to include all intronic and intergenic ORFs, one would reach a total of about 34% as the estimate of coding sequences in the mitochondrial genome of yeast. Some additional coding sequences might be discovered in the regions which have not yet been sequenced, but since the latter only account for 12–13% of the genome, this would not significantly change the conclusion that at least 2/3 of the mitochondrial genome of yeast is certainly made up of noncoding sequences.

(d) G + C levels of the genome

As shown by the values of Table VIII, summarizing those of Table VII, the G + C content of the mitochondrial genome is calculated as 17.8%, in good agreement with experimental results of 17.4% and 16.8% for strains A and B (Bernardi et al., 1970). The genes or exons have a G + C level of 24.6%, again in good agreement with our previous estimate (Prunell and Bernardi, 1974), whereas the intronic ORFs have a distinctly lower value of 22%. Increasingly lower values are shown by intronic CRFs, 20%, *ori* sequences, 19%, intergenic ORFs, 17.5%, intergenic sequences, 15%.

ACKNOWLEDGEMENTS

We thank Dr. R.P. Martin for providing us with sequence data prior to publication.

REFERENCES

Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G.: Sequence and organization of the human mitochondrial genome. Nature 290 (1981) 457–465.

Anderson, S., De Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G.: Complete sequence of bovine mitochondrial DNA. J. Mol. Biol. 156 (1982) 683–717.

Baldacci, G. and Bernardi, G.: Replication origins are associated with transcription initiation sequences in the mitochondrial genome of yeast. EMBO J. 1 (1982) 987–994.

Berlani, R.E., Pentella, C., Macino, G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system: isolation of mitochondrial transfer ribonucleic acid mutants and characterization of transfer ribonucleic acid genes of *Saccharomyces cerevisiae*. J. Bacteriol. 141 (1980a) 1086–1097.

Berlani, R.E., Bonitz, S.G., Coruzzi, G., Nobrega, M. and Tzagoloff, A.: Transfer RNA genes in the *cap-oxi1* region of yeast mitochondrial DNA. Nucl. Acids Res. 21 (1980b) 5017–5030.

Bernardi, G.: Molecular genetics of yeast mitochondria, in Bernardi, G. and Gros, F. (Eds.), Organization and Expression of the Eukaryotic Genome. Biochemical Mechanisms of Differentiation in Prokaryotes and Eukaryotes. Proceedings of the Tenth FEBS Meeting, 38 (1975) 41–56.

Bernardi, G.: The mitochondrial genome of yeast: organization and recombination, in Bücher, Th., Neupert, W., Sebald, W. and Werner, S. (Eds.), Genetics and Biogenesis of Chloroplasts and Mitochondria. Elsevier, Amsterdam, 1976, pp. 503–510.

Bernardi, G., Carnevali, F., Nicolaieff, A., Piperno, G. and Tecce, G.: Separation and characterization of a satellite DNA from a yeast cytoplasmic "petite" mutant. J. Mol. Biol. 37 (1968) 493–505.

Bernardi, G., Faures, M., Piperno, G. and Slonimski, P.P.: Mitochondrial DNA's from respiratory-sufficient and cytoplasmic respiratory-deficient mutant yeast. J. Mol. Biol. 48 (1970) 23–42.

Bernardi, G., Piperno, G. and Fonty, G.: The mitochondrial genome of wild-type yeast cells, I. Preparation and heterogeneity of mitochondrial DNA. J. Mol. Biol. 65 (1972) 173–189.

Bernardi, G., Prunell, A. and Kopecka, H.: An analysis of the mitochondrial genome of yeast with restriction enzymes, in Puiseux-Dao, S. (Ed.), Molecular Biology of Nucleocytoplasmic Relationships, Elsevier, Amsterdam, 1975, pp. 85–90.

Bernardi, G., Prunell, A., Fonty, G., Kopecka, H. and Strauss, F.: The mitochondrial genome of yeast: organization, evolution and the petite mutation, in Saccone, C. and Kroon, A.M. (Eds.), The Genetic Function of Mitochondrial DNA (Proc. 10th Int. Bari Conf.). Elsevier, Amsterdam, 1976, pp. 185–198.

Bernardi, G. and Timasheff, S.N.: Optical rotatory dispersion and circular dichroism. Properties of yeast mitochondrial DNA's. J. Mol. Biol. 48 (1970) 43–52.

Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A.: Sequence and gene organization of mouse mitochondrial DNA. Cell 26 (1981) 167–180.

Blanc, H.: Two modules from the hypersuppressive *rho*⁻ mitochondrial DNA are required for plasmid replication in yeast. Gene 30 (1984) 47–61.

Blanc, H. and Dujon, B.: Replicator regions of the yeast mitochondrial DNA responsible for suppressiveness. Proc. Natl. Acad. Sci. USA 77 (1980) 3942–3946.

Bonitz, S.G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. Sequences of yeast mitochondrial tRNA genes. J. Biol. Chem. 255 (1980) 9075–9081.

Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E. and Tzagoloff, A.:

Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for subunit 1 of yeast cytochrome oxidase. J. Biol. Chem. 255 (1980) 11927–11941.

Bonitz, S.G., Homison, G., Thalenfeld, B.E. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. Processing of the apocytochrome *b* precursor RNAs in *Saccharomyces cerevisiae*. J. Biol. Chem. 257 (1982) 6268–6274.

Bordonné, R.: Structures primaires de gènes mitochondriaux de la levure *Saccharomyces cerevisiae*. Thèse de 3è cycle, Univ. L. Pasteur de Strasbourg, 1982.

Borst, P., Bos, J.L., Grivell, L.A., Groot, G.S.P., Heyting, C., Moorman, A.F.M., Sanders, J.P.M., Talen, J.L., Van Kreijl, C.F. and Van Ommen, G.J.B.: The physical map of yeast mitochondrial DNA anno 1977, in Bandlow, W., Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), Mitochondria 1977: Genetics and Biogenesis of Mitochondria. De Gruyter, Berlin, 1977, pp. 213–254.

Broach, J.R., Li, Y.-Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A. and Hicks, J.B.: Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harbor Symp. Quant. Biol. 47 (1983) 1165–1173.

Butow, R.A.: Nonreciprocal exchanges in the yeast mitochondrial genome – review. Trends in Genet. 1 (1985) 81–84.

Christianson, T. and Rabinowitz, M.: Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by in vitro capping with guanylyltransferase. J. Biol. Chem. 258 (1983a) 14025–14033.

Christianson, T., Edwards, J.C., Mueller, D.M. and Rabinowitz, M.: Identification of a single transcriptional initiation site for the glutamic tRNA and *cob* genes in yeast mitochondria. Proc. Natl. Acad. Sci. USA 80 (1983b) 5564–5568.

Clary, D.O. and Wolstenholme, D.R.: The *Drosophila* mitochondrial genome, in Oxford Surveys on Eukaryotic Genes. 1984, Vol. 1.

Cobon, G.S., Beilharz, M.W., Linnane, A.W. and Nagley, P.: Biogenesis of mitochondria: mapping of transcripts from the *oli2* region of mitochondrial DNA in two grande strains of *Saccharomyces cerevisiae*. Curr. Genet. 5 (1982) 97–107.

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

Cosson, J. and Tzagoloff, A.: Sequence homologies of (guanosine + cytidine)-rich regions of mitochondrial DNA of *Saccharomyces cerevisiae*. J. Biol. Chem. 254 (1979) 42–43.

Corruzzi, G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. DNA sequence of subunit 2 of yeast cytochrome oxidase. J. Biol. Chem. 254 (1979) 9324–9330.

Coruzzi, G., Bonitz, S.G., Thalenfeld, B.E. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. Analysis of the nucleotide sequence and transcripts in the *oxi1* region of yeast mitochondrial DNA. J. Biol. Chem. 256 (1981) 12780–12787.

Dujardin, G., Labouesse, M., Netter, P. and Slonimski, P.P.: Genetic and biochemical studies of the nuclear suppressor *NAM2*: extraneous activation of a latent pleiotropic maturase, in Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), Mitochondria 1983, Nucleo–Mitochondrial Interactions, De Gruyter, Berlin, 1983, pp. 233–250.

Dujon, B.: Sequence of the intron and flanking exons of the mitochondrial *21S* rRNA gene of yeast strains having different alleles at the ω and *rib1* loci. Cell 20 (1980) 185–197.

Dujon, B., Bolotin-Fukuhara, M., Coen, D., Deutsch, J., Netter, P., Slonimski, P.P. and Weill, L.: Mitochondrial genetics, XI. Mutations at the mitochondrial locus ω affecting the recombination of mitochondrial genes in *Saccharomyces cerevisiae*. Mol. Gen. Genet. 143 (1976) 131–165.

Edwards, J.C., Osinga, K.A., Christianson, T., Hensgens, L.A.M., Janssens, P.M., Rabinowitz, M. and Tabak, H.F.: Initiation of transcription of the yeast mitochondrial gene coding for ATPase subunit 9. Nucl. Acids Res. 11 (1983) 8269–8282.

Ehrlich, S.D., Thiery, J.-P. and Bernardi, G.: The mitochondrial genome of wild-type yeast cells, III. The pyrimidine tracts of mitochondrial DNA. J. Mol. Biol. 65 (1972) 207–212.

Faugeron-Fonty, G., Le Van Kim, C., de Zamaroczy, M., Goursot, R. and Bernardi, G.: A comparative study of the *ori* sequences from the mitochondrial genomes of 20 wild-type yeast strains. Gene 32 (1984) 459–473.

Fonty, G., Goursot, R., Wilkie, D. and Bernardi, G.: The mitochondrial genome of wild-type yeast cells, VII. Recombination in crosses. J. Mol. Biol. 119 (1978) 213–235.

Gaillard, C. and Bernardi, G.: The nucleotide sequence of the mitochondrial genome of a spontaneous "petite" mutant of yeast. Mol. Gen. Genet. 174 (1979) 335–337.

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

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

Grivell, L.A.: Mitochondrial gene expression 1983, in Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), Mitochondria 1983: Nucleo–Mitochondrial Interactions, De Gruyter, Berlin, 1983, pp. 25–45.

Grivell, L.A.: Restriction and genetic maps of yeast mitochondrial DNA, in O'Brien, S.J. (Ed.), Genetic Maps – 1984, Vol. 3, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 234–247.

Grivell, L.A., Arnberg, A.C., Boer, P.H., Borst, P., Bos, J.L., van Bruggen, E.F.J., Groot, G.S.P., Hecht, N.B., Hensgens, L.A.M., van Ommen, G.J.B. and Tabak, H.F.: Transcripts of yeast mitochondrial DNA and their processing. ICN–UCLA Symp. Mol. Cell. Biol. 15 (1979) 305–324.

Grivell, L.A., Hensgens, L.A.M., Osinga, K.A., Tabak, H.F., Boer, P.H., Crusius, J.B.A., Van der Laan, J.C., De Haan, M., Van der Horst, G., Evers, R.F. and Arnberg, A.C.: RNA processing in yeast mitochondria, in Slonimski, P.P., Borst, P. and Attardi, G. (Eds.), Mitochondrial Genes. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, pp. 225–239.

Groudinsky, O., Carignani, G., Schiavon, E., Frezza, D., Bergantino, E. and Slonimski, P.P.: The first intron of the gene *oxi3* in yeast mitochondria encodes a mRNA-maturase, in Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), Mitochondria 1983; Nucleo–Mitochondrial Interactions. De

Gruyter, Berlin, 1983, pp. 227–232.

Hall, R.M., Nagley, P. and Linnane, A.W.: Biogenesis of mitochondria. XLII. Genetic analysis of the control of cellular mitochondrial DNA levels in *Saccharomyces cerevisiae*. Mol. Gen. Genet. 145 (1976) 169–175.

Hensgens, L.A.M., Bonen, L., De Haan, M., Van der Horst, G. and Grivell, L.A.: Two intron sequences in yeast mitochondrial *cox1* gene: homology among URF-containing introns and strain-dependent variation in flanking exons. Cell 32 (1983) 379–389.

Hudspeth, M.E.S., Ainley, W.M., Shumard, D.S., Butow, R.A. and Grossman, L.I.: Location and structure of the *var1* gene on yeast mitochondrial DNA: nucleotide sequence of the *40.0* allele. Cell 30 (1982) 617–626.

Jacq, C., Pajot, P., Lazowska, J., Dujardin, G., Claisse, M., Groudinsky, O., de la Salle, H., Grandchamp, C., Labouesse, M., Gargouri, A., Guiard, B., Spyridakis, A., Dreyfus, M. and Slonimski, P.P.: Role of introns in the yeast cytochrome-*b* gene: *cis*- and *trans*-acting signals, intron manipulation, expression and intergenic communications, in Slonimski, P.P., Borst, P. and Attardi, G. (Eds.), Mitochondrial Genes. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, pp. 155–183.

Labouesse, M. and Slonimski, P.P.: Construction of novel cytochrome *b* genes in yeast mitochondria by subtraction or addition of introns. EMBO J. 2 (1983) 269–276.

Lauer, G.D., Roberts, T.M. and Klotz, L.C.: Determination of the nuclear DNA content of *Saccharomyces cerevisiae* and implications for the organization of DNA in yeast chromosomes. J. Mol. Biol. 114 (1977) 507–526.

Lazowska, J., Jacq, C. and Slonimski, P.P.: Sequence of introns and flanking exons in wild-type and *box3* mutants of cytochrome *b* reveals an interlaced splicing protein coded by an intron. Cell 22 (1980) 333–348.

Li, M. and Tzagoloff, A.: Assembly of the mitochondrial membrane system: sequences of yeast mitochondrial valine and an unusual threonine tRNA gene. Cell 18 (1979) 47–53.

Macino, G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system: sequence analysis of a yeast mitochondrial ATPase gene containing the *oli2* and *oli4* loci. Cell 20 (1980) 507–517.

Martin, R.P., Bordonné, R. and Dirheimer, G.: The paromomycin in the yeast mitochondrial genome, in Akoyounoglou, G., Evangelopoulos, E.A., Georgatsos, J., Palaiologos, G., Trakatellis, A. and Tsiganos, C.P. (Eds.), Cell Function and Differentiation, Part B. Liss, New York, 1983, pp. 355–365.

Martin, R.P., Sibler, A.-P., Bordonné, R., Canaday, J. and Dirheimer, G.: Nucleotide sequence and gene localization of yeast mitochondrial initiator tRNA *fmet* and UGA-decoding tRNA *trp*, in Kroon, A.M. and Saccone, C. (Eds.), The Organization and Expression of the Mitochondrial Genome, Elsevier, Amsterdam, 1980, pp. 311–314.

Michel, F.: A maturase-like coding sequence downstream of the *oxi2* gene of yeast mitochondrial DNA is interrupted two GC clusters and a putative end-of-messenger signal. Curr. Genet. 8 (1984) 307–317.

Miller, D.L., Martin, N.C., Pham, H.D. and Donelson, J.E.: Sequence analysis of two yeast mitochondrial DNA fragments containing the genes for tRNA-*ser*/UCR and tRNA-*phe*/UUY. J. Biol. Chem. 22 (1979) 11735–11740.

Miller, D.L., Underbrink-Lyon, K., Najarian, D.R., Krupp, J. and Martin, N.C.: Transcription of yeast mitochondrial tRNA genes and processing of tRNA gene transcripts, in Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), Mitochondria 1983: Nucleo-Mitochondrial Interactions, De Gruyter, Berlin, 1983, pp. 151–164.

Morimoto, R. and Rabinowitz, M.: Physical mapping of the *Xba*I, *Hin*cII, *Bgl*II, *Xho*I, *Sst*I and *Pvu*II restriction endonuclease cleavage fragments of mitochondrial DNA of *S. cerevisiae*. Mol. Gen. Genet. 170 (1979a) 11–23.

Morimoto, R. and Rabinowitz, M.: Physical mapping of the yeast mitochondrial genome. Derivation of the fine structure and gene map of strain D273-10B and comparison with a strain (MH41-7B) differing in genome size. Mol. Gen. Genet. 170 (1979b) 25–48.

Morimoto, R., Lewin, A. and Rabinowitz, M.: Physical mapping and characterization of the mitochondrial DNA and RNA sequences from mit-mutants defective in cytochrome oxidase peptide 1 (*oxi3*). Mol. Gen. Genet. 170 (1979) 1–9.

Naggert, J.: Untersuchungen zur Struktur und zum Ausdruck mitochondrieller DNA in Hefe. Dissertation, Friedrich-Alexander-Universität, Erlangen-Nürnberg, 1983.

Nagley, P., Cobon, G.S., Linnane, A.W. and Seilharz, M.W.: Transcription of the *oli2* region of yeast mitochondrial DNA shows strain-dependent variation. Biochem. Int. 3 (1981) 473–481.

Nobrega, F.G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. DNA sequence and organization of the cytochrome *b* gene in *Saccharomyces cerevisiae* D273-10B. J. Biol. Chem. 255 (1980) 9828–9837.

Novitski, C.E., Macreadie, I.G., Maxwell, R.J., Lukins, H.B., Linnane, A.W. and Nagley, P.: Biogenesis of mitochondria: genetic and molecular analysis of the *oli2* region of mitochondrial DNA in *Saccharomyces cerevisiae*. Curr. Genet. 8 (1984) 135–146.

Osinga, K.A. and Tabak, H.F.: Initiation of transcription of genes for mitochondrial ribosomal RNA in yeast: comparison of the nucleotide sequence around the 5'-ends of both genes reveals a homologous stretch of 17 nucleotides. Nucl. Acids Res. 10 (1982) 3617–3626.

Osinga, K.A., De Vries, E., Van der Horst, G.T.J. and Tabak, H.F.: Initiation of transcription in yeast mitochondria: analysis of origins of replication and of genes coding for a messenger RNA and a transfer RNA. Nucl. Acids Res. 12 (1984a) 1889–1900.

Osinga, K.A., De Vries, E., Van der Horst, G. and Tabak, H.F.: Processing of yeast mitochondrial messenger RNAs at a conserved dodecamer sequence. EMBO J. 3 (1984b) 829–834.

Prunell, A. and Bernardi, G.: The mitochondrial genome of wild-type yeast cells. IV. Genes and spacers. J. Mol. Biol. 86 (1974) 825–841.

Prunell, A., Kopecka, H., Strauss, F. and Bernardi, G.: The mitochondrial genome of wild-type yeast cells, V. Genome evolution. J. Mol. Biol. 110 (1977) 17–52.

Rödel, G., Holl, J., Schmelzer, C., Schmidt, C., Schweyen, R.J., Weiss-Brummer, B. and Kaudewitz, F.: *Cob* intron 1 and 4:

studies on mutants and revertants uncover functional intron domains and test the validity of predicted RNA-secondary structures, in Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), Mitochondria 1983: Nucleo–Mitochondrial Interactions, De Gruyter, Berlin, 1983, pp. 191–201.

Sanders, J.P.M., Heyting, C., Verbeet, M.Ph., Meijlink, F.C.P.W. and Borst, P.: The organization of genes in yeast mitochondrial DNA, III. Comparison of the physical maps of the mitochondrial DNAs from three wild-type *Saccharomyces* strains. Mol. Gen. Genet. 157 (1977) 239–261.

Sor, F. and Fukuhara, H.: Séquence nucléotidique du gène de l'ARN ribosomique *15S* mitochondrial de la levure. C.R. Acad. Sci. Paris 291 (1980) D-933–936.

Sor, F. and Fukuhara, H.: Nature of an inserted sequence in the mitochondrial gene coding for the *15S* ribosomal RNA of yeast. Nucl. Acids Res. 11 (1982) 1625–1633.

Sor, F. and Fukuhara, H.: Complete DNA sequence coding for the large ribosomal RNA of yeast mitochondria. Nucl. Acids Res. 2 (1983) 339–348.

Thalenfeld, B.E. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. Sequence of the *oxi2* gene of yeast mitochondrial DNA. J. Biol. Chem. 255 (1980) 6173–6180.

Thalenfeld, B.E., Hill, J. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. J. Biol. Chem. 258 (1983) 610–615.

Tzagoloff, A., Nobrega, M., Akai, A. and Macino, G.: Assembly of the mitochondrial membrane system. Organization of yeast mitochondrial DNA in the *oli1* region. Curr. Genet. 2 (1980) 149–157.

Van Kreijl, C.F. and Bos, J.L.: The repeating nucleotide sequence in the repetitive mitochondrial DNA from a "low-density" petite mutant of yeast. Nucl. Acids Res. 4 (1977) 2369–2388.

Van Ommen, G.-J.B., Boer, P.H., Groot, G.S.P., De Haan, M., Roosendaal, E. and Grivell, L.A.: Mutations affecting RNA splicing and the interaction of gene expression of the yeast mitochondrial loci *cob* and *oxi3*. Cell 20 (1980) 173–183.

Wong, J.F.H., Ma, D.P., Wilson, R.K. and Roe, B.A.: DNA sequences of the *Xenopus laevis* mitochondrial heavy and light strand replication origins and flanking tRNA genes. Nucl. Acids Res. 11 (1983) 4977–4995.

de Zamaroczy, M.: L'organisation, et les origines de réplication du génome mitochondrial de la levure *Saccharomyces cerevisiae*. Thèse d'État, Université Paris VII, 1984.

de Zamaroczy, M., Marotta, R., Faugeron-Fonty, G., Goursot, R., Mangin, M., Baldacci, G. and Bernardi, G.: The origins of replication of the yeast mitochondrial genome and the phenomenon of suppressivity. Nature 292 (1981) 75–78.

de Zamaroczy, M., Faugeron-Fonty, G., Baldacci, G., Goursot, R. and Bernardi, G.: The *ori* sequences of the mitochondrial genome of a wild-type yeast strain: number, location, orientation and structure. Gene 32 (1984) 439–457.

Zassenhaus, H.P. and Butow, R.A.: Functions of nongenic DNA in yeast mitochondria, in Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), Mitochondria 1983: Nucleo–Mitochondrial Interactions, De Gruyter, Berlin, 1983, pp. 95–106.

Communicated by M.R. Culbertson.