Gene, 36 (1985) 1–13
Elsevier

GENE 1291

A new putative gene in the mitochondrial genome of *Saccharomyces cerevisiae*

(Yeast; petite mutants; origin of replication; transcriptional start point; codon usage)

Yves Colin *, Giuseppe Baldacci * 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 December 2nd, 1984)
(Accepted March 12th, 1985)

SUMMARY

The 2200-bp *ori*2–*ori*7 region of the mitochondrial (mt) genome of *Saccharomyces cerevisiae* has been sequenced on the genome of a petite, b7, excised at those *ori* sequences from wild-type strain B. The region contains an open reading frame, ORF5, which is transcribed into a 900-nucleotide (nt) RNA in both the parental wild-type strain and its derived petite, b7. This RNA uses as a template the strand used by most mt transcripts. Its start point is located 337 nt upstream of ORF5; and a messenger termination site has been found 900 nt downstream of the initiation site. These data suggest that ORF5 is a new mitochondrial gene. The G + C content of ORF5 is only 15.7%; 90% of the G + C base pairs of ORF5 are comprised in a palindromic G + C cluster similar to that present in the *var1* gene. The coding capacity of ORF5 is 46 amino acids (aa), mainly represented by methionine, phenylalanine, arginine, valine, asparagine, isoleucine and tyrosine. The aa composition and the codon usage of ORF5 are reminiscent of those of *var1* and of other intergenic ORFs.

INTRODUCTION

In the present work we have investigated the primary structure and the transcription pattern of one of the few extended regions not yet sequenced in the mt genome of *S. cerevisiae*, the *ori*2–*ori*7 region (Fig. 1).

The rationale for this research was threefold. First, we had previously observed (Marotta et al., 1982) that a spontaneous petite, b7, carrying an mt genome formed by repeat units excised between *ori*2 and *ori*7 was supersuppressive (de Zamaroczy et al., 1979; Goursot et al., 1980) and represented 50% of all spontaneous petites derived from wild-type strain B.

* Present addresses: (Y.C.) Centre National de Transfusion Sanguine, Unité INSERM U76, 6, rue A. Cabanel, 75015 Paris (France) Tel. (1)306-70-00, ext. 337; (G. Baldacci) Dipartimento di Biologia Cellulare e dello Sviluppo, Università di Roma I, 00100 Rome (Italy) Tel. (6)495-39-50.
** To whom reprint requests should be addressed.

Abbreviations: aa, amino acid(s); bp, base pair(s); EtBr, ethidium bromide; mt, mitochondrial; nt, nucleotide(s); ORF, open reading frame; *ori*, origin of replication; PA, polyacrylamide; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl, 0.015 M Na_3 · citrate, pH 7–8.

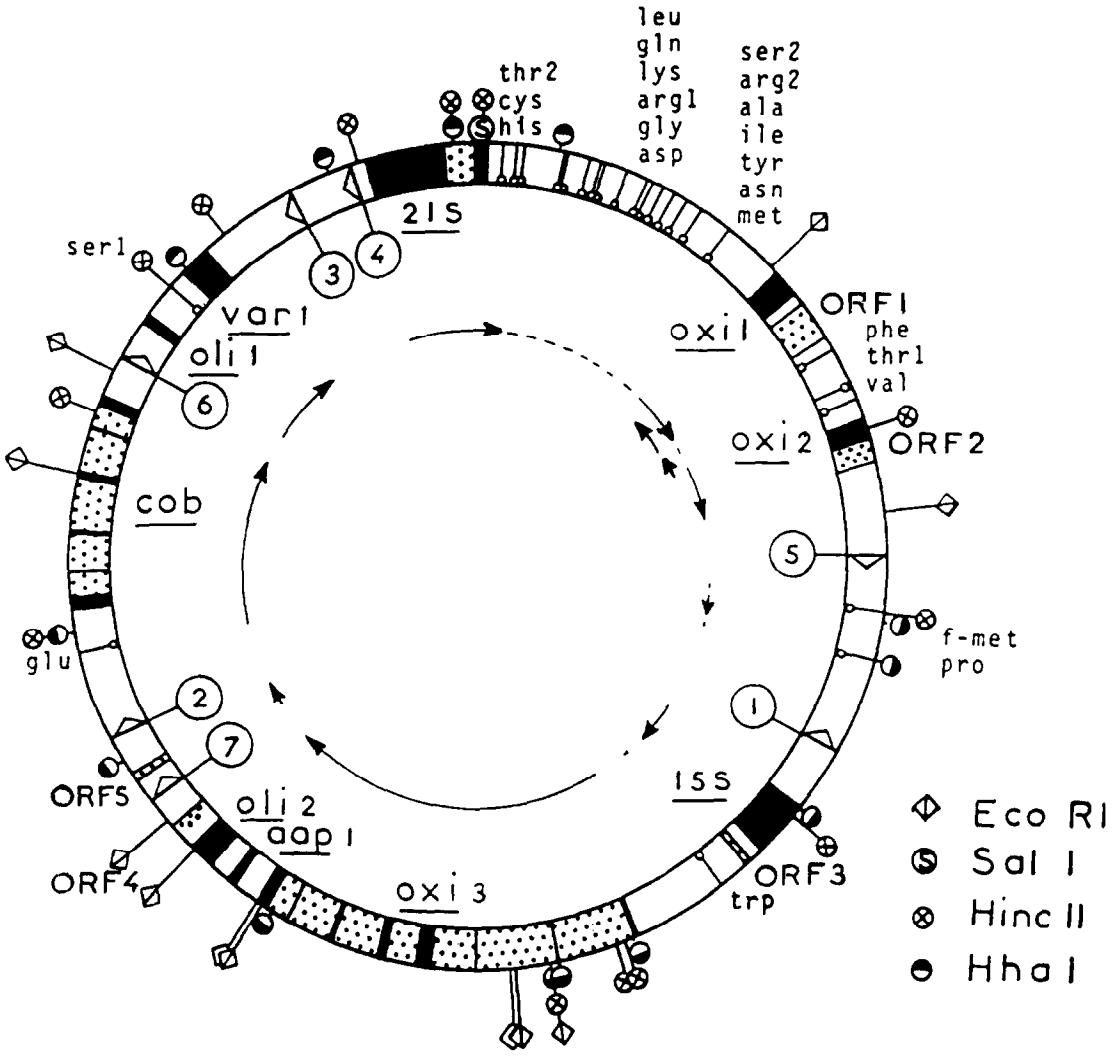


Fig. 1. Physical and genetic map of the mt genome of wild-type *S. cerevisiae* strain A (de Zamaroczy et al., 1984; Baldacci et al., 1984). Some restriction sites are indicated. Circled numbers indicate the location of *ori* sequences 1 to 7; arrowheads point in the direction from cluster C to cluster A. Black areas correspond to mitochondrial genes or their exons, stippled areas to introns and intergenic ORFs (see DISCUSSION, section c); thin radial lines indicate tRNA genes. Inner arrows specify the regions of the mt genome which have been shown to be transcribed (see Baldacci et al., 1984) and indicate the direction of transcription. The map of strain B, from which petite b7 was derived, is practically identical to that of strain A, except for a deletion of about 2000 bp including *ori*4 and for the lack of the intron of the 21S RNA gene (Faugeron-Fonty et al., 1984).

This high frequency could be understood by considering that (i) the mt genomes of spontaneous petites are generated by an excision process involving two direct repeats (Gaillard et al., 1980; Baldacci et al., 1980; de Zamaroczy et al., 1983); and (ii) *ori*2 and *ori*7 are two long, largely homologous tandem-repeated sequences separated by only 2200 bp. The b7 genome carries ori^h, a very efficient hybrid *ori* sequence formed by part of the cluster A–cluster B region of *ori*7 and by most of *ori*2 (de Zamaroczy et al., 1983). Similar petites can be easily obtained from different wild-type strains (Marotta et al., 1982), and are ideally suited to investigate the evolution of a well-defined intergenic region of the mt genome of *S. cerevisiae*.

Second, the repeat unit of a spontaneous petite, b, extensively studied in our laboratory, was derived from a secondary excision involving adjacent repeat units of petite b7 (Marotta et al., 1982). One could hope, therefore, to identify the sequences used in the excision of the mt genome of petite b.

Third, an mt gene involved in spore germination

was localized in the *cob-oli*2 region and possibly in the *ori*2–*ori*7 region of the wild-type strain, AP3, and a germination-deficient mutant, V17, derived from AP3, was available (Hartig et al., 1981; M. Breitenbach, personal communication). Hopefully, one could identify this gene, localize it, and learn about its primary structure.

Here we report the sequence of the *ori*2–*ori*7 region as determined in petite b7, and show the presence in it of a sequence coding for a 46-aa peptide. Transcription results suggest that this sequence corresponds to a new mt gene.

MATERIALS AND METHODS

(a) Yeast strains and media

Petites b, b7, b17 used in this work were spontaneous mutants derived from wild-type strain B (Marotta et al., 1982). Culture media were as in Marotta et al. (1982).

(b) Mitochondrial DNA

The mtDNA was prepared from early stationary-phase cells according to a procedure described elsewhere (Marotta et al., 1982). Restriction enzyme digestions, gel electrophoresis, transfer onto nitrocellulose and nick-translation of DNA probes were performed essentially as described (Faugeron-Fonty et al., 1979).

(c) Mitochondrial RNA

Protoplasts were prepared according to Faye et al. (1974) and lysed in the presence of 10 mM vanadium-adenosine, a ribonuclease inhibitor (Burger and Birkenmeier, 1979). The mt nucleic acids were extracted by the method of Locker (1979). DNase I and RNase I treatments were performed as described by Baldacci and Bernardi (1982). RNA electrophoresis was done on 1.5% agarose/6 M urea gels in Tris-phosphate buffer (Locker, 1979). Fractionated RNAs were transferred onto nitrocellulose following the method of Thomas (1980), as modified by Baldacci and Zennaro (1982). The mt RNAs were labelled at their 5′ ends by in vitro capping with guanylyltransferase (Levens et al., 1981).

(d) Hybridization

Hybridization conditions were as described by Thomas (1980). Final washing of filters was in 0.1 SSC, at 42°C.

(e) Nucleotide sequencing

(1) DNA

DNA sequencing after 5′- or 3′-end labelling and after secondary cleavage or strand separation was performed according to Maxam and Gilbert (1980). For details on the labelling strategy, see legend to Fig. 2.

(2) RNA

In vitro-capped mtRNAs were separated on 8% and/or 20% sequencing PA gels. After elution in 2 mM Tris pH 7.4, 1 mM EDTA, 0.01% SDS, 50 μg/ml tRNA, 0.5 M NH_4 · acetate they were sequenced with the RNases T1, U2, Phy M, and *Bacillus cereus* RNase (P.L. Biochemicals, Milwaukee, WI) following the protocols of D'Alessio (1982).

RESULTS

(a) The *ori*2–*ori*7 region

Fig. 2 presents the restriction map of the repeat unit of petite b7 and its flanking sequence, as determined for the longer repeat unit of petite b17 excised from the same region of the mt genome of wild-type strain B. The nt sequence comprised between *ori*2 and *ori*7 was determined on the repeat unit of petite b7, and is shown in Fig. 3.

(1) The excision sequences of petite b7

The length of the repeat unit of petite b7 is identical to the 2187-bp distance between equivalent points of *ori*2 and *ori*7. Because of differences present in *ori*2 and *ori*7 (de Zamaroczy et al., 1984), the excision sequences of b7 can be identified as the 65-bp perfect direct repeats forming the left ends (on the map of Fig. 2) of *ori*2 and *ori*7 (de Zamaroczy et al., 1983; Fig. 3).

(2) The excision sequences of petite b

Restriction mapping of this repeat unit (Fig. 2) had provided an approximate location of the se-

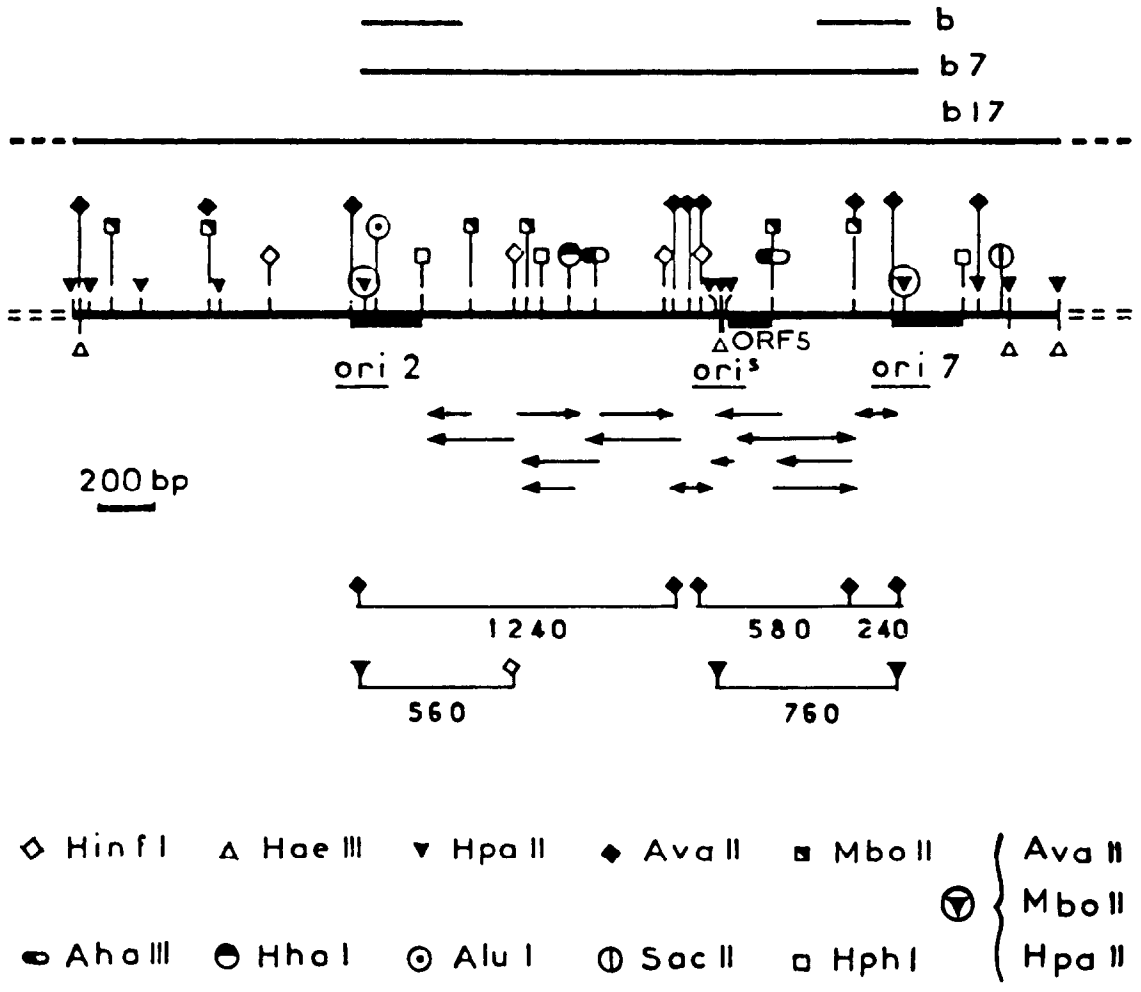


Fig. 2. Restriction map of the repeat unit of petite b7 and of its flanking sequences on the mt genome of parental strain B. These data expand those of Marotta et al. (1982). Horizontal arrows indicate the strategy used in sequencing experiments. Restriction fragments were 5′- and/or 3′-labelled with ^{32}P, and the end-labelled single-stranded fragments were separated by secondary cleavage (single-headed arrows) or by strand separation (double-headed arrows). The positions of *ori*2, *ori*7, and ORF5 are indicated as well as the sizes of some restriction fragments discussed in the text. The *ori*^{h}, the *ori* sequence of petite b7 (and of petite b), is formed from a part of cluster A–cluster B of *ori*7 (this region corresponds on the map to the *Ava*II-*Mbo*II-*Hpa*II site cluster; see also Fig. 3) and from most of *ori*2.

quences used in its excision (Marotta et al., 1982). At these positions two direct repeats, 15-bp long, with a single mismatch, were found (Fig. 3). Excision at these sequences, as present on two subsequent repeat units of the b7 genome, generates an 852-bp repeat unit, in perfect agreement with a previous estimate of an 853-bp unit derived from electrophoretic mobility (Faugeron-Fonty et al., 1979).

(3) An open reading frame

An ORF potentially coding for a sequence of 46 aa (Fig. 4) is present between positions 817 and 957 of the *ori*2–*ori*7 region. This sequence, which will be called ORF5 (see legend to Fig. 1), is the longest ORF found in the region; it starts with an ATA codon for methionine and shows a codon usage and an aa composition which are presented in Table I. ORF5 contains a G + C-rich complementary palindrome similar in secondary structure to a sequence found in the *var1* gene (Hudspeth et al., 1982; Fig. 5).

(b) Transcripts of the *ori*2–*ori*7 region in wild-type strain B and petite b7

To characterize the *ori*2–*ori*7 transcripts, the mtRNAs from petite b7 and its parental wild-type

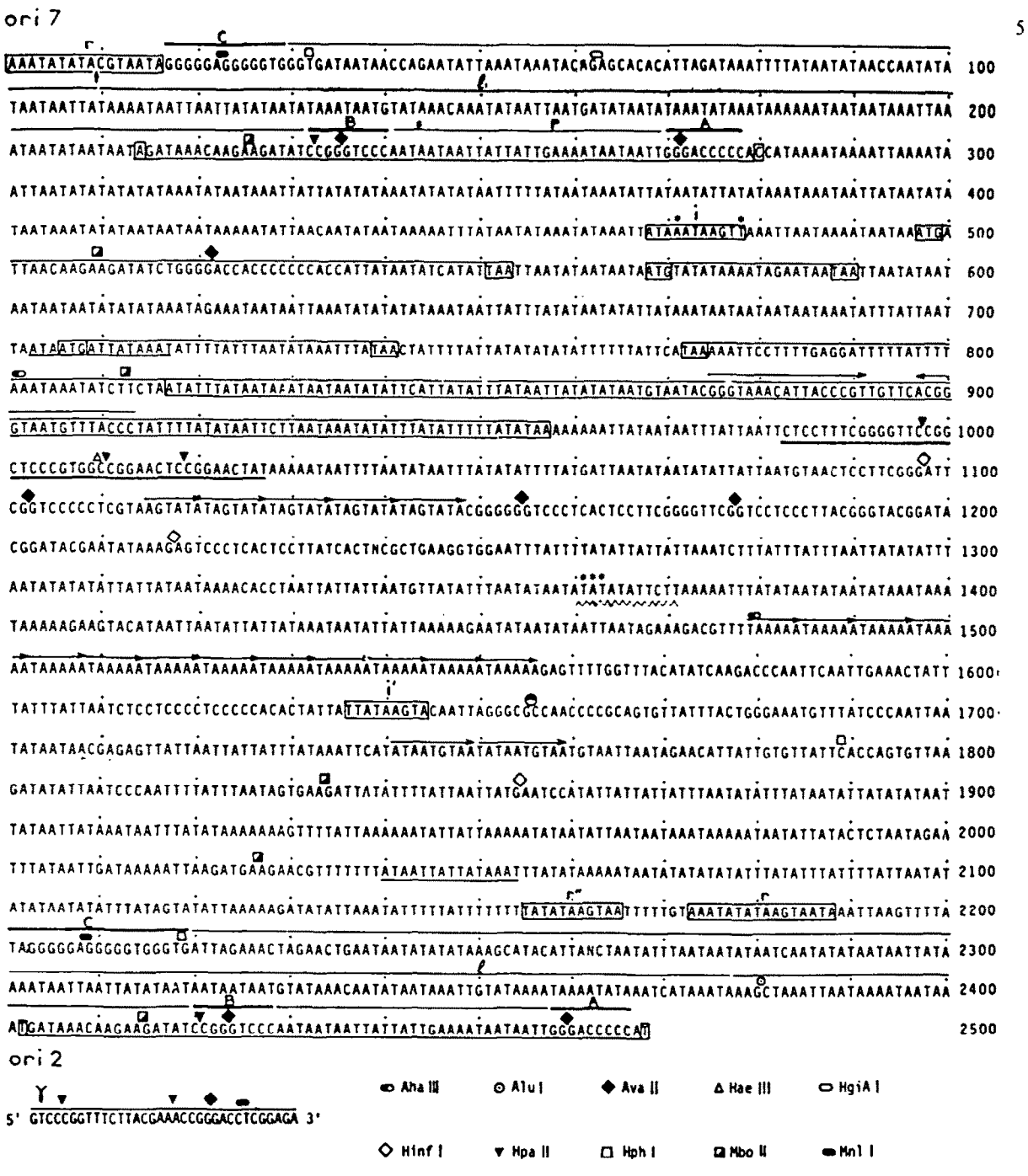


Fig. 3. Nucleotide sequence of the "nontemplate" strand of the repeat unit of the mt genome from petite b7. Both *ori*7 and *ori*2 are shown with their different segments (de Zamaroczy et al., 1984). The sequence of G + C-rich cluster *γ* from *ori*7 is also shown; its position is indicated by a vertical arrow at positions 9/10. The nucleotides delimiting the 65-bp direct repeats (underlined) used as the excision sequences (de Zamaroczy et al., 1983) of the repeat unit of petite b7 are boxed (positions 214, 280 and 2402, 2468). The two direct repeats (with one mismatch) used in the excision of the repeat unit of petite b are underlined (positions 703–717 and 2040–2054); the repeat sequences used in the excision belong, in fact, to two subsequent repeat units of b7. Arrows indicate conspicuous repeated sequences (positions 1115–1148; 1479–1556; 1740–1760). ORF5 is boxed (positions 817–957) and its G + C-rich complementary palindrome is indicated by two inverted arrows. The region upstream of ORF5 comprises the start point of transcription, *i* (positions 468–472, box), and three ORFs (overlined) starting with ATG codons and ending with TAA codons (boxes). The following overlined segment (positions 775–816) corresponds to an ORF which could extend ORF5 on the 5′ side (see text). The region downstream of ORF5 comprises (i) an *ori** sequence (heavy underlining; positions 983–1027), (ii) a sequence sharing a large homology with a messenger termination signal (Thalenfeld et al., 1983; Osinga et al., 1984b, wavy underlining), and (iii) another start point, *i*′ (positions 1636–1644, box), identical to those involved in the initiation of transcription of some tRNAs (Christianson and Rabinowitz, 1983; Osinga et al., 1984a).

```
--- met phe met met asn asn asn met phe ile met phe met ile met tyr asn val met arg val asn ile
CTA ATA TTT ATA ATA AAT AAT AAT ATA TTC ATT ATA TTT ATA ATT ATA TAT AAT GTA ATA CGG GTA AAC ATT

thr arg cys ser arg val met phe thr thr phe tyr met ile thr asn lys tyr ile tyr ile phe met *
ACC CGT TGT TCA CGG GTA ATG TTT ACC CTA TTT TAT ATA ATT CTT AAT AAA TAT ATT TAT ATT TTT ATA TAA
```

Fig. 4 Nucleotide sequence of ORF5 and the corresponding aa sequence.

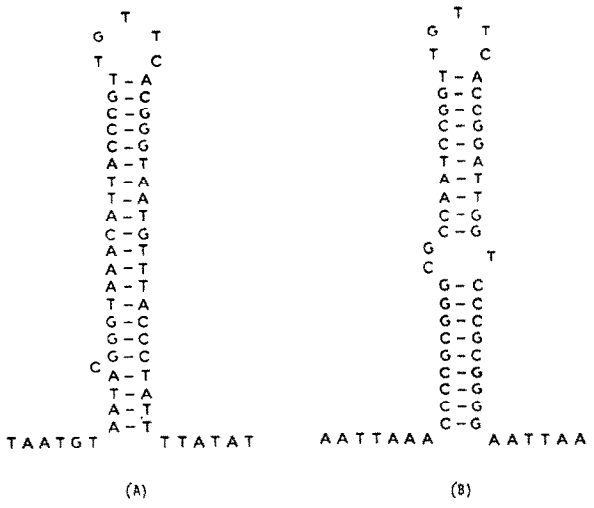


Fig. 5. G + C-rich complementary palindromes present (A) in ORF5 and (B) in *var1* (Hudspeth et al., 1982).

strain B were separated by gel electrophoresis (Fig. 6A), transferred onto nitrocellulose, and hybridized to mtDNA probes from petite b7 (Fig. 6, B–E).

When the whole repeat unit of b7 was used as a probe (Fig. 6B), the autoradiogram showed a major hybridization for a 900-nt transcript in the case of strain B. A faint transcript of this size could be detected in petite b7 where, however, the main hybridization band corresponded to a 1400-nt band; furthermore, two transcripts, 1200 and 500 nt in size, were also detectable (Fig. 6B).

To identify the sequence acting as a template for the 900-nt transcript of the B strain, *Ava*II restriction fragments of b7 DNA were used as probes. The 580-bp *Ava*II fragment carrying ORF5 strongly hybridized with the 900-nt transcript (Fig. 6C), whereas the 1240-bp fragment gave a weaker hybridization (Fig. 6D), and the 240-bp fragment gave no hybridization at all (Fig. 6E). In contrast, in the case of petite b7, all probes gave a positive hybridization (Fig. 6, B–E), indicating that the whole repeat unit is transcribed (see DISCUSSION). Hybridization with the separated strands of *Ava*II fragments from b7 showed that, both in strain B and in petite b7, the template strand was the one used by most mt transcripts (not shown).

A further characterization of the transcripts of petite b7 was done by hybridizing the mt RNA from this petite, labelled at the 5′-polyphosphate ends, to

TABLE I

Amino acid composition and codon usage in ORF5

Amino acid	Codon	%	Amino acid	Codon	%
Ala	GCA	0	Met	AUA	21.7
	U	0		G	0
	C	0	Phe	UUU	10.9
	G	0		C	2.2
Arg	AGA	0	Pro	CCA	0
	G	0		U	0
	CGA	0		C	0
	U	2.2		G	0
	C	0	Ser	UCA	2.2
	G	4.3		U	0
Asn	AAU	10.9		C	0
	C	2.2		G	0
Asp	GAU	0		AGU	0
	C	0		C	0
Cys	UGU	2.2	Thr	ACA	0
	C	0		U	0
Gln	CAA	0		C	4.3
	G	0		G	0
Glu	GAA	0		CUA	2.2
	G	0		U	2.2
Gly	GGA	0		C	0
	U	0		G	0
	C	0	Trp	UGA	0
	G	0		C	0
His	CAU	0	Tyr	UAU	8.7
	C	0		C	0
Ile	AUU	12.5	Val	GUA	6.5
	C	0		U	0
Leu	UUA	0		C	0
	G	0		G	0
Lys	AAA	2.2			
	G	0	Ter	TAA	2.2
				TAG	0
				TGA	0

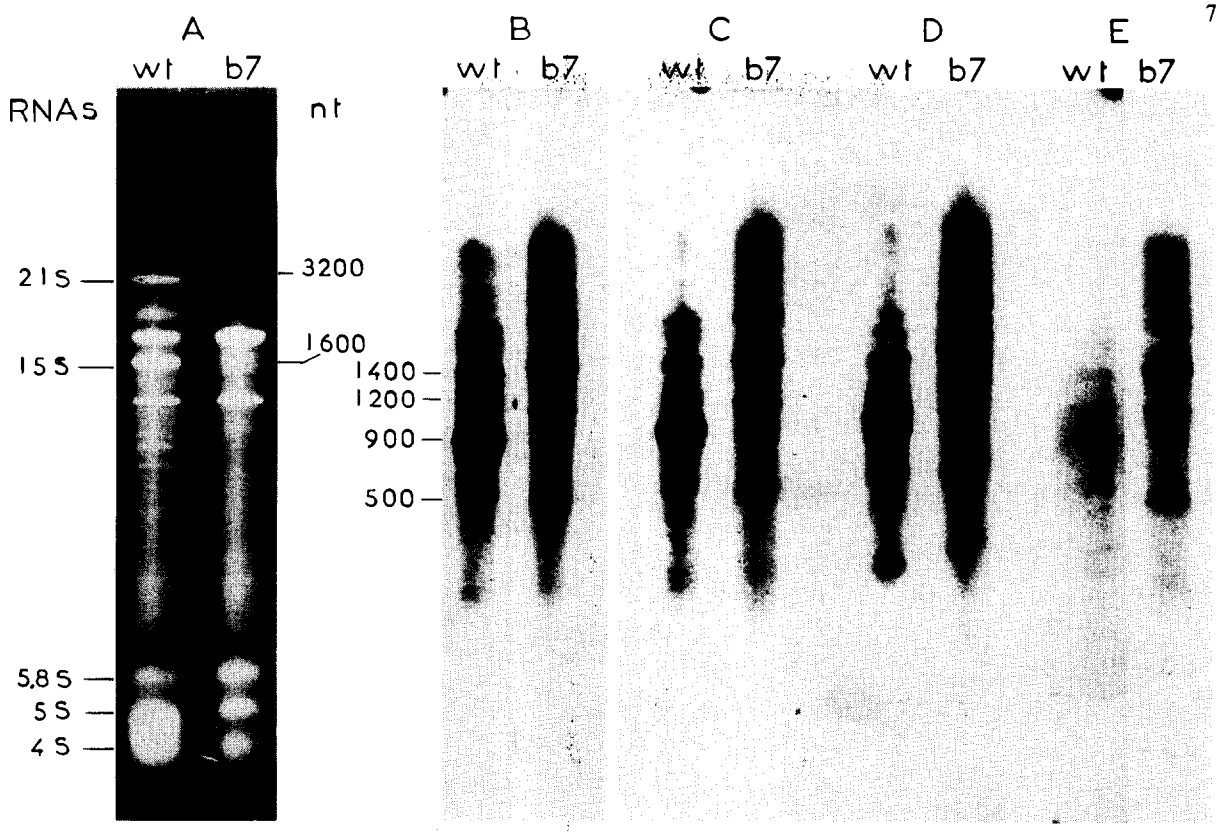
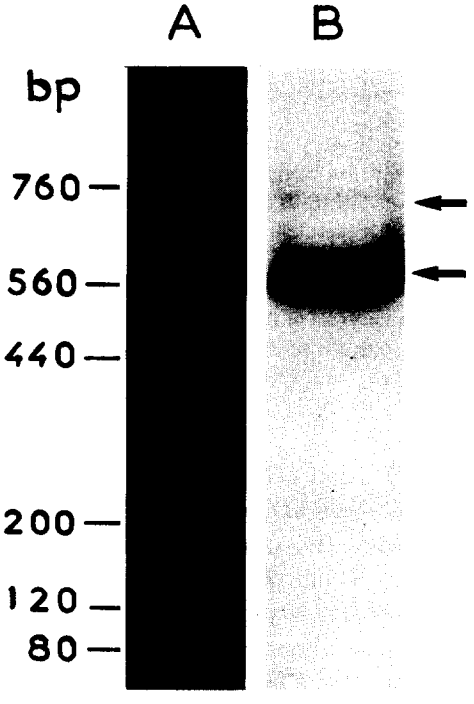


Fig. 6. Electrophoresis on agarose/urea gels of mtRNAs from strain B (wt) and petite b7, and their hybridization with labelled mtDNA probes from petite b7. Total mtRNAs from wild-type strain B and petite b7 were separated by electrophoresis in 1.5% agarose/6 M urea gels and stained with EtBr (A). RNAs were then transferred onto nitrocellulose and hybridized with the following ^{32}P-labelled probes from petite b7: total mtDNA (B); the 580-bp *Ava*II fragment carrying ORF5 (C); the 1240-bp fragment (D); and the 240-bp fragment (E); see also Fig. 2. 21S and 15S mt rRNAs (3200 and 1600 nt long, respectively) were used as size markers for the calibration of transcripts. Longer exposure times were used to reveal hybridization to RNAs from strain B compared to b7.



restriction fragments of b7 DNA. As shown in Fig. 7, this produced a very strong hybridization band on the 560-bp *Hin*fI-*Hpa*II fragment containing the transcription-initiation site *r* of *ori*^{h}, and a faint band on the 760-bp *Hpa*II fragment containing ORF5 (Fig. 3), which is contiguous to the former but located on the following repeat. These data indicate the existence of two RNA species: one of them starts at sequence *r* of *ori*^{h}, whereas the other starts in the 760-bp *Hpa*II fragment containing ORF5; neither of these RNAs, as obtained after capping, is long

Fig. 7. EtBr-stained restriction fragments from mtDNA of petite b7 and autoradiogram after hybridization with b7 mtRNA labelled by in vitro capping. (A) Electrophoretic pattern on 1.5% agarose, after EtBr staining, of mtDNA from petite b7 digested by *Hin*dI + *Hpa*II + *Hha*I. The sizes are indicated on the left margin (see also Figs. 2 and 3). (B) Autoradiogram of the same restriction fragments after hybridization with in vitro capped mtRNA from petite b7. Arrows specify the hybridizing fragments of 760 and 560 bp.

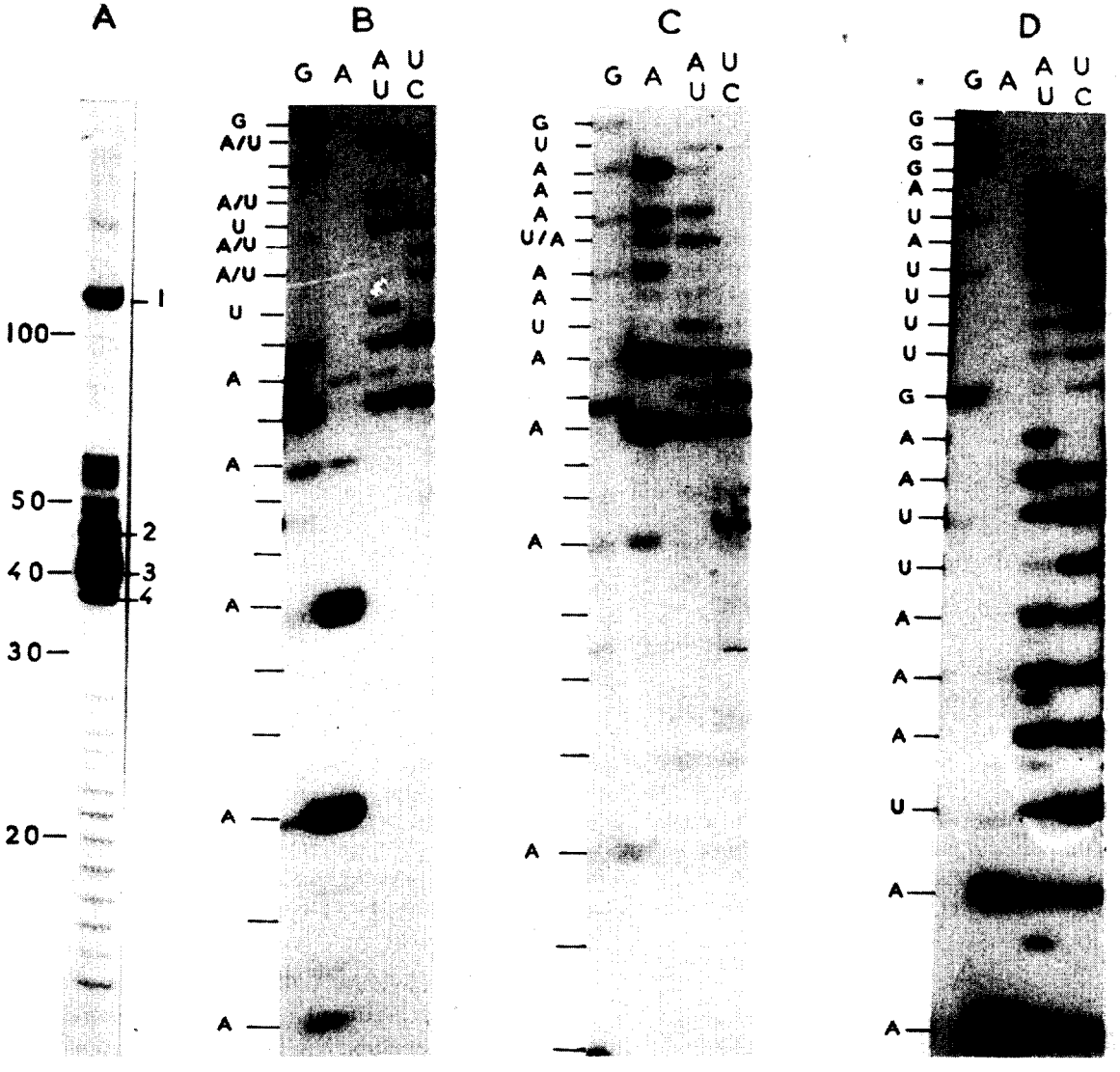


Fig. 8. Fractionation (A) and enzymatic sequencing (B–D) of mtRNA of petite b7. (A) Fractionation on a 20% PA/urea gel of in vitro capped mtRNAs from petite b7. The size (in nt) of radioactive fragments is indicated on the left margin; numbers on the right (1–4) specify fragments (or groups of fragments) which have been sequenced using specific RNases. (B, C) Sequencing patterns on 20% PA/urea gels of RNase-digested RNA fragment 1 from panel A. The ribonucleotides at which specific RNases cut are indicated at the top. Identified ribonucleotides are indicated to the left of the patterns; unassigned positions are blank. In (B) RNA was overdigested by (A-specific) U2 RNase, whereas it was underdigested by (A/U-specific) PhyM and (U/C-specific) *B. cereus* RNases; in (C), RNA was underdigested. Nonspecific breaks are evident at positions 10, 12 and 18. (D) Autoradiogram of RNase digestion products of in vitro capped RNAs 2–4 of panel A. Digestion by (A-specific) U2 RNase is subterminal. Identical results were obtained with fragments 2, 3 and 4 and with the fragments just above 2. All these fragments correspond to nonspecifically cut RNAs which share the same labelled 5′ end.

enough as to hybridize to any downstream DNA fragments.

The initiation sites for both transcripts were more precisely localized by sequencing b7 RNAs labelled at their 5′-polyphosphate ends by in vitro capping and separated by gel electrophoresis (Fig. 8A). These experiments confirmed that the majority species, RNAs 2, 3 and 4 (Fig. 8A), were initiated at sequence *r* of *ori*^{h} (Figs. 8D and 9B) and showed that the minority species, RNA 1 (Fig. 8A), had its starting point (Fig. 8, B and C; 9A) at nt 379 (Fig. 3), which immediately follows (on the DNA template) sequence *i*, a sequence homologous with known sites for the initiation of mt DNA transcription.

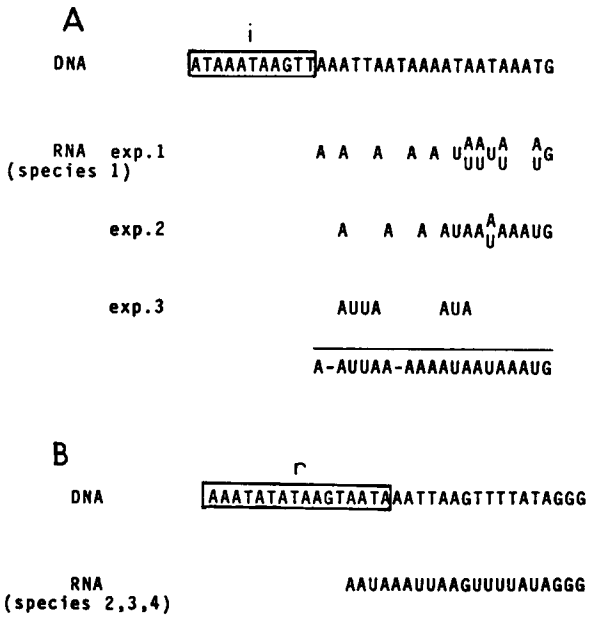


Fig. 9. The 5′ ends of transcripts initiated in the *ori*2–*ori*7 region. (A) The sequence of DNA between nt 468 and 499 of Fig. 3 is shown along with the results of three independent sequencing experiments indicating the 5′ ends of RNA species 1 (panel A of Fig. 8) from petite b7 mitochondria. Experiments 1 and 2 correspond to panels B and C, respectively, of Fig. 8. The sequence at the bottom has been constructed from well identified nucleotides, as determined in three experiments. *i* (box) indicates the sequence homologous with known transcription initiation sites. (B) The sequence of DNA between nt 2173 and 2188 of Fig. 3 is shown along with the sequence of the 5′ ends of RNA species 2, 3 and 4 (panel A of Fig. 8). *r* corresponds to the transcription initiation site of ori^h.

DISCUSSION

(a) ORF5: a new putative mitochondrial gene

The general features of the *ori*2–*ori*7 region of the mt genome of wild-type yeast cells are similar to those found in the intergenic sequences of this genome, in that the same AT spacer–G + C cluster organization (Prunell and Bernardi, 1977) is found (Fig. 3). As in the case of other intergenic regions, the *ori*2–*ori*7 region contains an ORF. When this ORF5 was discovered, the possibility existed that it might correspond to the mt gene involved in spore germination (Hartig et al., 1981). Recent results have not substantiated, however, the previously proposed localization of this gene (M. Breitenbach, personal communication), leaving this problem completely open.

In spite of the fact that its protein product is unknown, ORF5 might well correspond to a new mitochondrial gene. This conclusion is mainly supported by the presence in the mitochondria of wild-type strain B of a specific RNA detected by petite b7 DNA probes (Fig. 6). This RNA cannot be due to transcription initiated at the *r* sequence of *ori*2, because in the wild-type genome this signal is located downstream (in the direction of transcription) of the *ori*2–*ori*7 region (Figs. 1 and 2). This RNA must, therefore, be associated with a transcription initiation located either (i) in the *ori*2–*ori*7 region, or (ii) within *ori*7, or (iii) upstream of *ori*7. The latter possibility is contradicted by the absence of stable and abundant transcripts corresponding to *ori* sequences in wild-type mitochondria (Baldacci and Bernardi, 1982; Fig. 1). Initiation within *ori*7 is also ruled out, because the transcription initiation sequence *r* of *ori*7 is inactivated (Baldacci and Bernardi, 1982) by the insertion of the G + C-rich cluster γ. Therefore, the transcript must start in the *ori*2–*ori*7 region. Indeed, the results obtained by hybridizing restriction fragments of b7 DNA localize the template of the transcript in the middle of the *ori*2–*ori*7 region (Figs. 2, 3 and 6). The size of the transcript in wild-type mitochondria is about 900 nt. Its beginning and end correspond to a transcription start point and terminator, respectively, and the template DNA strand is that used by the vast majority of mt transcripts. All these results suggest the presence of a new protein-coding gene in the *ori*2–*ori*7 region of the mt genome of wild-type yeast.

The possibility that the 580-bp probe (which contains a G + C-rich *ori*5 sequence) is in fact revealing a wild-type transcript other than that discussed above because of nonspecific hybridization on G + C-rich clusters is contradicted by the fact that the 900-nt transcript from wild-type strain B has the same size as the well-characterized petite b7 transcript (see below). Moreover, the 900-nt RNA seen in Fig. 6C as a strong band upon hybridization with the 580-bp probe is weak and diffuse in Fig. 6D, where the 1240-bp probe was used, and absent in Fig. 6A, where the probe was the 240-bp fragment. These latter fragments overlap with the 900-nt RNA to the extent of 150 and 50 nt, respectively, accounting for the results just mentioned.

An unlikely possibility is that ORF5 is just an exon of a longer gene, since three short ORFs starting with

ATG codons are present between the RNA initiation site and ORF5 (see Fig. 3). Indeed, this explanation presents serious problems. Apart from the fact that the second ORF contains a TAG codon (namely a codon likely to act as a terminator codon), only 12 bp after the methionine codon, usage of the first and second ORFs would yield a messenger RNA which would be not 900 nt, but less than 800 nt long; such a difference would have been detected. In contrast, usage of the third ORF cannot be ruled out because the intronic segment eliminated would be too short to cause a significant difference; this difference could be even smaller if ORF5 were to extend from its 5′ end up to the TAA codon located at position 772 (Fig. 3).

(b) Transcription of the *ori*2–*ori*7 region in petite b7

The transcription pattern of the mtDNA from petite b7 is more complex than that of the corresponding region of the wild-type genome, because the *r* sequence of *ori*^{h} (which essentially corresponds to *ori*2) initiates very abundant transcripts which cover *ori*^{h} and extend over the contiguous repeat unit (leftwards in Fig. 2). To investigate the transcription initiation sites of b7, its transcripts were labelled at their 5′-polyphosphate ends and hybridized to restriction fragments of mtDNA from the same petite. The results show two hybridization bands, characterized by very different intensities. Since the two bands correspond to fragments that are contiguous, but located on subsequent repeat units of b7 (Fig. 2), the different intensity could be due to RNAs initiated at the *r* sequence of *ori*^{h} (contained in the strongly radioactive 560-bp band; see Figs. 2 and 7), but too short (because of degradation during capping labelling) to hybridize to the same extent with the 560-bp and the 760-bp fragment, which is downstream in the direction of transcription. Alternatively, the hybridization could be due to the presence of two RNA species initiated at different locations, one of them (that initiated at the *r* sequence of *ori*^{h}) being by far more abundant.

Sequencing of in vitro capped RNAs from b7 does indeed show the existence of two, and only two, distinct RNA species having different polyphosphate 5′-ends. The most abundant one corresponds to molecules initiated at the *r* sequence of *ori*^{h}, the others to molecules initiated just after sequence *i* (located at positions 468–478 of Fig. 3), which is largely homologous with other transcription initiation sequences of the mitochondrial genome (Osinga and Tabak, 1982; Baldacci and Bernardi, 1982; Christianson and Rabinowitz, 1983).

The results obtained with the transcripts starting at sequence *r* (see Fig. 9B) also provide an answer to another problem. In contrast to all other *ori* sequences, *ori*2 and *ori*3 contain, upstream (in the direction of transcription) of sequence *r*, an additional sequence *r″*, which is formed by a nonanucleotide identical to the transcription initiation site of *r* (Baldacci and Bernardi, 1982; de Zamaroczy et al., 1984). One may, therefore, wonder whether transcription starts at *r*, at *r″*, or both at *r* and *r″*. The sequencing results clearly indicate that transcription only starts at sequence *r*.

The nucleotide sequence downstream of the transcription start point *i* contains a site (nt 1360–1371) homologous with putative termination signals (Thalenfeld et al., 1983b; Osinga et al., 1984b). As already mentioned, its position, about 900 nt downstream of the initiation site, corresponds to the size of the transcript detected in wild-type mitochondria. This termination signal is, in fact, about 1400 nt downstream of the *r* sequence of *ori*^{h}, and accounts, therefore, for the size of the more abundant RNA species from b7. The results just described are of interest insofar as the general problem of mitochondrial transcription is concerned. They show that (i) sequences which differ from the most commonly used transcription initiation sequences may be used to initiate transcription, whereas sequences which, in primary structure, perfectly match well-established strong initiation sites may not be used, pointing to the importance of neighbouring sequences and of local secondary and tertiary structures; (ii) a transcript that is nonphysiological for wild-type cells extends from one repeat unit into the next in the petite genome and is stopped by a physiological termination signal. Interestingly, both situations had already been met at other genome locations (Baldacci and Bernardi, 1982).

(c) ORF5 and the other intergenic ORFs

ORF5 is the last discovered of the intergenic ORFs from the mitochondrial genome of yeast. ORF5 is different from the other ORFs in that it

produces a specific transcript. This is not the case with the other ORFs. Indeed, three of these are represented in the tails of transcripts originating from other genes. ORF1 is represented in the transcript of *oxi*1 and in another 1500-nt transcript copied from the opposite strand (Coruzzi et al., 1981). ORF4 is represented in the *oxi*3-*aap*1-*oli*2 transcript (Macino and Tzagoloff, 1980; Cobon et al., 1982; Osinga et al., 1984b) and ORF2 in the *oxi*2 transcript (Thalenfeld and Tzagoloff, 1980). It should be noted that (i) ORF4 is absent in some strains (Nagley et al., 1981; Cobon et al., 1982); (ii) 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 (iii) the corresponding transcript has been found shorter by 2000 nt on the 3′ 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).

(d) The codon usage and the evolutionary origin of ORF5

ORF5 exhibits a codon usage (Table I) similar to that shown by other mitochondrial ORFs, as present in both introns and intergenic regions, and by *var1*; this is different from the codon usage of mitochondrial genes and exons (Table I). The two most striking differences concern the use of AUA vs. AUG for methionine, and of UUU vs. UUC for phenylalanine, but many others can be seen in Table I, like the use of C- or G-ending codons for alanine, glycine, proline, leucine, threonine, and valine, and the use of CGN codons for arginine. The aa composition of ORF5 shows the same strong bias in favour of asparagine, isoleucine, and tyrosine already seen in *var1* and in other ORFs. Two features unique to ORF5 are the extremely high content of methionine (24%), phenylalanine (13%) and tyrosine (8.7%), and the absence of several amino acids (alanine, aspartic acid, glutamic acid, glycine, histidine, leucine, proline and tryptophan). This latter point is probably just related to the small size of ORF5, as suggested by the similar situation found with the very short *aap1* gene (Novitsky et al., 1983). The base composition of ORF5 is 15.2% G + C, but 90% of the G + C base pairs are concentrated in a complementary G + C-rich palindrome, similar in potential secondary structure to that found in *var1* (Fig. 5).

ORF5, *var1* and the other mitochondrial ORFs present in a number of intergenic regions are different from the respiratory-enzyme-encoding genes not only in codon usage, but also in other respects: (i) lower G + C content; (ii) lack of evolutionary conservation; and (iii) dispensability, at least for some of them. Two different explanations have been proposed so far for the origin of these ORFs, and more specifically of *var1*: that (i) 'they represent the past colonization of mitochondria by a different group of genes than those that participate in oxidative phosphorylation; this argument suggests that fungi were colonized at least twice, and that the marked differences in codon usage reflect these independent events' (Hudspeth et al., 1982); (ii) such genes, at least in the case of *var1*, 'were constructed over time from short, potentially functional domains embedded in otherwise nongenic A + T-rich spacer DNA; the recruitment of additional domains to graft onto the *var1* proto-gene occurred by recombination between homologous sequences of the spacer DNA' (Zassenhaus and Butow, 1983).

Both explanations rely on the assumption that codon usage is a fixed property which does not change in evolution, a point disproved by recent work showing that structural constraints may change it (Bernardi et al., 1985). A more reasonable explanation for the features exhibited by the A + T-rich mt 'genes' was proposed two years ago (Bernardi, 1983) and will be discussed in detail elsewhere (G. Bernardi, paper in preparation), namely that these 'genes' are very recent acquisitions derived by point mutations, insertions and deletions in A + T-spacer–G + C-cluster sequences in, namely of noncoding, internally repetitive sequences. In turn, these sequences have been proposed to have arisen by a genome expansion process involving replicase slippage at the *ori* sequences, as well as unequal crossing-over and insertions (Bernardi, 1982; 1983). This 'light compartment' of the mt genome of yeast is absent in other mt genomes, ranging from *Schizosaccharomyces pombe* to man (Bernardi, 1983). A strong argument in favour of this idea is that the only known protein originated from such genes, the Var1 protein, is the paradigm of a protein fulfilling some secondary role in ribosome structure; *var1* tolerates additions and deletions

without losing its function, and is not even conserved in another ascomycete, *Neurospora crassa.*

ACKNOWLEDGEMENTS

We thank our colleague M. de Zamaroczy for many useful discussions. One of us (G. Baldacci) thanks EMBO for the award of a long-term fellowship.

REFERENCES

Baldacci, G. and Zennaro, E.: Mitochondrial transcripts in glucose-repressed cells of *Saccharomyces cerevisiae.* Eur. J. Biochem. 127 (1982) 411–416.

Baldacci, G., de Zamaroczy, M. and Bernardi, G.: Excision sites in the GC clusters of the mitochondrial genome of yeast. FEBS Lett. 114 (1980) 234–236.

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.

Baldacci, G., Cherif-Zahar, B. and Bernardi, G.: The initiation of DNA replication in the mitochondrial genome of yeast. EMBO J. 3 (1984) 2115–2120.

Bernardi, G.: Evolutionary origin and the biological function of non-coding sequences in the mitochondrial genome of yeast, in Slonimski, P.P., Borst, P. and Attardi, G. (Eds.), Mitochondrial Genes. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, pp. 269–278.

Bernardi, G.: Genome instability and the selfish DNA issue. Folia Biol. 29 (1983) 82–92.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F.: The mosaic genome of warm-blooded vertebrates. Science 228 (1985) 953–958.

Burger, S.L. and Birkenmeier, C.S.: Inhibition of intractable nucleases with ribonuclease-vanadyl complexes: isolation of messenger ribonucleic acid from resting lymphocytes. Biochemistry 18 (1979) 5143–5149.

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 (1983) 14025–14033.

Cobon, G.S., Beilharz, M.W., Linnane, A.W. and Nagley, P.: Biogenesis of mitochondria: mapping of transcripts from the *oli*2 region of mitochondrial DNA in two grande strains of *Saccharomyces cerevisiae.* Curr. Genet. 5 (1982) 97–107.

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 *oxi*1 region of yeast mitochondrial DNA. J. Biol. Chem. 256 (1981) 12780–12787.

D'Alessio, J.M.: Gel electrophoresis of nucleic acids, in Richard, D. and Hames, B.D. (Eds.), Gel Electrophoresis of Nucleic Acids, IRL Press, Oxford, 1982, pp. 173–197.

de Zamaroczy, M., Baldacci, G. and Bernardi, G.: Putative origins of replication in the mitochondrial genome of yeast. FEBS Lett. 108 (1979) 429–432.

de Zamaroczy, M., Faugeron-Fonty, G. and Bernardi, G.: Excision sequences in the mitochondrial genome of yeast. Gene 21 (1983) 193–202.

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.

Faugeron-Fonty, G., Culard, F., Baldacci, G., Goursot, R., Prunell, A. and Bernardi, G.: The mitochondrial genome of wild-type yeast cells, VIII. The spontaneous cytoplasmic "petite" mutation. J. Mol. Biol. 134 (1979) 493–537.

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 twenty wild-type yeast strains. Gene 32 (1984) 459–473.

Faye, G., Kujawa, C. and Fukuhara, H.: Physical and genetic organization of petite and grande yeast mitochondrial DNA: in vivo transcription products of mitochondrial DNA and localization of 23S ribosomal RNA in petite mutants of *Saccharomyces cerevisiae.* J. Mol. Biol. 88 (1974) 185–203.

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

Goursot, R., de Zamaroczy, M., Baldacci, G. and Bernardi, G.: Supersuppressive "petite" mutants of yeast. Curr. Genet. 1 (1980) 173–176.

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.

Hartig, A., Schroeder, R., Mucke, E. and Breitenbach, M.: Isolation and characterization of yeast mitochondrial mutants defective in spore germination. Curr. Genet. 4 (1981) 29–36.

Hudspeth, M.E.S., Ainley, W.M., Shumard, D.S., Butow, R.A. and Grossmann, 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.

Levens, D., Ticho, B., Ackerman, E. and Rabinowitz, M.: Transcriptional initiation and 5′ termini of yeast mitochondrial RNA. J. Biol. Chem. 256 (1981) 5226–5232.

Locker, J.: Analytical and preparative electrophoresis of RNA in agarose-urea. Anal. Biochem. 98 (1979) 358–367.

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

Macreadie, I.G., Novitski, C.E., Maxwell, R.J., John, U., Ooi, B.-G., McMullen, G.L., Lukins, H.B., Linnane, A.W. and Nagley, P.: Biogenesis of mitochondria: the mitochondrial gene (*aap1*) coding for mitochondrial ATPase subunit 8 in *Saccharomyces cerevisiae.* Nucl. Acids Res. 11 (1983) 4435–4449.

Marotta, R., Colin, Y., Goursot, R. and Bernardi, G.: A region of extreme instability in the mitochondrial genome of yeast. EMBO J. 1 (1982) 529–534.

Martin, R.P., Bordonné, R. and Dirheimer, G.: The paramomycin 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.

Maxam, A.M. and Gilbert, W.: Sequencing end-labeled DNA with base specific chemical cleavages. Methods Enzymol. 65 (1980) 499–559.

Michel, F.: A maturase-like coding sequence downstream of the oxi-2 gene of yeast mitochondrial DNA is interrupted by two GC clusters and a putative end-of-messenger signal. Curr. Genet. 8 (1984) 307–317.

Nagley, P., Cobon, G.S., Linnane, A.W. and Beilharz, M.W.: Transcription of the *oli*2 region of yeast mitochondrial DNA shows strain-dependent variation. Biochem. Int. 3 (1981) 473–481.

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. and Tabak, H.F.: Processing of yeast mitochondrial messenger RNAs at a conserved dodecamer sequence. EMBO J. 3 (1984) 829–834.

Prunell, A. and Bernardi, G.: The mitochondrial genome of wild-type yeast cells, VI. Genome organization. J. Mol. Biol. 110 (1977) 53–74.

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: characterization of the *oxi2* transcript and localization of its promoter in *Saccharomyces cerevisiae* D273-10B. J. Biol. Chem. 258 (1983a) 610–615.

Thalenfeld, B.E., Bonitz, S.G., Nobrega, F.G., Macino, G. and Tzagoloff, A.: *oli1* transcripts in wild type and in a cytoplasmic "petite" mutant of yeast. J. Biol. Chem. 258 (1983b) 14065–14068.

Thomas, P.S.: Hybridization of denatured RNA and small DNA fragment transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA 77 (1980) 5201–5205.

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

Communicated by M.R. Culbertson.