# The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*

Françoise Foury*, Tiziana Roganti, Nicolas Lecrenier, Bénédicte Purnelle

*Unité de Biochimie Physiologique, Place Croix du Sud 2–20, 1348 Louvain-la-Neuve, Belgium*

Received 31 October 1998

**Abstract The currently available yeast mitochondrial DNA (mtDNA) sequence is incomplete, contains many errors and is derived from several polymorphic strains. Here, we report that the mtDNA sequence of the strain used for nuclear genome sequencing assembles into a circular map of 85 779 bp which includes 10 kb of new sequence. We give a list of seven small hypothetical open reading frames (ORFs). Hot spots of point mutations are found in exons near the insertion sites of optional mobile group I intron-related sequences. Our data suggest that shuffling of mobile elements plays an important role in the remodelling of the yeast mitochondrial genome.**



*Key words:* *Saccharomyces cerevisiae*; Mitochondrial DNA; Sequence; Polymorphism

## 1. Introduction

Since the discovery by Ephrussi [1] in 1949 of cytoplasmic heredity of the respiratory-deficient 'petite' mutants, *Saccharomyces cerevisiae* has been at the heart of mitochondrial genetics. The mitochondrial genes [2] and their mosaic intronic structure [3,4] were first identified in *S. cerevisiae* and the first mitochondrial gene sequenced was from this organism [5,6]. However, 20 years later, the sequence of the yeast mtDNA is still incomplete, contains many errors and is derived from 12 different strains.

The multi-copy mitochondrial genome from *S. cerevisiae* is characterized by low gene density and high A+T content. Its base composition is highly heterogeneous [7]; while the G+C content of the genes is approximately 30%, the intergenic spacers are composed of quasi-pure A+T stretches of several hundreds of base pairs, interrupted by more than 150 G+C-rich clusters, ranging from 10 to 80 bp in length [8]. These traits explain why scientists have sequenced the genes and neglected the intergenic regions.

The yeast mitochondrial genome contains the genes for cytochrome *c* oxidase subunits I, II and III (*cox1*, *cox2* and *cox3*), ATP synthase subunits 6, 8 and 9 (*atp6*, *atp8* and *atp9*), apocytochrome *b* (*cytb*), a ribosomal protein (*var1*) and several intron-related open reading frames (ORFs) [9,10]. The *cox1* and *cytb* genes contain several introns, some of which are translated, independently or in frame with their upstream exons, to produce maturases, reverse transcriptases or site-specific endonucleases [9,11]. In addition, the mitochondrial genome contains seven to eight replication origin-like (*ori*) elements and encodes 21S and 15S ribosomal RNAs, 24 tRNAs that can recognize all codons, and the 9S RNA component of RNase P [10]. All the genes are transcribed from the same strand, except $tRNA^{thr1}$. The published partial sequences, derived from a dozen different *S. cerevisiae* strains, have been assembled by de Zamaroczy and Bernardi [10] in 19 annotated contigs, constituting a remarkable basis for further studies (accession number L36885). However, this review does not correspond to any existing strain. Almost all of the protein-coding genes have been sequenced in the 'short mitochondrial genome' of D273-10B which contains two introns in the *cytb* gene and five introns in the *cox1* gene [9]. In the 'long' versions of the mitochondrial genome, only the sequences of the additional introns and immediate flanking exons have been published, so that the complete gene sequences are not available. Similarly, the *ori* elements, RNA genes and flanking regions have generally been sequenced in strains for which no protein-coding gene data are available.

## 2. Material and methods

Mitochondrial DNA was isolated from a purified clone of *S. cerevisiae* FY1679, an isogenic derivative of strain S288C. After growth on raffinose minimal medium cells were lysed and mitochondrial DNA was purified by centrifugation in a cesium chloride gradient in the presence of Hoechst dye 33258. A library was constructed from sheared mtDNA fragments inserted in the *Eco*RV site of Bluescript SK vector and random DNA sequencing of 0.5–0.7 kb fragments was performed by automatic sequencing using Bigdye terminators. Ten gaps, each of less than 200 bp, were filled by direct sequencing of polymerase chain reaction (PCR)-amplified products using mtDNA as template. The mitochondrial genome sequence is based on 1600 sequences performed on both strands with a 5.8-fold sequence coverage. Sequence assembly was achieved with DNASTAR software using 90% sequence identity. In a first step, 12 contigs were obtained and mapped with the help of the sequence published by de Zamaroczy and Bernardi [10]; this step greatly facilitated the design of the PCR primers necessary to fill the gaps. ORF1, ORF2, ORF3, ORF4 and ORF5 correspond to the nomenclature used in the review by de Zamaroczy and Bernardi [10].

EMBL accession number: AJ011856.

## 3. Results and discussion

### 3.1. Organization and sequence of FY1679 mitochondrial genome

The mtDNA sequence of strain FY1679, an isogenic derivative of S288C, is 85 779 bp in length and assembles into a circular contig (Fig. 1, right, Tables 1 and 2). Some 10 000 nucleotides are new sequences, essentially composed of long A+T stretches interrupted by many G+C clusters. In agreement with previous estimates, the average G+C content is 17.1%. The *cox1* gene and, to a lesser extent, the *cytb*, *21S*

*Corresponding author. Fax: (32) (10) 47 38 72.
E-mail: foury@fysa.ucl.ac.be

*Abbreviations:* ORF, open reading frame


PII: S0014-5793(98)01467-7

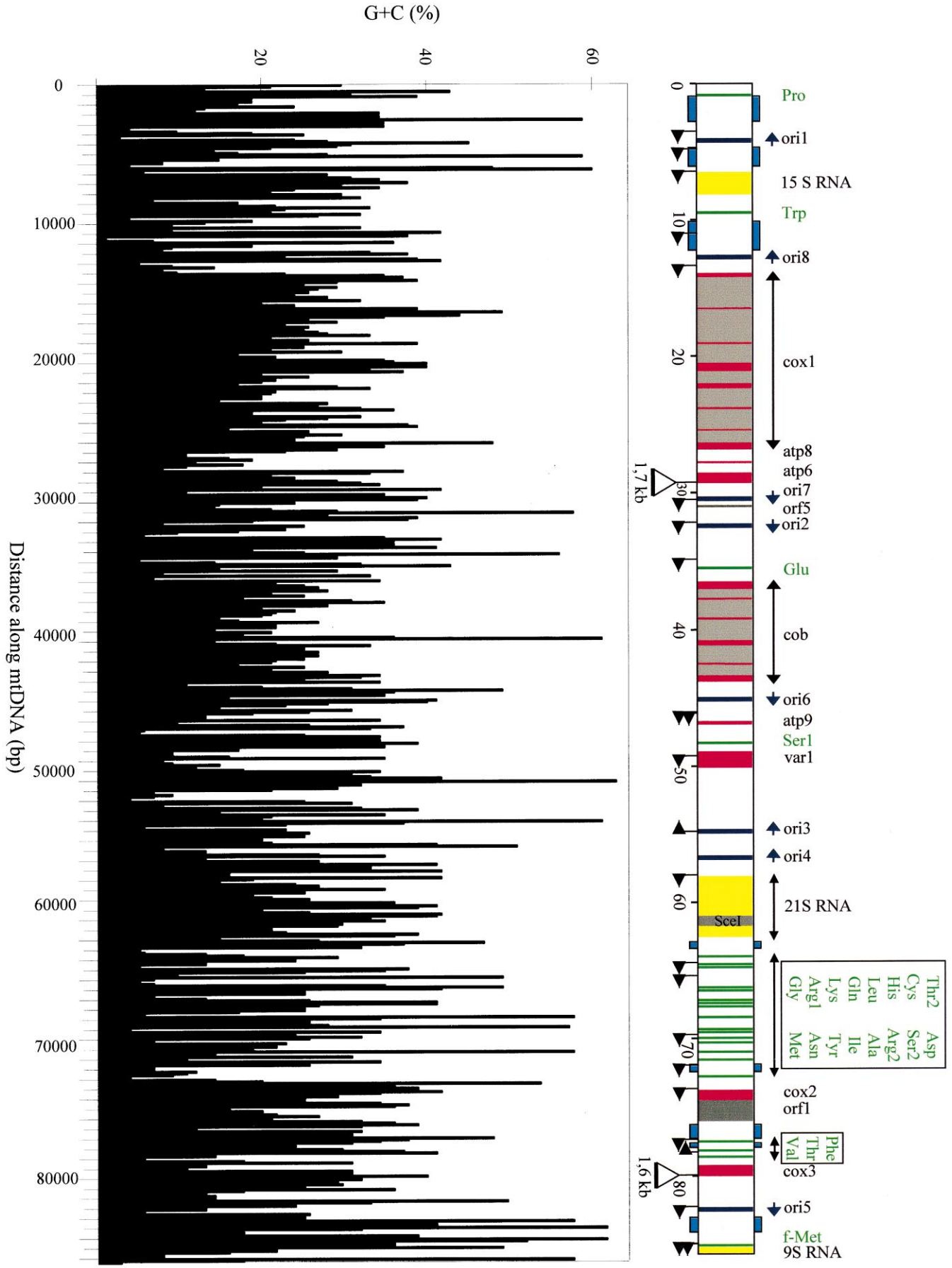
G+C (%)
20
40
60
0
10000
20000
30000
40000
50000
60000
70000
80000
Distance along mtDNA (bp)
0
10
20
30
40
50
60
70
80
1,7 kb
1,6 kb
Pro
ori1
15 S RNA
Trp
ori8
cox1
atp8
atp6
ori7
orf5
ori2
Glu
cob
ori6
atp9
Ser1
var1
ori3
ori4
21S RNA
SceI
Thr2 Asp
Cys Ser2
His Arg2
Leu Ala
Gln Ile
Lys Tyr
Arg1 Asn
Gly Met
cox2
orf1
Phe
Thr
Val
cox3
ori5
f-Met
9S RNA

Fig. 1. The mitochondrial genome of strain FY1679. Right: Linearized map. The genetic elements are indicated by boxes or bars. Red, exons of protein-coding genes; gray, introns and intron-related ORFs; green, tRNAs; yellow, 9S, 15S and 21S RNAs; dark blue, *ori* elements; light blue, main previous gaps filled during FY1679 mtDNA sequencing. The two large deletions (positions 29 295 and 80 030) are indicated by triangles. The flags indicate the transcription initiation sites and their orientation at positions 3504 (ORF6 and ORF7, potential site); 4814 (new potential site); 6461; 10 969 (ORF8, new potential site); 13 348; 30 493 (ORF5); 32 208; 34 878; 46 071; 46 149; 49 307; 54 839; 58 000; 64 392; 65 383; 69 705; 71 951 (new potential site); 73 697; 77 398; 78 300; 82 321; 85 000; 85 014. The potential transcription initiation consensus sites destroyed by G+C cluster insertion are at positions 4346 (*ori1*), 12 814 (*ori8*) and 56 866 (*ori4*). The mRNA processing sites downstream of *atp6* and *cox3* are at positions 29 353 and 80 123, respectively. The *ori7* sequence with accession number X59536 is similar to that of FY1679. Left: G+C profile of the sequence. Each bar represents the average G+C content in a 100-bp sliding window using steps of 50 bp.
←

*RNA* and *15S RNA* genes constitute the largest blocks of higher G+C density (Fig. 1, left). The *atp6*, *atp9*, *cox2*, *cox3* and *tRNA* genes appear as small G+C-enriched islands in the middle of A+T and G+C cluster-rich regions. The other high G+C density peaks correspond to the G+C clusters, their width depending on the number of these lying near one another (Fig. 1, left).

All the previously reported genetic elements that are essential for mitochondrial function are present in FY1679 mtDNA [10]. The sequence of the hypothetical ORF5 is also conserved [10]. Like strains KL14-4A and 777-3A, FY1679 contains 13 introns (five in *cytb*, seven in *cox1* and one in the *21S RNA* gene). Transcription of yeast mtDNA is polycistronic and is initiated at several sites characterized by the A/TTATAAGTA consensus sequence [12,13]. Nineteen initiation sites, previously shown to be functional, are present in FY1679; these include the promoter sites for *ori2*, *ori3* and *ori5*. Three other candidate sites are present in newly sequenced regions (Fig. 1, right). The previously described transcription processing sites were also found in FY1679 (data not shown).

However, comparison of the FY1679 mitochondrial genome with the sequences reported in the databases revealed important divergences. We detected numerous small 1–5 nucleotide additions (or deletions) in the *9S*, *15S* and *21S RNA* genes and intergenic regions. Although some of these can be accounted for as previous sequencing mistakes, in the main they represent strain polymorphisms that can easily be explained by template-primer misalignment during DNA replication in repetitive regions.

### 3.2. Two large deletions in FY1679

Compared with strain D273-10B, two fragments, each of more than 1.5 kb, immediately downstream of *cox3* and *atp6* stop codons (accession numbers X06706 and X02421), are missing in FY1679 (Fig. 1, right). As already reported for some strains [14–18], this results in the absence of ORF2 [14–16] and ORF4 [17,18], two proteins related to group I introns, and in the relocalization of their transcription termination sites immediately downstream of the *cox3* and *atp6* genes [10]. This polymorphism can result from DNA acquisition or loss. The sequences downstream of the *cox3* and *atp6* genes are completely different in the D273-10B and FY1679 strains. In contrast, in *Saccharomyces douglasii*, the region downstream of *cox3* (accession number X95975) bears homology to that of D273-10B and, moreover, contains an ORF2 orthologue. Since *S. cerevisiae* and *S. douglasii* diverged 50–80 millions years ago, it can be concluded that ORF2, and probably also ORF4, were initially present in both *Saccharomyces* species and have been lost in FY1679. The analysis of the new DNA joints produced in strain FY1679 at the deletion sites did not shed light on the mechanisms responsible for the deletion events. However, ORF4 is known to be a double-strand break DNA endonuclease, called Endo-*Sce*I, initiating homologous recombination at multiple cutting sites, and in particular in the *atp6* gene, near its 3′ end [17,18]. Thus, it is possible that the endonuclease activity of ORF2 and ORF4 has played a role in the excision process.

### 3.3. G+C clusters are a source of polymorphism

The major source of polymorphism is the G+C clusters. Previous studies have strongly suggested that G+C clusters are recombinogenic mobile elements [15,19–21]. The most frequent G+C cluster sequence is strictly delineated by the CT dinucleotide [15] (or AG, in the opposite orientation). In support of their recombinogenic and remodelling properties, we noticed that several G+C clusters have A+T stretches inside their CT (AG) boundaries (data not shown). We found that in FY1679, 15 G+C clusters are inserted at new sites and 20 G+C cluster sites, previously listed in other strains, are missing. In addition, 26 G+C cluster sites were identified in the newly sequenced regions. Even when inserted at identical positions, the G+C clusters from different strains show marked variability in their length and nucleotide sequence (data not shown). In the *21S RNA* gene, the optional G+C cluster located upstream of the ω intron is 18 bp longer in FY1679 than in previously sequenced strains [10]. Similarly, the G+C cluster present in the *9S RNA* gene [10] is 58 bp longer in FY1679.

In strain KL14-4A, the *cox1*-ai5β open reading frame [22] contains an in-frame G+C cluster. In FY1679, *cox1*-ai5β contains, 582 bp downstream of this G+C cluster, a second G+C cluster that introduces a frame-shift in the otherwise perfectly conserved open reading frame.

As previously noted [14,23], several uncertainties exist in the published ORF1 sequence (accession number J01484), a large open reading frame related to group I introns and located downstream of the *cox2* gene [23]. Our revised version shows that ORF1 start codon overlaps with the 3′ end of the *cox2* gene and that the reading frame extends 32 bp downstream of the previously reported stop codon. The ORF1 C-terminal domain matches that of its strongly conserved orthologue in *S. douglasii* (accession number X95973). In addition to the single incompletely sequenced G+C cluster present in ORF1 from D273-10B, there is, in FY1679, a second G+C cluster, 209 bp downstream of the first. Both of these break the frame, but, if the G+C cluster sequences delineated by the CT duplication are removed, a continuous frame is restored. It should be noted that, even though *S. douglasii* ORF1 is not interrupted by any G+C cluster, a frame-shift is also observed in the region corresponding to the first G+C cluster in *S. cerevisiae*. The organization of *S. cerevisiae* ORF1 is thus quite similar to that of ORF2 and ORF4 since in most *S. cerevisiae* strains, ORF2 and ORF4 open reading frames are also broken by intervening G+C clusters [14,16,18]. It has been shown

Table 1
Localization of the genes and replication origin elements of the yeast mitochondrial DNA

| Gene or ori | ORF | Localization (nt) | Function |
|---|---|---|---|
| tRNA pro | | 731–802 | |
| ORF6 | ORF6 | 3952–4338 | Hypothetical protein; unknown function |
| ori1 | | 4012–4312 (complement) | Replication origin-like |
| ORF7 | ORF7 | 4254–4415 | Hypothetical protein; unknown function |
| 15S RNA | | 6546–8194 | 15S ribosomal RNA |
| tRNA trp1 | | 9374–9447 | |
| ORF8 | ORF8 | 11667–11957 | Hypothetical protein; unknown function |
| ori8 | | 12510–12780 (complement) | Replication origin-like |
| cox1 | cox1 | Join: (13818–13986, 16435–16470, 18954–18991, 20508–20984, 21995–22246, 23612–23746, 25318–25342, 26229–26701) | Cytochrome *c* oxidase, subunit 1 |
| cox1-ai1 | Scai1 | 13987–16322 | Maturase/reverse transcriptase |
| cox1-ai2 | Scai2 | 16471–18830 | Maturase/reverse transcriptase |
| cox1-ai3 | I-SceIII | 18992–19996 | DNA endonuclease |
| cox1-ai4 | I-SceII | 20985–21935 | DNA endonuclease |
| cox1-ai5α | Scai5α | 22247–23167 | DNA endonuclease |
| cox1-ai5β | Scai5β | Join: (24156–24870, 24906–25255) | Hypothetical protein; unknown function |
| atp8 | atp8 | 27666–27812 | ATP synthase, subunit 8 |
| atp6 | atp6 | 28487–29266 | ATP synthase, subunit 6 |
| ori7 | | 30220–30594 | Replication origin-like |
| ORF5 | ORF5 | 30874–31014 | Hypothetical protein; unknown function |
| ori2 | | 32231–32501 | Active replication origin |
| tRNA glu | | 35373–35447 | |
| cytb | cytb | Join: (36540–36954, 37723–37736, 39141–39217, 40841–41090, 42508–42558, 43297–43647) | Apocytochrome *b* |
| cytbi2 | Scbi2 | Join: (36540–36954, 37723–38579) | Maturase |
| cytbi3 | Scbi3 | 39141–40265 | Maturase |
| cytbi4 | Scbi4 | 41094–42251 | Maturase |
| ori6 | | 44927–45227 | |
| atp9 | atp9 | 46723–46953 | ATP synthase, subunit 9 |
| tRNA ser | | 48201–48290 | |
| var1 | var1 | 48901–50097 | Ribosomal protein |
| ORF9 | ORF9 | 51052–51228 | Hypothetical protein; unknown function |
| ORF10 | ORF10 | 51277–51429 | Hypothetical protein; unknown function |
| ori3 | | 54567–54840 (complement) | Active replication origin |
| ori4 | | 56567–56832 (complement) | Replication origin-like |
| 21S RNA | | Join: (58009–60724, 61868–62447) | 21S ribosomal RNA |
| 21S RNA | SceI | 61022–61729 | DNA endonuclease |
| tRNA thr2 | | 63862–63937 | |
| tRNA cys | | 64415–64490 | |
| tRNA his | | 64596–64670 | |
| ORF11 | ORF11 | 65770–66174 | Hypothetical protein; unknown function |
| tRNA leu | | 66095–66179 | |
| tRNA gln | | 66210–66285 | |
| tRNA lys | | 67061–67134 | |
| tRNA arg1 | | 67309–67381 | |
| tRNA gly | | 67468–67542 | |
| tRNA asp | | 68322–68396 | |
| tRNA ser2 | | 69203–69288 | |
| tRNA arg2 | | 69289–69362 | |
| tRNA ala | | 69846–69921 | |
| tRNA ile | | 70162–70237 | |
| tRNA tyr | | 70824–70907 | |
| tRNA asn | | 71433–71503 | |
| tRNA met | | 72630–72705 | |
| cox2 | cox2 | 73758–74513 | Cytochrome *c* oxidase, subunit 2 |
| ORF1 | ORF1 | Join: (74495–75622, 75663–75872, 75904–75984) | Hypothetical protein; unknown function |
| tRNA phe | | 77431–77505 | |
| tRNA thr1 | | 78089–78162 (complement) | |
| tRNA val1 | | 78533–78608 | |
| cox3 | cox3 | 79213–80022 | Cytochrome *c* oxidase, subunit 3 |
| ori5 | | 82329–82600 | Active replication origin |
| tRNA f-met | | 85035–85112 | |
| ORF12 | ORF12 | 85554–85709 | Hypothetical ORF; unknown function |
| 9S RNA | | Join: (85290–85779, 1–11) | RNA component of RNase P |

The *cox1* and *cytb* genes of *S. cerevisiae* FY1679 are composed of eight and six exons, respectively. The complete *cox1* and *cytb* ORFs and the *21S RNA* gene are obtained by joining nucleotides at the positions indicated. Cox1-ai5β and ORF1 contain one and two G+C clusters, respectively, that break the frame of these ORFs, so that a continuous open reading frame is only obtained by joining nucleotides at the positions indicated. The term 'complement' indicates localization on the non-transcribed strand.


Fig. 2. Strain polymorphism in *cox1*, *cox3* and *atp6* genes. Mismatches are indicated by bold characters and stop codons are underlined. Exons 5 and 6 refer to FY1679. The insertion site of intron cox1-ai5α is indicated by an arrow.

that a continuous open reading frame, with no intervening G+C clusters, is required for the double-strand break endonuclease activity (Endo-*Sce*I) encoded by ORF4 [17,18]. These data raise the question whether ORF1 is a functional protein in *S. cerevisiae* and in *S. douglasii*. A continuous open reading frame of ~390 residues is conserved in both yeast species and contains the two dodecapeptide motifs that characterize the ORFs of group I introns [11]. This suggests that this large open reading frame is active; in contrast, G+C clusters may have invaded the 3′ end of the ORF1 gene, because it did not perform an essential function in the mitochondrion and there was, therefore, no selective pressure for its conservation.

The *ori* sequences are approximately 270 bp in length and are characterized by three conserved G+C blocks, A, B, and C, separated by A+T-rich stretches [24,25]. The additional G+C clusters present upstream, within, or downstream of some of these *ori* elements are conserved in FY1679 [26]. Several differences, however, are seen in FY1679. In *ori1*, an optional G+C cluster interrupts the A+T stretch located between blocks B and C and another one, downstream of block C, destroys the transcription initiation consensus. This G+C cluster is also present downstream of *ori4* and *ori8*, eliminating a potential initiation site. In *ori7*, the A+T-rich sequence located between blocks B and C contains a segment which is located downstream of the *ori7* element in other strains [10] and includes the ORF5 initiation transcription site. Some *ori* sequences may have been inactivated by intervening G+C clusters. Indeed, we have noted that in all strains, *ori2*, *ori3* and *ori5* possess a transcription initiation site and never contain intervening G+C clusters. These *ori* are active in mtDNA replication [24,25]. In contrast, *ori4* and *ori6*, which have no transcription initiation consensus and contain intervening G+C clusters, are probably not active [24,25]. Taken together, these data strongly suggest that G+C clusters play an important role in the remodelling of the yeast mitochondrial genome structure and activity.

### 3.4. Three mutational hot spots

The sequences of the protein-coding genes are, in general, highly conserved between *S. cerevisiae* strains. However, comparing strains FY1679 and D273-10B, we found a small number of mutational hot spots (Fig. 2). Two mutated clusters are located at the 3′ end of the *cox3* and *atp6* genes, respectively, upstream of the optional ORF2 and ORF4 genes. A third nucleotide divergence cluster is located in exons 5 and 6 of the *cox1* gene, at their boundary with the group I ai5α intron. This intron is absent in D273-10B and exons 5 and 6 are fused into a single exon. The same *cox1* sequence polymorphism has

Table 2
Localization of the introns in the *cox1*, *cytb* and *21S RNA* genes

| Gene | Intron group | Mobility | Position (nt) |
|---|---|---|---|
| cox1 | | | 13818–26701 |
| ai1 | Group II | Yes | 13987–16434 |
| ai2 | Group II | Yes | 16471–18953 |
| ai3 | Group I | Yes | 18992–20507 |
| ai4 | Group I | Yes | 20985–21994 |
| ai5α | Group I | Yes | 22247–23611 |
| ai5β | Group I | No | 23747–25317 |
| ai5γ | Group II | No | 25343–26228 |
| cytb | | | 36540–43647 |
| bi1 | Group II | No | 36955–37722 |
| bi2 | Group I | Yes | 37737–39140 |
| bi3 | Group I | No | 39218–40840 |
| bi4 | Group I | No | 41091–42507 |
| bi5 | GroupI | No | 42559–43296 |
| 21S RNA | | | 58009–62447 |
| ω | Group I | Yes | 60725–61867 |

The information concerning the classification of the introns and their mobility is taken from [11].

previously been reported for strains D273-10B and KL14-4A and has been suggested to be related to the loss of intron ai5α in D273-10B [22]. A few nucleotide changes are also observed in ω intron-containing and intron-less strains near the insertion site of the ω intron of the *21S RNA* gene [27]. Thus, mutational hot spots in yeast mtDNA protein-coding genes are specifically located in the vicinity of the insertion sites of optional mobile group I intron-related elements [15,16,22,28–30]. The gene nucleotide sequence in a given yeast strain is correlated with the presence or absence of this element in the vicinity of the gene (Fig. 2). All mutations are nucleotide substitutions, with the frequency of transitions being equal to, or higher than, that of transversions. These data are in sharp contrast with the excess of transversions that we recently reported for spontaneous mtDNA mutants of *S. cerevisiae* [31]. Similarly, a comparison between the genes of *S. cerevisiae* and those of several other yeast species revealed a large majority of A to T (T to A) transversions (data not shown). We have shown that in *S. cerevisiae*, while mismatch repair preferentially eliminates transitions and 3′-5′ exonucleolytic proofreading contributes to elimination of transversions, the latter are not efficiently repaired [31]. Our new data suggest that the mutational hot spots observed in *cox1*, *cox3* and *atp6* genes result from an unusual mutagenic process associated with the deletion of group I intron-related elements. This process may be linked to their double-strand DNA endonuclease activity; however, an alternative possibility would be that the deletion process involves the reverse transcriptase activity of the cox1-ai1 and -ai2 mobile group II introns acting in *trans*, as this has been shown to be the case for deletion of group I introns cox1-ai5α, -ai5β or -ai5γ [32]. This hypothesis would explain the nucleotide sequence differences described above since, while the yeast mtDNA polymerase is generally faithful, reverse transcriptases are characterized by low DNA replication fidelity.

### 3.5. New hypothetical small open reading frames

All the previously identified ORFs, except ORF5, start with the AUG codon [10]; however, it has been shown by site-directed mutagenesis of the *cox2* initiator codon [33] (Bonnefoy and Fox, personal communication) that GUG and, in a much less efficient manner, AUA can initiate translation in yeast mitochondria. We have eliminated all those hypothetical ORFs that could not be included in a transcriptional unit, and thus, we have selected seven small hypothetical ORFs, either starting with the AUG codon and having at least 50 codons, or starting with the extremely frequent AUA codon and having at least 100 codons. None bears homology to ORFs of known or unknown function and all are characterized by a strong bias in their amino acid composition and by the presence of rare codons. It must also be noted that ORF7 and ORF12 overlap with *ori1* and the *9S RNA* gene, respectively.

### 3.6. Conclusion

The complete determination of the mtDNA sequence of strain FY1679 confirms now that the yeast mtDNA map is circular and that the general organization of the mitochondrial genome, previously deduced from a dozen different strains, is conserved in this strain. Future experiments will determine whether the seven ORFs listed in this paper are expressed. This work also underlines the role played by mobile elements, in particular G+C clusters and group I intron-related ORFs, in the generation of polymorphism in yeast mtDNA. Moreover, the scientific community can now refer to a complete and reliable yeast mtDNA sequence determined in the single reference strain that has been used for the nuclear genomic sequence.

*Acknowledgements:* Dr. S. Salzberg (The Institute for Genomic Research, Rockville MD, USA) is gratefully acknowledged for the program 'GCCOUNT'. We thank Jaga Lazowska (Centre de Génétique Moléculaire, Gif-sur-Yvette, France) for discussion and help in intronic ORF sequence localization. We thank A. Goffeau for his constant support and critical reading of the manuscript. MIPS (Max-Planck Institute, Martinsried, Germany) is also acknowledged. This work was supported by the FINYS programme of the European Commission (BIO4-CT96-0558) and by the Fonds National de la Recherche Scientifique Belge.

## References

[1] Ephrussi, B. and Chimenes, A.M. (1949) Ann. Inst. Pasteur 76, 351–364.
[2] Tzagoloff, A., Akai, A., Needleman, R.B. and Zulch, G. (1975) J. Biol. Chem. 250, 8236–8242.
[3] Bos, J.L., Heyting, C., Borst, P., Arnberg, A.C. and Van Bruggen, E.F. (1978) Nature 275, 336–338.
[4] Slonimski, P.P., Claisse, M.L., Foucher, M., Jacq, C., Kochko, A., Lamouroux, A., Pajot, P., Perrodin, G., Spyridakis, A. and Wambier-Kruppel, M.L. (1978) in: Biochemistry and Genetics of Yeast: Pure and Applied Aspects (Bacila, M., Horecker, B.L. and Stoppani, A.O.M., Eds.), Academic Press, New York.
[5] Macino, G. and Tzagoloff, A. (1979) Proc. Natl. Acad. Sci. USA 76, 131–135.
[6] Hensgens, L.A.M., Grivell, L.A., Borst, P. and Bos, J.L. (1979) Proc. Natl. Acad. Sci. USA 76, 1663–1667.
[7] Bernardi, G., Piperno, G. and Fonty, G. (1972) J. Mol. Biol. 65, 173–189.
[8] De Zamaroczy, M. and Bernardi, G. (1986) Gene 41, 1–22.
[9] Tzagoloff, A. and Myers, A.M. (1986) Annu. Rev. Biochem. 55, 249–285.
[10] de Zamaroczy, M. and Bernardi, G. (1986) Gene 47, 155–177.
[11] Lambowitz, A.M. and Belfort, M. (1993) Annu. Rev. Biochem. 62, 587–622.
[12] Christianson, T. and Rabinowitz, M. (1983) J. Biol. Chem. 258, 14025–14033.
[13] Osinga, K.A. and Tabak, H.F. (1984) Nucleic Acids Res. 12, 1889–1900.
[14] Michel, F. (1984) Curr. Genet. 8, 307–317.
[15] Séraphin, B., Simon, M. and Faye, G.J. (1987) J. Biol. Chem. 262, 10146–10153.
[16] Séraphin, B., Simon, M. and Faye, G. (1985) Nucleic Acids Res. 13, 3005–3014.
[17] Nakagawa, K., Morishima, N. and Shibata, T. (1991) J. Biol. Chem. 266, 1977–1984.
[18] Nakagawa, K., Morishima, N. and Shibata, T. (1992) EMBO J. 11, 2707–2715.
[19] Butow, R., Perlman, P. and Grossman, L. (1986) Science 228, 1496–1501.
[20] Dieckmann, D.C. and Gandy, B. (1987) EMBO J. 6, 4197–4203.
[21] Weiller, G., Schueller, M.E. and Schweyen, R.J. (1989) Mol. Gen. Genet. 218, 272–283.
[22] Hensgens, L.A.M., Bonnen, L., De Haan, M., Van der Horst, G. and Grivell, L.A. (1983) Cell 32, 379–389.
[23] Coruzzi, G., Bonitz, S.G., Thalenfeld, B.E. and Tzagoloff, A. (1981) J. Biol. Chem. 256, 12780–12787.
[24] Blanc, H. and Dujon, B. (1980) Proc. Natl. Acad. Sci. USA 77, 3942–3946.
[25] de Zamaroczy, M., Marotta, R., Faugeron-Fonty, G., Goursot, R., Mangin, M., Baldacci, G. and Bernardi, G. (1981) Nature 292, 75–78.
[26] De Zamaroczy, M., Faugeron-Fonty, G., Baldacci, G., Goursot, R. and Bernardi, G. (1984) Gene 32, 439–457.

[27] Dujon, B. (1980) Cell 20, 185–197.
[28] Jacquier, A. and Dujon, B. (1985) Cell 41, 383–394.
[29] Moran, J.V., Wernette, C.M., Mecklenberg, K.L., Butow, R.A. and Perlman, P.S. (1992) Nucleic Acids Res. 20, 4069–4076.
[30] Séraphin, B., Faye, G., Hatat, D. and Jacq, C. (1992) Gene 113, 1–8.
[31] Vanderstraeten, S., Van den Brûle, S., Hu, J. and Foury, F. (1998) J. Biol. Chem. 273, 23690–23697.
[32] Levra-Juillet, E., Boulet, A., Séraphin, B., Simon, M. and Faye, G. (1989) Mol. Gen. Genet. 217, 168–171.
[33] Folley, L.S. and Fox, T.D. (1991) Genetics 129, 659–668.