

Sequence analysis of three mitochondrial DNA molecules reveals interesting differences among *Saccharomyces* yeasts

R. B. Langkjær, S. Casaregola¹, D. W. Ussery, C. Gaillardin¹ and J. Piškur*

BioCentrum-DTU, Technical University of Denmark, Building 301, DK-2800 Kgl. Lyngby, Denmark and
¹UMR Microbiologie et Génétique Moléculaire, CNRS INRA, INA-PG, F-78850 Thiverval-Grignon, France

Received February 28, 2003; Revised and Accepted April 11, 2003

DDBJ/EMBL/GenBank accession no. AJ430679

ABSTRACT

The complete sequences of mitochondrial DNA (mtDNA) from the two budding yeasts *Saccharomyces castellii* and *Saccharomyces servazzii*, consisting of 25 753 and 30 782 bp, respectively, were analysed and compared to *Saccharomyces cerevisiae* mtDNA. While some of the traits are very similar among *Saccharomyces* yeasts, others have highly diverged. The two mtDNAs are much more compact than that of *S.cerevisiae* and contain fewer introns and intergenic sequences, although they have almost the same coding potential. A few genes contain group I introns, but group II introns, otherwise found in *S.cerevisiae* mtDNA, are not present. Surprisingly, four genes (*ATP6*, *COX2*, *COX3* and *COB*) in the mtDNA of *S.servazzii* contain, in total, five +1 frame-shifts. mtDNAs of *S.castellii*, *S.servazzii* and *S.cerevisiae* contain all genes on the same strand, except for one tRNA gene. On the other hand, the gene order is very different. Several gene rearrangements have taken place upon separation of the *Saccharomyces* lineages, and even a part of the transcription units have not been preserved. It seems that the mechanism(s) involved in the generation of the rearrangements has had to ensure that all genes stayed encoded by the same DNA strand.

INTRODUCTION

The mitochondrial DNA (mtDNA) of bakers yeast, *Saccharomyces cerevisiae* is 85.8 kb long and encodes subunits I, II and III of cytochrome c oxidase (*COX1*, *COX2* and *COX3*), apocytochrome b (*COB*), subunits 6, 8 and 9 of ATPase (*ATP6*, *ATP8* and *ATP9*) and a ribosome-associated protein (*VARI*). It also contains a number of unidentified reading frames (URFs) and intron-related open reading frames (ORFs). In addition, the mitochondrial genome specifies the small and large ribosomal RNAs (SSU rRNA and LSU rRNA), 24 transfer RNAs (tRNAs) and the 9S RNA (Rpm1r) component of RNase P (1). All genes are encoded and

transcribed from one strand, except for *tRNA^{Thr1}_{CUN}* (1, reviewed in 2). The yeast mitochondrial genetic code is used in translation and differs from the universal code by AUA being read as methionine, UGA as tryptophan and CUN as threonine. The *S.cerevisiae* mtDNA is characterised by a very low GC content, 17–18%, and extensive intergenic regions, which comprise 62% of the genome (1,2). These regions are composed of long adenosine and thymidine (A+T) stretches, short guanosine and cytidine (G+C) clusters, and a special class of intergenic sequences, *ori/rep*, which are present in eight copies (2,3). The *ori/rep* repeats are involved in preferential transmission of mtDNA (4). A number of short repeats, direct or indirect, can be found within the *S.cerevisiae* mitochondrial sequence, especially within the intergenic regions (1,3). These short repeats are involved in the generation of petite mtDNAs. The initially proposed circular nature of the *S.cerevisiae* mtDNA molecule is still controversial (5).

Within the hemiascomycetous yeasts, the size of the mtDNA varies from ~18.9 kb in *Candida glabrata* (6) to ~101 kb in *Brettanomyces custersii* (7). *Pichia canadensis*, formerly *Hansenula wingei* (8), *Candida albicans* (accession no. NC002653) and *Yarrowia lipolytica* (9) have mtDNA sizes of 27.7, 40.2 and 47.9 kb, respectively. Even in the genus *Brettanomyces/Dekkera*, mitochondrial genome sizes have been shown to vary from ~28 to ~101 kb (7,10). The mitochondrial genomes of some ascomycetous yeasts have been characterised by the presence of a *VARI* gene as in the yeasts *S.cerevisiae*, *C.glabrata* and *P.canadensis* (8,11–14). On the other hand, *Y.lipolytica* and *Schizosaccharomyces pombe* both lack *VARI* suggesting that only a limited number of ascomycetous yeasts have mtDNA containing *VARI* (9,15,16). Apart from *Saccharomyces* and *Kluyveromyces* yeasts (17) as well as *S.pombe* (15), mitochondrial genomes of several Ascomycetes, like *Y.lipolytica* (9), contain several genes encoding hydrophobic subunits of NADH dehydrogenase complex I. Finally, the mtDNA of *Podospira anserina* lacks both *VARI* and *ATP9* (18).

The *Saccharomyces* genus contains several species that can be divided into a group of petite-positive and a group of petite-negative yeasts, of which the latter includes only *Saccharomyces kluyveri* (19). The petite-positive yeasts, which display both fermentative and respirative metabolisms, can be further divided into *Saccharomyces sensu stricto* yeasts

*To whom correspondence should be addressed. Tel: +45 45 25 25 18; Fax: +45 45 93 28 09; Email: jp@biocentrum.dtu.dk

(including *S.cerevisiae*) and *Saccharomyces sensu lato* yeasts (including *Saccharomyces castellii* and *Saccharomyces servazzii*). In addition, *C.glabrata* is very closely related to the *sensu lato* group (20). In the *sensu stricto* group the mtDNA sizes ranges from 64 to 86 kb, whereas the *sensu lato* group has mtDNA with sizes below 50 kb (19,21). The gene order also varies considerably within the *Saccharomyces* genus and is a result of a limited number of large rearrangements taking place during the yeast evolutionary history (21). Previously, these rearrangements were suggested to be created either by transposition- and/or inversion-like events (21,22). One should note that if a segment is inverted then the coding strand of the corresponding genes is changed. In the case of a transposition-like event, the coding strand can either be preserved or changed. Previously it has been demonstrated *in vivo* that the *S.cerevisiae* gene order could be rearranged by either homologous or illegitimate recombination (22). While the rearranged genomes have preserved the respiration capacity, they have exhibited a decreased transmission ability (4). Therefore, at least in *S.cerevisiae*, preservation of the gene order seems to be under strong selective pressure.

The complete sequence of the mitochondrial genome of *S.castellii* has recently been determined but not fully analysed (23). Most of the mtDNA sequence of *S.servazzii* has previously been obtained in the random sequencing tag project known as Génolevures (24). Along with the complete mtDNA sequence of *S.servazzii*, we present here a thorough analysis of the mitochondrial genomes of the two *Saccharomyces sensu lato* yeasts. The complete sequences of these two mitochondrial genomes, as well as the previously determined *S.cerevisiae* sequence, now provide sufficient data to reconstruct several events that have reshaped the *Saccharomyces* mtDNA molecules during yeast evolutionary history.

MATERIALS AND METHODS

Sequencing

The mitochondrial sequence of the *S.castellii* type strain (NRRL Y-12630 = CBS4309) was determined and deposited to GenBank (accession no. AF437291) (23).

Much of the sequence of the mitochondrial genome of *S.servazzii* type strain CBS4311 was generated by the Génolevures project (24). Additional sequence was obtained from clones of the *S.servazzii* type strain genomic DNA library used in Génolevures. Sequence assembly was performed using programs phred (version 0.980904.c) and phrap (version 0.960731) with a minscore of 14 and a minmatch of 30 (25,26). The sequence compilation was edited with consed, version 10.38 (beta) (27) and deposited to GenBank (accession no. AJ430679). Each base of the *S.servazzii* mtDNA was covered at least twice. As far as the +1 T frameshift-containing regions are concerned, except for the frameshift region of *COX3* obtained from the same strand on two independent clones, the frameshift region sequences were obtained in each case from both strands (phred score >40). The *COB* region has been sequenced from five independent clones, the *COX2* region from four independent clones and the *ATP6* from 10 independent clones.

Gene annotation

Searches of similarity to the already known and annotated mtDNA genomes of *S.cerevisiae* (1) and *P.canadensis* (8) were carried out by using variations of the Basic Local Alignment Search Tool (BLAST) (28), BLASTN (nucleotide similarity based search) and BLASTX (amino acid similarity based search), available on the world wide web. In some instances, sequences were analysed with various programs in the GCG environment (Genetics Computer Group, Madison, WI, USA), including FASTA (29). tRNAs were annotated using the program tRNAscan-SE 1.1 (30). Intron and endonuclease nomenclature follows that of Dujon (31) and Dujon *et al.* (32), respectively, and classification is according to Burke *et al.* (33) and Burke (34).

DNA atlases

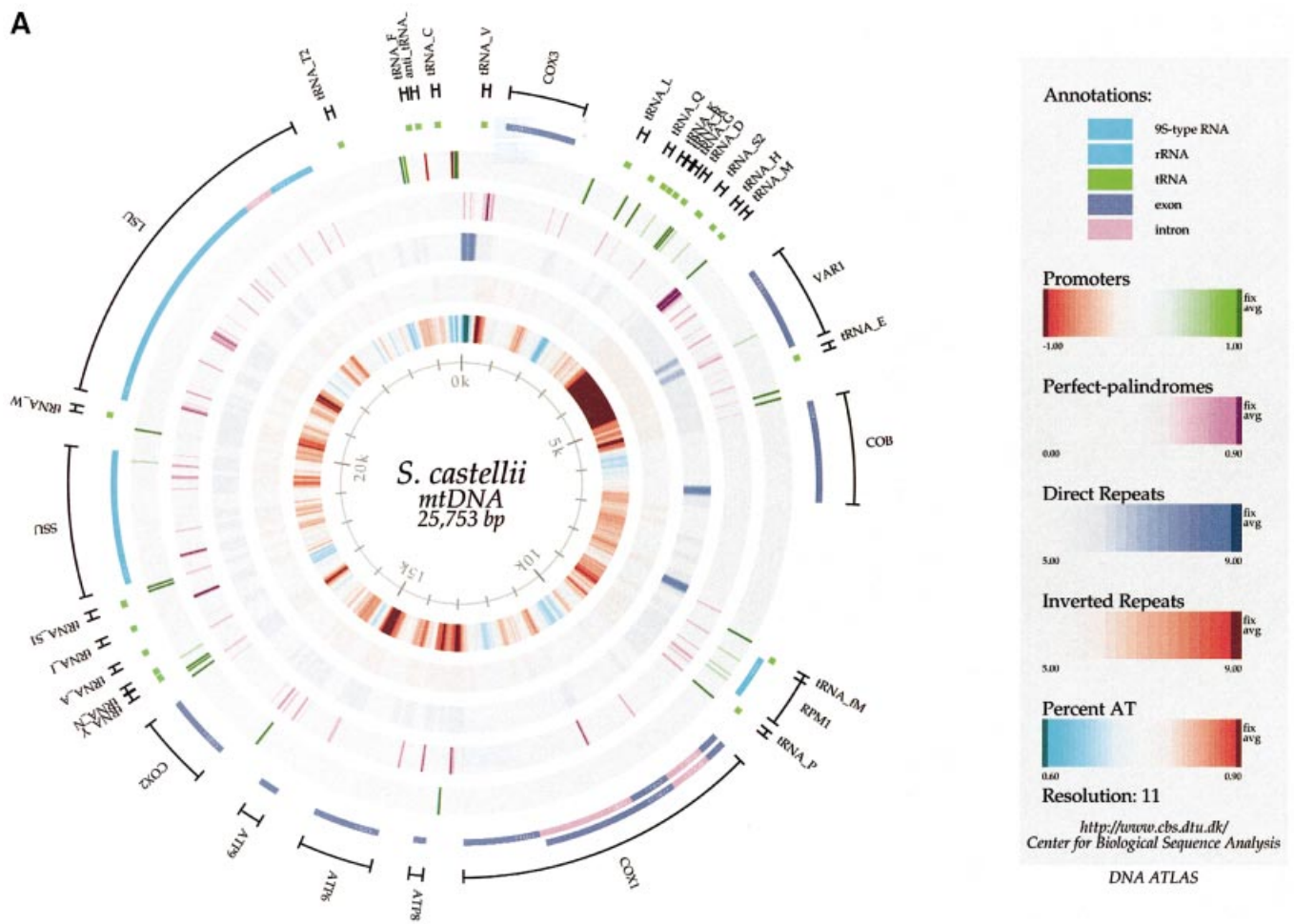
'DNA atlas' plots were used to visualise different features of the mtDNA molecules of *S.castellii* and *S.servazzii* (35). Perfect palindromes, which are two copies of a sequence located on opposite strands, were found using 30-nt windows and a repeat length of 7 nt. Global repeats within the entire genome were determined using a repeat length of 100 nt as the window. A direct repeat is a 100-bp sequence that is present in at least two copies and located on the same strand; inverted repeats are located on opposite strands (36). The DNA atlases of *S.castellii*, *S.servazzii* and other yeast mtDNA genomes are available on our web site (http://www.cbs.dtu.dk/services/GenomeAtlas/scast/mito_table.html).

RESULTS

General features

The size and gene order of the mitochondrial genomes of *S.castellii* and *S.servazzii* was previously determined by restriction analysis and mapping (21). Sequencing of the whole mitochondrial genomes has shown that these two mtDNAs consist of 25 753 and 30 782 bp, respectively, and in Figure 1 they are presented as circular molecules. For *S.castellii* the A+T content of the entire genome and of the exonic ORFs (the eight structural genes only) is 79.6 and 77.5%, respectively. *VARI* differs from the average by having an A+T content at 94.2%, and *ATP9* differs by having only 65.8% A+T. The intronic ORF encoded by *S.castellii* *COX1.2* has an A+T content (76.9%) close to the average. For *S.servazzii*, the A+T content of the entire genome is 77.5%, whereas the exonic ORFs (including *ORF1*) have 75.8% A+T. The *VARI* gene differs from the average by having 89.0% A+T, and *ATP9* differs by having only 69.3% A+T.

The *S.castellii* mtDNA molecule consists of 52.7% gene encoding regions (tRNAs, rRNAs and eight protein-encoding) and 7.6% introns, leaving 39.7% for intergenic sequences. Similarly, *S.servazzii* mtDNA consists of 45.6% genes (tRNAs, rRNAs and nine protein-encoding), 11.4% introns and 43.0% intergenic sequences. *Saccharomyces castellii* and *S.servazzii* mitochondrial sequences are analysed for their coding regions, transcription units, gene order, introns and intergenic regions, as described below. Both mitochondrial sequences are compared to each other and to the *S.cerevisiae* sequence, as well as to other relevant fungal sequences, in



order to deduce possible mechanisms, which operated during evolution of the yeast mitochondrial genome.

Protein-coding regions

The gene content of the mitochondrial genomes of *S. castellii* and *S. servazzii* is typical for the already characterised *Saccharomyces*, *Kluyveromyces* and *Schizosaccharomyces* yeasts, with the NADH dehydrogenase complex 1 subunits not being present. Due to similarity to the already known yeast mitochondrial genes, eight structural genes were identified as *COX1*, *COX2*, *COX3* (cytochrome oxidase subunits I, II and III), *ATP6*, *ATP8* and *ATP9* (ATPase subunits 6, 8 and 9), *COB* (apocytochrome b) and *VARI* (ribosomal associated protein) (Fig. 1). All ORFs encoding structural genes start at an ATG initiation codon and terminate at a TAA termination codon, with two exceptions in *S. servazzii*. One is the URF called *ORF1*, which shows homology to Q0255, a hypothetical ORF of the mitochondrial genome of *S. cerevisiae*. *Saccharomyces servazzii ORF1* probably encodes a maturase and has ATA as the start codon and TAG as the stop codon. The other exception is an intronic ORF that has ATG as the start and TAG as the stop codon. All ORFs in both species are encoded by the same strand.

Interestingly, *S. servazzii ATP6*, *COX2* and *COX3* ORFs each contain a +1 frameshift at amino acid positions 191, 229 and 249, respectively. The strong amino acid sequence conservation found between the *S. servazzii* deduced protein sequences and those of *S. cerevisiae* and *S. castellii* suggests that these frameshifts are due to one additional T residue within the 3' end (Fig. 2). Most strains of *C. glabrata*, a species closely related to *S. cerevisiae*, also contain a +1 frameshift in *COX2* (37). Interestingly, the +1 frameshifts in the *COX2* genes of *S. servazzii* and *C. glabrata* are located near the 3' end of *COX2* and are only separated by five amino acids. In addition, *S. servazzii COB* contains two +1 frameshifts also caused by additional T residues located after the annotated intron at positions 74 and 82 and separated by 14 bp (Fig. 2).

Approximately 1.5 kb downstream of the *COB* gene in the *S. castellii* mtDNA molecule, there is a pseudo 3' part of a *COB* sequence. This is seen as a global direct repeat in Figure 1 and indicates that a partial gene duplication has occurred. When the amino acid sequences of the two *S. castellii* repeat regions are compared to the 3' end of the *S. cerevisiae COB* gene, the *S. castellii* pseudo 3' part of *COB* shows higher similarity (84% identity) than the 3' end of the true *S. castellii COB* (72% identity), whereas the identity between the two *S. castellii*

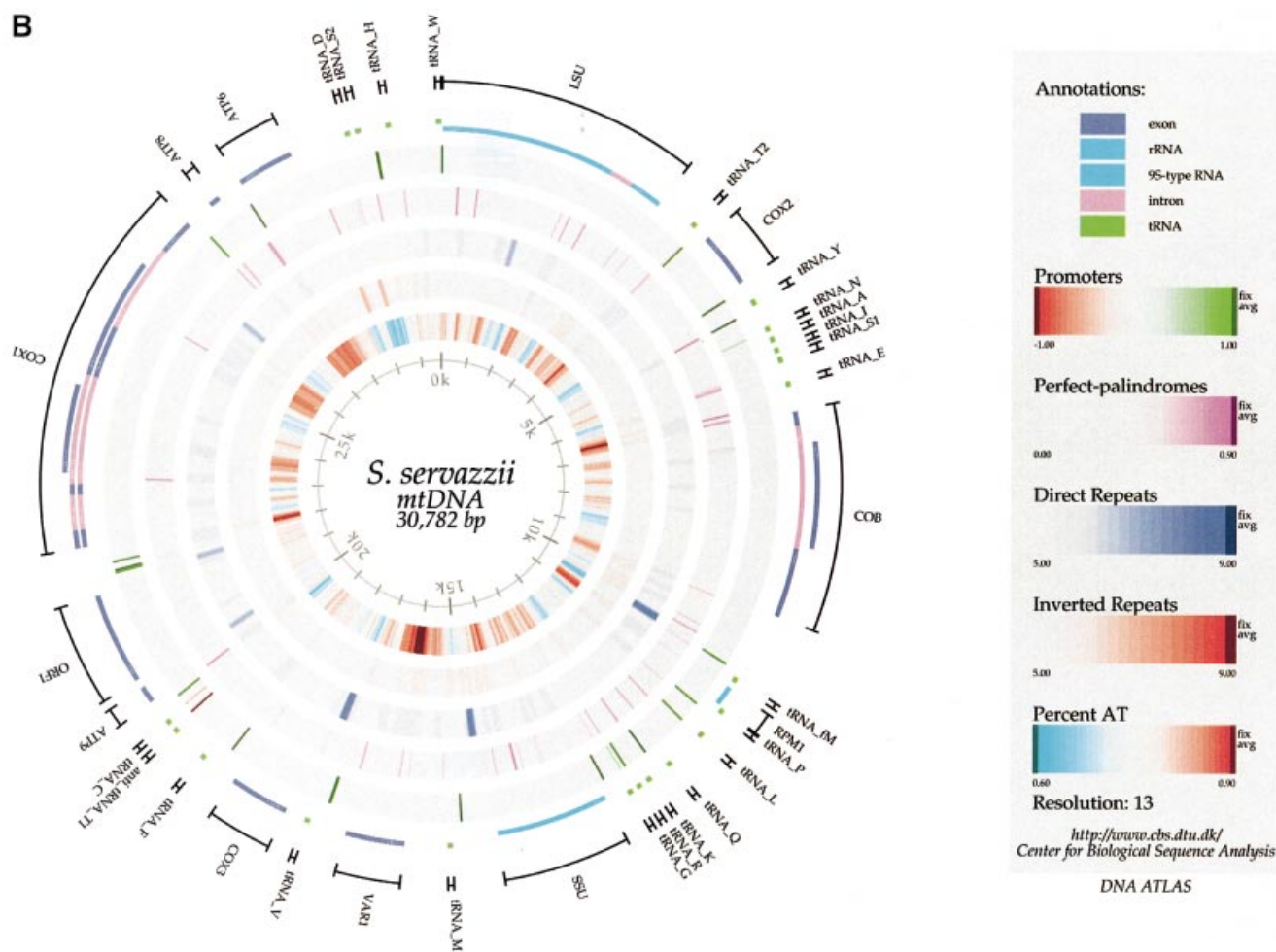


Figure 1. (Previous page and above) Circular DNA atlases of the mitochondrial genome of *S.castellii* (A) and *S.servazzii* (B). The outer ring shows the annotated genes and introns. All genes, but one (anti-*tRNA*^{Thr1}), are transcribed clockwise. Next are the locations of promoter motifs that are closer than 500 bp from the downstream gene (modified from a table in Supplementary Material). Green and red bands represent motifs on the positive (clockwise) and negative strand, respectively. Light green and light red bands represent motifs within coding regions. Perfect palindromes are indicated with purple in the following ring and since transcription termination regions often contain palindromes this feature indicates regions with possible transcription termination. The next two rings indicate global direct and inverted repeats as blue and red bands, respectively, ranging from 50 to 90%. The innermost ring is the calculated A+T content, ranging from 60% (turquoise) to 90% (red). The *S.castellii* map was modified from Petersen *et al.* (23).

3'-*COB* sequences is only 75%. Since both copies of the *COB* 3' part have a conserved amino acid sequence, they might both be functional. Another interesting observation is the lack of palindromic sequences (see later section) in the 1.5 kb region between the two 3' parts.

Ribosomal, transfer and RNase P RNA genes

Based on sequence similarity, the mitochondrial genomes of both *S.castellii* and *S.servazzii* contain the large subunit and small subunit ribosomal RNA (*LSU* and *SSU*) genes. The 5' and 3' boundaries are estimated from sequence comparison to rRNAs from *S.cerevisiae* (1). *tRNA* genes were predicted from their secondary structure. *Saccharomyces castellii* and *S.servazzii* each contain 23 *tRNA* genes. All annotated *tRNA* genes are encoded by the same strand as all the other genes, except for *tRNA*^{Thr1}. In comparison, *S.cerevisiae* and *P.canadensis* mtDNAs encode 24 and 25 tRNAs, respectively

(1,8). The codon usage within the exonic ORFs and the corresponding tRNAs can be found in the Supplementary Material. The anticodons for all tRNAs are identical between *S.castellii* and *S.servazzii*, and the genetic code seems to be the same as the one used by the *S.cerevisiae* mitochondria. The *tRNA* gene that is missing in *S.castellii* and *S.servazzii* is that corresponding to the CGN family, which in *S.cerevisiae* encodes arginine. This codon family is used neither in any of the predicted exonic ORFs, nor in the intronic ORFs. The same phenomenon is seen in other yeasts, like *C.glabrata* (13), *Kluyveromyces thermotolerans* (38) and in exonic ORFs of *Y.lipolytica* (9). The CUN family does not specify leucine but instead threonine, as in *S.cerevisiae* and *C.glabrata* (reviewed in 39). The corresponding *tRNA* gene is found on the opposite strand in *S.cerevisiae*, *S.castellii* and *S.servazzii*, but on the same strand as all other genes in *C.glabrata* (13). For all three *Saccharomyces* yeasts, the predicted secondary structure

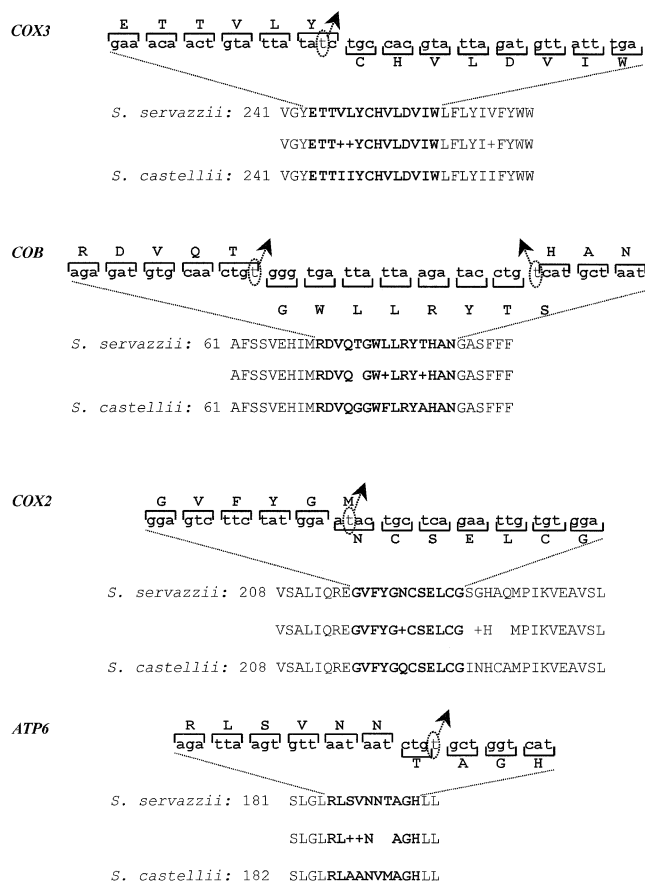


Figure 2. Representation of the DNA sequences surrounding the frameshifts in the four *S. servazzii* ORFs and the conceptual translation of these regions. In each case, a +1 T residue giving the most conserved amino acid sequence was chosen, although in *COX3*, a +1 C frameshift affecting the third base of the codon would preserve the tyrosine. In *COX2*, a +1 C frameshift affecting the third base of the codon would result in a methionine. On the other hand, in *COB* and in *ATP6*, any other combination would either impair the sequence conservation or introduce a CGN codon. Nucleotide sequences are shown in lowercase letters and amino acid sequence in capital letters. The dotted circles and arrows show the extra T residues in the nucleotide sequences that might be a target for post-transcriptional editing.

contains eight nucleotides and not the normal seven in the anticodon loop (40) and the anticodon is UAG. *Saccharomyces castellii* and *S. servazzii* differ from *S. cerevisiae* by having an additional nucleotide in the D-loop, but interestingly the extra nucleotide is differently positioned. Several codons are missing in the *S. castellii* and *S. servazzii* mitochondrial genomes, and, as expected for A+T-rich genomes, all missing codons are G+C-rich.

In *S. cerevisiae*, the gene encoding Rpm1r of RNase P (*RPM1*) is found between *tRNA^{Met}* and *tRNA^{Pro}* and the three genes are transcribed together as one unit (1,41). This organisation is also seen in other yeasts, like *C. glabrata*, where co-transcription also occurs (42), and in *Saccharomyces exiguus* (43) and *Kluyveromyces lactis* (44). In *S. castellii* the same *tRNA* genes are separated by 627 bp of which at least 525 nt presumably encode the Rpm1r, starting only a few nucleotides from the 3' end of *tRNA^{Met}*. Two regions, one near the 5' end and the other near the 3' end, have high identity to the conserved sequences that are found in the *RPM1* of

C. glabrata, *K. lactis*, *Saccharomycopsis fibuligera* and other *Saccharomyces* yeasts (42,43,45). These two regions make up the conserved helix P4 in the potential secondary structure (data not shown) that is similar to other known RNA components of RNase Ps (reviewed in 46). In the case of *S. servazzii*, the same two *tRNA* genes are separated by 382 bp of which 277 nt presumably encode the Rpm1r. The assumption is based on the same conserved features as mentioned for *RPM1* of *S. castellii*. For comparison, the Rpm1r of *K. lactis* is 191 nt (44), that of *C. glabrata* is 227 nt (42), and that of *S. exiguus* is 277 nt (43), while that of *S. cerevisiae* is 453 nt (41).

Putative transcription initiation sites and endonucleolytic cleavage sites

In *S. cerevisiae*, the nonanucleotide motif sequence 5'-TATAAGTAA-3' is considered as the transcription initiation site (47) and is similar to what is found in *C. glabrata*, 5'-TATAAGTA-3' (13) and *P. canadensis*, 5'-TATAAG(T/A)(A/T/c/g)-3' (8). Based on the consensus sequences, we used the motif TATAAG to detect putative transcription initiation sites in *S. castellii* and *S. servazzii* mtDNAs, and several nonanucleotide motif sequences 5'-TA(T/a)AAG(A/T/g/c)(A/T/c/g)(A/T/c/g)-3' and 5'-TA(T/a)AAG(A/T/g/c)(A/T/g/c)(A/c/t/g)-3' were detected in *S. castellii* and *S. servazzii*, respectively (Fig. 1 and Supplementary Material). However, several of these motifs, located within genes or far upstream, are likely not to be transcription initiation sites. Taking this into account the consensus of the motif sequences is 5'-TATAAG(A/T)(A/T)(A/T)-3' and 5'-TATAAG(T/A)(A/T)A-3' for *S. castellii* and *S. servazzii*, respectively. Motifs with the core sequence 5'-TATAAG-3' were not found upstream of *S. servazzii* anti-*tRNA^{Thr1}*, but instead two motifs with the core sequence 5'-TAAAAG-3' were located in this region. The putative transcription initiation sites that were located <500 bp away from the downstream gene are indicated in Figure 1.

In *S. cerevisiae* the dodecamer motif 5'-AATAATATTCTT-3' is found downstream of most protein-coding genes and believed to be an endonucleolytic cleavage site needed for RNA processing of multi-gene transcripts (48–50). Likewise, the motif 5'-TATAATATTCTT-3' has been found in *C. glabrata* (13). In *S. castellii* and *S. servazzii* similar motifs are found downstream of several genes and have the consensus sequences 5'-A/T/c ATAATATTC A/c/t T/A-3' and 5'-(A/t/c)(T/a/g)(TAATA(a)TTC(T/A)(A/T/c)-3', respectively (see Supplementary Material). When only the motifs downstream of protein-encoding genes are considered, the consensus sequences change slightly to 5'-(A/t)ATAATATTC(A/c)(T/A)-3' in *S. castellii* and 5'-(A/t)(T/a)TAATA(a)TTC(T/A/g)(T/A/C)-3' in *S. servazzii*. Notice that the pseudo 3' end of *COB* in the *S. castellii* mtDNA is followed by two perfect motifs. Interestingly, the consensus motif was found in the 3' end of RNA genes, like in *C. glabrata* (13). The main difference between the putative 3' processing sites in the two studied yeasts is that the second base is an A in *S. castellii* and a T in *S. servazzii*.

Transcription termination could occur in regions containing palindromic sequences. Therefore, the mitochondrial genomes of *S. castellii* and *S. servazzii* were examined for perfect palindromes (Fig. 1) and, in general, these were found in intergenic regions. As expected, the *RPM1* as well as rRNA

regions contain perfect palindromes to a much higher extent than the protein-coding regions, with the exception of the very AT-rich *VARI*, and, not surprisingly, perfect palindromes were found in some of the *S.castellii* and *S.servazzii* introns.

Gene order and transcription units

The gene order of the mitochondrial genomes of *S.castellii*, *S.servazzii* and *S.cerevisiae* was compared (Fig. 3). Several of the gene clusters are shared with other related species. The three genes, *COX1*, *ATP8* and *ATP6*, constitute the most conserved gene cluster among hemiascomycetous yeasts. This cluster has been found in *S.cerevisiae*, in all the *Saccharomyces* yeasts analysed by Groth *et al.* (21), in *C.glabrata* (13), *K.thermotolerans* (38), *K.lactis* (44, reviewed in 51), *P.canadensis* (8) and in *Y.lipolytica* (9), suggesting that this is a common ancestral unit. In *S.cerevisiae* they form a transcription unit (50), whereas in *Y.lipolytica* *ATP8*, *ATP6*, *COX3* and *ND4* are co-transcribed (9). The conservation of gene order and the presence of putative endonucleolytic cleavage sites downstream of *COX1* and *ATP6* in both *S.castellii* and *S.servazzii* supports co-transcription of these three genes (Fig. 3). However, promoter motifs are found upstream of the first two genes in *S.castellii* and upstream of all three genes in *S.servazzii* suggesting that at least *COX1* could also be transcribed independently.

In *S.cerevisiae*, the *tRNA^{Glu}* and *COB* genes are co-transcribed (see below), while *ATP9* is the first gene in a transcription unit containing *tRNA^{Ser1}* and *VARI* (reviewed in 52). The latter unit is not conserved in the *Saccharomyces* sensu lato yeasts (Fig. 1), but it seems to have emerged in the lineage leading to the *Saccharomyces* sensu stricto yeasts (21). The gene order of protein-encoding genes in the related yeasts, *S.exiguus* (53–55), *C.glabrata* and *K.thermotolerans* (38) provides some insight on the fate of *COX2*. Together with *S.castellii*, these yeasts share a cluster of six protein-encoding genes, *COB-COX1-ATP8-ATP6-ATP9-COX2*, which presumably existed already in the ancestral mtDNA. It could be that *COX2* in both *S.cerevisiae* and *S.servazzii*, as well as *ATP9* in *S.servazzii*, have moved to a new position. In *S.castellii* and *S.servazzii*, *COX2* is succeeded by a tRNA cluster with the conserved gene order (Fig. 3, cluster 6 and see below). Since promoter motifs are found upstream of both the tRNA cluster and the *COX2* gene, the latter seems to be transcribed alone. A cluster shared between *S.castellii* and *S.servazzii* contains the three genes, *tRNA^{Phe}*, *anti-tRNA^{Thr1}* and *tRNA^{Cys}* (Fig. 3, cluster 2). The same gene order is seen in *C.glabrata*, but in this yeast *tRNA^{Thr1}* is located on the same strand as all other genes and implies that *tRNA^{Thr1}* at some point in the evolution has changed the strand location. When the mtDNA of *S.servazzii* was examined for putative initiation sites upstream of the *anti-tRNA^{Thr1}*, only motifs with the altered core sequence 5'-TAAAAG-3' were found. In addition, *S.castellii* and *C.glabrata* share the gene order of the following two genes, *tRNA^{Val}* and *COX3*, with *S.cerevisiae*. Notice that in *S.cerevisiae*, *tRNA^{Cys}* has moved to a new position and that the cluster containing *tRNA^{Phe}*, *anti-tRNA^{Thr1}* and *tRNA^{Cys}* has moved downstream of *COX3* in *S.servazzii*. In conclusion, the gene order found in *S.castellii*, *S.servazzii*, *S.cerevisiae* and *C.glabrata* suggests that *tRNA^{Phe}-anti-tRNA^{Thr1}-tRNA^{Cys}-tRNA^{Val}-COX3* comprise a conserved ancient block.

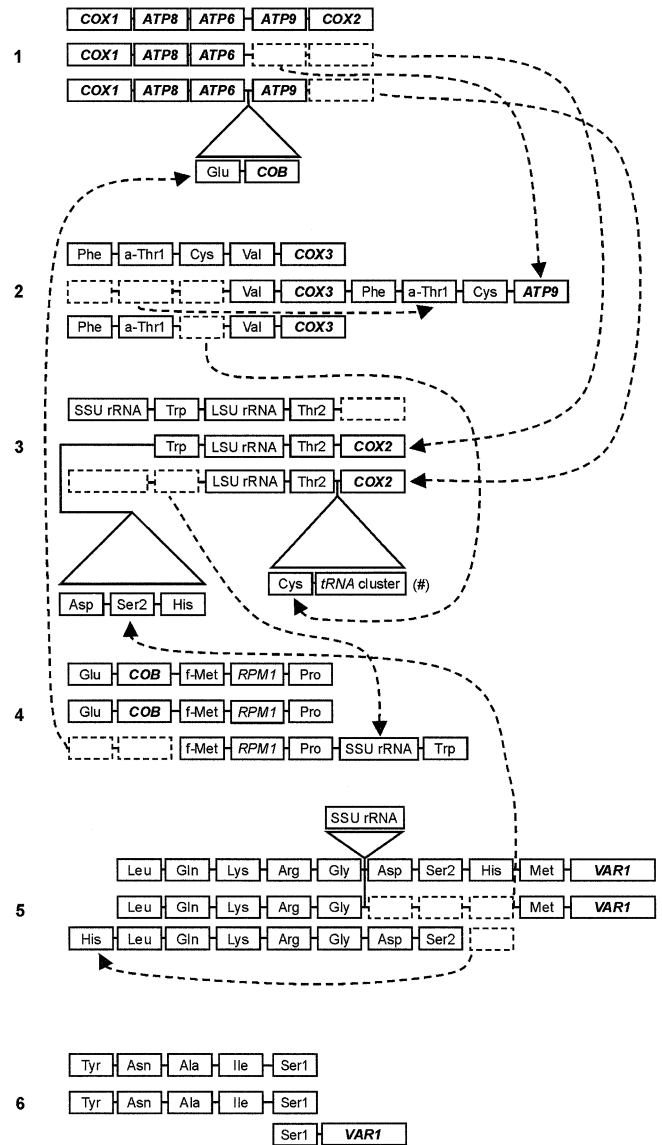


Figure 3. Schematic illustration of clusters (1–6) with conserved gene order as well as some possible rearrangements between *S.castellii* (top), *S.servazzii* (middle) and *S.cerevisiae* (bottom). Dotted lines with arrows indicate transposition/jumping of genes from positions marked with dotted boxes. Black lines connecting the boxes show adjacent genes. Some isolated tRNA genes from *S.cerevisiae* as well as all URFs are not included. tRNAs are named by the amino acid that they specify and numbers refer to those of *S.cerevisiae* (1) or homologues of *S.cerevisiae* tRNAs. Protein-encoding genes are bold. Figures not drawn to scale. a-Thr1, anti-*tRNA^{Thr1}*. (#) The large *S.cerevisiae* tRNA cluster consists of *tRNA^{His}*-*tRNA^{Leu}*-*tRNA^{Gln}*-*tRNA^{Lys}*-*tRNA^{Arg1}*-*tRNA^{Gly}*-*tRNA^{Asp}*-*tRNA^{Ser2}*-*tRNA^{Arg2}*-*tRNA^{Ala}*-*tRNA^{Ile}*-*tRNA^{Tyr}*-*tRNA^{Asn}*-*tRNA^{Met}*.

Whereas one of the transcription units in *S.cerevisiae* presumably contains *tRNA^{Phe}*, *tRNA^{Val}* and *COX3* (47, reviewed in 52), promoter motifs are found upstream of both *tRNA^{Phe}* and *tRNA^{Val}* in *S.castellii* (Fig. 1), and in *S.servazzii*, where rearrangements have occurred, promoter motifs are found upstream of *tRNA^{Val}*, *tRNA^{Phe}* and *ATP9*. While the genes in cluster 2 (Fig. 3) are contained in one transcription unit in *S.cerevisiae*, at least two transcription units are likely to be present in *S.castellii* and *S.servazzii*.

As in *S.cerevisiae* and *K.lactis*, the *SSU* and *tRNA^{Trp}* genes are adjacent in *S.castellii*. Whereas *S.castellii* has a perfect promoter motif upstream of *SSU* and another promoter motif just upstream of *tRNA^{Trp}*, *S.cerevisiae* only seems to have an active promoter upstream of *SSU*. Since the gene order of *SSU* and *tRNA^{Trp}* is conserved in several yeasts, *SSU* has supposedly moved to a new location in *S.servazzii* (Fig. 3, clusters 3 and 5). *Saccharomyces castellii* and *S.servazzii* share the gene order, *tRNA^{Trp}-LSU-tRNA^{Thr2}*, with *C.glabrata*. In *S.cerevisiae* one transcription unit contains the *LSU* gene followed by *tRNA^{Thr2}* (56,57). Whereas *S.servazzii* has a promoter motif just upstream of *LSU*, *S.castellii* does not and the transcription unit may begin with *tRNA^{Trp}* (Fig. 1).

Saccharomyces castellii and *S.servazzii* share the five-gene cluster, from *tRNA^{Glu}* to *tRNA^{Pro}*, with *C.glabrata*, whereas in *S.cerevisiae* this cluster has been split into two (Fig. 3, cluster 4), with the *tRNA^{Glu}* and *COB* genes probably being transcribed as one unit from a promoter upstream of *tRNA^{Glu}* (58). Whereas the arrangement of the two genes is the same in the *Saccharomyces sensu lato* yeasts, the presence of promoter motifs differs. Such motifs are found upstream of *tRNA^{Glu}* (but within *VARI*) and *COB* in *S.castellii*, while none is found in *S.servazzii*. The lack of suitable promoter motifs could suggest that these yeasts have transcription initiation sites with alternative motifs. Both *S.castellii* and *S.servazzii* have putative endonucleolytic cleavage sites downstream of *COB* (see Supplementary Material).

As previously mentioned, the gene order of *tRNA^{Met}*, *RPM1* and *tRNA^{Pro}* is conserved in many different yeast species: *S.cerevisiae* (41), *C.glabrata* (42), *S.exiguus* (43) and *K.lactis*. *Saccharomyces castellii* has one perfect promoter motif upstream of *tRNA^{Met}*, three promoter motifs within the *RPM1* and one putative endonucleolytic cleavage site downstream of *RPM1*, whereas *S.servazzii* has just one perfect promoter motif upstream of *tRNA^{Met}* (Fig. 1, see also Supplementary Material). As in *S.cerevisiae* and *C.glabrata*, these three genes are likely transcribed as one unit in both *S.castellii* and *S.servazzii*.

A large cluster containing seven tRNA genes from *tRNA^{Leu}* to *tRNA^{Ser2}* is shared not only between *S.castellii* and *S.cerevisiae* (Fig. 3, cluster 5) but also with *C.glabrata* and *K.lactis*, whereas it is split up into smaller pieces in *S.servazzii* (Fig. 3) and *P.canadensis*. The conservation continues between *S.castellii* and *K.lactis*, where *tRNA^{Ser2}* is followed by *tRNA^{His}* and *tRNA^{Met}*. In addition to *S.castellii* and *S.servazzii*, *tRNA^{Met}* and *VARI* are adjacent in *C.glabrata*. The comparison of the three *Saccharomyces* yeasts and *K.lactis* suggests that the three tRNA genes specifying asparagine, serine and histidine have switched position with *SSU* in *S.servazzii* and that *tRNA^{His}* has moved in front of *tRNA^{Leu}* in *S.cerevisiae* (Fig. 3, clusters 3 and 5).

In *S.castellii* and *S.servazzii* the tRNA cluster (Fig. 3, cluster 6) located downstream of *COX2* and containing *tRNA^{Tyr}*, *tRNA^{Asn}*, *tRNA^{Ala}*, *tRNA^{Ile}* and *tRNA^{Ser1}*, is also found in *K.lactis*. In *C.glabrata*, the first four tRNA genes have the same order, whereas in *S.cerevisiae* and *P.canadensis* these genes are shuffled. In *S.cerevisiae*, these four genes are part of the large tRNA cluster, which is not adjacent to *tRNA^{Ser1}* and *VARI* (Fig. 3, cluster 3). Based on the presence of promoter motifs, the five tRNA genes are likely to be

transcribed together in *S.castellii*, while *tRNA^{Tyr}* in *S.servazzii* might be transcribed alone (Fig. 1).

Introns and intronic ORFs

Mitochondrial introns are divided into two groups, I and II, based on their function, structure and splicing mechanism. Based on sequence homology of both the nucleotide sequence and the expected amino acid sequence, introns have been predicted to interrupt genes in both *S.castellii* and *S.servazzii* (Table 1). The *S.castellii COX1* gene is interrupted by two introns of 487 and 1123 nt in length. The ORF from the first exon continues only ~10 nt into the first intron (*Sca cox1.1*) and encodes neither a separated gene if not spliced, nor a free-standing ORF. The second intron in the *S.castellii COX1* gene (*Sca cox1.2*) shows high identity to the *S.cerevisiae* intron, *Sc cox1.4*. In addition, it shows limited identity to the *S.cerevisiae* intron *Sc cob.4* and the *P.canadensis* intron, *Pc cox1.1*. Together with exon1 and exon2, it makes a 1797 bp ORF encoding 598 amino acids that shows high similarity (73% identity) to the DNA endonuclease I-SceII from *S.cerevisiae* and 61% identity to a probable site-specific DNA endonuclease from *P.canadensis*. The deduced amino acid sequence contains two LAGLI-DADG motifs (Table 1).

Saccharomyces servazzii COX1 contains three introns of 475, 1471 and 1230 nt in length. The first intron, *Ss cox1.1* does not encode an ORF, while *Ss cox1.2* contains a 1167 bp free-standing ORF encoding 388 amino acids with one or maybe two LAGLI-DADG motif(s). The third intron in *S.servazzii COX1*, *Ss cox1.3* encodes together with the upstream exons a hypothetical protein of 675 amino acids with two LAGLI-DADG motifs and with homology to an intron-encoded endonuclease from *Y.lipolytica* (Table 1). The *S.castellii LSU* gene is interrupted by a 356 nt intron, *Sca lsu.1*, at the same position as the intron in the *S.cerevisiae LSU* gene, *Sc lsu.1*, also known as the ω (omega) intron (31,59). The *S.castellii* intron has previously been described as a group I intron resembling *Sc lsu.1*, and was shown not to encode a homing endonuclease gene (HEG) like I-SceI in *Sc lsu.1* (60). In *S.servazzii*, a 326 nt intron was found at the same position as described for other *Saccharomyces* yeasts, including *S.cerevisiae* and *S.castellii*. Apparently, this intron does not encode a I-SceI homologue (Table 1).

Finally, a 1669 nt intron is found in the *S.servazzii COB* gene, *Ss cob.1*, and contains a 1401 bp free-standing ORF encoding 466 amino acids that shows limited similarity to several intron-encoded ORFs, especially from mtDNA of *P.anserina*, and to a hypothetical ORF from the mtDNA of *S.cerevisiae*, Q0255. The deduced amino acid sequence contains one LAGLI-DADG motif (Table 1). In addition to the intronic ORFs, the URF (*ORF1*) in *S.servazzii* contains one or maybe two LAGLI-DADG motif(s).

Intergenic regions and repeats

The *S.cerevisiae* mtDNA consists of approximately two-thirds of intergenic sequences (2), which contains approximately 200 repetitive A+T-rich spacers with an average size of 190 bp, separated by G+C-rich clusters. The *S.cerevisiae ori/rep* sequences are formed by three GC-clusters having the conserved sequences, cluster A, GGGGGTCCCC, cluster B, GGGACCCGG and cluster C, CACCCACCCCCTCCCC (3). Cluster C or basic monomeric penta-C units and the

Table 1. Exon-intron junction and intronic ORFs in *S.castellii* and *S.servazzii* mtDNA

Intron name	Exon upstream	Intron 5' end	Intron 3' end	Exon downstream	Intron group	BLASTN	Intronic ORF	BLASTX
Group I	...T		...G		IA or IB	–	–	–
Group II		GUGCG...	...AY		II	–	–	–
Sca cox1.1	GGTTTGGT	AAATTATA	TTTTTATG	AATTATAT	IB	Sc cox1.3	None	–
Sca cox1.2	TCTTTGGT	CTTCTTAT	GCTGAATG	CATCCAGA	IB	Sc cox1.4 + Sc cob.4 + Pc cox1.1	2 LD	I-SceII
Ss cox1.1	GTTTTGGT	AACATACC	TAAATAAG	AATTATAT	IB	–	None	–
Ss cox1.2	TGAAGTGT	ATGAATAT	AAAAAATG	GTATCCTC	IB	–	1–2 LD, F	Aa COX1-i1 + Yl COX1-i3
Ss cox1.3	TCCATGAT	AGAAAAAA	TATAAAAAG	ACATACTA	IB	Kl cox1.4 + Yl cox1.7	2 LD	Kl COX1-i4 + Yl COX1-i7
Sca lsu.1	CTAGGGAT	AAATTACA	ATAAATTG	AACAGGGT	IA	Se lsu.1	None	–
Ss lsu.1	CTAGGGAT	AAGTACCC	AAAATTTG	AACAGGGT	IA	Tp lsu.1 + Kt lsu.1	None	–
Ss cob.1	TAGAACAT	ATTATTTA	TATTTTTG	ATCATGAG	IB	–	1 LD, F	Pa COB-i1

The names of the *S.castellii* and *S.servazzii* introns are shown along with the 5' and 3' end of exons and introns. The introns were classified as group IA or IB from schematic drawings of their possible secondary structure (data not shown) according to Burke (34) (reviewed in 80). Intronic encoded ORFs were examined for motifs like those found in different endonucleases. Last, sequence homology was searched by comparing the intron sequences to the GenBank database using BLASTN and their translated sequences if ORFs were present to the GenBank database using BLASTX. F, free-standing ORF; LD, LAGLI-DADG motif(s); Aa, *Agroclybe aegeria*; Kl, *K.lactis*; Kt, *K.thermotolerans*; Pa, *P.anserina*; Pc, *P.canadensis*; Sca, *S.castellii*; Sc, *S.cerevisiae*; Se, *S.exiguus*; Sp, *S.pombe*; Ss, *S.servazzii*; Tp, *Torulaspota pretoriensis*; Yl, *Y.lipolytica*.

flanking sequences, r* and r, are believed to be the most ancient sequence elements in the *ori* sequences, since bi-directional DNA replication is initiated by RNA primers starting at r* and r and continuing into DNA chains at cluster C (3,61). In *S.castellii* only one element was identical to cluster A, whereas similarities to all three GC clusters were found at several locations but never within reasonable distance and/or in the same direction. The mtDNA of *S.servazzii* was also examined for *ori/rep*-like GC clusters. Only two sequence elements identical to cluster A were found, one on each strand. Several penta-C and penta-G units were found in the mtDNA of *S.castellii* and *S.servazzii* but never followed by sequences homologous to r or r', which are described in de Zamaroczy *et al.* (62). The mtDNA of *S.servazzii* is approximately one-fifth larger than the mtDNA of *S.castellii*, has a higher GC content and almost three times as many GC-rich clusters.

The mtDNAs of *S.castellii* and *S.servazzii* were also examined for repeats. According to the analysis shown in Figure 1, a few global repeats were found, while several simple repeats and local repeats predominantly were found in the intergenic regions (D.W. Ussery, unpublished data; see http://www.cbs.dtu.dk/services/GenomeAtlas/scast/mito_table.html). In the *S.cerevisiae* *VARI* gene, two GC clusters (46 bp in length) of opposite orientation have been shown to be involved in recombination leading to size variation of this gene (63). No such GC cluster was found in the *S.castellii* or *S.servazzii* *VARI* genes. Instead, the *S.castellii* *VARI* gene contains one pair of large global (direct) repeats. This region contains a significant proportion of local repeats (direct, inverted, mirror, everted) and perfect palindromes (Fig. 1 and D.W. Ussery, unpublished data; http://www.cbs.dtu.dk/services/GenomeAtlas/scast/mito_table.html). In addition to *VARI*, two other pairs of global repeats are found in the *S.castellii* mtDNA; one pair is located in the intergenic region upstream of *tRNA^{Val}* and the other pair is seen as the implied duplication of the 3' end of the *COB* gene (Fig. 1). Global repeats are also seen in the *S.servazzii* mtDNA; one pair of inverted repeats and two pairs plus one set of three

direct repeats (Fig. 1). Most repeats are located in intergenic regions but one pair of direct repeats is located in the *COX1* and the *LSU* gene.

DISCUSSION

Whether the *S.cerevisiae* mtDNA is circular or linear is controversial (5), but the present sequence analysis suggests that *S.castellii* and *S.servazzii* mtDNAs can exist as circular molecules at least for a short period of their life cycle (Fig. 1). On the other hand, several yeasts have been shown to have linear mtDNA molecules (64). Whereas the coding potential of the *S.cerevisiae*, *S.castellii* and *S.servazzii* mtDNAs is similar, they differ significantly in their size (Fig. 1). The mtDNA of *S.cerevisiae* consists of 16% genes, 22% introns and 62% intergenic sequences (2). Considering the distribution in the mtDNA of *S.castellii* (53, 7 and 40%) and *S.servazzii* (46, 11 and 43%) it seems that as the genome size increases, the amount of intergenic regions and introns rises, confirming that the size variation of the mtDNA among yeasts belonging to the *Saccharomyces* genus is mainly caused by the variable length of intergenic sequences and the presence or absence of introns (19). As in *S.cerevisiae*, all annotated genes are located on one strand, except for *tRNA^{Thr1}* (Fig. 1).

Saccharomyces castellii and *S.servazzii* use the yeast mitochondrial genetic code for translation of their mitochondria-encoded genes but a few small differences in codon usage exist among the three *Saccharomyces* species (see Supplementary Material). Whereas *S.servazzii* almost entirely uses the CUA codon of the CUN family, *S.castellii* uses equal amounts of CUA and CUU. Another significant difference is the absence of the lysine specifying AAG codon in *S.castellii*, while almost one-tenth of the lysine codons in *S.servazzii* are AAG. The fact that the CGN family is missing in several yeasts, such as *S.castellii*, *S.servazzii*, *C.glabrata*, *K.thermotolerans* and *Y.lipolytica*, but is present in *S.cerevisiae*, contradicts the previously suggested time point in evolution where the CGN should have become non-coding

in the lineage leading to *S.cerevisiae* after the separation from *C.glabrata* (reviewed in 65). It is more likely that the CGN family first became non-coding somewhere in the lineage leading to the Hemiascomycetes and then later in evolution was regained in *S.cerevisiae*. Another explanation could be that the CGN codon independently became non-coding in several lineages. Interestingly, while mtDNA of *S.cerevisiae* encodes tRNA^{Arg} with the anticodon ACG, *S.pombe* and all metazoan mitochondria have a tRNA^{Arg} with the anticodon UCG (reviewed in 65). Since no tRNA^{Arg} is imported into the *S.cerevisiae* mitochondria (reviewed in 66) and A normally only pairs with U, it is not known if the tRNA^{Arg} is functional in *S.cerevisiae*. However, unconventional pairing of A with C, G, and A has been seen with *in vitro* translation using anticodon AGU in *Mycoplasma capricolum* (67). In *Y.lipolytica*, CGN codons have been found in intronic ORFs that also contain mutations and thereby suggest that these intronic ORFs are pseudo-genes (9).

The *S.servazzii* mitochondrial genome contains several +1 frameshifts that affect four (*ATP6*, *COX2*, *COX3* and *COB*) of the nine detected exonic ORFs. In *COX2*, a +1 C frameshift affecting the third base of the codon would result in a methionine. A +1 C frameshift in *COX2* was described in all the tested strains of *C.glabrata* except for CBS138 (37). Recently, sequencing of the mtDNA of the latter strain confirmed the presence of the same frameshift (A. Malpertuy, personal communication). Interestingly, the +1 frameshifts in *COX2* of *S.servazzii* and *C.glabrata* are located near the end of the ORF and are only separated by five amino acids. Another case of +1 frameshift in *COX2* was described in a mutant of *S.cerevisiae* that was shown to be suppressed to a high extent by a mutation in the anti-codon stem of a tRNA^{Ser}, through a likely alteration of the pairing between the tRNA and mRNA molecules (68,69). The mechanism described for *S.cerevisiae* is unlikely to be used in *S.servazzii* as this would require that different tRNAs would be specifically mutated to bypass the five detected frameshifts. Other suppression mechanisms involve the insertion of an extra nucleotide in the anticodon loop of the tRNA that would sterically modify the interactions between the codon and the anticodon and facilitate a 4-nt translocation (reviewed in 70). Except in the anti-tRNA^{Thr1}, which has an extra G in the anti-codon stem, as in that of other yeasts, no extra nucleotide was detected in the *S.servazzii* mtDNA tRNAs. In *COX3*, both a +1 T and a +1 C frameshift affecting the third base of the codon would conserve the tyrosine. In *COB* and in *ATP6*, any other combination than the suggested +1 T residues would either impair the sequence conservation or introduce a CGN codon. Removing the first T residue in *COB* (Fig. 2) ensures the conserved glycine at position 75 and prevents the occurrence of the unusual TGG codon, which is not used in the *S.servazzii* or *S.castellii* mtDNA (see Supplementary Material). Programmed frame-shifting was previously described for translation of *ORF2* of the *S.cerevisiae* retrotransposon Ty (71). Several mechanisms of RNA editing were also described in mitochondria of various organisms that rely on the post-transcriptional modification of tRNAs or mRNAs, like U-deletion in the mitochondria of trypanosomes (72).

Compared to *S.cerevisiae*, only a few introns were found in the mtDNA of *S.castellii* and *S.servazzii* and only introns belonging to group IA and IB were present, whereas no introns

had the 5' or 3' splice site consensus of group II introns (Table 1). It could be that the *Saccharomyces sensu lato* yeasts have lost the group II introns or that these have been acquired only by the *Saccharomyces sensu stricto* lineage. Some of the intronic ORFs (Table 1) and one URF (*ORF1*) in *S.servazzii* contain the LAGLI-DADG motif found in RNA maturases and DNA endonucleases (reviewed in 73). Although intron mobility has not been demonstrated in *S.castellii* and *S.servazzii*, the presence of LAGLI-DADG motifs in intronic ORFs suggest that these introns may be or may once have been mobile (74). In *S.cerevisiae*, the Sc cob.4 encodes a maturase, which excises both the Sc cob.4 and the Sc cox1.4 intron from the cob pre-mRNA and the cox1 pre-mRNA, respectively (75). Skelly and Maleszka (76) found that two out of 21 investigated species, namely *B.custersii* and *C.glabrata*, contain only a cox1.4-like intron and no cob.4-like intron, as in *S.castellii*. A similar situation is found in *P.canadensis* (8). However, in *S.cerevisiae* Sc cox1.4 encodes a potential maturase that can be triggered by a single point mutation (77) suggesting that the cox1.4-like introns in the mentioned yeasts are self-splicing.

The mtDNA of *S.cerevisiae* carries many promoter motifs but not all have been shown to be active (1,47,78). Apparently, *S.castellii* and *S.servazzii* have similar consensus sequences to that of *S.cerevisiae* (Fig. 1 and Supplementary Material). But, like in *S.cerevisiae*, not all these sites are likely to be active. The 3' end of all *S.cerevisiae* mitochondrial mRNAs is formed by endonucleolytic processing within the conserved sequence, 5'-AAUAAUUAUUCUU-3' (reviewed in 52). Surprisingly, the putative endonucleolytic cleavage sites differ not only between *S.castellii* and *S.servazzii*, but also from that of *S.cerevisiae* and that of *C.glabrata* (see Supplementary Material). The appearance of putative transcription initiation sites and endonucleolytic cleavage sites can nevertheless be used to propose possible transcription units in the investigated yeasts. Figure 3 shows the comparison of common gene clusters in the completely sequenced mtDNA of the *Saccharomyces* yeasts: *S.castellii*, *S.servazzii* and *S.cerevisiae*. Several of these clusters are also shared with other related species, like the *sensu lato* yeast, *C.glabrata* (13) and the less closely related petite-negative yeast, *K.lactis* (44, reviewed in 51). The sequence analysis clearly demonstrates that transcription units have not been completely conserved during evolution of the *Saccharomyces* yeasts. Apparently, transcription units containing several genes have been broken into separate units and/or two units have been fused into a single one in some yeast lineages.

The mtDNA molecules and their gene order can get rearranged if a segment, carrying mitochondrial genes, is (i) inverted or (ii) moved to a new position. Both of these events can be mediated by short intergenic repeats. In the case of an inversion only the repeats adjacent to the segment are involved, while in the case of a movement the repeats adjacent as well as those present at the new site are involved in the recombination event. If a segment is inverted, then the coding strand, regarding the whole genome, is changed. On the other hand, if a segment is moved to a new position in the genome (a transposition-like event), the coding strand is either preserved or changed. The orientation of genes is conserved in all the compared yeasts, with the exception of the *C.glabrata* tRNA^{Thr1}. This suggests that there has been a strong pressure to

keep the coding potential concentrated on one strand. Therefore, we propose that the rearrangements have likely occurred exclusively through transposition-like events, and not inversions, using short intergenic repeated sequences as sites of the illegitimate recombination or homologous recombination using *tRNA* genes in such a way that the jumping genes preserved their orientation and stayed on the same coding strand. This mechanism does not rely on the presence of a circular mtDNA form. While inversions seem to have been 'prohibited' through evolution of the *Saccharomyces* yeasts they could have operated in the *C.albicans* lineage. The complete sequencing of the mitochondrial genome of *C.albicans* revealed that approximately one-third of the genes are located on the opposite strand (79). Whereas the native mitochondrial genomes of the *Saccharomyces* yeasts concentrate all genes on a single strand, *S.cerevisiae* mutants having genes on both strands have been reported (22). These mutants are respiratory competent, but are less competitive in genetic crosses (reviewed in 4). An elevated transmission capacity of the mtDNAs may be the reason why the coding potential in *Saccharomyces* yeasts is preserved on one strand and, as a result, transposition is a preferred mechanism to create novel gene orders.

SUPPLEMENTARY MATERIAL

The following features obtained for *S.castellii* and *S.servazzii* are available at NAR Online: a table showing sequence homology between mitochondrial-encoded proteins from various ascomycetous fungi, a table showing the codon usage in the exonic ORFs in the two mitochondrial genomes, a table listing the putative transcription initiation sites and a table listing the putative endonucleolytic cleavage sites.

ACKNOWLEDGEMENTS

We are grateful to A. Malpertuy (Institut Pasteur, Paris) for communicating results before publication. This work has been partially supported by grants from the Danish Research Council, the Novo Nordisk Foundation, the Carlsberg Foundation, the Plasmid Foundation, the Institut National de la Recherche Agronomique, the Centre National de la Recherche Scientifique and the GDR/CNRS 2354 'Génolevures II'.

REFERENCES

- Foury,F., Roganti,T., Lecrenier,N. and Purnelle,B. (1998) The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.*, **440**, 325–331.
- de Zamaroczy,M. and Bernardi,G. (1985) Sequence organization of the mitochondrial genome of yeast—a review. *Gene*, **37**, 1–17.
- de Zamaroczy,M. and Bernardi,G. (1986) The GC clusters of the mitochondrial genome of yeast and their evolutionary origin. *Gene*, **41**, 1–22.
- Piškur,J. (1994) Inheritance of the yeast mitochondrial genome. *Plasmid*, **31**, 229–241.
- Williamson,D. (2002) The curious history of yeast mitochondrial DNA. *Nature Rev. Genet.*, **3**, 475–481.
- Clark-Walker,G.D., Sriprakash,K.S., McArthur,C.R. and Azad,A.A. (1980) Mapping of mitochondrial DNA from *Torulopsis glabrata*. *Curr. Genet.*, **1**, 209–217.
- Hoeben,P. and Clark-Walker,G.D. (1986) An approach to yeast classification by mapping mitochondrial DNA from *Dekkera/Brettanomyces* and *Eeniella* genera. *Curr. Genet.*, **10**, 371–379.
- Sekito,T., Okamoto,K., Kitano,H. and Yoshida,K. (1995) The complete mitochondrial DNA sequence of *Hansenula wingei* reveals new characteristics of yeast mitochondria. *Curr. Genet.*, **28**, 39–53.
- Kerscher,S., Durstewitz,G., Casaregola,S., Gaillardin,C. and Brandt,U. (2001) The complete mitochondrial genome of *Yarrowia lipolytica*. *Comp. Funct. Genom.*, **2**, 80–90.
- Hoeben,P., Weiller,G. and Clark-Walker,G.D. (1993) Larger rearranged mitochondrial genomes in *Dekkera/Brettanomyces* yeasts are more closely related than smaller genomes with a conserved gene order. *J. Mol. Evol.*, **36**, 263–269.
- Hudspeth,M.E., Ainley,W.M., Shumard,D.S., Butow,R.A. and Grossman,L.I. (1982) Location and structure of the var1 gene on yeast mitochondrial DNA: nucleotide sequence of the 40.0 allele. *Cell*, **30**, 617–626.
- Ainley,W.M., Macreadie,I.G. and Butow,R.A. (1985) var1 gene on the mitochondrial genome of *Torulopsis glabrata*. *J. Mol. Biol.*, **184**, 565–576.
- Clark-Walker,G.D., McArthur,C.R. and Sriprakash,K.S. (1985) Location of transcriptional control signals and transfer RNA sequences in *Torulopsis glabrata* mitochondrial DNA. *EMBO J.*, **4**, 465–473.
- Okamoto,K., Sekito,T. and Yoshida,K. (1994) The mitochondrial genome of yeast *Hansenula wingei* encodes NADH dehydrogenase subunit genes ND4L and ND5. *Mol. Gen. Genet.*, **243**, 473–476.
- Lang,B.F. (1993) The mitochondrial genome of *Schizosaccharomyces pombe*. In O'Brien,S.J. (ed.), *Genetic Maps: Locus Maps of Complex Genomes*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 3133–3135.
- Bullerwell,C.E., Leigh,J., Forget,L. and Lang,B.F. (2003) A comparison of three fission yeast mitochondrial genomes. *Nucleic Acids Res.*, **31**, 759–768.
- Génolevures (2000) Genomic exploration of the hemiascomycetous yeasts. *FEBS Lett.*, **487**, 1–149.
- Cummings,D.J., McNally,K.L., Domenico,J.M. and Matsuura,E.T. (1990) The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. *Curr. Genet.*, **17**, 375–402.
- Piškur,J., Smole,S., Groth,C., Petersen,R.F. and Pedersen,M.B. (1998) Structure and genetic stability of mitochondrial genomes vary among yeasts of the genus *Saccharomyces*. *Int. J. Syst. Bacteriol.*, **48**, 1015–1024.
- Kurtzman,C.P. and Robnett,C.J. (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie Van Leeuwenhoek*, **73**, 331–371.
- Groth,C., Petersen,R.F. and Piškur,J. (2000) Diversity in organization and the origin of gene orders in the mitochondrial DNA molecules of the genus *Saccharomyces*. *Mol. Biol. Evol.*, **17**, 1833–1841.
- Clark-Walker,G.D. (1989) In vivo rearrangement of mitochondrial DNA in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **86**, 8847–8851.
- Petersen,R.F., Langkjær,R.B., Hvidtfeldt,J., Gartner,J., Palmen,W., Ussery,D.W. and Piškur,J. (2002) Inheritance and organisation of the mitochondrial genome differ between two *Saccharomyces* yeasts. *J. Mol. Biol.*, **318**, 627–636.
- Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H., Artiguenave,F., Wincker,P. and Gaillardin,C. (2000) Genomic exploration of the hemiascomycetous yeasts: 7. *Saccharomyces servazzii*. *FEBS Lett.*, **487**, 47–51.
- Ewing,B. and Green,P. (1998) Base-calling of automated sequencer traces using *phred*. II. Error probabilities. *Genome Res.*, **8**, 186–194.
- Ewing,B., Hillier,L., Wendl,M.C. and Green,P. (1998) Base-calling of automated sequencer traces using *phred*. I. Accuracy assessment. *Genome Res.*, **8**, 175–185.
- Gordon,D., Abajian,C. and Green,P. (1998) *Consed*: a graphical tool for sequence finishing. *Genome Res.*, **8**, 195–202.
- Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.

30. Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
31. Dujon, B. (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene*, **82**, 91–114.
32. Dujon, B., Belfort, M., Butow, R.A., Jacq, C., Lemieux, C., Perlman, P.S. and Vogt, V.M. (1989) Mobile introns: definition of terms and recommended nomenclature. *Gene*, **82**, 115–118.
33. Burke, J.M., Belfort, M., Cech, T.R., Davies, R.W., Schweyen, R.J., Shub, D.A., Szostak, J.W. and Tabak, H.F. (1987) Structural conventions for group I introns. *Nucleic Acids Res.*, **15**, 7217–7221.
34. Burke, J.M. (1989) Sequences and classification of group I and group II introns. *Methods Enzymol.*, **180**, 533–545.
35. Pedersen, A.G., Jensen, L.J., Stærfeldt, H.-H., Brunak, S. and Ussery, D.W. (2000) A DNA structural atlas for *Escherichia coli*. *J. Mol. Biol.*, **299**, 907–930.
36. Jensen, L.J., Friis, C. and Ussery, D.W. (1999) Three views of microbial genomes. *Res. Microbiol.*, **150**, 773–777.
37. Sanson, G.F. and Briones, M.R. (2000) Typing of *Candida glabrata* in clinical isolates by comparative sequence analysis of the cytochrome c oxidase subunit 2 gene distinguishes two clusters of strains associated with geographical sequence polymorphisms. *J. Clin. Microbiol.*, **38**, 227–235.
38. Clark-Walker, G.D. and Weiller, G.F. (1994) The structure of the small mitochondrial DNA of *Kluyveromyces thermotolerans* is likely to reflect the ancestral gene order in fungi. *J. Mol. Evol.*, **38**, 593–601.
39. Fox, T.D. (1987) Natural variation in the genetic code. *Annu. Rev. Genet.*, **21**, 67–91.
40. Li, M. and Tzagoloff, A. (1979) Assembly of the mitochondrial membrane system: sequences of yeast mitochondrial valine and an unusual threonine tRNA gene. *Cell*, **18**, 47–53.
41. Miller, D.L. and Martin, N.C. (1983) Characterization of the yeast mitochondrial locus necessary for tRNA biosynthesis: DNA sequence analysis and identification of a new transcript. *Cell*, **34**, 911–917.
42. Shu, H.H., Wise, C.A., Clark-Walker, G.D. and Martin, N.C. (1991) A gene required for RNase P activity in *Candida (Torulopsis) glabrata* mitochondria codes for a 227-nucleotide RNA with homology to bacterial RNase P RNA. *Mol. Cell. Biol.*, **11**, 1662–1667.
43. Wise, C.A. and Martin, N.C. (1991) Dramatic size variation of yeast mitochondrial RNAs suggests that RNase P RNAs can be quite small. *J. Biol. Chem.*, **266**, 19154–19157.
44. Wilson, C., Ragnini, A. and Fukuhara, H. (1989) Analysis of the regions coding for transfer RNAs in *Kluyveromyces lactis* mitochondrial DNA. *Nucleic Acids Res.*, **17**, 4485–4491.
45. Sbisà, E., Pesole, G., Tullo, A. and Saccone, C. (1996) The evolution of the RNase P- and RNase MRP-associated RNAs: phylogenetic analysis and nucleotide substitution rate. *J. Mol. Evol.*, **43**, 46–57.
46. Frank, D.N. and Pace, N.R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.*, **67**, 153–180.
47. Biswas, T.K. (1998) Usage of non-canonical promoter sequence by the yeast mitochondrial RNA polymerase. *Gene*, **212**, 305–314.
48. Thalenfeld, B.E., Bonitz, S.G., Nobrega, F.G., Macino, G. and Tzagoloff, A. (1983) oli1 transcripts in wild type and in a cytoplasmic ‘petite’ mutant of yeast. *J. Biol. Chem.*, **258**, 14065–14068.
49. Zassenhaus, H.P., Martin, N.C. and Butow, R.A. (1984) Origins of transcripts of the yeast mitochondrial var1 gene. *J. Biol. Chem.*, **259**, 6019–6027.
50. Osinga, K.A., De Vries, E., Van der Horst, G. and Tabak, H.F. (1984) Processing of yeast mitochondrial messenger RNAs at a conserved dodecamer sequence. *EMBO J.*, **3**, 829–834.
51. Wésolowski-Louvel, M., Breunig, K.D. and Fukuhara, H. (1996) *Kluyveromyces lactis*. In Wolf, K. (ed.), *Nonconventional Yeasts in Biotechnology: A Handbook*. Springer-Verlag, Berlin, pp. 139–201.
52. Grivell, L.A. (1989) Nucleo-mitochondrial interactions in yeast mitochondrial biogenesis. *Eur. J. Biochem.*, **182**, 477–493.
53. Clark-Walker, G.D., McArthur, C.R. and Sriprakash, K.S. (1983) Order and orientation of genic sequences in circular mitochondrial DNA from *Saccharomyces exiguus*: implications for evolution of yeast mtDNAs. *J. Mol. Evol.*, **19**, 333–341.
54. Clark-Walker, G.D. (1985) Basis of diversity in mitochondrial DNAs. In Cavalier-Smith, T. (ed.), *The Evolution of Genome Size*. John Wiley & Sons Ltd, Chichester, pp. 277–297.
55. Clark-Walker, G.D. (1992) Evolution of mitochondrial genomes in fungi. *Int. Rev. Cytol.*, **141**, 89–127.
56. Locker, J. and Rabinowitz, M. (1981) Transcription in yeast mitochondria: analysis of the 21 S rRNA region and its transcripts. *Plasmid*, **6**, 302–314.
57. Palleschi, C., Francisci, S., Zennaro, E. and Frontali, L. (1984) Expression of the clustered mitochondrial tRNA genes in *Saccharomyces cerevisiae*: transcription and processing of transcripts. *EMBO J.*, **3**, 1389–1395.
58. Christianson, T., Edwards, J.C., Mueller, D.M. and Rabinowitz, M. (1983) Identification of a single transcriptional initiation site for the glutamic tRNA and COB genes in yeast mitochondria. *Proc. Natl. Acad. Sci. USA*, **80**, 5564–5568.
59. Dujon, B. (1980) Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. *Cell*, **20**, 185–197.
60. Goddard, M.R. and Burt, A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. USA*, **96**, 13880–13885.
61. Baldacci, G., Chérif-Zahar, B. and Bernardi, G. (1984) The initiation of tRNA replication in the mitochondrial genome of yeast. *EMBO J.*, **3**, 2115–2120.
62. de Zamaroczy, M., Faugeron-Fonty, G., Baldacci, G., Goursot, R. and Bernardi, G. (1984) The *ori* sequences of the mitochondrial genome of a wild-type yeast strain: number, location, orientation and structure. *Gene*, **32**, 439–457.
63. Butow, R.A. and Zinn, A.R. (1986) Mobile elements in the yeast mitochondrial genome. *Basic Life Sci.*, **40**, 29–37.
64. Nosek, J., Tomaska, L., Fukuhara, H., Suyama, Y. and Kovac, L. (1998) Linear mitochondrial genomes: 30 years down the line. *Trends Genet.*, **14**, 184–188.
65. Osawa, S., Jukes, T.H., Watanabe, K. and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol. Rev.*, **56**, 229–264.
66. Tarassov, I.A. and Martin, R.P. (1996) Mechanisms of tRNA import into yeast mitochondria: an overview. *Biochimie*, **78**, 502–510.
67. Inagaki, Y., Kojima, A., Bessho, Y., Hori, H., Ohama, T. and Osawa, S. (1995) Translation of synonymous codons in family boxes by *Mycoplasma capricolum* tRNAs with unmodified uridine or adenosine at the first anticodon position. *J. Mol. Biol.*, **251**, 486–492.
68. Fox, T.D. and Weiss-Brummer, B. (1980) Leaky +1 and –1 frameshift mutations at the same site in a yeast mitochondrial gene. *Nature*, **288**, 60–63.
69. Hüttenhofer, A., Weiss-Brummer, B., Dirheimer, G. and Martin, R.P. (1990) A novel type of +1 frameshift suppressor: a base substitution in the anticodon stem of a yeast mitochondrial serine-tRNA causes frameshift suppression. *EMBO J.*, **9**, 551–558.
70. Farabaugh, P.J. and Björk, G.R. (1999) How translational accuracy influences reading frame maintenance. *EMBO J.*, **18**, 1427–1434.
71. Kawakami, K., Pande, S., Faiola, B., Moore, D.P., Boeke, J.D., Farabaugh, P.J., Strathern, J.N., Nakamura, Y. and Garfinkel, D.J. (1993) A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in *Saccharomyces cerevisiae*. *Genetics*, **135**, 309–320.
72. Brennicke, A., Marchfelder, A. and Binder, S. (1999) RNA editing. *FEMS Microbiol. Rev.*, **23**, 297–316.
73. Belfort, M. and Roberts, R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
74. Cech, T.R. (1990) Self-splicing of group I introns. *Annu. Rev. Biochem.*, **59**, 543–568.
75. Labouesse, M., Netter, P. and Schroeder, R. (1984) Molecular basis of the ‘box effect’, a maturase deficiency leading to the absence of splicing of two introns located in two split genes of yeast mitochondrial DNA. *Eur. J. Biochem.*, **144**, 85–93.
76. Skelly, P.J. and Maleszka, R. (1991) Distribution of mitochondrial intron sequences among 21 yeast species. *Curr. Genet.*, **19**, 89–94.
77. Dujardin, G., Jacq, C. and Slonimski, P.P. (1982) Single base substitution in an intron of oxidase gene compensates splicing defects of the cytochrome b gene. *Nature*, **298**, 628–632.
78. Christianson, T. and Rabinowitz, M. (1983) Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by in vitro capping with guanylyltransferase. *J. Biol. Chem.*, **258**, 14025–14033.
79. Anderson, J.B., Wickens, C., Khan, M., Cowen, L.E., Federspiel, N., Jones, T. and Kohn, L.M. (2001) Infrequent genetic exchange and recombination in the mitochondrial genome of *Candida albicans*. *J. Bacteriol.*, **183**, 865–872.
80. Lambowitz, A.M. and Belfort, M. (1993) Introns as mobile genetic elements. *Annu. Rev. Biochem.*, **62**, 587–622.