Analysis of the regions coding for transfer RNAs in *Kluyveromyces lactis* mitochondrial DNA

Cathal Wilson, Antonella Ragnini and Hiroshi Fukuhara*

Institut Curie, Section de Biologie, Bâtiment 110. Centre Universitaire, Orsay 91405. France

Received April 7, 1989; Accepted May 16, 1989 EMBL accession nos. X14740–X14756, X14821–X14824 (incl.)

**ABSTRACT**

The major regions coding for the transfer RNA genes in the mitochondrial DNA of *K. lactis* were studied. Twenty one, out of a supposed twenty four tRNA genes were identified and localized with respect to other mitochondrial genes. Most of the tRNA genes were found in a cluster downstream of the large ribosomal RNA gene. The order of a few groups of genes is conserved with respect to *S. cerevisiae* and *T. glabrata*. The highly diverged intergenic sequences contained a large number of guanine-cytosine clusters which frequently formed long palindromic sequences.

**INTRODUCTION**

The mitochondrial genome of the yeast *Kluyveromyces lactis* is a circular DNA of 39 kilobase-pairs (kbp). In previous studies we have localized the seven major protein-coding genes and the two ribosomal RNA genes on the restriction site map. The present work focused on the identification and localization of individual transfer RNA (tRNA) genes. In fungal genomes, the arrangement of mitochondrial genes is extremely diverse (1) despite the high degree of conservation of individual gene sequences. To see how the mitochondrial genes have been reshuffled in the course of evolution, a convenient tool is the set of the two dozen tRNA genes present in the yeast mitochondrial DNA (mtDNA). As in the case of *Saccharomyces cerevisiae*, most of the tRNA genes in *K. lactis* mitochondria were found as a large cluster immediately downstream of the large subunit ribosomal RNA gene, the remainder being scattered in a few other regions. We describe here the sequence and the position of 21 tRNA genes, as well as the nature of the intergenic sequences.

**MATERIALS & METHODS**

**Strain.** The strain 2360/7 (alpha, lysA, K−) was the source of mtDNA. It is an auxotrophic subclone isolated by A. Algeri from the strain CBS 2360 (Centraalbureau voor Schimmelcultures, Delft).

**Preparation of mtDNA.** A modification of a previously published method (2) was used: unwashed mitochondria were prepared from the protoplasts, and, without DNase treatment, the mitochondria were dissolved in 1% sodium-dodecylsalcosinate/50 mM Tris-HCl, pH8/5mM EDTA; DNA was extracted with phenol and purified by two cycles of isopycnic centrifugation in CsCl-bisbenzimide solution to remove nuclear DNA contamination.

**Cloning and sequencing.** MtDNA was cleaved by single or combined digestions with restriction enzymes, and cloned into the bacterial plasmid pBR322. Almost the total genome was cloned in 13 fragments. The DNA from each fragment was spotted on nitrocellulose filters and hybridized with 32P-labelled tRNA probes which consisted of various tRNA gene clusters of the *S. cerevisiae* mtDNA (3). Hybridization conditions were 0.6 M NaCl, 60°C, 18 hours. The cloned mtDNA fragments were subcloned into pTZ18R/pTZ19R
RESULTS & DISCUSSION

Localization of tRNA gene regions

Figure 1 is the map of K. lactis mtDNA. In a previous study, hybridization of labelled 4S RNA with mtDNA restriction fragments had suggested that the majority of the tRNA genes should lie near the ribosomal RNA (rRNA) genes. A more refined location was obtained by mixed tRNA gene probes from S. cerevisiae mtDNA: the major tRNA gene cluster was between the 3' end of the 21S rRNA gene and the ATPase subunit 9 gene. We have therefore sequenced about 10,000 base pairs around and between the two rRNA genes. A total of 21 tRNA genes were identified in these regions. 15 were clustered within a 3 kbp segment downstream of the 21S rRNA gene, 4 were in the sequence between the two rRNA genes, and 2 more were found upstream of the 1S rRNA gene. Supposing a decoding pattern analogous to S. cerevisiae mitochondria, at least three tRNA genes still remain to be identified: valyl (anticodon UAC), threonyl 1 (UAG) and arginyl 2 (ACG) tRNAs.

Order of tRNA genes.

The arrangement of mitochondrial tRNA genes in yeast mitochondria has previously been determined in S. cerevisiae (3) and Torulopsis glabrata (Candida glabrata) (4). By comparison with these two species, we can see that K. lactis has conserved the same order
within several groups of tRNA genes, but the positions of other tRNA genes are completely different. The conserved groups are: phe-gln-lys-argl-gly-asp-ser2, tyr-asn, ala-ile and fmet-pro. Although the gene rearrangement seems to involve translocations of genomic segments rather than duplications, we can find the trace of such a duplication event. Within Region 2, a pseudo prolyl tRNA gene was found as a duplication of a part of the gene (3'terminal 21 bases plus 5 flanking bases) at a site 65 bp downstream. The duplicated sequence itself formed a part of a large palindrome.

It has previously been shown (5) that transcription in S.cerevisiae mitochondria often starts at the consensus sequence (A/T)TATAAGTA. The same sequence seems to be used in the transcription of rRNA genes in K.lactis (5) and of tRNA genes in Torulopsis glabrata (4). We found this sequence before a few groups of tRNA genes in K.lactis mtDNA. In all three species, the consensus sequence was found before phenyl tRNA and tyrosyl tRNA genes. It was also found before the cysteinyl tRNA gene in K.lactis and in S.cerevisiae. Possibly transcription units may have evolved from a common organization. However the nucleotide sequences between the consensus nonanucleotide and the downstream tRNA gene were completely divergent, and their length variable. One of the consensus nonanucleotide was found within the 3' end sequence of the methionyl tRNA gene.

In S.cerevisiae (6, 7, 8), and in T.glabrata (4), some protein-coding genes and rRNA genes are followed by a conserved dodecanucleotide sequence TATAATATTCTT, which is thought to be involved in the processing of transcripts from these genes. This sequence has not been found so far in the sequenced regions of K.lactis mtDNA.

Sequence of individual tRNA genes.

The primary structure of individual tRNA genes is highly conserved in S.cerevisiae, T.glabrata and K.lactis (cf. ref.9 for compilation of S.cerevisiae data, and ref.4 for T.glabrata data). For each isofunctional tRNA, the average number of nucleotide changes per gene was 6.5 between K.lactis and S.cerevisiae, 5.1 between S.cerevisiae and T.glabrata, and 8.3 between T.glabrata and K.lactis. Despite these changes, the characteristic deviations of each tRNA species with respect to the standard cloverleaf structure, as depicted by Sibler et al. (10) for S.cerevisiae genes, were also found in K.lactis. For example, there is a T:T pair in the anticodon stem of the leucyl tRNA gene; a bulged T in the TTC stem of the lysyl tRNA gene (this feature is absent from the T.glabrata gene); an A replacing the universal T8, an A replacing the universal Pyg and a T:T pair in the TTC stem of the prolyl tRNA gene. The sequences of four mitochondrial tRNA genes (cys, leu, gln and lys) from K.lactis have recently been reported (19). They agree with our data, except that a few discrepancies were found in the 3' flanking region of glutaminyl tRNA.

Intergenic sequences.

Although the sequencing of the spaces between the tRNA genes is not yet complete, a few typical features of intergenic spaces can be seen in the example shown Figure 2.

One of the characteristics of K.lactis mtDNA is the presence of a large number of guanine-cytosine (GC) rich sequence clusters (11). Many of them contain SacII restriction sites (CCGCGG) which often appear as a pair, because they are engaged in a palindromic sequence. The intergenic spaces were crowded with such palindromes. Several other species of Kluyveromyces also have the same type of mtDNA (many SacII sites, ref. 11). Presence of these repeats and their GC richness raised technical problems in the enzymatic sequencing of this particular mtDNA. Some palindromes tended to cause spontaneous deletions during cloning in E.coli. Therefore, the identity of the cloned fragments was verified by comparing them with the size and restriction sites of the fragments directly obtained from purified.
Region I.

NNNNNNNNNN NNNAGATCTA ATTAATATAT AATNCCNCGG ATTATCATAA

GAATGATAA CGCGGGGATAC ATTAATATAT TTTCCACNCG GGTCATCCAA 100

CCCTATGCGT GGACCCCGGT GGGGTTTAA TAGATTAAAT TTTATATTAT

AATATTATTA ATATATATAT ATATATTATTA TAATTTTGAT TCTCCCATATA 200

GACATCGGAA TACTTTTTNN NNNNTTATCC CTGGAATTAT CAGAGATAAA

GTCCCGGCGG TTTATTAAAT TTTAAAATAT AAATATATAG GTAATTATAT 300

AGGTATATAG TAAATTATAC TCCAAATTAC ATGTTGAATT CCCCCTCCCTA

CGCACTCGT GTTGCGGATA GGGAGGAATA ATGGTAATAA TCCNNNNCCA 400

TCACNNNACAT GTATGCGTGG TGTTAGCAGC GTTTAGCTTTA TATAACTAAT

ATATATATT ATAAATTTAA ATATTATAC CCGCGGAAT TCCGGATTTT 500

ATGGCTAAANN NNNNNTAAAT AAATATATAA TAAATAAATA TAAATATTAT 600

ATATATATT TAATATAAAT TATATTATTT ATAAATATTAA ATAAATTTA

CCCCGCCGCT CATCACCCCTA TATTGJAGG GTGAGAGACC GGGGGGGGAA 700

AAATATAAAA ATAAATTTGC ATCTCACCTT GTCGTAGTA GAAAAAATATI

TATAATAGAA TATTATGgcc tttataagctt agttagaaag caataaattg

aagtttagt tacattgagt tcatgtcctaa tttaagacgat ATAGAGTACTT 800

tiggtgaaag tggtagaaca gcaccaacctt aaagttggttt tgcgaagata 900

tgaagggttca aacctttctaa gttagaatAT ATATATATAT ATagatcgta 1000

gtaaatatg taaagtccac aaacctttggt tggtaagtgg tgggtaagca

tcaacccgtg tcataatAT TCCGGAGGCC CATTTGCGGA CCCCCCTAGA

GGTTNNNNNGA GTGCGGGGGT AAAATTATAT ATTAATAGTAG TACGTCGCTC

TTATTTATAA GAAAGATATAA CCGCGGAGGA gagaattttgt ttaatgatta 1100

aaacagttcg ctttttttgaacctt ggtgtcagtt ccacagtctct

tatATATATT ATAAATAAAAT TATAAAAAAT TATATAATAAT ATAAATCCTT 1200

CCCCCGGGCT CATCGCGGCT CCTCTATGAT GGGGGAGGAT

ATTATATATAA TATAATACTC CGGGGGATCA TATAATATT Agcttctcta

gcttaaatgtt taagaacctaa tactcttaat ataaagttcatt tgtggcagc 1300

tcataagag aagtaATAT TATAAATAAT ATataagttt aagtaaatgg

gtaaactctt gttgataaat gcaatgtcag tgggaagtttc tttggtagg

tataATATAT ATAAATTATATA ATAAATATAA ATAAATATTAT AATTATACCTT 1500

AATATATTGGA GATAAAATATT ACCCGGGCG CATCGCGGCT CATTTGCGGA

GGGATGAGTC CGGAAATTTG GGAATATAT AATAATATTAT ATTATATTAA 1600

TTATATATATA TATAATATAA ATAAATATAT ATAAATATAA ATAAATATAA

AAATGCGGCT TATATATATT TGATAATCC TGTACTTATT TTTTTTAAA 1700

GGATACACAG TTTATTTAAA ATTTTTTAAAT gtagttatct gtagcataat agtaaagttcc

cactctgctca cagatgtgga tggatgcagca aagcttctttt gttctttcta 1800

Aggaagattt actatggattta atgtagatatt tttgctcataat attgaacta

tgcttttagt atgctgatct actatgatttt ggtttcatttt TCTATATTTT

TATAATATAT ATAGTTGgaa tataacattaa aagtgtagaaa aatgttttgtg 2000

gcggcttaaa tcttgagtt aacctgctta tctaccccttt GCCGGGGGAA

CGACATCGCT ATAGGGATAG TGATTGGG GCTTATATAC ATTTATATAC

ACCCCGGCCA CCCCCCGCA GGGAGGTGACC GGGAGGAATT ACTTTAAAAA

TATTAGTTAA TTTTTATAAT ATTTAAATA ATATAATAC tttttaagttt

taatggttaa aaacctctgt tctataagctg aagtatatag ggtcaagttc

cattatang ATTTAAAAAT TATAATATAT TAAAGAATAT ATATATTTAT

TGCAATCTT TTTTTATAAT TGAGTTGATAA AAGTGATAAA TATAATATAA

GATTGATGTA ATAAATACAC GTTAGTTTTTG CGTATTATTA TAAAATATAC

CCCGCGGACCT CATCATTCTA CTTAAAGTAGT AATGAGAGGC GCGGAGATAA

TAAATTTTCC CCCCCGCTCT ACCGTTACA CCGGTTAGAC TGGGGGGGGA

AATATAATT ATATATATAT TATATATATA AAAAAATATAAAAAATATATAT

ATATATATCC ACTATATACA TATAATATT TTAGAAAGAG GGAATATTT

TTATATATT ATTTATAATA ATATAATAAT ATATAATTTAC TGGGAAATCG

ACACACCCAA CCGGAGGGTT GGGCGGATGC GGCGAGGGGT AATATATAT 2700
<table>
<thead>
<tr>
<th>Region</th>
<th>Nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>ATATATATAA TATAATTAAA TATATTTCTC ATTTCTTATG AATATTAGA GAAGGCTGTA TATATAATT</td>
</tr>
<tr>
<td></td>
<td>TAGGGTGAATA CCGGGGGGCT TATATAATAT TTAGATTATG AATATTAGA</td>
</tr>
<tr>
<td></td>
<td>ATGGGAAATT TAAATTTATC TATATAATTAA TATATATTAA</td>
</tr>
<tr>
<td></td>
<td>TAGATATGCA TATATATGCA TATATCCTCA CAGTCCGAT</td>
</tr>
<tr>
<td></td>
<td>GCTGTCATTAC TTAGTTATTT GAAGCTTCTT CAGATTATAT</td>
</tr>
<tr>
<td></td>
<td>CAGATTGCT TACGTTGTTT TATAGGAAAGG TATATAATAG</td>
</tr>
<tr>
<td></td>
<td>ATGATTATTAT CTTCTCTCCT GCCCGCGGGA ACTCCGAGC AAGATGATT</td>
</tr>
<tr>
<td></td>
<td>ATGATATTTT TATATATTAT AATATAGTATA AATACGTTTG</td>
</tr>
<tr>
<td></td>
<td>ATGATTATTA CACCGCGGGA GCGGATAGTA TAATATATG</td>
</tr>
<tr>
<td></td>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td></td>
<td>ATGATTATTT TATATATTAT AATATAGTATA AATACGTTTG</td>
</tr>
<tr>
<td></td>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td></td>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td></td>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
</tbody>
</table>

**Region II.**

(15S rRNA) - TTTTTTTATT CTACCAATT TATATGGAAT TTTTTTTATT

<table>
<thead>
<tr>
<th>Nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATATCCACCA ACCCATATAC GCTCCACCT CCGGGCAAC ACATGGTGG</td>
</tr>
<tr>
<td>CAGATAATGAA GGTAGAAAGG TGGGTAGGA ATTAATATAA ATATAATATAA</td>
</tr>
<tr>
<td>TTATATATAT CCGGGGAGG TCCAGGAGG GCGGATAGTAC</td>
</tr>
<tr>
<td>CAGATTGCTA TTTCTTCTT TATAGGAAAGG ATATATATAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>TTGGGGCTT CTATTTTTT TATAGGAAAGG ATATATATAT</td>
</tr>
<tr>
<td>GGAATGATCC GGAATATATA ATATATATATA ATATATATATA</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>GGAATGATTT ATGATATTT ATGATATTTA ATGATATTTA</td>
</tr>
<tr>
<td>AAGGGCTGTC CGCAGGCGGAG TTTCTCTTCT GTAAGCTCTT TTAAGGTTAT</td>
</tr>
<tr>
<td>TAGCCGGGGA GTTGTATATAA TATAAAATAT AATAAAATAT</td>
</tr>
<tr>
<td>ATTATATATAT TTTCTTCTT ATTTGATGAAA GAAATATAG</td>
</tr>
<tr>
<td>ATTATATATAT TTTCTTCTT ATTTGATGAAA GAAATATAG</td>
</tr>
</tbody>
</table>

**Region III.**

(35S rRNA) - ATGTTTTTA TATATTTCTC ATTTCTTATG AATATTAGA GAAGGCTGTA TATATAATT

<table>
<thead>
<tr>
<th>Nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
<tr>
<td>TAGCCGGGGA GTTGTATATAA TATAAAATAT AATAAAATAT</td>
</tr>
<tr>
<td>ATTATATATAT TTTCTTCTT ATTTGATGAAA GAAATATAG</td>
</tr>
<tr>
<td>ATTATATATAT TTTCTTCTT ATTTGATGAAA GAAATATAG</td>
</tr>
<tr>
<td>ATGATTATTT TATATATTAT ATGATTATTA TATATATTAT</td>
</tr>
</tbody>
</table>
Nucleic Acids Research

```
TATATAATAT ATATTTGATA TATATAATA TAAATATT AATAATAAA 1400
TATATAATAT ATATTTGATA TATATAAAAT TTATTATTAAT 1500 pro
TATACAGTAT aggcggaggtt gttacggtct tccttgaga gaagaagtta 1510 AsuII
GTATATATGAA atctatccaa tctgATATAA TAAAAATAAT AAAATAATAA 1600 AsuII
TATATACAC ATGGGTTAAA TCTTAAAAT ATATAAAAA 1700 pro'
```

Region III.

```
ACCCCCCGCG GCACACCCAA CTCTACGGTT GGGTGAGGCC GCGGGGGAAA 100 HaeIII
TCTATATATA AGTTAATAGTA gcttacaaat tgcctatgtg tcctgcttta 100 thr2
GTTTGTAT gtaagggttt gacgttggaa tcctaatat aacagacctt 100 AsuII
```

```
T.ACTATACAC ATTGGTAAA TTCTTAATT ATATAATCC CAATTCCGGC 200 glu
TATATATAAT ATATAATTAT ATTATATATA TATATATTCC CCAATGTGCA 700
```

Figure 2. Nucleotide sequence of the major tRNA regions. Lower case letters: tRNA sequence as specified in the right margin; partial duplication of the prolyl tRNA (region II) is indicated by (pro'); as threonyl and glutamyl tRNA genes (region III) are contiguous, the beginning of the latter is indicated by a dot. Thick letters: major restriction sites; N: undetermined or uncertain bases. Typical palindromes (some containing a few imperfections) are underlined. The putative promoter element TTATAAGTA is boxed.

```
mRNA. The sequencing of the unsolved parts may be facilitated by cloning them in recBC strains and by the use of Taq polymerase.
```

While *T.glabrata* mtDNA appears to be devoid of intergenic GC clusters (4), *S.cerevisiae* mtDNA is well known for the presence of many GC clusters (characteristically *HpaII/HaeIII* site clusters) (12). The function of these GC clusters is not understood (see a discussion in ref. 13). In *S.cerevisiae*, recombination and rho~ deletion often occur between these sequences (14, 15). Although rho~ type deletions are not known in *K.lactis*, the dissemination of GC clusters is a major source of sequence polymorphism of *K.lactis* mitochondrial genome. Various *K.lactis* strains analyzed so far all showed several changes in the SacII digestion pattern of mtDNA (an example is in ref.11).

RNAse P RNA-like sequence.

Other parts of the intergenic sequences are extremely rich in adenine and thymine (AT) bases, as in the case of many fungal mtDNAs. All of these AT sequences are highly divergent among the above mentioned yeast species. In *S.cerevisiae*, the AT rich region between f-methionyl and prolyl tRNA genes is known as the tRNA synthesis locus (TSl)

```
T.g. 5’(24) TATATAAAGAAAAAGTCATAAATA(161) ATACATAAATTAAGCTTATATATAGT(25)3’
S.c. 5’(39) TATATAAAGAAAAAGTCATAAATA(343) ATACATAAATTAAGCTTATATATAGT(23)3’
K.I. 5’(27) TATATAAAGAAAAAGTCATAAATA(117) ATACATAAATTAAGCTTATATATAGT(3)3’
E.c. 5’(54) GGAAGGGGAGAAAAAGTCGGGCTCT(267) CGACAGAACCCGGCTTACAGACGT(8)3’
B.s. 5’(36) TCTGTAAGAAAAAGTCGGGCTCT(307) GAACAAACACGCTTACAGACGT(10)3’
M.m. 5’(59) GGCCTAGGAAGAAAAAGTCGGGCTCT(160) CAACACACGGGCGCTTACAGACGT(8)3’
```

Figure 3. Analysis of a TSl locus-like sequence.

The sequence between f-methionyl and prolyl tRNA genes of *K.lactis* (K.I) mtDNA (region II) was compared to the corresponding region of *T.glabrata* (T.g.) and the 9S RNA coding sequence of the TSl locus of *S.cerevisiae* (S.c.) (16). *E.coli* (E.c.) and *B.subtilis* (B.s.) RNAse P RNA sequences were taken from ref.17, and Mouse (M.m.) nuclear-coded MRP-RNA gene sequence from ref. 18. In parentheses are the number of bases. * means identity of bases. Thick letters indicate the core of the conservative sequences.
which encodes a 9S RNA required for the 5' end processing of mitochondrial tRNA precursors (16). Within this particular region, a local similarity of sequence between *T. glabrata* and *S. cerevisiae* had been noticed by Clark-Walker et al. (4). Inspection of the corresponding region in *K. lactis* revealed the presence of two short segments, one near the 5' end and the other near the 3' end, which were common to all the three species, as shown in Figure 3. The core part of both segments coincided with the highly conserved sequences in the similar positions of known RNAse-P RNAs (17, 18). Although the similarity is probably significant, it remains to be shown whether this region in *K. lactis* and in *T. glabrata* represents a functional gene, considering its small size, too short (191 and 258 bp, respectively) to code for a 9S type RNA. Another possibility is that it is a vestige of a gene which might have been taken over by a nuclear function. Such a gene exists for mouse mitochondrial MRP-RNA (18).

ACKNOWLEDGEMENTS

We thank Dr. Claudio Palleschi for the gift of several tRNA gene probes, and Dr. Nancy Martin for a probe for the TSS locus of *S. cerevisiae*. This work was supported in part by the European Communities (BAP-0026F) and the Centre National de la Recherche Scientifique (AIP 6931).

*To whom correspondence should be addressed*

REFERENCES