Mitochondrial DNA of the yeast *Kluyveromyces*: guanine–cytosine rich sequence clusters

Antonella Ragnini and Hiroshi Fukuhara*

Institut Curie, Section de Biologie, Bâtiment 110, Centre Universitaire, Orsay 91406, France

Received June 8, 1988; Revised and Accepted July 29, 1988

Accession nos X12467–X12483 (incl.)

ABSTRACT

Mitochondrial DNA from the yeast *Kluyveromyces marxianus* var. *lactis* (*K.lactis*) is a circular molecule of 39 kilobase-pairs. A genetic and physical map was constructed. We found that this genome contained a large number of guanine–cytosine(GC)-rich sequence clusters, many of which are characterized by the presence of SacII restriction sites (CCGCGG). The primary sequence of the GC clusters often showed a palindromic structure. These GC clusters were present in several varieties of *K.marxianus*, but not in others. The presence of these clusters is a major feature that distinguishes *K.lactis* strains from those of *K.marxianus* var. *marxianus* (including *K.fragilis*).

INTRODUCTION

The mitochondrial DNA (mtDNA) of certain fungal species contains characteristic guanine–cytosine (GC)-rich clusters composed of short repetitive sequences. We have found that a new kind of such sequences was present in the mtDNA from all the independent isolates of *Kluyveromyces marxianus* var. *lactis* (*K.lactis*). Analysis of various strains belonging to the species *Kluyveromyces marxianus* showed that the presence of these GC clusters was a characteristic feature associated with some varieties of *K.marxianus* previously defined on the basis of other taxonomic criteria.

MATERIALS & METHODS

Strains. Our standard strain of *Kluyveromyces marxianus* var. *lactis* (called *K.lactis* in the text) used for the detailed analysis of mtDNA was the auxotrophic mutant 2360/7 (α lysA, K⁻), which had been isolated by Dr.A.Algeri (Institute of Genetics, University of Parma) by UV treatment of CBS 2360. Other strains come from the Centraalbureau voor Schimmelcultures (Delft, The Netherlands) and the American Type Culture Collection (Rockville, Maryland, USA).

DNA analysis. Construction of restriction maps and localization of genes followed previously described procedures (1). Known Saccharomyces cerevisiae mtDNA gene fragments were used as gene probes. K.lactis mtDNA fragments were cloned into the plasmid pBR322 before detailed analysis. DNA sequence analysis was carried out by several variations of the dideoxy chain termination method using commercial kit products; the GC-rich sequences often showed compressed gel electrophoresis patterns and had to be analyzed by several different procedures.

RESULTS

Physical and genetic structure of K.lactis mtDNA.

A preliminary physical map of this mtDNA has been reported previously (1). Using DNA fragments cloned into bacterial plasmids, a more complete description of genes and restriction sites has been obtained (Figure 1). The five strains of K.lactis of independent origin showed almost identical restriction patterns except for a few differences for the enzymes that cut at many sites.

It is known that the gene order of yeast mitochondrial genes differs considerably from one species to another, although some local similarities can be recognized. This is also the case for K.lactis. Hybridization experiments using the gene probes from Saccharomyces cerevisiae mtDNA showed the presence of all the equivalent genes in K.lactis: large and small ribosomal RNA genes separated by about 5 kbp, cytochrome-oxidase subunit 1, 2 and 3, ATPase subunits 6, 8 and 9, as well as cytochrome b. Most of the transfer RNA genes are localized near the two ribosomal RNA genes.

K.lactis mitochondrial DNA has an unexpectedly high number of SacII sites.

The mapping studies showed that K.lactis mtDNA is a circular molecule of 39 kilobase-pairs (kbp). Like most yeast mt DNA, it has a very low GC content. Judged from the buoyant density and the melting temperature, its GC content is less than 20 %. Therefore, any restriction sites composed of 6 G/C bases should appear statistically less than once in the molecule. However, digestion of K.lactis mtDNA by the restriction enzyme SacII or SstII (target sequence CCGCGG) showed at least 30 visible fragments by agarose gel electrophoresis. Figure 2 (left) illustrates such analyses of the mtDNA from several K.lactis strains of independent origin, as well as the strains of other varieties of K.marxianus. All K. lactis mtDNAs showed closely related patterns of SacII digestion.

Figure 1. Restriction sites and gene map of *K.lactis* mtDNA. Unique restriction sites are indicated at the top. The XhoI site was chosen as the map origin. The map is circular. In the lower part, the sequences homologous to *S. cerevisiae* genes are indicated by thick bars. Cox 1, 2 and 3 are the genes for cytochrome-oxidase subunits, Cytb for cytochrome b, ATP6, 8 and 9 for ATPase subunits, 15S and 21S rRNA for small and large subunit ribosomal RNAs. Most tRNA (4S RNA) genes were found between the two rRNA genes and between 21S rRNA and ATP9, but a few were found also in other regions. At the bottom, the number of SacII sites determined on each cloned fragment is indicated.

The presence of many SacII sites suggested that this mtDNA contained repetitive stretches of GC rich sequences. HpaII (CCGG) digestion also produced a relatively large number of fragments (Figure 2, right), though less than SacII.

GC rich sequence clusters.

MtDNA sequences of the strain 2360/7 has been cloned into the bacterial

Figure 2. SacII and HpaII digestion patterns of mtDNA. MtDNA was purified from various *Kluyveromyces* strains, digested with SacII (left part) or HpaII (right) and electrophoresed in 1% agarose gel. DNA bands were detected by ethidium bromide staining.

plasmid pBR322. In each cloned segment, the number of SacII sites was determined. The total number of SacII sites was thus estimated to be at least 60, ignoring those placed very close to each other. These sites are

Table 1. List of typical palindromic GC cluster sequences of *K.lactis* mitochondrial DNA.

Cat.no.		
KLMGC1	C GC CATTAGTCATACGACTAATG GC G	(26)
KLMGC2	CC GC CATATTGCCAATATTGGCAATATATG GC GG	(32)
KLMGC3	CC GC CATAATATATCCCTAAGGGATATATTATG GC GG	(37)
KLMGC4	CCC GC CATATATACCTAATAGGTATATATG GC GGG	(35)
KLMGC5	CCC GC CATTATTACCAATATTGGTAATAATG GC GGG	(36)
KLMGC6	CCC GC GAAATTATTATAATTTC GC GGG	(28)
KLMGC7	CCC GC CCCCATATATAATTGAATATGGGG GC GGG	(34)
KLMGC8	CCC GC GAACTCCCGATAATATCGGGAGTTC GC GGG	(35)
KLMGC9	CCC GC GCGGGTTCATCCACCCCTATATGGGGTAGATGACCCATTAGC GC GGG	(53)
KLMGC10	CCCC GC GGACTTCTCTCCTGTAATACCTTTTAAGGGTATTAGCC GC GGGG	(51) *
KLMGC11	CCCC GC GGTTCATCACCTACATTGCTTGGGTGAGAGACC GC GGGG	(46) *
KLMGC12	CCCC GC GACATGACCCCTCATAAAGGGTCATGTC GC GGGGG	(41)
KLMGC13	CCCCC GC GGCACACCAACCCTACGGTTGGGTGAGGCC GC GGGGGG	(47) *
General form: (C) _n GC (G/C).....(G/C) GC (G) _n		
Variants and other palindromes:		
KLMGC14	CCCGTGGCATCACCCGACCTCATAAAGGGTCGGGTGATCCGCGGG	(46)
KLMGC15	GCCCGCGGAATACCCCGAGATTGGGGTATTGCGCGGC	(39)
KLMGC16	GGGAGGGGGTAGCATAGAATGGATGGAATCATATATAGTAATACTAAATATATGATTTTC	(85)
KLMGC17	CATCCATTCTATGCTACCCCTCCC	(85)
KLMGC17	ATAGGATGGATTACGAACCTTGTTCTATGGAACCGACGTTCGAATCTATCTAT	(56) **

The 5' end tends to accumulate C stretches and the 3' end G stretches.

*: Contains a pair of SacII sites and the palindrome is internally imperfect (bulges). **: An example of large non-GC palindromes. In parentheses: number of nucleotides.

scattered over the total genome as shown in the lower part of Figure 1.

Nucleotide sequencing revealed that there were in fact many GC clusters, and the structures of some of them are shown in Table 1. In these sequences, the SacII sites were often found in pairs, making part of a palindromic structure. There are also many GC clusters that did not contain SacII sites. In these palindromes, with a potential loop and stem structure, the GC clusters occur mostly in the stem part forming the beginning and the end of the palindrome. The 5' end of the palindromes forms a C stretch of variable length, and the 3' end a G stretch. The central part of the palindrome is highly variable in sequence and in length. Most GC clusters belonged to this type of structure, but variant forms (often containing CCGCA in the stem in place of SacII sites), as well as very different kinds of GC rich palindromes were also observed. Some of them can be as large as 88 bases long. A few examples of the palindromic structures are illustrated in Figure 3.

Figure 3. Examples of palindromic GC clusters. A palindromic GC cluster containing SacII sites (a), a variant (b) and an unrelated type (c) are shown in folded forms.

SacII type GC clusters are associated with the mtDNA of a limited group of
yeast species.

The presence of characteristic repetitive sequences is indicative of evolutionary relatedness between genomes. We examined typical strains of all Kluyveromyces species for the presence or absence of the SacII-type GC clusters in mtDNA. The classification of Kluyveromyces has been changing in recent years. In particular, several previously defined species have been incorporated into K.marxianus on the basis of cellular hybridization experiments of Johannsen (2). All the species of Kluyveromyces as classified in the 1984 edition of The Yeast by Kreger-van Rij (3) are included in Table 2. Previous K.lactis is thus considered as a variety of K.marxianus.

MtDNA was extracted from each strain and SacII and HpaII restriction patterns were examined. Some examples are included in Figure 2. GC clusters were detected in four varieties of Cluyveromyces marxianus (var. lactis, drosophilae, vanudenii and dobzhanskii), but not in other varieties (var.

Table 2. Occurrence of SacII type GC clusters in various Kluyveromyces species.

Species §	Strains	Mt DNA	
		SacII	HpaII
<u>K.marxianus</u> var. <u>marxianus</u>	CBS 712 (T)	0	5
	CBS 397	2	3
	CBS 1553	2	3
	ATCC 12424	2	3
	ATCC 36907	0	5
<u>K.marxianus</u> var. <u>bulgaricus</u>	CBS 2762 (T)	2	3
	CBS 4857	0	5
	CBS 5671	0	5
<u>K.marxianus</u> var. <u>lactis</u>	CBS 683 (T)	+++	+++
	CBS 2359	+++	+++
	CBS 2360	+++	+++
	CBS 4574	+++	+++
	CBS 141	+++	+++
<u>K.marxianus</u> var. <u>vanudenii</u>	CBS 4372 (T)	+++	+++
<u>K.marxianus</u> var. <u>dobzhanskii</u>	CBS 2104 (T)	+++	+++
<u>K.marxianus</u> var. <u>drosophilae</u>	CBS 2105 (T)	+++	+++
	CBS 2896	+++	+++
	CBS 2103	+++	+++
<u>K.wickerhamii</u>	CBS 2745 (T)	+++	+++
<u>K.thermotolerans</u>	CBS 6340 (T)	(0)	5
	CBS 137	(0)	5
	CBS 2803	(0)	5
<u>K.waltii</u>	CBS 6430 (T)	(1)	5
<u>K.polysporus</u>	CBS 2163 (T)	(0)	a few
<u>K.phaffii</u>	CBS 4417 (T)	(0)	a few
<u>K.delphensis</u>	CBS 2170 (T)	(0)	+++
<u>K.blattae</u>	CBS 6284 (T)	(0)	a few
<u>K.lodderae</u>	CBS 2757 (T)	(1)	+++
<u>K.africanus</u>	CBS 2517 (T)	(0)	+++
<u>K.aestuarii</u>	CBS 4438 (T)	(1)	+++
<u>K.cellobiovorus</u>	CBS 7153	(0)	a few
<u>K.yarrowii</u>	CBS 6070	(0)	a few

§ Strains are classified according to van der Walt and Johannsen (3). T means type strain. K.cellobiovorus and K.yarrowii are not mentioned in ref.3. +++ means there are many sites as in the case shown for 2360/7. Uncertain numbers of sites (0-1) are in parentheses.

marxianus and bulgaricus). The striking difference between these two groups of varieties is in agreement with the results of the zymogramme comparison experiments (4) which point to the same conclusion. Besides K.marxianus, the only other species that showed the presence of the SacII type GC cluster was K.wickerhamii. Other species showed none or a very few SacII sites, although some contained many HpaII sites.

We noticed in this analysis that some strains (CBS 397, ATCC 12424) of K.marxianus var. marxianus had a mtDNA restriction pattern similar to the type strain CBS 2762 of K.marxianus var. bulgaricus, while some strains (CBS 4857 and CBS 5671) of K.marxianus var. bulgaricus had a mtDNA closely resembling that of K.marxianus var. marxianus strains (type strain CBS 712 and ATCC 36907). Therefore, it appears that the varieties marxianus and bulgaricus are very close to each other and can be clearly distinguished from other varieties of K.marxianus by the absence of SacII type GC clusters. The close relationship between var. marxianus and var. bulgaricus has now been confirmed by OFAGE chromosome banding analysis (Sor and Fukuhara, in preparation). The chromosome patterns also confirmed that the varieties lactis, drosophilae and vanudenii are closely related with each other, and clearly distinct from the varieties marxianus and bulgaricus.

DISCUSSION

GC clusters of mtDNA are one of the intriguing features of various organelle DNA in lower eukaryotes. For Saccharomyces cerevisiae, Prunell and Bernardi (5) were the first to describe the presence of many GC clusters in mtDNA. They were detected by an abnormally high frequency of HpaII/HaeIII restriction sites in this genome. In Neurospora crassa mtDNA, palindromic GC clusters characterized by the presence of PstI sites have also been described by Yin et al. (6). The function of these particular sequences with repetitive motifs has not been established. Various hypotheses have been proposed: sites of intramolecular recombination, duplication of replication origin, rho⁻ excision sites, vestige of transposed elements, etc. (see review in ref. 7), but convincing interpretations will need more mutational studies. In S.cerevisiae, the GC clusters were mostly found in the intergenic spacer regions, and can be transcribed as a part of polygenic transcripts which are then processed. Several cases are known where the GC clusters were integrated within coding sequences, hence producing modified forms of functional products (8, 9). For K.lactis, such cases have not been observed so far, the number of studied sequences being limited. In S.cerevisiae, there is also some evidence that the GC clusters play a role in RNA processing: specific adenine-thymine-oligonucleotide motifs seem to be preferentially used as RNA processing signals when they are placed close to the GC clusters (10).

We note that some of the GC clusters of S. cerevisiae, typically the a2 and v families as classified by de Zamaroczy and Bernardi (7), contain SacII

sites and multi C/multi G ends, suggesting that they may have emerged from related origins. The general repetitive motifs in GC clusters of these two genomes are however different as shown by the relative frequencies of specific GC rich restriction sites.

The distribution of the GC clusters among Kluyveromyces species is of particular interest in yeast classification. The fact that the presence of GC clusters can distinguish particular groups of yeast in accordance with the current classification which had been made on independent taxonomic criteria suggests that GC cluster analysis of mtDNA may be a useful key to assess the relatedness of species. The significance of mtDNA analysis in taxonomy has been emphasized recently by Hoeben and Clark-Walker (11). Besides the restriction sites and gene order analyses, the presence or absence of repetitive sequence motifs allows a quite different type of comparison, independent of small changes at the nucleotide level. Presence of characteristic repetitive sequences in a group of species may be an indication that either these sequences existed before the divergence of these species, or these clusters were transferred horizontally among hybridizable species. K.marxianus var. drosophilum, strain CBS 2105, contains a circular plasmid, pKD1 (12). We found that the host specificity of this plasmid was limited also to the above mentioned varieties characterized by the presence of GC-clusters in their mtDNA (Chen and Fukuhara, in preparation).

ACKNOWLEDGEMENTS

This work was supported in part by the Centre National de la Recherche Scientifique, Institut Curie and the Commission of the European Communities (BAP-0026F). A.R. has received a fellowship from the Fondation Cenci-Bolognetti-Institut Pasteur.

*To whom correspondence should be addressed

REFERENCES

1. Wésolowski, M., Algeri, A. and Fukuhara, H. (1981) *Current Genetics* 3, 157-162.
2. Johannsen, E. (1980) *Antonie van Leeuwenhoek* 46, 177-189.
3. van der Walt, J.P. and Johannsen, E. (1984) in *Yeast*, Kreger-van Rij, N.J.W., Ed., pp.224-251. Elsevier Science Publishers. Amsterdam.
4. Sidenberg, D. and Lachance M.-A. (1983) *Int.J. Systematic Bacteriol.* 33, 822-828.
5. Prunell, A., and Bernardi, G. (1977) *J.Mol.Biol.* 110, 17-52.
6. Yin, S., Heckman, J. and RajBhandary, U.L. (1981) *Cell* 26, 325-332.

7. de Zamaroczy, M. and Bernardi, G. (1986) *Gene* 41, 1-22.
8. Sor, F. and Fukuhara, H. (1982) *Nucl. Acids Res.* 11, 1625-1633.
9. Hudspeth, M.E.S., Ainley, W.M., Shumard, D.S. and Butow, R.A. and Grossman, L.I. (1982) *Cell* 30, 617-626.
10. Francisci, S., Palleschi, C. Ragnini, A. and Frontali, L. (1987) *Nucl. Acids Res.* 15, 6387-6403.
11. Hoeben, P. and Clark-Walker, G.D. (1986) *Current Genetics* 10, 371-379.
12. Falcone, C., Saliola, M., Chen, X.J., Bianchi, M.M., Frontali, L. and Fukuhara, H. (1985) *Plasmid* 15, 248-252.