Genome Instability and the Selfish DNA Issue*

G. BERNARDI

Laboratory of Molecular Genetics, Jacques Monod Institute, 75005 Paris

Both prokaryotes and eukaryotes contain repeated DNA sequences in their genomes. The relative amounts of such sequences, however, are very different. In the case of prokaryotes, a small number of families of insertion sequences form about 1—2% of the genome. In eukaryotes, families of interspersed repeated sequences generally represent 20—30% of the genome. It has been suggested that such repeated sequences play a regulatory role, or are essential to chromosome structure, pairing, rearrangements etc... An alternative view is that these sequences have no phenotypic or evolutionary function, their only function being their survival within the genome (9). According to Orgel and Crick (17) "a piece of selfish DNA, in its purest form, has two distinct properties: 1) it arises when a DNA sequence spreads by forming additional copies of itself within the genome; 2) it makes no specific contribution to the phenotype".

A preliminary remark should be made at the outset of the present discussion, namely that the two views just presented should be considered to represent two extremes, which are, however, not mutually exclusive. It is conceivable that some repeated sequences play a biological role, while others do not, and a discussion of the issue should ideally lead to an estimate of the relative amount of repeated sequences playing a biological role. Such a discussion will be difficult, if not impossible, if we do not know enough about the genomes under consideration. Unfortunately, this is the case for the nuclear genome of eukaryotes, which has been at the center of the debate. Here, I would like to discuss the selfish DNA issue using as a model the mitochondrial genome of yeast, which is almost completely sequenced and is well-known from a genetic point of view. This system is of special interest because it shares the basic sequence features of the nuclear genome of eukaryotes, namely an interspersion of unique and repeated sequences.

The organization of the mitochondrial genome of yeast

Investigations carried out in our laboratory between 1966 and 1976 [see Bernardi (3), for a brief review] showed that over 50% of the 25 (or 50) mitochondrial genome units present in every haploid (or diploid) wild-type *Saccharomyces cerevisiae* cell is made up of: (a) long AT spacers (GC $< 5\%$) which are formed by short dAT : dAT and dA : dT sequences with dG : dC base pairs occurring rarely, and which are internally repetitive in sequence and rich in palindromes (5), and (b) about 200 short GC clusters (GC $> 60\%$)

* Presented at a symposium "The Past, Present and Future of Genetics" held at Brno, Czechoslovakia, August 26—September 1, 1982, in honour of Gregor Mendel. The symposium was organized by the Institute of Molecular Genetics, Czechoslovak Academy of Sciences, Prague and the Moravian Museum, Brno, and sponsored by the International Genetics Federation with the assistance of UNESCO, Paris and the Czechoslovak National Commission for UNESCO.

characterized by sequences that are often palindromic and largely homologous to each other. These GC clusters are embedded in AT spacers (7, 11). AT spacers and GC clusters form the intergenic sequences of the mitochondrial genome and the closed reading frames of the intervening sequences of the cob and oxi 3 genes (Fig. 1). Like the nuclear genome of eukaryotes, the mitochondrial genome of yeast contains: (a) unique sequences, namely the genes for the rRNAs, the tRNAs and the mRNAs for polypeptide sub-units of respiratory enzyme complexes (the other sub-units of these complexes are encoded in the nucleus), one protein of mitochondrial ribosomes (var 1) and, possibly, some other proteins; and (b) interspersed repeated sequences, namely the AT spacers and the GC clusters.

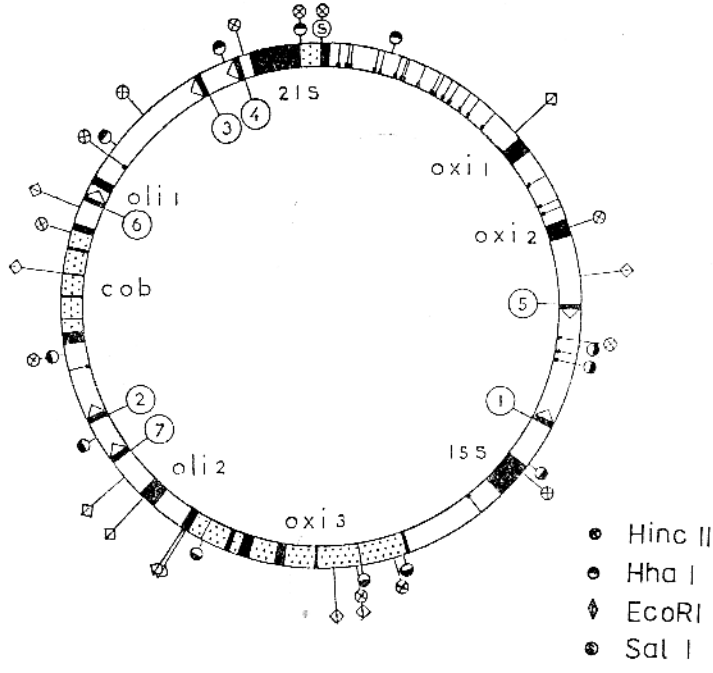


Fig. 1. Physical and genetical map of a mitochondrial genome unit of wild-type yeast (strain A). Some restriction sites are indicated. Circled numbers indicate the location of ori sequences 1—7 (arrowheads point in the direction cluster C to cluster A; see Fig. 4). Black and dotted areas correspond to exons and introns of mitochondrial genes, respectively. The introns of cob and oxi 3 genes contain closed reading frames (4). Thin radial lines indicate tRNA genes. The var 1 gene is located between ori 3 and ori 6. White areas correspond to long AT spacers embedding short GC clusters. [Modified from de Zamaroczy et al. (20).]

The instability of the mitochondrial genome of yeast: the spontaneous petite mutation

The direct sequence repeats present in the AT spacers and in the GC clusters of the mitochondrial genome of wild-type yeast cells are the source of a tremendous instability, since they can function as excision sequences (21) to cut out genome segments which then undergo a tandem amplification process to become the repeat units of the defective mitochondrial genome units of spontaneous suppressive petite mutants (Fig. 2; see below for the definition of suppressivity). An identical excision process can also occur in the genomes of petite mutants, leading to the production of the genomes of secondary suppressive petite mutants (Fig. 2). Petites totally deprived of mitochondrial genome (neutral petites) may also be produced. The very large number of direct repeats in the wild-type genome accounts for the extremely high frequency of petite mutants (1—5% per generation). The excision process leading to the production of petite genomes, in all likelihood

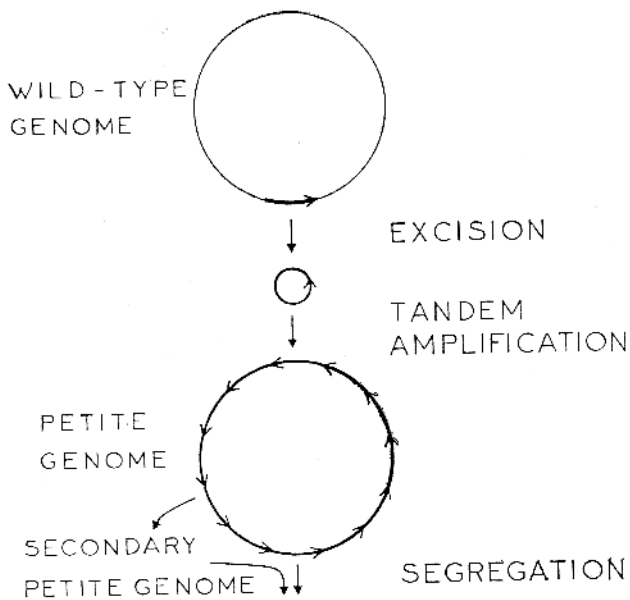


Fig. 2. Scheme depicting the excision-amplification process leading to the formation of the genome of a spontaneous petite mutant. A segment of a unit of a wild-type mitochondrial genome is excised and tandemly amplified into a defective genome unit. This then replicates and segregates into the buds to form the genome of a petite mutant; the petite genome can undergo further excisions leading to the formation of secondary petite genomes.

an internal site-specific recombination (Fig. 3), is probably just a by-product of the very frequent site-specific recombination events occurring in the wild-type genome (10).

The spontaneous petite mutants, whether or not they still contain a defective mitochondrial genome, are respiratory deficient. This condition is tolerated by the facultative aerobe *S. cerevisiae* and leads to the formation of small (petite) colonies because of the lower growth rate allowed by the use of fermentative instead of respiratory pathways.

The ori sequences

The vast majority of spontaneous suppressive petite mutants contain in each repeat unit of their mitochondrial genome units at least one of seven canonical ori sequences. These (Fig. 4) are highly homologous sequences about 300 base pairs in length characterized by two GC clusters, A and B,

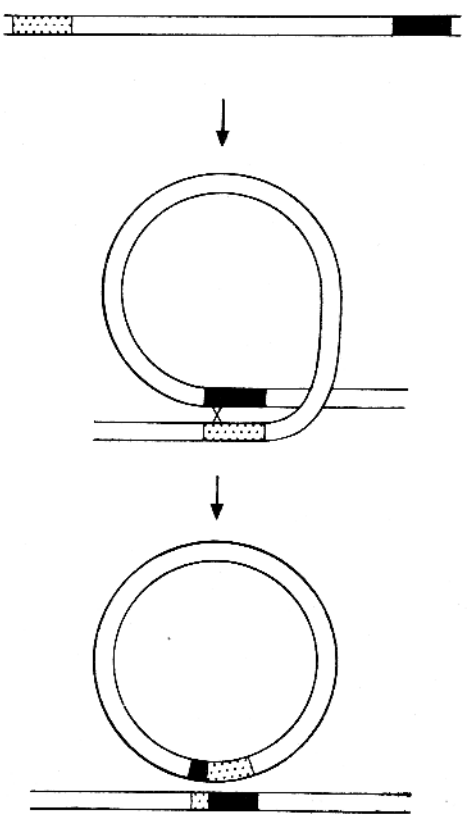


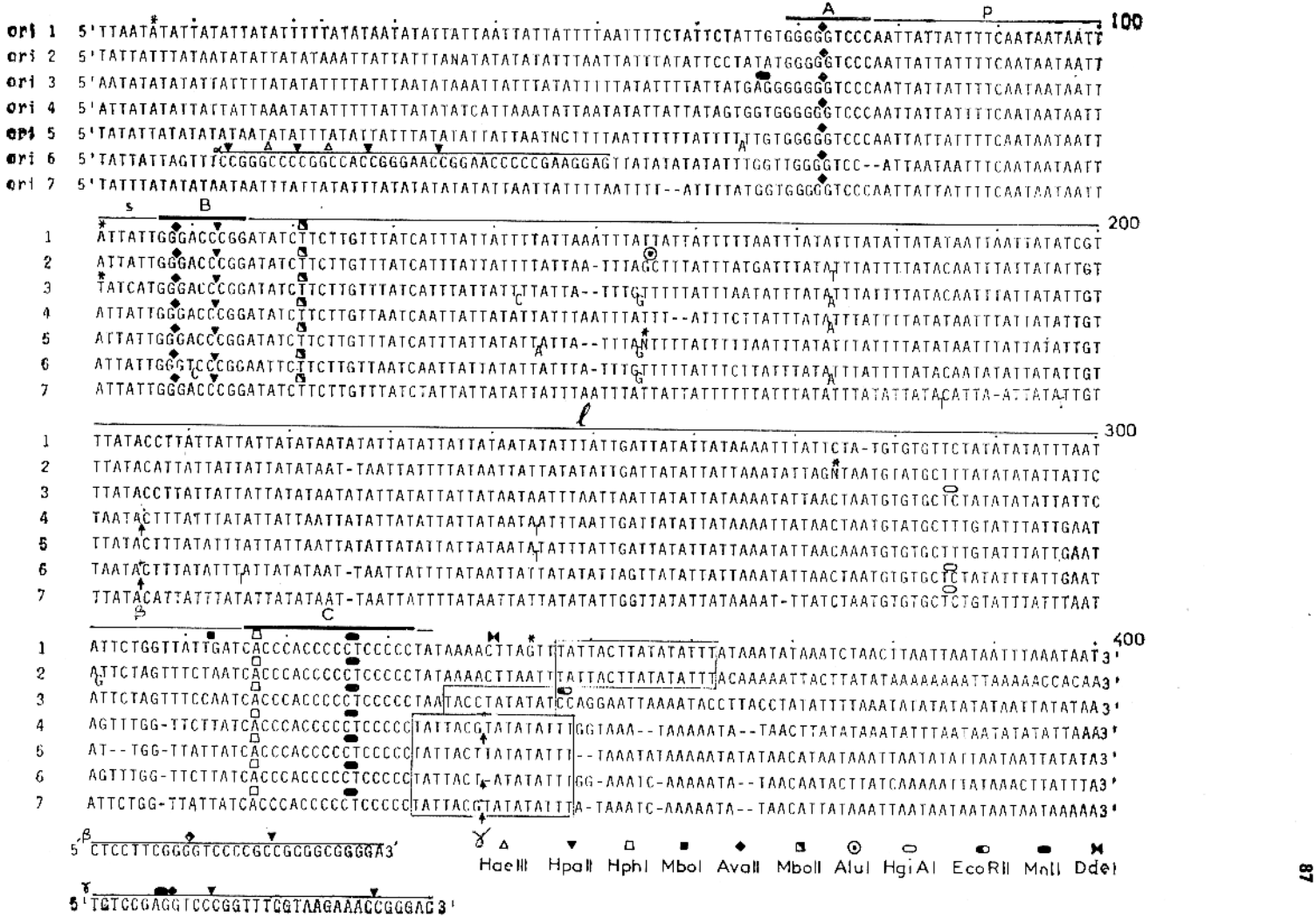
Fig. 3. Scheme of the excision process leading to the formation of a petite genome unit. Dotted and black areas represent two direct sequence repeats. Excision is followed by a tandem amplification (see Fig. 2).

at one end and one GC cluster ,C, at the other; the middle region is extremely high in AT. Two additional GC clusters, called β and γ, exist in some ori sequences; they are located in the middle region and just after cluster C, respectively (Fig. 4). The folding of clusters A and B into a hairpin loop and the sequence of cluster C are very similar to structures found in the origins of replication of mammalian mitochondrial genomes (Fig. 5). Partial or total deletions and rearrangements of ori sequences profoundly depress the suppressivity of the corresponding petites (called ori$^-$, ori$^\circ$ and ori^r, respectively), namely the level of transmission of the petite genome to the progeny of petite × wild-type cell crosses (22). Recent investigations (2) have shown that sequences contiguous to cluster C are used as transcription initiation sites, that transcription proceeds in the cluster C to cluster A direction using the strand containing the pyrimidines of cluster C as template, and that transcription initiation sites are highly homologous to those used for the transcription of rRNAs (18). The presence of both orientations in ori sequences (Fig. 1) suggests that both strands are transcribed, as in the mitochondrial DNA of animal cells. Interestingly, transcription is lost in ori$^\circ$ and in ori C$^-$ petite genomes as well as in the genomes containing cluster γ inserted in the transcription initiation sequence. In the case of ori$^\circ$ petite genome, it has been shown that surrogate origins, formed by palindromic GC clusters (the ori^s sequences), are used for replication (12).

It should be pointed out that the canonical ori sequences of the mitochondrial genome of yeast share two basic features with prokaryotic origins of replication (1), namely their length and the presence of very highly conserved recognition sequences, the GC clusters, and the less conserved spacer sequences, separating the GC clusters. It is conceivable that, as in prokaryotic ori sequences (13), a transcriptional activation facilitates the binding of a primase initiating the replicating strand. RNA priming at the origin of DNA replication has been demonstrated in the case of mammalian mitochondrial DNA (6).

The evolutionary origin of non-coding sequences

The very high homology of the seven canonical ori sequences indicates that they arose as the result of duplication and translocation events. As far as the intergenic sequences are concerned, if one considers that they are made up of AT spacers and GC clusters like the ori sequences, and that four out of eight of them contain one or two ori sequences in their middle, it is conceivable that they were derived from ori sequences by an expansion phenomenon. This might have taken place through three different mechanisms, all of which are likely to have played a role. First, a slippage of the replicase could occur at the ori sequences; this is a well-known phenomenon first studied in the reiterative replication of poly (dAT : dAT) by DNA polymerase I of *E. coli* (15). A second mechanism is unequal crossing-over; evidence for the high fre-


ori 1
ori 2
ori 3
ori 4
ori 5
ori 6
ori 7
A
P
100
s
B
200
300
400
C
HaeIII HpaII HphI MboI AvaII MboII AluI HgiAI EcoRII MnlI DdeI
87

quency of such a phenomenon in mitochondrial recombination is available (10). A third mechanism is insertion. Almost all GC clusters are inserted in AT spacers; some rare ones are inserted in AT-rich regions of rRNA genes (20) and even in a protein-coding gene, var 1 (14). Interestingly, these insertions are not only transcribed but, in the case of var 1, also translated.

The expansion hypothesis (4) was put to an experimental test by comparing the level of homology existing between ori sequences and intergenic sequences and that found between random sequences, having the same length and base composition of ori sequences, and intergenic sequences. Such a test, in which ori sequences or random sequences were compared with about 20 000 nucleotides of intergenic sequences (4) revealed that homology was 10 times larger in the first case than in the second. This supports the view that intergenic sequences were derived from ori sequences by an expansion process.

Another interesting result (4) was that the closed reading frames of the intervening sequences of oxi 3 and cob genes share all the features of intergenic non-coding sequences. This may either reflect an insertion of an intergenic sequence into an intron, or indicate that the intron itself originated from an intergenic sequence that at one time contained an ori sequence. The loss of ori sequences in the mitochondrial genomes of some yeast strains is well documented (Faugeron-Fonty, personal communication).

The case of the var 1 gene (14) is of special interest in connection with the idea of the expansion of ori sequences. This gene is 10% GC and contains a 46 bp GC cluster accounting for 38% of total GC. Its similarity with spacer-cluster sequences is so striking that it suggests that this gene arose from an intergenic sequence only recently.

Are the non-coding sequences of the mitochondrial genome of yeast selfish DNA sequences ?

An obvious question raised by the presence of such a large amount of non-coding sequences in the mitochondrial genome of yeast is whether these sequences can be considered selfish DNA. The origin of such sequences by a mechanism identical to that postulated for selfish DNA by Orgel and Crick (17) only makes the question more relevant. Our present understanding of the molecular genetics of yeast mitochondria allows a meaningful discussion of this issue, and the following case can be made.

Fig. 4. Primary structure of the ori sequences and their flanking sequences. Thick lines indicate GC clusters A, B, and C, thin lines AT regions p, s and l. The positions and the sequences of extra GC clusters β and γ are given (as well as that of GC clusters α, which is located outside the ori sequence). Restriction sites are indicated by the symbols shown. The sequences homologous to initiation transcription sites on the right side of cluster C, are indicated by boxes. [Taken from de Zamaroczy et al. (20), Baldacci and Bernardi (2), and unpublished results of M. de Zamaroczy.]

The non-coding sequences of the mitochondrial genome of yeast are the source of three disadvantages for the genome. The main one is that the abundant direct repeats which they contain are potential excision sequences; as such, they are responsible for the extreme instability of the genome. The other two are that they increase the genome replication time and energy expenditure. These disadvantages would quickly change wild-type yeast cells into suppressive and neutral petite mutants, if this intracellular selection were not counterbalanced by an intercellular selection in which the faster growing respiratory-competent wild-type cells compete out the respiratory-deficient petite mutants; (this accounts for the fact that in nature

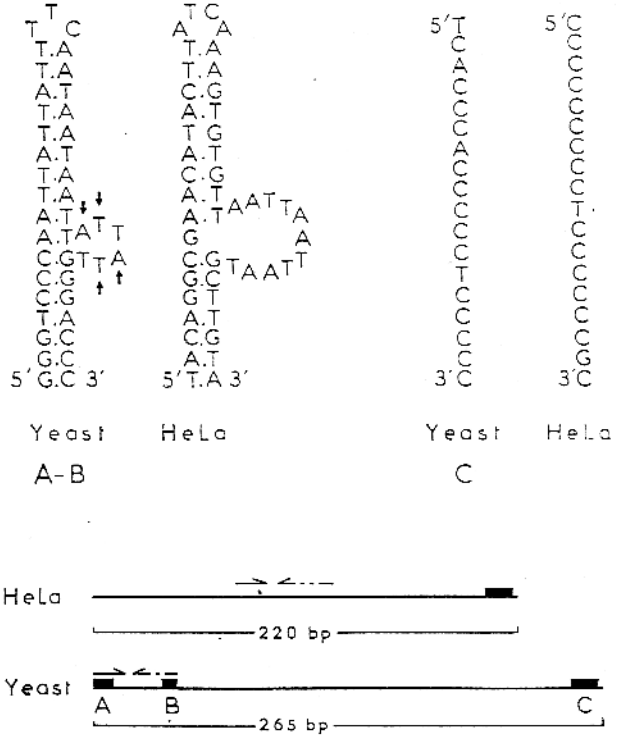


Fig. 5. Comparison of ori sequences of mitochondrial genomes from yeast (20) and HeLa cells (8). Homology of potential secondary structure is found for the inverted repeats in cluster A — cluster B region; arrows indicate the base changes found in this region in different petite genomes. Homology of primary structure is found for cluster C. The bottom compares the two ori sequences; the arrows indicate the inverted repeats of the A—B region, the broken line corresponding to the looped-out sequence; bp, base pairs. [Taken from de Zamaroczy et al. (20).]

only wild-type yeast cells are found). Even if the disadvantages associated with the non-coding sequences do not lead, therefore, to the elimination of the mitochondrial genome, they should at least lead to the elimination of the non-coding sequences themselves. We know, however, that, although *S. cerevisiae* strains exist which lack a number of intervening sequences and also ori 4, in general non-coding sequences tend to be largely conserved. This indicates that the removal of non-coding sequences is selectively disadvantageous or, in other words, that non-coding sequences provide selective advantages which compensate for the disadvantages associated with them. This obviously raises the question of the nature of these advantages, namely of the physiological roles played by non-coding sequences.

It is clear from the map in Fig. 1 that the excision of intergenic sequences, where most excision sequences used in the spontaneous mutation are located (21), will frequently remove canonical ori sequences from the wild-type genome. Even if a wild-type genome lacking ori 4 has been found, it is evident that in general such elimination will affect replication and also transcription. There is therefore a selective advantage in keeping ori sequences in the wild-type genome. As far as non-coding sequences outside ori sequences are concerned, it should first of all be stressed that the expansion process does not propagate non-sense sequences, but propagates instead sequences which have been highly selected and conserved in evolution and whose primary role is to interact specifically with enzymes involved in DNA replication and transcription. Thus, the expansion of ori sequences leads to the propagation of potential regulatory signals, which may be used in the regulation of gene expression, in repression (anaerobiosis or glucose can shut off transcription) and derepression, in the processing of primary transcripts, and in the regulation of nucleo-mitochondrial interactions. Another physiological role of non-coding sequences concerns recombination. Indirect evidence exists (10) that repeated and palindromic non-coding sequences are involved in mitochondrial site-specific recombination. Finally, it should be recalled that some non-coding sequences appear to be inserted into transcribed genes or even to be transformed into genes. In summary, a number of physiological roles can be demonstrated, or thought of, for non-coding sequences; these roles apparently provide selective advantages compensating for the disadvantages inherent in the non-coding sequences.

A final point on which the mitochondrial genome of yeast is relevant to the selfish DNA issue is the occurrence of functionless genomes in suppressive petites (19). Many of these genomes contain no gene, and yet replication, transcription and even transcript processing may still go on. In nature, as already pointed out, these genomes rapidly disappear as petites are competed out by the faster growing wild-type cells. Many of these genomes have such a replicative advantage over wild-type genomes that they could spread out through crosses with wild-type cells, if haploid. This does not

occur either in nature because parental wild-type cells and the derived petites have the same mating type. When isolated from competition with wild-type cells in the laboratory, however, petites not only survive, but frequently end up with very stable genomes which are the result of a selection on the basis of replication efficiency. These "genomes without genes" are practically made up of repeat units containing barely more than an ori sequence; replication is most efficient, the corresponding petites being super-suppressive, and transcription is preserved probably because of the role played in replication. These "selfish genomes", exemplifying primordial self-replicating systems, will only be lost in the long run, when mitochondrial or nuclear mutations will inactivate the initiation of replication.

In conclusion, what we know about the non-coding sequences of the mitochondrial genome of yeast suggests that their conservation in the genome is due to selective advantages associated with their physiological roles; these sequences cannot, therefore, be considered selfish DNA sequences, at least in their majority. On the other hand, functionless genomes like the mitochondrial genomes of suppressive petites not only can exist and be quite stable, but they undergo a selection favoring those which are closest to the ultimate situation of being just a set of replication origins; this in vivo selection is very much the same found in the Qβ-replicase in vitro system by Mills et al. (16).

REFERENCES

1. Asada K., Sugimoto, K., Oka, A., Takanami, M., Hirota, Y., Nucleic Acids Res., *10* : 3745, 1982.
2. Baldacci, G., Bernardi, G., EMBO J., *1* : 987, 1982.
3. Bernardi, G., Trends Biochem. Sci., *4* : 197, 1979.
4. Bernardi, G., in: *Mitochondrial genes*. Eds. Slonimski, P. P. et al. Cold Spring Harbor 1982 (p. 269).
5. Bernardi, G., Bernardi, G., FEBS Lett., *115* : 159, 1980.
6. Clayton, D. A., Cell, *28* : 693, 1982.
7. Cosson, J., Tzagoloff, A., J. Biol. Chem., *254* : 42, 1979.
8. Crews, S., Ojala, D., Posakony, J., Nishiguchi, J., Attardi, G., Nature, *277* : 192, 1979.
9. Doolittle, W. F., Sapienza, C., Nature, *284* : 601, 1980.
10. Fonty, G., Goursot, R., Wilkie, D., Bernardi, G., J. Mol. Biol., *119* : 213, 1978.
11. Gaillard, C., Bernardi, G., Mol. Gen. Genet., *174* : 335, 1979.
12. Goursot, R., Mangin, M., Bernardi, G., EMBO J., *1* : 705, 1982.
13. Hobom, G., Curr. Topics Microbiol. Immunol., *94/95* : 93, 1981.
14. Hudspeth, M. E. S., Ainley, W. M., Shumard, D. S., Butow, R. A., Grossman, L. I., Cell, *30* : 617, 1982.
15. Kornberg, A., Bertsch, L. L., Jackson, J. F., Khorana, H. G., Proc. Nat. Acad. Sci., *51* : 315, 1964.
16. Mills, D. R., Peterson, R. L., Spiegelman, S., Proc. Nat. Acad. Sci., *58* : 217, 1967.
17. Orgel, L. E., Crick, F. H. C., Nature, *284* : 604, 1980.
18. Osinga, K. A., Tabak, H. F., Nucleic Acids Res., *10* : 3617, 1982.

19. Reid, R. A., Nature, *285* : 620, 1980.
20. Sor, F., Fukuhara, H., Nucleic Acids Res., *10* : 1625, 1982.
21. Zamaroczy, M. de, Faugeron-Fonty, G., Bernardi, G., Gene, 1983 in press.
22. Zamaroczy, M. de, Marotta, R., Faugeron-Fonty, G., Goursot, R., Mangin, M., Baldacci, G., Bernardi, G., Nature, *292* : 75, 1981.