

Genome instability and the selfish DNA issue

Giorgio BERNARDI *

1. Introduction
2. The organization of the mitochondrial genome of yeast
3. The instability of the mitochondrial genome of yeast: the spontaneous *petite* mutation
4. The *ori* sequences
5. The evolutionary origin of non-coding sequences
6. Are the non-coding sequences of the mitochondrial genome of yeast selfish DNA ?

ABSTRACT: The mitochondrial genome of yeast is formed by 25-50 circular DNA molecules containing, next to single-copy gene sequences, abundant interspersed, non-coding repeated sequences, the long AT spacers and the short GC clusters. These are the source of a tremendous genome instability, since the direct repeats they contain are used as excision sequences releasing DNA segments which undergo tandem amplification to become the repeat units of the defective genomes of petite mutants. Such repeat units usually contain at least one of the seven origins of DNA replication of the parental genome. The homology between ori sequences and the non-coding sequences suggests that the latter arose by an expansion of the former. Notwithstanding their direct responsibility in genome instability, the non-coding sequences are conserved, showing that they are associated with a selective advantage having to do with regulation of gene expression and recombination. The interspersed repeated sequences of yeast cannot therefore be considered 'selfish DNA'.

I. INTRODUCTION

Both prokaryotes and eukaryotes contain repeated DNA sequences in their genomes. The relative amounts of such sequences are, however, very different. In the case of prokaryotes, a small number of families

* Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris.

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 play no phenotypic or evolutionary role, their only function being their survival within the genome (DOOLITTLE and SAPIENZA, 1980). According to ORGEL and CRICK (1980) « 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 outlined 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. Therefore 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.

2. THE ORGANIZATION OF THE MITOCHONDRIAL GENOME OF YEAST

Investigations carried out in our laboratory between 1966 and 1976 (see Bernardi, 1979, 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 (BERNARDI and BERNARDI, 1980); and b) about 200 short GC clusters (GC > 60%) characterized by sequences that are often palindromic and largely homologous to each

other. These GC clusters are embedded in AT spacers (COSSON and TZAGOLOFF, 1979; GAILLARD and BERNARDI, 1979). AT spacers and GC clusters form the intergenic sequences of the mitochondrial genome and the closed reading frame regions which are localized (BERNARDI, 1982) in the middle 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.

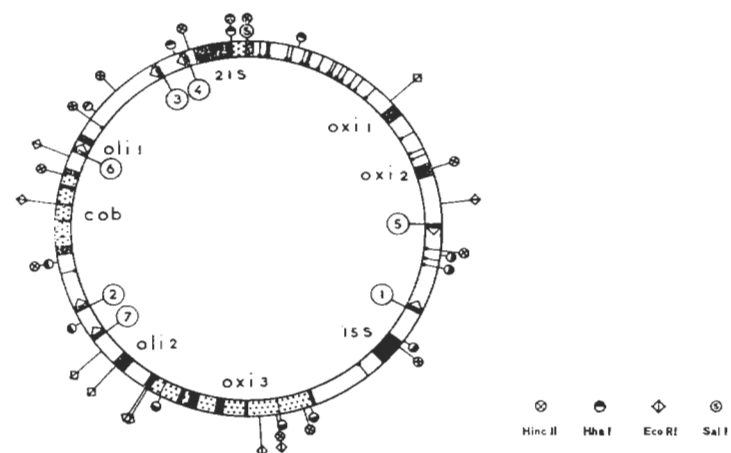


Fig. 1 — Physical and genetical map of a mitochondrial genome unit of wild-type yeast (strain A). Some restriction sites are indicated. Circled numbers indicate the location of *ori* sequences 1–7 (arrowheads point in the direction cluster C to cluster A; see Fig. 4). Black and dotted areas correspond to exons and introns of mitochondrial genes, respectively. The introns of *cob* and *oxi 3* genes contain closed reading frames (BERNARDI, 1982). 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., 1981).

— Carta fisica e genetica di una unità genomica mitocondriale di cellule di tipo selvaggio di lievito (ceppo A). Alcuni siti di restrizione sono indicati. I numeri entro circoli indicano la posizione delle sequenze *ori* 1–7 (le frecce sono puntate nella direzione nucleo C → nucleo A; vedi Fig. 4). Le zone nere e punteggiate corrispondono rispettivamente a esoni e introni di geni mitocondriali. Gli introni dei geni *cob* e *oxi 3* contengono quadri di lettura chiusi (BERNARDI, 1982). Linee radiali sottili indicano i geni di tRNA. Il gene *var 1* è situato tra *ori 3* e *ori 6*. Le aree bianche corrispondono agli spaziatori AT e ai nuclei di GC (modificato da DE ZAMAROCZY et al., 1981).

3. 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 (DE ZAMAROCZY et al., 1983) 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

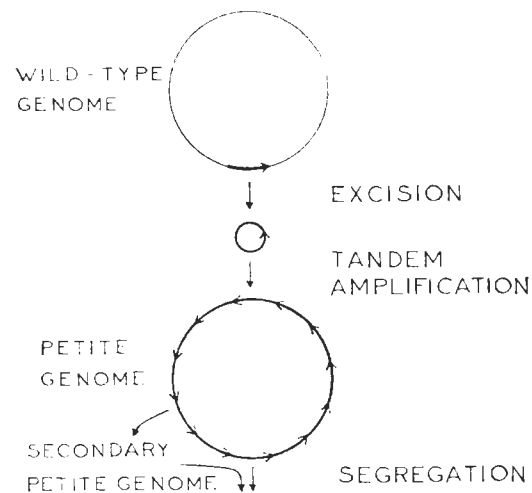


Fig. 2 — Scheme depicting the excision–amplification process leading to the formation of the genome of a spontaneous petite mutant. A segment of a unit of a wild-type mitochondrial genome is excised and tandemly amplified into a defective genome unit. This then replicates and segregates into the buds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

— Schema del processo di escissione–amplificazione che conduce alla formazione del genoma di un mutante spontaneo 'petite colonie'. Un segmento di una unità genomica di tipo selvaggio è escisso e amplificato in tandem per dare una unità genomica difettiva. Questa poi si replica e segrega nelle gemme per formare il genoma di un mutante petite. Questo genoma difettivo può subire ulteriori escissioni che conducono alla formazione di genomi 'petite colonie' secondari.

sive 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 spontaneous petite mutants (1–5% per generation). The excision process leading to the production of petite genomes, in all likelihood an internal site-specific recombination (Fig. 3), is probably just a side-line of the very frequent site-specific recombination occurring in the wild-type genome (FONTY et al., 1978).

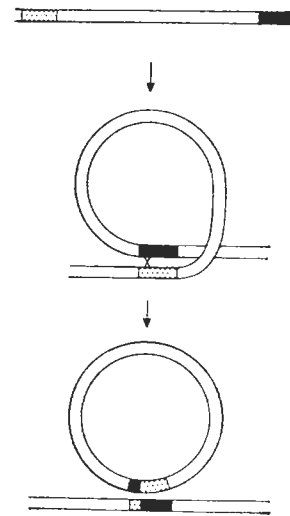


Fig. 3 — Scheme of the excision process leading to the formation of a petite genome unit. Dotted and black areas represent two direct sequence repeats. Excision is followed by a tandem amplification (see Fig. 2).

— Schema del processo di escissione che conduce alla formazione di una unità genomica petite. Le aree nere e punteggiate rappresentano due sequenze ripetute dirette. L'escissione è seguita da una amplificazione in tandem (vedi Fig. 2).

The spontaneous petite mutants, whether or not they still contain a defective mitochondrial genome, are respiratory deficient. This condition is tolerated by the facultative anaerobe *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.

4. 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 (Fig. 4). These are highly homologous

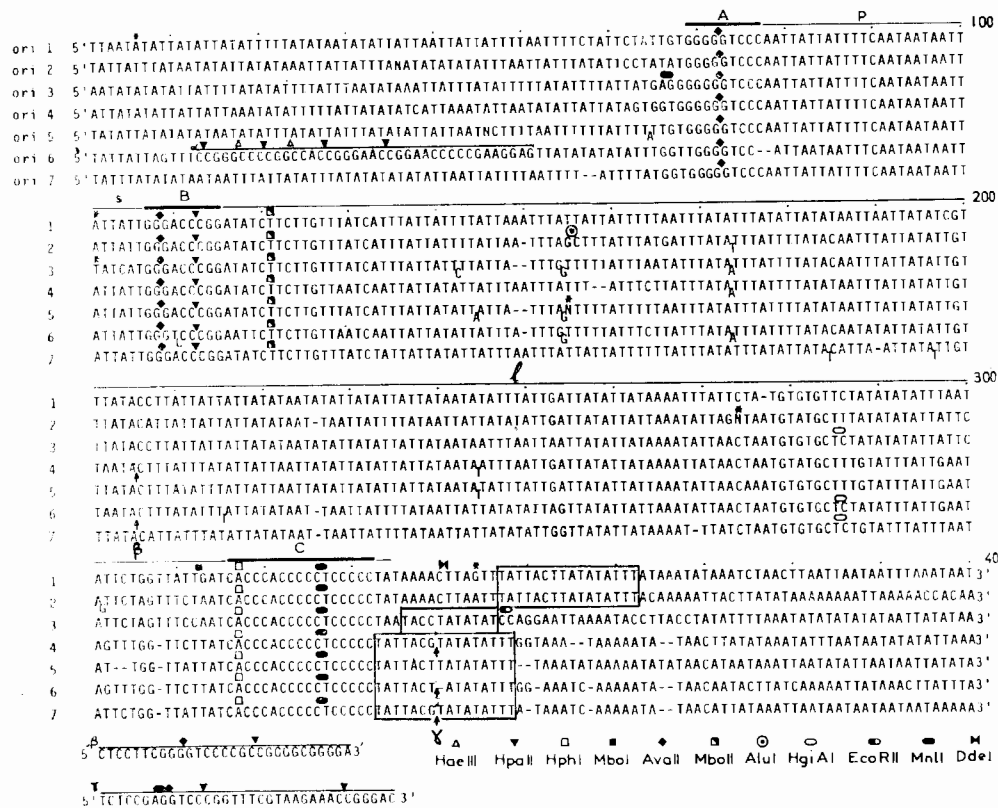


Fig. 4 — Primary structure of the *ori* sequences and their flanking sequences. Thick lines indicate GC clusters A, B and C, thin lines AT regions *p*, *s*, and *l*. The positions and the sequences of extra GC clusters β and γ are given (as well as that of GC cluster α , which is located outside the *ori* sequence). Restriction sites are indicated by the symbols shown. The sequences homologous to initiation transcription sites at the right of cluster C, are indicated by boxes (from DE ZAMAROCZY et al., 1981; BALDACCI and BERNARDI, 1982; and unpublished results of M. DE ZAMAROCZY).

— *Struttura primaria delle sequenze ori. I tratti spessi indicano i nuclei di GC: A, B e C, i tratti sottili le regioni AT p, s e l. Le posizioni e la sequenza dei nuclei di GC supplementari β e γ sono indicati, come anche quella del nucleo di GC α , che si trova all'esterno della sequenza ori. I siti di restrizione sono indicati dai simboli presentati in figura. Le sequenze omologhe ai siti di iniziazione della trascrizione a destra del nucleo C sono inquadrate (da DE ZAMAROCZY et al., 1981; BALDACCI e BERNARDI, 1982; e risultati non pubblicati di DE ZAMAROCZY).*

sequences, about 300 base pairs in length, characterized by two GC clusters, A and B, 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*^o and *ori*^r, respectively), namely the level of transmission of the petite genome to the progeny of petite x wild-type cell crosses (DE ZAMAROCZY et al., 1981). Recent investigation (BAL-

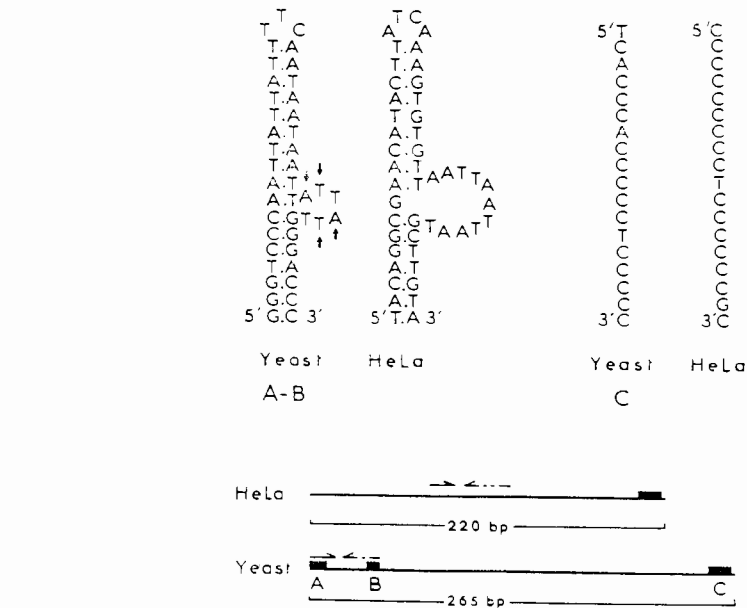


Fig. 5 — Comparison of *ori* sequences of mitochondrial genomes from yeast (DE ZAMAROCZY et al., 1981) and HeLa cells (CREWS et al., 1979). Homology of potential secondary structure is found for the inverted repeats in cluster A — cluster B region; arrows indicate the base changes found in this region 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 (from DE ZAMAROCZY et al., 1981).

— *Confronto delle sequenze ori dei genomi mitocondriali di lievito (DE ZAMAROCZY et al., 1981) e cellule HeLa (CREWS et al., 1979). Una omologia di struttura secondaria potenziale esiste per le ripetizioni inverse della regione nucleo A — nucleo B; le frecce indicano i cambiamenti di base trovati in questa regione in diversi genomi 'petite'. Una omologia di struttura primaria esiste per il nucleo C. La parte inferiore della figura presenta un confronto tra le due sequenze ori; le frecce indicano le ripetizioni inverse della regione A-B; la linea interrotta corrisponde all'ansa; bp, paio di basi (da DE ZAMAROCZY et al., 1981).*

DACCI and BERNARDI, 1982) 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 (OSINGA and TABAK, 1982). 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*^o 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*^o petite genome, it has been shown that surrogate origins, formed by palindromic GC clusters (the *ori*^s sequences), are used for replication (GOURSOT et al., 1982).

It should be pointed out that the canonical *ori* sequences set of the mitochondrial genome of yeast share two basic features with prokaryotic origins of replication (ASADA et al., 1982), namely their length and the presence of very highly conserved recognition sequences, the GC clusters, and the less conserved spacer sequence, separating the GC clusters. It is conceivable that, as in prokaryotic *ori* sequences (HOBOM, 1981), 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 (CLAYTON, 1982).

5. 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 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* (KORNBERG et al., 1964). A second mechanism is unequal crossing-over; evidence for the

high frequency of such a phenomenon in mitochondrial recombination is available (FONTY et al., 1978). 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 (SOR and FUKUHARA, 1982) and even in a protein-coding gene, *var 1* (HUDSPETH et al., 1982). Interestingly these insertions are not only transcribed but, in the case of *var 1*, also translated.

The expansion hypothesis (BERNARDI, 1982) 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 (BERNARDI, 1982) revealed that homology was 10 times larger in the first case than in the second one. This supports the view that intergenic sequences were derived from *ori* sequences by an expansion process.

Another interesting result (BERNARDI, 1982) 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, pers. comm.).

The case of the *var 1* gene (HUDSPETH et al., 1982) 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.

6. 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 (1980) 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. 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 was 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 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 of Fig. 1 that the excision of intergenic sequences, where most excision sequences used in the spontaneous mutation are located, (DE ZAMAEOCZY et al., 1983) 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 (FONTY et al., 1978) 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 (see REID, 1980). 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 supersuppressive, 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 Q β -replicase *in vitro* system by MILLS et al. (1967).

Paris, December 4th 1982

L'instabilità genomica e la questione del DNA egoista

1. *Introduzione.* Sia i procarioti che gli eucarioti contengono sequenze ripetute di DNA nei loro genomi. Le quantità relative di queste sequenze, tuttavia, sono molto diverse. Nei procarioti, un piccolo numero di famiglie di sequenze di inserzione formano circa l'1-2% nel genoma. Negli eucarioti, le famiglie di sequenze ripetute 'intersperse' generalmente rappresentano 20-30% del genoma. È stato suggerito che queste sequenze ripetute riempiono un ruolo di regolazione, o sono essenziali per la struttura, l'appaiamento, i riarrangiamenti dei cromosomi. Un altro punto di vista è che queste sequenze non abbiano una funzione che si esprime nel fenotipo o nel corso dell'evoluzione: la loro sola funzione sarebbe invece la loro propria sopravvivenza nel genoma (DOOLITTLE e SAPIENZA, 1980). Secondo ORGEL e CRICK (1980) « un pezzo di DNA egoista, nella sua più pura espressione, ha due proprietà distinte: 1) si origina quando una sequenza di DNA si propaga formando copie di se stessa nel genoma; 2) non contribuisce in modo specifico al fenotipo ».

Una osservazione preliminare deve essere fatta all'inizio di questa discussione, e cioè che i due punti di vista menzionati rappresentano due estremi, e non si escludono a vicenda. Infatti, è concepibile che certe sequenze ripetute abbiano un ruolo biologico, ed altre no, per cui una discussione del problema in questione dovrebbe condurre idealmente a stimare la quantità relativa di sequenze ripetute che riempiono un ruolo biologico. Questa discussione sarà difficile, se non impossibile, se noi non conosciamo sufficientemente il genoma o i genomi presi come modelli. Sfortunatamente, questa è esattamente la situazione del genoma nucleare degli eucarioti, che è stato finora al centro del dibattito. In questo articolo discuterò il problema del DNA egoista usando come modello il genoma mitocondriale del lievito, un genoma quasi completamente sequenziato e ben conosciuto dal punto di vista della genetica. Questo sistema è particolarmente interessante perché presenta la stessa caratteristica fondamentale del genoma nucleare degli eucarioti, cioè una intersperzione di sequenze 'uniche' (cioè presenti una sola volta nel genoma) e ripetute.

2. *L'organizzazione del genoma mitocondriale del lievito.* Ricerche effettuate nel nostro laboratorio tra il 1966 e il 1976 (vedi BERNARDI, 1979, per una breve rassegna) dimostrarono che più del 50% delle 25 (o 50) unità genomiche mitocondriali presenti in ogni cellula aploide (o diploide) di tipo selvaggio di *Saccharomyces cerevisiae* consiste di: a) lunghi 'AT spacers' (il contenuto in GC di questi spaziatori AT è inferiore al 5%) formati da brevi sequenze dAT:dAT e dA:dT con rare paia di basi dG:dC; queste sequenze sono ripetitive e ricche in palindromi (BERNARDI e BERNARDI, 1980); b) circa 200 corti 'GC clusters' (il contenuto in GC di questi nuclei di GC è superiore al 60%); questi sono caratterizzati da sequenze che sono spesso palindromiche e in grande misura omologhe. I nuclei di GC sono all'interno degli spaziatori AT (COSSON e TZAGOLOFF, 1979; GAILLARD e BERNARDI, 1979). Gli spaziatori AT e i nuclei di GC formano le sequenze intergeniche del genoma mitocondriale e i quadri di lettura chiusi degli introni dei geni *cob* e *oxi 3* (Fig. 1). Come il genoma nucleare degli eucarioti, il genoma mitocondriale del lievito contiene: a) sequenze uniche, cioè i geni per gli RNA ribosomici, i tRNA e gli mRNA delle sub-unità polipeptidiche di enzimi respiratori (altre sub-unità di questi stessi enzimi sono codificate dal nucleo), di una proteina dei ribosomi mitocondriali (*var 1*) e, forse, di qualche altra proteina; b) sequenze ripetute intersperse, cioè gli spaziatori AT e i nuclei di GC.

3. *Instabilità del genoma mitocondriale del lievito: la mutazione spontanea 'petite colonie'.* Le sequenze ripetute dirette (cioè aventi la stessa orientazione) presenti negli spaziatori AT e nei nuclei di GC del genoma mitocondriale di cellule di lievito di tipo selvaggio sono la fonte di una tremenda instabilità, poiché possono funzionare come sequenze di escissione (DE ZAMAROCZY et al., 1983). Segmenti di genoma così escissi vanno incontro ad una amplificazione in tandem (cioè con la stessa orientazione) e diventano le *unità di ripetizione* delle *unità genomiche* difettive dei mutanti spontanei soppressivi 'petite colonie' (Fig. 2; vedi sotto per la definizione di soppressività). Un processo di escissione identico può avvenire anche nei genomi dei mutanti 'petite', e condurre alla produzione dei genomi di petites soppressive secondarie (Fig. 2). Petites completamente prive di genoma mitocondriale (petites neutre) possono anche essere prodotte. Il grandissimo numero di sequenze ripetute dirette nel genoma di tipo selvaggio spiega l'altissima frequenza della mutazione petite spontanea (1-5% per generazione). Il processo di escissione che conduce alla produzione dei genomi delle petites, con ogni verosimiglianza una ricombinazione interna sito-specifica

(Fig. 3), è probabilmente solo una via collaterale della frequentissima ricombinazione sito-specifica che ha luogo nel genoma di tipo selvaggio (FONTY et al., 1978).

I mutanti spontanei 'petite colonie', che contengano o meno un genoma mitocondriale difettivo, non respirano. Questa situazione è tollerata dall'anaerobio facoltativo *S. cerevisiae* e conduce alla formazione di piccole (petite) colonie a causa della più lenta velocità di crescita permessa dall'uso delle vie metaboliche della fermentazione invece di quelle della respirazione.

4. *Le sequenze ori.* La grandissima maggioranza delle petites spontanee soppressive contengono in ciascuna unità di ripetizione delle loro unità genomiche mitocondriali almeno una delle sette sequenze *ori* canoniche (Fig. 4). Queste sono sequenze ad alta omologia, lunghe circa 300 paia di basi, e caratterizzate da due nuclei di GC, A e B, ad una estremità, e da un nucleo, C, all'altra; la regione intermedia è estremamente ricca in AT. Due nuclei di GC supplementari, β e γ , esistono in certe sequenze *ori*, e sono localizzati rispettivamente nella regione intermedia e subito dopo il nucleo C (Fig. 4). La struttura secondaria potenziale dei nuclei A e B e la struttura primaria del nucleo C sono molto simili a strutture esistenti nelle origini di replicazione dei genomi mitocondriali di mammiferi (Fig. 5). Delezioni parziali o totali e riarrangiamenti delle sequenze *ori* diminuiscono fortemente la soppressività delle petites corrispondenti (chiamate rispettivamente *ori*⁻, *ori*^o e *ori*^r), cioè il livello di trasmissione del genoma petite alla discendenza di incroci petite x tipo selvaggio (DE ZAMAROCZY et al., 1981). Ricerche recenti (BALDACCI e BERNARDI, 1982) hanno dimostrato che sequenze contigue al nucleo C sono usate come siti di iniziazione della trascrizione, che la trascrizione procede nella direzione nucleo C \rightarrow nucleo A usando come modello la catena che contiene le pirimidine del nucleo C e che i siti di iniziazione della trascrizione sono quasi identici a quelli usati per la trascrizione degli RNA ribosomici (OSINGA e TABAK, 1982). Siccome le sequenze *ori* presentano le due orientazioni possibili sul genoma (Fig. 1), è verosimile che tutte e due le catene del DNA siano trascritte come nel DNA mitocondriale dei mammiferi. La trascrizione non ha luogo in genomi *ori*^o e *ori*^{C-}, o contenenti il nucleo γ nella sequenza di iniziazione della trascrizione. Nel caso dei genomi *ori*^o, è stato dimostrato che origini surrogate di replicazione sono usate; queste origini sono formate da nuclei di GC palindromici (le sequenze *ori*^s: GOURSOT et al., 1981).

Le sequenze *ori* canoniche del genoma mitocondriale di lievito posseggono

due caratteristiche delle origini di replicazione procariotiche (ASADA et al., 1982), cioè la lunghezza e la presenza di sequenze di riconoscimento molto conservate, i nuclei di GC, e di una sequenza intermedia meno conservata che separa i nuclei di GC. È concepibile che una attivazione trascrizionale faciliti la fissazione sulla sequenza *ori* di una primasi che inizia la nuova catena di DNA, come nelle sequenze *ori* procariotiche (HOBOH, 1981). Che un piccolo oligoribonucleotide sia all'origine della replicazione del DNA è stato dimostrato nel caso del DNA mitocondriale dei mammiferi (CLAYTON, 1982).

5. *L'origine evolutivistica delle sequenze non-codificanti.* L'altissima omologia delle sette sequenze *ori* indica che queste si sono originate per via di duplicazioni e traslocazioni. Per quanto riguarda le sequenze intergeniche, se si considera che sono costituite da spaziatori AT e nuclei di GC, come le sequenze *ori*, e che quattro su otto contengono una o due sequenze *ori*, è concepibile che esse sono derivate da sequenze *ori* attraverso un fenomeno di espansione. Questo può aver luogo grazie a tre meccanismi. Innanzitutto, uno slittamento della replicasi può avvenire a livello delle sequenze *ori*; questo è un fenomeno ben noto, studiato per la prima volta nella replicazione reiterativa del poli (dAT : dAT) da parte della polimerasi I di *E. coli* (KORNBERG et al., 1964). Un secondo meccanismo è il crossing-over disuguale, fenomeno molto frequente nel genoma mitocondriale (FONTY et al., 1978). Un terzo meccanismo è l'inserzione. Quasi tutti i nuclei di GC sono inseriti negli spaziatori AT, con rare eccezioni inserite in regioni ricche in AT di geni di RNA ribosomico (SOR e FUKUHARA, 1982) ed anche in un gene che codifica una proteina, *var 1* (HUDSPETH et al., 1982). È interessante che queste inserzioni sono non solo trascritte, ma anche, nel caso di *var 1*, tradotte. L'ipotesi della espansione (BERNARDI, 1982) è stata messa alla prova confrontando il livello di omologia esistente tra le sequenze *ori* e le sequenze intergeniche e quello esistente tra sequenze aleatorie aventi la lunghezza e la composizione delle sequenze *ori* e le sequenze intergeniche. Questo confronto, tra le sequenze *ori* o le sequenze aleatorie e circa 20.000 nucleotidi di sequenze intergeniche ha dimostrato che l'omologia era 10 volte più grande nel primo caso che nel secondo. Questo risultato è quindi in favore dell'idea che le sequenze intergeniche sono derivate da sequenze *ori* attraverso un processo di espansione.

Un altro risultato di questo confronto (BERNARDI, 1982) è che i quadri di lettura chiusi degli introni dei geni *oxi 3* e *cob* hanno le stesse caratteristiche delle sequenze non codificanti intergeniche. Questo può riflettere l'inserzione

di una sequenza intergenica in un introne o indicare che l'introne stesso è derivato da una sequenza intergenica che in passato conteneva una sequenza *ori*. La perdita di sequenze *ori* nei genomi mitocondriali di certi ceppi di lievito è documentata (FAUGERON-FONTY, com. pers.).

Il caso del gene *var 1* (HUDSPETH et al., 1982) è particolarmente interessante nel contesto dell'idea della espansione delle sequenze *ori*. Questo gene ha un contenuto in GC del 10% e contiene un nucleo di GC di 46 paia di basi che corrisponde a 38% di tutte le paia di basi dG : dC. La sua somiglianza con le sequenze spaziatori AT-nuclei di GC fa pensare che questo gene si è originato recentemente da una sequenza intergenica.

6. Sono le sequenze non codificanti del genoma mitocondriale del lievito sequenze di DNA egoista? Una ovvia domanda sollevata dalla presenza di una tal quantità di sequenze non codificanti nel genoma mitocondriale di lievito è se queste sono sequenze di DNA egoista. Il fatto che le sequenze mitocondriali han preso origine attraverso un meccanismo identico a quello postulato per il DNA egoista (ORGEL e CRICK, 1980) rende la domanda ancora più pertinente. La nostra comprensione della genetica molecolare dei mitocondri di lievito permette una discussione non puramente speculativa del problema. Le sequenze non codificanti del genoma mitocondriale del lievito presentano tre svantaggi per il genoma. Lo svantaggio principale è che le abbondanti ripetizioni dirette che esse contengono sono sequenze di escissione potenziali; in quanto tali, queste sequenze sono direttamente responsabili dell'estrema instabilità del genoma. Due altri svantaggi, meno gravi, sono che esse provocano un aumento del tempo di replicazione del genoma e un aumento del consumo di energia. Questi svantaggi porterebbero rapidamente alla trasformazione delle cellule di tipo selvaggio in petites neutre e soppressive, se questa selezione intracellulare non fosse controbilanciata da una selezione intercellulare in cui le cellule di tipo selvaggio, capaci di respirare e di crescere più rapidamente, eliminano le petites incapaci di respirare (questo spiega come in natura si trovino solo cellule di tipo selvaggio). Anche se gli svantaggi legati alle sequenze con codificanti non conducono all'eliminazione del genoma mitocondriale, dovrebbero almeno condurre all'eliminazione delle sequenze non codificanti. Noi sappiamo, tuttavia, che, sebbene esistano dei ceppi di *S. cerevisiae* che mancano di un certo numero di introni e anche della sequenza *ori 4*, in generale le sequenze non codificanti tendono ad essere conservate. Questo indica che l'eliminazione delle sequenze non codificanti è selettivamente svantaggiosa o, in altre parole, che le sequenze non codificanti forniscono dei vantaggi selettivi che compensano per gli svantaggi

che esse comportano. Ciò solleva ovviamente il problema della natura di questi vantaggi, cioè del ruolo fisiologico delle sequenze non codificanti.

È chiaro dalla carta della Fig. 1 che la escissione di sequenze intergeniche, in cui sono localizzate la maggior parte delle sequenze di escissione usate nella mutazione spontanea (DE ZAMAROCZY et al., 1983) eliminerà frequentemente delle sequenze *ori* canoniche dal genoma di tipo selvaggio. Anche se è stato trovato un genoma di tipo selvaggio privo di *ori 4*, è evidente che in generale questa eliminazione inciderà sulla replicazione e anche sulla trascrizione. Esiste quindi un vantaggio selettivo nel conservare le sequenze *ori* nel genoma di tipo selvaggio. Per quanto riguarda le sequenze non codificanti esterne alle sequenze *ori*, bisogna innanzitutto sottolineare che il processo di espansione non propaga sequenze senza senso, ma propaga invece sequenze che sono state altamente selezionate e conservate nella evoluzione e il cui ruolo primordiale è di interagire specificamente con gli enzimi attivi nella replicazione e nella trascrizione del DNA. Così, la espansione delle sequenze *ori* conduce alla propagazione di segnali di regolazione potenziali, che possono essere usati nella regolazione della espressione dei geni, nella repressione (la anaerobiosi o il glucosio possono arrestare la trascrizione) e nella derepressione, nella maturazione dei trascritti primari e nella regolazione delle interazioni nucleocitoplasmatiche. Un altro ruolo fisiologico delle sequenze non codificanti riguarda la ricombinazione. Esistono prove indirette (FONTY et al., 1978) che sequenze ripetute e palindromiche non codificanti intervengono nella ricombinazione mitocondriale sito-specifica. Infine, bisogna ricordare che alcune sequenze non codificanti sono inserite in geni trascritti o anche trasformate in geni. Insomma, diversi ruoli fisiologici possono essere dimostrati, o immaginati, per le sequenze non codificanti; questi ruoli sembrano fornire vantaggi selettivi capaci di compensare gli svantaggi dovuti alle sequenze non codificanti.

Un ultimo punto in cui il genoma mitocondriale di lievito può contribuire degli elementi di risposta al problema del DNA egoista è la presenza di genomi senza funzione nelle petites soppressive (vedi REID, 1980). Molti di questi genomi non contengono geni; ciò non ostante, la replicazione, la trascrizione e qualche volta anche la maturazione dei trascritti continuano. In natura, come è già stato detto, questi genomi spariscono rapidamente con l'eliminazione delle petites da parte delle cellule di tipo selvaggio. Molti di questi genomi hanno un tale vantaggio replicativo sui genomi di tipo selvaggio che potrebbero propagarsi attraverso incroci con cellule di tipo selvaggio, se aploidi. Ciò non avviene in natura tuttavia poiché le cellule parentali di tipo selvaggio e le petites derivate da loro hanno lo stesso tipo

sessuale. Quando sono isolate dalla competizione con le cellule di tipo selvaggio, in laboratorio, tuttavia le petites non solo sopravvivono, ma spesso finiscono per avere dei genomi difettivi molto stabili che sono il risultato di una selezione sulla base della efficienza di replicazione. Questi 'genomi senza geni' sono praticamente fatti di unità di ripetizione che contengono poco più di una sequenza *ori*; la replicazione di questi genomi è estremamente efficiente (le petites corrispondenti sono supersoppressive), e la trascrizione è mantenuta probabilmente a causa del suo ruolo nella replicazione. Questi 'genomi egoisti', esempi di sistemi autoreplicativi primordiali, saranno perduti solo dopo un tempo considerevole, quando mutazioni mitocondriali o nucleari avranno inattivato la iniziazione della replicazione. In conclusione, ciò che sappiamo sulle sequenze non codificanti del genoma mitocondriale di lievito suggerisce che la loro conservazione nel genoma è dovuta a vantaggi selettivi legati alle loro funzioni fisiologiche. Queste sequenze non possono perciò essere considerate sequenze di DNA egoista, almeno nella loro maggioranza. D'altro canto, genomi senza funzione come i genomi mitocondriali delle petites soppressive non solo possono esistere ed essere notevolmente stabili, ma subiscono una selezione che favorisce quelli che sono più prossimi alla situazione estrema di essere soltanto una serie di origini di replicazione. Questa selezione *in vivo* è molto simile a quella trovata nel sistema *in vitro* della Q β -replicasi (MILLS et al., 1967).

Parigi, 4 dicembre 1982

(Traduzione dell'autore)

REFERENZE

- ASADA K., SUGIMOTO K., OKA A., TAKANAMI M., HIROTA Y., 1982 — Structure of replication origin of the *Escherichia coli* K-12 chromosome: the presence of spacer sequences in the *ori* region carrying information for autonomous replication. *Nucleic Acids Res.*, **10**, 3745-3755.
- BALDACCI G., BERNARDI G., 1982 — Replication origins are associated with transcription initiation sequences in the mitochondrial genome of yeast. *EMBO J.*, **1**, 987-994.
- BERNARDI G., 1979 — The petite mutation in yeast. *Trends Biochem. Sci.*, **4**, 197-201.
- BERNARDI G., 1982 — Evolutionary origin and the biological function of non-coding sequences in the mitochondrial genome of yeast. In *Mitochondrial Genes* (SLONIMSKI P. P., et al., eds.), pp. 269-278, Cold Spring Harbor.

- BERNARDI G., BERNARDI G., 1980 — Repeated sequences in the mitochondrial genome of yeast. *FEBS Letters*, **115**, 159-162.
- CLAYTON D. A., 1982 — Replication of animal mitochondrial DNA. *Cell*, **28**, 693-705.
- COSSON J., TZAGOLOFF A., 1979 — Sequence homologies of (guanosine-cytidine) mitochondrial DNA of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **254**, 42-43.
- CREWS S., OYALA D., POSAKONY J., NISHIGUCHI J., ATTARDI G., 1979 — Nucleotide sequence of a region of human mitochondrial DNA containing the precisely identified origin replication. *Nature*, **277**, 192-198.
- DE ZAMAROCZY M., FAUGERON-FONTY G., BERNARDI G., 1983 — Excision sequences in the mitochondrial genome of yeast. *Gene* (in press).
- DE ZAMAROCZY M., MAROTTA R., FAUGERON-FONTY G., GOURSOT R., MANGIN M., BALDACCI G., BERNARDI G., 1981 — The origins of replication of the mitochondrial genome of yeast and the phenomenon of suppressivity. *Nature*, **292**, 75-78.
- DOOLITTLE W. F., SAPIENZA C., 1980 — Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**, 601-603.
- FONTY G., GOURSOT R., WILKIE D., BERNARDI G., 1978 — The mitochondrial genome of wild-type yeast cells. VII. Recombination in crosses. *J. Mol. Biol.*, **119**, 213-235.
- GAILLARD C., BERNARDI G., 1979 — The nucleotide sequence of the mitochondrial genome of a spontaneous 'petite' mutant of yeast. *Mol. Gen. Genet.*, **174**, 335-337.
- GOURSOT R., MANGIN M., BERNARDI G., 1982 — Surrogate origins of replication in the mitochondrial genome of *ori^o* petite mutants of yeast. *EMBO J.*, **1**, 705-711.
- HOBOM G., 1981 — Replication signals in prokaryotic DNA. In *Current Topics in Microbiology and Immunology* (W. HENLE, et al., eds.) **94/95**, 93-142, Springer-Verlag, Berlin.
- HUDSPETH M. E. S., AINLEY W. M., SHUMARD D. S., BUTOW R. A., GROSSMAN L. I., 1982 — Location and structure of the var I Gene on yeast mitochondrial DNA: nucleotide sequence of the 40.0 allele. *Cell*, **30**, 617-626.
- KORNBERG A., BERTSCH L. L., JACKSON J. F., KHORANA H. G., 1964 — Enzymatic synthesis of deoxyribonucleic acid. XVI. Oligonucleotides as templates and the mechanism of their replication. *Proc. Natl. Acad. Sci. USA.*, **51**, 315-323.
- MILLS D. R., PETERSON R. L., SPIEGELMAN S., 1967 — An extracellular darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. USA.*, **58**, 217-224.
- ORGEL L. E., CRICK F. H. C., 1980 — Selfish DNA: the ultimate parasite. *Nature*, **284**, 604-607.
- OSINGA K. A., TABAK H. F., 1982 — Initiation of transcription of genes for mitochondrial ribosomal RNA in yeast: comparison of the nucleotide sequence around the 5'-ends of both genes reveals a homologous stretch of 17 nucleotides. *Nucleic Acids Res.*, **10**, 3617-3626.
- REID R. A., 1980 — Selfish DNA in 'petite' mutants. *Nature*, **285**, 620.
- SORF., FUKUHARA H., 1982 — Nature of inverted sequence in the mitochondrial gene coding for the 15S ribosomal RNA of yeast. *Nucleic Acids Res.*, **10**, 1625-1633.