Atti Accademia Nazionale dei Lincei

Classe Scienze Fisiche Matematiche Naturali

Rendiconti

Francesco Amaldi, Pedro A. Lava-Sanchez, Mario Buongiorno-Nardelli

Redundancy of genetic information and evolution of living systems

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **58** (1975), n.6, p. 939–944.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1975_8_58_6_939_0>

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Accademia Nazionale dei Lincei, 1975.

SEZIONE III

(Botanica, zoologia, fisiologia e patologia)

Genetica. — *Redundancy of genetic information and evolution of living systems.* Nota di FRANCESCO AMALDI (*), PEDRO A. LAVA-SANCHEZ (*) e MARIO BUONGIORNO-NARDELLI (**), presentata (***) dal Socio G. MONTALENTI.

RIASSUNTO. — La reiterazione di alcuni geni nelle cellule eucariotiche è qui interpretata come una ridondanza informazionale in senso cibernetico, comparsa nel corso dell'evoluzione in concomitanza con la distinzione tra linea somatica e linea germinale. Questa ridondanza genica tamponerebbe gli effetti fenotipici delle mutazioni somatiche e la sua comparsa sarebbe stata quindi essenziale per l'evoluzione di organismi viventi complessi. Questo punto di vista implica una efficiente rettificazione dei geni ridondanti i cui possibili meccanismi sono qui discussi.

It is well known that in Eukaryotes the cellular DNA content varies, both in plants and animals, over a range of about 10^3 without any relation with variations of apparent developmental, morphological and physiological complexity of the organisms. One of the possible explanations of this fact was offered by Callan's hypothesis which assumes the existence of multiple copies of each gene tandemly arranged along the chromosomes [10, 25].

Reassociation kinetics studies of denaturated DNA revealed the presence in the genome of all Eukaryotes of a consistent fraction of DNA sequences repeated many times [6]. By this techniques the DNA can be generally resolved in three distinct components; highly repetitive, intermediate and unique sequences. The first one is not transcribed and cannot be considered to represent real genes but rather sequences involved in some special functions [27]. It is important to consider what is the relative arrangement within the genome of the other two components which are, at least in part, transcribed. It seems that a fraction of repeated DNA is composed of short sequences interspersed with unique sequences [13]. It is possible to envisage several reasons for such sequences to be scattered along the genome; they could represent recognition sites for factors and enzymes for replication, transcription, regulation etc. (see for instance the theory of Britten and Davidson [5]). Divergence of duplicated sequences or convergence of originally different ones could have originated, under selective pressure, such repeated sequences which are in fact rather similar than identical to each other.

(*) Centro Acidi Nucleici del C.N.R., Istituto di Fisiologia Generale, Università di Roma.

(**) Istituto di Istologia ed Embriologia, Facoltà di Scienze, Università di Roma.

(***) Nella seduta dell'11 giugno 1975.

On the other hand another fraction of the intermediate repeated DNA is surely composed of clusters of repeated sequences as suggested by " circularization " experiments [18] and, for some genes at least, clearly proved by RNA-DNA molecular hybridization. By this technique in fact it is possible to study individual genes. For many reasons those for ribosomal RNA have been the most deeply investigated in this last decade (for a review see [3]). Other genes received also some attentions, those for 5 s RNA, tRNAs, histones globin, fibroin and ovoalbumin [7, 12, 11, 17, 4, 15, 23, 14]. These studies suggest that in eukaryotic cells some genes are unique while others are present in multiple copies from few tens to several thousands. Furthermore it is known, most but not only from studies on rRNA genes, that the repeated genes tandemly arranged in clusters (gene families). At variance with the interspersed repeated sequences, the multiple gene copies within a clusters seem to be precisely repeated.

Gene Reiteration as informational redundancy

We will discuss now which could be the physiological role of gene reiteration and why, once established, it has been maintained through evolution, thus suggesting its indispensability for eukaryotic life. Its most obvious role might be to enhance the rate of gene product synthesis. In other words multiple gene copies do their job faster than a single one. This generally accepted interpretation is mainly based on the special case of the rRNA genes which provide the cell of an enormous amount of ribosomes and in fact have been found to be repeated in multiple copies in all eukaryotic cells.

As mentioned, some other genes have been studied for multiplicity; while those for 5 s RNA, tRNA and for histone mRNA have been demonstrated to be present in multiple copies [7, 12, 11, 17], the genes for globin, fibroin and ovoalbumin mRNAs are not reiterated [4, 15, 23, 14]. Surprisingly even in the differentiated cells specialized in the production of great amounts of the specific proteins, these genes are present in single or very few copies per haploid genome. On the other hand it is not clear why Eykaryotes, e.g. *Drosophila*, need hundreds of rRNA genes while *E. coli* can manage so well with very few ones, where genetic complexity (total number of different genes) is in the same order of magnitude. Thus it seems to us that to relate gene multiplicity only with quantitative requirements is a too simple interpretation not sufficiently supported by experimental evidences.

In consideration of the fact that gene reiteration appears in Eukaryotes where a " soma " is present, we think that this condition might represent a " redundancy " in informatic sense. It is well known that the malfunction or error probability of a machine is proportional both to the malfunction probabilities of the elements composing it and to the number of them. The way cybernetics bypass this difficulty to allow the construction of very

complex reliable machines is the reduction of error probability of the single elements and/or the introduction of redundancy for each element of the machine. It is thus possible to build reliable machines even with low reliability elements if the degree of redundancy is opportunely regulated.

An higher organism can be viewed as a complex system made up of many elements, the cells. These again are complex structures whose " reliability " depends on numerous functions genetically encoded in many genes. In this sense the organism is a machine for which, at various levels of its organisation, the laws of informatics and cybernetics apply. Random mutations occurring in the genome represent errors of the biological machine; here, at variance with the non biological machine, such errors are essential to occur in the germinal line to generate the variability which is the basis for selection and thus evolution. On the other hand mutations, when occurring in the somatic line, are either dangerous both for individual cells and for the whole organism or at least useless for evolution. Redundancy at different organisation levels minimizes the phenotypic effects of mutations occurring in the soma. Those functions necessary for the life of the organism as a whole are in fact performed from many equally differentiated cells; this represents a sort of " cell redundancy ". At another organisation level the enhancement of the reliability of the individual cells is attained by " gene redundancy ". Polyploidy and polyteny, which frequently occur in somatic cells, probably have this functional role. The macronucleus of Protozoa is another case of gene redundancy; in these organisms the " clone " corresponds to the soma of pluricellular organisms. The above discussed gene reiteration (tandemly arranged copies of a gene; rRNA, tRNA genes etc.) is a particular mechanism by which redundancy is introduced in the genome. Reiteration can be regulated at different levels for different genes according to their functions. In this context it is striking the fact that a particularly high level of gene reiteration has been demonstrated up to now for genes (rRNA, 5 sRNA, tRNAs and histone mRNAs) whose functions, besides being necessary to the life of all individual cells display pleiotropic effects. Only one mutation affecting one of the many components of the protein synthesis machinery could have deleterious effects on all synthesized proteins. The genes found to be not appreciably reiterated (those for globin, fibroin and ovoalbumin) are responsible for the synthesis of proteins which, being specific of differentiated states of the cells, are relevant mainly for the life of the organism as a whole. From our point of view these functions are sufficiently protected in the organism by the mentioned " cell redundancy " or at the gene level by diploidy itself and eventually by somatic polyploidy and polyteny.

Let's consider *ad absurdum* a multicellular organism with no redundancy, both at gene and cell levels. In this hypothetical organism each differentiated function, including reproductive one, whould be carried out by a single cell, whose genome would consist only of unique genes, as prokaryotic cells. The occurrence during somatic life of random mutations would make the soma of this organism not to be the true phenotypic counterpart of the

germinal line so that selection would eventually act paradoxically; it would select against a "bad" soma carrying a "good" germ cell or viceversa. Different is the case of Prokaryotes; here the somatic and germinal lines are not distinct and selection acts directly on individual cells. Gene redundancy is thus superfluous. On the contrary the existence in Eukaryotes of redundancy, buffering the effect of somatic mutations, guarantees that, in spite of them, the phenotype represents the original genotype of the individual through the lifetime. A quite similar interpretation of gene reiteration has been recently discussed in relation with the problem of ageing by Medvedev [19].

Genome size and evolution

Let's consider now the variation of genome size (DNA content per genome) during evolution. It has been often observed that the organisms with a particularly high DNA content occupy peculiar phylogenetic positions. For example among Vertebrates, the Dipnoi and the Urodeles which have the highest DNA content per cell, are representative of those ancestor Vertebrates which signed the macroevolutive change from aquatic to terrestrial life. An interpretation [20, 16] attributes to the higher DNA content, attained by gene duplication followed by divergence, the value of a reservoir of rough material for new gene production. (It is clear that gene duplication is different from gene reiteration, the difference consisting in divergence due to lack of rectification as discussed below).

We think that besides gene duplication, true gene redundancy may have also some relevance in the understanding of the DNA content paradox. The variability, basis for darwinian evolution, is introduced by random mutational changes of genetic material. The rate of this random events is known to depend not uniquely from external factors, but to be largely due to genetically specified factors such as DNA-polymerases accuracy, repair systems efficiency and other "mutator genes". In other words mutation rate, as a real genetic trait, must be itself under selective pressure. Organisms with a high mutation rate are advantaged in those evolutive situations which require an higher variability, as for instance the ancestors of terrestrial Vertebrates to which we referred above. In such organisms the higher mutability, required for evolutive purposes, must be accompanied by an increased protection of the soma against mutations. This is obtained by a higher gene redundancy which will contribute to the increase of the DNA content per genome.

Experimental evidences showing a rough correlation between genome size and mutation rate per locus have been recently discussed [1]. Following our above presented hypothesis, the increased mutation rate per locus would not be a consequence of the increased genome size but rather the contrary; genome size would increase as a protective adaptation to the higher mutation rate necessary for evolution, obviously both parameters changing under selective pressure.

RECTIFICATION

If the role of gene reiteration is to stabilize the phenotypic expression of some genetic information during somatic life, the occurrence of a periodic efficient rectification event is compellent. Gene rectification proposed by Callan [10] and so named by Thomas [25], is any process able to keep all the members of a gene family equal to each other, allowing though the possibility of evolutive changes for the family as a whole.

The concept of rectification, which can be compared with the " horizontal evolution " discussed by Brown [8] in opposition to the " vertical evolution ", must be accepted, while the mechanism(s) by which it is attained and its relationship with meiotic recombination [28] are still matter of discussion. Several alternative models have been proposed; some of them involve a correction of the multiple gene copies by comparison, as the original " master-slaves " theory by Callan [10] and others derived from it [26]. Another suggested mechanism is the " contraction-expansion " of the cluster attained by unequal crossing-over (e.g. the deletion-magnification of rRNA genes in Drosophila), which would result in a rectification at low efficiency [21, 24, 22]; that is a mutation could be spread to the whole cluster only after many generations. Moreover, neutral mutations, not submitted to selective pressure, would not be rectified. Finally a third kind of models is based on the periodic substitution of the whole cluster of a reiterated gene with a set of new genes produced by a series of replication of a single one [9, 2] resulting in a very efficient rectification. The most suited moment for such rectification to occur is meiosis which thus would have another basic function, that is to " rejuvenate " the " aged " families of reiterated genes.

Note added in proof: following submission of this paper, we became aware of a paper by Jukes and Gatlin (Progr. Nucl. Ac. Res. and Mol. Biol., (1971) 11, 303) in which a cybernetic intepretation of genetic redundancy was also proposed.

REFERENCES

[1] ABRAHAMSON S., BENDER M. A., CONGER A. D. and WOLFF S. (1973) – *Uniformity of radiation-induced mutation rates among different species*, « Nature », *245*, 460.

[2] AMALDI F., LAVA-SANCHEZ P. A. and BUONGIORNO-NARDELLI M. (1973) – *Nuclear* DNA *content variability in* Xenopus laevis: *a redundancy regulation common to all gene families*, « Nature », *242*, 615.

[3] BIRNSTIEL M. L., CHIPCHASE M. and SPEIRS J. (1971) – *The ribosomal cistrons*, « Progress in Nucleic Acid Research and Molecular Biology » (J. N. Davidson and W. E. Cohn eds., New York: Acad. Press), *11*, 351.

[4] BISHOP J. O., PEMBERTON R. and BAGLIONI C. (1971) – *Reiteration frequency of haemoglobin genes in the duck*, « Nature New Biol. », *235*, 231.

[5] BRITTEN R. J. and DAVIDSON E. H. (1969) – *Gene regulation for higher cells: a theory*, « Science », *165*, 349.

[6] BRITTEN R. J. and KOHNE D. E. (1968) – *Repeated sequences in* DNA, « Science », *161*, 529.

[7] BROWN D. D. and SUGIMOTO K. (1973) – 5s DNAs *of* Xenopus laevis *and* Xenopus mulleri: *evolution of a gene family*, « J. Mol. Biol. », *78*, 397.

[8] BROWN D. D., WENSINK P. C. and JORDAN E. (1972) – *A comparison of the ribosomal* DNA's *of* Xenopus laevis *and* Xenopus mülleri: *the evolution of tandem genes*, « J. Mol. Biol. », *63*, 57.

[9] BUONGIORNO-NARDELLI M., AMALDI F. and LAVA-SANCHEZ P. A. (1972) – *Amplification as a rectification mechanism for the redundant* rRNA *genes*, « Nature New Biol. », *238*, 134.

[10] CALLAN H. G. (1967) – *The organization of genetic units in chromosomes*, « J. Cell. Sci », *2*, 1.

[11] CLARKSON S. G., BIRNSTIEL M. L. and PURDOM I. F. (1973) – *Clustering of transfer* RNA *genes of* Xenopus laevis, « J. Mol. Biol. », *79*, 411.

[12] CLARKSON S. G., BIRNSTIEL M. L. and SERRA V. (1973) – *Reiterated transfer* RNA *genes of* Xenopus laevis, « J. Mol. Biol. », *79*, 391.

[13] DAVIDSON E. H., HOUGH B. R., AMENSON C. S. and BRITTEN R. J. (1973) – *General interspersion of repetitive with non-repetitive sequence elements in the* DNA *of* Xenopus, « J. Mol. Biol. », *77*, 1.

[14] HARRIS S. E., MEANS A. R., MITCHELL W. M. and O'MALLEY B. W. (1973) – *Synthesis of* (^{3}H) DNA *complementary to ovalbumin messenger* RNA: *evidence limited copies of the ovalbumin gene in chick oviduct*, « Proc. Natl. Acad. Sci. US », *70*, 3776.

[15] HARRISON P. R., HELL A., BIRNIE G. D. and PAUL J. (1972) – *Evidence for single copies of globin genes in the mouse genome*, « Nature », *239*, 219.

[16] HINEGARDNER R. and ROSEN D. E. (1972) – *Cellular* DNA *content and the evolution of teleostean fishes*, « Amer. Natur. », *106*, 621.

[17] KEDES L. H. and BIRNSTIEL M. L. (1971) – *Reiteration and clustering of* DNA *sequences complementary to histone messenger* RNA, « Nature New Biol. », *230*, 165.

[18] LEE C. S. and THOMAS C. A. (1973) – *Formation of rings from* Drosophila DNA *fragments*, « J. Mol. Biol. », *77*, 25.

[19] MEDVEDEV ZH. A. (1972) – *Possible role of repeated Nucleotide sequences in* DNA *in the evolution of life spans of differentiated cells*, « Nature », *237*, 453.

[20] OHNO S., WOLF U. and ATKIN N. B. (1968) – *Evolution from fish to mammals by gene duplication*, « Hereditas », *59*, 169.

[21] RITOSSA F. (1973) – *Crossing-over between* X *and* Y *chromosomes during ribosomal* DNA *magnification in* Drosophila melanogaster, « Proc. Natl. Acad. Sci. US. », *70*, 1950.

[22] SMITH G. P. (1973) – *Unequal crossover and the evolution of multigene families*, « Cold Spring Harbor Symp. Quant. Biol. », *38*, 507.

[23] SUZUKI Y., GAGE L. P. and BROWN D. D. (1972) – *The genes for silk fibroin in* Bombyx mori, « J. Mol. Biol. », *70*, 637.

[24] TARTOF K. D. (1973) – *Unequal mitotic system chromatic exchange and disproportionate replication as mechanisms regulating ribosomal* RNA *gene redundancy*, « Cold Spring Harbor Symp. Quant. Biol. », *38*, 491.

[25] THOMAS C. A. (1970) – *The theory of the master gene*, « The neurosciences » (F. D. Schmitt ed., New York: The Rockefeller University Press), p. 973.

[26] THOMAS C. A. (1973) – *The rolling helix: a model for the eukaryotic gene*, « Cold Spring Harbor Symp. Quant. Biol. », *38*, 347.

[27] WALKER P. M. B., FLAMM W. G. and MCLAREN A. (1969) – *Highly repetitive* DNA *in rodents*, « Handbook of molecular cytology » (A. Lima de Faria ed., Amsterdam: North Holland), p. 53.

[28] WHITEHOUSE H. L. K. (1973) – *Hypothesis of post-recombination resynthesis of gene copies*, « Nature », *245*, 295.