SO MUCH "JUNK" DNA IN OUR GENOME

Susumu Ohno
City of Hope National Medical Center

The mammalian genome (haploid chromosome complement) contains roughly 3.0×10^{-9} mg of DNA which represents about 3.0×10^{9} base pairs. This is at least 750 times the genome size of *E. coli*. If we take the simplistic assumption that the number of genes contained is proportional to the genome size, we would have to conclude that 3 million or so genes are contained in our genome. The falseness of such an assumption becomes clear when we realize that the genome of lowly lungfish and salamanders can be 36 times greater than our own (Ohno and Atkin, 1966).

In fact, there seems to be a strict upper limit for the number of gene loci which we can afford to keep in our genome. Consequently, only a fraction of our DNA appears to function as genes. The observations on a number of structural gene loci of man, mice and other organisms revealed that each locus has a 10^{-5} per generation probability of sustaining a deleterious mutation. It then follows that the moment we acquire 10^{5} gene loci, the overall deleterious mutation rate per generation becomes 1.0 which appears to represent an

unbearably heavy genetic load. Taking into consideration the fact that deleterious mutations can be dominant or recessive, the total number of gene loci of man has been estimated to be about 3×10^4 (Muller, 1967; Crow and Kimura, 1970). Even if an allowance is made for the existence in multiplicates of certain genes, it is still concluded that, at the most, only 6% of our DNA base sequences is utilized as genes (Kimura and Ohta, 1971). Aside from conventional structural genes and regulatory genes, this 6% should include the *promotor* region and *operator* region which are situated adjacent to each structural gene, for these regions can certainly sustain deleterious mutations. More than 90% degeneracy contained within our genome should be kept in mind when we consider evolutional changes in genome sizes. What is the reason behind this degeneracy?

Certain untranscribable and/or untranslatable DNA base sequences appear to be useful in a negative way (the importance of doing nothing). If functional genes customarily occupied the region around the centromere, evolutional changes of chromosome complements would not have occurred as often as has been observed. Because the centromeric heterochromatin which represents a long tandem repeat of a short untranscribable sequence (Jones, 1970; Southern, 1970) can be lost or duplicated without deleterious consequences, speciation more often than not has been accompanied by chromosomal changes. The same can be said of those DNA base sequences which are used as partitions between the genes. It may be of selective advantage to space adjacent genes far enough apart by inserting a stretch of untranscribable and/or

untranslatable DNA base sequence as a partition. In this way, the deleterious effect of *nonsense* or *frame-shift* mutations can be confined to a single locus, instead of allowing it to spread to other genes. Indeed, Miller and Beatty (1969) have shown long partitioning sequences between genes for 18S and 28S ribosomal RNA of the nucleolar organizing region. Furthermore, the recent recovery of longer than usual human hemoglobin α- and β-chains (Milner *et al.*, 1971; Flatz *et al.*, 1971) can be interpreted to mean that the hemoglobin α- as well as β-chain gene is followed by a partitioning sequence and that a mutation can bring about a partial translation of a normally silent partitioning sequence.

Inasmuch as the only requirement to be qualified as partitioning sequences is to be untranscribable and/or untranslatable, it is not likely that these sequences came into being as a result of positive selection. Our view is that they are the remains of nature's experiments which failed. The earth is strewn with fossil remains of extinct species; is it a wonder that our genome too is filled with the remains of extinct genes?

So long as a particular function is assigned to a single gene locus in the genome, those mutations which affect the *active* site of a gene product would produce deleterious consequences and, therefore, be eliminated by natural selection. Thus, natural selection permits only trivial changes, while forbidding changes in the basic character of a gene. Indeed, the amino acid sequences of fibrinopeptide A and B in which only C-terminal arginine appears to

represent a real *active* site evolved 1500 times faster than histone IV where the *active* site appears to consist of an entire molecule (Kimura and Ohta, 1971).

It then follows that the creation of a new gene with hitherto nonexistent function is possible only if a gene becomes sheltered from relentless pressure of natural selection. This shelter has apparently been provided either by polyploidization or by tandem duplication. Redundant copies of genes thus produced are now free to accumulate formerly forbidden mutations and thereby to acquire new functions (Ohno, 1970).

The chance of acquiring a new function by unrestricted accumulation of mutations, however, should be as small as that of an isolated population emerging triumphant as a new species. Degeneracy is the more likely fate. The creation of every new gene must have been accompanied by many other redundant copies joining the ranks of silent DNA base sequences, and these silent DNA base sequences may now be serving the useful but negative function of spacing those which have succeeded. Triumphs as well as failures of nature's past experiments appear to be contained in our genome.

References

1. S. Ohno and N.B. Atkin, Chromosoma 18, 455 (1966).
2. H.J. Muller, in Heritage from Mendel (University of Wisconsin Press, Madison, 1967) pp. 419-447.
3. F. Crow and M. Kimura, An Introduction to Population Genetics Theory (Harper & Row, New York, 1970).
4. M. Kimura and T. Ohta, Nature 229, 467 (1971).

5. K.W. Jones, Nature 225, 912 (1970).

6. E.M. Southern, Nature 227, 794 (1970).

7. O.L. Miller, Jr. and B.R. Beatty, Science 164, 955 (1969).

8. P.F. Milner, J.B. Clegg and D.J. Weatherall, Lancet I, 729 (1971).

9. G. Flatz, J.L. Kinderlerer, J.V. Kilmartin and H. Helmann, Lancet I, 732 (1971).

10. S. Ohno, Evolution by Gene Duplication (Springer-Verlag, Heidelberg, 1970).

DISCUSSION

BOYER: The calculation of the permissible number of structurally active loci (a la King and Jukes) from lethal mutation rates depends on how well the value of 10^{-5} mutations/locus/generation represents the whole of the genome. We only measure what we see. Immutable or nearly immutable loci are not examined. We don't yet know the real proportions. It thus seems to me that the permissible number of structural loci is - as yet - a somewhat suspect way to arrive at figures of 1% structural utility to 99% junk.

OHNO: Although deleterious mutations affecting such gene products as actin, myosin and microtubule protein have not been detected, I believe that this is simply because these mutations are early embryo lethals. The estimate has been made on inbred strains that the mouse genome contains 10^4 gene loci which mutate at the rate of 10^{-5}/locus/generation to recessive lethals. Furthermore, if immutable gene loci really exist, by their very definition, they would not have contributed to evolution.

EVOLUTION OF GENETIC SYSTEMS

Editor
H. H. SMITH

Editorial Committee
H. J. PRICE
A. H. SPARROW
F. W. STUDIER
J. D. YOURNO

BIOLOGY DEPARTMENT
BROOKHAVEN NATIONAL LABORATORY
UPTON, NEW YORK

GORDON AND BREACH
New York • London • Paris

(1972).