(1973).
In: Modern Aspects of Cytogenetics: Constitutive Heterochromatin in Man. R.A. Pfeiffer, ed. F.K. Schattauer Verlag, Stuttgart, Germany, pp. 169-173.

Evolutional Reason for Having so Much Junk DNA*

S. Ohno

Over the years, I have learned that there is no such thing as a fact. What passes for a fact is in truth a set of observations and its interpretation. Therefore, the interpretation is just as important to a fact as the observation itself.

I have also learned that a discussion tends to be more fruitful if each person polarizes his view to an extreme; not because he is an extremist, but merely for the sake of discussion. Even in science, the adversary system seems the best way to bring out a glimpse of truth.

The points I wish to make are: 1) Natural selection is an extremely conservative force. So long as a particular function is assigned to a single gene locus in the genome, natural selection only permits trivial mutations of that locus to accompany evolution. 2) Only a redundant copy of a gene can escape from natural selection and while being ignored by natural selection can accumulate meaningful mutation to emerge as a new gene locus with a new function. Thus, evolution has been heavily dependent upon the mechanism of gene duplication. 3) The probability of a redundant copy of an old gene emerging as a new gene, however, is quite small. The more likely fate of a base sequence which is not policed by natural selection is to become degenerate. My estimate is that for every new gene locus created about 10 redundant copies must join the ranks of functionless DNA base sequence. 4) As a consequence, the mammalian genome is loaded with functionless DNA which is the subject of this meeting. The only contributions which these base sequences make may be stated as »the importance of doing nothing«.

Conservative nature of natural selection and an escape from natural selection as a prerequisite of evolution

The complete amino acid sequences of increasing numbers of polypeptide chains have become known in recent years. This has given us the chance to look at the direct products of genes and has further enabled us to deduce the evolutional history of individual gene loci. The extremely conservative nature

* Supported in part by a grant (CA 05138) from the National Cancer Institute, U.S. Public Health Service.

(+ discussion at back).

of natural selection became immediately apparent. For example, there is the fact that histone IV (110 amino acid residues) of cattle and garden peas differ from each other only by two amino acid substitutions (5, 7). It appears that an entire molecule of histone IV represents a functionally critical active site, so that almost any amino acid substitution makes a mutant protein delinquent in the performance of its assigned function which is to attach itself side-by-side to a DNA strand. Because of such a deleterious effect, natural selection has eliminated almost all the bearers of mutations affecting this gene locus since the creation of eukaryotes. On the other hand, fibrinopeptides A (15 to 19 amino acid residues) and B (14 to 21 residues) are the fastest evolving of all the known peptide chains (2, 4). During blood clot formation, fibrinopeptides A and B are split off from the inert fibrinogen molecule by the trypsin-like action of thrombin. Their role being passive, most amino acid substitutions would be harmless as only the carboxyl terminal arginine and a few other sites are functionally critical (10).

So long as a particular function is assigned to a single gene locus in the genome, natural selection never permits character-changing mutations to affect that locus, for such mutations are deleterious to individuals which bear them. If continuously surveyed by natural selection, a gene which began as a pyruvate kinase gene would forever remain a pyruvate kinase gene. How then was the evolution from prokaryotes to eukaryotes and from fish to mammals made possible? We come to realize that it is the periodical escape from natural selection which is *conditio sine qua non* of evolution. This escape is provided for a redundant copy of a gene created by gene duplication. There is little doubt that the creation of many new gene loci via gene duplication through polyploidy as well as unequal crossing-over contributed greatly to evolution from fish to mammals (Ohno, 1970).

For example, most of the secreted vertebrate proteins, although widely divergent in function, appear to have originated from a few digestive enzymes secreted by entoderm of earlier and simpler forms of life (1). Not only are the amino acid sequences of different peptide hormones similar, but also there exists recognizable homology in the amino acid sequences even between growth hormone, on one hand, and trypsin and chymotrypsin, on the other.

The fate of a redundant copy: A new gene or degenerate sequence

In the absence of natural selection, formerly forbidden mutations affecting the active site sequence can accumulate in a redundant copy. Thus, it has a chance of becoming a new gene with a previously nonexistent function. A new function is contingent upon the acquisition of a new active site sequence. The

chances of a redundant copy emerging triumphant as a new gene, however, are exceedingly small. The more likely fate which awaits a redundant copy is degeneracy.

Of randomly sustained mutations, a *frameshift* mutation due to deletion or addition of a single or two consecutive DNA bases would almost surely lead to degeneracy of a redundant copy, since the complete alteration of the amino acid sequence of a gene product is a consequence. The three consequences of a mutational single DNA base substitution are *samesense*, *missense* and *nonsense*. These three types should occur in the ratio of 6:17:1 (15). Therefore, there is one in 24 chance that a first bases substituting mutation sustained by a redundant copy would be a chain-terminating *nonsense* mutation which would again lead to degeneracy.

In order to avoid degenerating into a worthless DNA base sequence, a redundant copy must acquire a new function useful to the well-being of an organism in a relatively short period and thus again come under the protection of natural selection. It is natural selection which protects a newly emerged gene locus by preventing further accumulation of randomly sustained mutations. It is in this role that natural selection becomes an active contributor to evolutional change instead of being a conservative force. Since a redundant copy cannot be long ignored by natural selection, however, a group of genes which shared a common ancestry, as a rule, show only a limited degree of functional divergence; i.e., myoglobin and hemoglobin genes (8).

From the above, we come to realize that while temporary escape from natural selection provided by the mechanism of gene duplication was *conditio sine qua non* of evolution, the process of acquiring new gene loci was a rather expensive one. For every redundant copy which has succeeded in becoming a new gene, there must have been a great number which degenerated. My estimate of the success/failure ratio is $^1/_{10}$. It follows then that doubling of the number of gene loci may have been accompanied by 10-fold increase in the genome size (DNA content).

Functionless DNA and the importance of doing nothing

The realistic number of functionally significant gene loci in the mammalian genome has been estimated to be $4–5 \times 10^4$ (3, 11). This many gene loci account for no more than 2% of the total genomic DNA (3×10^9 base pairs). Although *ribosomal* and *transfer* RNA genes which exist in multiple copies have to be added to the above number, it indeed appears that the bulk of our genomic DNA has no definable function in that these sequences are either never transcribed or if transcribed are not translated to meaningful amino acid

sequences. Accordingly, these sequences are undergoing very rapid evolutional changes by unrestricted accumulation of randomly sustained mutations (14).

A functional gene locus is defined as that DNA base sequence which may sustain deleterious mutations. A DNA base sequence in which all sorts of mutational changes are permissible is obviously not contributing to the well-being of an organism, and for this very reason, it has no function. Attempts have often been made in the past to attribute a vaguely defined regulatory function to apparently functionless DNA base sequence. The function of a regulatory base sequence (*operator*), however, is to offer a binding site to a regulatory gene product (repressor or activator protein or RNA); thus, it can not afford to accumulate randomly sustained mutations. Mutational changes in an *operator* base sequence which lead to either increase or decrease in its binding affinity to a regulatory gene product would be deleterious and therefore be eliminated by natural selection. Since the simultaneous occurrence of two independent mutations one affecting the regulatory gene locus and the other affecting the *operator* base sequence is mathematically improbable, a regulatory base sequence would be conserved by natural selection to the same extent as any respectable structural gene base sequence. Rapidly evolving DNA base sequences cannot possibly have regulatory functions of any kind.

Deleterious mutations randomly affect all the functional gene loci at the spontaneous rate of 10^{-5} per locus per organism generation. Thus, there is a finite upper limit to the number of functional gene loci an organism can afford to keep. This is what HALDANE (6) termed the cost of natural selection. 5×10^4 gene loci per genome is permissible since the overall deleterious mutation rate per generation is 0.5. On the other hand, 3×10^6 gene loci per genome would surely lead to the extinction of a species for the overall deleterious mutation rate becomes 30. If any of the 30 is dominant, it would immediately lead to the extermination of an affected individual, while recessive deleterious mutations would accumulate in the genome generation after generation at an alarming rate until the final extinction of a species. No matter how we look at it, we come to the inevitable conclusion that the bulk of DNA in the mammalian genome must have no function aside from that of »the importance of doing nothing«.

If functional genes customarily occupied the region around the centromere, the Robertsonian type of evolutional changes of chromosome complements would not have occurred as often as has been observed. Because the centromeric heterochromatin which represents long tandem repeats of a short sequence (9, 13) can be partially lost or duplicated without deleterious consequences, speciation more often than not has been accompanied by chromosomal changes. Functionless DNA base sequences in the euchromatic region which space adjacent functional genes apart may also be useful in a negative way.

Chromosome breaks occur spontaneously at a high rate (10^{-2} per cell generation or so); rejoining of broken ends may be accompanied by deletion, addition and substitution of a few base pairs. Yet, they are harmless so long as a break occurs at a partitioning functionless base sequence.

If $4–5 \times 10^4$ functional gene loci are neatly packed into a proportionally small genome, mammalian cells may divide as fast prokaryotic cells which would not be conducive to organogenesis during embryonic development or to the relatively long life span of individuals. The bulk of functionless DNA in the mammalian genome may serve as a damper to give a reasonably long cell generation time (12 hours or so instead of several minutes).

References

(1) Adelson, J. W.: Enterosecretory proteins. Nature *229:* 321 (1971).
(2) Blombäck, B., M. Blombäck, N. J. Grondahl: Studies on fibrinopeptides from animals. Acta Chem Scand. *19:* 1789 (1965).
(3) Crow, F., M. Kimura: An Introduction to Population Genetics Theory. New York, Harper & Row 1970.
(4) Dayhoff, M. O. (ed.): Atlas of Protein Sequence and Structure; Vol. 4. National Biomedical Research Foundation, Silver Springs, Maryland, 1969.
(5) Delange, R. L., D. M. Fambrough: Identical COOH-terminal sequences of an arginine-rich histone from calf and pea. Fed. Proc. *27:* 392 (1968).
(6) Haldane, J. B. S.: The cost of natural selection. J. Genet. *55:* 511 (1957).
(7) Hunt, L. T.: Histone IV – Bovine and pea. In: Atlas of Protein Sequence and Structure; p. 190. National Biomedical Research Foundation, Silver Springs, Maryland, 1969.
(8) Ingram, V. M.: The Hemoglobin in Genetics and Evolution. Columbia University Press, New York 1963.
(9) Jones, K. W.: Chromosomal and nuclear location of mouse satellite DNA in individual cells. Nature *225:* 912 (1970).
(10) Kimura, M., T. Ohta: Protein polymorphism as a phase of molecular evolution. Nature *229:* 467 (1971).
(11) Muller, H. J.: The gene material as the initiator and the organizing basis of life. In: Heritage from Mendel. (R. A. Brink, ed.); p. 419. The University of Wisconsin Press, Madison 1967.
(12) Ohno, S.: Evolution by gene duplication. Springer, Heidelberg 1970.
(13) Southern, E. M.: Base sequence and evolution of guinea-pig α-satellite DNA. Nature, *227:* 794 (1970).
(14) Walker, P. M. B.: How different are the DNA's from related animals? Nature *219:* 228 (1968).
(15) Whitfield, H. J., R. G. Martin, B. N. Ames: Classification of aminotransferase (C gene) mutants in the histidine operon. J. molec. Biol. *21:* 335 (1966).

Discussion

OHNO: Dr. PARDUE, *messenger* RNA for immunoglobulin light chain as well as for hemoglobin α- or β-chain appears to be at least twice longer than the length expected from the polypeptide chain length. It appears as though mammalian *messenger* RNA's contain a great deal of untranslatable *junk* sequence, only a part of which is poly-A. Are histone *messengers* considerably longer than 330-bases?

PARDUE: Our calculations of the histone messenger size are compatible with the size of the histone proteins. I should say one thing which I forgot and that is that the *Drosophila* histones are unusually different from the histones of other species. COHEN and GOTCHEL[1]) have compared *Drosophila* histones from birds, mammals and amphibia, and find that there are at least three of the *Drosophila* histones which show a quite different electrophoretic pattern. They have been able to show that this difference in electrophoretic mobility is caused both by changes in size and by changes in some of the aminoacid content. Therefore it is almost certain that we are not looking at complete cross-hybridization of the sea-urchin histone message to *Drosophila* DNA.

YUNIS: Is there a poly-A sequence for any of the histone messages?

PARDUE: No, there is no poly-A sequence in the histone message. At least it has'nt been found yet. We don't know whether the message, which we see binding to that one region represents all five of the histone messages or whether perhaps only one of the messages is actually cross-hybridizing. We are in the process of characterizing the hybrid. Also we would like to measure the amount of *Drosophila* DNA, which actually codes for histones. From our experiments, making a lot of assumptions, we can say that the amount of DNA which we think might be necessary to code for all of the histones is enough DNA to make up one large band in *Drosophila*. These bands are a lot larger than average and might accomodate the spacer also.

---

[1]) J. Biol. Chem. *246:* 1841 (1971).

question, I would say that we have not yet looked at the centromeric regions of cattle chromosomes under the E. M., but we would expect to observe a failure of dye uptake rather than a loss of material following the ASG schedule. I say this since there certainly is no evidence of a loss from the positive staining results obtained in these regions with C-banding techniques.

EVANS: Dr. OHNO, I suppose the take-home message was that in your view certainly the highly repetitive DNA is junk and you said that lots of the unique sequences are also junk. This point of view is now open to discussion.

YUNIS: I wonder if you really mean »junk«. You are equating non-translational and non-transcriptional DNA with junk. I agree that you must be right up to some extent, but I wonder whether you have ignored the proven polyploidization as a way of evolution.

OHNO: If there is any gene which is doing some good for your general well-being, you will suffer when you lose that gene. For this very reason a fraction of randomly sustained mutations of that locus would be deleterious. There is simply no way of having a useful gene without paying a certain price for the cost of natural selection. If, on the other hand, there is a gene which is totally irrelevant, you will lose that gene sooner or later, for natural selection would not police that gene.

YUNIS: We know that constitutive heterochromatin is rich in repetitive DNA and that satellite DNA spaces essential regions such as the centromere and nucleolar organizer. Isn't this an important role?

OHNO: Yes, spacer is important in the same negative way as fibrinopeptides A and B. Only a short stretch of base sequence at its end would have to be conserved as a signal to be nicked by ribonuclease.

HENNIG: I feel one could accept to some extent both views. From all what is known so far we can conclude that probably the nucleotide sequence as such does not matter. Furthermore the actual amount of simple sequence DNA (within some limits) seems not to be important. But since this kind of DNA is there one has to correlate it to some kind of function. That means that either simply *the presence* of some portion of this material is essential for structural or functional reasons. Or one could imagine that this kind of DNA is a product of certain molecular mechanisms which as such are essential for the eukaryote genome. This, of course, should be some kind of multiplication mechanism.

Such a mechanism may, for example, exist in order to keep ribosomal cistrons or 5S cistrons etc. by occasional multiplications alike. Such multiplication steps could, accidentally or not, include adjacent DNA sequences, thus producing simple sequence DNA. It is remarkable that nucleoli usually are associated with heterochromatin which contains simple sequence DNA. Also the histone genes in *Drosophila* seem to be associated with heterochromatin as Dr. PARDUE has shown.

YUNIS: It is important to emphasize that, in general, there is a certain constancy in the amount and distribution of satellite DNA.

HENNIG: It is certainly true that there is a constancy in certain limits of simple sequence DNA. But these limits could be governed by simple mechanical requirements of the chromosomes, for example in segregation of the chromosomes. Extremely large chromosomes do have difficulties during cell division and thus an upper limit could be introduced by the size of chromosomes. I think the variability of the heterochromatic arm of the X chromosome of the *Drosophila* species, which we are studying, is a good example for a block of simple sequence DNA which seems not to be essential. Deletion of the long heterochromatic arm of the X in *D. hydei* has no obvious consequences for the flies. This could mean that there is some DNA which may not be necessary but is there and is kept. Of course, this deletion stock has not been tested for its success in a population in competition with the wild type.

FORD: I think the word »junk« is a powerful word. The only thing that I would seriously question is this assumption that perhaps just 10 units would be sufficient to act as a spacer when 100 or more were there. Do we really know enough to be sure of that point?

OHNO: If we argue that a given spacer can change its base sequence any way it likes, but that a length of it has to be conserved rigidly, deletions would become deleterious to spacer function. It follows then that spacer sequences, too, contribute to the overall mutation load, and for this very reason, we cannot even afford to keep too many spacers.

FORD: I think it just would'nt be there unless it would do it. Something was a functional reason of some kind for it.

YUNIS: This is what I emphasized earlier, that this DNA must have a functional value since nothing is known so widespread and universal in nature that has proven useless.

FRACCARO: Well, there is an exception to that rule. A lot of us have permanent positions at the University but are considered by others (mainly by students) meaningless and of no utility whatsoever.

EVANS: Well with that very splendid comment I think we should now draw today's discussion to a close. I should remind you, however, that we have at least a full hour available tomorrow to continue this discussion and would like to end by thanking all the speakers for their excellent presentations and the discussants for taking part in the discussion.