Scientific Insights into the Evolution of the Universe and of Life
Pontifical Academy of Sciences, Acta 20, 2009
www.pas.va/content/dam/accademia/pdf/acta20/acta20-bernardi.pdf

THE ROLE OF CHANCE IN EVOLUTION

GIORGIO BERNARDI

I would like to start this contribution on a personal note by mentioning that I come from one of the few, perhaps the only Institute in the world, the Stazione Zoologica of Naples, which was established in order to prove a theory, in our case Darwin's theory (1). After its foundation by Anton Dohrn in 1873, investigations at the Stazione concentrated on what was possible to investigate at that time, namely the morphology, the physiology and the embryology of marine organisms, their great biodiversity being the main reason for the choice of Naples as the seat of the Institute. For a century after the death of Anton Dohrn in 1909 practically no work on evolution was done. At the beginning of 1998 I took the direction of the Stazione Zoologica and started a Laboratory of Molecular Evolution which still is very active. I will report here on our work on genome evolution and its general implications.

The role of chance in evolution

The first question one may raise about the role of chance in evolution is why this issue is so important. One may think about a number of explanations, but I prefer here to use a shortcut, by concentrating on the position presented in 1970 by Jacques Monod in his famous book *Le hasard et la nécessité* (2). There are three main reasons for this choice. The first one is the clarity of the ideas, the second the extreme stand and the third the discussion of its implications. These points make it easier to understand the problem under consideration here. Some key sentences clearly summarize the stand of the author: (i) *The origin of life on earth was due to a single chance event* and, since all living organisms descend from a common ancestor, (ii) 'the biosphere is completely separated from the inanimate environment', and 'Man knows to be alone in the indifferent immensity of the Universe, from which he emerged

by chance'. As far as the evolution of living organisms was concerned, Monod expressed the opinion that (iii) 'Mutations are accidents that happen at random. Since they represent the only source of changes in the genetic text, which is the only repository of inherited structures of organisms, it necessarily follows that chance is responsible for any novelty, for any creation in the biosphere', the conclusion being that 'Chance only is the source of every novelty, of every creation in the biosphere. Sheer chance, chance only, absolute but blind freedom at the very roots of evolution: this central notion in modern biology is not anymore a hypothesis among other possible or at least conceivable ones. This hypothesis is the only conceivable one, since it is the only one which is compatible with observation and experience. And nothing allows us to imagine (or to hope) that our ideas on this point will need, or will be subject to, revision'. Finally, Monod considered the implications of his conclusions and proposed an 'ethics of knowledge', which will be discussed at the end of this paper.

The best comment on Monod's book was made by Eigen (3) 'The only thing lacking in molecular biology was its integration into a general understanding of Nature. So far, such an attempt has been undertaken only once, by Jacques Monod. This was a fascinating and ambitious attempt, in which Monod did not shrink from drawing philosophical conclusions. It culminated in an apotheosis of chance'.

The classical evolutionary theories

The role of chance in evolution was not, however, a new problem. Let us look at which way mutations were visualized by the classical evolutionists. The most famous sentence in *The Origin of Species* (1) was the following: 'I have called Natural Selection, or the Survival of the Fittest, this preservation of favorable individual differences and variations and the destruction of those which are injurious variations'. This statement looks extremely simple, but Crick (4) remarked that 'Natural Selection is the basic mechanism that makes biology different from all other sciences. Of course anyone can grasp the mechanism itself, though remarkably few people actually do so'. Indeed, Darwin's sentence seemed to indicate a dichotomy, and was widely interpreted that way. The sentence was, however, immediately followed by another one, which is only rarely quoted: 'Variations neither useful nor injurious would not be affected by natural selection and would be left either a fluctuating element … or would ultimately become

fixed'. This still is the best definition of neutral changes. In other words, Darwin distinguished not two but three kinds of changes or mutations (which he called 'variations'): advantageous, deleterious and neutral.

Advantageous changes will tend to expand in the progeny, because the carriers and their progeny will reproduce more abundantly than average (this is the positive or Darwinian selection). In contrast, deleterious changes will tend to disappear from the population, because the carriers and their progeny will reproduce less abundantly (this is the negative or purifying selection). Finally, neutral changes may be fixed in the population (like advantageous changes) or disappear (like deleterious changes).

The idea of neutral changes was later obliterated by the neo-darwinians, the selectionists Fisher (5) and Haldane (6), only to be resurrected, later, by Kimura (7, 8) in his *mutation-random drift theory*. According to this neutral theory 'the main cause of evolutionary change at the molecular level – change in the genetic material itself – is random fixation of selectively neutral or nearly neutral mutants'; therefore, 'increases and decreases in the mutant frequencies are due mainly to chance'. As a logical consequence, this theory eventually replaced the *survival of the fittest* with the *survival of the luckiest* (9). Along the same line, King and Jukes (10) claimed in their *non-darwinian evolution* that 'most evolutionary changes in proteins may be due to neutral mutations and genetic drift' (the random changes in gene frequencies in a population). A significantly different position was taken by Ohta (11, 12) who proposed her *nearly neutral theory* according to which 'a substantial fraction of changes are caused by random fixation of nearly neutral changes, namely changes that are intermediates between neutral and advantageous, as well as between neutral and deleterious classes'. Fig. 1 (see p. 601) summarizes the points just mentioned.

It is now of interest to look at the experimental approaches used to develop the classical theories on evolution because of the tight links that exist between approaches, results and conclusions. Natural selection acts on the phenotype, namely the detectable characters (traits, features, properties) of living organisms. It is, therefore, understandable that the first approach to the study of evolution was based on morphological traits, a classical case being that of the beaks of the Galapagos finches, which show adaptations to different kinds of food, from hard seeds to soft vegetal tissues. After the rediscovery of Mendel's laws, the neo-darwinians relied on genetic characters. Only later a molecular approach was developed on the basis of the early protein and gene sequences, and this led to the neutral theory of Kimura. Indeed, the view that amino acids change linearly with

time in proteins (the molecular clock of Zuckerkandl and Pauling, 13), provided the very first hint in that direction.

The organization of the eukaryotic genome

A totally different approach moving from the molecular level of a few proteins and genes to the genome level was the one I started in 1959 by degrading DNA from mammals and birds with a DNase (14), and by fractionating DNA on hydroxyapatite columns (15). These experiments (probably the first ones in genomics) produced important results, such as the breakage of the genome into large fragments and the separation of double- from single-stranded DNA. Most of the following work was done, however, after our development (16) in 1968 of density gradient ultracentrifugation of DNA in the presence of sequence-specific DNA ligands (such as Ag+ ions), and our discovery in 1973 of the compositional heterogeneity of the bovine genome (17). Our *compositional approach* to the study of the genome, incidentally the only one that was possible at that time, was easily moved from the analysis of buoyant density profiles to nucleotide sequences as soon as these became available. The rationale of the compositional approach was that the base composition of the genome, the most elementary property of DNA, (i) is altered by mutations, insertions and deletions; (ii) influences DNA, RNA, protein and chromatin structure (see below); and (iii) can be precisely assessed on whole genomes and their domains. The conceptual simplicity of the approach is such that the results can be easily understood.

The compositional approach led to three major discoveries: (i) the vertebrate genomes (the only ones discussed here) are *mosaics of isochores* (18, 19), megabase regions (1 Mb is one million base pairs; the human genome is 3200 Mb in size) of fairly homogeneous GC level (Fig. 2, see p. 602); GC is the molar ratio (the percentage of the molecules) of guanine and cytosine in DNA); (ii) isochores belong in a few families, characterized by different levels of GC, dinucleotides and trinucleotides, and define a *genome phenotype* (20), namely the *compositional landscape* of the genome (see Fig. 3, p. 603); the GC-rich, gene-rich and the GC-poor, gene-poor isochores define two *gene spaces*, the *genome core* and the *genome desert,* that are correlated with all the basic structural and functional properties of the genome, the main ones being chromatin compaction, DNA methylation, gene distribution on the one hand, gene expression, recombination, replication timing on the other (see Fig. 4, p. 604); (iii) a *genomic code* (20; not to be confused with the

genetic code) correlates the compositions a) of coding sequences with those of contiguous non-coding sequences (*i.e.*, of 1% of the genome with the remaining 99%), b) of the three codon positions among themselves, and c) of coding sequences with the hydrophobicity and the secondary structure of the encoded proteins.

These discoveries (summarized in a book; 21) led to our conclusion that the genome is an *integrated ensemble*, with little or no room left for *junk* (22) or *selfish DNA* (23, 24). This is a completely new vision of the vertebrate (and more generally of the eukaryotic) genome, which has far-reaching implications. Indeed, (i) there is no way to create a compositionally compartmentalized genome, the mosaic of isochores, by random point mutations (namely, single base-pair changes); (ii) again no random process can lead to a genome phenotype or compositional landscape that is correlated with all basic structural and functional properties of the genome, and lastly, (iii) no random evolutionary process can lead to the compositional correlations mentioned above. In other words, the discoveries just presented rule out the *bean-bag view* of the genome (to paraphrase Mayr, 25), namely of a genome in which genes are randomly distributed in the bulk of non-coding sequences, a genome that is only endowed with additive and not with cooperative properties (21).

Genome evolution and the neo-selectionist theory

The ground was now ready to investigate genome evolution. The simple comparison of our early data (26) on vertebrate genomes (that we recently confirmed on the basis of full genome sequences, 27-30); led us to the discovery of two modes of evolution: the *conservative mode* and the *transitional mode* (31). The *conservative mode* is exemplified by a comparison of the isochore patterns of the genomes of Primates and Carnivores (Fig. 5, see p. 605). At least 50% base pairs changed during the time, 100 million years, comprised between their common ancestor and these two mammalian orders that independently diverged from it. The expectation from the randomness of neutral changes was a partial or total disappearance of the isochore families that were present in the common ancestor. Moreover, since nucleotide substitutions in vertebrates (and other organisms) favor GC→AT over AT→GC changes, this 'AT-bias' should also lead to lower GC levels. Instead, a remarkable conservation of isochore families was found in terms of GC levels and relative amounts.

This led us in a straightforward way to the *neo-selectionist theory* (32). As shown in Fig. 6 (see p. 606), this theory postulates a series of steps: (i) first of all, among AT-biased changes a number will accumulate to form local clusters; (ii) the 'last' AT-biased changes in the clusters, the *critical changes* transform clustered point mutations into regional changes that trespass a lower GC threshold; and (iii) cause changes in chromatin structure that expand over long distances. Fig. 1 (see p. 601) shows that the neo-selectionist theory incorporates the features of the nearly neutral theory of Ohta, adding, as a novelty, the critical changes, namely the superdeleterious changes that convert the clustered AT-biased point mutations into regional changes. It should be stressed that regional changes may also be caused by large insertions and deletions. The main point, however, is that chromatin changes are deleterious in that they affect some expression of genes located within the altered regions or in their neighborhood and may lead to negative selection of the carriers and of their progeny.

Since fish, amphibian and many reptilian genomes do not show the presence of the very GC-rich isochores that characterize the genomes of warm-blooded vertebrates (see Fig. 5, p. 605), a *transitional mode* of evolution in which isochore families underwent changes, must have taken place (see Fig. 7, p. 607). Back in 1986 we proposed (20) that: 'The formation and maintenance of the GC-rich isochores of warm-blooded vertebrates is due to natural selection, the selective advantages being the increased thermodynamic stability of DNA, RNA and proteins (GC-rich codons encoding aminoacids that stabilize proteins). In other words, the environment can mould the genome through natural selection'. The transitional mode involved both negative and positive selection, as discussed elsewhere (32).

An explanation as to why changes essentially affected the gene-rich isochores, is that these isochores are located (in the interphase nucleus) in an open chromatin structure, whereas the gene poor isochores are in a closed chromatin structure (33). Then, only the genome core needs to be stabilized by GC increases, the genome desert being stabilized by its own compact chromatin. While body temperature certainly is the *primum movens* of the compositional transitions that took place at the emergence of mammals and birds, other factors such as oxygen, salinity, pH, CO_2, may play a role in the compositional transitions which were found among fishes (see Fig. 5, p. 605).

To sum up, the *neo-selectionist theory* (i) provides a solution to the neutralist/selectionist debate, since it reconciles the nearly neutralist view of point mutations with selection at the regional level; (ii) is an epigenomic theory, in that the compositional changes in DNA affect chromatin struc-

ture and, as a consequence, gene expression, so leading to negative selection of the carriers and their progeny; and (iii) is an extension of Darwin's theory; in fact, the neo-selectionist theory may be visualized as an ultra-darwinian theory since even neutral and nearly neutral changes are eventually controlled by natural selection over evolutionary time. Needless to say, the neo-selectionist theory brings us back from Kimura's survival of the luckiest to Darwin's survival of the fittest (incidentally, a matter of satisfaction for somebody working at the Stazione Zoologica).

As any good theory, the neo-selectionist theory also made predictions: (i) that genome phenotype differences should be found in populations; and (ii) that some of them may affect the genomic fitness and cause genomic (not genetic) diseases (a typical one being cancer). The first prediction was confirmed by comparing two individual genomes: Venter's genome differs from the reference human genome because of a number of insertions and deletions that accumulate in GC-rich isochores (34). These may generate genomic diseases by affecting chromatin structure and, as a consequence, the expression of genes located within or next to altered regions, so reducing the genomic fitness of the carriers, without necessarily affecting the primary structure of coding and regulatory sequences.

Conclusions

We should now go back to our initial questions and see the answers that we can provide today. First of all, a currently accepted view is that in all likelihood the origin of life was not so much the *single chance event* visualized by Monod, as a necessity under the prevailing conditions (35). This establishes *a primordial link* between the inanimate world from which life arose and the living organisms. These are connected to each other by their common descent, and, far from being completely separated from the inanimate environment, are moulded by it through natural selection. In fact, we have shown that the genome itself is moulded by physical agents like temperature, oxygen, salinity, pH, etc. through natural selection.

Our findings lead to a largely deterministic vision of evolution, which is in contrast with the fully stochastic vision of Monod. Chance still plays a role in evolution through (i) *environmental chance events*, such as meteorite impacts, volcanic eruptions; (ii) *random drift,* the random changes in gene frequencies in populations*;* and (iii) *neutral and nearly neutral changes*; as in the case of random drift, these changes are evident when recent, or looked at on a limited time scale, but they vanish over longer time spans,

because they are eliminated by natural selection. Obviously, we are very far from the overwhelming role of chance postulated by Monod.

As a consequence, we are also very far from Monod's view on the ethical implications. Given his premises, Monod claimed that *true knowledge ignores values* and invoked an *ethics of knowledge,* whose only value is the objective knowledge itself. In contrast, knowledge contains values: knowledge of common descent of all living organisms links us with them and dictates our respect and love for them; knowledge of the moulding of living organisms by the environment, trough natural selection, links all of them to the inanimate world from which they derived in the first place. The *'old alliance'* with Nature, proposed by the *'animistic conception'*, far from being *'a projection of our brain on the inanimate world'* (as suggested by Monod), is the age-old intuition of links now established by Science.

I would like to finish as I started, on a personal note. I had the good luck of being acquainted with Jacques Monod over many years until his premature death in 1976. My admiration for him led me to change the name of the Institut de Biologie Moleculaire that I was directing in Paris to Institut Jacques Monod, as well as to organize several meetings in his memory (see, for instance, ref. 36). I would like to stress that the contrasting vision presented here was built on the scene set up by *Le hasard et la necéssité*, I could say on the shoulders of Jacques Monod. It is a great pity that we cannot have his viewpoint on our conclusions. I dare say, however, that he would have accepted them, based as they are on new facts, which were not available or conceivable at the time his book was published. I also venture to guess that he would have liked them, since one can feel that the pessimistic conclusions of the book were imposed by its internal logics but not necessarily liked by its author.

REFERENCES

1. Darwin C. (1859) *On the Origin of Species, a Facsimile of the First Edition* (Harvard Univ Press, Cambridge, MA); reprinted (1964).
2. Monod J. (1970) *Le hasard et la nécessité* (Editions du Seuil, Paris).
3. Eigen M. (1992) *Steps towards Life* (Oxford University Press, Oxford).
4. Crick F. (1988) *What Mad Pursuit* (Basic Books, New York).
5. Fisher R.A. (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford).
6. Haldane J.B.S. (1932) *The Causes of Evolution* (Longmans, Green, and Co., London).

7. Kimura M. (1968) *Nature* 217:624-625.
8. Kimura M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge).
9. Kimura M. (1986) *Philos Trans R Soc Lond* (Biol) 312:343-354.
10. King J.L., Jukes T.H. (1969) *Science* 164:788-798.
11. Ohta T. (1973) *Nature* 246:96-98.
12. Ohta T. (2002) *Proc Nat Acad Sci USA* 99:16134-16137.
13. Zuckerkandl E. and Pauling L. (1962) *Molecular disease, evolution, and genetic heterogeneity* in Kasha M. and Pullman B., eds. *Horizons in Biochemistry* 189-225 (Academic Press, New York).
14. Bernardi G., Champagne M., Sadron C. (1960) *Nature* 188:228-229.
15. Bernardi G. (1965) *Nature* 206:779-783.
16. Corneo G., Ginelli E., Soave C., Bernardi G. (1968) *Biochemistry* 7:4373-4379.
17. Filipski J., Thiery J.P., Bernardi G. (1973) *J Mol Biol* 80:177-197.
18. Macaya G., Thiery J.P., Bernardi G. (1976) *J Mol Biol* 108:237-254.
19. Bernardi G., Olofsson B., Filipski J., Zerial M., Salinas J., Cuny G., Meunier-Rotival M., Rodier F. (1985) *Science* 228:953-958.
20. Bernardi G., Bernardi G. (1986) *J Mol Evol* 24:1-11.
21. Bernardi G. (2004) *Structural and Evolutionary Genomics, Natural Selection in Genome Evolution* (Elsevier, Amsterdam); reprinted (2005).
22. Ohno S. (1972) *Brookhaven Symp. Biol.* 23:366-370.
23. Orgel L.E. and Crick F. (1980) *Nature* 284:604-607.
24. Doolittle W.F. and Sapienza C. (1980) *Nature* 284:601-603.
25. Mayr E. (1976) See letter.
26. Thiery J.P., Macaya G., Bernardi G. (1976) *J Mol Biol* 108:219-235.
27. Costantini M., Clay O., Auletta F., Bernardi G. (2006) *Genome Res.* 16:536-541.
28. Costantini M., Auletta F., Bernardi G. (2007) *Genomics* 90:364-371.
29. Costantini M., Di Filippo M., Auletta F., Bernardi G. (2007) *Gene* 400:9-15.
30. Costantini M., Cammarano R., Bernardi G. (submitted).
31. Bernardi G. (1995) *Annu Rev Genet* 29:445-476.
32. Bernardi G. (2007) *PNAS* 104:8385-8390.
33. Saccone S., Federico C., Bernardi G. (2002) *Gene* 300:169-178.
34. Costantini M., Bernardi G. (submitted).
35. de Duve C. (1995) *Vital Dust* (Basic Books, New York).
36. Quagliariello E., Bernardi G., Ullmann A., eds. (1987) *From enzyme adaptation to natural philosophy: heritage from Jacques Monod* (Elsevier, Amsterdam).

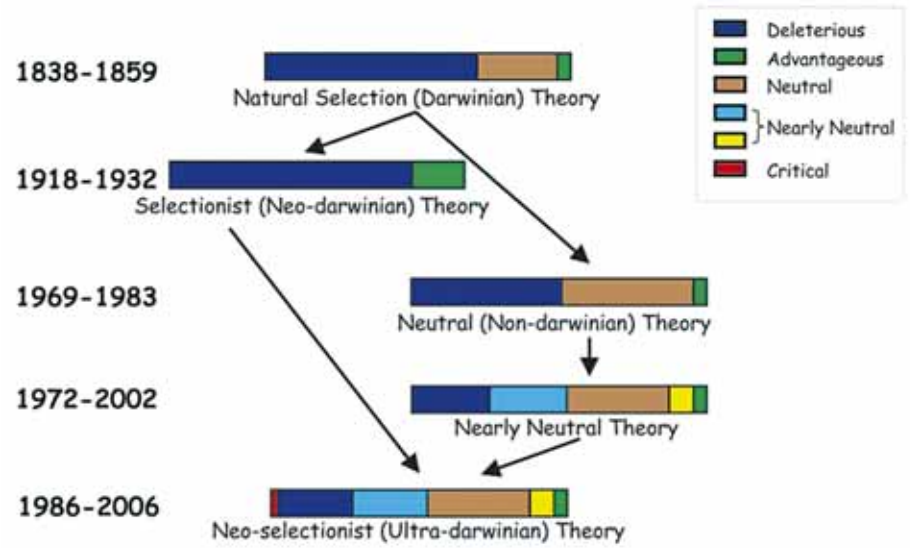

Figure 1. Darwin postulated the existence of deleterious, advantageous, and neutral changes. The neo-Darwinians (or selectionists) neglected neutral changes. These were reintroduced and amplified by Kimura (7, 8), who developed the neutral theory of evolution (a non-Darwinian evolution, according to King and Jukes,10). The nearly neutral theory was proposed by Ohta (11, 12) to include intermediates between neutral and advantageous, as well as between neutral and deleterious changes. In the neo-selectionist theory, nearly neutral theory is fully accepted and critical changes are responsible for the transition from point mutations to regional changes (from ref. 32).

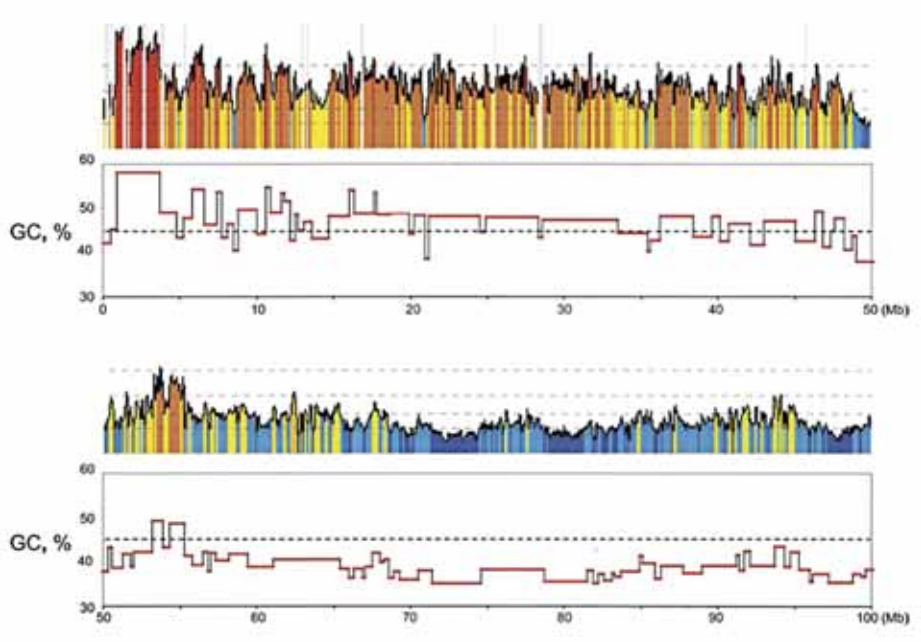

Figure 2. Overview of isochores on 100 Mb of chromosome 1 as a representative region of human chromosomes. The *top* frames represent GC profiles. Red to blue colours in the profiles correspond to decreasing GC levels. Horizontal red stretches in the *bottom* frames represent isochores (from ref. 27).

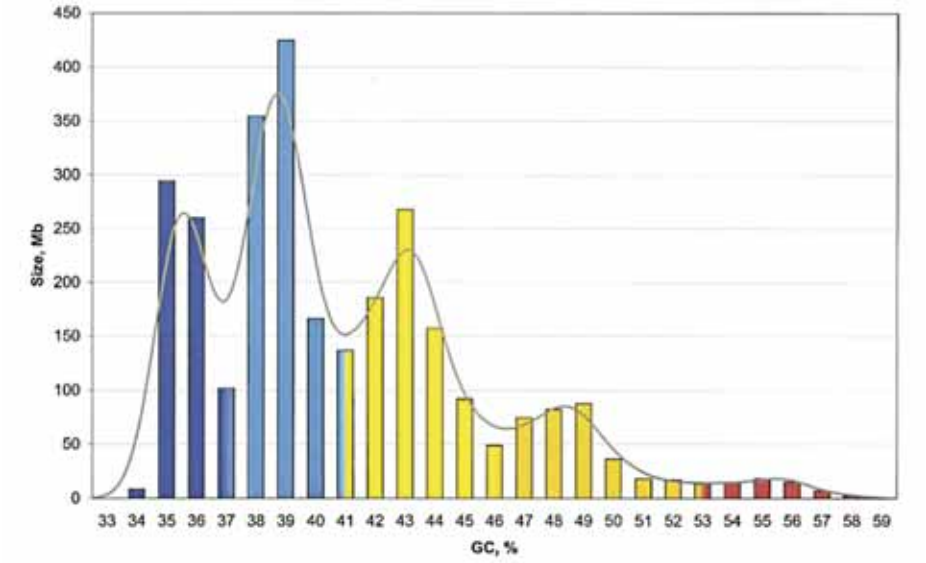

Figure 3. Distribution of human isochores according to GC levels (from ref. 27).

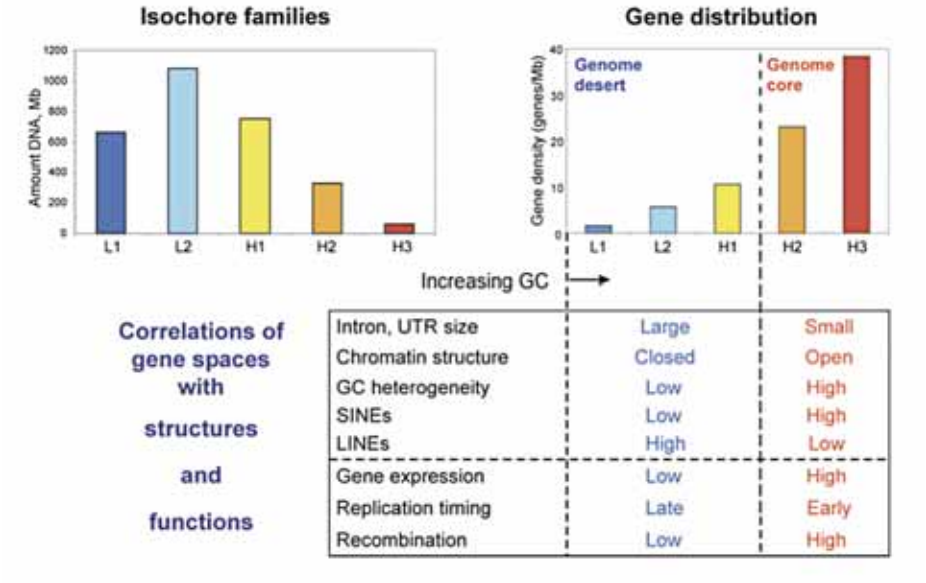

Figure 4. DNA and gene distribution in the isochore families of the human genome. The major structural and functional properties associated with each gene space are listed (in blue for the genome desert and in red for the genome core). SINEs are short interspersed sequences; LINEs, long interspersed sequences (from ref. 32).

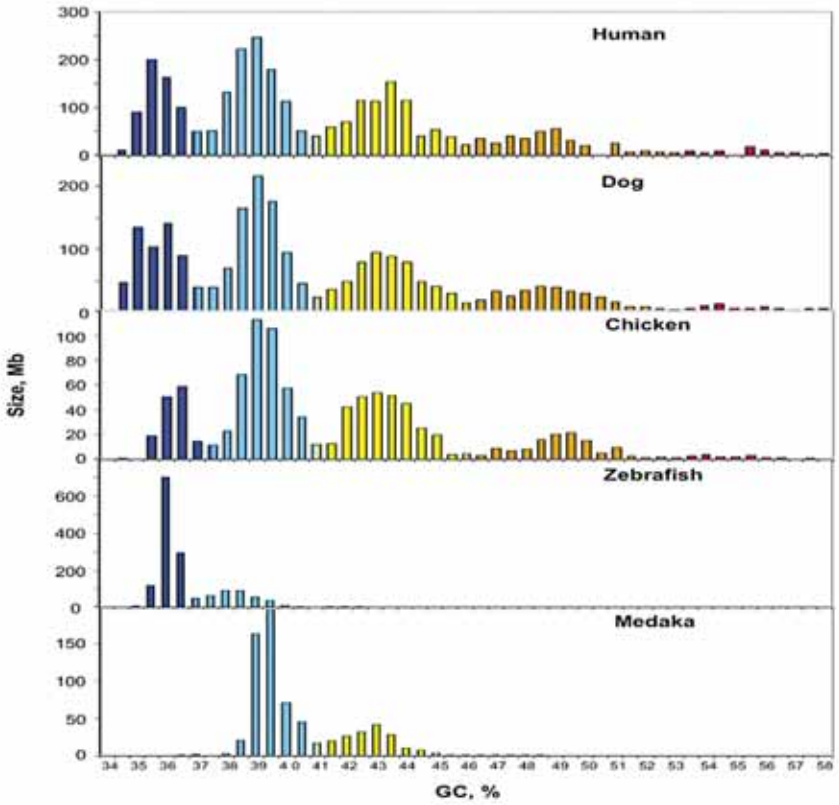

Figure 5. A comparison of the isochore families from several vertebrate genomes (from ref. 30).

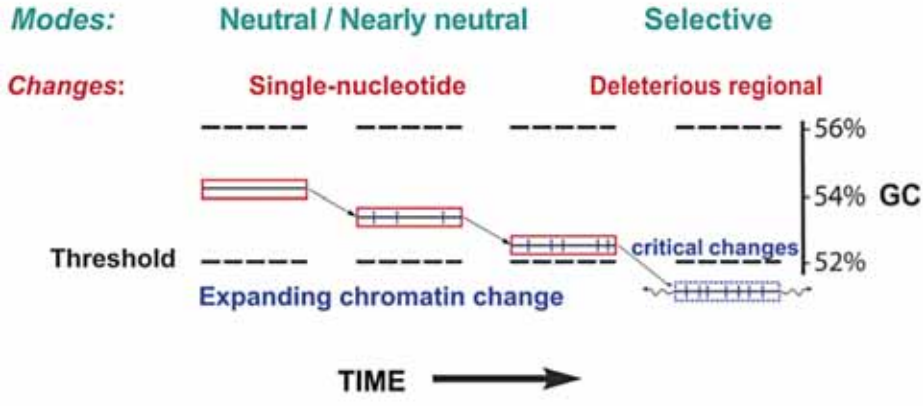

Figure 6. Time course of typical compositional changes of a GC-rich region from a warm-blooded vertebrate in the conservative mode of evolution. In an early phase, the average GC level of the region, initially visualized at its compositional optimum (arbitrarily set here at 54% GC), is decreasing because of the mutational AT bias (the vertical blue bars crossing the black DNA line in the chromatin red boxes represent the 'excess' GC → AT changes), but remains within a tolerated range (whose arbitrary thresholds are indicated by the thick horizontal broken lines). In a late phase, the average GC level trespasses the lower threshold (arbitrarily fixed here at 52% GC), because of the last changes, the critical changes. The corresponding chromatin (red boxes) then undergoes a structural change (broken blue box) that is deleterious for transcription (see text). Until then, the changes may be neutral or, more frequently, nearly neutral (from ref. 32).

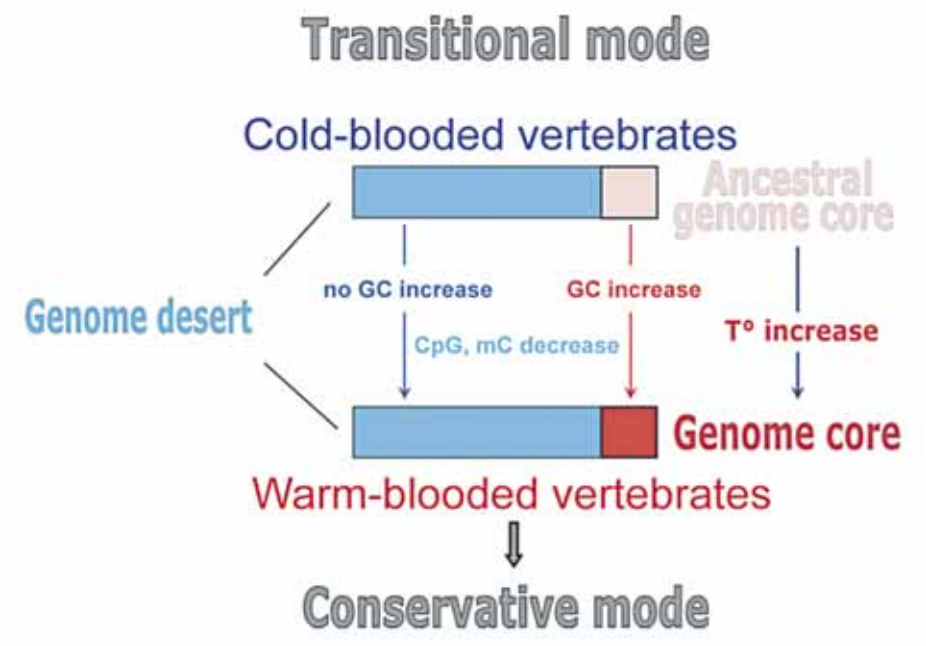

Figure 7. Scheme of the compositional evolution of vertebrate genomes. At the transition from cold- to warm-blooded vertebrates, the gene-dense, moderately GC-rich 'ancestral genome core' (pink box) became the gene-dense, GC-rich genome core (red box), but the GC-poor and gene-poor (blue box) genome desert did not undergo any major compositional change. This transitional (or shifting) mode, which was accompanied by an overall decrease of CpG doublets and methylcytosine, was followed by a conservative mode of genome evolution in which compositional patterns were maintained (from ref. 32).