ENCYCLOPEDIA OF THE HUMAN GENOME

2000

The evolutionary history of the Human Genome

Natural selection

Isochores

Neutral theory

Compositional genomics

Contents list:

Bernardi, Giorgio

Giorgio Bernardi

Stazione Zoologica Anton Dohrn

Villa Comunale, 80121 Naples Italy

The compositional evolution of the vertebrate genome, from fishes to human, is of very general interest because of its bearing on the role of natural selection in the evolutionary process.

1. Introduction

Compositional genomics is an approach to the problem of the organisation of eukaryotic genomes. Initially, this approach consisted in analysing the base composition of complex eukaryotic genomes (such as the nuclear genomes of vertebrates), as fractionated by density gradient centrifugation in the presence of sequence-specific ligands. Basically, the fractionation takes place because of the differential binding of the sequence-specific ligands to DNA fragments derived from different isochores. This approach revealed that vertebrate genomes are mosaics of very long stretches of DNA with essentially uniform base-composition (as judged at a size level of several Kb, kilobases). Isochores belong to a small number of families characterised by different base compositions (which are usually expressed as GC, the percentage of guanine+cytosine in any given isochore). This compositional approach, which could be immediately extended to nucleotide sequences, as soon as they became available, led to new insights into the organisation of the eukaryotic genome (a subject presented in another article). When comparative studies were made on the organisation of the nuclear genomes of vertebrates, compositional differences were shown to have an evolutionary relevance.

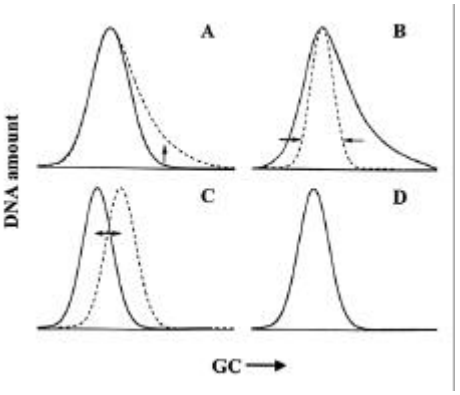
From a compositional viewpoint, vertebrate genomes evolve, as a general rule, according to a conservative mode: nucleotide substitutions do occur, but the base composition of sequences in general, and of coding sequences in particular, does not change. Under some circumstances, however, vertebrate genomes evolve according to a transitional, or shifting, mode: nucleotide substitutions do occur and the base composition of sequences changes. Figure 1 presents a scheme of the conservative mode of evolution and of the three kinds of compositional shifts that were observed.

This review will present first the intragenomic shifts, the "major shifts" and the "minor shift", and then the "whole-genome", or "horizontal", shifts. In each case, the conservative mode of evolution preceding and following the shifts will also be commented upon. Investigations on these evolutionary modes have shed new light on a central problem in molecular evolution, namely the role played by natural selection in modulating the mutational input.

2. The major shifts.

Two major, intragenomic, shifts took place in the genomes of the ancestors of present-day mammals and birds. These shifts were originally observed at the DNA level (see Fig. 1A). Indeed, DNAs from all warm-blooded vertebrates exhibit high compositional heterogeneities and strongly asymmetrical CsCl bands, whereas DNAs from cold-blooded vertebrates are generally characterised by low compositional heterogeneities and by only slightly asymmetrical CsCl bands. Both differences, in heterogeneity and asymmetry, are due to the presence in the genomes of warm-blooded vertebrates of a small percentage (about 15%) of GC-rich DNA molecules that are absent, or scarcely represented, in the genomes of most cold-blooded vertebrates. Since the genomes of mammals and birds derive from those of two separate lines of ancestral reptiles, these findings indicated that two major

compositional changes had independently occurred in the distinct ancestral lines leading to warm-blooded vertebrates.



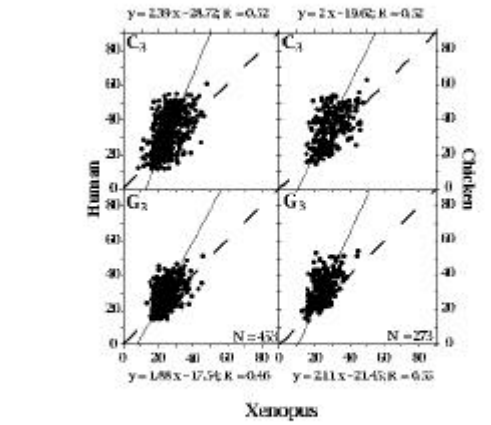
The two major shifts were then detected at the coding sequence level. Indeed, coding sequences from cold-blooded vertebrates are compositionally relatively homogeneous and generally characterised by low GC levels, whereas coding sequences from warm-blooded vertebrates are compositionally much more heterogeneous and reach very high GC levels, up to 100% GC in the third codon positions of genes.

The best evidence for the major shifts was obtained by comparing the nucleotides in third codon positions of orthologous genes from human and *Xenopus*. When the orthologous genes of human/*Xenopus* were investigated in their G_3 and C_3 values (the G and C levels in the third codon positions; see Fig. 2), points were scattered about the diagonal in the low GC range, showing no directional changes between the two species, whereas human gene values were increasingly higher, on the average, compared to the corresponding *Xenopus* gene values, as GC_3 values increased. This resulted in regression lines having slopes that were close to 2. Very similar results were obtained when comparing orthologous genes from chicken and *Xenopus*.

The major shifts were followed by compositionally conservative modes of evolution, as indicated by G_3 and C_3 plots concerning genes from human and calf (Fig. 3). The latter is a good representative of mammals (other than human) that share the "general mammalian pattern", namely the most widespread mammalian pattern as defined by the CsCl profiles and by GC_3 plots of orthologous sequences. In Fig. 3, the regression lines of the human/calf plots pass through the origin and are characterised by unity slopes and very high correlation coefficients, 0.88-0.89. In other words, G_3 and C_3 values of orthologous genes from human and calf are very close to each other over the entire GC_3 range. A very similar result was obtained when comparing orthologous genes among different avian orders. Since mammalian orders diverged some 100 Myrs ago from a common ancestor according to a star-like phylogeny (i.e., they evolved independently of each other), one should conclude that the many nucleotide changes that occurred during this long time interval did not lead to any change in the compositional patterns of either DNA or coding sequences.

The compositional transitions just described can be summarized as follows. (i) These transitions essentially concerned the genes and the intergenic sequences that are located in the GC-richest isochores H2 and H3 of the genomes of warm-blooded vertebrates (these isochores correspond to the gene-dense regions of the vertebrate genome, the "genome core"); in contrast, they did not concern the genes and the intergenic sequences from the gene-poor "empty space" of the genome. (ii) They

occurred (and were similar) in the independent ancestral lines of mammals and birds (see Fig. 2), but in almost no cold-blooded vertebrate (see Section 2.2, below). (iii) They stopped before the appearance of present-day mammals and birds, as indicated by the very similar patterns found in different mammalian and avian orders, respectively (see the human/calf comparison of Fig. 3, for an example).



These findings indicate that the compositional transitions affecting the "genome core" of the ancestors of mammals and birds, had reached an equilibrium at least as early as at the times of appearance of present-day mammals and birds, and that, from then on, the compositional patterns resulting from the cold- to warm-blooded transitions were maintained until present (except for some small "whole-genome" shifts; see section 4). This conservation is remarkable, if one considers that about 55% of the human genes (the genes of the "genome core") had undergone a compositional transition in which extremely high G_3 and C_3 values were reached, and maintained over 100 million years, the time of divergence of mammals.

An additional point to be made here is that the compositional transition involved more than the compositional changes just described. Indeed, (i) DNA methylation and CpG doublet concentration are lower by a factor of two in mammals, or birds compared to fishes, or amphibians (these changes apparently occurred already between amphibians and reptiles, which seem to exhibit the lower methylation pattern of warm-blooded vertebrates); (ii) CpG islands (regulatory sequences about 1 kb in size, located 5' of GC-rich genes, and characterized by high levels of GC and unmethylated CpG doublets) were formed; (iii) changes took place in the AUG initiator context of GC-rich human genes relative to genes from cold-blooded vertebrates; (iv) T (or $H3^+$) bands appeared in metaphase chromosomes; (v) karyotype changes and speciation increased. Interestingly, changes (ii) and (iii) indicate that the major shifts also altered regulatory sequences.

2.1 Explanations for the major shifts: natural selection.

The first explanation proposed for the "major shifts" and for the maintenance of the new patterns was that they were due to natural selection, namely "the preservation of favourable variations and the rejection of injurious variations" (as defined by Darwin); in addition, it was speculated that the selective advantages for the changes were the increased thermal stability of proteins, RNA and DNA.

As far as DNA stability is concerned, it should be recalled that the GC-richest and gene-richest isochores have their highest concentration in a set of R(everse) chromosomal bands that largely coincide with the T(elomeric) bands previously identified as particularly resistant to thermal denaturation. As for RNA, abundant evidence indicates that high GC levels stabilize RNA structures.

As far as protein stability is concerned, the supporting evidence is that the increase in GC_3 of coding sequences is paralleled by an increase in GC_2 and GC_1, by an increase in quartet codons, a decrease in duet codons, an increase in certain amino acids (alanine, valine), a decrease in other amino acids (lysine) and an overall increase in the hydrophobicity of the encoded proteins. Not surprisingly, therefore, the hydrophobicity of the proteins encoded by the orthologous genes that underwent a GC change was found to be higher in human than in *Xenopus*, indicating that the compositional genome transition between cold- and warm-blooded vertebrates was accompanied by changes in the structural features of the encoded proteins that stabilise them.

Another finding in favour of the functional significance of the changes is that regions of coding sequences corresponding to ? helix, ? sheet and coil structures in the proteins are characterised by different levels of individual nucleotides in all codon positions and by different substitution rates. The first result suggests that changes in the nucleotide composition lead to changes in the secondary structure of proteins. The second result indicates that synonymous and non-synonymous rates are correlated with the secondary structure of proteins, higher rates being found in regions corresponding to coil and a helix compared to regions corresponding to ß sheet.

2.2 Alternative explanations for the major shifts: regional mutational biases.

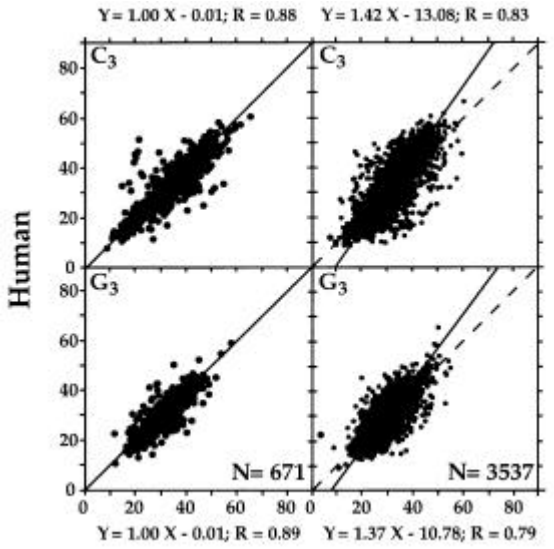
The fact that some genome regions of the ancestors of present-day mammals and birds underwent the transition while others did not might also be explained by "regional" mutational biases (other explanations have been discussed and ruled out). It was pointed out, however, that since mutational biases are the result of mutations in the replication machinery, and since there is just a single replication machinery in eukaryotic cells, additional hypotheses would be needed to explain why the changes were regional, instead of concerning the totality of the genome. Such hypotheses have been proposed. Indeed, different chromatin structures were postulated to be responsible for the "regional" mutational biases. Alternatively, repair could be more efficient in some regions, for instance in transcribed sequences.

This line of reasoning does not answer, however, several questions which are answered by the explanation based on natural selection: (i) why "regional" changes never appeared in cold-blooded vertebrates (with the possible exception of some

reptiles and fishes which were submitted to high temperature environments), which also are characterised by gene-poor and gene-rich regions and by biphasic (early-late) replication timings; (ii) why changes correlate intragenically with exon/intron structures, exons being systematically GC-richer than introns, and with different secondary structures of proteins; (iii) why changes were consistently in the direction of an increased thermodynamic stability; (iv) why similar compositional changes that could also be related to thermal stability were found in organisms phylogenetically distant from vertebrates (like *Gramineae*); (v) why G and R bands, replication banding, and synonymous rate regions (namely regions characterised by similar synonymous substitution rates), do not coincide with isochores; (vi) why an increase in substitution rates (such as that exhibited by murids; see the following section) did not lead to increased GC_3 levels of the genes that underwent the major transition; (vii) why the mutational bias GC? AT that was found by analyzing mutational substitution matrices of both GC-poor and GC-rich human genes did not lead to a GC-poor compositional pattern in the human genome; and (viii) why very similar C_3, G_3, and codon frequencies of orthologous genes were found in human and chicken. Incidentally, the finding that the majority of genes underwent the major shift, whereas a minority did not, may be related to the stringency of structural and functional requirements of the encoded proteins.

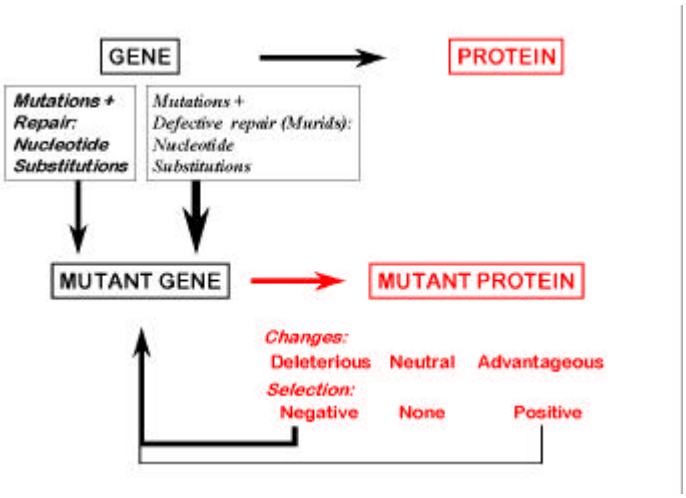
3. The minor shift

The GC-richest and GC-poorest genes of the murids are less GC-rich and less GC-poor, respectively, than their orthologs from other mammals, whereas the genes with intermediate GC values remain unchanged (see Fig. 3). Now, it is known that murids exhibit (i) rates of synonymous substitutions that are higher, by a factor of 5 to 10, relative to human coding sequences; (ii) a defective repair system; and (iii) a compositional pattern that is derived from the general mammalian pattern. Under these circumstances, it is interesting to observe that the higher mutational input "randomizes" the composition of synonymous positions by reducing the difference between the extreme GC values found in the general mammalian pattern. Indeed, under a mutational bias model one would expect instead an increase of the differences between extreme values (see end of preceding section).



It is also remarkable that this also leads to the "randomization" of non-coding sequences, causing for instance, an "erosion" of CpG islands. In the case of the ? globin, for example, the human gene is very GC-rich and is associated with a CpG island, whereas the mouse gene is less GC-rich and has lost the island.

As already mentioned, the very similar compositional patterns of genomes from mammalian orders that were separated for 100 million years indicates that the "general" mammalian pattern was already present in the common ancestor of present-day mammals and was the result of an equilibrium between the mutational input and negative selection. When the mutational input was increased, as in the case of murids, a new equilibrium was reached (see Fig. 4), as witnessed by the very similar compositional pattern of different murids, and a conservative mode of evolution took over. This new equilibrium already existed in the common ancestor of murids and was, therefore, reached 30 million years ago or earlier.



4. The whole-genome shift

Whole-genome (or horizontal) shifts were observed in the genomes of cold-blooded vertebrates, mainly of fishes (possibly only because fishes were studied more extensively compared to amphibians and reptiles). They consist in shifts of the entire distribution of DNA molecules towards higher or lower GC levels (see Fig. 1C). This suggests that the changes under discussion are due to "mutator mutations", namely to mutations in the sequences coding for protein sub-units of the replication machinery, that lead to mutational biases; as already noted, "regional", intragenomic changes have not been observed in fish genomes (possible exceptions being under study; see also below).

Whole-genome shifts are characterised by three important features: (i) they show different GC ranges in fish species belonging to different families or orders; (ii) they exhibit no increase with the geological time of appearance of the families or orders under consideration; and (iii) they are much more frequent and much larger compared to the horizontal shifts found in mammalian and avian orders.

The first point indicates an essential "randomness" in the primary events responsible for the changes and the possibility of variations in their directionality (AT? GC; GC? AT). The second one indicates, in addition, that there is no cumulative effect, in

that the genomes of species from ancient orders do not show more spreading of their average composition compared to those of recent ones. The third point suggests that the homeostasis of warm-blooded vertebrates leads to more stable compositional patterns of the genome; in other words, negative selection appears to restrict the range of tolerated mutations in warm-blooded vertebrates, that are more controlled in their environment, compared to cold-blooded vertebrates, which use more diverse ecological niches (see below).

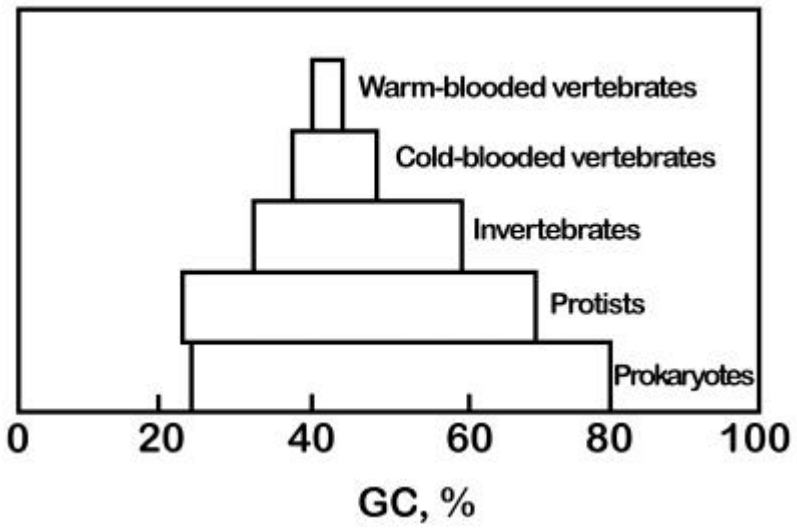
5. The horizontal shift of bacterial genomes

The whole-genome shifts of cold-blooded vertebrates are important not only in themselves, but also because they shed light on the mechanism(s) and the cause(s) of the wide compositional spectrum of bacterial genomes. The explanation originally proposed to account for the different base compositions of prokaryotic genomes is that they may shift because of directional mutations due to biases in replication enzymes. It has been argued, however, that while these mutational biases are acceptable as the mechanism of the compositional changes, they are not necessarily their cause. Indeed, "mutator mutations" lead not only to strong biases in base substitutions, but also to highly increased mutation rates. In the laboratory, such biases have only been detected in mutational hot spots. In spite of this, it is conceivable that in nature overall changes in the base composition of bacterial genomes can be achieved through mutational biases. The question remains open, however, whether the resulting compositional changes are only determined by the vagaries of random mutations in the genes encoding the protein sub-units of the replication machinery, or are under the control of natural selection.

Two major reasons point towards the second explanation. One reason is that changes in nucleotide composition correspond to changes in amino acids; in turn, these correspond to changes in the hydrophobicity and in the secondary structure of proteins (see Section 2). If changes only depended upon mutational bias, then one should accept the untenable viewpoint that the corresponding functional changes are selectively irrelevant. A second reason is that GC changes cover a much wider range in the genomes of prokaryotes and unicellular eukaryotes than in those of invertebrates, cold-blooded vertebrates and warm-blooded vertebrates, these three groups showing a progressively narrower range (Fig. 5). Since the potentially relevant mutations in the replication machinery responsible for different mutational biases presumably are comparable in all classes of organisms, and since the spread of GC levels is so different, the only explanation for the different ranges of Fig. 5 is that there is a different degree of selection on the mutations in the replication machinery of these different classes of organisms. Moreover, this different degree of selection seems to be correlated with the variety of intra- and extracellular environments of the organisms under consideration.

Two points are relevant here. (i) Since the horizontal shifts are presumably due to some specific mutations in the genes encoding protein sub-units of the replication machinery, whereas the major and the minor shifts are due to mutations occurring in other genes and other genome regions, these two kinds of mutations may be superimposed on each other; indeed, whole-genome shifts have been observed in mammalian genomes which had undergone either major or minor shifts. (ii) The effect of selection on the mutations responsible for the intragenomic major and minor

shifts has already been discussed; in the case of the whole-genome shifts, selection appear to act specifically on the mutations in the genes encoding the replication machinery.



6. Conclusions

Of the three transitional or shifting modes exhibited by vertebrate genomes, two are intragenomic and one concerns the whole genome. The two intragenomic major shifts took place in the ancestors of present-day mammals and birds and affected the gene-dense 15% of their genomes, which became GC-richer compared to the remaining 85%. These changes occurred separately in the independent ancestral lines of mammals and birds. These intragenomic changes which largely occurred in the same set of genes (the genes present in the genome core of present-day mammals and birds), reached an equilibrium before the time of appearance of present-day warm-blooded vertebrates, but did not take place in cold-blooded vertebrates (see, however, below). Interestingly, compositional changes also affected regulatory sequences and, therefore, in all likelihood, gene expression. This is a whole area that deserves further work.

The explanation originally proposed for the major shifts, namely natural selection (the advantages being associated with the increased thermal stability of proteins, RNA and DNA), is supported by the increased hydrophobicity of the proteins encoded by GC-rich genes, as well as by the different frequencies of individual nucleotides (and by the different substitution rates) in sequence regions coding for ? helix, ? sheet and coil, respectively. On the other hand, increases in GC lead to increased stability of DNA and RNA. Finally, transitions similar to those found in orthologous genes from *Xenopus* and human were also found when comparing sequences from dicots and *Gramineae*. In each case, the genome that underwent the transition experienced over a long period of time a higher temperature compared to the one that did not. It is not impossible that the genomes of some reptiles may have undergone similar compositional changes. In fact, if this point could be firmly established, it would provide a strong additional argument for the case made here. In contrast, the main alternative explanation for the major shift, namely regional mutational biases,

encounters an impossibly large number of problems, which have been detailed in a previous section.

The second intragenomic shift, the minor shift, took place in the common ancestor of murids, and "randomized" the extreme regions of the compositional distribution of DNA and of coding sequences. In contrast to expectations based on the assumption of a mutational bias, the increased mutation rate did not lead to an increased difference between the two ends of the compositional distribution of coding sequences, but to the opposite result. Interestingly, a new equilibrium between mutational input and negative selection was reached, compared to that shown by mammals exhibiting the "general pattern".

The whole-genome shifts shown by cold-blooded vertebrates are much more evident than the intragenomic shifts observed in warm-blooded vertebrates. Negative selection of a number of these mutations in the genes encoding the replication machinery can account for the very different ranges exhibited by warm-blooded vertebrates, cold-blooded vertebrates, invertebrates, unicellular eukaryotes and bacteria.

As far as the conservative mode is concerned (Fig. 1D), the available data indicate that the maintenance of the genome patterns resulting from the major and minor intragenomic compositional shifts is due to natural selection at the nucleotide level. Moreover the discovery of a GC? AT mutational bias in the human genome reinforces the need for natural selection to maintain the mammalian compositional pattern of mammals. This need seems to be less strong for the GC-poor isochores.

Finally, the conservative compositional patterns following horizontal shifts of bacteria and unicellular eukaryotes might just be maintained by the mutational input and its bias. In this case, natural selection would still control the "mutator mutations" responsible for the bias. It is obvious, however, that natural selection does more than that. Indeed, if mutational bias were the only factor responsible for a given compositional pattern in a bacterial genome, it would be difficult to understand, especially in the bacterial genomes characterized by extreme compositions, how deleterious mutations could be avoided. A role played by natural selection at the nucleotide level of genes in general is, therefore, inescapable even in the case of "horizontal shifts".

7. Some general remarks

(i) Natural selection, as discussed here in connection with the major shifts, is essentially negative (or stabilising) selection. This can be easily understood in the case of the maintenance of GC-rich isochores, because nucleotide substitutions leading to decreasing GC are counter-selected. It may also apply, however, to the transitional or shifting mode. Indeed, it is conceivable that body temperature increased progressively over large time spans in the ancestors of warm-blooded vertebrates. This shifting threshold of optimal base composition might have led to the counter-selection of increasingly higher numbers of GC? AT substitutions, a process causing a progressive net increase of the GC level and an adaptation.

Negative selection at the nucleotide level also applies to the short non-coding sequences of the genome core. This can also be understood, since some of the

intergenic sequences have a well-defined regulatory role (as is the case of CpG islands and untranslated sequences), and since other non-coding sequences may just influence, by their primary structure, not only chromatin structure and nucleosome density, but also the expression of neighbouring genes. This is indicated by results showing that the stability and the transcription of proviral sequences is optimal only in compositionally matching chromosomal environments, namely in the environments in which the host genes having the composition of integrated viral sequences are located.

(ii) As far as the selective advantages and disadvantages associated with compositional genome changes are concerned, it is impossible to identify them except in the most general terms, because compositional genome changes are the result of many superimposed factors. The situation may be different, however, when comparing closely related organisms because of the similarity of most of these superimposed factors. For example, in the case of a small taxon, such as that of vertebrates, the fact that the major shift took place at the transition between cold- and warm-blooded vertebrates and the fact that the changes led to an increased stability of the proteins, DNA and RNA, suggested that body temperature was the major selective force for GC increases in vertebrate genomes. Indeed, while cold-blooded vertebrates may cope with higher environmental temperatures over relatively short times using physiological responses, long-term high temperatures, such as those experienced at the transition from cold- to warm-blooded vertebrates, may lead to regulatory adaptations (such as changes in promotor structure) and, eventually, to genomic adaptations (such as the major shifts).

Obviously, there is no reason to think that the wide compositional spectra of prokaryotes and unicellular eukaryotes and even the narrower ones of vertebrate genomes are also related to temperature, since they may instead be due to a number of other overlapping factors that are difficult, if not impossible, to assess.

(iii) It should be stressed that the comparative compositional approach is substantially different from the approaches that are most often used to investigate orthologous mammalian sequences. Indeed, first of all, it deals with genomes which are in a state of compositional equilibrium. The compositional transitions under consideration here occurred in the remote past and were followed by a conservative mode of compositional evolution. Second, it looks at the genome forest and not at couples or small sets of similar trees. By contrast, current molecular evolutionary work is centred on differences, usually in nucleotide substitution rates, as exhibited by orthologous genes that have a common compositional background. As such, they are bound to miss the very existence of the selection phenomena discussed here. Indeed, nucleotide substitutions in orthologous GC-rich mammalian genes, whose overall base composition is the result of natural selection, may well appear to be neutral.

Figures Legends

Fig. 1. The "major" (**A**), the "minor" (**B**) and the "whole-genome" or "horizontal" (**C**) compositional transitions (shifts) in the genomes of vertebrates. This scheme displays CsCl profiles and the changes undergone as a consequence of the compositional shift under consideration. CsCl profiles are good approximations of the compositional, or GC, distributions of DNA molecules. **A** refers to the transition between cold- and warm-blooded vertebrates; **B** to the transition between the general mammalian pattern and the murid pattern; **C** to the transitions among cold-blooded vertebrates. **D** refers to the conservative mode of evolution. (From Bernardi, *Gene*, in press, 2000).

Fig. 2. Correlation between G_3 and C_3 values of orthologous genes from human, or chicken, and *Xenopus*. The orthogonal regression lines are shown together with the diagonals (dashed lines). The equations of the regression lines and the correlation coefficients are indicated. N is the number of gene pairs explored. (From Cruveiller et al., *Gene*, in press, 2000).

Fig. 3. Correlation between G_3 and C_3 values of orthologous genes from human and calf or mouse. Other indications are as in Fig. 2. (From Bernardi, *Gene*, in press, 2000).

Fig. 4. Scheme of the equilibrium between mutational input (nucleotide substitutions) and negative selection. The equilibrium on the left is that of genes from mammals showing the general mammalian pattern, the equilibrium on the right is the case of genes from murids. (From Bernardi, *Gene*, in press, 2000).

Fig. 5. GC level ranges of DNAs from warm- and cold-blooded vertebrates, invertebrates, protists, and prokaryotes. (From Bernardi and Bernardi, 1990).

Further reading

Bernardi G (2000) The compositional evolution of vertebrate genomes. *Gene* (in press).

Bernardi G (2000) Isochores and the evolutionary genomics of vertebrates *Gene* **241**: 3-17.

Bernardi,G (1995) The human genome : organization and evolutionary history *Annu. Rev. Genet.* **29**: 445-476.

Bernardi G and Bernardi G (1990) Compositional transitions in the nuclear genomes of cold-blooded vertebrates *J. Mol. Evol.* **31:** 282-293.

Cruveiller S, D'Onofrio G and Bernardi G (2000) The compositional transition between the genomes of cold- and warm-blooded vertebrates: codon frequencies in orthologous genes *GENE* (in press).

Freese E (1962) On the evolution of base composition of DNA *J. Theor. Biol.* **3:** 82-101.

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition *Proc. Natl. Acad. Sci. USA* **48:** 582-592.

Sueoka, N (1988) Directional mutation pressure and neutral molecular evolution *Proc. Natl. Acad. Sci. USA* **85:** 2653-2657.

Graur, D and Li WH (2000) *Fundamentals of Molecular Evolution* (2nd edition) Sinauer Associates, Inc, publishers, Sunderland, Massachusetts.

Carroll RL (1997) *Patterns and Processes of Vertebrate Evolution* Cambridge University Press, Cambridge, UK.